



Universidade de Aveiro Departamento de Comunicação e Arte



Universidade do Porto Faculdade de Letras

2017

José Manuel  
Rodrigues Nunes

Visualização de Interação em Cenários de  
Comunicação Humano-Computador





Universidade de Aveiro Departamento de Comunicação e Arte



Universidade do Porto Faculdade de Letras

2017

José Manuel  
Rodrigues Nunes

Visualização de Interação em Cenários de  
Comunicação Humano-Computador

estudo aplicado a soluções de comunicação e fluxo de informação interna  
nas organizações, sustentadas em contextos de mediação tecnológica.





Universidade de Aveiro Departamento de Comunicação e Arte



Universidade do Porto Faculdade de Letras

2017

José Manuel  
Rodrigues Nunes

## Visualização de Interação em Cenários de Comunicação Humano-Computador

dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Doutor em Informação e Comunicação em Plataformas Digitais, realizada sob a orientação científica do Prof. Doutor Óscar Mealha, Professor Associado com Agregação do Departamento de Comunicação e Arte e da Prof<sup>a</sup>. Doutora Beatriz Sousa Santos, Professora Associada com agregação do Departamento de Eletrónica, Telecomunicações e Informática, ambos da Universidade de Aveiro.



dedicatória

Aos meus pais, por toda a compreensão, apoio incondicional e infindável paciência.





o júri

presidente

**Prof. Doutor Eduardo Anselmo Ferreira da Silva**  
Professor Catedrático da Universidade de Aveiro

vogais

**Prof. Doutor João Manuel Gonçalves Duarte Cunha**  
Professor Catedrático convidado aposentado da Faculdade de Ciências da Universidade de Lisboa

**Prof<sup>a</sup>. Doutora Cândida Fernanda Antunes Ribeiro**  
Professora Catedrática da Faculdade de Letras da Universidade do Porto

**Prof. Doutor Fernando Manuel dos Santos Ramos**  
Professor Catedrático da Universidade de Aveiro

**Prof. Doutor Luís Manuel Borges Gouveia**  
Professor Catedrático da Faculdade de Ciências e Tecnologia da Universidade Fernando Pessoa

**Prof. Doutor Óscar Emanuel Chaves Mealha**  
Professor Associado com agregação da Universidade de Aveiro (orientador)

**Prof<sup>a</sup>. Doutora Graça Maria Bordalo Rocha Simões**  
Professora Auxiliar da Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa



## agradecimentos

Gostaria de agradecer a todos os que, de alguma forma, contribuíram para tornar possível a realização desta tese:

- Aos meus orientadores, Prof. Doutor Óscar Mealha e Prof<sup>a</sup>. Doutora Beatriz Sousa Santos, pela amizade, conselhos, orientação, constante incitamento e por sempre terem acreditado neste projeto.
- Ao Eng. Florin Zamfir, pela sua amizade e colaboração em todo o trabalho desta tese, em especial no desenvolvimento do protótipo que culminou na sua dissertação de mestrado.
- Ao Prof. Doutor Carlos Ferreira, pela sua ajuda e contribuição, especialmente no processamento estatístico.
- Aos sTIC pelas preciosas opiniões e colaboração na avaliação do protótipo, bem como pelo material de trabalho (ficheiros de log) fornecido.
- Ao DeCA e IEETA por me acolherem e pelos recursos disponibilizados durante a realização do trabalho desta tese.
- À minha família e à Elsa, pela sua paciência e que, de forma incondicional, sempre me apoiaram e incentivaram a prosseguir, mesmo nos momentos de maior dúvida e incerteza.
- Aos meus amigos, em particular à Cristina e também ao Carlos, André, Geninha, Constança, Sérgio, Rossana, Elsa, Mané, Tim e Gena, por todo o apoio e incentivo.
- Por fim, para todos os meus colegas e alunos que de alguma forma contribuíram para a realização deste trabalho, o meu sincero obrigado.



palavras-chave

avaliação de usabilidade, visualização da interação humano-computador, análise de logs web, visualização de informação.

resumo

Os contextos infocomunicacionais suportados em mediação tecnológica (no contexto geral da comunicação mediada por computador) estão a tornar-se cada vez mais presentes nas atividades do dia a dia de um número crescente de indivíduos e instituições. Especificamente, as tecnologias e serviços da internet/web têm uma presença marcante nas instituições, um pouco por todo lado. Os sites internos das instituições (vulgo intranets) são desenvolvidos de acordo com as estratégias de comunicação internas, refletindo os fluxos internos de informação e respetivos serviços de comunicação que lhes estão associados. Um problema emergente tem a ver com a gestão destas plataformas infocomunicacionais internas (intranets) e da relação com os seus interlocutores externos (extranets), ambas em crescimento constante. Os especialistas em comunicação organizacional sentem a falta de ferramentas que lhes permita analisar (padrões de atividade e comportamento) e perceber o que realmente se está a passar dentro da instituição. De facto, estes instrumentos tendem a basear-se em métricas de cariz técnico clássico, em muitos casos, para afinações técnicas e não para gerir a comunicação organizacional ou para a análise e gestão da informação.

Esta tese centra-se na conceção e avaliação destas ferramentas de análise e diagnóstico para que possam contribuir para um desenvolvimento destas infraestruturas sofisticadas e, conseqüentemente, melhorar a eficiência dos processos infocomunicacionais que lhes são intrínsecos. Um dos problemas está na identificação dos desajustes utilizador-sistema ao nível da interação humano-computador, que têm que ser completamente identificados, e os problemas prontamente apresentados à equipa que procede ao desenho e desenvolvimento das plataformas infocomunicacionais. O sistema tem que servir a organização e manter com eficácia os seus padrões de fluxo de informação e respetivas tarefas.

O conceito sistémico de feedback apresenta-se aqui como fundamental e necessariamente eficiente para a rigorosa identificação de problemas na plataforma infocomunicacional de uma determinada instituição. As propostas apresentadas demonstram capacidade de diagnosticar problemas estruturais e de conteúdo a dois níveis: ao nível da própria interface dos serviços infocomunicacionais e ao nível da estrutura interna, ou de organização relacional de informação. Os serviços de diagnóstico apresentados baseiam-se em pressupostos de análise contextual fortemente suportados em técnicas de análise visual e revelam, através de algumas experiências de cariz empírico, conseguir dar resposta ao desafio lançado por esta tese.



## keywords

usability evaluation, human-computer interaction visualization, web log analysis, information visualization.

## abstract

Technologically mediated info-communicational scenarios are becoming more and more pervasive in the day-to-day activity of a growing number of individuals and institutions. Specifically, internet/web technologies and services have a strong presence in institutions worldwide. Internal web sites (also known as intranets) are developed in compliance with internal communication strategies, reflecting internal information, workflow and related communication services. An emerging problem concerns the management of these constantly growing internal info-communicational platforms (intranets) and its external counterparts (extranets). Organizational communication specialists lack efficient tools to analyze (activity and behavioral patterns) and understand what is really going on inside the institutions. In fact, these instruments tend to be based on classical technical metrics, in most situations, for technical tuning and not for organizational communication and information analysis.

This thesis is focused on the conception and evaluation of these diagnostic tools in order to contribute to the development of these sophisticated infrastructures and, consequently, improve the efficiency of their internal info-communicational processes. One of the issues lies in identifying user-system mismatch at the human-computer interaction level, which must be thoroughly identified, and the problems pinpointed to the design team. The system must serve the organization and adapt perfectly to its internal communication strategies, sustaining efficiently its information and workflow patterns. Efficient feedback instruments are fundamental to identify info-communicational platform problems inside an institution. The offered proposals demonstrate the ability to diagnose structural and content issues at two levels: at the level of its own info-communication services interface, and at the level of the internal structure or relational layout of information. The presented diagnostic services are based upon assumed contextual analysis, strongly supported in visual assessment methods, and manage to provide a response to the challenge issued by this thesis, through some empirical experiments.





## Índice

Índice.....	xvii
Índice de figuras.....	xxiii
Índice de quadros.....	xxxii
<b>CAPÍTULO I. INTRODUÇÃO</b> .....	<b>1</b>
1. Enquadramento .....	2
2. Estrutura do documento.....	2
3. Abordagem metodológica .....	3
3.1. Terminologia e considerações .....	3
3.2. Descrição do problema .....	3
3.3. Contexto.....	5
3.4. Topologia de intervenção.....	6
3.5. Proposta de um modelo.....	8
3.6. Validação do modelo.....	8
3.6.1. Protótipo.....	8
3.6.2. Aplicação do protótipo.....	10
4. Pergunta de Investigação .....	10
<b>CAPÍTULO II. LEVANTAMENTO BIBLIOGRÁFICO</b> .....	<b>11</b>
1. Introdução .....	12
2. O que já existe relacionado com o tema desta tese.....	12
3. Comunicação organizacional mediada por computador .....	15
3.1. Percurso histórico .....	19
3.2. As tecnologias e o seu enquadramento organizacional .....	25
3.2.1. Serviços de comunicação assíncrona.....	26
3.2.2. Serviços de comunicação síncrona .....	32
3.2.3. Serviços de suporte e difusão de informação.....	34
3.2.4. Serviços de segurança.....	40
4. Visualização de informação – principais aspetos no desenvolvimento duma solução .....	41
4.1. Aplicação de cor em visualização.....	43
4.1.1. Cor para classificar.....	44
4.1.2. Sequência de cores .....	46
4.2. Taxonomias para a visualização de informação.....	47
4.2.1. Tipos de dados.....	49
4.2.2. Métodos de visualização.....	51
4.2.3. Técnicas de apresentação e interação .....	52
4.3. Visualização de estruturas de dados lineares .....	55
4.4. Visualização de redes e grafos.....	56

4.5. Visualização de redes .....	57
4.6. Visualização de informação multidimensional .....	57
4.7. Visualização de sites web .....	58
5. Usabilidade de sites web .....	60
5.1. Ficheiro de log de site web .....	60
5.2. Métrica.....	60
5.3. Estudos de usabilidade aplicados a sites web.....	61
6. Ferramentas de análise de log's.....	62
6.1. FastStats Analyzer .....	63
6.2. WebTrends Log Analyzer.....	66
6.3. Algumas Conclusões .....	68
6.4. Outras abordagens.....	69
6.5. Evolução das ferramentas de análise (2016) .....	70

**CAPÍTULO III. MODELO DE ANÁLISE E VISUALIZAÇÃO DE INFORMAÇÃO – UMA PROPOSTA** **77**

1. Introdução .....	78
2. Modelo conceptual .....	79
2.1. Elementos constituintes .....	79
2.2. Articulação entre módulos.....	81
2.3. Instrumentos de manipulação direta .....	82
3. Raw data .....	83
3.1. Estrutura de dados .....	84
3.2. Recolha de informação .....	85
3.2.1. Ficheiro de log de site web.....	86
3.2.2. Páginas da intranet .....	87
3.2.3. Eye Tracker .....	87
3.2.4. Intercetor .....	89
3.3. Caracterização da informação recolhida – Taxonomia de informação.....	90
3.3.1. Ficheiro de log de site web.....	90
3.3.2. Páginas da intranet .....	91
3.3.3. Movimento Ocular (Eye Tracker) .....	91
3.3.4. Interação Humano-Computador (Intercetor) .....	92
3.4. Condicionantes tecnológicas e representação de informação .....	93
3.4.1. Fontes de informação.....	94
3.4.2. Limitações no preenchimento da estrutura de dados .....	95
3.5. Extração de informação de um ficheiro log de um servidor web .....	96
3.6. Identificação de uma sessão de um utilizador .....	98
4. Métodos de visualização e análise de informação .....	99
4.1. Algumas definições .....	100
4.2. Taxonomia.....	102

4.3. Aplicação de cor – propostas de tabelas de cor .....	106
4.4. Visualização da organização do campo visual de interação .....	107
4.4.1. Visualização das áreas (blocos) de uma página .....	107
4.4.2. Visualização de detalhes de zonas interativas .....	109
4.4.3. Visualização de detalhes de zonas interativas em 3D.....	110
4.4.4. Relações de uma página .....	111
4.4.5. Informação sobre uma área da página .....	112
4.4.6. Visualização da página associada ao link .....	113
4.4.7. Apresentação de informação de eye-tracking .....	114
4.4.8. Apresentação de informação de mouse-traking .....	115
4.4.9. Visualização 3D dos links utilizados por um utilizador.....	116
4.4.10. Visualização dos links usados por um utilizador.....	117
4.4.11. Visualização da distribuição espacial de áreas de um site.....	118
4.5. Visualização da estrutura do site / sessão.....	120
4.5.1. Visualização das páginas de um site .....	120
4.5.2. Visualização da estrutura de um site (versão 1).....	121
4.5.3. Visualização da estrutura de um site (versão 2).....	122
4.5.4. Visualização da estrutura de um site (versão 3).....	124
4.5.5. Visualização da estrutura de um site em 3D .....	125
4.5.6. Informação sobre uma página .....	126
4.5.7. Visualização da página ampliada.....	127
4.5.8. Informação sobre uma ligação.....	128
4.5.9. Visualização / enaltecimento de informação relacional de uma ligação ou nodo .....	129
4.5.10. Visualização dos trajetos possíveis .....	130
4.5.11. Visualização da taxa de utilização de cada página de um site.....	131
4.5.12. Visualização das páginas de uma sessão de um utilizador .....	132
4.5.13. Visualização em 3D das páginas de uma sessão de um utilizador.....	132
4.5.14. Visualização das páginas de uma sessão de um utilizador (por níveis) .....	134
4.5.15. Identificação das ligações entre um conjunto de páginas .....	136
5. Integração e correlação visual de métodos de visualização .....	137
6. Algumas conclusões .....	139

## **CAPÍTULO IV. PROTÓTIPO** .....

1. Introdução .....	142
2. Tipos de protótipos .....	143
2.1. Protótipo evolutivo.....	144
2.2. Protótipo descartável .....	145
3. Trabalho prévio exploratório .....	147
3.1. Plataformas de prototipagem rápida em visualização .....	147
3.2. Linguagens, ambientes ou plataformas de desenvolvimento.....	150
4. Primeira proposta de protótipo – versão 0.....	152

4.1. Parser Métrico de ficheiro de log .....	152
4.2. Visualização .....	154
4.3. Correlação e integração de módulos .....	156
4.4. Interface gráfica de utilizador .....	156
5. Segundo protótipo – versão 1 .....	160
5.1. Interface base.....	161
5.2. Métodos de visualização .....	164
5.3. Evolução.....	170
6. Protótipo final – versão 2 .....	171
6.1. Base de dados .....	174
6.2. Estrutura de suporte (Framework) .....	175
6.3. Intercetor .....	176
6.4. Analisador .....	177
6.5. Compilador .....	178
6.6. Visualizador.....	179
6.6.1. Modelo.....	179
6.6.2. Interface de utilizador base .....	182
6.6.3. Métodos de visualização .....	190
7. Algumas conclusões .....	198
<b>CAPÍTULO V. ESTUDO DE APLICABILIDADE DO PROTÓTIPO .....</b>	<b>199</b>
1. Introdução .....	200
2. Avaliação das várias versões do protótipo .....	201
2.1. Introdução .....	201
2.2. Características gerais da aplicação .....	203
2.3. Metodologia de avaliação .....	204
2.3.1. Caracterização de utilizadores.....	205
2.3.2. Instrumentos utilizados na experiência .....	207
2.3.3. Objetivos e tarefas .....	207
2.3.4. Procedimentos .....	208
2.3.5. Informação recolhida.....	211
2.4. Resultados .....	212
2.4.1. Resultados obtidos com estudantes .....	212
2.4.2. Resultados obtidos com profissionais.....	216
2.5. Algumas conclusões .....	216
3. Repositórios web de estudo.....	217
3.1. Portal CPJ [1ª fase] .....	218
3.2. Disciplina do WebCT (e-learning) [2ª fase].....	219
3.3. Site da disciplina IHC [3ª fase].....	221
4. Resultados .....	222
4.1. Dados do CPJ [1ª fase].....	222

---

4.2. Dados do WebCT [2ª fase] .....	223
4.3. Dados da disciplina IHC [3ª fase] .....	224
5. Algumas conclusões .....	225
<b>CAPÍTULO VI. CONCLUSÕES E TRABALHO FUTURO</b> .....	<b>227</b>
1. Conclusões .....	228
2. Reflexão crítica .....	228
3. Contribuição original .....	230
4. Trabalho futuro.....	231
<b>BIBLIOGRAFIA</b> .....	<b>235</b>
<b>ANEXOS</b> .....	<b>245</b>
1. Organização/arquitetura da informação recolhida.....	246
1.1. Representação do Site.....	246
1.2. Representação de uma sessão de um utilizador.....	248
1.3. Representação de múltiplas sessões .....	251
1.4. Informação estatística .....	251
2. Extração de informação de um ficheiro de log de um servidor web.....	254
2.1. Exemplo para um servidor Microsoft IIS 5.0 (aderente ao formato W3C).....	254
2.2. Exemplo para um servidor Unix Apache genérico.....	256



## Índice de figuras

Figura II-1: Exemplo de uma tabela de texto. Informação obtida utilizando o programa FastStatsAnalyser v2.77 da Mach5 Software, Inc. ....	13
Figura II-2: Exemplo de um gráfico simples (duas séries de valores em simultâneo). Informação obtida utilizando o programa FastStatsAnalyser v2.77 da Mach5 Software, Inc. ....	13
Figura II-3: Exemplo de um digrama de barras (duas séries de valores em simultâneo). Informação obtida utilizando o programa FastStatsAnalyser v2.77 da Mach5 Software, Inc. ....	14
Figura II-4: Exemplo de uma mistura de um gráfico / diagrama de barras. Informação obtida utilizando o programa FastStatsAnalyser v2.77 da Mach5 Software, Inc. ....	14
Figura II-5: Exemplo de um diagrama de barras. Informação obtida utilizando o programa FastStatsAnalyser v2.77 da Mach5 Software, Inc. ....	14
Figura II-6: Exemplo de um diagrama tipo queijo. Informação obtida utilizando o programa FastStatsAnalyser v2.77 da Mach5 Software, Inc. ....	14
Figura II-7: Modelo de referência da Visualização (adaptado de (Card et al., 1999)) .....	42
Figura II-8: Sistemas operativos dos visitantes de um site web. Informação obtida utilizando o programa FastStatsAnalyser v2.77 da Mach5 Software, Inc. ....	64
Figura II-9: Horas mais populares do dia na utilização de um site web. Informação obtida utilizando o programa FastStatsAnalyser v2.77 da Mach5 Software, Inc. ....	65
Figura II-10: Utilização da largura de banda ao longo de um período de tempo de um site web. Informação obtida utilizando o programa FastStatsAnalyser v2.77 da Mach5 Software, Inc. ....	65
Figura II-11: Frases utilizadas em motores de busca para chegar a um site web. Informação obtida utilizando o programa FastStatsAnalyser v2.77 da Mach5 Software, Inc. ....	66
Figura II-12: Utilizadores autenticados mais ativos num site web. Informação obtida utilizando o programa WebTrends Log Analyser v6.0a da WebTrends Corp. Esta ferramenta produz os resultados em formato HTML.....	67
Figura II-13: Nível de atividade (número de sessões) por hora do dia num site web. Informação obtida utilizando o programa WebTrends Log Analyser v6.0a da WebTrends Corp. Esta ferramenta produz os resultados em formato HTML. ....	67
Figura II-14: Vista geral da utilização de um site web. Ilustração retirada de (Hochheiser & Shneiderman, 1999). ....	69
Figura II-15: Página mais usadas como ponto de entrada de um site web. Informação obtida utilizando o programa FastStat Analyzer v4.1.7 da Mach5 Software, Inc. ....	71
Figura II-16: Duração média das sessões dos visitantes de um site web. Informação obtida utilizando o programa FastStat Analyzer v4.1.7 da Mach5 Software, Inc. ....	71
Figura II-17: Vista das ligações de e para uma determinada página de um site web. Informação obtida utilizando o programa FastStat Analyzer v4.1.7 da Mach5 Software, Inc. ....	72
Figura II-18: Exemplo de uma vista integrada (Analytics Website Space) da análise da utilização de um site web. Imagem retirada do site da Webtrends. ....	73
Figura II-19: Exemplo de uma vista integrada (Analytics for SharePoint) da análise da utilização de um site web interno de uma empresa. Imagem retirada do site da Webtrends.....	73

Figura II-20: Exemplo de uma vista integrada (Facebook Page and App Spaces) da análise da utilização de um site web. Imagem retirada do site da Webtrends. ....	74
Figura II-21: Exemplo de uma vista integrada (Mobile Website Spaces) da análise da utilização de um site web. Imagem retirada do site da Webtrends. ....	74
Figura II-22: Exemplo de relatório da análise da utilização de um site web. Imagem retirada do site da Google ( <a href="https://ga-dev-tools.appspot.com/embed-api/">https://ga-dev-tools.appspot.com/embed-api/</a> ).....	75
Figura II-23: Exemplo de relatório da análise da utilização de um site web. Imagem retirada do site da Google ( <a href="https://ga-dev-tools.appspot.com/polymer-elements/">https://ga-dev-tools.appspot.com/polymer-elements/</a> ). ....	75
Figura III-1: Modelo do processo de visualização de um site web, usando uma arquitetura de representação estática. ....	79
Figura III-2: Modelo do processo de visualização de um site web, usando uma arquitetura de representação dinâmica. Note-se a ligação adicional entre o módulo “Técnicas de exploração de informação” e o módulo “Parser Métrico”. ....	83
Figura III-3: Representação gráfica da taxonomia proposta, em que cada eixo corresponde a um tipo de classificação. No geral, um método de visualização corresponderá a um ponto de interceção nos três eixos apresentados (um ponto num espaço 3D). Baseado nas propostas de (Shneiderman, 1996), (Keim, 2000, 2001) e (Andrews, 2002). ....	103
Figura III-4: Tabela de cinzentos, de preto a branco com variação linear. Nesta tabela só há variação da luminância. ....	106
Figura III-5: Tabela de tons de ciano. Varia desde preto até ao ciano mais saturado. A luminância tem uma variação linear. ....	106
Figura III-6: Tabela de cor com variação entre azul e amarelo. Esta tabela tem a particularidade de associar uma variação de luminância e de crominância. ....	107
Figura III-7: Tabela de pseudocor usada em algumas modalidades imagiológicas médicas. ....	107
Figura III-8: Níveis de luminância da tabela de pseudocor apresentada na Figura III-7. ....	107
Figura III-9: Exemplo de visualização, recorrendo a cor, da classificação das zonas identificadas numa página web. ....	108
Figura III-10: Exemplo de visualização das zonas de interação de uma página web e representação das páginas às quais estão ligadas. ....	109
Figura III-11: Exemplo de visualização das zonas de interação de uma página web e representação 3D das páginas às quais estão ligadas. ....	110
Figura III-12: Exemplo de visualização de uma página web com as suas zonas de interação. Do lado esquerdo, encontram-se as páginas que contêm links (com as respetivas áreas identificadas) que apontam para a página corrente. Do lado direito, estão as páginas para as quais apontam cada um dos links da página corrente. ....	111
Figura III-13: Exemplo de visualização dos detalhes referentes à área sob o cursor do rato. Os detalhes aparecem como uma pequena janela de texto (tool-tip window), opaca ou semitransparente, junto ao cursor do rato. ....	112
Figura III-14: Exemplo de visualização da miniatura da página a que se refere o link associado à área sob o cursor do rato. A miniatura da página como uma pequena janela (tool-tip window), opaca ou semitransparente, junto ao cursor do rato. ....	113
Figura III-15: Exemplo de visualização do foco e sentido do olhar de um utilizador enquanto observa uma página web. ....	114
Figura III-16: Exemplo de visualização dos eventos relacionados com o rato. ....	115



Figura III-17: Exemplo de visualização do percurso de um utilizador num site com identificação dos links utilizados representados ao longo de uma linha temporal... 116	116
Figura III-18: Exemplo de visualização do percurso de um utilizador num site com identificação dos links utilizados, representados num plano..... 117	117
Figura III-19: Exemplo de visualização da distribuição espacial das áreas de um dado tipo ao longo do site. Os valores (número de ocorrências, para um dado ponto) poderão ser codificados com recurso a uma tabela de cor..... 119	119
Figura III-20: Interface de manipulação dos limites da tabela de pseudocor a aplicar a um esquema de visualização. Os limites inferior e superior podem ser deslocados de forma a só ficar visível a parte correspondente aos valores entre os limites..... 119	119
Figura III-21: Exemplo de visualização da estrutura de um site em 2D. Tem-se simplesmente uma representação de cada página constituinte do site. .... 120	120
Figura III-22: Exemplo de visualização da estrutura de um site em 2D, com as páginas organizadas por níveis e respetivas ligações..... 122	122
Figura III-23: Exemplo de visualização da estrutura de um site em 3D. Cada página é representada como uma folha nesse espaço 3D. As páginas estão organizadas por níveis. Uma página de um dado nível pode estar ligada a zero, uma ou mais páginas do nível seguinte – links da página. Cada página poderá ser representada mais que uma vez, em níveis diferentes. .... 123	123
Figura III-24: Exemplo de visualização da estrutura de um site em 3D. Cada página é representada como uma folha nesse espaço 3D. As páginas estão organizadas por níveis. Uma página de um dado nível pode estar ligada a uma ou mais páginas – links da página – que podem encontrar-se em qualquer nível. Cada página é representada uma única vez..... 125	125
Figura III-25: Exemplo de visualização da estrutura de um site em 3D. Cada página é representada também em 3D. As páginas estão organizadas por níveis. Uma página de um dado nível pode estar ligada a uma ou mais páginas do nível seguinte – links da página. .... 126	126
Figura III-26: Exemplo de visualização dos detalhes referentes à página sob o cursor do rato. Os detalhes aparecem como uma pequena janela de texto (tool-tip window), opaca ou semitransparente, junto ao cursor do rato. .... 127	127
Figura III-27: Exemplo de visualização de uma ampliação da página a que se refere a miniatura sob o cursor do rato. A ampliação aparece como uma pequena janela (tool-tip window), opaca ou semitransparente, junto do cursor do rato..... 128	128
Figura III-28: Exemplo de visualização dos detalhes referentes à ligação entre duas páginas sob o cursor do rato. Os detalhes aparecem como uma pequena janela de texto (tool-tip window), opaca ou semitransparente, junto ao cursor do rato..... 128	128
Figura III-29: Exemplo de visualização das páginas relacionadas com uma ligação. Tanto a ligação, sob o rato, como as páginas afetadas aparecem enaltecidas. O efeito pode ser conseguido usando por exemplo, um traço mais grosso e/ou mudando a cor dos objetos em causa..... 129	129
Figura III-30: Exemplo de visualização das páginas relacionadas com uma página (nodo). Tanto a página, sob o rato, como as ligações afetadas aparecem enaltecidas. O efeito pode ser conseguido usando, por exemplo, um traço mais grosso e/ou mudando a cor dos objetos em causa..... 129	129
Figura III-31: Exemplo de visualização dos percursos possíveis entre duas páginas selecionadas. No caso ilustrado, existem quatro percursos possíveis com um número díspar de cliques. .... 130	130

Figura III-32: Exemplo de visualização da taxa de utilização de cada página de um site, bem como das suas respectivas ligações. O número de visitas de cada página e/ou ligação está codificada com recurso a uma tabela de cor. ....	131
Figura III-33: Exemplo de visualização de uma sessão de um utilizador, identificando cada página visitada e o tempo decorrido entre cada página utilizados pelo utilizador....	132
Figura III-34: Exemplo de visualização em 3D das páginas de um site web, correspondentes a uma visita de um utilizador. ....	133
Figura III-35: Exemplo de visualização em 3D das páginas de um site web, correspondentes a uma visita de um utilizador. ....	133
Figura III-36: Exemplo de visualização em 3D das páginas de um site web, correspondentes a uma visita de um utilizador. ....	134
Figura III-37: Exemplo de representação de uma sessão de um utilizador, identificando cada página visitada pelo utilizador e respetivo nível. ....	135
Figura III-38: Exemplo de identificação das ligações existentes entre um conjunto de páginas previamente selecionadas. ....	136
Figura III-39: Exemplo de integração de quatro métodos de visualização. ....	140
Figura IV-1: Modelo do processo de visualização de um site web, usando uma arquitetura de representação dinâmica. ....	142
Figura IV-2: Aspeto geral da interface do Iris Explorer. Estão representadas várias componentes da aplicação. ....	149
Figura IV-3: Modelo do processo de visualização de um site web, usando uma arquitetura de representação dinâmica (com os blocos “Parser Métrico” e “Visualização” enaltecidos). ....	152
Figura IV-4: Modelo do processo de visualização de um site web, usando uma arquitetura de representação dinâmica (bloco Parser Métrico). ....	153
Figura IV-5: Modelo do processo de visualização de um site web, usando uma arquitetura de representação dinâmica (bloco “Visualização”). ....	154
Figura IV-6: Vista 3D das páginas de uma sessão de um utilizador. As páginas estão empilhadas recorrendo a semitransparência. Imagem obtida com a versão 0 do protótipo. ....	155
Figura IV-7: Vista 3D das páginas de uma sessão de um utilizador. As páginas estão empilhadas recorrendo a um operador do tipo XOR. Imagem obtida com a versão 0 do protótipo. ....	155
Figura IV-8: Visualização de uma sessão de um utilizador. Do lado esquerdo estão representadas miniaturas das várias páginas visitadas. Do lado direito está representada a página selecionada com a identificação dos Hotspots. Imagem obtida com a versão 0 do protótipo. ....	157
Figura IV-9: Diagrama com a representação de uma sessão de um utilizador. Imagem obtida com a versão 0 do protótipo. ....	158
Figura IV-10: Interface de navegação de um site. Imagem obtida com a versão 0 do protótipo. ....	159
Figura IV-11: Diagrama com uma representação de um site. São representados os percursos de duas pequenas sessões de utilizador. Imagem obtida com a versão 0 do protótipo. ....	159
Figura IV-12: Vista 3D das páginas usadas numa sessão de um utilizador. Imagem obtida com a versão 0 do protótipo. ....	160

Figura IV-13: Vista manipulada dos vários elementos da interface base do protótipo, com os menus visíveis. Imagem obtida com a versão 1 do protótipo, publicada em (Nunes et al., 2002). .....	161
Figura IV-14: Aplicação com todas as vistas ativas. Note-se que as vistas estão dispostas em cascata, ficando a maior parte das janelas ocultas pelas que estão no topo. Imagem obtida com a versão 1 do protótipo. ....	162
Figura IV-15: Aplicação com todas as vistas ativas. Note-se que estão dispostas de maneira a que todas sejam visíveis em simultâneo. Imagem obtida com a versão 1 do protótipo. ....	162
Figura IV-16: Múltiplas vistas sincronizadas. Ao selecionar uma página num vista, são todas atualizadas de forma sincronizada em conformidade. Imagem obtida com a versão 1 do protótipo. ....	163
Figura IV-17: Vista das páginas de um site. Na parte esquerda está a lista das páginas do site e do lado direito está uma representação real da página selecionada. Imagem obtida com a versão 1 do protótipo, publicada em (Nunes et al., 2002). ...	164
Figura IV-18: Visualização em 3D da estrutura de um site. Cada página é representada por um cubo e as ligações entre cada página estão também representadas. Imagem obtida com a versão 1 do protótipo, publicada em (Mealha et al., 2004). .	165
Figura IV-19: Visualização em 2D da estrutura de um site. Cada página é representada por uma miniatura e as ligações entre cada página estão também representadas. Imagem obtida com a versão 1 do protótipo. ....	166
Figura IV-20: Interface de exploração de um site. Em baixo aparece a lista das páginas do site. Na parte superior, no centro está a página corrente, do lado direito uma lista com os links e do lado esquerdo uma lista com os seus referers. Imagem obtida com a versão 1 do protótipo, publicada em (Nunes et al., 2002). ....	167
Figura IV-21: Apresentação da informação estatística. No quadrante superior esquerdo seleciona-se o tipo de informação pretendida e à direita a sua representação num gráfico de barras. Na parte inferior é possível selecionar os elementos constituintes do gráfico. Imagem obtida com a versão 1 do protótipo, publicada em (Nunes et al., 2002). ....	168
Figura IV-22: Interface que permite selecionar uma ou várias sessões. Permite ainda pesquisar e/ou ordenar segundo os critérios pretendidos. Imagem obtida com a versão 1 do protótipo, publicada em (Nunes et al., 2002). ....	168
Figura IV-23: Visualização de uma sessão de utilizador. Na parte esquerda está a lista de páginas visitadas. No lado direito pode observar-se a página selecionada, neste caso com informação de EyeTracking sobreposta. Imagem obtida com a versão 1 do protótipo. ....	169
Figura IV-24: Representação 3D das páginas de uma sessão. As páginas são apresentadas por ordem cronológica. É ainda possível selecionar uma página. Imagem obtida com a versão 1 do protótipo, publicada em (Nunes et al., 2002). ....	170
Figura IV-25: Modelo simplificado do sistema. Adaptada de (F. L. Zamfir, 2007). ....	172
Figura IV-26: Linha temporal da análise a um site. Adaptada de (F. L. Zamfir, 2007). ....	173
Figura IV-27: Representação hierárquica de um site web na base de dados. Retirado de (F. L. Zamfir, 2007). ....	173
Figura IV-28: Informação de um site web na base de dados. Retirado de (F. L. Zamfir, 2007). ..	174
Figura IV-29: Informação de uma página web na base de dados. Retirado de (F. L. Zamfir, 2007). ....	175
Figura IV-30: Modelo simplificado das várias partes constituintes da aplicação. Adaptado de (F. Zamfir, Nunes, Teixeira, Santos, et al., 2004). ....	176

Figura IV-31: Interface da aplicação SiteAnalyzer. Na parte superior podem controlar-se os detalhes do que se pretende analisar. Na parte inferior é possível utilizar a página, como se de um navegador normal se tratasse. Imagem obtida com a versão 2 do protótipo, publicada em (F. L. Zamfir, 2007). .....	177
Figura IV-32: Interface da aplicação Compiler. Imagem obtida com a versão 2 do protótipo, publicada em (F. L. Zamfir, 2007). .....	178
Figura IV-33: Entidade de sessão. ....	179
Figura IV-34: Modelo das camadas da aplicação Visualizador. Adaptado de (F. L. Zamfir, 2007). .....	180
Figura IV-35: Subcomponentes da aplicação Visualizador. Adaptado de (F. L. Zamfir, 2007). ....	181
Figura IV-36: Passos do processo de configuração inicial do Wizard (da esquerda para a direita e de cima para baixo). Imagem obtida com a versão 2 do protótipo, publicada em (F. L. Zamfir, 2007). .....	183
Figura IV-37: Possibilidades de manipulação da interface de utilizador do Visualizador. Imagem obtida com a versão 2 do protótipo, adaptado de (F. L. Zamfir, 2007). .....	184
Figura IV-38: Modelo da interface de utilizador do Visualizador. Imagem obtida com a versão 2 do protótipo, adaptado de (F. L. Zamfir, 2007). .....	185
Figura IV-39: Ilustração da flexibilidade dos objetos constituintes da interface de utilizador do Visualizador. Imagem obtida com a versão 2 do protótipo, adaptado de (F. L. Zamfir, 2007). .....	186
Figura IV-40: Padrão da vista corrente. ....	186
Figura IV-41: Seleção tipo visualização. ....	186
Figura IV-42: Barra ferramentas. ....	187
Figura IV-43: Barra ferramentas web. ....	187
Figura IV-44: Menu seleção método visualização. ....	188
Figura IV-45: Seleção da janela com um conjunto de esquemas de visualização relacionados. .	188
Figura IV-46: Seleção da janela com um conjunto de esquemas de visualização relacionados – vista só com os ícones. ....	188
Figura IV-47: Vista da tabela de cor ativa. ....	189
Figura IV-48: Manipulação da tabela de cor ativa. ....	189
Figura IV-49: Detalhes de alguns aspetos da interface de utilizador. Imagem obtida com a versão 2 do protótipo, retirado de (F. L. Zamfir, 2007). .....	189
Figura IV-50: Vista da componente Explorer (Exploração do Site). Imagem obtida com a versão 2 do protótipo, retirado de (F. L. Zamfir, 2007). .....	190
Figura IV-51: Representação da informação relacional das páginas de um site web. Imagem obtida com a versão 2 do protótipo .....	191
Figura IV-52: Representação dos percursos possíveis para se chegar a uma página destino a partir de uma página de partida. Imagem obtida com a versão 2 do protótipo, retirado de (F. L. Zamfir, 2007). .....	192
Figura IV-53: Localização do Hotspots numa página e informação hipermedia relacionada. Imagem obtida com a versão 2 do protótipo. ....	193
Figura IV-54: Representação dos resultados de uma pesquisa. As páginas aparecem desenhadas em miniatura (thumbnail). Imagem obtida com a versão 2 do protótipo. ....	194

---

Figura IV-55: Representação dos tempos envolvidos no percurso pela estrutura em forma de árvore de um site por um utilizador durante uma sessão. Imagem obtida com a versão 2 do protótipo, retirado de (F. L. Zamfir, 2007). .....	195
Figura IV-56: Representação de uma sessão de utilizador (linha temporal vs. posição da página na estrutura em árvore do site). Imagem obtida com a versão 2 do protótipo. ....	196
Figura IV-57: Representação sobreposta dos Hotspots e da utilização das ligações. Imagem obtida com a versão 2 do protótipo. ....	197
Figura V-1: Utilização simultânea, em sincronia, de quatro esquemas de visualização diferentes. Imagem obtida com a versão 2 do protótipo. ....	204
Figura V-2: O papel dos estudantes. Retirado de (Santos et al., 2004). ....	209
Figura V-3: Diagrama de caixas correspondente aos tempos gastos pelos utilizadores #1 e #2. Retirado de (Santos et al., 2004). ....	213
Figura V-4: Diagrama de caixas correspondente ao tempo e satisfação de cada tarefa. Retirado de (Santos et al., 2004). ....	214



## Índice de quadros

Quadro II-1: Nível 1: Intranet básica (adaptado de (Hinrichs, 1997)).	22
Quadro II-2: Nível 2: Intranet para publicação de repositórios (adaptado de (Hinrichs, 1997)).	23
Quadro II-3: Nível 3: Intranet para colaboração (adaptado de (Hinrichs, 1997)).	23
Quadro II-4: Nível 4: Intranet para transações (adaptado de (Hinrichs, 1997)).	24
Quadro II-5: Nível 5: Extranet (adaptado de (Hinrichs, 1997)).	24
Quadro V-1: Número de utilizadores que completou, ou não, tarefas sem perguntas. Retirado de (Santos et al., 2004).	213
Quadro V-2: Número de utilizadores que completou corretamente, erradamente ou não fez tarefas que incluíam uma questão. Adaptado de (Santos et al., 2004).	214





# **Capítulo I.**

## **INTRODUÇÃO**

# 1. Enquadramento

As organizações têm vindo a adotar soluções de comunicação e fluxo de informação interna integralmente sustentadas em contextos de mediação tecnológica. A estratégia que lhes está subjacente pressupõe um aumento do desempenho humano e dos processos que lhe estão associados. Uma consequência natural deste facto corresponderá a um aumento de qualidade na prossecução da missão da organização.

Um contexto de comunicação interna desta natureza proporciona a utilização de instrumentos de gestão bastante elucidativos da dinâmica organizacional. Fundamental será, contudo, correlacionar e analisar, de forma visualmente integrada e contextual, a informação para poder extrair soluções com vista a uma intervenção pragmática onde forem identificados problemas, ou seja, uma previsível otimização de desempenho, da atividade e/ou comportamento infocomunicacional de cada um dos agentes internos das instituições.

Este trabalho enquadra e propõe um conjunto de metodologias e técnicas de visualização do comportamento infocomunicacional, suas propriedades, funcionalidades, inter-relações ou complementaridades no contexto de um instrumento de gestão da comunicação e informação nas organizações. As fontes de informação utilizadas para a construção do espaço de visualização e análise são os ficheiros de *log* provenientes das plataformas digitais (servidores) convenientemente configuradas para efetuar o registo do comportamento infocomunicacional dos respetivos utilizadores em contexto de uso.

# 2. Estrutura do documento

Este documento está estruturado nos seguintes grandes blocos:

- Abordagem metodológica adotada;
- Levantamento bibliográfico nas áreas envolvidas;
- Apresentação da proposta de modelo de análise e visualização de informação em plataforma digitais;
- Descrição do processo de desenvolvimento, implementação e avaliação de um protótipo;
- Estudo de aplicabilidade do protótipo
- Conclusões e trabalho futuro;
- Lista de bibliografia referida.

## 3. Abordagem metodológica

### 3.1. Terminologia e considerações

Este contexto de doutoramento impõe algumas considerações de natureza terminológica e conceptual atendendo às características transdisciplinares (L. d. Freitas et al., 1994) de algumas matérias abordadas. Um dos conceitos dominantes é o de sistema, nomeadamente sistema infocomunicacional sustentado em soluções de mediação tecnológica, fundamentalmente com recurso a tecnologias e serviços *world wide web* (*web*). Por vezes, e atendendo ao registo científico do trabalho desta tese na área das Ciências e Tecnologias da Comunicação (CTC), é habitual usar-se o termo “plataformas digitais” numa alusão a estes mesmos sistemas. Considerando CTC a área científica dominante deste trabalho, é importante referir que as áreas pertinentes de referência para definir métodos científicos, gestão projetual, conceitos e termos são as áreas de Ciências da Comunicação, Ciência da Informação e Ciência da Computação (Mealha, 2012), sem prejuízo de outras que contribuem mas talvez em menor dimensão para esta tese como a psicologia, sociologia, antropologia, etc. O termo infocomunicacional é recorrente em muitos trabalhos de investigação no Centro de Investigação em Comunicação, Informação e Cultura Digital (CIC.Digital<sup>1</sup>) e sublinha a pertinência da forte correlação existente entre os conceitos comunicação e informação quando se quer compreender o comportamento de interação do ser humano nestes contextos de mediação tecnológica. Numa perspetiva, para compreender as características da comunicação interpessoal mediada tecnologicamente e por outra, mais relacionada com esta tese, a relação do ser humano com a informação nos vários formatos e tipologias de acesso com que hoje nos debatemos.

### 3.2. Descrição do problema

Tem-se vindo a assistir, de forma crescente, à proliferação de soluções de mediação tecnológica de apoio a sistemas de comunicação e informação em vários contextos. A comunicação interna nas instituições, para o efeito desta tese é o contexto mais relevante, contudo outros como as comunidades *online*, formação e educação a distância, teletrabalho, etc., são exemplos que podem vir a usufruir dos contributos deste trabalho.

---

<sup>1</sup> CIC.Digital: <http://www.cicdigital.org/>

A mediação tecnológica considerada nesta tese permite uma facilitação dos processos infocomunicacionais, com alguns benefícios. Para enunciar alguns, pode-se referir o contributo para um aumento da eficiência da comunicação interpessoal mediada, uma melhor gestão do trabalho e da disponibilidade, aumento da abrangência da comunicação sem redução da eficácia, melhor gestão integrada da informação, contribuir diretamente para a redução de custos operacionais atendendo à redução do tempo nos ciclos processuais (produção) e de trabalho e na gestão genérica de vários recursos mediados tecnologicamente.

Contudo, este espaço tecnológico pode apresentar alguns problemas que se manifestam de várias formas e em várias situações infocomunicacionais. É possível identificar problemas tanto ao nível da mediação tecnológica da comunicação interpessoal como ao nível da interação Humano-Tecnologia/Computador (Norman, 2002). É habitual verificar que: i) existe uma nítida desadequação da solução tecnológica relativamente ao processo de comunicação em causa; ii) que existem problemas de conforto ergonómico físico e cognitivo, até por falta de literacia para algumas soluções de mediação adotadas; iii) existem problemas na forma como a informação é disponibilizada, nomeadamente ao nível da organização e disponibilização (partilha) de documentos internos de natureza e em contextos operacionais diversos; em suma, verifica-se habitualmente uma desadequação entre o modelo de operação infocomunicacional adotado e o modelo expectado por cada um dos seres humanos que integra estes sistemas de informação e comunicação institucionais, com um prejuízo claro para o resultado esperado da missão da respetiva instituição.

Estes problemas vêm lembrar que para melhorar o processo de comunicação, independentemente do contexto, não é suficiente introduzir instrumentos de mediação tecnológica, é necessário garantir a sua adequação ao contexto em que se enquadra, fundamentalmente garantir uma matriz de eficácia do instrumento técnico num contexto específico de satisfação (ISO, 1998) do ser humano (utilizador) em causa. Para ultrapassar esta situação é necessário conceber, desenvolver e validar instrumentos de monitorização e diagnóstico que permitam aferir esta (des)adequação e que façam sugestões conducentes a soluções para os problemas detetados. Idealmente, estes instrumentos poderiam mesmo intervir no sistema, de forma automática (subsistema de alertas), por forma a indicar e corrigir os problemas detetados.

Em suma, o trabalho inerente a esta tese tem por base a pesquisa, reflexão, conceptualização de um modelo de base visual para monitorização e análise de processos infocomunicacionais em plataformas *web*. Modelo que será validado através da prototipagem de um instrumento capaz de monitorizar, registar e diagnosticar a atividade de interação num sistema de informação e comunicação por mediação tecnológica (tecnologias e serviços *internet/web*).

### 3.3. Contexto

Alguns sistemas de comunicação por mediação tecnológica fazem parte do dia a dia das pessoas e são usados habitualmente com alguma frequência. Exemplo disto são o telefone e a *internet*, conhecidos e utilizados com alguma frequência (INE-Família, 2012) pela generalidade das pessoas.

Um olhar crítico ao contexto de uso do telefone começa logo por revelar que a taxa de penetração deste dispositivo é bem superior a 1 dispositivo por pessoa em Portugal e que, de entre os muitos tipos e modelos diferentes existentes, é, na generalidade das opiniões, fácil de usar para efeitos de comunicação interpessoal síncrona (voz) e assíncrona (SMS). A título de exemplo, e para melhor compreender o percurso de um dispositivo ou sistema de mediação da comunicação interpessoal, apresenta-se uma breve explicação das fases mais pertinentes inerentes à evolução do telefone. Quando se quer falar com alguém (comunicar), pega-se no telefone, digita-se o número da pessoa pretendida e estabelece-se a ligação. Do outro lado da linha, o telefone toca e a outra pessoa atende. A ligação está estabelecida e pode-se comunicar com essa pessoa. Este gesto, hoje bastante vulgar, pode parecer trivial, mas na verdade integra um conjunto de processos técnicos e de aferição ergonómica de dispositivo bastante mais complexos e historicamente evolutivos, do que à primeira vista pode parecer.

Mesmo não sendo objeto de estudo deste trabalho, conforme foi referido, e a título de suporte à contextualização, vai analisar-se este percurso e os processos que lhe estão inerentes com um pouco mais de detalhe. Em primeiro lugar, os terminais (os telefones que estão lá em casa) foram objeto de estudos de ergonomia, por forma a adequá-los aos utilizadores (as pessoas). Houve um longo percurso desde os primeiros modelos, com microfone fixo e manivela para chamar o operador, até aos mais recentes modelos portáteis, sem fios, com marcação automática a partir de uma longa lista de contactos guardada em memória ou com marcação por voz, em que basta dizer o nome da pessoa pretendida para que a ligação seja desencadeada e, eventualmente, estabelecida.

No que diz respeito à infraestrutura de ligação (linhas, centrais de comutação, etc.) houve também um longo percurso de evolução. Das linhas de cobre ligadas a uma central com operadores que estabeleciam as ligações manualmente, às modernas centrais digitais de comutação usando ligações de fibra ótica concentrando centenas ou milhares de ligações simultâneas num único cabo, tem-se assistido a um percurso galopante e continuado!

O próprio planeamento da rede foi objeto de estudos detalhados para determinar a capacidade de cada troço, em função das necessidades dos utilizadores desse troço. Num troço subdimensionado, haveria o risco de um utilizador pretender telefonar e não ter nenhuma linha disponível. Em contrapartida, num sobredimensionado, estar-se-ia a desperdiçar recursos e, conseqüentemente, dinheiro.

Para a determinação do dimensionamento ótimo de um dado troço, as companhias de telecomunicações procedem à monitorização permanente do tráfego nos seus troços de rede. Determinam o número médio de chamadas simultâneas em cada instante, detetam picos de utilização e tentam correlacionar esta informação com dados estatísticos que caracterizam a comunidade abrangida (número de habitantes, número de assinantes, eventos como festividades entre outros, são especialmente relevantes). Depois de ponderarem toda a informação disponível, estariam em condições de decidir qual a capacidade a atribuir a cada troço.

No que diz respeito às tecnologias e serviços da *internet*, a situação não é muito diferente. Quando se implementa, por exemplo, um *site web*, é necessário atender a um conjunto de considerações. A capacidade do servidor, a velocidade da ligação, os serviços e funcionalidades a disponibilizar, para não falar nas tecnologias a utilizar na implementação desse *site*, são considerações importantes e a ter em mente.

Para além destes aspetos, de carácter mais tecnológico, há ainda questões que têm a ver com o conteúdo do *site*. Estruturação da informação (documentos) e serviços, a forma como é apresentada, adequação aos públicos-alvo, entre outros aspetos, são questões igualmente importantes e que, por vezes, não recebem atenção suficiente.

Os objetivos do *site*, enquadrados na missão da respetiva instituição, determinarão muitas das condicionantes. Um *site* de promoção de uma empresa tem objetivos e implicações diferentes de um *site* de comércio eletrónico ou de um *site* de apoio à comunicação interna numa organização (Ceruzzi, 2003).

Na linha desta problemática, este trabalho irá debruçar-se, de forma especial, sobre as questões inerentes aos processos infocomunicacionais Humano-Computador num contexto de comunicação interna numa instituição sustentada em tecnologias e serviços de mediação tecnológica. Em analogia ao que foi anteriormente referido no percurso evolutivo do telefone, esta tese tem por objetivo verificar (testar) se um instrumento de monitorização e diagnóstico adequado poderá detetar e melhorar a eficiência dos processos infocomunicacionais em sistemas de informação e comunicação institucionais mediados tecnologicamente. Especificamente, deverá focar-se em definir métodos e técnicas para captar a experiência de acesso e uso dos documentos nos sistemas internos pelos utilizadores e fazer um levantamento do comportamento infocomunicacional ao nível das respetivas interfaces, de forma visualmente correlacionada no espaço e no tempo.

### **3.4. Topologia de intervenção**

Cada vez mais as instituições usam cenários de comunicação interna sustentados em contextos de mediação tecnológica, recorrendo às vulgarmente chamadas *intranets* e

respetivos *sites web* internos. No que concerne à monitorização e avaliação de uso destes sistemas, o mercado disponibiliza um leque, mais ou menos alargado, de ferramentas de avaliação de *sites web*. Podem encontrar-se desde aplicações de domínio público com um conjunto limitado de funcionalidades, até aplicações comerciais sofisticadas de monitorização da atividade dos utilizadores durante o seu uso e análise dos dados obtidos.

Contudo, este tipo de aplicações tende a só permitir fazer uma análise parcial do problema. Habitualmente, limita-se a analisar, numa perspetiva técnica, o que se passa relativamente ao servidor *web* (nomeadamente, através dos seus ficheiros de *log*). Este tipo de abordagem técnica normalmente não permite estudar a forma como o utilizador “interage” com cada página ou serviço, no fundo não permite registar e compreender os “comportamentos de interação infocomunicacionais”.

Podem ainda encontrar-se estudos de usabilidade aplicados a serviços / páginas *web*, que estudam as questões da interação humano-computador. Contudo, também esta abordagem está habitualmente focada em algum aspeto particular do processo infocomunicacional de uma instituição em particular.

É necessária uma solução de monitorização integradora, que permita abordar e observar os vários aspetos envolvidos nos processos de fluxo de informação e de comunicação dentro de uma organização:

- Aspetos técnicos da comunicação;
- Registo e fluxo de informação dentro da organização;
- Comportamento infocomunicacional na Interação humano-tecnologia.

No contexto desta problemática, este trabalho pretende contribuir com métodos e técnicas inovadoras de:

- Otimização da gestão dos processos infocomunicacionais internos das organizações sustentados em contextos de mediação tecnológica;
- “Visualização” da atividade de mediação tecnológica, tentando perceber o comportamento infocomunicacional no processo de interação humano-computador/tecnologia;
- Aferição do exercício de conceção de interfaces e respetivos sistemas;
- Validação duma proposta de mediação tecnológica, tendo por base um processo de conformidade com uma heurística de usabilidade.

Um levantamento bibliográfico exaustivo e a consulta a projetos e produtos já existentes constituirá um ponto de partida desta tese, contudo com a consciência de que apenas servirá de referência de base e inspiradora para a construção de um modelo, referência conceptual para o desenho concreto das soluções e respostas às necessidades e dificuldades enunciadas anteriormente.

## **3.5. Proposta de um modelo**

Nas secções anteriores viu-se que os processos infocomunicacionais em geral, e nas organizações em particular, têm vindo a ser objeto de um crescente investimento em soluções de mediação tecnológica. Esta mediação permite acelerar a difusão e o acesso à informação e pode contribuir para um aumento da eficácia nos processos de comunicação.

Viu-se também que existem alguns problemas que urge identificar e resolver nestes processos. O mercado fornece algumas soluções que se têm mostrado insuficientes para lidar com o problema eficazmente. Habitualmente lidam com a questão de forma parcelar, não estando ainda disponíveis soluções capazes de lidar com esta problemática de forma suficientemente abrangente.

Neste espírito, e como objetivo deste trabalho, propõe-se a conceptualização de um modelo com a conseqüente prototipagem de um instrumento que complemente o espectro de instrumentos de gestão da comunicação interna nas organizações. Um instrumento que integre as várias abordagens necessárias para uma visão holística dos processos de comunicação numa organização, que permita identificar possíveis problemas nesses processos e que oriente na procura de soluções para esses problemas. Uma das funções determinantes do protótipo no contexto deste trabalho será proporcionar algumas situações experimentais com intuito de validar determinados pressupostos do modelo proposto.

## **3.6. Validação do modelo**

O protótipo, conforme foi referido na secção anterior, foi aplicado a vários tipos de situações, tanto em ambientes controlados (no contexto de estudos de laboratório) como ambientes reais onde foi possível registar procedimentos de interação humano-computador que traduzem experiência de uso em contexto de uso efetivo.

### **3.6.1. Protótipo**

Importa sublinhar que a atividade de pesquisa e conceptualização inerente a esta tese estimulou, desde muito cedo, a necessidade de implementar um protótipo que permitisse testar alguns dos conceitos do modelo em desenvolvimento. Esta secção sistematiza as fases evolutivas do protótipo e a intenção de desenvolvimento associada a cada uma delas.

A primeira abordagem consistiu em tentar utilizar uma ferramenta de visualização de informação que permitisse tratar a informação recolhida e inerente ao modelo e testar



determinados métodos de visualização da mesma. A ferramenta inicialmente escolhida para esse efeito foi o IRIS Explorer da NAG. Contudo, esta solução veio a mostrar-se pouco adequada e pouco flexível para o problema em causa, pelo que foi necessário encontrar soluções alternativas mais flexíveis no tratamento e visualização da informação e com capacidade de integrar várias funções previstas para um protótipo desta natureza.

Como não foi possível identificar uma ferramenta no mercado capaz de responder a todas as necessidades, a solução adotada foi avançar para uma implementação de raiz. Para o efeito, começou por desenvolver-se pequenos programas / módulos em linguagem C / C++.

A primeira fase do protótipo, designada de “versão 0”, destinava-se a testar a extração de informação de um ficheiro de *log* de um *site web* com a pertinência de identificar e registar uma sessão de um utilizador. Este aspeto será discutido com mais detalhe e é uma das situações mais críticas relacionada com a recolha de informação de uso através de ficheiros de *log*. Esta fase também foi utilizada para testar a adequabilidade de alguns métodos de visualização da informação recolhida.

A versão seguinte do protótipo, “versão 1”, tinha por objetivo fundamental a integração de várias funções, num único instrumento (aplicação), de forma a garantir funcionalidades e robustez para a fase de testes experimentais prevista. Um dos problemas identificados nesta versão e que levou a uma revisão da arquitetura global do protótipo foi a falta de modularidade do mesmo. Esta versão do protótipo revelou-se incapaz de integrar eficazmente outros módulos com funcionalidades específicas para responder ao desenho conceptual e validação do modelo de referência. À medida que novos conceitos iam sendo estudados e testados, havia a necessidade de refinar e ampliar as capacidades operativas do protótipo. Compreendeu-se que era imperativo uma nova plataforma que permitisse fazer novas experiências ou introduzir novas funcionalidades, sem ser necessário alterar uma quantidade incomportável de código anteriormente escrito. Esta necessidade implicava conceber um novo protótipo totalmente de raiz, com uma filosofia e arquitetura completamente diferente da adotada até então.

Foi também este o momento escolhido para fazer uma avaliação, do ponto de vista de usabilidade, do protótipo em utilização. Na nova versão (ver. 2.0), era importante manter os pontos fortes de interação da versão anterior mas, ainda mais importante, era necessário resolver os problemas identificados.

Este processo de desenvolvimento do protótipo (versão 2) deu lugar ao desenvolvimento de uma tese de mestrado por parte de um dos elementos da equipa de investigação (F. L. Zamfir, 2007), sempre que oportuno serão referidos os contributos, e impacto, deste mesmo trabalho de mestrado neste projeto de investigação.

### 3.6.2. Aplicação do protótipo

O protótipo, nas suas várias fases de desenvolvimento, foi aplicado a vários contextos infocomunicacionais para efeitos de teste e conseqüente validação do modelo proposto. Um dos testes mais significativos consistiu no acompanhamento de uma experiência em ambiente controlado. Tratou-se da monitorização e análise do processo de comunicação e informação mediados por tecnologia *web* num repositório documental de material de apoio a uma unidade curricular ministrada na Universidade de Aveiro.

Para otimizar o registo de informação da experiência de uso de cada um dos utilizadores do repositório documental e tornar a experiência tão profícua quanto possível, foi criado um *site web* exclusivo para este efeito. O servidor onde estava alojado foi configurado para registar (nos seus ficheiros de *log*) toda a informação possível relacionada com a atividade dos utilizadores neste *site* e as páginas do próprio *site* foram submetidas a ligeiras alterações por forma a permitir identificar, de uma forma ainda mais detalhada, cada ação do utilizador enquanto utilizava o *site* em causa.

Foram ainda feitas experiências noutros contextos de comunicação e informação, que também serão descritas ao longo deste trabalho. No contexto do relato destas experiências empíricas e de avaliação do protótipo (para as diversas versões do mesmo), com a finalidade de validar o modelo proposto nesta tese, serão apresentadas as técnicas de recolha de dados de avaliação, o seu tratamento e discussão.

## 4. Pergunta de Investigação

Considerando a pertinência da dimensão humana, quanto aos seus comportamentos na definição dos percursos de interação nos repositórios de informação/documentais em plataformas digitais e considerando ainda a relevância do impacto da interface gráfica nestes repositórios, nasce a pergunta:

É possível conceber um modelo de análise do fluxo de interação infocomunicacional, de forma integrada, síncrona e holística, tanto na interface gráfica, como na estrutura dos repositórios de informação, para orientar a conceção, avaliação e gestão otimizada de repositórios *web*?

# **Capítulo II.**

## **LEVANTAMIENTO BIBLIOGRÁFICO**

# 1. Introdução

Conforme já foi mencionado no Capítulo I, o presente trabalho, com características transdisciplinares, depende de várias áreas de saber como as Ciências da Comunicação, Ciência da Informação e Ciências da Computação, só para referir as mais dominantes. Estas áreas estruturam as várias secções deste capítulo onde será discutida alguma da bibliografia mais relevante destas mesmas áreas no contexto da problemática deste trabalho.

Na primeira secção deste capítulo será dada uma ideia dos projetos/sistemas e aplicações existentes, em termos comerciais e de investigação, na linha deste trabalho.

Nas secções seguintes será feito um levantamento bibliográfico dos aspetos relevantes de cada uma das seguintes áreas:

- Comunicação organizacional mediada por computador, onde se fará um pequeno levantamento do percurso histórico e do enquadramento organizacional das tecnologias da informação e comunicação;
- Visualização de informação, onde se exploram algumas classificações taxonómicas propostas por alguns autores, bem como métodos de visualização de vários tipos de informação;
- Usabilidade de *sites web*, com especial atenção às questões relacionadas com plataformas e *sites web*, aplicação de métricas e utilização de estudos de usabilidade neste contexto;
- Ferramentas de análise de *log's* de *sites web*, onde se analisam alguns produtos presentes no mercado e as suas abordagens.

## 2. O que já existe relacionado com o tema desta tese

A interação humano-computador tem sido objeto de estudo desde o surgimento dos primeiros computadores, mais acentuadamente desde a comercialização e posterior massificação do conceito de “computador pessoal”. Contudo há ainda muitos aspetos que carecem de um estudo mais aprofundado. A *internet* é uma proposta tecnológica relativamente recente, porém o seu conteúdo é cada vez mais abrangente. A *internet* é cada vez mais usada como plataforma para a implementação de sistemas de informação

e comunicação, tanto no interior como na relação com o exterior, de um número crescente de organizações.

Face a esta evolução, torna-se imprescindível estudar a interação dos utilizadores com as tecnologias e serviços *internet*, nomeadamente com os portais e *sites web* e as suas respetivas páginas ou interfaces.

### Abordagem tradicional

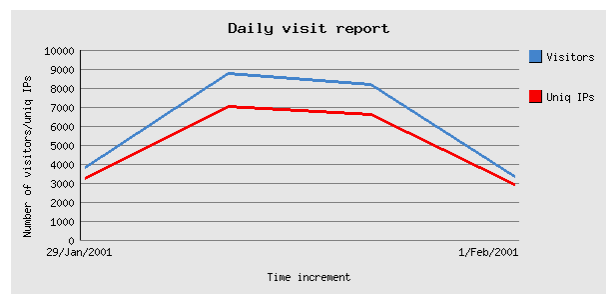
A generalidade das ferramentas de análise de interação em contextos *web*, existentes no mercado, permite fazer uma análise da utilização de um *site web* pelos seus utilizadores com recurso a tabelas e gráficos. Em geral, estas ferramentas conseguem avaliar o número de páginas visitadas, o número de sessões e/ou utilizadores únicos, alguns detalhes sobre os visitantes (como o endereço IP, sistema operativo ou navegador utilizados), tráfego de dados produzido e eventualmente alguma informação estatística adicional. Existem, contudo, algumas abordagens alternativas, tipicamente aplicadas a cenários específicos e que serão apresentadas mais tarde, neste capítulo, juntamente com alguns exemplos que refletem a sua natureza muito particular.

Uma das formas de se perceber a abordagem tradicional das aplicações de análise de *sites web* é através dos seus paradigmas visuais de representação de informação. Tomando isso em consideração, e a título de exemplo, foram compiladas imagens com algumas das formas mais comuns de representação dos resultados obtidos. Basicamente são usadas tabelas (Figura II-1), gráficos simples (Figura II-2) ou de barras (Figura II-3 e Figura II-5), podendo por vezes misturar mais que um tipo (Figura II-4), ou recorrer a outras formas como digramas tipo queijo (Figura II-6).

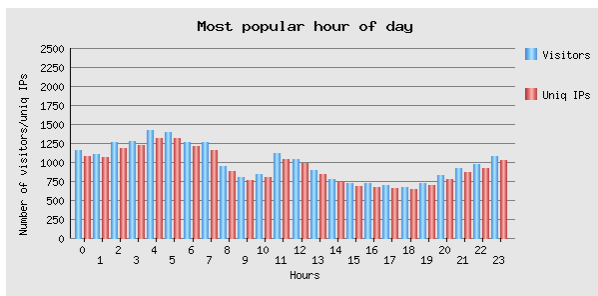
Estes exemplos foram retirados de aplicações comerciais disponíveis no mercado e que serão abordadas com mais detalhe neste capítulo.

Domain Name	Hits	Users	Bandwidth
proxy.aol.com	95	7	7.01 megabytes
strako.net	58	1	1.44 megabytes
ithaca.edu	56	1	134.70 kilobytes
shadow.net	51	1	1.22 megabytes
195.209.39.173	51	1	1.08 megabytes
ix.netcom.com	48	1	173.24 kilobytes
flash.net	45	4	3.86 megabytes
mt.us.ibm.net	44	2	1.24 megabytes
193.78.121.12	43	1	145.11 kilobytes
dialup.mindspring.com	43	3	419.33 kilobytes
tvp.com.pl	39	1	140.92 kilobytes
ts-2.dc.idt.net	31	1	2.10 megabytes
dialsprint.net	31	2	115.71 kilobytes
greenville-01.sc.dial-access.att.net	28	1	131.59 kilobytes
new-york-01.ny.dial-access.att.net	28	1	1.12 megabytes
n78.dip.aha.edu	27	1	1.14 megabytes
tn3.tampa.fl.gt.uu.net	26	1	109.21 kilobytes
sag-mi.concentric.net	25	1	84.79 kilobytes
206.42.107.51	25	1	83.02 kilobytes
birmingham-01.al.dial-access.att.net	24	1	1.15 megabytes
195.209.39.160	24	1	1.63 megabytes
mediawars.or.com	23	1	1.13 megabytes
austin.texas.net	22	1	168.64 kilobytes

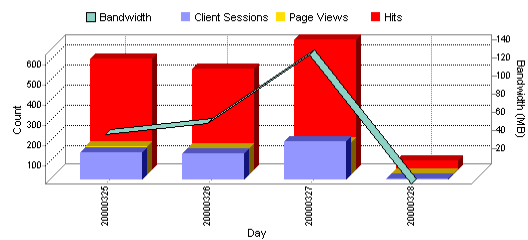
**Figura II-1:** Exemplo de uma tabela de texto. Informação obtida utilizando o programa FastStatsAnalyser v2.77 da Mach5 Software, Inc.



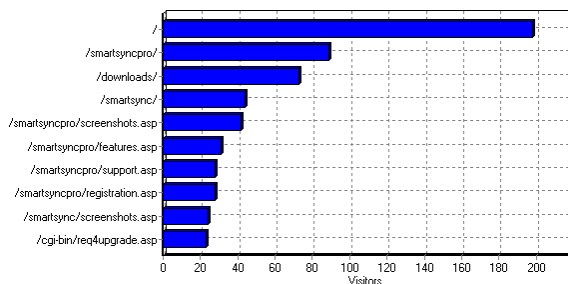
**Figura II-2:** Exemplo de um gráfico simples (duas séries de valores em simultâneo). Informação obtida utilizando o programa FastStatsAnalyser v2.77 da Mach5 Software, Inc.



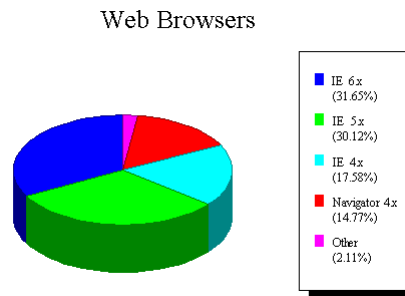
**Figura II-3:** Exemplo de um digrama de barras (duas séries de valores em simultâneo). Informação obtida utilizando o programa FastStatsAnalyser v2.77 da Mach5 Software, Inc.



**Figura II-4:** Exemplo de uma mistura de um gráfico / diagrama de barras. Informação obtida utilizando o programa FastStatsAnalyser v2.77 da Mach5 Software, Inc.



**Figura II-5:** Exemplo de um diagrama de barras. Informação obtida utilizando o programa FastStatsAnalyser v2.77 da Mach5 Software, Inc.



**Figura II-6:** Exemplo de um diagrama tipo queijo. Informação obtida utilizando o programa FastStatsAnalyser v2.77 da Mach5 Software, Inc.

## Projetos de Investigação

Existem alguns grupos de investigação com preocupações e objetivos que intersectam a problemática deste trabalho. Destacam-se 3 exemplos, pela proximidade da abordagem e das questões de investigação e desenvolvimento.

No artigo “*The Scent of a Site*” de um grupo de investigação da *Xerox Palo Alto Research Center* (Chi et al., 2000) pode encontrar-se uma mesma abordagem, com objetivos semelhantes a esta tese. Há, contudo, uma diferença importante: no artigo em causa, o paradigma de interface da análise não usa como referência o espaço *web* inicial nem o toma em consideração no contexto de representação de resultados. Contudo, as perguntas a que o sistema responde são muito semelhantes às desta tese (Chi et al., 2000):

- **Site em geral.** Qual é fluxo de tráfego geral atual? Quais são os caminhos, efetivos e previstos, adotados durante a navegação? Qual o desempenho do *site* em termos de facilidade e custo de acesso à informação?

- **Página em concreto.** De onde vêm os visitantes (caminho que percorrem)? Onde vão efetivamente? Que outras páginas estão relacionadas?
- **Utilizadores.** Quais os interesses do visitante (reais ou simulados) nesta página? Onde se pensa que deverão ser conduzidos, atendendo aos seus interesses? Será que os dados recolhidos coincidem com os previstos e porquê? Qual o custo (em termos de tempo de download) da navegação para esses utilizadores?

Outra abordagem, de investigadores chineses (Zhou et al., 2001), tem muitas semelhanças com esta tese mas apenas recai sobre a análise da estrutura do *site* com base nos padrões de navegação dos visitantes. Tal como nesta tese, também usam *web logs* para extrair a informação que necessitam e apresentam um conjunto de critérios para detetar e legitimar uma “sessão” de visita de um utilizador.

Para terminar esta secção, falta referir um trabalho desenvolvido no Laboratório de Interação Humano-Computador da Universidade de Maryland (Hochheiser & Shneiderman, 1999) onde se apresenta um sistema que procura formular respostas a questões similares às desta tese. Tal como no sistema proposto nesta tese, usa os ficheiros de *web log* como fonte de informação base. Refere ainda que a modelação e caracterização de “*web site access patterns*” é uma área ativa de pesquisa e investigação. No contexto desta tese, a expressão “*web usage patterns*” é preferida por dar um maior enfoque na dimensão de utilização e não exclusivamente de acesso às várias páginas. Um dos problemas do método descrito tem a ver com a metáfora visual de análise diferente e descontextualizada da informação original que leva a que se perca o contexto, nomeadamente holístico visual (imagem), de página do *site* assim como da sua relação com outras páginas e/ou documentos.

### 3. Comunicação organizacional mediada por computador

Arménio Rego em (Rego, 2003) inserido no livro (Cunha et al., 2003) afirma: “A comunicação é uma condição *sine qua non* da vida social e, a *fortiori*, da vida organizacional. Sem comunicação, não pode haver organização, gestão, cooperação, motivação, vendas, oferta ou procura, *marketing*, ou processos de trabalho coordenados. [...] De facto, uma organização humana é simplesmente uma rede comunicacional: se a comunicação falha, uma parte de estrutura organizacional também falha. [...] Dado que os indivíduos gastam perto de 70% do seu tempo comunicando – escrevendo, lendo, falando, ouvindo – parece razoável concluir que uma das forças mais inibidoras do desempenho bem-sucedido dos grupos é a ausência de comunicação eficaz.”

McPhee (McPhee & Zaug, 2009) sublinha bem a pertinência desta necessidade ao apresentar o seu modelo assente em quatro fluxos da comunicação para a constituição das organizações, que designa de “*The Communicative Constitution of Organizations*”.

Não é de estranhar, portanto, que o computador apareça como um instrumento mediador e supostamente facilitador da comunicação, tanto nas organizações como na sociedade em geral.

Arménio Rego (Rego, 2003) refere também as dificuldades do estudo e compreensão desta temática: “O fascínio exercido pela comunicação eletrónica (CE) é quase tão grande quanto a dificuldade em recolher na literatura especializada elementos consistentes acerca das suas características e efeitos. [...] O facto de muitas pesquisas terem sido realizadas em meios laboratoriais e de os seus resultados não serem pacificamente generalizáveis às organizações, contribui para esse adensamento das dificuldades do campo de estudo. [...] Daqui decorre que a realidade organizacional é apenas parcialmente refletida nas pesquisas.”

Lança, ainda, luz sobre alguns temas pertinentes relacionados com a comunicação eletrónica (Rego, 2003):

- a) A CE é a “mãe” da **organização virtual**: um conjunto de entidades geograficamente dispersas, ligadas por formas eletrónicas de comunicação.
- b) A CE facilita a emergência do **teletrabalho** e oferece novas oportunidades de relacionamento com clientes e fornecedores.
- c) A CE contribui para o incremento no **volume** das comunicações, ainda que haja o risco de sobrecarga de informação, de existência de **mensagens-lixo** e de perda de **eficiência** na comunicação
- d) A CE promove a **rapidez** das comunicações, mas não se pode esquecer o fator humano que lhe está associado.
- e) Quando a comunicação é mediada por computadores, as mensagens são passíveis de **memorização** para análise, manuseamento ou alteração posterior.
- f) Não é garantido que a CE seja mais contributiva da **eficácia** e do desempenho organizacional do que outras formas de comunicação, mas é inquestionável o seu contributo para a eficácia em diversas atividades.
- g) Relativamente à comunicação cara-a-cara, os meios de CE geralmente não comportam sinais não-verbais como o contacto visual, o tom de voz, as expressões faciais etc., o que pode dificultar a regulação da interação, a perceção acerca dos parceiros de comunicação ou a sensibilidade para o contexto social.
- h) A CE induz uma maior **equalização das participações** dos vários intervenientes.
- i) Contudo não eliminam os constrangimentos relacionados com a hierarquia de *status* na organização.
- j) A resolução de problemas **não** é mais rápida quando mediada eletronicamente.



- k) A **identificação** das pessoas com a organização pode esmorecer devido à separação geográfica. Contudo a CE (*e-mail* por exemplo), ao facilitar a partilha de opiniões, pode também contribuir para o incremento do sentido de pertença.
- l) A **confiança** é uma condição necessária em qualquer relação social harmoniosa, que necessita do “toque”, proximidade, contatos e laços pessoais. Isso pode justificar a necessidade que muitos colaboradores de organizações virtuais sentem de se encontrar pessoalmente.

No entanto a questão da tecnologia nas organizações não é pacífica. Basta olhar para os pontos de vistas que os vários departamentos de uma organização têm em relação a esta matéria para se perceber que se trata de um assunto tudo menos consensual (Andriole, 2004). Gestores, comerciais ou informáticos podem ter visões dispares da importância da tecnologia (de informação e comunicação) para a organização onde se inserem. Enquanto uns podem considerar que se trata de algo indispensável à própria sobrevivência da organização, outros podem considerar que não passam de luxos ou mesmo algo que não traz qualquer mais-valia para a organização (Andriole, 2004).

Um sistema de informação e comunicação interna/externa de determinada organização (vulgo *intranet/extranet*) consiste numa rede de computadores que usa a tecnologia e os protocolos da *internet* para, de forma segura, partilhar qualquer parte da informação da organização, serviços de comunicação ou sistemas operativos de rede dentro dessa organização. O termo *intranet* é usado em contraponto com *internet*, uma rede global entre organizações, referindo-se a uma rede dentro de uma organização. Por vezes, o termo *intranet* refere-se somente ao *site web* interno da organização, mas frequentemente refere-se a uma parte mais extensa da infraestrutura tecnológica e de serviços infocomunicacionais da organização. Pode albergar múltiplos *sites web* privados e ser o ponto central da comunicação e colaboração internas. Numa *intranet* podem ser encontrados vários protocolos bem conhecidos da *internet*, tais como HTTP (serviços *web*), SMTP (email) ou FTP (transferência de ficheiros). As tecnologias da *internet* são frequentemente aplicadas para fornecer uma interface mais moderna com maior acessibilidade e usabilidade a sistemas de informação proprietários que alojam informação corporativa.

Uma *intranet* pode ser entendida como um análogo privado de *internet*, ou como uma extensão privada da *internet* confinada a uma organização. Os primeiros *sites* e páginas de *intranets* começaram a aparecer em organizações por volta de 1994 (Jacob Nielsen & Sano, 1994) (Hinrichs, 1997). Ainda que sem reconhecimento oficial, o termo *intranet* tornou-se lugar-comum entre organizações pioneiras como universidades e empresas de tecnologia por volta de 1991.

Pode-se também fazer um contraponto entre *intranets* e *extranets*. Enquanto as *intranets* estão geralmente restritas aos funcionários de uma determinada organização, as *extranets* podem também ser acedidas por clientes, fornecedores ou outros parceiros

aprovados (Callaghan, 2002). As extranets estendem a rede privada para a *internet* com cuidados especiais na autenticação, autorização e monitorização.

Em qualquer organização, as *intranets* são protegidas dos acessos exteriores não autorizados por meio de uma *gateway* e *firewall*. Para companhias mais pequenas, as *intranets* podem ser criadas usando simplesmente gamas de endereços IP privados, como os da gama 192.168.\*.\*. Nestes casos, a *intranet* só pode ser diretamente acedida a partir de um computador na rede local; contudo, as organizações podem fornecer acesso a funcionários remotos usando uma rede virtual privada (VPN). Também podem ser usadas outras medidas de segurança como autenticação dos utilizadores e encriptação.

Alternativamente, o domínio da *intranet* pode ser publicamente acessível, mas os utilizadores têm que se registar (fazer *login*) antes de poderem ter acesso à maior parte do conteúdo.

A *intranet* pode definir-se numa perspetiva técnica ou organizacional (Hinrichs, 1997):

- **Definição Técnica:** Uma *intranet* consiste num ambiente computacional heterogéneo que liga diferentes plataformas de hardware, ambientes de sistemas operativos e interfaces de utilizador de maneira a comunicar, colaborar, transacionar e inovar perfeitamente.
- **Definição Organizacional:** Uma *intranet* é uma organização que aprende, capaz de integrar as pessoas, processos, procedimentos e princípios para formar uma cultura intelectualmente criativa dedicada a implementar uma total eficácia organizacional.

Para efeitos de clarificação das possíveis correlações de organização e tecnologia é relevante manter presente que para a concretização da missão de uma organização é determinante considerar-se as pessoas, os processos e os procedimentos. É neste contexto que as tecnologias se inserem e devem proporcionar mais eficiência aos processos mantendo a eficácia num quadro acrescido de experiência e satisfação de uso. As semelhanças com a definição de usabilidade ditada por vários autores (Jordan, 2001; Jakob Nielsen, 1993) e registada na norma ISO 9241 – Parte 11 (ISO, 1998), são naturais pois o objetivo desta norma é garantir que os instrumentos tecnológicos, com feedback através de um campo visual, sejam concebidos para corresponder de forma otimizada aos propósitos a que se destinam, num contexto de uso que envolve o ser humano.

## 3.1. Percurso histórico

Para um entendimento mais aprofundado do conceito de *intranet* apresenta-se nesta secção um breve percurso pela história moderna da computação (Ceruzzi, 2003), nomeadamente com enfoque na *internet*, e da tecnologia que lhe está associada.

Nesta secção começa por se fazer um breve levantamento histórico do surgimento da *internet* e sua evolução até ao surgimento da *World Wide Web*. No contexto das organizações surgiu uma outra evolução, as *intranets*, que será também abordada.

### **Internet**

Desde muito cedo, na história da computação, que se ligaram computadores entre si, primeiro ponto a ponto e mais tarde formando redes. O passo seguinte consistiu na interligação dessas redes locais. A *Advanced Research Projects Agency* (ARPA), mais tarde rebatizada de *Defense Advanced Research Projects Agency* (DARPA)<sup>2</sup>, encomendada pelo Departamento de Defesa dos Estados Unidos da América, foi pioneira e responsável por conceber o que se tornaria a tecnologia base para a *internet*, sobretudo depois da adoção dos protocolos de comunicação TCP/IP e de os ter tornado públicos no pós-guerra fria.

Desde o momento em que a *internet* emergiu das suas raízes na ARPA, que começou a alterar-se. Os serviços infocomunicacionais iniciais resumiam-se à possibilidade de estabelecer uma sessão de trabalho numa máquina remota (Telnet<sup>3</sup>), à transferência de ficheiros/documentos de uma máquina para outra (FTP<sup>4</sup>) e ao envio de correio (e-mail<sup>5</sup>). Rapidamente começaram a surgir outros serviços como fóruns/grupos de discussão (News<sup>6</sup>) e *bulletin-board systems* (BBS<sup>7</sup>). Com o crescimento da *internet*, a quantidade

---

<sup>2</sup> DARPA: <http://www.darpa.mil>

<sup>3</sup> Telnet é um protocolo cliente-servidor. Tipicamente, este protocolo é usado para estabelecer uma ligação à porta 23 do *Transmission Control Protocol* (TCP), onde uma aplicação de servidor Telnet (telnetd) está à escuta.

<sup>4</sup> *File Transfer Protocol* (FTP) é um protocolo de rede standard usado para transferir ficheiros de uma máquina para outra sobre redes baseadas em TCP, como é o caso da *internet*.

<sup>5</sup> *Electronic mail*, referido frequentemente como email ou e-mail, é um método de troca de mensagens digitais de um autor para um ou mais destinatários.

<sup>6</sup> *NetNews* ou *Usenet* é um sistema de discussão na *internet* distribuído pelo mundo inteiro.

<sup>7</sup> Um *Bulletin Board System*, ou BBS, é um sistema computacional a correr software que permite que os utilizadores se liguem via telefone e se validem no sistema usando um programa terminal. Uma vez ligado, um utilizador pode efetuar funções como enviar ou descarregar ficheiros, ler notícias, trocar mensagens com outros utilizadores via email, ou mesmo *chatting* direto. Muitos BBS's também oferecem jogos on-line, em que os utilizadores podem competir entre si.

de informação disponível era cada vez maior, mas com um problema grave: parecia estar-se perante uma grande biblioteca sem qualquer catalogação que ajudasse a encontrar o pretendido.

No final da década de 1980 surgiu o serviço WAIS<sup>8</sup> que, em conjugação com o serviço Gopher<sup>9</sup>, que surgiu em 1990, pretendiam permitir que mesmo utilizadores com pouca experiência com computadores pudessem procurar e encontrar informação/documentos. Contudo o recurso a listas de menus de texto organizados de forma hierárquica tinha muitas limitações.

Não demorou muito tempo para que surgisse a *World Wide Web* (WWW<sup>10</sup> ou simplesmente *web*), inventada num lugar inesperado: no Laboratório de física de partículas Europeu (CERN<sup>11</sup>) localizado na fronteira entre a França e a Suíça. Neste contexto, Tim Berners-Lee, um investigador do CERN, apresenta uma proposta em 1989 (Berners-Lee, 1989) e em 1990 concebeu o protótipo original da *web* (Agnello & Levkowitz, 2015; Berners-Lee, 2016; Velez et al., 2015). Uma das suas preocupações tinha a ver com a possibilidade de ligação entre diferentes tipos de sistemas e software. Criou um identificador universal de recursos (*URL*<sup>12</sup>) que permitisse apontar para qualquer documento (ou qualquer outro tipo de recurso) no universo da informação. Para atingir estes objetivos criou um novo protocolo de transferência de hipertexto mais sofisticado (*HTTP*<sup>13</sup>) que era mais rápido e tinha mais funcionalidades que os precedentes (como o *FTP*). Finalmente, definiu o *HTML*<sup>14</sup> (uma linguagem de marcação

---

<sup>8</sup> *Wide Area Information Servers* ou WAIS é um sistema de pesquisa de texto do tipo cliente-servidor que usa o standard ANSI Z39.50 "*Information Retrieval Service Definition and Protocol Specifications for Library Applications*" para pesquisar índices de bases de dados em computadores remotos.

<sup>9</sup> O protocolo *Gopher* é uma camada de aplicação do protocolo TCP/IP desenhado para distribuir, pesquisar e recolher documentos sobre a *internet*. Com um design muito orientado por listas de menus, o protocolo *Gopher* apresentava-se como uma alternativa à *World Wide Web* (WWW ou *web*) nos seus primeiros tempos, contudo o HTTP tornou-se o protocolo dominante. O ecossistema *Gopher* é frequentemente referido como o predecessor efetivo da *World Wide Web*.

<sup>10</sup> A *World Wide Web* (WWW, W3 ou *web*) é um sistema de interligação de documentos de hipertexto acedidos via *internet*. Com um navegador *web*, é possível ver páginas que podem conter texto, imagens, vídeos e outros elementos multimédia e navegar entre páginas via hiperligações.

<sup>11</sup> O *European Organization for Nuclear Research*, conhecido como CERN, é uma organização internacional cujo propósito consiste em operar o maior laboratório de partículas do mundo, localizado nos subúrbios a noroeste de Génova, na fronteira Franco-Suíça: <http://home.web.cern.ch> ou <http://home.cern>.

<sup>12</sup> O *Universal Resource Locator* (*URL*) é um identificador universal de recursos na *web*.

<sup>13</sup> O *Hypertext Transfer Protocol* (*HTTP*) é um protocolo de aplicação para sistemas de informação hipermédia distribuídos e colaborativos (Fielding et al., 1999).

<sup>14</sup> *HyperText Markup Language* (*HTML*) é a principal linguagem baseada em marcas para mostrar páginas *web* e outra informação que possa ser apresentada num navegador *web*.

de hipertexto) para potencializar o fluxo do hipertexto através da rede. Consegue desta forma uma solução para a divulgação e partilha de informação e documentos dentro do CERN num âmbito alargado e tornando o acesso a estes recursos muito mais eficiente. Atitude que rapidamente se generalizou para todo o mundo ao tornar o HTTP e o HTML públicos.

No início de 1993, Andreessen e Bina tinham escrito um navegador (*browser*) a que viriam mais tarde a chamar *Mosaic*. Este juntava a facilidade de utilização com todas as funcionalidades do hipertexto da *web*. Pela primeira vez foi utilizado o rato para seleccionar itens com uma hiperligação identificados pela sua cor. Outra característica que sobressaiu foi a integração eficaz de texto e imagens. O *Mosaic* acabou por ser portado para várias plataformas (PC, Mac, etc.) e disponibilizado gratuitamente ou mediante um valor nominal.

Em setembro de 1994, a *Netscape* (uma empresa entretanto criada por Andreessen e Jim Clark) lança o seu primeiro navegador que rapidamente suplanta o *Mosaic*. Desde esta altura, várias empresas lançaram navegadores com níveis de sofisticação e funcionalidades crescentes.

## ***Intranet***

O uso da tecnologia *web*, numa forma que hoje se designaria por *intranet*, teve início por volta de 1993 em algumas empresas como Amdahl<sup>15</sup>, Schlumberger<sup>16</sup>, Lockheed<sup>17</sup> e NASA AMES<sup>18</sup> (Hinrichs, 1997). O fenómeno teve início antes de o termo ter surgido para o descrever.

Várias pessoas reclamam a autoria do termo “*intranet*” para descrever o desenvolvimento de uma solução interna cliente/servidor para organizações baseada em tecnologia *web*. Steve Telleen (Hinrichs, 1997; Telleen, 1995) usou o termo “*IntraNet*” em 1994 num artigo interno que escreveu sobre “metodologia da *IntraNet*” na empresa Amdahl.

A primeira aparição do termo numa publicação escrita comercial ocorreu num artigo de Stephen Lawton sobre *intranets* na revista *Digital News & Review* em Abril de 1995

---

<sup>15</sup> A Amdahl foi uma empresa da área das tecnologias de informação que em 1997 passou a ser uma subsidiária da Fujitsu: <http://www.fujitsu.com>.

<sup>16</sup> A Schlumberger é uma empresa da área da exploração petrolífera: <http://www.slb.com>.

<sup>17</sup> A Lockheed foi uma empresa da área da construção aeronáutica que em 1995 se fundiu com a Martin Marietta dando origem à Lockheed Martin: <http://www.lockheedmartin.com>.

<sup>18</sup> A NASA (*National Aeronautics and Space Administration*: <http://www.nasa.gov>), agência espacial americana, tem vários centros de investigação e desenvolvimento, nomeadamente em AMES, localizado no coração do *Silicon Valley* da Califórnia, nos Estados Unidos da América. Este centro chama-se NASA AMES: <http://www.nasa.gov/centers/ames/home/index.html>.

(Hinrichs, 1997; Lawton, 1995). Neste artigo o autor discute a publicação de páginas *web* e a instalação de servidores Telnet e FTP pelas companhias pertencentes ao *Fortune 1000*. Os pioneiros apontados eram Boeing<sup>19</sup>, Schlumberger Ltd, Weyerhaeuser Corp.<sup>20</sup>, Sun Microsystems<sup>21</sup> e Digital Equipment Corp<sup>22</sup>. As vantagens apontadas eram o baixo custo de implementação, a facilidade em escrever HTML e o acesso a vários tipos de documentos on-line como manuais dos funcionários, material de investigação e páginas pessoais (Hinrichs, 1997).

## Evolução da *intranet*

As *intranets* começaram como páginas *web* básicas partilhadas por um grupo restrito de pessoal técnico. Estes deram formação a mais algumas pessoas e rapidamente se percebeu a sua amplitude. Nasceram equipas multifuncionais e no decurso deste processo surgiu uma tomada de consciência geral da alteração de paradigma e uma nova forma de negócio substituiu a atitude empresarial prevalente. Com cada fase de evolução, foram sendo descobertas novas capacidades. Algumas capacidades abriram novas ideias de inovação e o processo repetiu-se na fase evolutiva seguinte (Hinrichs, 1997).

Hinrichs (Hinrichs, 1997) identifica cinco níveis na evolução da *intranet*. Muitas organizações, nomeadamente empresas e instituições, por exemplo as universidades, passaram apenas pelos primeiros três níveis, o que fica aquém do que pode ser perspectivado no quarto e quinto nível. Os níveis são os seguintes: *intranets* básicas, *intranets* para publicação de repositórios, *intranets* para colaboração, *intranets* para transações e *extranets*. Pode ver-se uma breve descrição de cada um desses níveis nos Quadro II-1 a Quadro II-5, abaixo.

**Quadro II-1:** Nível 1: *Intranet* básica (adaptado de (Hinrichs, 1997)).

<b><i>Intranet</i> básica</b>	
<i>Descrição:</i>	Pequeno <i>site web</i> (+50 páginas) com texto, <i>link</i> e gráficos.
<i>Arquitetura:</i>	Baseada em TCP/IP, servidor central e navegador. Necessidade de software do lado do servidor e de ferramentas de edição.

<sup>19</sup> A Boeing é uma empresa da construção aeronáutica: <http://www.boeing.com>.

<sup>20</sup> A Weyerhaeuser é uma empresa madeireira: <http://www.weyerhaeuser.com>.

<sup>21</sup> A Sun Microsystems é uma empresa fabricante de computadores, semicondutores e software que foi adquirida em 2009 pela Oracle: <http://www.oracle.com/sun/>.

<sup>22</sup> A Digital Equipment Corp (também conhecida como DEC) foi uma empresa da indústria de computadores. Em 1998 foi adquirida pela Compaq que posteriormente, em 2002, se fundiu com a Hewlett-Packard (HP): <http://www.compaq.com> e <http://www.hp.com>.

<i>Pessoas e processos:</i>	Todo o design e desenvolvimento efetuado por uma única pessoa ou um pequeno grupo que tem ainda a responsabilidade de dar formação aos restantes elementos da organização.
<i>Mais-valia:</i>	Presença on-line.
<i>Custos:</i>	Mínimo relacionado com hardware e software.
<i>Limitações:</i>	Capacidade, aptidão.

**Quadro II-2:** Nível 2: *Intranet* para publicação de repositórios (adaptado de (Hinrichs, 1997)).

<b><i>Intranet</i> para publicação de repositórios</b>	
<i>Descrição:</i>	Produção e acesso a documentos e recursos da empresa. <i>Site</i> razoavelmente grande, ainda que estático.
<i>Arquitetura:</i>	<i>Intranet</i> básica mais software de publicação de gestão documental, motor de pesquisa e integração de bases de dados.
<i>Pessoas e processos:</i>	Exige pessoas que procedam à publicação, edição e escrita. Necessita de processos de submissão do trabalho e de um <i>site web</i> centralizado para funcionar como ponto único de contacto.
<i>Mais-valia:</i>	Acesso à informação da organização.
<i>Custos:</i>	Com funcionários e ferramentas. Necessidades acrescidas de equipamento e de processamento.
<i>Limitações:</i>	Precisão, exatidão.

**Quadro II-3:** Nível 3: *Intranet* para colaboração (adaptado de (Hinrichs, 1997)).

<b><i>Intranet</i> para colaboração</b>	
<i>Descrição:</i>	Grupo de <i>sites</i> interdependentes. Páginas de departamento. Página de entrada ( <i>homepage</i> ) centralizada. Conjunto de ferramentas avançadas.
<i>Arquitetura:</i>	<i>Intranet</i> completa com serviços de diretório, ficheiros, impressão e correio. Integrada com gestão de grupos, software de gestão de projeto e sistemas de correio eletrónico.
<i>Pessoas e processos:</i>	Requer uma infraestrutura sofisticada. Uma equipa multidisciplinar com gestor de <i>sites</i> , responsável de segurança, programadores de aplicações e produtores de conteúdos. Precisam de modelos, linhas orientadoras e princípios para desenvolver as suas atividades.
<i>Mais-valia:</i>	Ciclos de desenvolvimento reduzidos.

<i>Custos:</i>	Crescimento exponencial nos custos operacionais. Custos de desenvolvimento depende dos pivots.
<i>Limitações:</i>	Manutenção.

**Quadro II-4:** Nível 4: *Intranet* para transações (adaptado de (Hinrichs, 1997)).

<b><i>Intranet</i> para transações</b>	
<i>Descrição:</i>	Integrada com sistemas e aplicações de rede proprietários ( <i>legacy</i> ). Usa Java, ActiveX e outras APIs ( <i>Application Programming Interfaces</i> ).
<i>Arquitetura:</i>	Serviço de <i>intranet</i> completo. Integração segura com bases de dados proprietárias e todas as aplicações.
<i>Pessoas e processos:</i>	Necessidade de novas competências para a maioria dos processos de trabalho, a todos os níveis. Requer a integração dos processos com as ferramentas de rede, com bases de dados e com as métricas. Necessária formação para orientar a força de trabalho.
<i>Mais-valia:</i>	Eficácia organizacional.
<i>Custos:</i>	Integração completa de todos os custos "MIS" ( <i>Management Information Systems</i> ).
<i>Limitações:</i>	Segurança.

**Quadro II-5:** Nível 5: *Extranet* (adaptado de (Hinrichs, 1997)).

<b><i>Extranet</i></b>	
<i>Descrição:</i>	<i>Intranet</i> completa, capaz de permitir a utilização plena da rede empresarial a partir de fora dos limites da empresa.
<i>Arquitetura:</i>	Serviço de <i>intranet</i> completo complementado com software de encriptação que comunica com os parceiros externos.
<i>Pessoas e processos:</i>	Processos definidos para executar tarefas. Sistemas com métricas fortes para determinar a eficácia. Processos integradores e parceria aumentada.
<i>Mais-valia:</i>	Criação de bem-estar institucional.
<i>Custos:</i>	Largura de banda. Parcerias. Segurança.
<i>Limitações:</i>	Competitividade.



Mas antes disto já o computador pessoal prestava um auxílio administrativo nas organizações ainda que sem um *standard* de rede inter-organizacional, por vezes nem intra-organizacional. Existiam várias redes e tecnologias ao serviço das organizações tornando a comunicação entre elas, e dentro delas, um caso crítico de funcionalidade. Pela primeira vez na história do fenómeno infocomunicacional nas organizações, aparece uma tecnologia capaz de garantir a integração e coesão de pessoas e recursos de forma a melhorar consideravelmente a eficiência das sinergias existentes (e potenciar outras) para o desenvolvimento da missão inerente a cada organização.

## 3.2. As tecnologias e o seu enquadramento organizacional

Quando se pretende analisar uma *intranet* corporativa, em toda a sua extensão, é importante observar (Hinrichs, 1997):

- Arquitetura da *intranet* (hardware, componentes de software e processos);
- Infraestrutura da *intranet* (ferramentas, pessoas e processos);
- Interatividade / interoperabilidade da *intranet* (ligação a sistemas proprietários, a bases de dados e aos clientes);
- Segurança da *intranet* (*firewall*, políticas, risco);
- Design da *intranet* (quatro camadas: criação do *site*, criação do conteúdo, manutenção do *site*, análise do *site*).

Esta lista fornece uma visão global dos problemas chave que devem ser endereçados quando se pretende implementar uma *intranet*. Permite fazer uma análise da exequibilidade e determinar a melhor estratégia de desenvolvimento da *intranet*. No contexto deste trabalho apenas uma parte destas preocupações será analisada, em concreto, algumas das tecnologias/serviços envolvidos e que serão abordados nas subsecções abaixo.

As *intranets* utilizam várias tecnologias, de forma mais ou menos integradas que, em conjunto, fornecem um leque de funcionalidades essenciais para as organizações. Estas tecnologias podem ser agrupadas em função de vários critérios temporais, espaciais, funcionais, entre outros. Contudo, para este trabalho, não será utilizada uma organização formal, optando-se por um agrupamento baseado na funcionalidade prática de cada uma das tecnologias no contexto de uma organização.

### 3.2.1. Serviços de comunicação assíncrona

Abaixo são listadas algumas tecnologias que permitem a comunicação assíncrona entre elementos, individuais ou em grupos, de uma organização.

#### Correio eletrónico (email)

O serviço de correio eletrónico foi um dos primeiros serviços digitais a ser criados e disponibilizados. Surgiu nos primórdios dos computadores, muito antes do surgimento da *internet*. A sua forma de funcionamento não sofreu grandes alterações, no entanto recebeu algumas evoluções ao longo do tempo que lhe trouxeram novas funcionalidades, tornando-o ainda mais pático e funcional, mantendo-o como um dos serviços mais utilizados no mundo, apesar da sua longa idade.

#### Objetivos com enquadramento institucional

O serviço de email permite a comunicação assíncrona entre agentes da organização, na forma de troca de mensagens textuais. Permite também enviar ficheiros, sob a forma de anexos, o que também possibilita a veiculação de informação ou documentos de natureza multimédia.

#### Caracterização técnica e funcional

Este serviço permite a composição e envio de uma mensagem, para vários destinatários com indicação de perfil de receção (to, cc e bcc), de um registo de assunto, de prioridade da mensagem, entre outros detalhes. Originalmente só era permitido texto ASCII<sup>23</sup> simples, foi posteriormente expandido para suportar texto e formatação mais sofisticados, bem como conteúdo MM, com recurso a anexos suportados por extensões como MIME<sup>24</sup>.

O serviço de email é assegurado por servidores que usam os protocolos SMTP<sup>25</sup>, POP<sup>26</sup> e/ou IMAP<sup>27</sup>, MAPI<sup>28</sup> e HTTP/HTTPS em muitos casos.

---

<sup>23</sup> ASCII (*American Standard Code for Information Interchange*) é um esquema de codificação de caracteres, que servem para representam texto em computadores e outros equipamentos digitais.

<sup>24</sup> MIME (*Multipurpose Internet Mail Extensions*) é standard da *internet* que estende o formato do email para suportar: texto com codificação diferente da ASCII, anexos de imagens, áudio, vídeo, etc., entre outras funcionalidades.

<sup>25</sup> SMTP (*Simple Mail Transfer Protocol*) é um protocolo standard da *internet* para a transmissão de email.

<sup>26</sup> POP (*Post Office Protocol*) é um protocolo standard da *internet* usado por programas clientes de email para obterem os emails de servidores remotos sobre ligações TCP-IP. Atualmente na versão 3 (POP3), está a tornou-se obsoleto com o surgimento do mais avançado IMAP.

<sup>27</sup> IMAP (*Internet Message Access Protocol*) é um protocolo standard da *internet* usado por programas clientes de email para obterem os emails de servidores remotos sobre ligações TCP-IP. Comparativamente ao POP, resolve várias limitações e fornece um conjunto de funcionalidades adicionais.

## Tecnologias existentes

Existem vários servidores de suporte a este serviço e várias aplicações terminais para o seu processamento (receção, leitura, composição e envio).

Estes servidores podem ser suportados em vários sistemas operativos (Microsoft Windows, Apple Macintosh, Linux/Unix, entre outros) com vários serviços de correio (Microsoft Exchange Server, Postfix, sendmail, Exim, entre outros) e ainda disponibilizar serviços adicionais como o serviço *web* (Apache, IIS, nginx, entre outros).

## Vantagens e/ou desvantagens da sua adoção

O aumento inquestionável da eficiência em relação ao suporte em papel faz com que por vezes seja paradoxalmente mal utilizado. Por exemplo a função “*reply*” a uma mensagem de *mailing-list*<sup>29</sup> quando a resposta se destina apenas ao emissor e não ao fórum.

A resposta a uma mensagem que trazia documentos anexos normalmente não sinaliza que a mensagem original tinha documentos anexos, a sua designação e características dos mesmos. Pode não ser necessário, em processo de troca de mensagens, continuar a veicular documentos anexos mas é bastante relevante manter o contexto de comunicação informado sobre as instâncias do processo de comunicação em que ocorreram transmissão de documentos anexos e suas características.

Há a necessidade de “classificar” o assunto das mensagens de forma a otimizar a receção. Por exemplo, caso uma mensagem estivesse “classificada” como “ordem de serviço”, teria um percurso pré-definido no servidor de correio-electrónico institucional (*mailing-lists* e pessoas) com uma dada prioridade. Determinadas “classificações” poderiam mudar (forçar em contexto institucional) a prioridade de leitura de mensagens nas aplicações cliente (terminal de leitura) e até pré organizá-las de forma diferente, em pastas virtuais individuais e institucionais.

A gestão da comunicação assíncrona interna e externa da organização, especificamente o seu registo processual e pesquisa de situação infocomunicacional, passam a ser feitos de forma bastante mais eficiente.

O email pode ser bastante ineficiente se for utilizado como instrumento de conversação por falta de capacidade de representação da pontuação da comunicação. Esta situação

---

<sup>28</sup> MAPI (*Messaging Application Programming Interface*) é uma arquitetura de passagem de mensagens e uma *Application Programming Interface* (API) baseada no *Component Object Model* (COM) do Microsoft Windows. Apesar da MAPI ter sido desenhada para ser independente do protocolo, é usada habitualmente com MAPI/RPC, o protocolo proprietário que o Microsoft Outlook usa para comunicar com o Microsoft Exchange.

<sup>29</sup> As lista de email (*mailling-lists*), permitem o envio de email a um grupo de utilizadores alargado com a simples indicação da lista, sendo o nome da lista substituído automaticamente pelos emails respetivos.

torna-se particularmente problemática quando o email é utilizado com o dispositivo de entrada teclado, situação que traduz a prática habitual. As alternativas de interação com voz melhoram esta situação de falta de eficiência contudo fica sempre por resolver o problema de pontuação. Alguns clientes de email atuais resolveram, em parte, este problema com a adoção de representações hierárquicas das mensagens trocadas considerando o assunto, o emissor-receptor e a hora da mensagem, muitas das vezes em fórum com mais de 2 pessoas.

Outro problema deste serviço tem a ver com o spam<sup>30</sup> que consiste no envio em massa de mensagens não-solicitadas. Frequentemente são utilizados programas que facilitam ou automatizam a obtenção de endereços de email e o envio de mensagens para um grande número de destinatários. Existem vários métodos para um *spammer* obter uma lista de endereços. Entre eles podem encontrar-se programas de interpretação de texto que fazem o varrimento de locais com um número potencialmente elevado de endereços disponíveis, como páginas de *internet*, Fóruns ou registos de DNS<sup>31</sup>. Mesmo quando o *spammer* não tem acesso a esse tipo de programas, pode recorrer a uma variedade de *sites* na *internet* que vendem listas de email prontas a usar.

## Fórum de Discussão

Os fóruns de discussão (do inglês *Internet Forum*) tiveram origem num serviço que tem vindo a perder relevância, tendo atualmente uma utilização bastante reduzida. Este serviço, de nome *Net-News* ou *Usenet*, surgiu em 1980 na Universidade da Carolina do Norte e na Universidade de Duke (Lueg & Fisher, 2003) quase uma década antes da *World Wide Web*. Os utilizadores podem ler e submeter mensagens (chamadas de artigos) em uma ou mais categorias, conhecidas como *newsgroups*. Por se tratar de um serviço que é cada vez menos utilizado, não irá ser detalhado.

Um *Internet Fórum* consiste num *site* de discussão online onde as pessoas podem manter conversas na forma de mensagens publicadas (vBulletin, 2013).

### Objetivos com enquadramento institucional

Os Fóruns permitem a comunicação assíncrona entre agentes da organização. Ao contrário do correio eletrónico, em que as mensagens são transmitidas quase

---

<sup>30</sup> O termo SPAM pode significar *Sending and Posting Advertisement in Mass* (enviar e publicar publicidade em massa), ou *Stupid Pointless Annoying Messages* (mensagem ridícula, sem propósito e irritante).

<sup>31</sup> DNS (*Domain Name System*) é um sistema de gestão de nomes hierárquico e distribuído para computadores, serviços ou qualquer recurso ligado à *internet* ou rede privada. Serve para traduzir os nomes de domínio, mais facilmente memorizáveis, para endereços IP necessários para localizar e identificar serviços e dispositivos computacionais com os seus protocolos inerentes.

diretamente do remetente para o/s destinatário/s, os artigos publicados num Fórum ficam disponíveis a todos os participantes desse Fórum.

Para além da publicação de texto simples, é muitas vezes também possível a aplicação de formatação a esse texto e permitida a publicação de informação MM como imagens, sons e vídeos.

### **Caracterização técnica e funcional**

Os Fóruns permitem a publicação de mensagens, inicialmente só texto simples, posteriormente foram evoluindo para suportar também formatação no texto, bem como conteúdo MM.

Tradicionalmente as mensagens publicadas ficam visíveis a toda a comunidade mas alguns fóruns permitem o envio de mensagens privadas, só visíveis por um destinatário ou pequeno grupo de destinatários.

### **Tecnologias existentes**

Apesar de existirem alguns servidores dedicados ao suporte a este tipo de serviço, habitualmente não é necessário mais que um servidor *web* convencional e um CMS<sup>32</sup> pois muitos já fornecem este serviço de raiz.

Para utilizar a maioria dos Fóruns não é necessário qualquer aplicação cliente especial – basta um navegador *web*.

### **Vantagens e/ou desvantagens da sua adoção**

A rapidez com que se pode pedir ajuda a uma comunidade para um determinado problema e a facilidade com que se pode contribuir para uma solução é, sem qualquer dúvida, umas das grandes vantagens deste serviço. Contudo, tal como já visto com o email, estes serviço é muitas vezes mal utilizado, levando a uma perda de produtividade.

O spam também pode ser um problema. O spam nos Fóruns consiste na criação de mensagens de publicidade indesejadas ou não-solicitadas em fóruns na *internet*. É geralmente feito por *spam-bots*, ou seja, automatizado. A maior parte do spam em fóruns é composto por *link* para *sites* externos, com o duplo objetivo de aumentar a visibilidade nos motores de buscas em áreas com elevado nível de competição, como a perda de peso, produtos farmacêuticos, jogos, pornografia, imóveis ou empréstimos, e gerar mais tráfego para esses *sites* comerciais.

## **Blogue**

Um blogue ou *blog*<sup>33</sup> é um *site* cuja estrutura permite a atualização rápida a partir da publicação de artigos ou *posts*. Estes são, em geral, organizados de forma cronológica

---

<sup>32</sup> CMS – Content Management System (Sistema de Gestão de Conteúdos).

(normalmente invertida), tendo como foco a temática proposta do blogue, podendo ser escritos por uma ou várias pessoas, de acordo com a política do blogue.

### **Objetivos com enquadramento institucional**

Um blogue pode ser privado, como na maioria dos casos, ou pode ter objetivos institucionais ou de negócio. Blogues usados internamente para melhorar a comunicação e cultura institucional, ou usados externamente para fins de marketing, *branding*, ou relações públicas são chamados de blogues corporativos.

Blogues internos podem ser usados em vez de reuniões e discussões por emails e podem ser especialmente úteis quando as pessoas envolvidas estão em locais diferentes ou têm horários incompatíveis. Os blogues também podem permitir que pessoas possam contribuir na sua área de conhecimento/especialidade, em situações que de outra maneira não teriam conhecimento ou não seriam convidadas a participar numa discussão.

### **Caracterização técnica e funcional**

Os primeiros blogues não passavam de simples componentes de *sites*, atualizados manualmente no próprio código da página. A evolução das ferramentas de auxílio à produção e manutenção de artigos colocados por ordem cronológica facilitaram o processo de publicação, ajudando em muito na popularização do formato. Isso levou ao surgimento e aperfeiçoamento de ferramentas, bem como de plataformas de alojamento específico para blogues.

### **Tecnologias existentes**

Numa pesquisa rápida na *internet* pode encontrar-se vários serviços de alojamento, tanto pagos como gratuitos, que permitem a criação de um novo blogue a qualquer pessoa, sem necessidade de conhecimentos técnicos.

Caso se pretenda alojamento próprio, apesar de existirem alguns servidores dedicados ao suporte a este tipo de serviço, habitualmente não é necessário mais que um servidor *web* convencional e um CMS na medida em que este serviço está frequentemente incluído nas funcionalidades fornecidas de base.

No geral, não é necessário qualquer aplicação cliente especial – basta um navegador *web*.

### **Vantagens e/ou desvantagens da sua adoção**

Os blogues usados internamente podem ajudar a promover a cultura organizacional, permitindo que os vários elementos dessa organização tenham uma participação mais ativa e facilitada. Um exemplo poderia ser um “diário de bordo” da organização.

---

<sup>33</sup> *Blog* – contração do termo inglês *web log*, “diário da rede”.

Contudo, este instrumento de comunicação é facilmente mal usado, com contribuições descontextualizadas, reivindicativas ou intencionalmente desestabilizadoras.

No caso dos blogues externos, podem contribuir para uma maior proximidade entre uma organização e os seus parceiros (clientes, fornecedores, etc.), trazendo benefícios nas áreas de marketing, *branding*, ou relações públicas. Contudo, também são frequentes os casos de difamação, de implicações políticas e de funcionários que, ao fazerem comentários sobre elementos do seu local de trabalho, podem afetar negativamente o reconhecimento de *branding* da organização.

## **Wiki**

Uma *Web Wiki* consiste num software colaborativo que permite a edição coletiva dos documentos usando um sistema que não necessita que o conteúdo seja revisto antes da sua publicação. Permite que os documentos sejam editados coletivamente com uma linguagem de marcação muito simples, através da utilização de um navegador *web*.

### **Objetivos com enquadramento institucional**

Uma *Wiki* convida todos os utilizadores da organização a editar qualquer página ou a criar novas páginas dentro do *site* da *Wiki*, usando somente um navegador *web* sem necessidade de qualquer acrescento.

As *Wikis* promovem associações significativas de tópicos entre diferentes páginas, tornando a criação de ligações entre páginas muito fácil e intuitiva, mostrando se uma determinada página alvo pretendida existe ou não.

A *Wiki* não é um *site* cuidadosamente elaborado para os visitantes ocasionais. Em vez disso, procura envolver o visitante num processo contínuo de criação e de colaboração que constantemente atualiza o *site web*.

### **Caracterização técnica e funcional**

O software de *Wiki* é um tipo de software colaborativo que executa um sistema de *Wiki*, permitindo que páginas *web* sejam criadas e editadas usando um navegador *web* comum. Este serviço pode ser implementado com uma série de scripts por trás de um servidor *web* ou como uma aplicação independente do lado do servidor que corre em um ou mais servidores *web*. O conteúdo é armazenado num sistema de ficheiros e as alterações no conteúdo são guardadas num sistema de gestão de base de dados relacional.

### **Tecnologias existentes**

Existem vários pacotes de software para a gestão de Wikis como MediaWiki<sup>34</sup>, WikidPad<sup>35</sup> ou TiddlyWiki<sup>36</sup>, entre muitas outras soluções. As Wikis também podem ser alojadas numa “Wiki farm” (“quinta Wiki”), onde o software do lado do servidor é executado pelo detentor da “quinta”. Alguns exemplos de “Wiki farm”: PBwiki, Socialtext, Wetpaint, e Wikia. Algumas Wikis também podem ser pessoais e protegidas por uma senha (*password*).

### **Vantagens e/ou desvantagens da sua adoção**

A facilidade e rapidez com que se consegue criar novas páginas ou alterar páginas existentes é uma vantagem óbvia ao permitir a colaboração da comunidade da organização. Contudo, esta facilidade é simultaneamente a sua principal desvantagem ao permitir que as alterações possam acontecer sem qualquer supervisão o que pode levantar questões de segurança. Um bom exemplo deste problema pode ser observado com frequência na Wikipedia<sup>37</sup> em que os seus gestores são obrigados a bloquear páginas (artigos) na sequência de alterações efetuadas por pessoas menos bem-intencionadas.

## **3.2.2. Serviços de comunicação síncrona**

Os primeiros serviços/protocolos de comunicação síncrona só permitiam a troca de mensagens de texto. Contudo, este tipo de serviços evoluiu imenso sendo hoje comum a disponibilização de serviços de voz e vídeo.

### **Serviço de IRC**

O *Internet Relay Chat* (IRC) é um protocolo de comunicação utilizado na *internet* que, basicamente, é utilizado para conversas online em tempo real, permitindo conversas em grupo ou privadas. Este protocolo permitiu o surgimento de vários serviços de conversação, designados genericamente como serviços de IRC. Originalmente, este protocolo só permitia a transferência de texto puro, tendo posteriormente sido estendido o que veio permitir serviços adicionais como a transferência de ficheiros.

Os serviços baseados neste protocolo, hoje já pouco usados, foram sendo substituídos por outros que, além de texto, permitiam também conversas de voz e vídeo em tempo

---

<sup>34</sup> MediaWiki: <https://www.mediawiki.org/>

<sup>35</sup> WikiPad: <http://wikidpad.sourceforge.net/>

<sup>36</sup> TiddlyWiki: <http://tiddlywiki.com/>

<sup>37</sup> Wikipedia (<https://www.wikipedia.org/>) é uma enciclopédia na *internet* de acesso livre.



real, entre outras funcionalidades. O mercado disponibiliza várias destas soluções com uma variedade de funcionalidades, tais como: *chat* de texto, conversação de áudio, conferência de vídeo, transferência de ficheiros, tudo em tempo real com um ou vários interlocutores em simultâneo.

### **Objetivos com enquadramento institucional**

Os serviços de *chat* permitem o contacto, em tempo real, entre duas ou mais pessoas que podem estar geograficamente distantes. A possibilidade de comunicação com voz e vídeo torna o processo comunicacional mais rico, tornando-o mais próximo do contacto face-a-face, que, associado à capacidade de *chat* textual e de envio de ficheiros, permite uma otimização significativa da eficiência desse processo comunicacional.

### **Caracterização técnica e funcional**

Os serviços de *chat* mais simples (só texto) podem ser suportados por um serviço que pode ser adicionado a um servidor existente e num dos programas cliente, compatíveis com o servidor escolhido, disponíveis no mercado.

Quando se pretende usar os serviços mais avançados, como áudio e vídeo, as exigências técnicas aumentam consideravelmente, tanto do lado dos servidores como do lado dos clientes, não só pela capacidade de processamento acrescida como pela largura de banda necessária.

### **Tecnologias existentes**

Existem várias implementações deste tipo de serviços, no entanto, o mercado é dominado por um número reduzido de empresas especializadas neste tipo de serviços. Algumas destas empresas, para além dos serviços pagos (*premium*), apresentam muitas vezes também versões gratuita dos seus serviços, como é o caso do Skype<sup>38</sup>.

A maioria das soluções comerciais oferecidas por estas empresas recorrem à utilização de protocolos proprietários, pelo que é frequente o serviço só funcionar no cliente fornecido pela própria empresa.

### **Vantagens e/ou desvantagens da sua adoção**

A possibilidade de promover reuniões em conferência em que os seus participantes se encontram geograficamente separados é uma vantagem pela redução de custos e de tempo associados à deslocação das pessoas.

Ao complementar a escrita, do serviço tradicional de *chat* textual, com a audioconferência e videoconferência (multiutilizador), e com a partilha de ficheiros permite uma flexibilidade e funcionalidade excelentes.

---

<sup>38</sup> Skype (<http://www.skype.com/>): serviço que permite a realização troca de mensagens instantâneas de texto e a realização de conferências de áudio ou vídeo com múltiplos intervenientes.

Comparativamente com outros métodos de comunicação remota, a utilização de áudio e vídeo melhora o processo comunicacional entre intervenientes ao permitir alguma comunicação não-verbal, impossível, por exemplo, na comunicação escrita.

A utilização deste tipo de ferramentas, em detrimento do contacto face-a-face entre as pessoas, pode ter impacto negativo no sentimento de pertença e na cultura corporativa da organização. Esta situação é especialmente relevante no caso do teletrabalho a partir de casa.

A possibilidade de registo (gravação de áudio e/ou vídeo) pode ser simultaneamente uma vantagem, ao permitir o registo para memória futura da organização, e uma desvantagem se utilizado indevidamente.

### 3.2.3. Serviços de suporte e difusão de informação

A maioria dos serviços de informação e comunicação estão alojados em servidores especializados. De seguida irão ser abordados alguns dos mais importantes.

#### Serviço FTP

O *File Transfer Protocol* (FTP), é um protocolo de rede standard usado para transferir ficheiros de um sistema computacional para outro sobre uma rede baseada no protocolo TCP, como a *internet*. Este protocolo é baseado numa arquitetura do tipo cliente-servidor em que os utilizadores se podem autenticar usando um *login* e *password*, sendo também possível a ligação anónima se o servidor estiver configurado para o permitir. Para ligações seguras, de modo a proteger o *login* e *password*, bem como encriptar o conteúdo, o FTP pode ser protegido com SSL/TLS<sup>39</sup> (FTPS) ou, alternativamente, sobre SSH<sup>40</sup> (SFTP).

O termo FTP tanto se pode referir ao protocolo como ao programa que implementa este protocolo (Servidor FTP) ou mesmo ao programa cliente.

As primeiras aplicações clientes FTP eram baseadas em linha de comando, desenvolvidas antes de os sistemas operativos disponibilizarem interfaces gráficas. Ainda hoje este tipo de aplicações (baseadas em linha de comando) são incluídas nas instalações base de muitos sistemas operativos como o Windows, Mac OS, Unix ou Linux. Desde então, muitas aplicações clientes e utilitários FTP foram desenvolvidos

---

<sup>39</sup> SSL/TLS: *Secure Sockets Layer* (SSL) e o seu sucessor *Transport Layer Security* (TLS) são protocolos criptográficos desenvolvidos para assegurar comunicações seguras sobre rede de computadores, como a *internet*.

<sup>40</sup> FTP sobre SSH: Consiste na utilização do FTP através de um túnel SSH que encripta todo o tráfego que circula nesse túnel.

para desktops, servidores e dispositivos móveis, entre outros sistemas computacionais. O suporte para FTP foi também incorporado em muitas aplicações de produtividade, como editores de páginas (*sites*) *web*. A maioria dos navegadores *web* também incorporam suporte para o FTP.

### **Objetivos com enquadramento institucional**

O serviço FTP permite a troca / partilha de ficheiros com recurso a um sistema de armazenamento central (servidor FTP). A possibilidade de armazenamento de qualquer tipo de ficheiro dá-lhe uma flexibilidade e funcionalidade importantes para as organizações.

### **Caracterização técnica e funcional**

O serviço / protocolo FTP permite a validação dos utilizadores por intermédio de um par *login + password*. Contudo, não recorre a qualquer tipo de encriptação dos dados que coloca / recebe da rede, incluindo os dados de validação do utilizador.

De maneira a permitir o *login* seguro e a transmissão do conteúdo de forma encriptada, o FTP é frequentemente protegido com SSL/TLS (FTPS). Uma alternativa passa por usar o FTP sobre SSH (SFTP) ou sobre VPN<sup>41</sup>. Contudo, esta diversidade de soluções levanta alguns problemas de compatibilidade.

### **Tecnologias existentes**

Para o lado do servidor existem várias implementações, desde as fornecidas diretamente com alguns sistemas operativos, passando pelas soluções gratuitas e/ou de código aberto, e terminando em soluções comerciais proprietárias. A solução mais adequada a determinado cenário dependerá de vários fatores tais como as funcionalidades pretendidas, a dimensão e tráfego previstos ou o nível de segurança.

Para o lado do cliente também existem muitas soluções disponíveis no mercado, começando pelas aplicações disponibilizadas diretamente pelo SO (sistema operativo) (como Windows, Mac OS ou Linux), passando pelo suporte que a maioria dos navegadores *web* disponibilizam de forma nativa, e terminado em aplicações específicas FTP. Normalmente, tanto as aplicações que vêm com o SO como os navegadores *web* têm algumas limitações, disponibilizando um conjunto de funcionalidades limitado. Felizmente a oferta de aplicações clientes FTP é bastante variada, não sendo difícil encontrar soluções gratuitas bastante competentes.

---

<sup>41</sup> VPN: Rede privada virtual (*Virtual Private Network*) é uma rede de comunicações privada construída em cima de uma rede de comunicações pública (como por exemplo, a *internet*). Por questões de segurança, o tráfego é normalmente encriptado.

## Vantagens e/ou desvantagens da sua adoção

A possibilidade de dispor de um sistema de armazenamento central onde se pode guardar ou consultar qualquer tipo de ficheiro é uma mais-valia.

Contudo, é necessário ter um cuidado especial na organização da informação armazenada sob pena de ser virtualmente impossível geri-la passado pouco tempo.

Outro problema importante pode ser, no caso de documentos partilhados, a definição de permissões (quem tem acesso a quê), já que o FTP não suporta este tipo de funcionalidades diretamente.

A possibilidade de encriptar o conteúdo transferido é uma mais-valia, contudo levanta também alguns problemas devido à diversidade de soluções existentes incompatíveis entre si. Acresce que a maioria de tanto servidores como clientes só implementam uma parte destes protocolos de segurança, podendo ter implicações / impondo limitações nas escolhas possíveis.

## Servidor web

A história dos servidores *web* está intimamente ligada à história da *World Wide Web* pois os servidores *web* são o seu motor, sendo responsáveis pelo conteúdo que pode ser consultado por intermédio de, por exemplo, um navegador *web*. Este percurso começou num local inesperado (Ceruzzi, 2003): o laboratório de física de alta energia do CERN<sup>42</sup>, na fronteira Franco-Suíça.

Estes servidores usam normalmente o protocolo HTTP<sup>43</sup> para comunicar com os clientes (por exemplo, os navegadores *web*) e usam o HTML<sup>44</sup> para formatar o conteúdo apresentado. As comunicações entre um servidor e um cliente podem ser protegidas com o recurso ao protocolo HTTPS<sup>45</sup> que permite encriptar todo o conteúdo transferido.

Os primeiros servidores *web* só permitiam o alojamento de páginas estáticas. Hoje quase todos os servidores suportam páginas dinâmicas, o que permite executar código

---

<sup>42</sup> CERN: Organização Europeia para a Pesquisa Nuclear (*Conseil Européen pour la Recherche Nucléaire*) – <https://home.cern/>.

<sup>43</sup> HTTP: Protocolo de Transferência de Hipertexto (*Hypertext Transfer Protocol*) é um protocolo de comunicação usado por sistemas de informação hipermédia. Ele é a base para a comunicação de dados da *World Wide Web*.

<sup>44</sup> HTML: Linguagem de Marcação de Hipertexto (*HyperText Markup Language*) é uma linguagem de marcação utilizada para produzir páginas na *web*. Os documentos HTML podem ser interpretados por navegadores *web*.

<sup>45</sup> HTTPS: Protocolo de transferência de Hipertexto Seguro (**HTTP Secure**, HTTP over TLS ou HTTP over SSL).

do lado do servidor recorrendo a tecnologias como o PHP<sup>46</sup>, ASP<sup>47</sup>, entre outras. Este código ao ser executado permite o acesso a recursos, como bases de dados, fazer algum tipo de processamento sobre os dados e, no final, produzir o código HTML que será enviado para o sistema cliente.

### Objetivos com enquadramento institucional

Por trás de qualquer *site*, existe sempre um servidor *web*, independentemente deste fazer parte de uma *intranet* de uma organização ou estar exposto ao acesso público. Este tipo de servidor dá suporte a uma multiplicidade de serviços, desde a simples partilha de informação com um *site* estático até às mais sofisticadas redes sociais, como o Facebook<sup>48</sup>. Uma organização pode disponibilizar um vasto leque de serviços com suporte num servidor *web*, bastando para tal proceder à sua instalação e respetiva configuração.

### Caracterização técnica e funcional

Os servidores *web* permitem mecanismos sofisticados de validação de utilizadores. A possibilidade de encriptar as comunicações, utilizando o protocolo HTTPS, proporciona-lhe um nível de segurança considerável nas comunicações entre o servidor e o cliente.

A possibilidade de ligação a bases de dados (alojadas no mesmo servidor ou noutra) aumenta em muito as funcionalidades disponibilizadas.

As funcionalidades disponibilizadas dependem sobretudo dos serviços instalados no servidor *web*. Um bom exemplo são os CMS<sup>49</sup> como por ex. WordPress<sup>50</sup> ou Joomla<sup>51</sup> que disponibilizam um conjunto de funcionalidades que facilitam a criação e gestão de um *site* sofisticado só necessitando de suporte a por ex. PHP e acesso a uma base de dados.

---

<sup>46</sup> PHP: Acrónimo para processador de hipertexto (“*PHP: Hypertext Preprocessor*”, originalmente “*Personal Home Page*”), é uma linguagem de *script* (de código aberto) interpretada do lado do servidor *web*, capaz de produzir conteúdo dinâmico.

<sup>47</sup> ASP: Acrónimo para “*Active Server Pages*” é uma estrutura de bibliotecas básicas para processamento de linguagens de *script* no lado servidor para geração de conteúdo dinâmico. Exemplos de linguagens suportadas são VBScript e JScript.

<sup>48</sup> Facebook é uma rede social disponível em <https://www.facebook.com/>

<sup>49</sup> CMS: Sistema de gestão de conteúdo (*Content Management System*)

<sup>50</sup> WordPress: sistema de gestão de conteúdo gratuito e de código aberto baseado em PHP e MySQL. Site oficial: <https://wordpress.org/>

<sup>51</sup> Joomla: sistema de gestão de conteúdo gratuito e de código aberto. Site oficial: <https://www.joomla.org/>

## **Tecnologias existentes**

Para o lado do servidor existem várias implementações, desde as fornecidas diretamente com alguns sistemas operativos, passando pelas soluções gratuitas e/ou de código aberto, e terminando em soluções comerciais proprietárias. A solução mais adequada a determinado cenário dependerá de vários fatores tais como as funcionalidades pretendidas, a dimensão e tráfego previstos ou o nível de segurança.

Para o lado do cliente também existem muitas soluções disponíveis no mercado, começando pelos navegadores *web* muitas vezes disponibilizados junto com o SO (como Windows, Mac OS ou Linux), passando por aplicações específicas para determinadas funções como algumas *apps* disponíveis para *smartphones*.

## **Vantagens e/ou desvantagens da sua adoção**

A facilidade com que pode proceder à instalação de um novo serviço num servidor *web* é um detalhe importante para um gestor de informática de uma organização.

Contudo, também se corre o risco de serem instalados e disponibilizados serviços sem o planeamento necessário que, no limite, podem ser desadequados à organização

A possibilidade de partilhar informação de forma fácil é, naturalmente, uma mais-valia. Uma interface bem projetada pode ajudar a vida dos utilizadores.

Contudo, a qualidade dessa interface não depende diretamente do servidor mas do serviço instalado nesse servidor *web*. Adicionalmente, como acontece com muitas outras ferramentas, há a possibilidade de utilização incorreta e/ou abusiva dos serviços disponibilizados.

A utilização de canais de comunicação seguros (HTTPS) permite que qualquer colaborador possa ter acesso a material sensível mesmo quando se encontre fora das instalações da organização.

Contudo, essa possibilidade também expõe o servidor a potenciais ataques perpetrados por, por exemplo, *hackers*.

## **Servidores de Base de Dados**

Uma base de dados consiste num conjunto de arquivos relacionados entre si com registos de qualquer tipo (ex. sobre pessoas, locais, produtos). Mantém coleções organizadas de dados que se relacionam de forma a criar algum significado (Informação) e proporciona uma maior eficiência quando se efetua uma pesquisa ou estudo.

Antes do surgimento das bases de dados, as aplicações recorriam a ficheiros armazenados diretamente pelo sistema operativo para armazenar os seus dados.

Existem várias soluções de bases de dados no mercado, incluído algumas de código aberto e sem custos. Entre as bases de dados mais populares podem listar-se as

seguintes: MySQL<sup>52</sup>, Oracle DB<sup>53</sup>, Microsoft Access<sup>54</sup> ou Microsoft SQL Server<sup>55</sup>, entre muitas outras.

### **Objetivos com enquadramento institucional**

A utilização de uma base de dados permite o armazenamento, organização e pesquisa de dados / informação mais eficiente. Quando o servidor de base de dados é usado em conjugação com um servidor *web* o acesso à informação pode ser muito facilitado pelas funcionalidades que uma interface *web* pode propiciar.

### **Caracterização técnica e funcional**

Os servidores de base de dados são normalmente computadores dedicados que alojam os dados da base de dados e que muitas vezes só correm o software relacionado com a gestão da base de dados. Estes computadores geralmente recorrem a hardware poderoso: capacidade de multiprocessamento (múltiplos CPUs<sup>56</sup>), quantidade de memória generosa e sistemas de armazenamento de alto desempenho (muitas vezes recorrendo a soluções com SSDs<sup>57</sup> e discos em RAID<sup>58</sup>).

### **Tecnologias existentes**

O tipo de dados determina a estrutura lógica da base de dados e no fundo, determina a maneira como a informação pode ser armazenada, organizada e manipulada. O modelo mais popular é a base de dados relacional que usa um formato baseado em tabelas.

### **Vantagens e/ou desvantagens da sua adoção**

A utilização de um servidor de base de dados permite otimizar o processo de armazenamento, organização e, sobretudo, de pesquisa de dados / informação. Contudo, uma base de dados mal planeada / desenhada pode apresentar limitações crescentes (de desempenho, por exemplo) à medida que vai crescendo e pode tornar virtualmente impossível atualizações futuras para, por exemplo, acrescentar à estrutura existente novos campos de informação.

---

<sup>52</sup> MySQL é um sistema de gestão de bases de dados relacionais de código aberto. Site da empresa: <http://www.mysql.com/>

<sup>53</sup> A Oracle é umas das empresas que há mais tempo está no mercado das bases de dados. Site da empresa: <http://www.oracle.com/>

<sup>54</sup> O Access faz parte do pacote Office da Microsoft. É uma solução para bases de dados para uso em sistemas clientes, não servidores. Página oficial do produto: <http://office.microsoft.com/access>

<sup>55</sup> O SQL server é a solução de base de dados para servidor da Microsoft. Página oficial do produto: <http://www.microsoft.com/sqlserver>

<sup>56</sup> CPU: Unidade central de processamento (*Central Processing Unit*).

<sup>57</sup> SSD: Unidade de armazenamento de estado sólido (*Solid-State Drive*).

<sup>58</sup> RAID: Sistema redundante de discos independentes (*Redundant Array of Independent Disks*).

### 3.2.4. Serviços de segurança

A segurança dos serviços de informação e comunicação disponibilizados pelas organizações é vital. De maneira a assegurar essa segurança e simultaneamente manter os serviços com o mínimo de restrições e constrangimentos recorre-se a soluções como as apresentadas de seguida.

#### **Servidor *Proxy***

O servidor *proxy* funciona como intermediário entre os pedidos dum sistema cliente e os recursos disponibilizados por um servidor. Basicamente o cliente liga-se ao servidor *proxy*, pedindo qualquer serviço, como um ficheiro, ligação, página *web*, ou outro recurso disponibilizado por outro servidor diferente. O servidor *proxy* avalia o pedido de maneira a simplificar e controlar a sua complexidade, permitindo adicionar estruturação e encapsulamento a sistemas distribuídos. O tipo de *proxies* mais conhecidos e banais são as *proxies web* que facilitam o acesso a conteúdo da *World Wide Web*, fornecendo segurança e/ou anonimato.

Os servidores *proxy* com *cache* conseguem acelerar os pedidos ao recorrer a conteúdo armazenado dum pedido anterior efetuado pelo mesmo cliente ou mesmo um diferente. As *proxies* com *cache* mantêm cópias locais dos recurso frequente solicitados, permitindo às organizações grandes reduzir significativamente a utilização de largura de banda e respetivo custo, e simultaneamente aumentar a *performance*.

Uma das maiores vantagens do uso de um servidor *proxy web* numa organização consiste em ser possível, de dentro da rede informática da organização, ter acesso a conteúdo da *World Wide Web* sem expor a rede interna ao exterior da organização – o único ponto exposto é o servidor *proxy*.

#### **Servidor *Firewall***

Uma *firewall* é um sistema de segurança de rede que monitoriza e controla a tráfego de entrada e saída da rede baseado num conjunto de regras pré-determinadas. Tipicamente, uma *firewall* funciona como uma barreira entre uma rede interna segura e confiável, e uma rede externa, como a *internet*, que se assume como não sendo nem segura nem de confiança.

A *firewall* aparece muitas vezes associada a serviços de *proxy* permitindo, além das funcionalidades da *proxy*, juntar um serviço de filtragem que permite que determinados pacotes de dados provenientes de determinado serviço ou dispositivo passem enquanto outros pacotes são barrados.



## Servidor VPN

O serviço VPN<sup>59</sup> estende a rede privada de uma organização através de uma rede pública, como a *internet*. Permite aos utilizadores enviar e receber dados através de redes partilhadas ou públicas como se o seu dispositivo computacional estivesse diretamente ligado à rede privada, beneficiando da funcionalidade, segurança e políticas de gestão dessa rede privada. Uma VPN é criada estabelecendo uma ligação ponto a ponto virtual recorrendo a ligações dedicadas, protocolos de encaminhamento (*tunnelling*) ou encriptação de tráfego.

Este serviço permite que qualquer colaborador tenha acesso, de forma segura, aos serviços de informação e comunicação disponibilizados na rede da organização, mesmo quando o colaborador está no exterior.

## 4. Visualização de informação – principais aspetos no desenvolvimento duma solução

Esta secção começará por uma pequena introdução à visualização de informação, abordando de seguida a aplicação de cor em visualização e algumas propostas de taxonomia para visualização de informação resultantes da pesquisa bibliográfica e no contexto de necessidade deste projeto, nomeadamente considerando três eixos fundamentais: tipos de dados, métodos de visualização e técnicas de apresentação e interação. Depois será discutida a visualização de diversos tipos de informação: estruturas lineares, hierarquias, redes e informação multidimensional. Finalmente serão abordados alguns contextos de aplicação.

Várias definições de Visualização têm sido propostas desde que se constituiu como área científica independente:

- “Método de Computação. Transforma o simbólico no geométrico, permitindo aos investigadores observar simulações e computações ... fornecendo uma visão científica com recurso a métodos visuais” (McCormick et al., 1987).
- “O uso de representações visuais, interativas e suportadas por computador, de dados para amplificar a cognição” (Card et al., 1999).

---

<sup>59</sup> VPN: Rede privada virtual (*Virtual Private Network*).

- “Um pensamento visual computacionalmente intenso” (Rhyne, 2000, 2003).

Assim, uma visualização pode ser considerada uma representação visual interativa de um conjunto de dados que permite aos seus utilizadores uma melhor compreensão do fenómeno a eles subjacente. O seu desenvolvimento deve ser feito com base no conhecimento dos utilizadores, das tarefas que vão desempenhar e do contexto de utilização e implica um conjunto de fases incluídas no modelo de referência da Visualização de Informação (Figura II-7). Dado um conjunto de dados em bruto (*raw data*), em primeiro lugar é necessário pré processá-los, depois estabelecer as estruturas visuais a usar para os representar (i.e., fazer o *mapping*) e produzir as imagens correspondentes às vistas selecionadas pelo utilizador. O utilizador pode atuar em todas as fases, desde escolher as transformações a aplicar aos dados (por exemplo retirar *outliers*, ou modificar a escala), a técnica de visualização a usar (por exemplo escolher uma escala de cor) e a vista a ser mostrada no ecrã (por exemplo escolher o ponto de vista, ou quais os dados a apresentar).

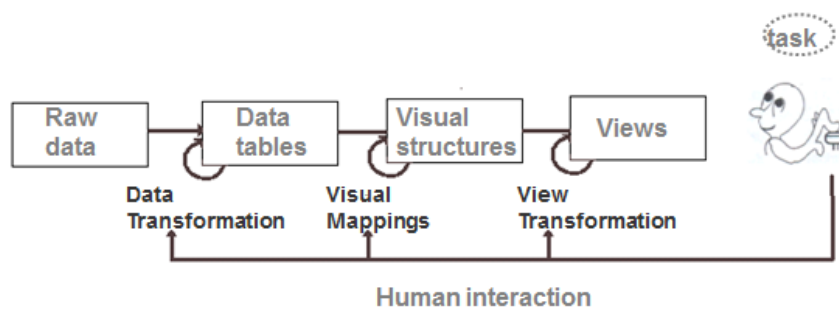


Figura II-7: Modelo de referência da Visualização (adaptado de (Card et al., 1999))

Nesta secção apresentam-se aspetos fundamentais no desenvolvimento de qualquer solução de visualização ou aplicação de visualização (Spence, 2000, 2007): tipos de dados, representação, apresentação e interação. As características dos dados influenciam diretamente os métodos de visualização a usar; em face dos dados (obtidos diretamente do utilizador ou derivados), coloca-se o problema de selecionar a forma de os representar visualmente (ou fazer *visual mapping*), depois será necessário ponderar a melhor forma de os apresentar em cada momento ao utilizador, sendo frequentemente necessário apresentar apenas parte dos dados, e finalmente será necessário permitir ao utilizador interagir com os dados.

Todos estes passos estão representados no modelo de referência (Figura II-7) e em todo este processo é fundamental a escolha das estruturas visuais (*visual structures*) a usar para representar visualmente os atributos dos dados (através do *visual mapping*). Esta escolha deve incidir em três tipos de estruturas: o substrato espacial, os elementos gráficos e as propriedades gráficas (Mazza, 2009), que devem ser escolhidas de acordo com o tipo de dados e os objetivos da visualização. O substrato espacial define as dimensões do espaço físico em que a visualização é criada e pode ser definido em

termos dos eixos (xx, yy...). Os elementos gráficos correspondem ao que é visível no espaço da visualização e podem ser pontos, linhas, superfícies ou volumes. Estes elementos podem ter diferentes propriedades gráficas a que a retina humana é sensível (pelo que são geralmente chamadas de propriedades retinianas). As propriedades gráficas mais comuns são o tamanho, a cor, a forma, a textura e a orientação, sendo o tamanho considerado como a propriedade que permite representar com mais precisão dados quantitativos (Mazza, 2009; Spence, 2007). A cor, não sendo tão eficaz na representação deste tipo de dados, é uma pista visual muito forte, frequentemente utilizada para esse fim através de tabelas de cor, bem como para ajudar em tarefas de seleção e associação; no entanto, a sua correta utilização é essencial para que não degrade o desempenho do utilizador e deve ser ponderada com cuidado. Devido à complexidade desta escolha, apresentam-se algumas considerações sobre a utilização da cor em visualização.

Nas subsecções seguintes descrevem-se, com base no modelo de referência, várias classificações propostas para os dados e para os métodos de visualização mais importantes para esta tese.

Antes de avançar, duas notas no que diz respeito à importância das características do ecrã onde é efetuada a visualização e à diferença de desempenho existente entre género feminino e masculino em ambientes 3D.

Em (Tan, Gergle, et al., 2003) é comparado o desempenho de ecrãs de grandes dimensões com o de ecrãs normais, mantendo o mesmo ângulo de visão. De acordo com este autor, ecrãs grandes, como os projetados numa parede, permitem uma melhoria significativa no desempenho em tarefas de orientação espacial.

Em (Tan, Czerwinski, et al., 2003) vão mais longe e apresentam um estudo onde se demonstra que a conjugação de ecrãs de grandes dimensões com vista panorâmica (“wide”) favorece o desempenho, sendo a vista panorâmica especialmente benéfica para o género feminino. Os autores deste estudo adicionam alguns princípios, em relação à abordagem tradicional, a aplicar ao design de sistemas virtuais de navegação 3D.

Nesta sequência, logo à partida, é de esperar que a proposta apresentada neste trabalho possa, por exemplo, apresentar desempenhos distintos consoante seja utilizado por gestores do sexo feminino ou masculino.

## **4.1. Aplicação de cor em visualização**

Muitos métodos de visualização tiram partido da cor para organizar, codificar e apresentar informação ao observador. A cor é usada nas mais variadas tarefas, desde

aplicações tão comuns como nas interfaces de utilizador, até outras mais específicas como na codificação da classificação de valores ou classes de objetos.

Neste contexto, irá fazer-se uma breve abordagem à problemática da cor, ainda que não sendo o tema, nem o contributo central deste trabalho, é, contudo, um instrumento fundamental em visualização. Assim, será discutida a questão em dois cenários, apresentados a seguir:

- Utilização da cor para classificação;
- Utilização de sequência (paletas ou tabelas) de cores para representação de variáveis quantitativas.

A abordagem apresentada tem como base a proposta de Colin Ware em (Ware, 2000, 2003), influenciado pelo trabalho desenvolvido por Eduard Tufte em (Tufte, 1994, 1998).

### 4.1.1. Cor para classificar

O nome técnico para a classificação (ou etiquetagem) de um objeto é “codificação nominal de informação”. Um código nominal não tem que ser ordenável; tem somente que ser lembrado e reconhecido. A cor é frequentemente muito eficiente quando utilizada com código nominal (Ware, 2003). Quando se pretende facilitar a vida a alguém que tem que classificar objetos visuais em diferentes categorias, atribuir cores distintas aos objetos é frequentemente a melhor opção.

Existem alguns fatores perceptuais a ter em consideração quando se pretende escolher um conjunto de cores para classificar (Ware, 2003).

1. **Capacidade de distinção.** O espaço de cor CIE<sup>60</sup> permite determinar o grau da diferença percebida entre duas cores colocadas próximas uma da outra. Quando se pretende distinguir rapidamente uma cor de um conjunto de outras cores podem ser aplicadas várias regras. Uma delas passa por escolher uma cor afastada das restantes no espaço de cor CIE.
2. **Cores primárias.** As cores primárias, vermelho, verde, amarelo e azul, bem como o preto e o branco, são especiais a vários níveis. Certamente, estas cores apresentam-se como escolhas naturais quando se necessita de um leque reduzido de cores. Adicionalmente, o trabalho na área da confusão cromática sugere que não devem ser escolhidas duas cores da mesma categoria. Deve ser evitado o uso de múltiplas tonalidades de verde para codificação, por exemplo.
3. **Contraste com o fundo (*background*).** Em muitas situações, espera-se que objetos codificados com cor possam aparecer numa variedade de fundos. O contraste simultâneo com cores de fundo pode alterar dramaticamente o aspeto

---

<sup>60</sup> CIE: Commission Internationale de l'Eclairage (Comissão Internacional da Iluminação) <http://www.cie.co.at/>

da cor, fazendo com que uma cor se pareça com outra. Esta é uma razão pela qual se deve ter um conjunto reduzido de códigos de cor. Um método de reduzir os efeitos do contraste consiste em colocar uma linha fina branca ou preta a envolver o objeto codificado com cor. Adicionalmente, nunca se devem usar códigos usando diferenças puramente cromáticas relativamente ao fundo. Deve existir uma diferença de luminância significativa, adicionalmente à diferença de cor.

4. **Cegueira cromática.** Uma vez que uma parte significativa da população tem algum tipo de distúrbio relacionado com a perceção das cores (daltonismo, etc.), pode ser desejável usar cores que possam ser distinguidas mesmo por pessoas com esses problemas. Note-se que a maioria das pessoas com problemas visuais relacionados com cor não conseguem distinguir cores que variam na direção vermelho-verde. Contudo, quase toda a gente consegue distinguir variações na direção amarelo-azul. Infelizmente, isto reduz drasticamente as opções de design disponíveis.
5. **Número.** Apesar de a codificação de cor ser uma forma excelente de mostrar informação de categorias, somente um número reduzido de categorias pode ser rapidamente percebido. Estimativas variam entre cinco e dez cores.
6. **Tamanho do objeto.** Os objetos codificados com cor não devem ser muito pequenos; especialmente se as diferenças de cor estiverem na direção amarelo-azul, provavelmente o tamanho mínimo será cerca de meio grau do ângulo visual. Em geral, quanto maior for a área codificada com cor, mais fácil será distinguir as cores. Objetos pequenos codificados com cor devem ter cores fortes e saturadas para uma discriminação máxima. Quando são usadas grandes áreas codificadas com cor, por exemplo, regiões de um mapa, as cores devem ser pouco saturadas e variar pouco umas das outras. Isto permite que pequenos objetos codificados com cores vivas possam ser percebidos no meio das regiões de fundo.
7. **Convenções.** Por vezes, as convenções para a codificação de cor devem ser tidas em consideração. Algumas convenções comuns: vermelho=quente, vermelho=perigo, azul=frio, verde=vida, etc.. Contudo, é importante lembrar que estas convenções não são necessariamente universais para todas as culturas. Por exemplo, na cultura Chinesa, vermelho significa vida ou sorte e verde significa morte.

A seguir é apresentada uma lista de doze cores recomendadas para uso em codificação:

- |             |           |              |
|-------------|-----------|--------------|
| 1. Vermelho | 5. Preto  | 9. Cinzento  |
| 2. Verde    | 6. Branco | 10. Laranja  |
| 3. Amarelo  | 7. Rosa   | 11. Castanho |
| 4. Azul     | 8. Ciano  | 12. Púrpura  |

## 4.1.2. Sequência de cores

A pseudocor é a técnica de representar valores com variações contínuas usando uma sequência de cores (Ware, 2003). A pseudocor é largamente usada nos diagramas de radiação astronômica, imagens médicas e em muitas outras aplicações científicas. Os geógrafos usam uma sequência bem definida de cores para mostrar a altura acima do nível do mar; as terras baixas são sempre verdes, que invoca a imagem da vegetação, e a escala continua à medida que a altitude aumenta, passando pelo castanho e o branco para os picos das montanhas.

O esquema de codificação de cor mais usado pelos físicos consiste numa sequência de cor que se aproxima do espectro físico da luz visível (Ware, 2003). Apesar de esta sequência ter algumas propriedades úteis, não é uma sequência perceptual. Isto pode ser facilmente demonstrado com o seguinte teste. Dê-se um conjunto de pastilhas pintadas com vários tons de cinzento a pessoas, pedindo que as coloquem por ordem. Elas rapidamente as ordenarão pela ordem escuro-claro ou claro-escuro. Agora, dê-se às mesmas pessoas pastilhas pintadas de vermelho, verde, amarelo e azul e peça-se que as coloquem por ordem e o resultado será variado. Para muitas pessoas, o pedido não fará muito sentido. Pode, até, haver quem decida ordenar por ordem alfabética.

Seguidamente é apresentado um conjunto de questões importantes que devem ser atendidas quando se pretende usar codificação de cor (Ware, 2000):

- Algumas sequências de cores não serão percebidas por pessoas que sofram de distúrbios de percepção cromática.
- Se for necessária uma sequência ordenável do ponto de vista perceptual, pode ser usada uma sequência de preto-branco, vermelho-verde, amarelo-azul ou saturação (esbatido – vívido). Em geral, uma série de cores que cresça ou decresça monotonicamente relativamente a um ou mais canais oponentes de cor, resultará numa sequência ordenada do ponto de vista perceptual.
- Diferentes sequências podem ser apropriadas dependendo do nível de detalhe dos dados. É sugerido que quando é necessário mostrar um elevado nível de detalhe, a sequência de cor deve ser baseada, sobretudo, na luminância, para tirar partido da capacidade deste canal para comunicar altas frequências espaciais.
- Espaços de cor uniformes podem ser úteis para gerar sequências de cor nas quais passos perceptuais iguais correspondem a passos métricos iguais. Contudo, em alguns casos pode ser desejável exagerar deliberadamente alguns atributos dos dados, usando uma sequência perceptual não-uniforme.
- Quando for importante recuperar valores a partir de uma tabela de cores, a melhor sequência de cores para minimizar erros devido aos efeitos do contraste consiste em usar uma que circule através de muitas cores.

- Em muitos casos, a melhor sequência de cores pode ser uma espiral no espaço de cor. A sequência pode variar através uma gama de cores, tendo o cuidado de escolher cada cor com uma luminância mais alta que a anterior.
- Mesmo que as cores sejam criadas numa sequência contínua e suave, há a tendência para a perceber como um conjunto discreto de cores colocadas em ordem.
- Uma vez que a cor é tridimensional, é possível mostrar duas ou mesmo três dimensões usando cor. Na verdade, isto é usado de forma habitual no caso de imagens de satélite, onde partes invisíveis do espectro são mapeadas para as cores primárias do monitor: vermelho, verde e azul.
- A cor é só uma das maneira de revelar as formas das superfícies. Com frequência, o *shading* da superfície com uma fonte de luz artificial, usando técnicas convencionais de computação gráfica, é uma melhor alternativa.

## 4.2. Taxonomias para a visualização de informação

Vários autores apresentam algumas propostas de taxonomia para visualização de informação. Tendo em conta o contexto de necessidade deste projeto, são considerados os seguintes três eixos fundamentais: tipos de dados, métodos de visualização e técnicas de apresentação e interação.

Daniel Keim (Keim, 2001) apresenta uma classificação taxonómica assente em 3 critérios que são apresentados numa representação cartesiana 3D em que cada eixo caracteriza, respetivamente:

- i) Dados a visualizar, onde os dados são classificados por tipo (ex. textual, unidimensional, bidimensional, etc.);
- ii) Técnicas de interação e distorção, onde as técnicas são agrupadas por tipo de ação aplicada sobre os dados (ex. filtragem, zoom, projeção, etc.);
- iii) Métodos de visualização, onde os métodos são agrupados com base na estratégia de visualização adotada (ex. visualização 2D/3D, densidade de pixéis, etc.).

Stephen Eick (Eick, 2001), faz uma abordagem mais centrada em *sites* de comércio eletrónico (*e-commerce*) e apresenta propostas de visualização para:

- i) Mostrar a estrutura do *site* e sua navegação;
- ii) Percursos dos visitantes,
- iii) Análise de fluxos;

- iv) Monitorização da atividade comercial;

No caso do trabalho desta tese, o último item, monitorização da atividade comercial, é substituído por monitorização da comunicação interna na organização e é ainda acrescentada um novo item à lista:

- v) Monitorização da comunicação interna na organização;
- vi) Mostrar / despistar problemas no design da interface.

Lisa Tweedie (Tweedie, 1999) tem por objetivo, neste seu trabalho, uma caracterização e organização das técnicas subjacentes ao processo de visualização de forma a chegar a uma espaço de sistematização que responda da seguinte forma:

- i) Conhecer/caracterizar os dados utilizados para criar a representação;
- ii) Identificar as formas de interação disponibilizadas ao utilizador;
- iii) Caracterizar a informação de I/O explicitamente representada.

Ben Shneiderman (Shneiderman, 1996) propõe uma taxonomia baseada em tarefas por tipos de dados (*task by data type*) com 7 tipos de dados:

- i) Dados unidimensionais (*1 dimensional data*);
- ii) Dados 2D (*2 dimensional data*);
- iii) Dados 3D (*3 dimensional data*);
- iv) Dados temporais (*temporal data*);
- v) Dados multidimensionais (*multi-dimensional data*);
- vi) Dados estruturados em árvore (*tree-structured data*);
- vii) Dados estruturados em rede (*network data*).

E 7 tarefas:

- i) Vista geral (*overview*);
- ii) Zoom (ver detalhes ou perspetivas gerais);
- iii) Filtragem;
- iv) Obtenção de detalhes por ação direta;
- v) Possibilidade de efetuar relacionamentos;
- vi) Construção de um histórico de ações;
- vii) Possibilidade de extração de resultados (relatório e resultados).

Keith Andrews, em (Andrews, 2002), ainda sistematiza um pouco mais os tipos de dados:

- i) Linear: Tabelas código computador, listas alfabéticas, etc.;
- ii) Hierarquias: Estruturas em árvore;



- iii) Redes: Estruturas gráficas em geral, como ligações entre nodos hipermédia, redes semânticas, etc.;
- iv) Multidimensional: Atributos de meta-informação como tipo, tamanho, autor, data de alteração, etc. Itens com  $n$  atributos tornam-se pontos num espaço  $n$ -dimensional;
- v) Espaço de vetores: A partir da extração de informação (*information retrieve*), corpos de documentos de texto, frequência de palavras, etc.;
- vi) Espacial: Inerentemente dados 2D e 3D como plantas, mapas, modelos CAD, etc..

### 4.2.1. Tipos de dados

As características dos dados, tal como os objetivos e contexto de utilização e as características dos utilizadores, são muito importantes pois influenciam diretamente a escolha dos métodos de visualização a usar numa situação específica. Têm sido apresentadas várias classificações para os dados no contexto da Visualização de Informação (Andrews, 2002; Mazza, 2009; Shneiderman, 1996; Spence, 2000, 2007), sendo uma das mais recentes a apresentada em (Munzner, 2014). Esta classificação distingue entre semântica, tipos de atributos e tipos dos dados. A semântica corresponde ao significado no mundo real, enquanto o tipo de atributo está relacionado com as escalas de medida. Por vezes a mera observação dos dados permite inferir a semântica e os tipos de atributos; no entanto, esta observação pode não ser suficiente e, nesse caso, será necessária informação adicional (geralmente designada como meta-dados). Os atributos podem ser categóricos (sem qualquer ordenação implícita) ou ordenados (que possuem uma ordenação implícita), podendo estes últimos ser ordinais ou quantitativos (que permitem operações aritméticas). Ainda de acordo com (Munzner, 2014), ao nível do conjunto de dados (*data set*) também se podem considerar vários tipos: tabelas, redes e árvore, texto e *logs*, podendo ser combinados em tipos derivados. A obtenção de dados derivados doutros tipos permite alargar a gama de métodos de visualização que podem ser aplicadas a uma situação específica, uma vez que existe uma relação estreita entre os tipos de dados e os métodos de visualização adequados à sua representação visual. Nesta perspetiva, um aspeto com implicações importantes na escolha dos métodos de visualização a usar é o facto do *data set* poder ser inerentemente espacial (como os dados do corpo humano obtidos a partir de uma modalidade auxiliar de diagnóstico como a Ressonância Magnética), ou não (como os dados provenientes de um inquérito). Esta questão está na base da subdivisão da Visualização em Visualização Científica (SciVis) e Visualização de Informação (InfoVis) (Rhyne, 2003); sendo que a primeira trata tradicionalmente situações em que os dados são inerentemente espaciais, o que implica em geral a necessidade de compreensão da

sua estrutura espacial, fornecendo uma “metáfora” intuitiva para a representação visual dos dados, e na segunda a metáfora é escolhida pelo designer.

A seguir apresentam-se tipos de dados que são relevantes, classificados segundo o seu tipo. Foi seguida uma classificação proposta em (Shneiderman, 1996) e sistematizada em (Andrews, 2002).

## 1D

Tipo de dados linear que inclui documentos textuais, código fonte de programas, e listas de nomes ordenadas alfabeticamente que estão organizados numa forma sequencial. Cada item da lista consiste numa linha de texto que contém uma *string* de caracteres. Atributos adicionais da linha podem aparecer com a data da última atualização ou o nome do autor. Pormenores do design da interface incluem que fontes, cores, tamanho a usar ou que métodos podem ser usados para ver, percorrer ou selecionar. Alguns dos problemas que o utilizador pode sentir passam por determinar o número de itens, ver os itens que contenham um determinado atributo (por exemplo, mostrar só as linhas que sejam títulos de uma secção, linhas que foram alteradas desde a última versão, ou pessoas com mais de 18 anos), ou ver um item com todos os seus atributos.

## 2D

Dados planares ou mapas podem incluir mapas geográficos, plantas de casas, ou *layouts* de um jornal. Cada item de um conjunto cobre uma parte ou a totalidade da área e pode ser retangular ou não. Cada item tem atributos relacionados com a tarefa tais como nome, dono, valor, etc., e atributos relacionados com as propriedades tais como tamanho, cor, opacidade, etc. Enquanto muitos sistemas optam pela abordagem da utilização de múltiplas camadas para lidar com dados do tipo mapa, cada camada é bidimensional. Os utilizadores normalmente têm problemas em encontrar itens adjacentes, quando um item contém outro, caminhos entre itens ou em tarefas básicas como contagem, filtragem ou detalhes a pedido.

## 3D

Objetos do mundo real como moléculas, o corpo humano e edifícios têm itens com volume e alguma relação potencialmente complexa com outros itens. Os sistemas de *design*, assistidos por computador, usados por pelos arquitetos, pelos modeladores de sólidos, e pelos engenheiros mecânicos são feitos para manipular relações tridimensionais complexas. Os utilizadores têm que lidar com tarefas de identificação de relações de proximidade, acima – abaixo, ou dentro – fora, bem como com as tarefas mais básicas. Nas aplicações 3D os utilizadores têm que perceber a sua posição e orientação quando veem objetos e ainda lidar com o problema da oclusão, que pode ser

sério. Muitos protótipos apresentam soluções para alguns destes problemas recorrendo a técnicas como vista geral, recurso a marcas, perspetiva, estereoscopia, transparência e codificação de cor.

## **Multidimensional**

A maioria das bases de dados relacionais e estatísticas podem ser convenientemente manipuladas como dados multi-dimensionais em que itens com  $n$  atributos aparecem como pontos num espaço  $n$ -dimensional.

## **Temporal**

As linhas temporais são usadas largamente em registos médicos, gestão de projetos ou apresentações históricas para criar um tipo de dados que é separado de tipo de dados unidimensional. A distinção nos dados temporais está no facto de que os itens têm um tempo de início e de fim e que os itens se podem sobrepor.

## **Hierarquias**

As hierarquias ou estruturas em árvore consistem num conjunto de itens em que cada item tem uma ligação a um item “pai” (exceto o item “raiz”). Os itens e as ligações entre “pais” e “filhos” podem ter múltiplos atributos.

## **Rede / malha**

Por vezes, as relações entre itens não são convenientemente capturadas recorrendo a uma estrutura em árvore e é útil ter itens ligados a um número arbitrário de outros itens. Apesar de existirem muitos casos especiais de redes (acíclicas, *lattices*, com raiz vs. sem raiz, direcionada vs. não direcionada) parece conveniente considerá-las com um único tipo de dados.

## **Espaço vetorial**

Os espaços vetoriais são usados para representar informação como a obtida quando se analisa o corpo de um documento de texto ou se tenta caracterizar as medidas de semelhança entre documentos.

### **4.2.2. Métodos de visualização**

Os métodos de visualização podem ser agrupados em várias categorias. A seguir apresenta-se uma classificação tendo por base a proposta usada em (Keim, 2001). Outra

possível taxonomia encontra-se em (Spence, 2007), baseada no número de atributos e no tipo de estrutura dos dados.

### **Representação baseada em gráficos (2D, 3D normais)**

Usa métodos de visualização 2D (como gráficos x-y) ou 3D standard para visualizar os dados.

### **Representação hierárquica**

Permite visualizar os dados de forma hierárquica. No caso de dados multidimensionais, as dimensões dos dados a ser usadas têm que ser cuidadosamente selecionadas. Para se obterem visualizações úteis, as dimensões mais importantes têm que corresponder aos primeiros níveis da hierarquia.

### **Representação orientada para o pixel (densidade)**

Esta representação permite visualizar os valores de cada dimensão sob a forma de uma cor de um pixel ou de um grupo de pixéis de uma área adjacente. Ao organizar e colorir os pixéis de forma apropriada, a visualização resultante fornece informação detalhada sobre as correlações e dependências locais, bem como os pontos de maior interesse.

### **Representação baseada em ícones**

Permite visualizar cada item de dados como um ícone (como por exemplo figuras de ligação) e os valores de dimensão como propriedades dos ícones.

### **Representação baseada em transformações geométricas**

Permite usar transformações geométricas e projeções para produzir visualizações úteis. Podem incluir-se aqui as coordenadas paralelas, projeção para um ponto de fuga e os vários métodos de visualização de grafos e redes. Também (Spence, 2007) apresenta uma possível organização dos métodos de visualização.

## **4.2.3. Técnicas de apresentação e interação**

Para além dos métodos de visualização que resolvem o problema da representação visual, como referido anteriormente, a exploração dos dados também implica a utilização de técnicas de apresentação e de interação; as primeiras resolvem o problema do que apresentar em cada momento ao utilizador, as segundas permitem-lhe alterar dinamicamente as visualizações de acordo com os seus objetivos, bem como relacionar e

combinar múltiplas visualizações independentes. Em muitos casos, as técnicas de apresentação e interação estão ligadas de forma estreita, não sendo possível compreendê-las apenas dum dos pontos de vista. A sua escolha num caso concreto depende muito do perfil dos utilizadores, das tarefas que terão que desempenhar com base nas visualizações e do contexto em que o farão. Referem-se a seguir algumas das técnicas de apresentação e interação mais comuns (Keim, 2001, 2002; Shneiderman, 1996; Spence, 2007).

## Projeção

Numa projeção, são mantidos os itens com perda de informação correspondente a alguma(s) dimensão(ões). Há muitos tipos de projeções 3D em 2D, nomeadamente as projeções geométricas planas muito usadas em Computação Gráfica. Estas projeções podem dividir-se em paralelas e perspectivas, sendo que as últimas são mais complexas mas permitem um maior realismo que as projeções paralelas (Hearn & Baker, 2004). As projeções utilizadas em mapas projetam a superfície da Terra numa outra superfície e também neste caso a escolha da projeção a usar deve atender às características das tarefas a realizar pelos utilizadores (Munzner, 2014).

## Filtragem

Permite reduzir o número de elementos mostrados através da supressão de alguns itens e/ou atributos. Uma forma útil de aplicar a filtragem consiste em permitir ao utilizador a seleção de uma ou mais gamas de interesse. Em contexto de visualização interativa é frequente a utilização de *queries* dinâmicas, permitindo ao utilizador ver imediatamente os resultados da filtragem (Munzner, 2014). No entanto, numerosas técnicas de filtragem interativa têm sido desenvolvidas com o objetivo de proporcionar uma melhor exploração dos dados, sendo as *Magic Lenses* (Bier et al., 1993) outra técnica de filtragem interativa.

## Zoom

Consiste na ampliação dos itens com interesse para análise de detalhe, permitindo a sua melhor compreensão; ou na sua diminuição obtendo uma visão mais geral (*overview*), o que corresponde a uma agregação. Além do *zoom* meramente geométrico, existe o *zoom* semântico em que a representação dos objetos se adapta ao espaço disponível para a sua representação podendo variar bastante para diferentes escalas (Cockburn et al., 2008; Munzner, 2014; Spence, 2007). Em conjunto com o *zoom*, usa-se frequentemente o *pan* (movimento suave de enquadramento numa imagem) que permite navegar na representação dos dados e pode considerar-se que resulta numa filtragem com base no *layout* espacial 2D.

## ***Linking and Brushing***

É uma forma muito usada para estabelecer uma relação entre itens que são selecionados numa vista e a sua representação noutras vistas. Também é conhecido por *linked highlighting* e, relacionando várias vistas, permite ultrapassar limitações ao número de atributos que podem ser representados simultaneamente numa única vista. A maioria das aplicações de visualização que permitem exploração interativa dos dados permite alguma forma de *linking and brushing*, sendo frequentemente usado para combinar várias visualizações como gráficos de barras, coordenadas paralelas ou *scatterplots* múltiplos (Keim, 2002; Munzner, 2014; Spence, 2014).

## ***Scrolling***

Permite deslocar o conteúdo no ecrã (vertical ou horizontal) de maneira a ser possível visualizar uma vista que, de outra maneira, precisaria de uma área de representação maior do que a disponível no ecrã. O *scrolling* aparece muitas vezes associado a outras técnicas, como o *zoom*, permitindo a observação de detalhes de uma determinada zona, sendo fácil passar a observar qualquer zona adjacente fazendo *scrolling* (arrastando a vista) na direção pretendida.

## ***Overview + Detail***

O conhecido mantra de Ben Shneiderman, “*Overview First, Zoom and Filter, Details on Demand*” (Shneiderman, 1996) enfatiza a relação entre a necessidade de fornecer uma visão geral (*overview*) e de mostrar detalhes, bem como a importância da redução de dados e dos mecanismos de navegação. O *overview* permite ao utilizador obter uma noção geral dos dados, sem ser necessário usar métodos de navegação (*pan* ou *scroll*), e descobrir onde será mais proveitoso procurar detalhes. Uma forma simples de criar *overviews* consiste em fazer *zooming out* geométrico mostrando todos os itens (Munzner, 2014).

## ***Distorção***

Consiste em apresentar uma parte dos dados com um elevado grau de detalhe integrada numa única vista que também mostra uma maior porção dos dados. Apoia o processo de exploração dos dados mantendo uma vista mais geral (*overview*) durante as operações de obtenção de detalhe (*drill-down*). De acordo com Tamara Munzner (Munzner, 2014), as técnicas de distorção pertencem ao tipo mais geral *focus+context* e têm como objetivo diminuir a desorientação potencial que pode ser provocada pelo simples *zoom* geométrico fornecendo informação de contexto que reduz a carga cognitiva do utilizador. Têm sido propostas muitas técnicas de distorção que, de acordo com a mesma autora, podem ser agrupadas em cinco tipos: *3D perspective*, *fish eye lenses*, *hyperbolic*

*geometry, stretch and squish navigation, e magnification fields*, sendo algumas das mais conhecidas ilustradas em (Furnas, 1986, 2006; Jonh Lamping & Rao, 1996; Mackinlay et al., 1991; Munzner, 1998; Sarkar & Brown, 1994; Slack & Munzner, 2006).

### 4.3. Visualização de estruturas de dados lineares

Quando se fala de estruturas de dados lineares normalmente estão em causa dados lineares que podem ter uma ou mais dimensões.

Entre as técnicas disponíveis para a visualização de estruturas lineares podem encontrar-se os gráficos como os de barras, colunas, linhas, área, superfície, diagramas de queijo, tipo radar, ou com codificação de cor.

O mercado dispõe de várias soluções, tanto comerciais como gratuitas, para a visualização deste tipo de dados:

- Folhas de cálculo como Microsoft Excel<sup>61</sup>, OpenOffice.org Calc<sup>62</sup>, Google Sheets<sup>63</sup> ou Apple Numbers for Mac<sup>64</sup>, permitem produzir gráficos a partir de dados lineares 1D ou 2D de forma muito fácil. Estes programas costumam dispor de um leque bastante alargado de tipos de gráficos que podem ser utilizados, além de permitirem fazer vários tipos de ajustes sobre esses gráficos, tudo de uma forma relativamente simples;
- Programas de computação técnica como Matlab<sup>65</sup> ou Mathematica<sup>66</sup>, permitem produzir gráficos a partir de dados 1D, 2D ou mesmo 3D. Neste tipo de programas é possível, adicionalmente, efetuar algum processamento / manipulação sobre os dados a visualizar e de seguida criar um gráfico;
- Programas como o Processing<sup>67</sup> conseguem ir ainda mais longe, permitindo a visualização de dados de vários tipos, não só de dados lineares como muitos outros tipos de dados mais complexos.
- Existem ainda dezenas de outras aplicações e bibliotecas de visualização que permitem a produção de gráficos a partir de vários tipos de dados (Munzner, 2014; Spence, 2014).

---

<sup>61</sup> Microsoft Excel: <https://products.office.com/excel>

<sup>62</sup> OpenOffice Calc: <https://www.openoffice.org/>

<sup>63</sup> Google Sheets: <https://www.google.com/sheets/about/>

<sup>64</sup> Apple Numbers: <http://www.apple.com/mac/numbers/>

<sup>65</sup> Matlab: <http://www.mathworks.com/products/matlab/>

<sup>66</sup> Mathematica: <https://www.wolfram.com/mathematica/>

<sup>67</sup> Processing: <https://processing.org/>

## 4.4. Visualização de redes e grafos

Existem vários tipos de estruturas hierárquicas (como redes ou grafos) passíveis de ser visualizadas. No âmbito desta tese, há um tipo de estruturas deste tipo particularmente importante para este trabalho: a estrutura de um *site*, com as suas páginas e respetivas ligações.

Na bibliografia, para além das abordagens mais tradicionais, podem ser encontradas algumas soluções interessantes. Stephen Eick (Eick, 2001) apresenta algumas soluções para a visualização de *sites*. Uma, desenvolvida originalmente na Xerox PARC (John Lamping & Rao, 1994), propõe uma solução de árvores hiperbólicas em que as páginas são colocadas num espaço hiperbólico. Outra, desenvolvida por Tamara Munzner (Munzner, 1998), propõe uma solução de exploração de gráficos de grandes dimensões num espaço hiperbólico 3D em que usa um espaço hiperbólico para distribuir os ramos da árvore, reduzindo a confusão visual, acelerar o desenho e fornecer uma interação mais fluida.

Daniel Keim (Keim, 2001) sugere uma alternativa, apresentada originalmente por Matthias Kreuzler (Kreuzler et al., 2000), de uma técnica de visualização de hierarquias num hemisfério (esférico) que utiliza uma técnica de processamento e visualização escaláveis que permite a seleção de nós de um cluster e/ou dos níveis da hierarquia.

Há situações em que a estrutura da informação pode ser composta por múltiplas hierarquias que se intersectam que podem ser chamadas de poli-hierarquias. George Robertson (Robertson et al., 2002) propõe métodos de visualização com o objetivo de permitir descobrir relações entre as hierarquias que de outra maneira seria muito difícil.

Benjamin Bederson (Bederson, 2001) propõe uma função de zoom para a análise de informação do tipo hierárquica recorrendo a variações de algoritmos existentes de *treemaps* (mapas em árvore) e de *bubblemaps* (mapas de bolha).

Steve Benford (Benford et al., 1999) compila um conjunto de métodos para a visualização da estrutura da *World Wide Web*. Aborda métodos para a visualização de informação do tipo hierárquica e rede, recorrendo sobretudo a representações 3D.

Agnello (Agnello & Levkowitz, 2015) apresenta uma solução que tem uma grande proximidade na abordagem ao que é proposto, suportado em "*user-driven analytics*" e o facto de suportar o *engine* de processamento em dados pré-classificados para ajudar a estudar e comparar "*profiles*". Ainda que num contexto de *Big Data*, esta aplicação "Quick Vis" continua a considerar pertinente o comportamento de interação, na verdade considerando-o um dos fatores orientadores da pesquisa visual e na definição da estrutura de "*profiles*" proposto.



## 4.5. Visualização de redes

James Abello apresenta em (Abello & Korn, 2000), complementado em (Abello & Korn, 2002), um sistema para visualizar estruturas multi-digráficas, apresenta um interface integrado que consiste na soma de várias janelas contendo esquemas diferentes num mesmo ecrã, organizadas manualmente de forma a não existir sobreposição. Utiliza teoria de grafos e visualização científica, usa a taxonomia de Ben Shneiderman (Shneiderman, 1996) como orientação.

Carl Gutwin (Gutwin, 2002) e (Gutwin & Skopik, 2003) apresenta um método que melhora o desempenho da técnica *fisheye view* (vista olho de peixe) para visualizar dados em rede ou árvore. Aposta sobretudo na melhoria da função de focagem com a definição do contexto de observação/análise.

Benjamin Bederson (Bederson, 2000) utiliza uma técnica semelhante, *fisheye menus* (menus em olho de peixe), aplicada a menus em que o tamanho dos itens dos menus são dinamicamente redimensionados para proporcionar uma zona de foco à volta da posição do cursor do rato.

## 4.6. Visualização de informação multidimensional

A visualização de informação foca-se em conjuntos de dados que nem sempre tem uma semântica 2D ou 3D inerente e, como tal, sem um mapeamento direto dos dados abstratos para um espaço físico como o papel ou o ecrã. Apesar de existirem vários métodos bem conhecidos que permitem a visualização (ainda que possa ser só parcial) de dados, estão limitadas a dados com um número de dimensões relativamente reduzidas. Para ajudar a resolver este problema, alguns autores avançam com soluções para a visualização de dados com mais de três dimensões.

Alfred Inselberg (Inselberg & Dimsdale, 1990) referido em (Keim, 2001) apresenta uma técnica de coordenadas paralelas, em que se faz o mapeamento (não projetivo) de um espaço n-dimensional para um 2D. Nesta técnica, as hiper-superfícies são representadas pelas suas imagens planares que têm algumas das propriedades geométricas análogas às propriedades das hiper-superfícies que representam.

Kent Wittenburg (Wittenburg et al., 2001) apresenta técnicas de visualização e interação aplicadas a um sistema de apoio à decisão baseado em diagramas de barras paralelos, com a possibilidade de fazer pesquisas de forma dinâmica.

Eleftherios Koutsofios (Koutsofios et al., 1999), referido por Keim (Keim, 2001) aborda uma técnica de visualização, sistema *SWIFT-3D*, aplicada a informação georreferenciada.

Daniel Keim (Keim, 2000, 2001) apresenta uma técnica de visualização, segmentos circulares orientados ao pixel (*pixel-oriented*). Esta técnica mapeia valores em pixels coloridos, em que, por exemplo, valores mais altos correspondem a cores mais claras.

## 4.7. Visualização de *sites web*

Na bibliografia é possível encontrar vários trabalhos que abordam problemas semelhantes aos tratados nesta tese. As soluções adotadas por esses trabalhos são, nuns casos semelhantes, noutros completamente dispares das adotadas nesta tese. Esta seção aborda alguns desses trabalhos.

John Cugini (Cugini & Scholtz, 1999), referido em (Eick, 2001), apresenta uma aplicação que permite visualizar em 3D os percursos dos utilizadores num *site*. Esta permite, de forma dinâmica, personalizar e simplificar a representação gráfica do *site* e selecionar quais os percursos de utilizador a visualizar. Usa a terceira dimensão para representar o tempo despendido na visita de cada página.

Stephen Eick (Eick, 2001) apresenta uma solução de representação simbólica da monitorização da atividade em tempo real dum *site* de comércio eletrónico.

Hendley (Hendley et al., 1996) apresenta uma ferramenta, o sistema Narcissus, que permite a visualização de espaços de dados de grandes dimensões, como a *World Wide Web*. A solução usa uma representação 3D na qual o utilizador pode navegar proporcionando uma visualização mais rica.

Hochheiser (Hochheiser & Shneiderman, 1999) apresenta um conjunto de ferramentas de visualização *Starfield* (“campo de estrelas”) como a aplicação *Spotfire* que combina a visualização simultânea de um número elevado de pontos de dados com uma interface integrada que proporciona funcionalidade de zoom, filtragem e pesquisa dinâmica.

Carpendale (M. S. T. Carpendale & Montagnese, 2001) apresenta um objetivo idêntico ao desta tese para a construção de uma interface de visualização “*unified presentation space*” (com um “espaço de apresentação unificado”), recorrendo a uma metáfora de lente. Usando uma metáfora semelhante, menus *fisheye*, Bederson (Bederson, 2000) propõe um *toolkit* gráfico (conjunto de ferramentas) em JAVA 2.0. Igarashi (Igarashi & Hinckley, 2000) propõe uma função complementar para a operação de zooming, tornando-a dependente da ação do utilizador em concreto, da operação de *scroll*.

Também na visualização de padrões de utilização de serviços de comunicação e informação existem algumas abordagens na bibliografia com muitas semelhanças à adotada nesta tese.

Danyel Fisher (Fisher, 2000) apresenta uma proposta para estudar a eficiência de um fórum, em concreto se um determinado assunto/questão publicada está a ter ou teve

adesão e qual a dimensão da participação. Tenta também perceber as ramificações autónomas que gerou a partir do assunto original.

Andreas Dieberger (Dieberger & Lönnqvist, 2000) propõe um instrumento de registo de “histórico de navegação” e propõe uma solução para tentar tornar o *site* mais “social” de maneira a os utilizadores se envolvam no que designa por “navegação social”.

A visualização da “evolução de uma ecologia *web*” representa o problema central da investigação de Ed Chi (Chi et al., 1998). Conceptualmente, procura implementar um conjunto de “árvores em disco” (*disk tree*) que formam um “tubo temporal” (*time tube*), representando a evolução da *web* em longos períodos de tempo.

Uma abordagem muito semelhante, com os mesmos objetivos, à desta tese é referida por Ed Chi em (Chi et al., 2000). Neste caso, a grande diferença situa-se no paradigma de interface de análise que não usa como referência o espaço *web* inicial nem o toma em consideração no contexto de representação de resultados. O sistema tenta responder a um conjunto de questões muito semelhantes às colocadas por esta tese, nomeadamente no que respeita ao *site* como um todo, páginas específicas e aos utilizadores, que passamos a citar (Chi et al., 2000):

- *Site como um todo*. Qual o fluxo de tráfego geral atual? Quais são os caminhos de navegação, atuais e previstos? Como mede o *site* a facilidade com que se encontra a informação pretendida e o seu custo?
- *Página em concreto*. De onde vêm os visitantes (que caminhos percorrem)? Para onde vão? Que outras páginas estão relacionadas?
- *Utilizadores*. Quais os interesses dos visitantes desta página? Onde se pensa que deverão ir, tendo em conta os seus interesses? A informação de uso confirma a predição e porquê? Qual o custo (por exemplo, em termos de tempo de tempo de download) de navegação para esses visitantes?

A proposta apresentada por Baoyao Zhou em (Zhou et al., 2001) tem muitas semelhanças com a apresentada nesta tese, ainda que recaia apenas sobre a análise da estrutura do *site*, avalia a integridade com base em informação de uso. Também usam ficheiros de *log* para extrair a informação que necessitam e definem os critérios para detetar e legitimar sessões de utilizador.

Em (Hochheiser & Shneiderman, 1999) é apresentada uma solução chamada *Starfield Visualizations of WWW Log Data*. Esta representa um sistema que procura formular respostas a questões similares à desta tese. Também usa como base ficheiros de *log*, tal como o sistema desenvolvido no âmbito desta tese. A caracterização e modelação dos padrões de uso de um *site* continua a ser uma área atual de investigação. Uma das limitações da abordagem destes autores está no facto de usarem uma metáfora visual de análise diferente e descontextualizada da informação original. Perde-se o contexto (de imagem) de página do *site* assim como da sua relação com outras páginas.

## 5. Usabilidade de *sites web*

Nesta secção serão apresentados alguns trabalhos que abordam a questão da usabilidade de um *site*. Muitos trabalhos encontrados na bibliografia recorrem à análise ficheiros de *log* do *site* para fazer a análise de usabilidade. As métricas usadas também serão abordadas. Na parte final serão também referidos alguns estudos de usabilidade aplicados a contextos de uso específicos.

### 5.1. Ficheiro de *log* de *site web*

Melody Ivory (Ivory et al., 2001) propõe um método empírico para se chegar às “*design guidelines*” de um *site web* e define quais as métricas mais adequadas para efetuar estudos de usabilidade deste género. Analisa os ficheiros de *log* dos *sites* ainda que admita esta análise possa ser inconclusiva porque os ficheiros fornecem rastros incompletos do comportamento dos visitantes além de que as estimativas temporais podem ser enviesadas pelas latências da rede.

Ed Chi (Chi et al., 2000), também faz uma abordagem à utilização de ficheiros de *log* de *sites* assim como a outros métodos alternativos.

Também Harry Hochheiser (Hochheiser & Shneiderman, 1999) utiliza ficheiros de *log* de *sites* para extrair a informação de análise. Discutem e apresentam algumas condicionantes e ambiguidades típicas da informação registada nos ficheiros de *log*.

Laila Paganelli (Paganelli & Paternò, 2002) apresenta um método que recorre à inclusão de rotinas javascript em cada página HTML para detetar eventos de interação. Esta estratégia é usada em complementaridade com os ficheiros de *log* do *site*. O resultado é semelhante ao do “*interceptor*” proposto nesta tese (secção 3.2 do capítulo seguinte). Há, contudo algumas diferenças: a metodologia de Paganelli exige a inclusão prévia de código javascript em cada página de maneira a garantir uma recolha de informação de qualquer equipamento terminal (garantidas as opções de execução de javascript). Em contraposição, a proposta desta tese obriga a uma instalação autorizada e controlada do “*interceptor*” no dispositivo a monitorizar, permitindo, contudo, pela natureza do método, recolher muito mais informação do processo de interação.

### 5.2. Métrica

Melody Ivory (Ivory et al., 2001) propõe um conjunto de métricas que têm em consideração a composição da página (número de palavras, *link*, gráficos), formatação da

página (texto enfatizado, sua posição, agrupamento) e características gerais da página (tamanho da página, velocidade de download).

Chris Coulston (Coulston & Vitolo, 2001) apresenta uma métrica para detetar e comparar a “profundidade de navegação” versus “percurso teórico ótimo”.

### 5.3. Estudos de usabilidade aplicados a *sites web*

O estudo apresentado por Fogg (Fogg et al., 2001) remete para uma área afim a esta tese e referente à “credibilidade de um *site web*”. Em concreto, apresenta como conclusão do estudo empírico elementos que contribuiriam para dar e retirar credibilidade a um determinado *site*.

Elementos com contributo para a credibilidade do *site*:

- Relação / semelhança com o mundo real;
- Facilidade de utilização;
- Nível de experiência / discernimento;
- Grau de confiança;
- Formatação / adequação à necessidade do utilizador;

Elementos com contributo para a descredibilização do *site*:

- Implicações comerciais;
- Amadorismo (ex.: erros ortográficos, links que não funcionam, *site* por vezes indisponível, baixa taxa de atualização do *site*).

A equipa de Fogg recorreu a um questionário como instrumento de base para o estudo.

Melody Ivory (Ivory & Hearst, 2002) apresenta um sistema que tem por objetivo efetuar uma análise de usabilidade semiautomática sobre um dado *site*. Usa como referência as recomendações e princípios de Nielsen (Jakob Nielsen, 1993; Jakob Nielsen & Mack, 1994) entre outros. Estabelece métricas sobre o conteúdo de uma página e a sua relação com avaliação de usabilidade geral sustentada em informação estatística. Este trabalho poderia ser complementado com a abordagem usada nesta tese através do uso do elemento unitário “*hotspot*” (introduzido na secção 3.1 do capítulo seguinte.).

Em (Kelly et al., 2002) é apresentado um estudo de usabilidade no âmbito de duas comunidades on-line com o intuito de obter “informação de utilização” das respetivas comunidades. Tem por objetivo melhorar o desempenho das comunidades on-line dando-lhes conhecimento dos estudos de utilização que são feitos sobre as próprias.

Em (Riedl & Amant, 2002) é apresentada uma solução de usabilidade para produtos de software que também poderia ser utilizada em *sites web*, mas que apenas extrai uma representação relacional da interface e os percursos possíveis até um determinado comando. Esta solução utiliza um modelo de base (para interface e *widgets*) que assenta

no paradigma Windows, sendo, por isso, pouco robusto e condicionado para que seja aplicado a outros produtos, como por exemplo o Adobe-Photoshop ou muitos *sites web*.

A solução apresentada em (Faraday & Sutcliffe, 1998) utiliza tecnologia de *eye tracking* para estudar o percurso do olhar num processo de estudo de um manual para apoio à substituição de *toner* numa impressora. Estuda como são formados os pontos de contacto entre sequências de texto e imagem quando o sujeito estuda o material fornecido. Apesar de o objeto de estudo ser diferente, a abordagem deste trabalho tem algumas semelhanças com a deteção de “mancha” usado nesta tese.

## 6. Ferramentas de análise de *log's*

Na secção 5.1 acima neste capítulo, os ficheiros de *log* já foram abordados numa perspetiva de fonte de informação para a aplicação de estudos de usabilidade. Nesta secção faz-se a avaliação de um par de aplicações comerciais usando ficheiros de *log* de *sites* reais conhecidos.

Existem no mercado inúmeras aplicações que permitem a análise de ficheiros de *log* de servidores *web*, aplicações estas normalmente denominadas de “*Log Analyzers*”. Algumas destas aplicações são específicas para determinada plataforma, outras mais universais, podem ser utilizadas para a análise de ficheiros obtidos dum leque mais alargado de combinações entre servidores *web*, sistemas operativos e plataformas de hardware. É possível ainda encontrar aplicações especialmente desenhadas para fazer um tipo de análise muito específico, havendo outras que permitem análises mais abrangentes, fornecendo um leque de informação mais alargado.

Entre a informação obtida e analisada por este tipo de programas pode encontrar-se informação estatística mais geral como o número de páginas visitadas (*page hits*), páginas mais e menos visitada, tráfego de rede provocado por essas visitas entre outros detalhes. Algumas aplicações conseguem ir mais longe e usam um algoritmo para identificar sessões de utilizador, fornecendo informação adicional como o número de visitantes únicos, o ponto de entrada de cada sessão e respetivo percurso dentro do *site*, tempo de permanência da sessão, entre outros detalhes.

A informação que é possível obter não depende só do programa usado, depende também do nível de detalhe da informação contida nos ficheiros de *log* que tem a ver com a configuração dos servidores que os produzem, bem como pela informação disponibilizada pelos sistemas clientes dos visitantes (que depende da configuração do sistema operativo, navegador, *firewall*, do possível uso de um serviço de *proxy*, entre outros detalhes).

De entre as aplicações a que os autores tiveram acesso (início do ano 2001), escolheram-se duas das mais universais, tanto do ponto de vista das plataformas suportadas (usaram-se ficheiros de *log* de duas plataformas distintas), como pela informação que disponibilizam.

A seguir passa-se a fazer uma breve descrição do FasStats Analyzer v2.7.7 da Mach5 Development LLC<sup>68</sup> e do WebTrends Log Analyzer v6.0a da WebTrends Corporation<sup>69</sup>.

Para esta primeira experiência foi usado o *site* <http://www.cpj.ua.pt>, no desenvolvimento do qual houve um envolvimento pessoal dos autores do trabalho desta tese. Numa fase posterior usou-se também um *site* de *eLearning* da UA, <http://webct.ua.pt> (plataforma entretanto abandonada), com atenção especial virada para uma unidade curricular específica, PEDDA (Programação, Estrutura de Dados e Algoritmos) ministrada ao 1º ano comum das licenciaturas em Engenharia da Universidade de Aveiro, pela sua base alargada de utilizadores.

Após a utilização das duas aplicações conclui-se que estas têm muitas semelhanças. A configuração inicial conta com um conjunto de passos mais ou menos comuns e a informação obtida (tipo e quantidade) é também muito semelhante, diferindo sobretudo na forma como é apresentada e, ainda assim, com muitas semelhanças.

## 6.1. *FastStats Analyzer*

Esta aplicação pode apresentar a informação sob duas formas:

- 1) num formato proprietário, usando a sua própria interface;
- 2) em formato HTML, ainda que uma versão reduzida da informação e com uma apresentação baseada, sobretudo, em tabelas de texto.

A informação é apresentada de forma agradável à vista, com gráficos de vários tipos (como de barras, cartesianos, tipo queijo) dando também a possibilidade de consultar as tabelas correspondentes.

Um dos aspetos apreciados foi a função “*Hyperlink Tree View*” que permite ficar com uma ideia da estrutura do *site*, mostrando as ligações entre as várias páginas. Infelizmente não foi possível fazer esta funcionalidade funcionar com o *site* em estudo. Não foi possível confirmar se o problema se devia ao formato do ficheiro de *log*, se ao tipo de servidor em causa, se devido à estrutura particular deste *site*, ou ainda devido a algum erro do próprio programa. Se tivesse havido disponibilidade financeira para a sua

---

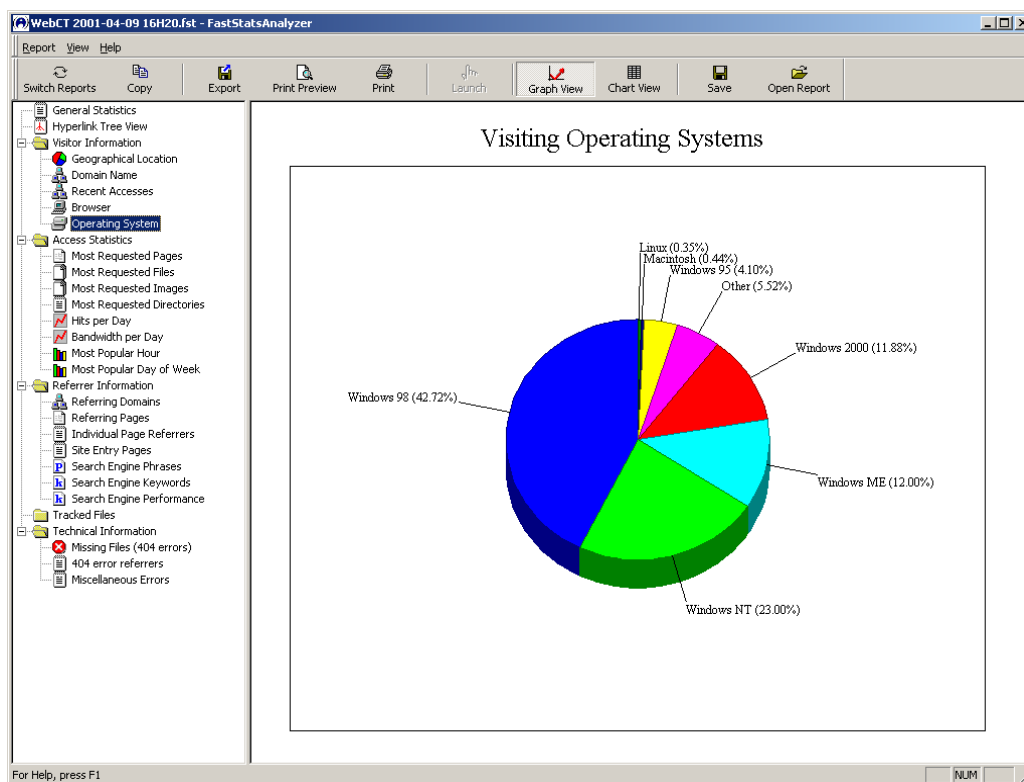
<sup>68</sup> FasStats Analyzer: <http://www.mach5.com/products/faststats/index.html>

<sup>69</sup> WebTrends Log Analyzer: <http://www.webtrends.com>

aquisição, poderia ter sido interessante tentar usar esta funcionalidade novamente, com uma versão posterior da aplicação.

Foi detetado um problema na forma como esta aplicação lida com os ficheiros de *log* obtidos de um servidor IIS da Microsoft. No *site* <http://www.cpj.ua.pt>, que corre sobre este tipo de plataforma, e com a sua configuração específica, os ficheiros de *log*, para além da informação normal correspondente a cada pedido servido, contêm ainda informação adicional, registada periodicamente, referente ao desempenho do servidor, tais como uma medida de utilização de CPU, de memória, etc., consumidos pelo servidor *web*. Esta informação adicional, supérflua do ponto de vista do *Log Analyzer*, impede a análise dos dados registados a partir desse ponto nesse ficheiro específico. Por exemplo, se for a primeira coisa a aparecer num ficheiro de *log*, a informação nele registada será simplesmente totalmente ignorada. Assim houve a necessidade de filtrar previamente os ficheiros de maneira a retirar toda a informação desnecessária.

As figuras seguintes (Figura II-8, Figura II-9, Figura II-10 e Figura II-11) ilustram algumas vistas da funcionalidade disponível no *FastStatsAnalyser*:



**Figura II-8:** Sistemas operativos dos visitantes de um *site web*. Informação obtida utilizando o programa *FastStatsAnalyser v2.77* da Mach5 Software, Inc.



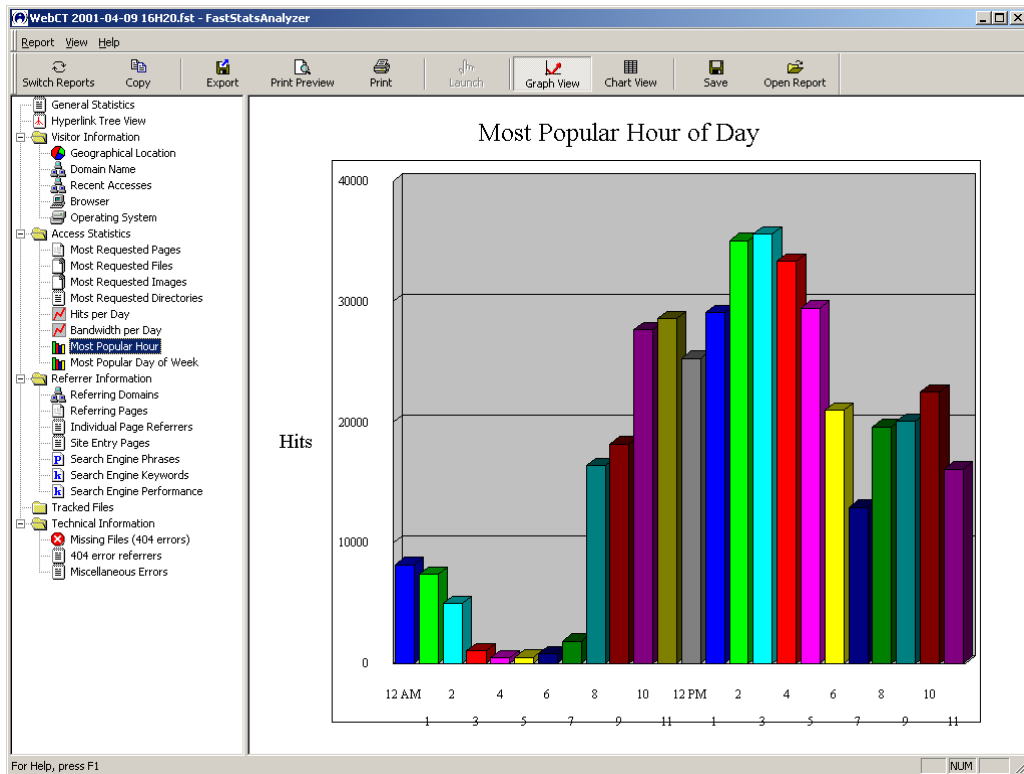


Figura II-9: Horas mais populares do dia na utilização de um *site web*. Informação obtida utilizando o programa FastStatsAnalyzer v2.77 da Mach5 Software, Inc.

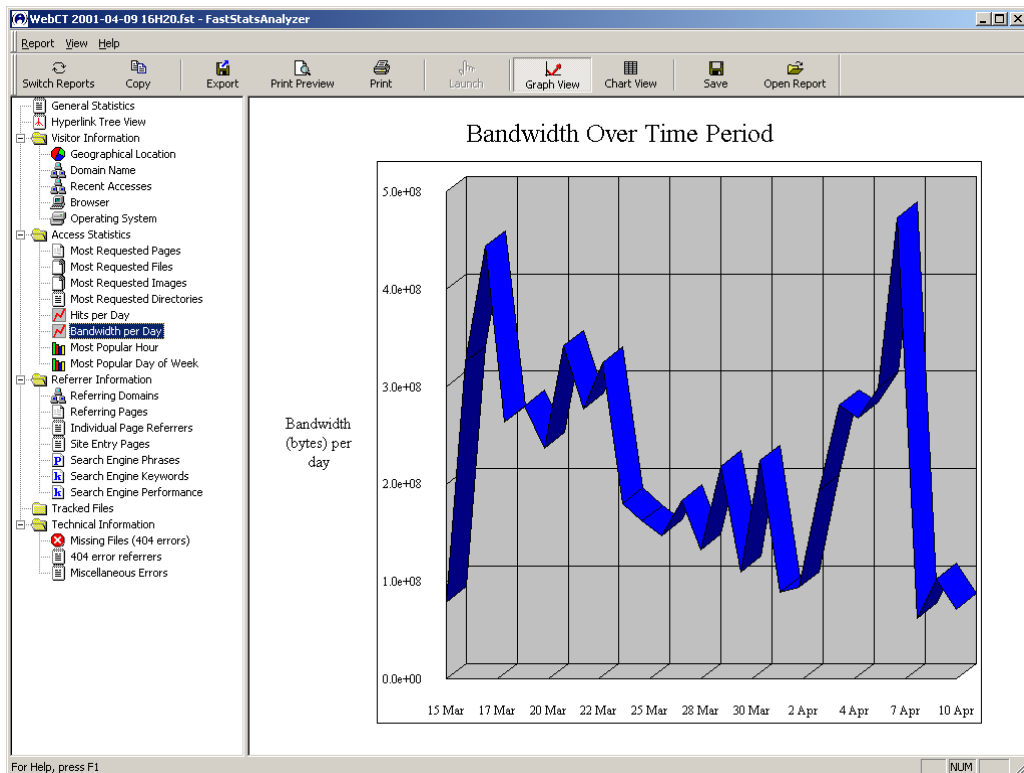


Figura II-10: Utilização da largura de banda ao longo de um período de tempo de um *site web*. Informação obtida utilizando o programa FastStatsAnalyzer v2.77 da Mach5 Software, Inc.

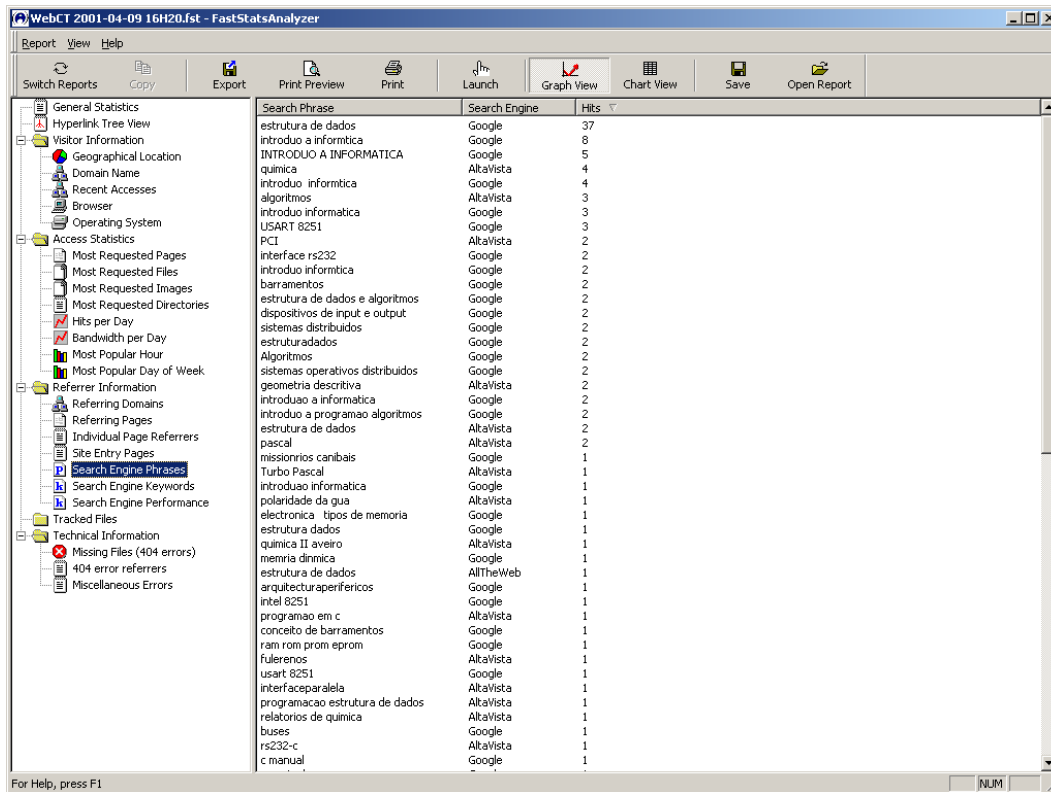


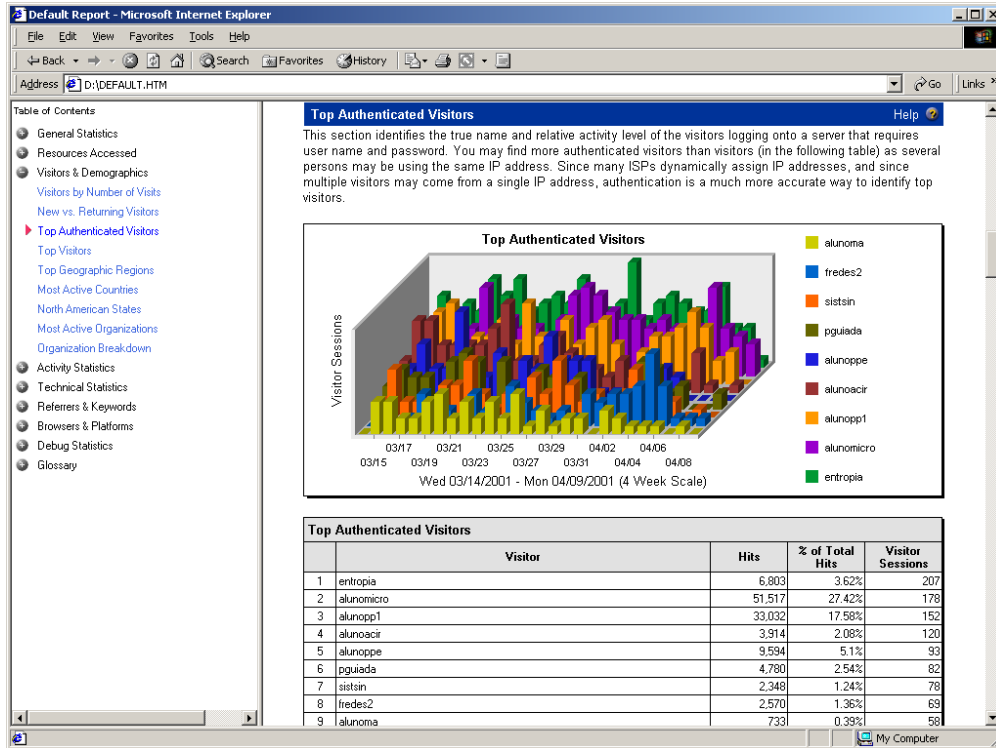
Figura II-11: Frases utilizadas em motores de busca para chegar a um *site web*. Informação obtida utilizando o programa FastStatsAnalyser v2.77 da Mach5 Software, Inc.

## 6.2. WebTrends Log Analyzer

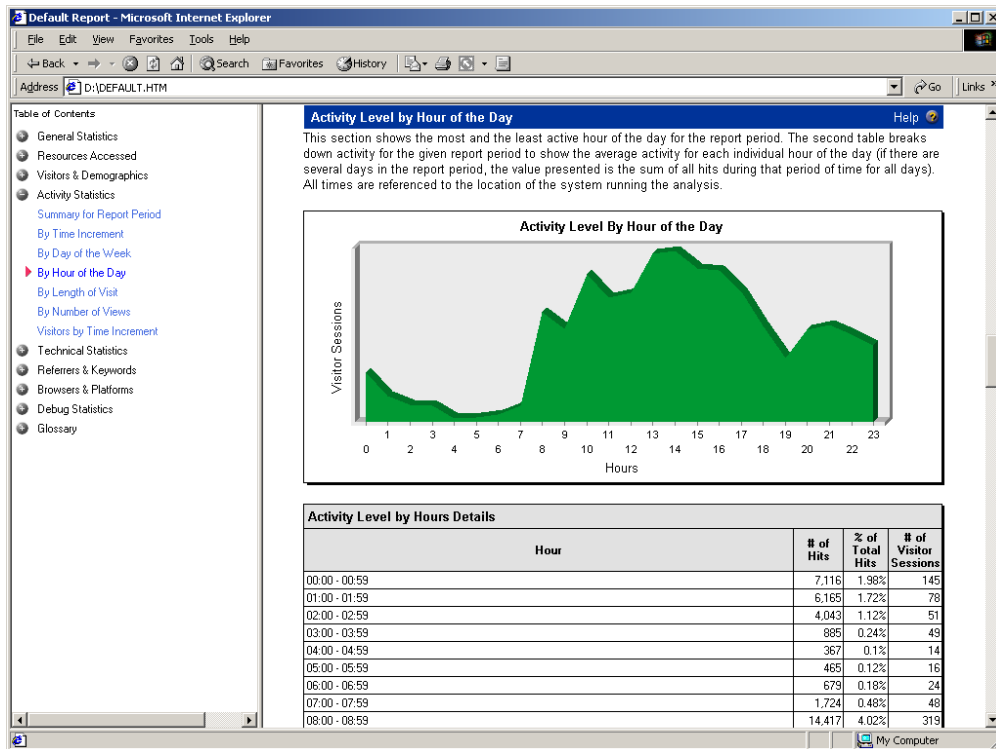
Esta aplicação apresenta os resultados integralmente em formato HTML ou outros, como MS Excel. Este facto não apresenta qualquer limitação à qualidade da apresentação dos resultados. A apresentação de resultados é constituída por gráficos e respetivas tabelas para um estudo mais detalhado, se assim se pretender.

Esta aplicação não apresentou as dificuldades apontadas à anterior, conseguindo ignorar todas as linhas que não é capaz de interpretar corretamente.

As figuras seguintes (Figura II-12e Figura II-13) ilustram algumas vistas da funcionalidade disponível no *WebTrends Log Analyzer*:



**Figura II-12:** Utilizadores autenticados mais ativos num *site web*. Informação obtida utilizando o programa WebTrends Log Analyser v6.0a da WebTrends Corp. Esta ferramenta produz os resultados em formato HTML.



**Figura II-13:** Nível de atividade (número de sessões) por hora do dia num *site web*. Informação obtida utilizando o programa WebTrends Log Analyser v6.0a da WebTrends Corp. Esta ferramenta produz os resultados em formato HTML.

## 6.3. Algumas Conclusões

Como já referido anteriormente, as duas aplicações disponibilizam um conjunto de informação muito semelhante. Em qualquer dos casos é produzida estatística de utilização de *site* em geral, as páginas mais visitadas, os pontos de entrada no *site* mais comuns, a origem das visitas, e sua distribuição demográfica, entre outras informações.

Saliente-se que a possibilidade da análise do “*Referer*” para cada página visitada é uma mais valia importante, na medida em que permite identificar como e de onde vêm os visitantes do *site* em causa.

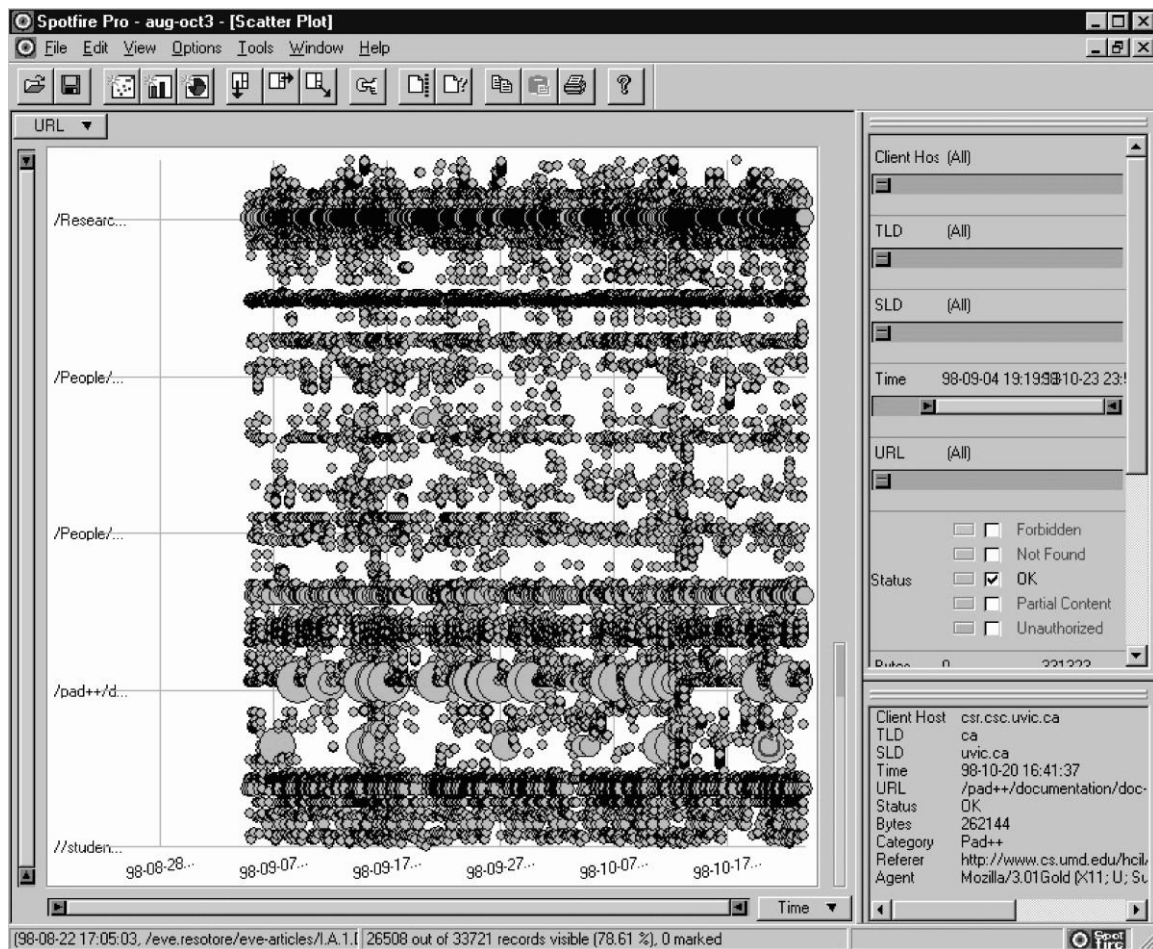
Analisando os resultados obtidos do estudo do *site* <http://www.cpj.ua.pt>, foi possível obter algumas conclusões:

- O grande número de pontos de entrada (para além da página principal, onde normalmente isso acontece) sugere que as sessões estarão a terminar prematuramente, sendo depois retomadas no ponto onde se encontravam (ponto mais ou menos aleatório). De facto, a maioria dos utilizadores deste *site* estavam ligados à *internet* através de uma linha telefónica (“*dial-up*”), por intermédio de um router RDIS. Este router, com o objetivo de reduzir custos, estava configurado para desligar (“*hang-up*”) passados 3 minutos de inatividade, tempo este demasiado curto se se tiver em consideração que estavam disponíveis jogos on-line. Esta situação levava a que durante o processo de navegação no *site*, se o utilizador se detivesse numa página um pouco mais tempo, poderia levar a que o router interrompesse a ligação voltando a estabelecê-la quando outra página fosse solicitada. No entanto esta nova ligação poderia acontecer sob um novo endereço IP, situação mais que provável num ambiente “*dial-up*” com atribuição de endereços IP dinâmica, levando a que seja identificada como uma nova sessão por parte do servidor.
- O elevado número de abandonos logo na página de entrada sugere que pode dever-se ao facto de os computadores da maioria dos utilizadores ligados a este projeto (dois computadores em cada uma das 15 instituições parceiras) estarem configurados com <http://www.cpj.ua.pt> como página pré-definida (“*Home Page*”) dos seus programas de navegação (“*Browsers*”). Assim a simples abertura do navegador para utilização da *internet* provoca um pedido de página (“*Hit*”), ainda que o utilizador só pretendesse visitar páginas noutros locais.

No Capítulo V, abaixo, estas questões serão abordadas com maior profundidade.

## 6.4. Outras abordagens

Podem ainda encontrar-se outras abordagens para este mesmo problema, como o *Starfield* usado na aplicação *Spotfire Pro* desenvolvida por Hochheiser (Hochheiser & Shneiderman, 1999) e ilustrado na Figura II-14, abaixo. No gráfico podem ver-se as páginas identificadas no lado esquerdo, as datas na parte inferior e a dimensão dos círculos representa o número de visitas de uma determinada página num determinado dia.



**Figura II-14:** Vista geral da utilização de um *site web*. Ilustração retirada de (Hochheiser & Shneiderman, 1999).

## 6.5. Evolução das ferramentas de análise (2016)

Desde a realização da parte experimental do trabalho descrito nesta tese, as ferramentas descritas sofreram alguma evolução e apareceram novas soluções no mercado, como o Google Analytics<sup>70</sup>.

A maioria das soluções de análise e monitorização atuais deixaram de utilizar os ficheiros de *log* produzidos pelos servidores *web* como fontes de informação. Este tipo de soluções, para além das limitações já descritas, não permitia análises em tempo real. Em sua substituição, passou a fazer-se a inclusão de código adicional em cada página que se pretende monitorizar, que comunica os dados necessário para a empresa que faz a monitorização (por exemplo a WebTrends ou a Google). Desta maneira é possível, por exemplo, saber quantos utilizadores estão a visitar um *site*, os seus países de origem, em que secção do *site* se encontram a dado momento e, em alguns casos, até o *mac address*<sup>71</sup> do dispositivo de interação de cada um deles.

### ***FastStats Analyzer***

A Mach5, empresa produtora do FastStatAnalyzer, deixou de evoluir este produto, tendo lançado a sua última versão em 2006 (v4.1.7). Esta versão não difere muito da versão estudada em 2001 (v2.77), tendo sido adicionadas algumas novas funcionalidades que permitem analisar detalhes adicionais da utilização do *site web*, como as páginas pontos de entrada e saída mais comuns (Figura II-15) ou a duração média das sessões dos visitantes (Figura II-16). Também já não apresenta o problema encontrado na função “*Hyperlink Tree View*”, que permite ficar com uma ideia da estrutura do *site*, mostrando as ligações de e para uma determinada página, bem como do número de vezes que cada uma delas foi utilizada (Figura II-17). Tem ainda uma nova funcionalidade de análise de cenários que permite relacionar, por exemplo durante uma campanha, o número de visitas oriundas de um determinado motor de busca com o número de visitas à página de encomendas e desta maneira determinar o sucesso dessa campanha.

---

<sup>70</sup> Google Analytics: <http://www.google.com/analytics/>

<sup>71</sup> MAC Address: Endereço MAC (*Media Access Control*) é um endereço físico associado à interface de comunicação, que conecta um dispositivo à rede. O MAC é um endereço “único”, não havendo duas portas com a mesma numeração, é usado para controle de acesso em redes de computadores.

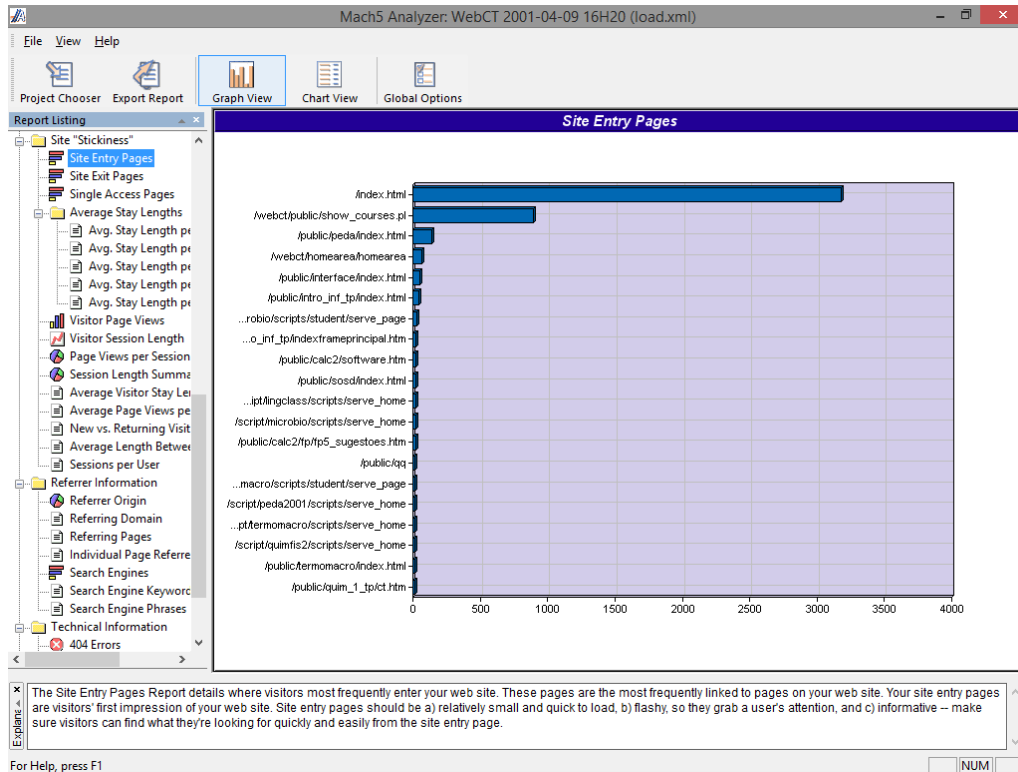


Figura II-15: Página mais usadas como ponto de entrada de um *site web*. Informação obtida utilizando o programa FastStat Analyzer v4.1.7 da Mach5 Software, Inc.

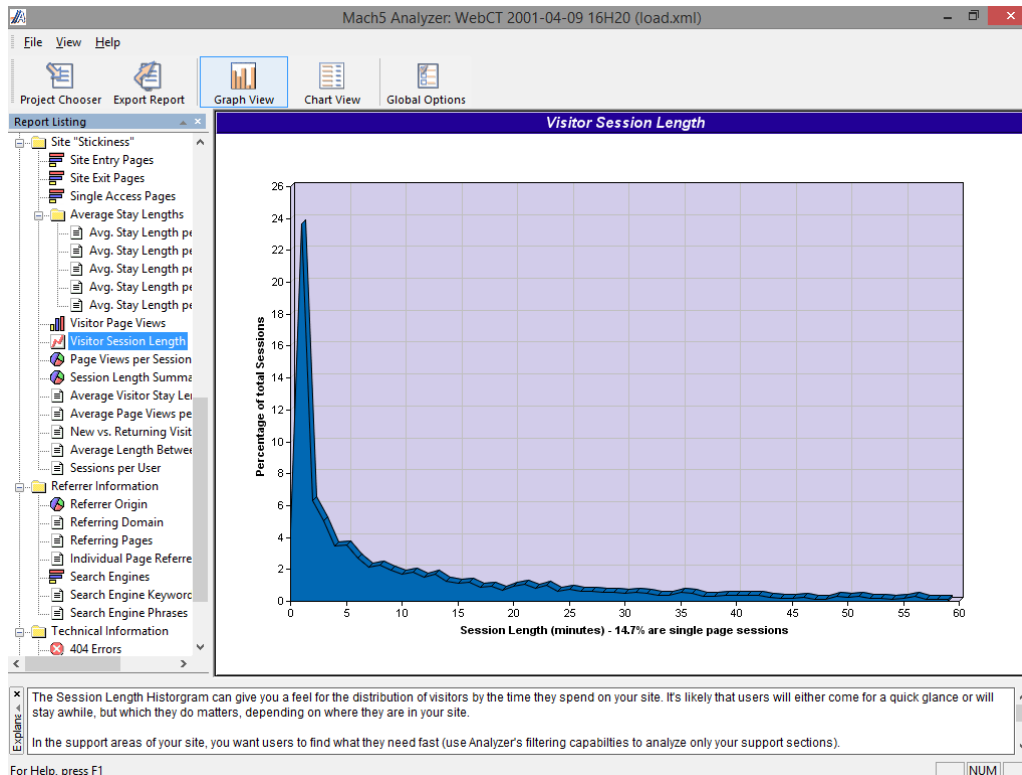


Figura II-16: Duração média das sessões dos visitantes de um *site web*. Informação obtida utilizando o programa FastStat Analyzer v4.1.7 da Mach5 Software, Inc.

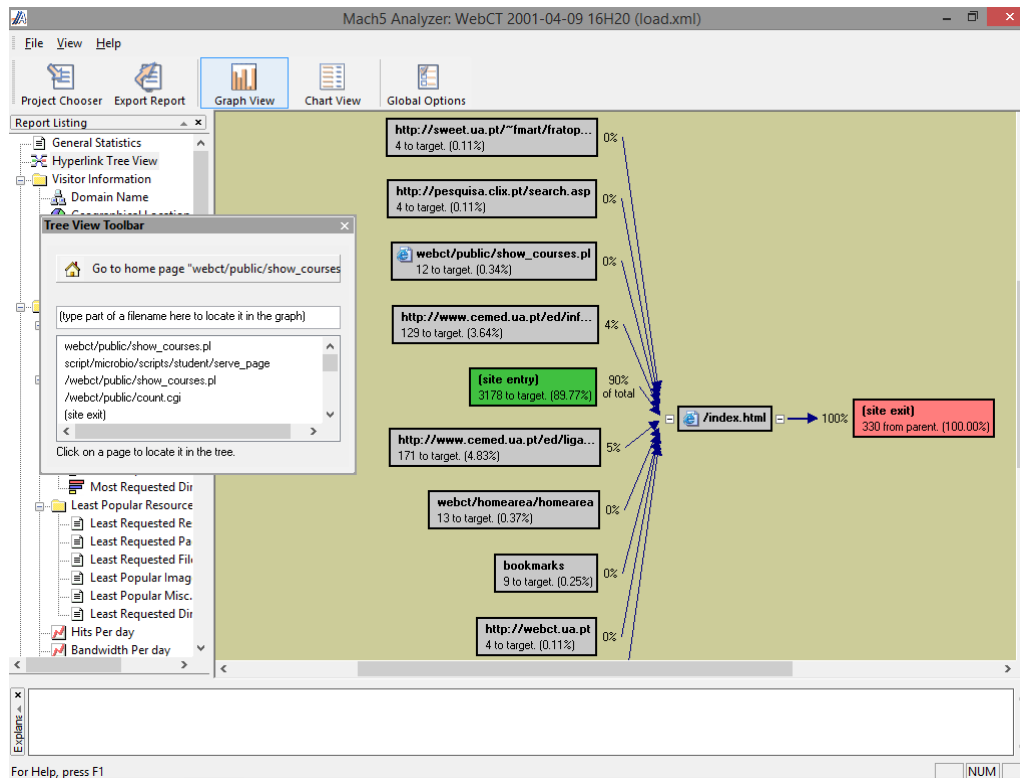


Figura II-17: Vista das ligações de e para uma determinada página de um *site web*. Informação obtida utilizando o programa FastStat Analyzer v4.1.7 da Mach5 Software, Inc.

## WebTrends Log Analyzer

A Webtrends, empresa produtora do WebTrends Log Analyzer, extinguiu este produto tendo-o substituído por uma solução baseada na *web* chamada Digital Analytics<sup>72</sup>. Esta solução exige a utilização de marcas JavaScript<sup>73</sup>, SDKs<sup>74</sup> mobile e/ou API<sup>75</sup> de recolha de dados, não dependendo dos dados registados nos ficheiros de *log* do servidor. Esta abordagem permite a monitorização / análise de *sites* alojados em qualquer plataforma (independentemente do sistema operativo, servidor *web* e de outras tecnologias utilizadas), obrigando, contudo, que seja colocado código adicional em todas as páginas que se pretenda monitorizar.

Visitando uma área reservada no *site* da empresa, podem ser obtidos vários tipos de relatórios, dependendo do tipo de contrato (módulos contratados) e tipo de *site* em análise. Por exemplo, a Figura II-18 ilustra uma vista integrada obtida pelo módulo

<sup>72</sup> Digital Analytics: <https://www.webtrends.com/products-solutions/digital-analytics/>

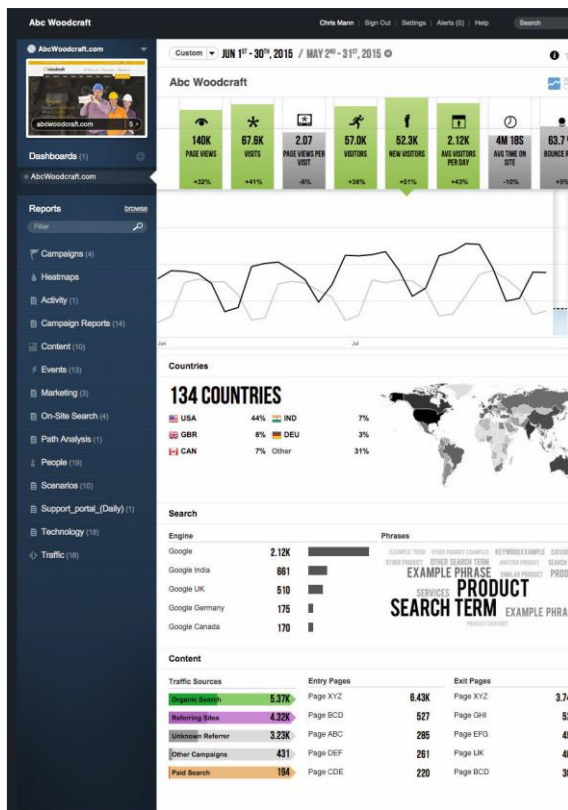
<sup>73</sup> **JavaScript**: linguagem de programação interpretada utilizada em páginas *web* e que corre diretamente no navegador do cliente.

<sup>74</sup> **SDK**: **Software Development Kit** – Kit de desenvolvimento de software.

<sup>75</sup> **API**: **Application Programming Interface** – Interface de Programação de Aplicações.



“Analytics Website Space” da análise da utilização de um *site* público de uma empresa. Uma empresa com um *site* interno (*intranet*) baseado na tecnologia *Sharepoint* pode recorrer ao módulo “Analytics for SharePoint” (Figura II-19). Muitas organizações têm presença nas redes sociais como Facebook ou Youtube e investem recursos avultados nessa presença. Naturalmente estas organizações estão interessadas em conhecer o nível de retorno desse investimento, existindo módulos específicos para a monitorização da atividade nessas redes como por exemplo o módulo “Facebook Page and App Spaces” (Figura II-20). Tendo em conta a grande proliferação das plataformas móveis, com soluções que vão dos *sites*, que podem ter sido adaptados para a dimensão do ecrã desses dispositivos, até a APPs<sup>76</sup> especialmente desenvolvidas para tirar o máximo partido desse formato. A Figura II-21 ilustra uma vista integrada obtida pelo módulo “Mobile Website Spaces” da análise de uma solução móvel.

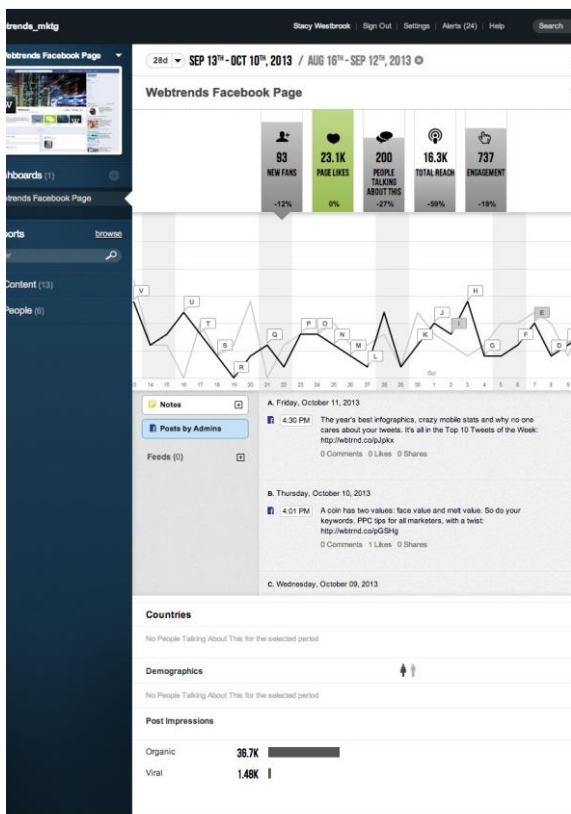


**Figura II-18:** Exemplo de uma vista integrada (Analytics Website Space) da análise da utilização de um *site web*. Imagem retirada do *site* da Webtrends.

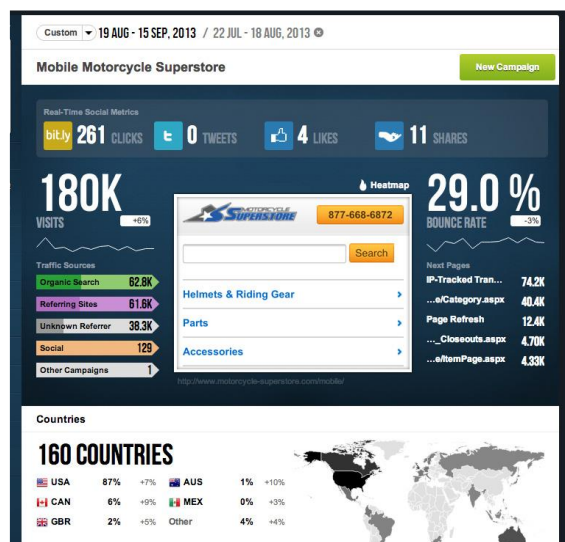


**Figura II-19:** Exemplo de uma vista integrada (Analytics for SharePoint) da análise da utilização de um *site web* interno de uma empresa. Imagem retirada do *site* da Webtrends.

<sup>76</sup> **APP:** Aplicação móvel – software desenvolvido para ser instalado em um dispositivo eletrónico móvel como telefones, PDAs ou tablets.



**Figura II-20:** Exemplo de uma vista integrada (Facebook Page and App Spaces) da análise da utilização de um *site web*. Imagem retirada do *site* da Webtrends.



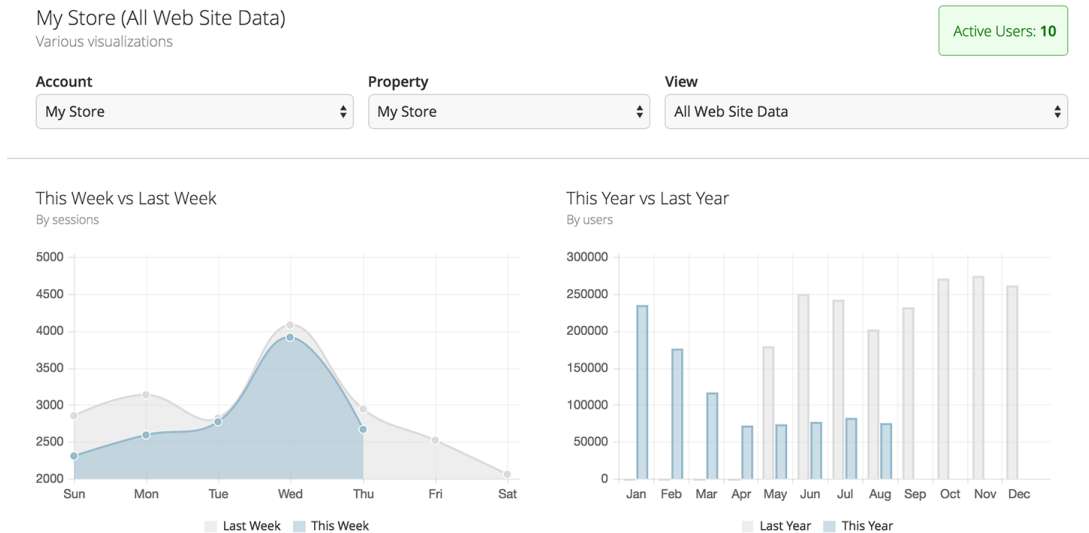
**Figura II-21:** Exemplo de uma vista integrada (Mobile Website Spaces) da análise da utilização de um *site web*. Imagem retirada do *site* da Webtrends.

## Google Analytics

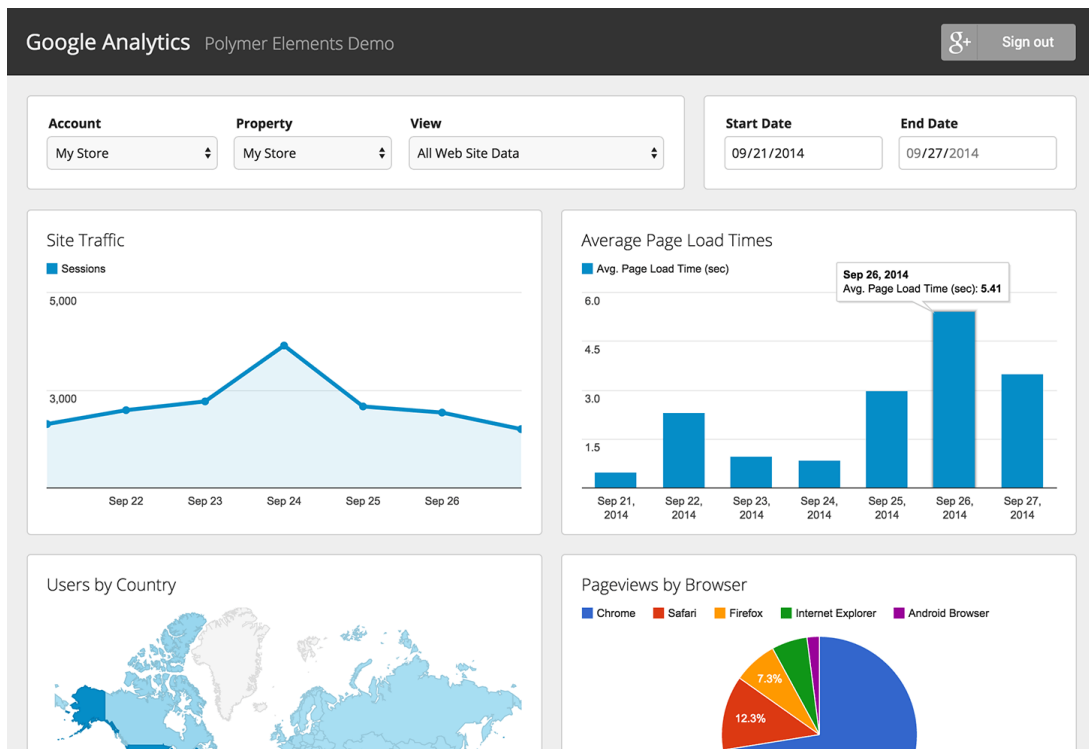
A Google foi uma empresa que teve um crescimento imenso nos últimos anos, tendo estendido a sua atividade muito para além dos motores de pesquisa com os quais se iniciou no mercado. A sua atividade também se estendeu à monitorização e análise da atividade dos utilizadores na *web*. Começou por fazê-lo primeiro nos seus próprios serviços e posteriormente passou a fornecer esse serviço aos seus clientes, disponibilizando versões gratuitas e pagas.

À semelhança da solução da WebTrends, também o Google Analytics exige a colocação de marcas JavaScript (*"The Google Analytics Embed API"*, que é uma biblioteca JavaScript), para recolha de dados, não dependendo dos dados registados nos ficheiros de *log* do servidor. Esta abordagem permite a monitorização / análise de *sites* alojados em qualquer plataforma (independentemente do sistema operativo, servidor *web* e de outras tecnologias utilizadas), obrigando, contudo, que seja colocado código adicional em todas as páginas que se pretenda monitorizar.

Os relatórios podem ser obtidos visitando uma área reservada no *site* da empresa ou com recurso a uma APP que pode ser instalada num dispositivo móvel. Os relatórios disponíveis dependem da solução contratada e do tipo de *site* monitorizado. Nas Figura II-22 e Figura II-23 podem ver-se exemplos de relatórios que podem ser obtidos.



**Figura II-22:** Exemplo de relatório da análise da utilização de um *site web*. Imagem retirada do *site* da Google (<https://ga-dev-tools.appspot.com/embed-api/>).



**Figura II-23:** Exemplo de relatório da análise da utilização de um *site web*. Imagem retirada do *site* da Google (<https://ga-dev-tools.appspot.com/polymer-elements/>).

Como se pode observar, apesar da evolução tecnológica das técnicas de monitorização utilizadas, as soluções de visualização permanecem muito semelhantes. Continuam a mostrar, basicamente, informação semelhante recorrendo a tipos de gráficos também não muito diferentes.

As propostas apresentadas nesta tese acrescentam uma mais-valia importante ao permitir fazer o cruzamento entre a informação de navegação entre as várias páginas com o conteúdo e utilização dos elementos constituintes dessas mesmas páginas.

# **Capítulo III.**

## **MODELO DE ANÁLISE E VISUALIZAÇÃO DE INFORMAÇÃO – UMA PROPOSTA**

# 1. Introdução

A análise da informação da utilização de um *site web* é essencial para a deteção de possíveis problemas, bem como para ajudar a perceber a forma como o *site* está a ser usado. De maneira a facilitar esta análise usam-se métodos de visualização de informação para representar, de forma visual, informação que de outra maneira seria de leitura muito mais difícil.

O ambiente hipermédia, que está na base de qualquer *site*, tem algumas limitações (Bieber et al., 1997). Os autores deste artigo enumeraram várias limitações e apresentaram algumas sugestões. O trabalho descrito nesta tese, nomeadamente o modelo de análise e visualização de informação proposto neste capítulo tem o potencial de complementar algumas das propostas apresentadas.

Neste capítulo vai ser abordada toda a problemática relacionada com a visualização e análise de informação.

Começar-se-á pela descrição da proposta de modelo conceptual genérico do processo de visualização de um *site web*, descrevendo cada módulo, a articulação entre módulos e os instrumentos de manipulação direta.

Na secção seguinte serão abordadas questões relacionadas com a *raw data*. Será apresentada a estrutura de dados, a forma como a informação é recolhida, caracterização da informação e condicionantes tecnológicas relacionadas. Esta secção termina com uma descrição do processo de extração de informação de ficheiros de *log* de *sites web* e com a identificação de uma sessão de um utilizador na mesma.

A parte principal deste capítulo surge na secção 4 onde são descritos os métodos de visualização e análise de informação. Inicia-se com a apresentação de algumas definições, de uma proposta de taxonomia e algumas considerações sobre utilização de cor. A secção termina com a apresentação dos métodos de visualização e análise de informação agrupados em métodos que permitem explorar o design de interação e métodos que permitem explorar a estrutura da informação.

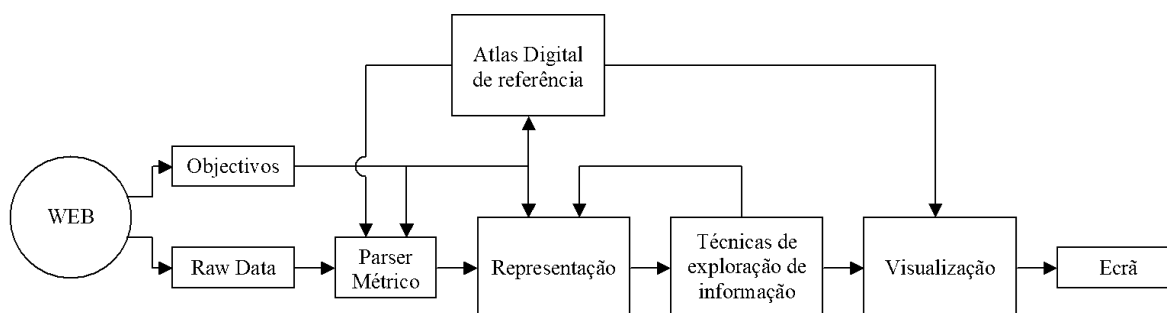
O capítulo termina com uma abordagem à integração e correlação visual de métodos de visualização e análise de informação.

## 2. Modelo conceptual

A visualização e análise de informação, no contexto deste trabalho, não é trivial e apresenta um nível de complexidade considerável. O processo de visualização implica um conjunto de passos, com ligações e dependências entre si que começa na *site web*, e informação nele contida, e termina no ecrã do utilizador, onde o resultado da visualização pode ser observado e explorado interativamente.

De maneira a sistematizar o processo de visualização de um *site web*, propõe-se um modelo conceptual que começou como representado na Figura III-1 e depois evoluiu como representado na Figura III-2. Este modelo foi construído tendo como base os vários passos necessários para, partindo da informação contida num *site web* e dos dados da sua utilização, se processe essa informação de maneira a obter uma imagem com a representação pretendida.

Como pode ser observado nas Figura III-1 e Figura III-2, este modelo apresenta blocos que representam os vários passos envolvidos no processo de visualização, bem como as ligações entre estes blocos.



**Figura III-1:** Modelo do processo de visualização de um *site web*, usando uma arquitetura de representação estática.

Nas subsecções seguintes serão descritos os vários módulos do modelo, bem como a sua articulação. Na última subsecção será introduzida uma alteração, ilustrada na Figura III-2, de maneira a permitir manipulação direta e, portanto, interatividade no processo de visualização.

### 2.1. Elementos constituintes

De forma a melhor perceber o modelo, sua representação simbólica expressa na Figura III-1 e os esquemas relacionais em causa, as secções seguintes descrevem com detalhe cada um dos módulos presentes no modelo.

## **Web**

Trata-se do *site web*, como elemento central num sistema de informação e comunicação (*intranet*) de uma organização. Este concentra / representa todos os serviços / páginas presentes no sistema em referência.

## **Objetivos**

Este é o elemento que comporta os objetivos a considerar em todo o processo de visualização e exploração da informação. Veja-se um exemplo: um utilizador entra no *site* da *intranet*, pela sua página principal e pretende executar uma dada tarefa: encontrar uma informação específica presente numa dada página.

Naturalmente, para se conseguir atingir um dado objetivo será necessário executar um conjunto de passos intermédios.

## **Raw Data**

Este bloco representa a *raw data* – antes de qualquer processamento – disponibilizada diretamente por um servidor web, por um monitor ocular, monitor da atividade do rato ou teclado, entre outros dispositivos computacionais com capacidade de gerar informação desta natureza (comportamento na interação).

A informação obtida necessita de ser tratada antes que possa ser utilizada de forma útil. Para tal será necessário proceder ao seu processamento e filtragem.

## **Parser Métrico**

Este módulo é o responsável pela filtragem e extração dos elementos relevantes da amálgama de dados proveniente do bloco *Raw Data*. Imagine-se que se pretende estudar uma sessão de um utilizador no *site* da *intranet* de uma instituição. Conhecendo a identidade do utilizador e a janela temporal será possível extrair unicamente os eventos relevantes da sessão em causa. Por exemplo: páginas visitadas, tempo decorrido entre duas páginas, etc. Desta forma será possível reduzir substancialmente a informação com que é necessário lidar em cada instante.

## **Atlas Digital de referência**

Este módulo representa uma biblioteca de referência construída previamente e que contém toda a informação relevante. Para um *site*, estará armazenado o conteúdo de todas as suas páginas e para cada página poderá ter-se o seu código HTML, uma imagem da sua representação, a identificação e classificação dos objetos que a



constituem, as zonas clicáveis, lista dos seus *links*, lista de páginas que têm um *link* para esta página, entre outra informação.

## Representação

Este módulo é responsável pelo método de representação de cada tipo de informação a ser usado para produzir a visualização.

Há vários métodos de representação disponíveis (2D, 3D, pseudo-cor, etc.) que podem ser utilizados consoante o tipo de informação em causa e o objetivo que se pretende atingir.

## Técnicas de exploração de informação

As representações obtidas pelo módulo anterior são combinadas com elementos adicionais de forma a permitir que seja possível manipular alguns parâmetros responsáveis pelo resultado da visualização.

Sem este módulo não seria possível qualquer interação com o sistema de visualização.

Aqui podem alterar-se as técnicas de representação e os parâmetros a utilizar no *Parser Métrico* de forma a obter a visualização pretendida.

## Visualização

Este módulo é o culminar de todo o processo: a criação de uma imagem que corresponde à etapa final do processo de visualização.

Naturalmente, o resultado deste módulo depende de todos os outros módulos, nomeadamente da representação dada à informação combinada com as técnicas de exploração de informação utilizando de forma exaustiva o atlas digital de referência.

## Ecrã

Não é mais que o ecrã do utilizador, onde se poderá observar o resultado da visualização e onde são indicadas as áreas / zonas de interação para alterar algum parâmetro de forma a obter a visualização pretendida.

## 2.2. Articulação entre módulos

A partir do conteúdo do *site web* são definidos os objetivos do processo de visualização, bem como parte da *raw data*.

Os objetivos definidos irão influenciar, de forma direta, a forma de representação da informação e a forma como o Atlas digital de referência será utilizado nesse processo de representação, no processo de filtragem da *raw data* e, ainda, de forma indireta, o resultado da visualização.

A *raw data* provém não só do *site web* – sob a forma de ficheiros de *log*, por exemplo – como também de outros processos de monitorização da atividade dos utilizadores.

O Atlas Digital de referência é uma biblioteca construída previamente e que terá influência direta na filtragem e extração de elementos relevantes da *raw data*, no método de representação da informação e, finalmente, na criação da imagem correspondente ao resultado do processo de visualização.

A *raw data* é então filtrada (no módulo *Parser Métrico*) para extrair os elementos relevantes, respeitando os objetivos e utilizando informação proveniente do Atlas digital de referência. Estes elementos relevantes são então processados de forma a encontrar uma representação apropriada.

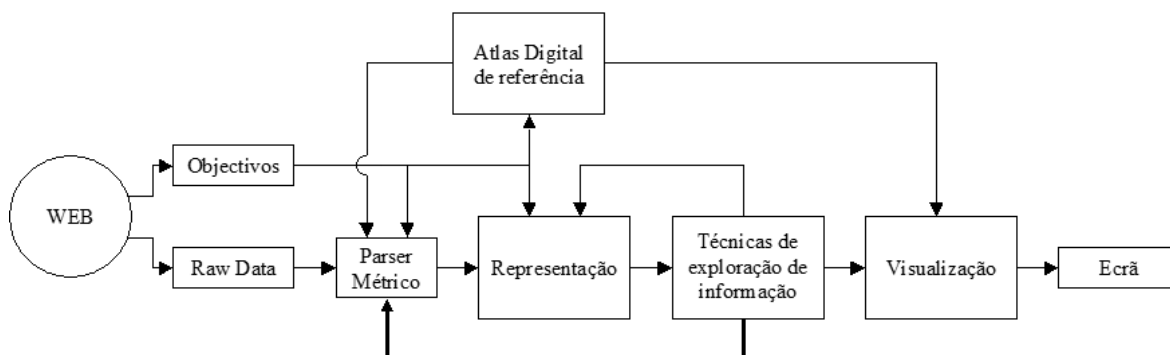
A representação da informação proveniente do módulo de filtragem (*Parser Métrico*) depende dos objetivos pretendidos, dos dados do Atlas digital de referência e dos parâmetros obtidos do processo de exploração de informação (sob a forma de realimentação).

A representação produzida para os vários blocos de informação que se pretendem visualizar num dado instante pode ser alterada de forma interativa, refinando desta forma a visualização final, que será apresentada no ecrã do utilizador.

## **2.3. Instrumentos de manipulação direta**

O processo de visualização apresentado sofre, contudo, de uma limitação importante. Depois de definidos os objetivos, só será possível alterar as técnicas de representação da informação, não havendo a possibilidade de refinar os limites da informação, propriamente dita.

Por forma a superar esta limitação, o modelo do processo de visualização de um *site* foi melhorado, com uma nova ligação entre os módulos das Técnicas de exploração de informação e o módulo *Parser Métrico*, como ilustrado na Figura III-2.



**Figura III-2:** Modelo do processo de visualização de um *site web*, usando uma arquitetura de representação dinâmica. Note-se a ligação adicional entre o módulo “Técnicas de exploração de informação” e o módulo “Parser Métrico”.

A ligação adicional entre o módulo “Técnicas de exploração de informação” e o módulo “Parser Métrico” vem permitir que durante o processo de exploração, idealmente por manipulação direta, se possa alterar não só os métodos de visualização aplicados a cada tipo de informação, como também alterar os parâmetros da filtragem e extração dos elementos relevantes, podendo na prática alterar a informação a ser visualizada de forma integrada e síncrona.

### 3. *Raw data*

A *raw data* usada, depois de processada, em vários métodos de visualização propostos neste trabalho e um dos elementos de entrada do modelo proposto, pode ser obtida de várias fontes recorrendo a múltiplos métodos de recolha.

De entre as fontes de informação usadas neste contexto, podem destacar-se os ficheiros de *log* de servidores de *sites web*, dados do registo da atividade ocular ou dados do registo da atividade do utilizador enquanto usa o seu computador (rato, teclado, monitor, etc.).

Os métodos de captura e recolha de dados a utilizar dependem da fonte de informação em causa, bem como da utilização que se pretende fazer dessa informação. A recolha de ficheiros de *log*, a utilização de dispositivos de monitorização da atividade ocular do utilizador (*eye trackers*) ou a utilização de serviços capazes de registar a atividade de interação do utilizador (*interceptor*) enquanto este utiliza o computador (*front-end, user-side*) são alguns dos métodos disponíveis.

Normalmente, os dados obtidos por estes métodos não são passíveis de ser utilizados diretamente, carecendo de tratamento e processamento prévio. Algumas fontes fornecem dados, dos quais é necessário extrair os eventos ou informação relevante,

noutros casos obtém-se dados que é necessário converter num formato passível de ser analisado.

É importante analisar os vários aspetos envolvidos em todo o processo:

- Fontes de informação e os processos de recolha da informação;
- Caracterização da informação, observando com detalhe o que é possível extrair da *raw data*;
- Análise das condicionantes / limitações associadas a cada tipo de informação.

### 3.1. Estrutura de dados

Alguma da informação necessária será recolhida diretamente, outra será obtida após processamento da *raw data*, como os ficheiros de *log* obtidos de um servidor *web*.

Podem identificar-se os seguintes tipos de informação:

- Estrutura do *site*, mantendo várias versões ao longo do tempo;
- Classificação do *site* em áreas de interesse / tipo de conteúdo;
- Conteúdo de cada página – *raw data* e/ou HTML e restantes objetos constituintes da mesma, bem com uma imagem da sua representação no ecrã;
- Lista de ligações (*links*) de cada página e respetiva posição dentro da mesma (*Hotspots*);
- Catalogação e classificação de cada área em cada página, bem como o seu *layout*;
- Registo da evolução de cada página ao longo do tempo;
- Informação de monitorização do olhar do utilizador (*eye-tracking*, por exemplo, a 50 amostras/segundo) – pontos (x,y) de fixação do olhar no ecrã, tempo de fixação (em milissegundos) num determinado ponto (x,y) do ecrã;
- Informação da monitorização da atividade do utilizador:
  - Rato (*mouse tracking*) – pontos (x,y), eventos, tempos;
  - Teclado (*key logger*) – teclas usadas, tempos;
  - Ecrã (*screen capture*) – imagens de ecrã, tempos, e/ou vídeo;
  - Contexto – páginas visitadas, percursos, tempos;
- Informação referente a uma sessão de um utilizador:
  - Percurso do utilizador – lista de páginas visitadas;
  - Tempos de cada página;
  - Eventos associados a determinada página.
- Informação referente a um conjunto de sessões de um ou vários utilizadores:
  - Lista de percursos alternativos usados pelos vários utilizadores;
  - Informação estatística dos percursos usados;
  - Tempos médios, máximos e mínimos de navegação entre páginas.

- Informação referente aos tipos de utilizadores / grupos de utilizadores e ao tipo de exploração que efetuaram:
  - Utilizadores / grupos de utilizadores que visitam cada área;
  - Classificação baseada nos tipos de conteúdo explorados;
  - Classificação relacionada com o idioma / alfabeto utilizado (símbolos gráficos do alfabeto);
  - Classificação baseada em níveis de segurança / acesso às várias áreas do *site*;
  - Classificação comportamental recolhida da análise estatística da utilização.

Esta informação foi utilizada na elaboração da estrutura de dados. Em anexo (na secção Organização/arquitetura da informação recolhida, página 246) pode encontrar-se uma representação da estrutura de dados e uma breve descrição das componentes da mesma.

## 3.2. Recolha de informação

Antes de se avançar para os processos de recolha de informação, vai começar-se por analisar as fontes de informação.

O objetivo é analisar o processo de interação, usando métodos de visualização, entre o utilizador humano e o sistema de comunicação e informação com mediação tecnológica. Posto de outra forma, poderá dizer-se que se pretende analisar a interação de uma pessoa com um dispositivo computacional enquanto utiliza os serviços infocomunicacionais de uma *intranet* de determinada organização.

Contudo, esta análise do processo de interação e respetiva recolha de informação são mais complexos do que à primeira vista se poderia supor. Por um lado, estes sistemas infocomunicacionais das organizações, são pautados de alguma complexidade fruto das diversas tecnologias, formatos e protocolos de comunicação que integram. Por outro, a experiência das pessoas que usam esses sistemas pode ser alterada por alguma perturbação introduzida pelo processo de análise / recolha de informação.

Em primeiro lugar, é necessário acautelar que o processo de recolha de informação não vá interferir no normal funcionamento do sistema em observação. Alterações no normal funcionamento do sistema ou no processo de interação entre o utilizador e o sistema poderão ter como consequência a obtenção de resultados adulterados ou sem validade, dado que as condições foram alteradas pelo processo de recolha de informação.

Estamos perante um processo sensível, que exige que as condições de funcionamento do sistema e que a interação do utilizador com o sistema permaneçam, tanto quanto possível, inalteradas. O processo deve ter o mínimo nível de intrusão possível.

Naturalmente que as limitações referidas dependem do tipo de fonte de informação.

Por exemplo, no caso de um servidor *web*, de onde se pode obter um ficheiro de *log* com informação referente a cada pedido recebido por esse servidor, é importante que a produção deste ficheiro não cause limitações em termos de desempenho ou de armazenamento no servidor em causa.

Já no caso da utilização de um *eye tracker*, os problemas podem surgir por causa das limitações de mobilidade da cabeça impostas ao utilizador por alguns destes dispositivos. Estas limitações poderão condicionar a experiência de interação e desta forma produzir dados e resultados de interação que não representam o uso natural e habitual das tecnologias por cada um dos sujeitos no seu dia-a-dia.

Algumas fontes de informação passíveis de serem utilizadas são:

- Ficheiros de *log* de *sites web*;
- Informação referente a cada página da *intranet* – o conteúdo multimédia e hipermédia do *site* propriamente dito;
- Ficheiros resultantes do *eye tracker*;
- Ficheiros resultantes do *interceptor*, nomeadamente da utilização do rato e teclado, bem como do conteúdo apresentado no monitor.

### **3.2.1. Ficheiro de *log* de *site web***

A maioria dos servidores *web* dispõe de mecanismos de monitorização que permitem o registo de vários detalhes relacionados com cada pedido recebido pelo servidor. Esta informação é normalmente depositada numa base de dados ou, ainda mais comum, num ficheiro de texto, correspondendo cada linha a um pedido. Cada entrada / linha contém vários campos, que dependem da configuração do servidor e da informação fornecida em cada pedido. É importante salientar que, do ponto de vista do servidor, uma página *web* pode corresponder a uma série de pedidos ao servidor: um para a página em si, mais um por cada elemento adicional constituinte dessa página (objetos gráficos presente na página, elementos multimédia como som e vídeo, folhas de estilo CSS, entre outros).

Finalmente, para ter acesso à informação basta descarregar os ficheiros de *log* do servidor e processá-los. Note-se, no entanto, que em *sites* com um nível de atividade muito elevado, a quantidade de informação poderá, também ela, ser muito elevada, sobretudo se o servidor estiver configurado para fazer um *log* detalhado de toda a atividade. Não é raro obterem-se ficheiros com tamanhos na ordem dos Gigabytes depois de um dia de monitorização.

### 3.2.2. Páginas da *intranet*

Por forma a complementar a informação obtida nos ficheiros de *log*, é importante saber o que realmente está a ser enviado aos utilizadores. Este aspeto é tanto mais importante quanto maior for a velocidade com que o conteúdo do *site* for atualizado / alterado. Por outras palavras, é importante reter uma imagem (ou amostra – *snapshot*) do conteúdo do *site* e isto significa guardar o conteúdo total de cada página para futura referência.

O mercado oferece algumas ferramentas que permitem fazer uma imagem integral (ou espelho – *mirror*) de um *site*. Contudo isto pode revelar-se uma tarefa complicada, sobretudo quando se está perante um *site* com conteúdo produzido dinamicamente (o código HTML da página é produzido por algum tipo de programa ou *script* na sequência do pedido). Como estas ferramentas usam técnicas de pesquisa sistemática, que têm pouco a ver com uma utilização normal de um utilizador humano, muitas vezes obtêm resultados desfasados da realidade, sobretudo em *sites* em que o padrão de utilização influencia o conteúdo das páginas a serem apresentadas.

De forma a tentar colmatar algumas falhas do método apresentado, pode reter-se o resultado de cada página apresentada numa sessão que simule uma utilização normal de um utilizador.

O conteúdo de cada página pode ser retido sobre múltiplas formas, como sejam:

- Código *raw* (conteúdo em bruto – HTML, imagens, *applets*, PDFs, etc.) que permita uma posterior observação do conteúdo apresentado;
- Resultado observado no navegador – o que o utilizador estará realmente a ver no seu monitor – usando para o efeito algum formato de imagem.

### 3.2.3. *Eye Tracker*

Uma tecnologia que registre o movimento ocular (*eye tracker*) pode ser uma ferramenta inestimável quando se pretende detetar problemas numa interface homem-máquina. Saber para onde um utilizador está a olhar num dado instante, quanto tempo detém o olhar em dado objeto ou o padrão de análise visual são informações de grande valor na deteção de erros ou inconsistências numa interface.

A maioria dos dispositivos deste tipo é bastante cara e, até há poucos anos, eram bastante intrusivos para o utilizador. Felizmente, hoje existem soluções baseadas em ecrãs de computador, mais ou menos convencionais, que impõem poucas limitações ao utilizador durante a sua utilização.

O princípio básico de funcionamento destes dispositivos consiste na utilização de uma ou mais câmaras de vídeo especiais que observam a posição da íris dos olhos do utilizador. Estas câmaras estão ligadas a um sistema de processamento, muitas vezes integrado no

próprio dispositivo de *eye-tracking*, e permitem identificar a direção do olhar em cada instante.

Uma das primeiras soluções que surgiram no mercado consistia na utilização de uma pequena câmara junto a cada olho sustentada por suportes fixos a uma estrutura do tipo capacete que o utilizador colocava na cabeça. Esta solução permitia um elevado grau de mobilidade da cabeça, uma vez que além de detetar a posição do globo ocular também determinava a posição da cabeça, sendo possível determinar para onde o utilizador estava a olhar graças à combinação das duas informações.

Existiam também soluções mais baratas, mas mais limitadas, que obrigavam o utilizador a imobilizar a cabeça (encostando o queixo e testa num apoio fixo) em frente ao objeto a observar. Na zona do objeto a observar (o monitor do computador, no caso em estudo) encontrava-se a câmara que determinaria a posição do globo ocular. Qualquer movimento na cabeça do utilizador poderia provocar erros significativos na informação recolhida.

Mais recentemente, passaram a estar disponíveis óculos bastante leves com capacidade de *eye-tracking* que nascem desta linha de investigação e desenvolvimento e já apresentam características de registo muito detalhadas, embora com desvantagens associadas, principalmente com questões de calibração no contexto de universo a ser observado e com o preço.

A solução mais comum, menos intrusiva e com processos de calibração mais simples (na relação do registo do movimento do globo ocular com o objeto a ser observado), é apresentada num dispositivo de *eye-tracking* desktop em que é colocado um sensor sobre a mesa ou junto do monitor com computador. Estão também disponíveis soluções que apostam num formato integrado no ecrã LCD.

Qualquer destes sistemas, depois de devidamente calibrado, fornece informação com uma precisão apreciável.

A *raw data* resultante de qualquer destes dispositivos, consiste num ficheiro com as coordenadas (x,y) da posição do olhar em cada instante (25 ou 50 amostras por segundo, por exemplo – há soluções que permitem 300 amostras por segundo). Alguns fabricantes, como por exemplo a Tobii<sup>77</sup>, fornecem ainda ferramentas que permitem converter esta informação em relatórios de base imagética, para uma mais rápida observação e análise dos resultados.

---

<sup>77</sup> Tobii: <http://www.tobii.com/>



### 3.2.4. *Intercetor*

A atividade do utilizador enquanto este usa o computador e respetivos periféricos pode fornecer dados de importância inestimável quando se pretende estudar a interação do utilizador com um sistema de informação e comunicação.

O utilizador interage com o rato e teclado e observa o resultado das suas ações no monitor... Isto corresponde também à informação que poderá ser interessante registar: atividade do teclado (teclas pressionadas), atividade do rato (movimento / posição do cursor e utilização das teclas), bem como o conteúdo do ecrã em cada instante.

No entanto, o processo de recolha desta informação não é trivial. Em primeiro lugar, há a destacar as questões éticas envolvidas – não é eticamente aceitável monitorizar a atividade de um utilizador sem o seu prévio conhecimento e consentimento. Em segundo lugar, aparecem as questões técnicas – ainda que monitorizar e registar a atividade do teclado ou rato não levante problemas de maior, o mesmo não se pode dizer relativamente ao conteúdo do ecrã, devido à grande quantidade de informação que pode estar em causa.

A este tipo de aplicações, que permitem capturar e registar a atividade do utilizador, pode chamar-se de *intercetores*. Normalmente, são aplicações que correm em segundo plano no computador que se pretende monitorizar e registam toda a atividade no próprio computador ou enviam para outro local onde será registada.

Existem no mercado várias aplicações deste tipo, mais ou menos sofisticadas, muitas delas oriundas das comunidades de *hackers* e usadas de forma ética e legalmente questionáveis. Podemos encontrar aplicações especificamente concebidas para análises de usabilidade. Contudo, as mais frequentes aparecem sob a forma de vírus e cavalos de troia que registam a atividade do teclado, com o objetivo de recolher números de cartões de crédito, senhas de acesso, etc. e que os fazem chegar, de alguma forma, ao seu autor.

O resultado deste tipo de aplicações é habitualmente um ou vários ficheiros contendo a informação relativa ao que se pretende monitorizar, por exemplo um ficheiro com a atividade do rato, outro com a atividade do teclado e um vídeo com registo do conteúdo do ecrã. Uma das aplicações mais conhecidas (usadas a nível profissional em estudos de usabilidade) é oferecida pelo grupo Nielsen Norman Group<sup>78</sup>.

---

<sup>78</sup> Nielsen Norman Group: <https://www.nngroup.com/>

### 3.3. Caracterização da informação recolhida – Taxonomia de informação

Para se poder concretizar o objetivo deste trabalho – analisar o processo de interação, usando métodos de visualização, entre o utilizador humano e o sistema de comunicação e informação com mediação tecnológica – é necessário recolher vários tipos de dados, o que só é possível obter a partir de múltiplas fontes. Cada fonte caracteriza-se por fornecer um conjunto de tipos de dados que pode ser complementado com dados provenientes de outras fontes.

Depois da descrição do processo de recolha de informação de cada fonte, vai passar-se a uma breve caracterização da informação passível de ser obtida de cada uma dessas fontes com o objetivo de conseguir sistematizar e enunciar um quadro classificativo da mesma.

#### 3.3.1. Ficheiro de *log* de *site web*

Os campos presentes num ficheiro de *log* proveniente de um servidor *web* permitem a obtenção de alguma informação importante, como se pode ver na secção 2 dos Anexos.

Na lista seguinte pode observar-se alguma da informação obtida de uma entrada de um ficheiro de *log*:

Date	2003-02-18
Time	10:12:38
Client IP Address	192.168.187.211
User Name	ihc-ect
Server IP	193.137.85.3
Server Port	80
Method	GET
URL Requested	/ihc/Contactos.htm
URL Query	param=top
Protocol Status	200
Protocol Version	HTTP/1.1
Server Host	www2.ca.ua.pt
User Agent	Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+5.0)
Cookie	ASPSESSIONIDSAQDDADD= BEEJIDABBOOMKMBGHPIIINDC
Referrer	http://www2.ca.ua.pt/ihc/Index.htm

Note-se, contudo, que alguns destes campos poderão não estar sempre presentes, como se pode ver na secção 3.4.1, abaixo neste capítulo.

### 3.3.2. Páginas da *intranet*

Ainda que os ficheiros de *log* já forneçam uma quantidade apreciável de informação, continua a faltar informação importante: o conteúdo de cada página. A melhor forma de obter esse conteúdo é mesmo recorrer ao próprio *site* e recolher essa informação, descarregando cada uma das páginas necessárias do *site* em causa.

A maioria das páginas é construída usando HTML, o formato mais utilizado na *internet*. O HTML permite representar texto e respetiva formatação, mas não suporta diretamente informação do tipo gráfico. Permite, no entanto, incluir uma panóplia alargada de objetos externos, como sejam imagens, sons, vídeos e *applets*, que a complementam aumentando dramaticamente as suas potencialidades dinâmicas e encapsulando-as também. Alguns destes conteúdos classificam-se hoje como aplicações *web* (*web app* ou simplesmente *App*) pela sua natureza modular e encapsulada, resolvendo por si mesmas muitas situações de interação sem recurso ao restante ambiente HTML.

Mas nem todas as páginas são construídas com base em HTML. É possível ter-se páginas noutros formatos, como em PDF ou mesmo só texto.

A *internet* é um meio em permanente evolução. Aliás, essa é umas das suas grandes potencialidades, sendo possível manter a informação disponibilizada permanentemente atualizada. Este facto tem como consequência a necessidade de obter o estado do *site* (imagem ou *snapshot*) do momento em que se efetuar o estudo. Assim, há a necessidade de arquivar toda a informação constituinte da página, o que no caso de uma página em HTML significa: a página em si e todos os objetos adicionais (imagens, *applets*, etc.) constituintes.

De forma a facilitar o processo, além de se guardar a informação enviada para o computador do utilizador, arquiva-se também uma imagem com o aspeto do conteúdo observado no navegador, ficando registado o que efetivamente é apresentado ao utilizador.

### 3.3.3. Movimento Ocular (*Eye Tracker*)

Os dispositivos de monitorização ocular fornecem, tipicamente, um ficheiro tipo *log* que pode apresentar vários formatos, sendo o formato texto um dos mais comuns.

Este ficheiro contém o registo da posição (x,y) do olhar, correspondente à direção do olhar do utilizador, o tempo que despendeu num determinado ponto e o tempo de deslocação entre pontos do olhar. Alguns dispositivos de *eye tracking* conseguem recolher e registar estes dados com uma frequência que pode superar as 300 amostras por segundo, ainda que a maioria se fique pelas 25 a 50 amostras por segundo, ainda

assim, uma resolução temporal suficientemente rigorosa para registo de comportamentos humanos desta natureza.

Cada amostra registada na *log file* do *eye tracker* pode corresponder a uma linha com o seguinte formato, representativo da posição no ecrã (calibrada) para onde o utilizador olhou no momento temporal registado:

```
2004-02-20 09:15:09:537 1203 605
```

Correspondente a:

2004-02-20	Data (ano-mês-dia)
09:15:09:537	Hora (hora:minuto:segundo:milésimo)
1203	X (coordenada relativa ao ecrã – horizontal)
605	Y (coordenada relativa ao ecrã – vertical)

Muitos dispositivos de monitorização ocular vêm acompanhados com um sistema que permite fazer algum processamento a esta informação, extraíndo e identificando de forma automática um conjunto de eventos associados ao comportamento de observação de um ecrã de computador.

Neste caso, obtém-se um ficheiro adicional, em que cada entrada pode ter o seguinte formato:

```
2004-02-20 09:15:09:537 1203 1003 HOLD 2987
```

Correspondente a:

2004-02-20	Data (ano-mês-dia)
09:15:09:537	Hora (hora:minuto:segundo:milésimo)
1203	X (coordenada relativa ao ecrã – horizontal)
1003	Y (coordenada relativa ao ecrã – vertical)
HOLD	Evento (neste caso – olhar fixo)
2987	Tempo (duração do evento – em milésimos de segundo)

### 3.3.4. Interação Humano-Computador (*Intercetor*)

O processo de monitorização da atividade do utilizador (*intercetor*), enquanto este utiliza o computador, pode fornecer vários tipos de informação. A informação mais comum diz respeito à atividade do rato, do teclado e o conteúdo do ecrã em cada instante.

Cada tipo de atividade aparece, normalmente, contido em ficheiros de *log* separados.

## Rato

Um ficheiro de *log* proveniente da monitorização da atividade do rato pode ter o seguinte formato:

```
2004-02-20 09:15:09:537 1203 1003 LeftDown
```

Correspondente a:

2004-02-20	Data (ano-mês-dia)
09:15:09:537	Hora (hora:minuto:segundo:milésimo)
1203	X (coordenada relativa ao ecrã – horizontal)
1003	Y (coordenada relativa ao ecrã – vertical)
Left	Evento (neste caso – pressão da tecla esquerda do rato)

## Teclado

Um ficheiro de *log* proveniente da monitorização da atividade do teclado pode ter o seguinte formato:

```
2004-02-20 09:15:09:537 5A CTRL
```

Correspondente a:

2004-02-20	Data (ano-mês-dia)
09:15:09:537	Hora (hora:minuto:segundo:milésimo)
5A	Carácter (código ASCII em formato Hexadecimal)
CTRL	Modificadores (Alt, Ctrl, Shift ou combinação destes)

## Ecrã

A monitorização do conteúdo do ecrã do utilizador pode ser efetuada de duas formas distintas: a captura de um vídeo, em que cada *frame* está indexada com uma marca temporal, ou a captura de uma imagem do ecrã sempre que for detetada uma alteração do mesmo.

O resultado final, em ambos os casos, não é muito diferente. Num caso, obtém-se um vídeo, no outro, pode construir-se um, se tal for pretendido.

## 3.4. Condicionantes tecnológicas e representação de informação

Apesar de toda a informação disponível, há situações em que não é suficiente. Umhas vezes, porque a informação pretendida não está disponível de todo – porque não foi registada nos ficheiros de *log* –, outras vezes é possível obtê-la de forma indireta a partir

da informação disponível, ainda que com alguns problemas devido à dificuldade em extraí-la ou pela precisão / rigor que, por vezes, não é possível assegurar.

### **3.4.1. Fontes de informação**

Algumas fontes de informação apresentam algumas limitações que são descritas a seguir.

#### **Ficheiro de *log de site web***

Ao nível dos ficheiros de *log*, o problema começa com o facto de existirem múltiplos formatos dependentes do fabricante do servidor *web* em causa. Para piorar a situação, o gestor de um dado servidor pode configurá-lo a seu gosto, podendo não só especificar que informação pretende ver registada como até pode alterar o formato como essa informação é registada. As consequências disto levam a que seja extremamente difícil criar um método universal para a extração da informação contida neste tipo de ficheiro *log*. A informação que se obtém também não pode ser garantida, pois podem faltar elementos importantes para a análise em curso, que dificilmente podem ser encontrados de outra forma.

#### **Páginas de *intranet***

Também aqui podem haver problemas. Um *site*, regra geral, não é estático ao longo do tempo. É normal as várias páginas serem atualizadas de forma a corrigir erros encontrados ou a completar a informação disponibilizada. É igualmente fácil adicionar ou remover páginas.

Para além desta situação, deve ter-se em consideração que há *sites* que são, por definição, dinâmicos. O conteúdo de cada página é produzido em tempo real, podendo depender de fatores como o momento do pedido, a identidade do utilizador, o percurso utilizado antes de chegar à página pretendida, enfim, o contexto de interação do momento.

Mas o relato desta dificuldade não termina aqui. Isto tem só a ver com a obtenção de uma dada página. Adicione-se o facto de o conteúdo de uma dada página poder ter elementos dinâmicos – que se alteram com a interação do rato, por exemplo – e ficar-se-á com uma ideia da complexidade técnica do registo que está em causa.

#### ***Eye Tracker***

Os dispositivos de monitorização ocular também não estão isentos de problemas.

Os maiores problemas estão relacionados com as limitações impostas ao utilizador. Muitos sistemas implicam algum nível de intrusão na experiência de interação do utilizador. A calibração do sistema também é um ponto crítico.

Ainda que o volume de dados produzido pelos sensores possa ser significativo, a maioria dos fabricantes já conseguem processar em tempo real esses dados brutos produzidos pelos sensores e devolver só os dados / eventos relevantes.

### ***Intercetor***

Ainda que respeitando escrupulosamente as questões éticas relacionadas com a utilização deste tipo de sistemas, eles apresentam algumas questões relacionadas com a quantidade de informação produzida e com possíveis limitações de desempenho da máquina monitorizada.

O rato e o teclado não apresentam problemas de maior, mas quanto ao ecrã, o método do vídeo tem a vantagem da facilidade de implementação (câmara de vídeo ou software de captura de ecrã) e existem muitas aplicações que permitem o visionamento (basicamente, qualquer reprodutor de vídeo). Tem o inconveniente de produzir um fluxo de informação relativamente elevado, o que pode causar constrangimentos ao nível do desempenho da máquina monitorizada e de armazenamento, em última análise.

O método da captura de imagens, quando são detetadas alterações, tem a vantagem de se poder programar os eventos com interesse e dessa maneira reduzir dramaticamente o fluxo de informação, com os benefícios evidentes em termos de desempenho e armazenamento. Como inconvenientes pode apontar-se a possibilidade de haver eventos que não são registados, nomeadamente, o processo de carregamento de uma página.

Um dos maiores constrangimentos deste método, caso afete o desempenho do dispositivo cliente sob monitorização, prende-se com a quebra do desempenho nos procedimentos naturais de interação por parte do utilizador. É legítimo concluir que se o tempo de feedback de cada uma das ações inerentes às narrativas de interação de determinado utilizador decaírem, a fiabilidade dos dados registados será posta em causa e consequentemente todo o processo de estudo de interação.

## **3.4.2. Limitações no preenchimento da estrutura de dados**

Já foram analisadas limitações de algumas fontes de dados bem como dos respetivos processos de recolha. Contudo, estas não são as únicas limitações que é necessário enfrentar quando se pretende analisar e visualizar informação relacionada com um *site*

*web*. Se não for possível recolher os dados, a ausência de informação complica ainda mais a situação.

A estrutura de dados apresentada na secção 3.1, acima neste capítulo, representa a informação base necessária para o processo de análise e visualização de um *site web*. Porém nem sempre é possível preencher todos os campos desta estrutura de dados. Estas faltas podem ser devidas a vários motivos:

- Algumas fontes podem não estar disponíveis em alguns estudos, por algum motivo, como aconteceu durante este trabalho em que não foi possível ter acesso a um monitor ocular devido ao seu custo;
- Algumas fontes podem não estar a fornecer toda a informação possível, como acontece muitas vezes com os ficheiros de *log* obtidos dos servidores *web*, na medida em que depende da sua configuração;
- Em alguns estudos não é possível recolher toda a informação, devido a limitações como tempo disponível, recursos, permissões, etc..

### 3.5. Extração de informação de um ficheiro *log* de um servidor *web*

A extração de informação relevante existente num ficheiro *log* de um servidor *web* depende dos diversos formatos de *log* existentes, dependendo ainda do fabricante do servidor e das especificações escolhidas pelo gestor.

Analise-se um exemplo de uma linha de um ficheiro de *log* com o formato W3C extraído de um servidor IIS a correr em Windows 2000 Server:

```
2003-02-18 10:12:38 192.168.187.211 ihc-ect W3SVC4 CPJ-SRV 193.137.85.3
80 GET /ihc/Contactos.htm param=top 200 0 8225 511 16 HTTP/1.1
www2.ca.ua.pt
Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+5.0;+.NET+CLR+1.0.3705)
ASPSESSIONIDSAQDDADD=BEEJIDABBOOMKMBGHPIIINDC
http://www2.ca.ua.pt/ihc/Index.htm
```

O processamento desta linha permite identificar a seguinte informação (como mencionado na secção 3.3.1, acima):

```
Date Time Client_IP_Address User_Name Site_Name Server_Name Server_IP
Server_Port Method URL_Request URL_Query Protocol_Status Win32_Status
Bytes_Sent Bytes_Received Time_Taken Protocol_Version Server_Host
User_Agent Cookie Referrer
```

Este exemplo é relativamente simples de processar, dado que os vários campos estão separados por um carácter “espaço”. Além disso, é garantido que o carácter “espaço” não fará parte de nenhum campo, sendo substituído pelo carácter “+” quando aparece no



conteúdo de um dos campos. Assim, será suficiente usar o carácter “espaço” como separador (*token*) para ler um campo, passar ao seguinte, ler novo campo e assim sucessivamente.

Contudo, em muitas circunstâncias, há campos de informação que não estão disponíveis. Campos como `User_Name`, `URL_Query`, `User_Agente`, `Cookie` ou `Referrer` estão ausentes com alguma frequência. Isto pode acontecer por vários motivos:

- Por não existir ou não ser necessária, pelo que não é fornecida – exemplo dos campos `User_Name` e `URL_Query`;
- Porque é bloqueada / suprimida devido à configuração dos programas de navegação (*browser*) do utilizador ou devido à presença de dispositivos de segurança, como uma *firewall* – exemplo dos campos `User_Agent`, `Cookie` ou `Referrer`;

A ausência de campos pode ser assinalada colocando um carácter específico, como “-“, no seu lugar ou o campo pode ser simplesmente suprimido. No primeiro caso é trivial detetar a sua ausência – ao encontrar um campo cujo conteúdo se resume ao carácter específico (por exemplo “-“) sabe-se automaticamente que não existe informação disponível para esse campo. Contudo, não se pode dizer o mesmo do segundo caso – é necessário proceder a uma análise de contexto, para poder determinar qual / quais o(s) campo(s) em falta e proceder em conformidade.

No entanto, o formato W3C é um exemplo de um formato relativamente “bem comportado” pois os campos estão todos delimitados com o carácter “espaço” que não é usado... Infelizmente, muitos servidores presentes no mercado produzem ficheiros com um formato proprietário que não respeita esta norma e o processo de extração da informação disponível não é tão simples. Analisemos um exemplo obtido de um servidor *Apache* a correr numa máquina *Linux*, para o mesmo pedido analisado acima:

```
192.168.187.211 ihc-ect@cpj1.ca.nat.ua.pt ihc-ect [18/Feb/2003:10:12:38
+0000] "GET /ihc/Contactos.htm?param=top HTTP/1.1" 200 8225
"http://www2.ca.ua.pt/ihc/Index.htm"
"Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+5.0;+.NET+CLR+1.0.3705)"
```

O processamento desta linha permite identificar a seguinte informação:

```
Client_IP_Address Ident_Info User_Name [Date:Time TimeZone] "Method
URL_Request?URL_Query Protocol_Version" Protocol_Status Bytes_Sent
"Referrer" "User_Agent" "Cookies"
```

Neste caso, a extração da informação de cada campo é muito mais complicada. Alguns campos nem sempre estão presentes – como por exemplo `Ident_Info`, `User_Name`, `URL_Query`, `Referrer`, `User_Agent` ou `Cookies` – e nem sempre é fácil determinar a sua ausência. Isto acontece porque quando alguns campos não estão presentes, estes

são substituídos pelo carácter “-“, enquanto outros simplesmente são suprimidos. Acresce o facto de alguns campos poderem conter espaços ou outros caracteres que são também utilizados como separadores entre campos.

Um exemplo: os campos `User_Name`, `URL_Request`, `URL_Query`, `Referrer`, `User_Agent` e `Cookies` podem conter espaços e outros caracteres usados como delimitadores dos campos, dificultando de sobremaneira a identificação dos limites dos campos e a extração da informação contida nos mesmos.

### 3.6. Identificação de uma sessão de um utilizador

Quando se está a analisar a utilização de um *site* há informação que é importante obter. Sem dúvida que os dados estatísticos mencionando as páginas mais e menos visitadas, distribuição geográfica dos visitantes, etc. fornecem informação inestimável. Mas é manifestamente insuficiente para caracterizar uma sessão de interação infocomunicacional com o sistema sob análise.

É importante saber como cada visitante interage com um determinado sistema infocomunicacional, ou seja, como utiliza o *site web* respetivo. Detalhes como o seu percurso dentro do *site*, que páginas visita, quanto tempo se detém em cada uma, que *links* utiliza, etc. podem dar pistas inestimáveis para a identificação e resolução de problemas que de outra maneira seria virtualmente impossível.

Contudo, para se poder estudar o comportamento de utilizador enquanto visita um *site* é necessário, antes de mais, identificar os limites de cada sessão, de cada visita, de cada utilizador – em termos práticos, identificar o instante em que um utilizador inicia uma visita, as páginas percorridas e quando abandona o *site web*.

Ora, isto não é um processo trivial, nomeadamente, porque normalmente os servidores não fornecem essa informação nos seus ficheiros de *log*, não sendo possível obter essa informação de forma direta. Assim, a única forma é tentar deduzir o "fim de uma sessão de utilização" a partir da informação proveniente desses ficheiros de *log*.

A identificação dos limites de uma sessão de um utilizador envolve imensos fatores que é necessário ter em consideração. Eis alguns dados que são normalmente utilizados para o efeito:

- Endereço IP da máquina do utilizador;
- *User ID*, quando disponível;
- Informação de *referrer*, quando esta está disponível;
- Informação contida nos *cookies*, quando podem ser utilizados e estão disponíveis;
- Limites temporais da visita;

Uma sessão é constituída por um conjunto de páginas que foram visitadas por um utilizador dentro de uma janela temporal, tendo normalmente um dado ponto de entrada (página inicial). Assim, pode-se impor um conjunto de restrições que permitam identificar os limites de uma sessão:

- Um utilizador durante uma sessão usa sempre o mesmo computador, pelo que é seguro afirmar que o endereço IP da sua máquina não se alterará durante a sessão.
- Se existir um processo de validação de utilizador, recorrendo à exigência de um *login* e *password*, o processo fica ainda mais fácil, na medida em que passa estar disponível a identidade do utilizador (*User ID*). Pode assumir-se que um utilizador não irá mudar de identidade durante uma sessão normal de visita a um *site*.
- A informação de *referrer* é de grande importância pois permite identificar de onde o utilizador veio para chegar a uma determinada página – a página que contém o *link* que permitiu chegar à página presente. Desta forma, pode identificar-se por exemplo se o utilizador seguiu um *link* presente numa página ou se saltou diretamente para a página (digitando o *URL* ou usando um favorito no navegador).
- Quando disponíveis, os *cookies* também podem dar uma ajuda, nomeadamente os *cookies* de sessão. Trata-se de pequenos ficheiros de texto, contendo informação fornecida pelo servidor e armazenados na máquina cliente, que são transmitidos com cada pedido ao servidor. Podem conter, por exemplo, um ID de sessão que permite identificar de forma inequívoca a sessão corrente. Quando o utilizador fechar o seu navegador ou passado um tempo pré-determinado, esta informação expira, sendo fornecido um ID novo numa nova visita posterior.
- Pode também assumir-se que uma visita não tem uma duração ilimitada. Assim, pode definir-se que:
  - Uma sessão tem duração máxima (por exemplo uma hora);
  - Se o utilizador não registar qualquer atividade durante um certo período de tempo (por exemplo 15 minutos), considerar que abandonou o *site*.

## 4. Métodos de visualização e análise de informação

Nesta secção serão apresentados vários métodos de visualização e análise de dados. Os vários métodos estão agrupados em dois grandes grupos, correspondentes a duas subsecções:

- Na subsecção 4.4. Visualização da organização do campo visual de interação serão apresentados os métodos que permitem visualizar / analisar a organização do campo visual de interação de uma dada página;
- Na secção 4.5. Visualização da estrutura do *site* / sessão serão apresentados os métodos que permitem visualizar / analisar a estrutura / organização do *site* ou sessão de utilização.

Cada método apresentado será classificado de acordo com a taxonomia apresentada na subsecção 4.2.

Antes de começar, importa introduzir algumas definições. Assim, começa-se por apresentar algumas definições e formalismo matemático na secção 4.1.

Na secção 4.2 é apresentada uma taxonomia utilizada para classificar os métodos de visualização.

Seguidamente, faz-se uma breve abordagem à utilização da cor em visualização na secção 4.3.

Para terminar esta secção, são apresentados os métodos de visualização, aparecendo na secção 4.4 os métodos relacionados com a análise da organização do campo visual e na secção 4.5 os métodos relacionados com a análise da estrutura do *site*.

## 4.1. Algumas definições

A proposta apresentada tem como base a abordagem, nomenclatura e definições expostas por Healey em (Healey et al., 2001) e as adaptações introduzidas em (Mealha et al., 2004).

Um *site*  $S$  de uma *intranet* tem um conjunto de páginas  $P$  :

$$S = [P_1 \ \cdots \ P_t], \quad \text{onde } t = \text{número de páginas do site};$$

Cada página  $P$  tem um conjunto de áreas  $H$  :

$$P = [H_1 \ \cdots \ H_n], \quad \text{onde } n = \text{número de áreas da página};$$

Cada área  $H$  da página tem um conjunto de atributos  $A$  :

$$H = [A_1 \ \cdots \ A_m], \quad \text{onde } m = \text{número de atributos};$$

A visualização começa com a construção de um mapeamento entre dados e atributos,  $M(V, \phi)$ , que converte a *raw data* (proveniente das fontes, em estado bruto) em imagens que são apresentadas ao utilizador.

$V = [V_1 \ \dots \ V_r]$  identifica um conjunto de atributos visuais  $V_i$  usados para visualizar um atributo da informação  $A_i \xrightarrow{\phi_i} V_i$  que faz o mapeamento do domínio  $A_i$  para a gama de valores visualizáveis em  $V_i$  e em que  $\phi = [\phi_1 \ \dots \ \phi_s]$  identifica um conjunto de instruções de representação  $\phi_i$ .

Com base nestas definições, uma visualização é a seleção de  $M_s$ , onde  $s$  = número de métodos de visualização. Uma visualização efetiva escolhe  $M_s$  para suportar a tarefa de exploração e análise que o utilizador pretenda efetuar.

Aplicando as definições apresentadas a um cenário concreto: um *site* de uma *intranet* tem um conjunto de páginas  $S = [P_1 \ \dots \ P_t]$  onde  $t$  = número de páginas do site. Cada página  $P = [H_1 \ \dots \ H_n]$  contém  $n$  áreas. Cada área  $H = [A_1 \ \dots \ A_m]$ , de qualquer página, tem  $m$  atributos, dos quais se podem identificar, pelo menos, os seguintes:

- $A_1$  = URL da página à qual pertence;
- $A_2$  = Imagem da página à qual pertence (que foi apresentada ao utilizador);
- $A_3$  = Posição e tamanho da área na página;
- $A_4$  = Indicação se a área tem características interativas (*Hotspot*);
- $A_5$  = URL da página que é referenciada (*link*);
- $A_6$  = Número de vezes que este *Hotspot* foi utilizado;
- $A_7$  = Imagem da página que é referenciada, da forma que foi apresentada ao utilizador;
- $A_8$  = Tempo decorrido para ir de uma página para outra, usando este *Hotspot*.
- $A_9$  = Utilizadores que usaram este *Hotspot*.

Estes atributos são apresentados visualmente sempre que necessário. Não são necessariamente usados em simultâneo num esquema de visualização, mas são úteis para representações intimamente relacionadas e interpretação entre diferentes esquemas de visualização.

A matriz  $[M_1(V_1, \phi_1) \ M_2(V_2, \phi_2) \ \dots \ M_s(V_s, \phi_s)]$  representa um vetor de funções que transformam a *raw data* em atributos visuais, uma para cada um dos métodos de visualização propostos.

## 4.2. Taxonomia

“Existem muitas linhas de orientação para o design visual de métodos de visualização de informação mas o princípio básico pode ser resumido como o Mantra da Pesquisa de Informação Visual: Primeiro, uma vista geral (holística), ampliar e filtrar, e depois detalhes a pedido” (Shneiderman, 1996).

De facto, é possível identificar um conjunto de técnicas que é necessário aplicar em qualquer método de visualização de informação. Estas técnicas, juntamente com as características da informação, pela sua natureza distintiva, são usadas habitualmente para classificar os métodos de visualização. Os tipos de classificação utilizados nesta tese são uma combinação das propostas de (Andrews, 2002; Keim, 2000, 2001; Shneiderman, 1996) e tem algumas semelhanças com as propostas de (Munzner, 2014; Spence, 2014):

1. Tipo de dados;
2. Métodos de visualização;
3. Técnicas de interação e distorção.

O trabalho inovador e distintivo destes autores representa claramente o quadro de necessidades previsto para o processamento e visualização da informação no contexto desta tese. Para clarificar o potencial taxonómico destes três tipos de classificação propõe-se uma representação gráfica combinada, como ilustrado na Figura III-3, fazendo corresponder cada tipo de classificação a um dos eixos cartesianos. Desta maneira, um método de visualização poderá ser identificado por um ponto distinto que corresponde à interseção de três vetores (correspondentes aos três critérios de classificação usados pela taxonomia).

### Tipos de dados

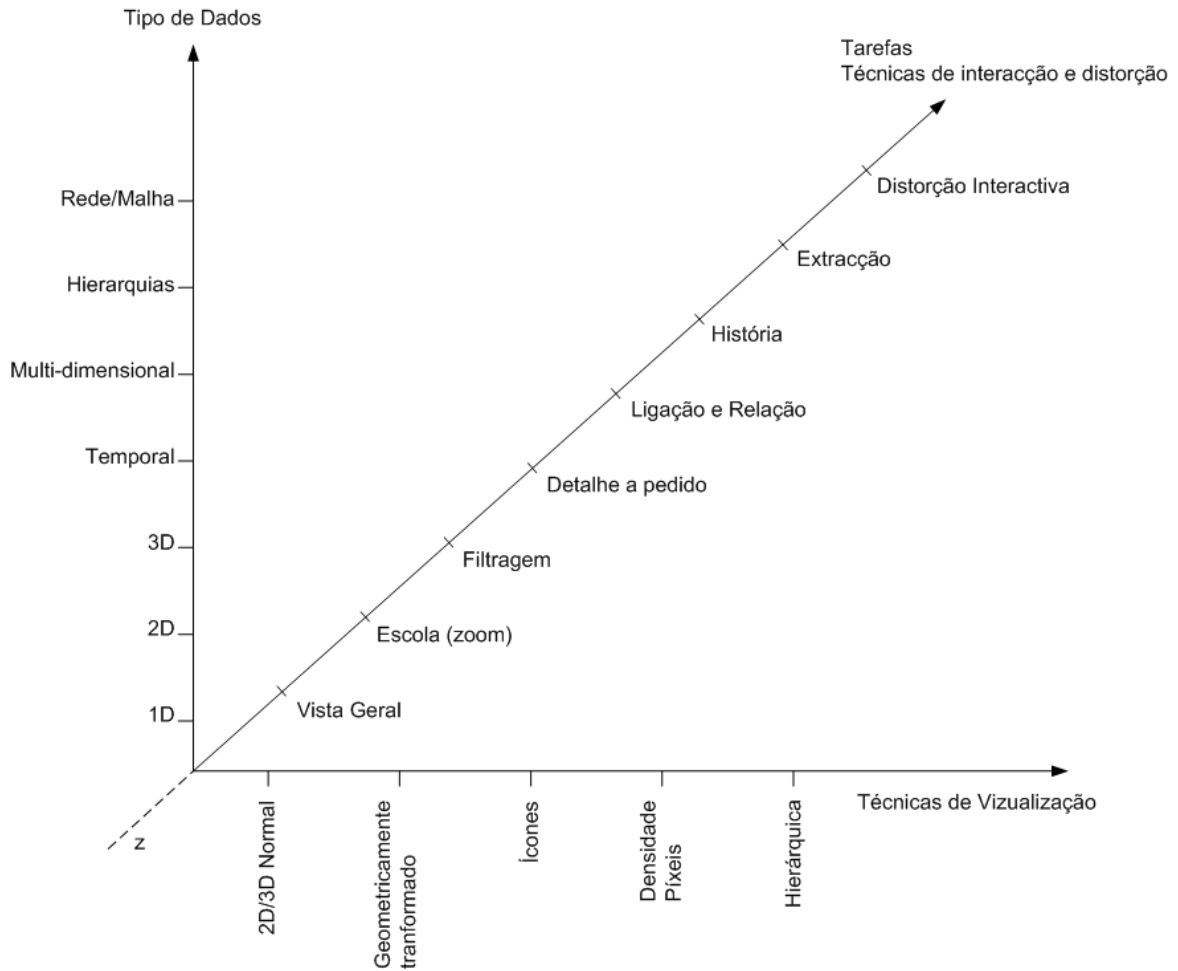
Os tipos de dados podem ser classificados segundo as seguintes categorias (Shneiderman, 1996):

- 1D

Tipo de dados lineares como documentos textuais ou código fonte de um programa. Cada item consiste numa linha de texto contendo uma *string* de caracteres.

- 2D

Tipo de dados planares como um mapa geográfico ou o *layout* de um jornal ou de uma página *web*. Cada item tem um conjunto de atributos relacionados com a tarefa como nome, proprietário, valor, etc. e atributos relacionados com a interface como tamanho, cor, opacidade, etc.



**Figura III-3:** Representação gráfica da taxonomia proposta, em que cada eixo corresponde a um tipo de classificação. No geral, um método de visualização corresponderá a um ponto de interceção nos três eixos apresentados (um ponto num espaço 3D). Baseado nas propostas de (Shneiderman, 1996), (Keim, 2000, 2001) e (Andrews, 2002).

- 3D

Tipo de dados relacionados com objetos físicos, nomeadamente do mundo real como o corpo humano ou edifícios que contêm características de volume e relações com outros itens potencialmente complexas.

- Temporal

As linhas temporais são usadas largamente em registos médicos, gestão de projetos ou apresentações históricas para criar um tipo de dados que é separado de tipo de dados unidimensional. A distinção nos dados temporais está no facto de que os itens têm um tempo de início e de fim e que os itens se podem sobrepor.

- **Multidimensional**

A maioria das bases de dados relacionais e estatísticas podem ser convenientemente manipuladas como dados multi-dimensionais em que itens com  $n$  atributos aparecem como pontos num espaço  $n$ -dimensional.

- **Hierarquias**

As hierarquias ou estruturas em árvore consistem num conjunto de itens em que cada item tem uma ligação a um item “pai” (exceto o item “raiz”). Os itens e as ligações entre “pais” e “filhos” podem ter múltiplos atributos.

- **Rede / malha**

Por vezes, as relações entre itens não são convenientemente capturadas recorrendo a uma estrutura em árvore e é útil ter itens ligados a um número arbitrário de outros itens. Apesar de existirem muitos casos especiais de redes (acíclicas, *lattices*, com raiz vs sem raiz, direcionada vs não direcionada) parece conveniente considerá-las com um único tipo de dados.

## **Métodos de visualização**

Os métodos de visualização podem ser agrupadas nas seguintes categorias (Keim, 2000, 2001):

- **Representação 2D / 3D normais**

Usando métodos de visualização 2D ou 3D *standard* (como gráficos  $x$ - $y$ ) para visualizar os dados.

- **Representações geometricamente transformadas**

Usando transformações geométricas e projeções para produzir visualizações úteis. Podem incluir-se aqui as coordenadas paralelas, projeção para um ponto de fuga e os vários métodos de visualização de grafos e redes.

- **Representação baseada em ícones**

Que visualizam cada item de dados com um ícone (como por exemplo figuras de ligação) e os valores de dimensão como propriedades dos ícones.

- **Representação por densidade de pixéis**

Que visualizam os valores de cada dimensão sob a forma de uma cor de um pixel ou de um grupo de pixéis de uma área adjacente. Ao organizar e colorir os pixéis de forma apropriada, a visualização resultante fornece



informação detalhada sobre as correlações e dependências locais, bem como os pontos de maior interesse.

- Representação hierárquica

Que permitem visualizar os dados de forma hierárquica. No caso de dados multidimensionais, as dimensões dos dados a ser usadas têm que ser cuidadosamente selecionadas. Para se obter visualizações úteis, as dimensões mais importantes têm que corresponder aos primeiros níveis da hierarquia.

## Técnicas de interação e distorção

As tarefas a realizar com os dados / técnicas de interação e distorção podem ser incluídas numa das seguintes categorias (Shneiderman, 1996):

- Vista geral (*overview*)

Ter uma vista geral, holística, de todo o conjunto de dados.

- Escala (*zoom*)

Ampliação dos itens com interesse, para análise de detalhes e com isso compreender a dimensão de complexidade da informação em causa.

- Filtragem

Filtrar para retirar os itens que não suscitem interesse.

- Detalhe a pedido

Selecionar um item ou um grupo de itens e obter detalhes quando necessário.

- Ligação e relacionamento

Observar as ligações e relações existentes entre itens.

- História

Manter um historial das ações para poder suportar ações de *undo* (desfazer), *replay* (repetir) e refinamento progressivo. É raro que uma única ação do utilizador produza o resultado pretendido. A exploração de informação é, inerentemente, um processo com vários passos, pelo que manter a história das suas ações e permitir que as reutilize é importante.

- Extração

Permitir a extração de um subconjunto de dados e dos parâmetros de pesquisa. Depois de o utilizador obter o item ou conjunto de itens que

deseja, pode ser útil ser possível extrair este conjunto e guardá-lo para futura utilização.

- Distorção interativa (Keim, 2000, 2001)

Suporta o processo de exploração dos dados conservando uma vista geral durante as operações de refinamento. Basicamente, mostra partes dos dados com um elevado grau de detalhe, enquanto outras partes são mostradas com um nível de detalhe mais reduzido. As distorções hiperbólica e esférica são algumas das técnicas de distorção mais populares.

### 4.3. Aplicação de cor – propostas de tabelas de cor

Na sequência do apresentado no Capítulo II.4.1. (Aplicação de cor em visualização), foram produzidas algumas tabelas de cor. Durante a sua utilização, algumas tabelas mostraram-se mais adequadas para uma dada tarefa, enquanto outras se mostraram desadequadas.

Uma das tabelas mais simples e mais usadas é a apresentada na Figura III-4. Consiste numa variação do nível de luminosidade do branco, resultado num gradiente de tons de cinzento, do preto mais escuro ao branco mais luminoso.



**Figura III-4:** Tabela de cinzentos, de preto a branco com variação linear. Nesta tabela só há variação da luminância.

A Figura III-5 corresponde a uma variante, usando-se uma variação de ciano em vez de branco. Pode ser utilizada a variação de qualquer outra cor.



**Figura III-5:** Tabela de tons de ciano. Varia desde preto até ao ciano mais saturado. A luminância tem uma variação linear.

Para ajudar a resolver os problemas relacionadas com a percepção das cores de que algumas pessoas sofrem (deuteranopia), pode usar-se a tabela da Figura III-6, com uma variação entre azul e amarelo. Tem ainda a vantagem de a variação de luminância também ser sempre crescente.



**Figura III-6:** Tabela de cor com variação entre azul e amarelo. Esta tabela tem a particularidade de associar uma variação de luminância e de crominância.

A Figura III-7 representa uma tabela de pseudocor elaborada para utilização na visualização de dados obtidos de algumas modalidades imagiológicas, usadas como meios auxiliares de diagnóstico em medicina. Apesar de permitir um nível de discriminação maior do que a verificada por algumas das tabelas anteriormente apresentadas, tem alguns inconvenientes, sobretudo para pessoas com deficiências ao nível de percepção das cores, pois o nível de luminância tem algumas variações ao longo da tabela, como pode ser observado na Figura III-8.



**Figura III-7:** Tabela de pseudocor usada em algumas modalidades imagiológicas médicas.



**Figura III-8:** Níveis de luminância da tabela de pseudocor apresentada na Figura III-7.

## 4.4. Visualização da organização do campo visual de interação

Nesta secção serão apresentados os métodos propostos para a visualização da organização do campo visual da interação de uma dada página.

### 4.4.1. Visualização das áreas (blocos) de uma página

#### Classificação taxonómica

Tipo de dados: **2D**

Técnica de visualização: **Normal 2D**

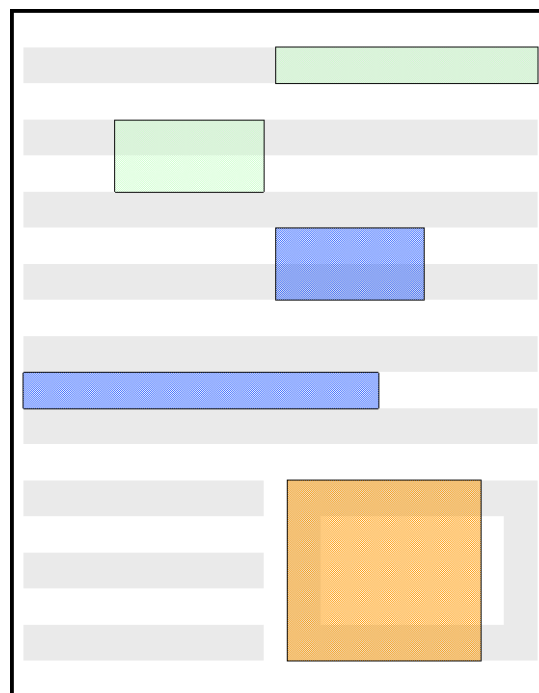
Técnica de interação e distorção: **Vista geral**

## Descrição

Uma página *web* é normalmente constituída por vários elementos textuais e gráficos, como imagens. Esta informação pode ser agrupada em blocos, como por exemplo blocos de texto, blocos de imagem, blocos de interação ou ainda baseados em algum critério que permita identificar algum aspeto pertinente que se pretenda analisar na página.

Tipicamente, os blocos são identificados por um conjunto de atributos e por um conjunto de coordenadas que definem a área (a forma retangular é a mais comum) da página em que esse bloco se localiza. Utiliza-se como base uma representação da página (como ela aparecerá no ecrã do utilizador) sobre a qual serão desenhados os blocos identificados. Os atributos poderão ser utilizados para codificar a cor e/ou transparência dos mesmos. Por exemplo, pode atribuir-se uma cor para os blocos de texto, outra para os blocos de imagem e outra ainda para os que identificam zonas interativas. De notar que as zonas podem ficar sobrepostas, pois pode haver critérios que não são mutuamente exclusivos.

Na Figura III-9 pode observar-se um esquemático ilustrativo da representação de um conjunto de cinco áreas em três classes diferentes (por ex. texto, *link* e imagem), cada uma codificada com a sua cor.



**Figura III-9:** Exemplo de visualização, recorrendo a cor, da classificação das zonas identificadas numa página *web*.

## 4.4.2. Visualização de detalhes de zonas interativas

### Classificação taxonómica

Tipo de dados: **1D**

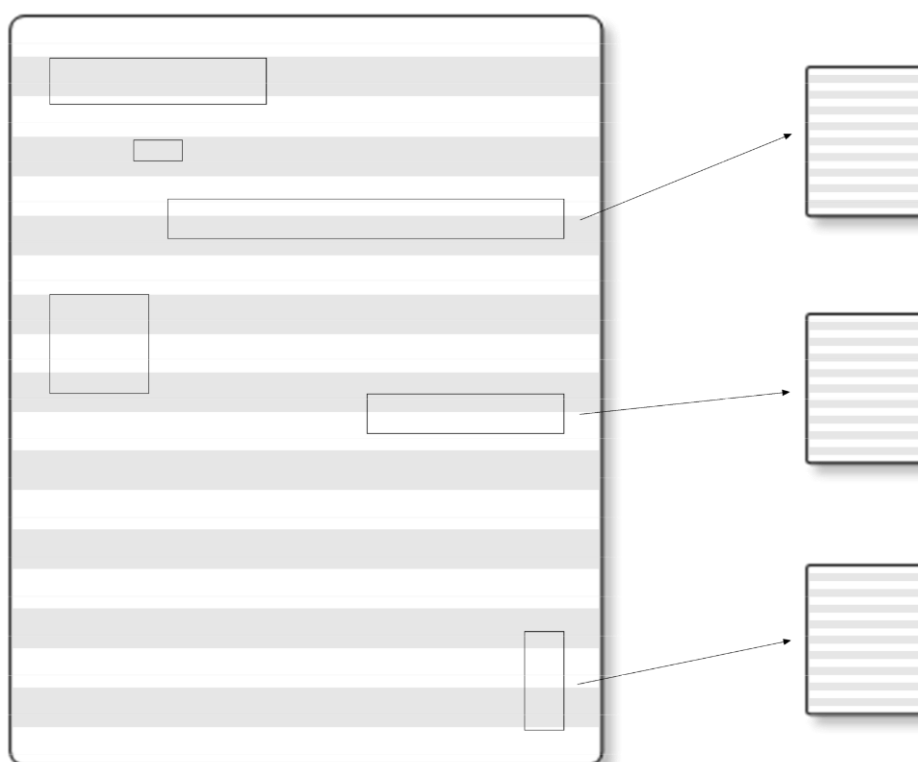
Técnica de visualização: **Hierárquica**

Técnica de interação e distorção: **Ligação e Relação**

### Descrição

Das áreas contidas numa página, é comum algumas identificarem zonas interativas (*Hotspots*) que podem permitir que o utilizador seja transportado para outra página do *site*.

Este método de visualização acrescenta, em relação ao apresentado acima, uma representação das páginas às quais se referem os *links* contidos na página. Na Figura III-10 pode observar-se um esquemático ilustrativo da representação de algumas áreas, das quais três contêm *links* para outras páginas que estão representadas em miniatura do lado direito.



**Figura III-10:** Exemplo de visualização das zonas de interação de uma página *web* e representação das páginas às quais estão ligadas.

### 4.4.3. Visualização de detalhes de zonas interativas em 3D

#### Classificação taxonómica

Tipo de dados: 1D

Técnica de visualização: **Hierárquica**

Técnica de interação e distorção: **Ligação e Relação**

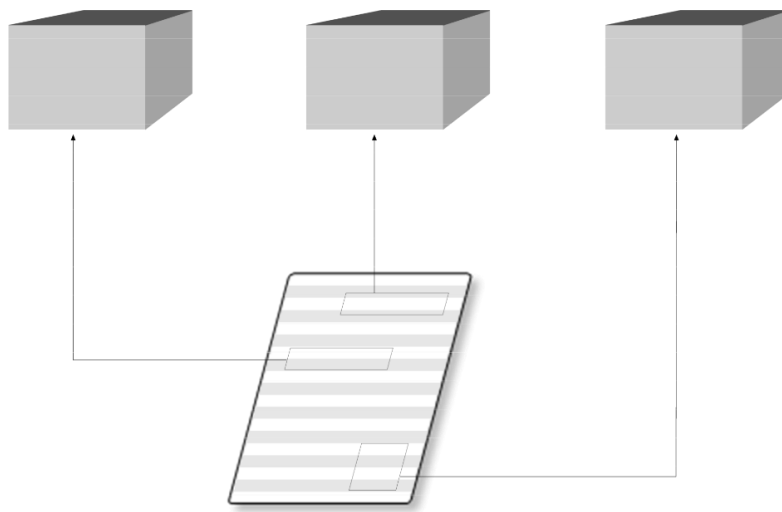
#### Descrição

Este método de visualização é em tudo semelhante ao anterior, com a única diferença que as representações em miniatura das páginas associadas aos *links* das zonas interativas são apresentadas em 3D.

Este método pode ser útil para situações em que uma página tem uma grande quantidade de *links*. Nestes casos, uma representação em 3D pode permitir ter uma visão global do conjunto alargado de objetos, ainda que perdendo os detalhes. O que ganha em visão global do conjunto, perde em detalhe do particular.

Na Figura III-11 pode ver-se um esquemático ilustrativo deste método de visualização em que cada uma das três áreas identificadas na página se encontra ligada a uma representação 3D da página destino.

As representações miniatura das páginas referentes aos *links* da página podem ser representados por uma simples folha ou por formas mais complexas como um paralelepípedo, como apresentado na Figura III-11.



**Figura III-11:** Exemplo de visualização das zonas de interação de uma página *web* e representação 3D das páginas às quais estão ligadas.

#### 4.4.4. Relações de uma página

##### Classificação taxonómica

Tipo de dados: **2D**

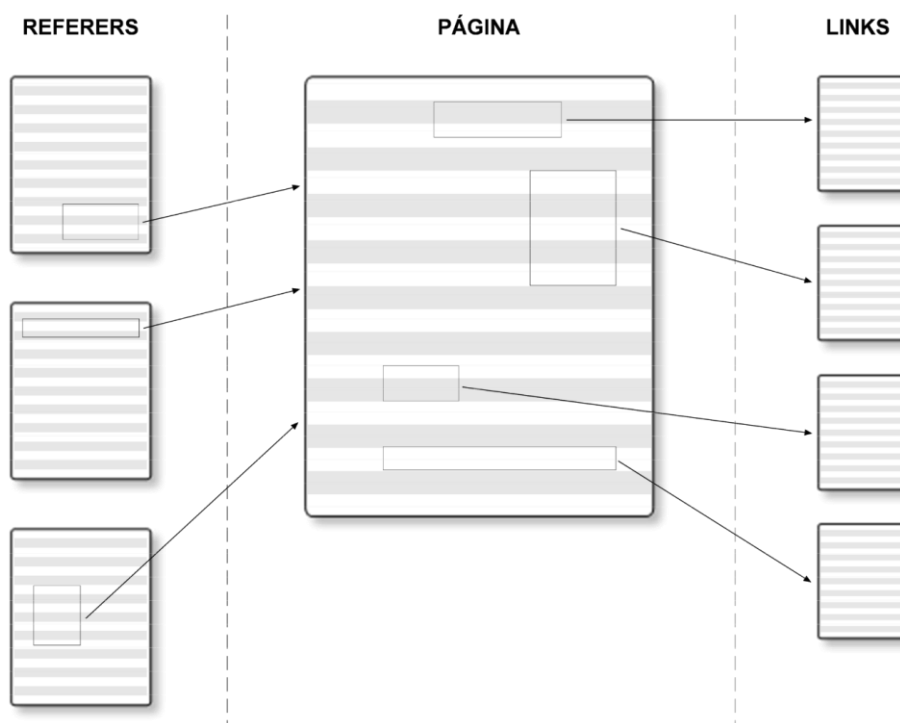
Técnica de visualização: **Hierárquica**

Técnica de interação e distorção: **Ligações e Relações**

##### Descrição

Este método de visualização acrescenta aos anteriores a informação de *referer*. Desta maneira, é possível não só visualizar para onde se pode ir a partir da página corrente, como também de onde se pode chegar.

Na Figura III-12 pode ver-se um esquemático exemplificativo. Na parte central da figura encontra-se a página corrente com as suas áreas interativas identificadas. Os *links* associados a estas áreas aparecem no lado direito da figura, sob a forma de representações miniatura das respetivas páginas. No lado esquerdo, aparecem todas as páginas que contêm *links* que apontam para a página corrente, podendo ser identificadas as áreas interativas contendo esses *links* em cada uma delas.



**Figura III-12:** Exemplo de visualização de uma página *web* com as suas zonas de interação. Do lado esquerdo, encontram-se as páginas que contêm *links* (com as respetivas áreas identificadas) que apontam para a página corrente. Do lado direito, estão as páginas para as quais apontam cada um dos *links* da página corrente.

## 4.4.5. Informação sobre uma área da página

### Classificação taxonómica

Tipo de dados: **2D**

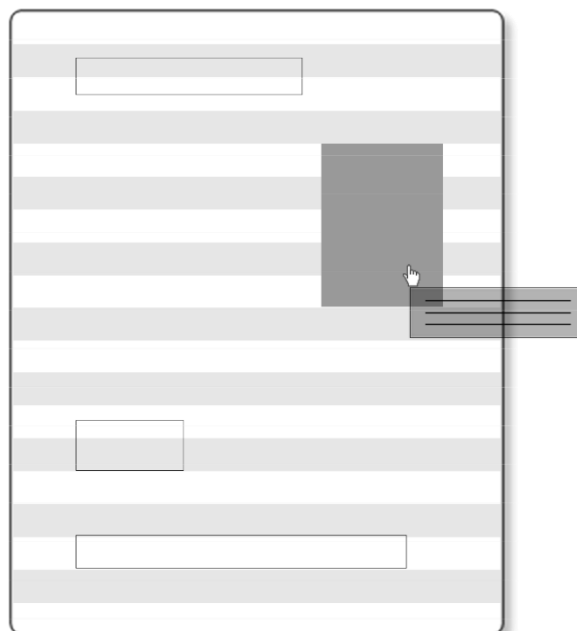
Técnica de visualização: **Normal 2D**

Técnica de interação e distorção: **Detalhe a pedido**

### Descrição

Este é mais um esquema de visualização que pode ser utilizado como complemento à maioria dos apresentados atrás. Aqui, quando o analista desloca o cursor do rato sobre uma área identificada, será apresentada informação adicional sobre essa área em particular. Pode ser apresentada informação como o *URL* do *link* associado, a categoria da área, informação estatística tal como o número de utilizadores que utilizaram esse *link* nesta página, etc.. A informação é apresentada numa pequena janela que se sobrepõe a tudo o resto e que aparece junto do cursor do rato. Esta janela poderá ter um fundo e texto opacos ou apresentar algum nível de transparência de forma a não provocar oclusão.

A Figura III-13 apresenta um esquemático que representa o conceito apresentado.



**Figura III-13:** Exemplo de visualização dos detalhes referentes à área sob o cursor do rato. Os detalhes aparecem como uma pequena janela de texto (*tool-tip window*), opaca ou semitransparente, junto ao cursor do rato.



## 4.4.6. Visualização da página associada ao *link*

### Classificação taxonómica

Tipo de dados: **2D**

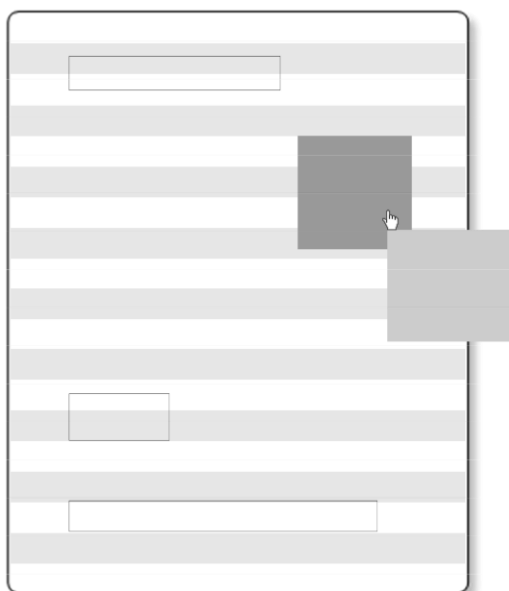
Técnica de visualização: **Normal 2D**

Técnica de interação e distorção: **Detalhe a pedido**

### Descrição

Este esquema é idêntico ao anterior e pode ser utilizado separadamente ou em simultâneo com ele. Um dos tipos de áreas de grande interesse na análise de uma página tem a ver com a interatividade. Estes normalmente têm *links* associados, pelo que pode ser útil ter uma ideia da página para a qual o *link* aponta. É aqui que este esquema aparece, permitindo observar uma janela com uma representação miniatura da página indicada pelo *link* com a simples colocação do cursor do rato numa área interativa que contenha um *link*. Esta representação, tal como no caso anterior, aparecerá junto do cursor do rato e poderá ser semitransparente – o que neste caso é de particular importância, pois pode ser necessário que esta representação tenha um tamanho considerável.

A Figura III-14 apresenta um esquemático ilustrativo do esquema descrito.



**Figura III-14:** Exemplo de visualização da miniatura da página a que se refere o *link* associado à área sob o cursor do rato. A miniatura da página como uma pequena janela (*tool-tip window*), opaca ou semitransparente, junto ao cursor do rato.

## 4.4.7. Apresentação de informação de *eye-tracking*

### Classificação taxonómica

Tipo de dados: **Temporal**

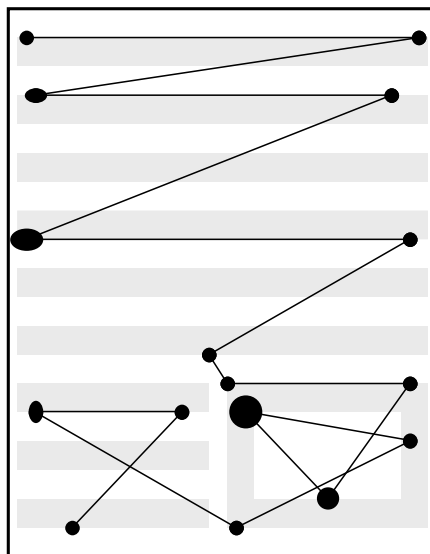
Técnica de visualização: **Normal 2D**

Técnica de interação e distorção: **Vista geral**

### Descrição

Este método de visualização permite observar o comportamento do utilizador enquanto ele utiliza uma dada página de um *site*. O local da página para o qual o utilizador olha em cada momento pode fornecer informação de grande utilidade. Para atingir este objetivo, cada evento detetado pelo equipamento de *eye-tracking* é colocado sobre uma representação da página em estudo. Cada tipo de evento pode ser codificado com uma forma geométrica ou uma cor diferente. Os vários eventos são ligados por linhas para se ter a noção do percurso utilizado. O tempo que o utilizador se deteve em cada ponto pode também ser codificado alterando a área desse ponto.

A Figura III-15 apresenta um esquemático ilustrativo de parte do conceito apresentado. Na situação ilustrada, o utilizador começou a olhar para a página no canto superior esquerdo, passando rapidamente para o canto superior direito e voltando depois para a esquerda, agora para o início da segunda linha. É possível observar que o utilizador se deteve bastante tempo numa zona do canto inferior direito, o que pode significar que aí se encontra algo particularmente interessante ou difícil de compreender.



**Figura III-15:** Exemplo de visualização do foco e sentido do olhar de um utilizador enquanto observa uma página web.

## 4.4.8. Apresentação de informação de *mouse-tracking*

### Classificação taxonómica

Tipo de dados: **Temporal**

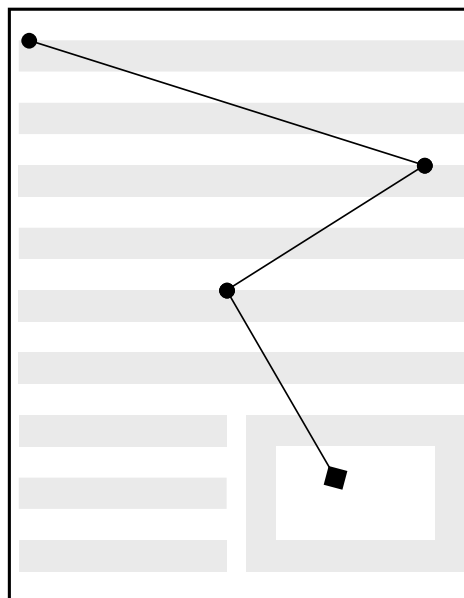
Técnica de visualização: **Normal 2D**

Técnica de interação e distorção: **Vista geral**

### Descrição

A observação da atividade do rato tem muitas semelhanças com o esquema anterior (posição do olhar). Aqui, é necessário representar a posição do cursor do rato a cada momento, bem como de qualquer evento relacionado (clique, indicando qual a tecla utilizada, movimento, arrastar, etc.). Tal como no caso anterior, cada evento pode ser codificado com um ponto com uma forma geométrica ou uma cor diferente, ligando-se os pontos com linhas para se ter a noção do percurso.

A Figura III-16 é um esquemático ilustrativo do esquema apresentado. Neste caso, o utilizador poderia ter começado com o cursor do seu rato no canto superior esquerdo, onde esteve parado uns instantes, depois deslocou para a direita e para baixo, tendo parado mais um instante. Voltou a deslocar, desta vez para baixo e para a esquerda, por mais um instante e depois novamente para baixo e para a direita, onde decide fazer um clique com a tecla esquerda do rato (num *link* por exemplo).



**Figura III-16:** Exemplo de visualização dos eventos relacionados com o rato.

## 4.4.9. Visualização 3D dos *links* utilizados por um utilizador

### Classificação taxonómica

Tipo de dados: **Temporal**

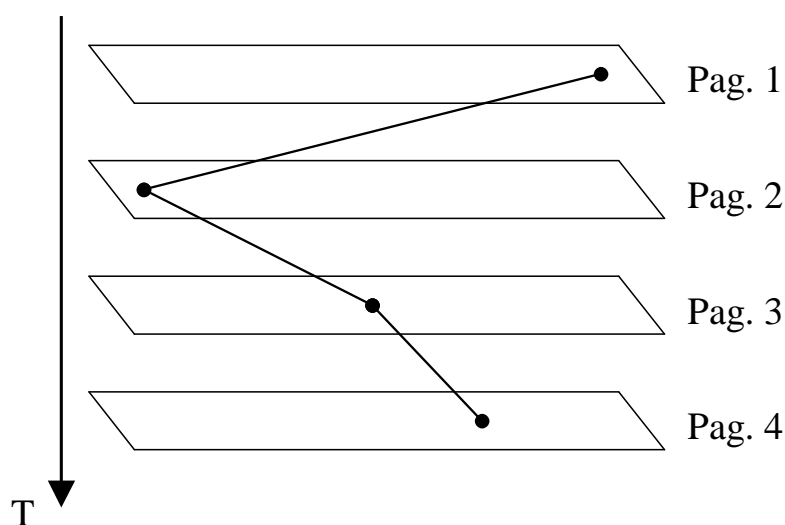
Técnica de visualização: **Transformação geométrica**

Técnica de interação e distorção: **Vista geral**

### Descrição

Um utilizador, quando visita um *site*, abre uma página, seleciona um *link* e conseqüentemente salta para outra página, volta a selecionar outro *link* e assim sucessivamente. Desta maneira vai percorrendo as páginas de um *site*. Por vezes tem interesse poder analisar este processo. Como evolui a zona onde o utilizador interage com as páginas?

Para ajudar a estudar este aspeto pode usar-se um esquema de visualização como o ilustrado de forma esquemática na Figura III-17. Representam-se as várias páginas visitadas pelo utilizador, empilhadas num arranjo 3D, de forma a permitir vislumbrar o seu conteúdo. De seguida, assinala-se o ponto ou área onde se clicou com o rato em cada uma das páginas. Para se ficar com uma melhor perceção da sequência, podem traçar-se linhas que unem esses pontos. Esta representação permite ainda ficar com a perceção do tempo que o utilizador levou a ir de uma página para outra, bastando que as páginas na pilha 3D sejam colocadas a uma distância em função do tempo.



**Figura III-17:** Exemplo de visualização do percurso de um utilizador num *site* com identificação dos *links* utilizados representados ao longo de uma linha temporal.

## 4.4.10. Visualização dos *links* usados por um utilizador

### Classificação taxonómica

Tipo de dados: **Temporal**

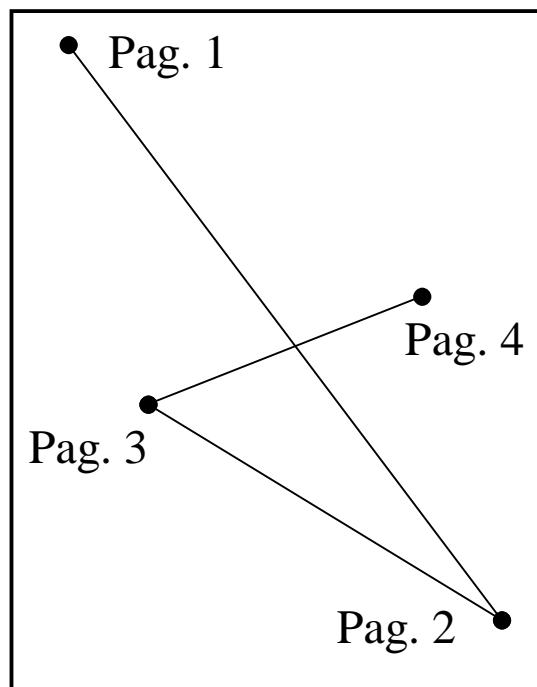
Técnica de visualização: **Transformação geométrica**

Técnica de interação e distorção: **Vista geral**

### Descrição

Este método de visualização é um caso particular da representação anterior. Colocando o ponto de vista alinhado com um eixo, no espaço 3D, de tal maneira que a pilha de páginas apareça completamente de topo, umas páginas sobre as outras, fica-se com a vista sobre uma única página. Pode aproveitar-se este aspeto e projetar todos os pontos ou áreas de interação de forma a obter-se uma única imagem 2D. Esta irá permitir, por exemplo, verificar a coerência das áreas de interação ao longo do *site* / sessão de utilização.

A Figura III-18 apresenta um exemplo esquemático representativo do esquema de visualização descrito. A avaliar pela representação, poderia concluir-se que os pontos de interação nesta sessão deste utilizador estão muito dispersos.



**Figura III-18:** Exemplo de visualização do percurso de um utilizador num *site* com identificação dos *links* utilizados, representados num plano.

## 4.4.11. Visualização da distribuição espacial de áreas de um *site*

### Classificação taxonómica

Tipo de dados: **2D**

Técnica de visualização: **Normal 2D**

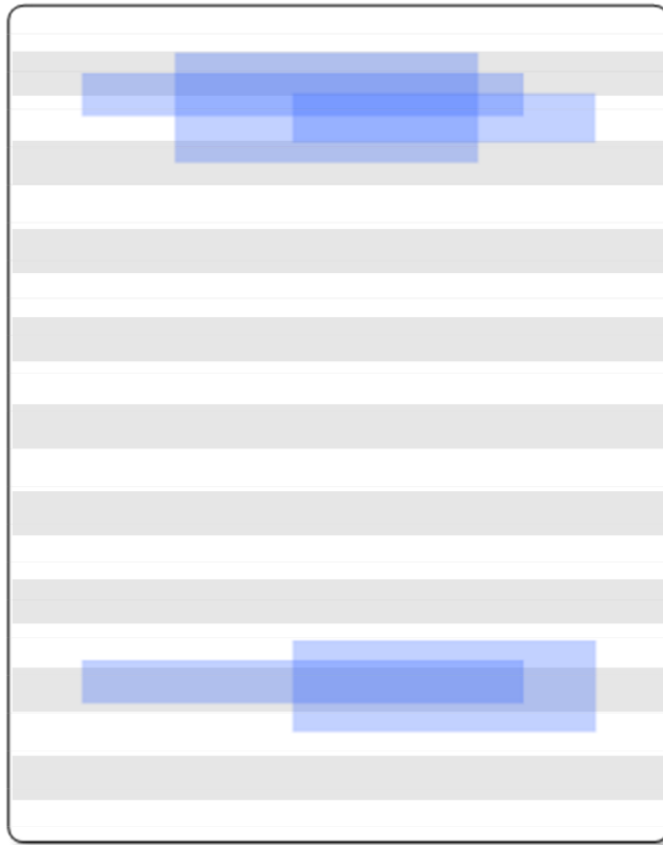
Técnica de interação e distorção: **Vista geral**

### Descrição

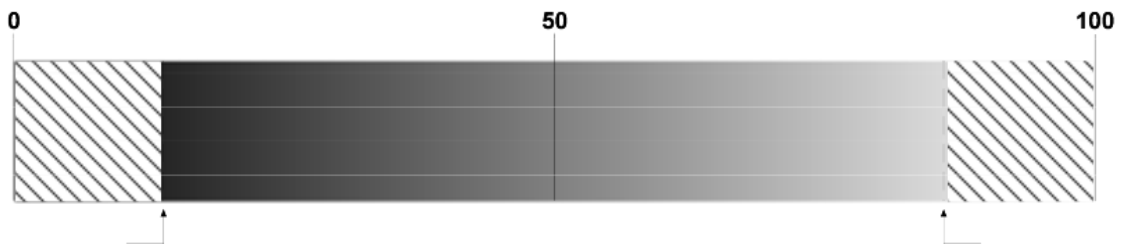
Na secção 4.4.1 foi vista uma técnica para observar as áreas classificadas de uma dada página. Contudo, ainda que a observação das áreas de uma página em particular possa ser útil, há situações em que é importante perceber a distribuição das mesmas nas páginas de um *site* completo. Para atingir esse objetivo, pode sobrepor-se cada página e respetivas áreas. Calculando, para cada ponto, o número de páginas que têm alguma área que cubra esse ponto, é possível criar um mapa de densidade. Aplicando uma tabela de cor para codificar os valores encontrados, obtém-se uma imagem que permite identificar rapidamente se um dado tipo de área está distribuído de forma coerente ao longo das páginas de um *site*.

Por exemplo, as áreas de interação (*Hotspots*) – zonas clicáveis, são de grande importância. Um *site* que apresente as suas zonas de interação completamente dispersas poderá indiciar algum problema de design. Já uma zona concentrada poderá testemunhar um *site* com um design cuidado.

A Figura III-19 apresenta um esquemático ilustrativo do método apresentado, em que se podem observar zonas com valores de 0 a 3. A Figura III-20 mostra uma interface para a manipulação dos limites da tabela de pseudocor a aplicar ao método de visualização. Ao deslocar os limites inferior e superior, é possível isolar uma gama de valores que se pretenda visualizar, ocultando todos os restantes. Por exemplo, pode observar-se só as zonas com valor entre 2 e 3.



**Figura III-19:** Exemplo de visualização da distribuição espacial das áreas de um dado tipo ao longo do *site*. Os valores (número de ocorrências, para um dado ponto) poderão ser codificados com recurso a uma tabela de cor.



**Figura III-20:** Interface de manipulação dos limites da tabela de pseudocor a aplicar a um esquema de visualização. Os limites inferior e superior podem ser deslocados de forma a só ficar visível a parte correspondente aos valores entre os limites.

## 4.5. Visualização da estrutura do *site* / sessão

Nesta secção serão apresentados os métodos propostos para a visualização da representação da estrutura do *site*, suas páginas e ligações entre elas, bem como a representação e exploração do percurso dos utilizadores durante uma ou mais sessões.

### 4.5.1. Visualização das páginas de um *site*

#### Classificação taxonómica

Tipo de dados: **Hierarquia**

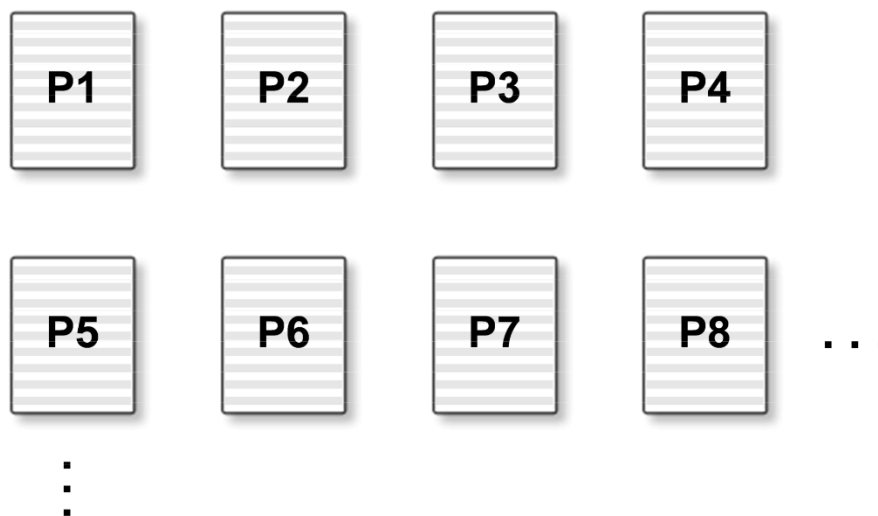
Técnica de visualização: **Ícones (*dashboard* estruturado)**

Técnica de interação e distorção: **Vista geral**

#### Descrição

Este método de visualização permite observar o conjunto de páginas que constituem um *site* ou ainda o conjunto de páginas visitadas durante uma ou várias sessões de um ou mais utilizadores. Nesta vista, cada página é representada na forma de uma miniatura, ordenadas por linhas e colunas, formando uma matriz bidimensional. Podem utilizar-se vários critérios para definir a ordem com que as páginas são colocadas: por ordem do número de visitas, sendo primeiro colocadas as páginas mais visitadas, pela ordem que foram visitadas numa dada sessão de um utilizador, entre outros critérios.

Na Figura III-21 pode observar-se uma ilustração esquemática do método apresentado.



**Figura III-21:** Exemplo de visualização da estrutura de um *site* em 2D. Tem-se simplesmente uma representação de cada página constituinte do *site*.



## 4.5.2. Visualização da estrutura de um *site* (versão 1)

### Classificação taxonómica

Tipo de dados: **Hierarquia**

Técnica de visualização: **Hierárquica**

Técnica de interação e distorção: **Vista geral**

### Descrição

Há muitas ocasiões em que é importante analisar a estrutura de um *site*.

Um *site* normalmente tem uma página principal ou de entrada, a partir da qual se pode chegar a outras páginas seguindo aos seus *links*. Cada uma destas páginas poderá também ter o seu conjunto de *links*, que permite chegar a ainda mais páginas e assim sucessivamente. Podemos assim definir níveis, em que no primeiro nível está a página inicial, no 2º nível estão todas as páginas referentes aos *links* presentes na página inicial, no 3º nível estão todas as páginas referentes aos *links* das páginas do 2º nível e assim sucessivamente.

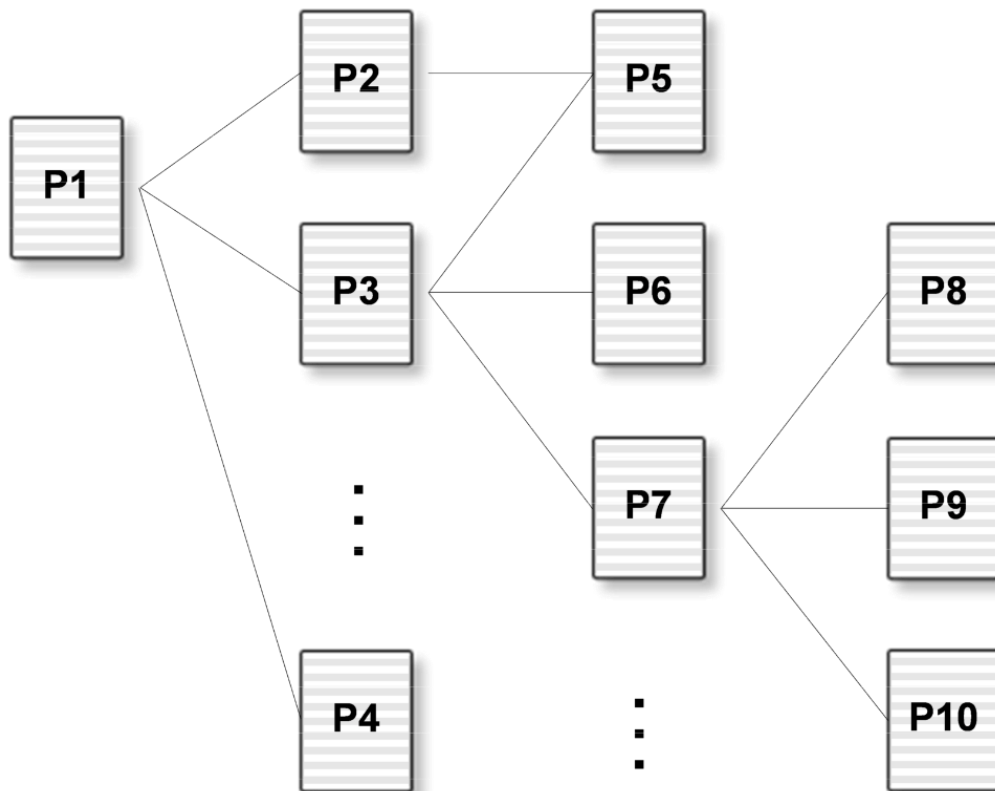
Esta estrutura resultará numa representação em árvore, em que cada página de um nível se pode encontrar, ou não, ligada a uma ou mais páginas do nível seguinte. O número de ligações será igual ao número de *links* diferentes que essa página contenha.

Para assegurar que a estrutura em árvore se mantém coerente, só existem ligações de páginas de um nível para páginas um nível imediatamente superior. Para que isto possa acontecer poderá ser necessário representar uma mesma página mais que uma vez, em níveis diferentes. Um exemplo é quando num dado *site*, todas as páginas têm um *link* para a página principal.

Para evitar situações de ciclo fechado que originariam níveis sem fim, sempre que aparece uma página previamente representada num nível anterior, esta é identificada (com uma cor diferente, por exemplo) e os *links* desta não voltam a produzir ligações para um novo nível.

Alternativamente, poderá não se impor esta restrição, especificando somente limites quanto aos níveis que se pretendem observar. Por exemplo, visualizar somente os níveis 3 a 5, permitindo observar todas as possibilidades dentro desses níveis.

A Figura III-22 ilustra de forma esquemática o método apresentado. Nesta ilustração pode observar-se que a página P1 tem *links* para as páginas P2, P3 e P4. A página P2 tem um *link* para a página P5 e assim sucessivamente. De notar que a página P5 é referida tanto por P2 como por P3.



**Figura III-22:** Exemplo de visualização da estrutura de um *site* em 2D, com as páginas organizadas por níveis e respectivas ligações.

### 4.5.3. Visualização da estrutura de um *site* (versão 2)

#### Classificação taxonómica

Tipo de dados: **Hierarquia**

Técnica de visualização: **Hierárquica (3D)**

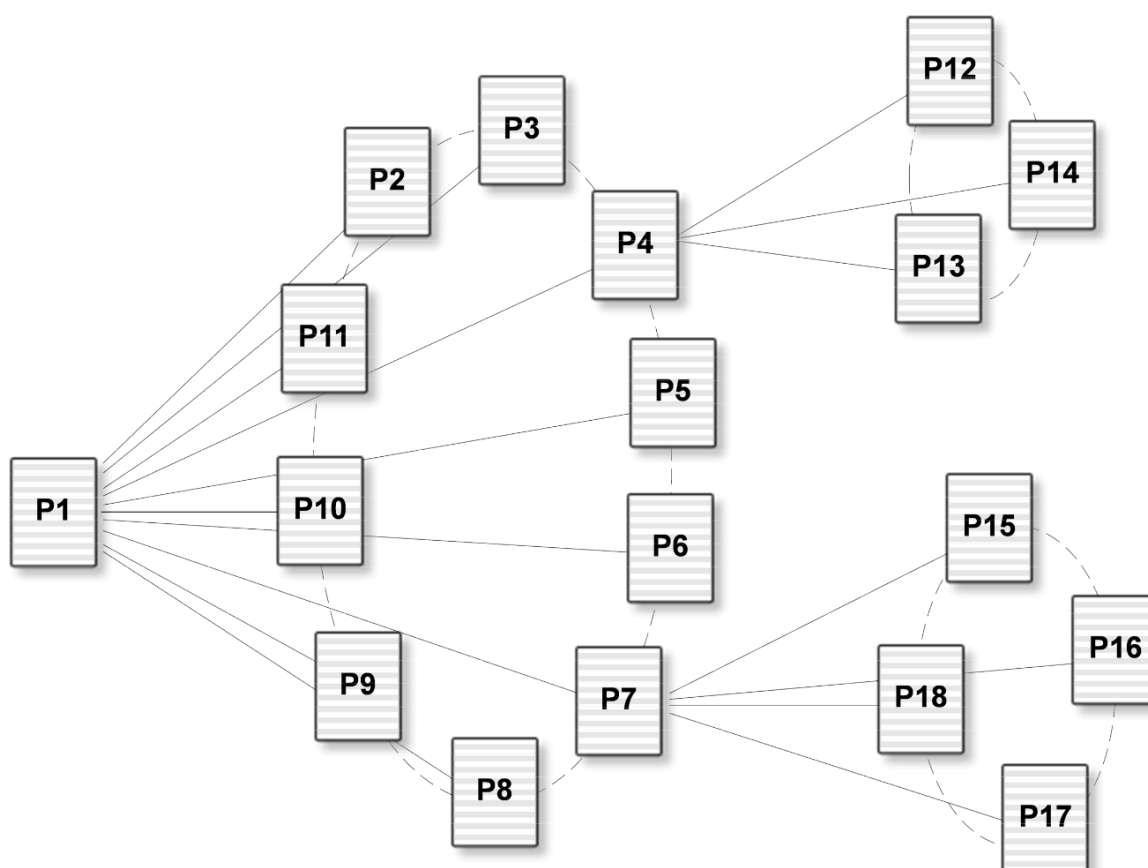
Técnica de interação e distorção: **Vista geral**

#### Descrição

Este método de visualização é muito semelhante ao método apresentado imediatamente acima. Neste caso, as páginas são representadas como ícones de páginas, organizadas por níveis, colocadas num espaço 3D. Cada folha pode estar ligada a zero, uma ou mais páginas do nível seguinte, dependendo dos *links* que contenha. Também aqui, uma dada página pode aparecer mais que uma vez, aplicando-se as restrições descritas no método anterior.

A vantagem deste método só é evidente quando se está a lidar com uma grande quantidade de páginas em cada nível. A utilização de uma estrutura 3D permite que as páginas possam ser parcialmente sobrepostas, podendo mesmo usar-se semitransparência para permitir uma visão global do contexto, mesmo das páginas parcialmente ocultadas por outras.

A Figura III-23 ilustra esquematicamente o método apresentado. Note-se que as páginas referentes aos *links* de uma dada página estão organizadas numa forma circular, formando um cone, quando ligadas a esta página. Este fenómeno pode ser observado, por exemplo, no caso da página P4 e as suas ligações com as páginas P12 a P14 e no caso da página P7 e suas ligações com as páginas P15 a P18. Repare-se que as página P4 e P7 fazem parte do nível 2, enquanto as páginas P12 a P14 e P15 a P18 se encontram no nível 3.



**Figura III-23:** Exemplo de visualização da estrutura de um *site* em 3D. Cada página é representada como uma folha nesse espaço 3D. As páginas estão organizadas por níveis. Uma página de um dado nível pode estar ligada a zero, uma ou mais páginas do nível seguinte – *links* da página. Cada página poderá ser representada mais que uma vez, em níveis diferentes.

## 4.5.4. Visualização da estrutura de um *site* (versão 3)

### Classificação taxonómica

Tipo de dados: **Hierarquia**

Técnica de visualização: **Hierárquica (3D)**

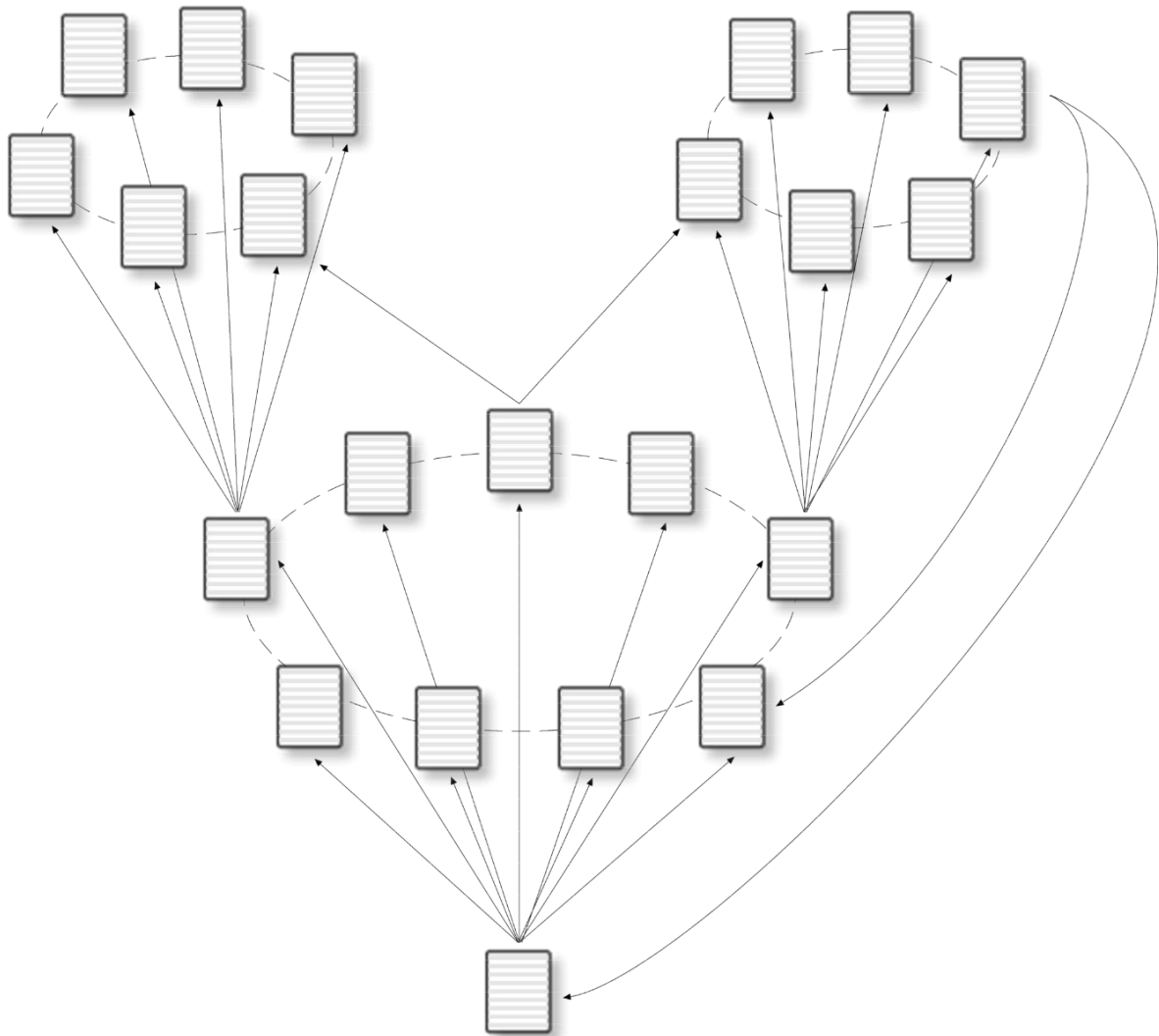
Técnica de interação e distorção: **Vista geral**

### Descrição

Este método é semelhante aos dois anteriores. As páginas podem ser representadas num plano (2D), como descrito em 4.5.2, ou no espaço (3D), como descrito em 4.5.3. Como diferenças tem-se que cada página é representada uma única vez e que podem haver ligações de uma página de um nível para outra de um nível anterior. Os níveis são definidos tendo em consideração a primeira vez que esta aparece referida por um *link* de outra página. Assim, se aparece referida por um *link* da página de entrada (1º nível), essa página será de segundo nível. Isto acontece porque raramente se tem acesso à estrutura hierárquica utilizada na conceção do *site*. Contudo, se essa informação estiver disponível, poderá ser utilizada para definir os níveis.

É importante referir que, para um *site* de dimensão e/ou número de ligações apreciável, poderá ter como resultado uma vista muito complexa devido ao “emaranhado” de ligações entre as páginas que terão que ser representadas.

A Figura III-24 ilustra de forma esquemática o método descrito para uma situação relativamente simples.



**Figura III-24:** Exemplo de visualização da estrutura de um *site* em 3D. Cada página é representada como uma folha nesse espaço 3D. As páginas estão organizadas por níveis. Uma página de um dado nível pode estar ligada a uma ou mais páginas – *links* da página – que podem encontrar-se em qualquer nível. Cada página é representada uma única vez.

## 4.5.5. Visualização da estrutura de um *site* em 3D

### Classificação taxonômica

Tipo de dados: **Hierarquia**

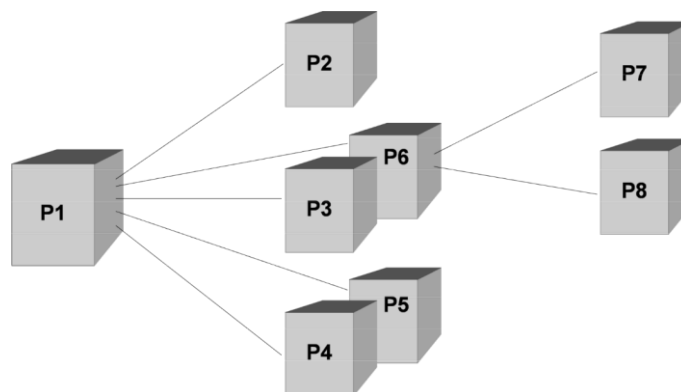
Técnica de visualização: **Hierárquica**

Técnica de interação e distorção: **Vista geral**

## Descrição

Este método tem exatamente a mesma funcionalidade e objetivos apresentados nas secções anteriores. Aqui, cada página é representada sob a forma de um paralelepípedo. Pode desenhar-se o conteúdo da página em todas as faces ou escolher só uma das faces para o efeito, podendo usar-se as restantes faces para representar outro tipo de informação relacionada com a página. Posteriormente, pode manipular-se o ponto de observação 3D para escolher a vista mais interessante para executar uma dada tarefa.

A Figura III-25 ilustra, de forma esquemática, o método apresentado.



**Figura III-25:** Exemplo de visualização da estrutura de um *site* em 3D. Cada página é representada também em 3D. As páginas estão organizadas por níveis. Uma página de um dado nível pode estar ligada a uma ou mais páginas do nível seguinte – *links* da página.

### 4.5.6. Informação sobre uma página

#### Classificação taxonómica

Tipo de dados: **2D**

Técnica de visualização: **Normal 2D**

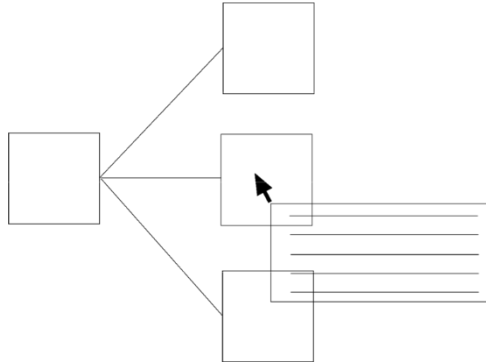
Técnica de interação e distorção: **Detalhe a pedido**

#### Descrição

Este é um esquema de visualização que pode ser utilizado como complemento à maioria dos métodos de visualização da estrutura de um *site*. Aqui, quando a/o analista desloca o cursor do rato sobre uma página, será apresentada informação adicional sobre essa página em particular. Pode ser apresentada informação como o *URL* da página, o nome ou título da mesma, o nível em que está classificada ou ainda informação estatística, tal como o número de utilizadores que visitaram a página, etc.. Este esquema é em tudo

semelhante ao apresentado na secção 4.4.5, aplicado a um cenário diferente. A informação é apresentada da mesma maneira.

A Figura III-26 apresenta um esquemático ilustrativo do método apresentado.



**Figura III-26:** Exemplo de visualização dos detalhes referentes à página sob o cursor do rato. Os detalhes aparecem como uma pequena janela de texto (*tool-tip window*), opaca ou semitransparente, junto ao cursor do rato.

## 4.5.7. Visualização da página ampliada

### Classificação taxonómica

Tipo de dados: **2D**

Técnica de visualização: **Normal 2D**

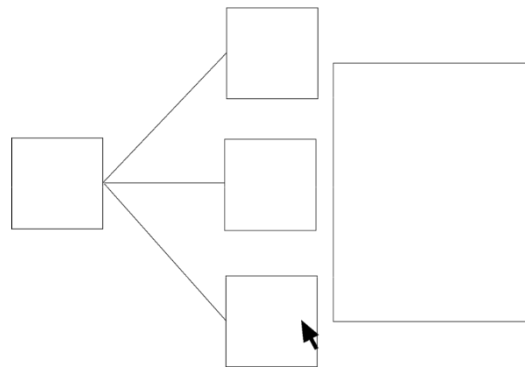
Técnica de interação e distorção: **Detalhe a pedido**

### Descrição

Este esquema é semelhante ao anterior e pode ser utilizado separadamente ou em simultâneo com ele. É ainda em tudo análogo ao apresentado na secção 4.4.6, contudo aplicável a um cenário diferente.

As miniaturas utilizadas para representar as páginas em muitos esquemas de visualização são demasiadamente pequenas para se conseguir mais que uma ideia grosseira da distribuição do seu conteúdo. Uma forma de permitir observar alguns detalhes, de forma rápida e sem perder o contexto, pode passar pela utilização de uma janela com uma representação ampliada da miniatura. Esta janela, tal como no caso anterior, aparecerá junto do cursor do rato e poderá ser opaca ou semitransparente – o que pode ser de grande utilidade, já que pode pretender-se que esta representação ampliada tenha um tamanho considerável.

A Figura III-27 ilustra, de forma esquemática, o método descrito.



**Figura III-27:** Exemplo de visualização de uma ampliação da página a que se refere a miniatura sob o cursor do rato. A ampliação aparece como uma pequena janela (*tool-tip window*), opaca ou semitransparente, junto do cursor do rato.

## 4.5.8. Informação sobre uma ligação

### Classificação taxonómica

Tipo de dados: **2D**

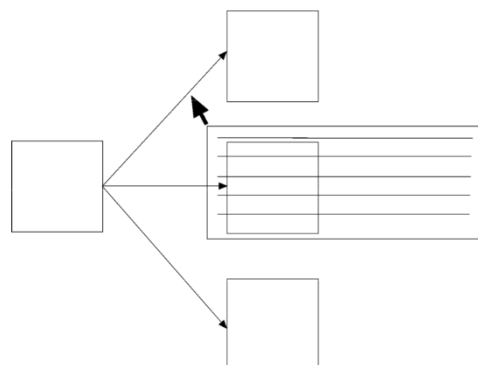
Técnica de visualização: **Normal 2D**

Técnica de interação e distorção: **Detalhe a pedido**

### Descrição

Este é um esquema de visualização exatamente igual ao apresentado na secção 4.5.6, aplicado a uma ligação entre duas páginas em vez de uma. Pode ser apresentada informação como o *URL* das páginas origem e destino da ligação ou ainda informação estatística, tal como o número de utilizadores que percorreram esta ligação, etc..

A Figura III-28 apresenta um esquemático ilustrativo do método apresentado.



**Figura III-28:** Exemplo de visualização dos detalhes referentes à ligação entre duas páginas sob o cursor do rato. Os detalhes aparecem como uma pequena janela de texto (*tool-tip window*), opaca ou semitransparente, junto ao cursor do rato.



## 4.5.9. Visualização / enaltecimento de informação relacional de uma ligação ou nodo

### Classificação taxonómica

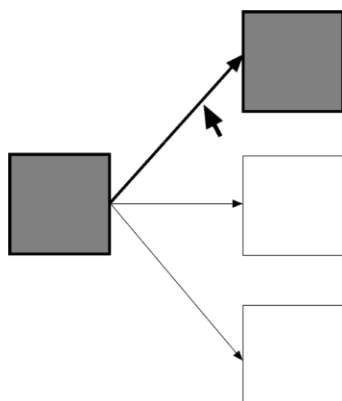
Tipo de dados: **2D**

Técnica de visualização: **Normal 2D**

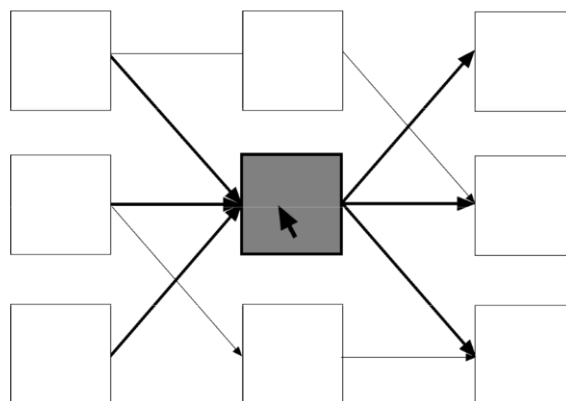
Técnica de interação e distorção: **Ligação e relação**

### Descrição

Este esquema de visualização permite observar com maior facilidade todos os objetos relacionados com o objeto (nodo) selecionado. No esquemático ilustrado na Figura III-29 podem observar-se as páginas relacionadas com a ligação sob o cursor. De igual modo, seria possível observar todas as ligações relacionadas com uma dada página (Figura III-30). A identificação dos objetos relacionados consegue-se enaltecendo-os, usando por exemplo um traço mais grosso e/ou mudando a cor desses objetos.



**Figura III-29:** Exemplo de visualização das páginas relacionadas com uma ligação. Tanto a ligação, sob o rato, como as páginas afetadas aparecem enaltecidas. O efeito pode ser conseguido usando por exemplo, um traço mais grosso e/ou mudando a cor dos objetos em causa.



**Figura III-30:** Exemplo de visualização das páginas relacionadas com uma página (nodo). Tanto a página, sob o rato, como as ligações afetadas aparecem enaltecidas. O efeito pode ser conseguido usando, por exemplo, um traço mais grosso e/ou mudando a cor dos objetos em causa.

## 4.5.10. Visualização dos trajetos possíveis

### Classificação taxonômica

Tipo de dados: **2D**

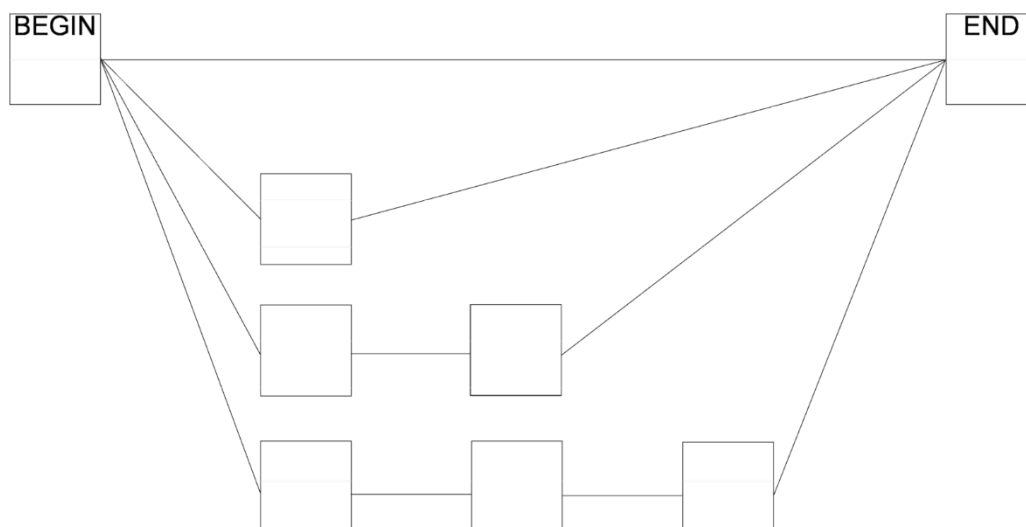
Técnica de visualização: **Normal 2D**

Técnica de interação e distorção: **Vista geral**

### Descrição

Quando se estuda a forma como os utilizadores usam um *site*, é frequente observar-se o percurso que estes utilizam para irem de uma dada página para outra. Normalmente, um *site* é constituído por um conjunto de páginas que têm *links* entre si. É comum encontrar muitas páginas com múltiplos *links*, o que pode proporcionar vários percursos alternativos entre duas páginas. Este método de visualização surge com o objetivo de ajudar a identificar estes percursos alternativos. De notar que os percursos não têm necessariamente o mesmo tamanho, podendo acontecer a existência de um percurso direto (um único clique) simultaneamente com outros mais longos (e portanto com um maior número de cliques e páginas visitadas).

Na Figura III-31 pode observar-se um esquemático ilustrativo deste método, onde se pode identificar quatro percursos possíveis para ir da página “Begin” para a página “End”. No caso apresentado, os vários percursos têm tamanhos diferentes, mas pode acontecer existirem vários percursos com a mesma extensão (número de cliques e páginas intermédias percorridas), contudo com níveis de eficiência diferenciados (representados pelo tempo médio utilizado para percorrer cada percurso).



**Figura III-31:** Exemplo de visualização dos percursos possíveis entre duas páginas selecionadas. No caso ilustrado, existem quatro percursos possíveis com um número díspar de cliques.

## 4.5.11. Visualização da taxa de utilização de cada página de um *site*

### Classificação taxonómica

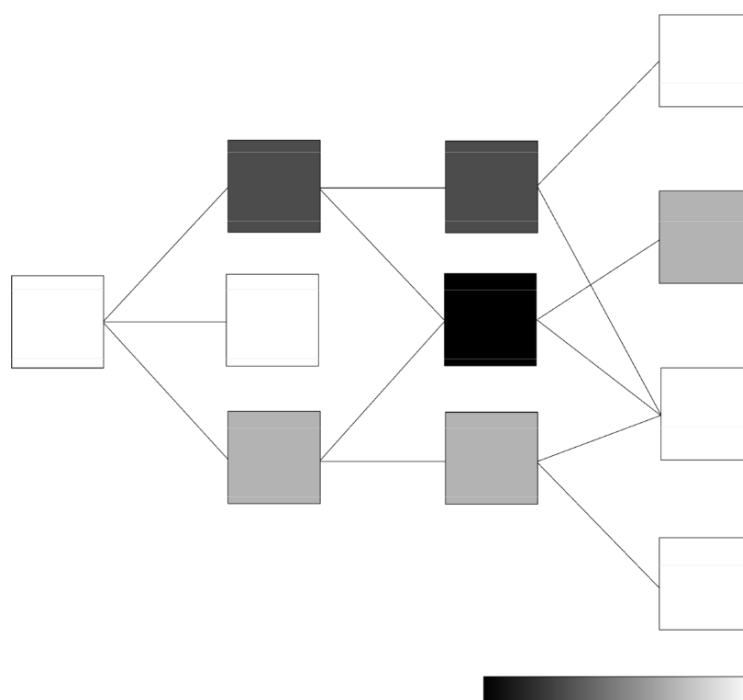
Tipo de dados: **1D**

Técnica de visualização: **Normal 2D**

Técnica de interação e distorção: **Vista Geral**

### Descrição

Quando se analisa um *site*, há informação pertinente que não se fica pela observação da estrutura desse *site*. A informação respeitante à utilização do mesmo pode dar indicações importantes quanto à forma como deve ser melhorado, por exemplo. A informação de utilização do *site* permite identificar as páginas mais visitadas e um esquema como o ilustrado pela Figura III-32 pode ser um instrumento de grande valia. Este método de visualização permite observar quais as páginas mais visitadas, o seu enquadramento na estrutura do *site* e a relação entre si. Para o conseguir, é utilizada uma tabela de cor para codificar o valor correspondente ao número de visitas de cada página.



**Figura III-32:** Exemplo de visualização da taxa de utilização de cada página de um *site*, bem como das suas respetivas ligações. O número de visitas de cada página e/ou ligação está codificada com recurso a uma tabela de cor.

## 4.5.12. Visualização das páginas de uma sessão de um utilizador

### Classificação taxonómica

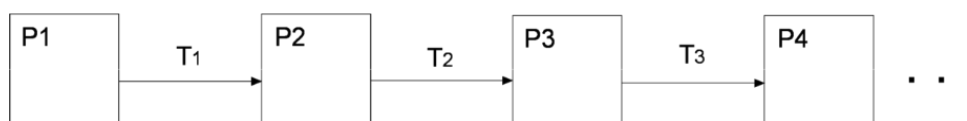
Tipo de dados: **Temporal**

Técnica de visualização: **Hierárquica**

Técnica de interação e distorção: **Vista geral**

### Descrição

Este método de visualização permite observar as páginas visitadas por um utilizador durante uma sessão de utilização de um *site*. As páginas são representadas por ícones de páginas colocadas ao longo de uma linha, respeitando a ordem por que foram visitadas, sendo também indicado o tempo decorrido entre cada página. O tempo decorrido entre cada página, além de poder ser indicado explicitamente junto das ligação entre páginas, pode ser codificado pela distância entre a representação de cada página. A Figura III-33 ilustra, esquematicamente, uma visualização do método descrito.



**Figura III-33:** Exemplo de visualização de uma sessão de um utilizador, identificando cada página visitada e o tempo decorrido entre cada página utilizados pelo utilizador.

## 4.5.13. Visualização em 3D das páginas de uma sessão de um utilizador

### Classificação taxonómica

Tipo de dados: **Temporal**

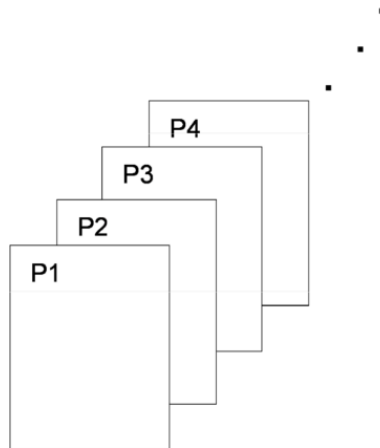
Técnica de visualização: **Normal 3D**

Técnica de interação e distorção: **Vista geral**

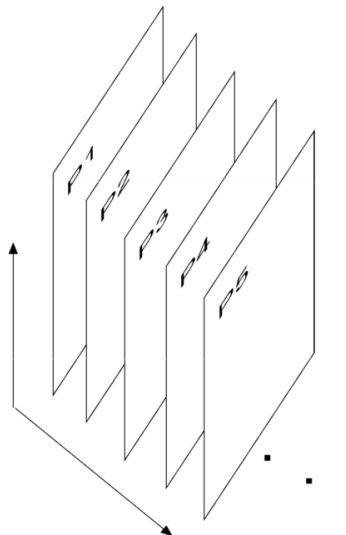
### Descrição

Este método de visualização permite observar as páginas visitadas por um utilizador durante uma sessão de utilização de um *site*. As páginas são representadas por ícones

de páginas empilhadas num espaço 3D, permitindo perceber a ordem por que foram visitadas. O tempo decorrido entre cada página pode ser codificado pela distância entre a representação de cada página. Cada página pode ser representada numa folha opaca ou pode utilizar-se transparência, sendo assim possível vislumbrar a parte oculta das páginas. As Figura III-34, Figura III-35 e Figura III-36 ilustram, esquematicamente, algumas possibilidades do método descrito.



**Figura III-34:** Exemplo de visualização em 3D das páginas de um *site web*, correspondentes a uma visita de um utilizador.



**Figura III-35:** Exemplo de visualização em 3D das páginas de um *site web*, correspondentes a uma visita de um utilizador.

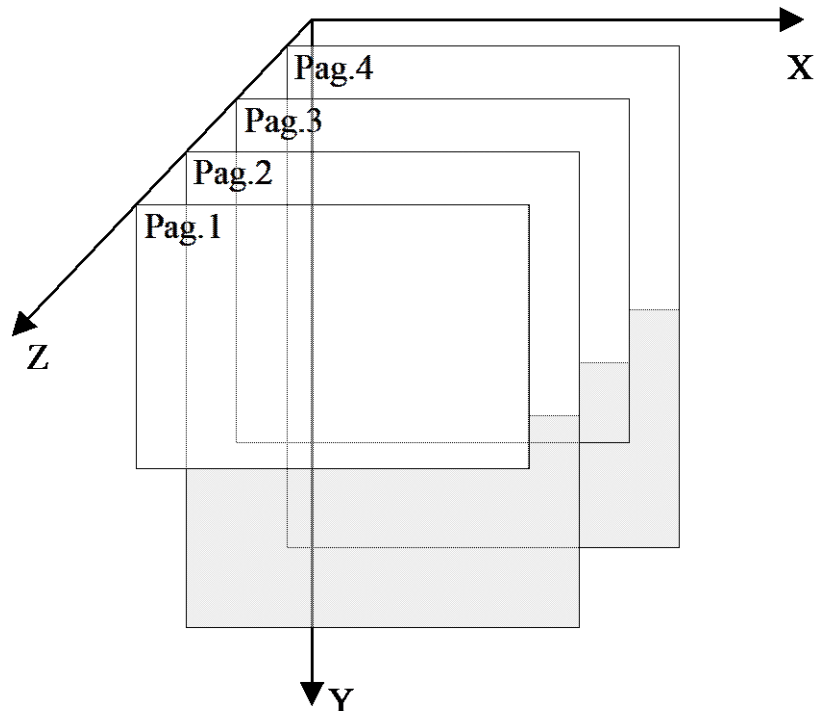


Figura III-36: Exemplo de visualização em 3D das páginas de um *site web*, correspondentes a uma visita de um utilizador.

#### 4.5.14. Visualização das páginas de uma sessão de um utilizador (por níveis)

##### Classificação taxonómica

Tipo de dados: **Temporal**

Técnica de visualização: **Hierárquica**

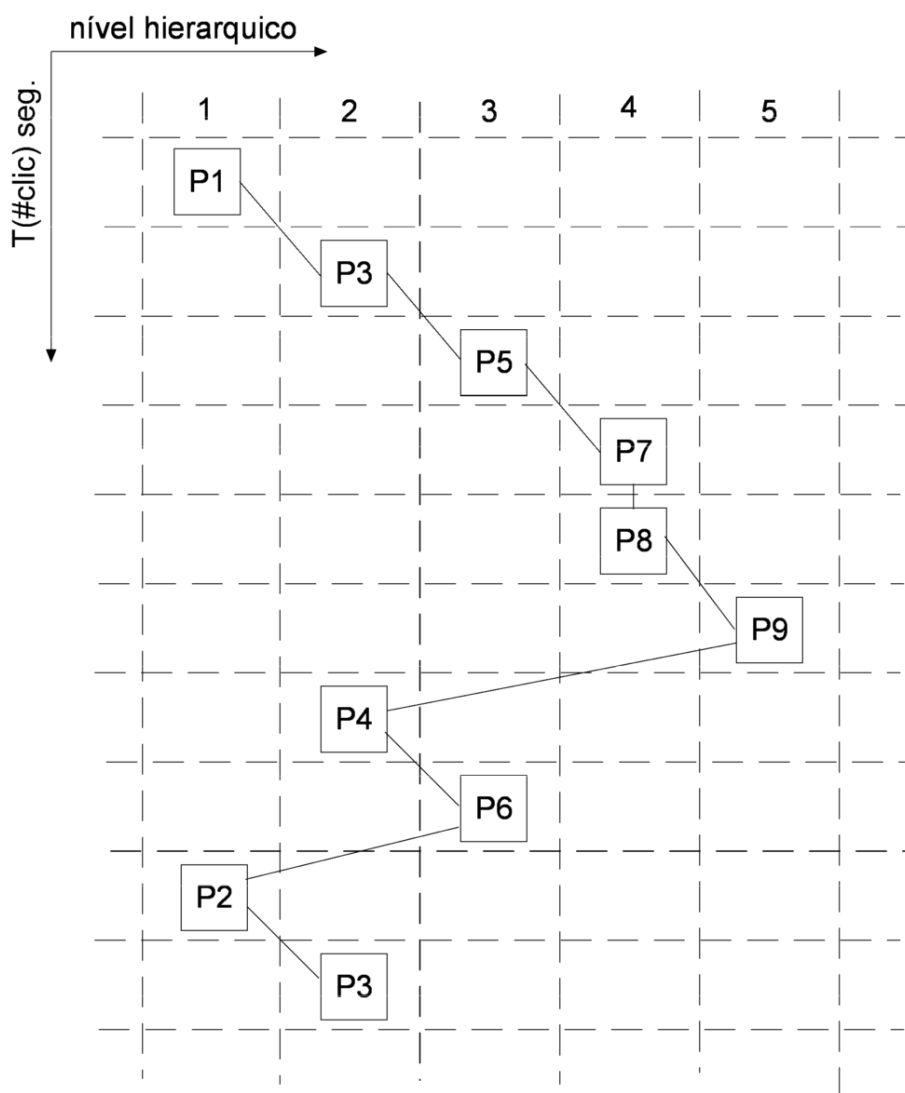
Técnica de interação e distorção: **Vista geral**

##### Descrição

Este método de visualização conjuga a informação e as propriedades do método descrito na secção 4.5.12 (Visualização das páginas de uma sessão de um utilizador) com as dos métodos descritos nas secções 4.5.2, 4.5.3 e 4.5.4 (Visualização da estrutura de um site – nas suas 3 versões). Pretende-se visualizar a atividade de um utilizador durante uma sessão de navegação de um *site*, tendo presente simultaneamente qual a página visitada num dado instante, com o seu conteúdo, bem como com a sua posição relativa na estrutura do *site* (nível). Desta maneira será possível detetar saltos atípicos dentro do *site*, o que poderá ajudar a identificar problemas na estrutura do *site*.

O nível hierárquico de cada página poderá ser atribuído de forma automática seguindo os links existentes em cada página. Um exemplo: considerando uma página P1 de nível 1, todas as páginas correspondentes aos *links* que esta contenha seriam de nível 2, e assim sucessivamente. O nível também poderia ser atribuído tendo em consideração a estrutura conceptual do *site* usada pelo seu criador.

A Figura III-37 ilustra de forma esquemática este método. No caso apresentado, um utilizador começa com numa página de nível 1, seguindo um *link* que o leva para uma página de nível 2 e assim sucessivamente, indiciando que o utilizador se está a deslocar de forma consistente para o seu objetivo. Já o salto existente entre a página P9 e a P4 pode indicar que o utilizador atingiu o seu objetivo e iniciou outra tarefa a partir do ponto onde terminou a anterior. Contudo, também pode significar que andou perdido e que teve necessidade de voltar para trás para tentar outro caminho.



**Figura III-37:** Exemplo de representação de uma sessão de um utilizador, identificando cada página visitada pelo utilizador e respetivo nível.

## 4.5.15. Identificação das ligações entre um conjunto de páginas

### Classificação taxonómica

Tipo de dados: **Rede ou Malha**

Técnica de visualização: **Normal 2D**

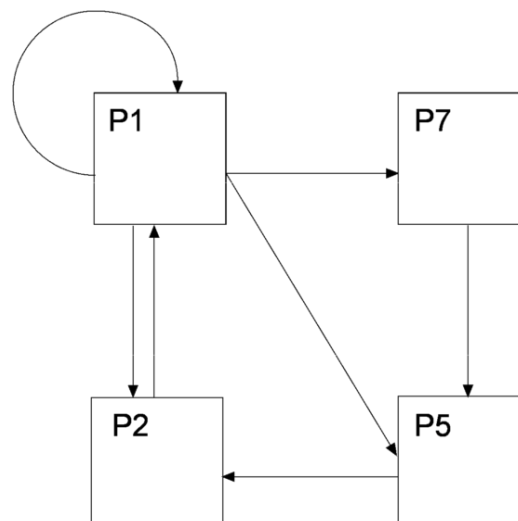
Técnica de interação e distorção: **Vista geral**

### Descrição

Este método de visualização permite identificar todas as ligações existentes entre um conjunto de páginas especificado. Há situações em que este tipo de estudo pode ser de grande utilidade para estudar a forma de relacionamento entre páginas. As páginas selecionadas podem ser parte das páginas utilizadas por um utilizador durante uma sessão de utilização de um *site* ou simplesmente um conjunto de páginas selecionadas pelo operador por algum motivo.

Neste método, são representadas miniaturas de cada página e as ligações entre elas, indicando o sentido das mesmas ligações. De notar que pode ser possível uma página ter uma ligação para si própria. Esta situação é comum quando um conjunto de páginas é produzido recorrendo a um modelo que contém um conjunto de *links* permanente.

A Figura III-38 ilustra de forma esquemática este método, com um conjunto de páginas e suas ligações.



**Figura III-38:** Exemplo de identificação das ligações existentes entre um conjunto de páginas previamente selecionadas.



## 5. Integração e correlação visual de métodos de visualização

Um método de visualização pode permitir a observação de um processo, fenómeno ou informação. Contudo, há muitas situações em que, apesar de se dispor de vários métodos de visualização, há aspetos que são de difícil observação ou apreensão. Como forma de colmatar alguns destes problemas podem usar-se vários métodos de visualização em simultâneo, permitindo ter vários pontos de vista / técnicas de observação do mesmo problema num mesmo ecrã.

Será necessário conseguir uma interface unificada que integre os métodos de visualização necessários para implementar a informação e conseguir extrair conhecimento vital à luz dos objetivos propostos de manutenção do sistema de informação e comunicação em causa.

Contudo, para que o processo seja efetivamente eficiente, é necessário que os vários métodos estejam intimamente ligados, para que quando se interage com uma das janelas contendo um dos métodos de visualização, as alterações sejam propagadas de forma sincronizada e imediata a todas as outras janelas.

Para atingir este objetivo é necessário ter em atenção alguns conceitos fundamentais:

1. Existir uma correlação visual dos métodos de visualização e análise de informação. Os vários métodos de visualização utilizados deverão ter uma interface semelhante, partilhando o mesmo paradigma de interação e de apresentação de resultados.
2. Permitir uma sincronização dos vários métodos com base no filtro de sessão(ões) e com base no objetivo de utilização especificado (página inicial do estudo e página final, contém o objetivo de utilização);
3. Garantir a utilização de uma mesma tabela de codificação de cor de um determinado atributo nos vários métodos de visualização. Ao garantir que é utilizado o mesmo esquema de cor para codificar, por exemplo, a taxa de visualização de uma página nos vários métodos de visualização visíveis num dado momento, garante-se uma maior coerência.
4. Dar prioridade à contextualização visual e redução da taxa de esforço cognitivo, através da integração numa só “janela” de múltiplos métodos de visualização estrategicamente sincronizados. Os métodos de visualização visíveis num dado momento deverão estar integrados, em justaposição, como se de uma única janela se tratasse, para que estejam todos permanentemente visíveis. Devem ainda estar perfeitamente sincronizados, de maneira a que, ao tomar alguma ação

relativamente a um dos métodos de visualização, tenha as repercussões no resultado dos outros métodos de visualização.

5. Reduzir a necessidade de memorização de curto prazo, permitindo a observação simultânea de múltiplas vistas com as janelas justapostas para que não se ocultem umas às outras. Ao manter todas as vistas permanentemente visíveis em simultâneo, reduz-se a carga da utilização da memória de curto prazo, necessária quando o utilizador é obrigado a trocar de janela, ocultando a atual para poder passar a ver uma nova.
6. Permitir ao utilizador a escolha de quantas e quais janelas pretende, bem como a sua distribuição (seleção do tamanho relativo de cada uma) pelo ecrã.
7. O tamanho e resolução do ecrã, bem como o número de ecrãs disponíveis, são fatores que condicionam a informação que pode ser observada num dado momento. Um utilizador com mais recursos, por exemplo com dois monitores de alta resolução, poderá ter mais janelas abertas e, portanto, ter a possibilidade de observar mais métodos de visualização em simultâneo, além de poder dedicar uma área maior de ecrã para cada método.

Nesta linha, Ben Shneiderman em (Shneiderman, 1998) apresenta um conjunto de princípios a ter em consideração:

- **Scrolling sincronizado** – Uma coordenação simples é o *scrolling* sincronizado, em que a barra de *scroll* de uma janela é ligada à barra de *scroll* de outra janela e a ação numa barra de *scroll* provoca o *scroll* do conteúdo associado à outra janela. Esta técnica é útil para, por exemplo, fazer a comparação de duas versões de um programa ou documento.
- **Navegação hierárquica** – Janelas ligadas podem ser utilizadas para suportar navegação hierárquica. Se uma janela contiver a tabela de conteúdos de um documento, a seleção do título de um capítulo deve conduzir à visualização, numa janela adjacente, do conteúdo do capítulo selecionado. A navegação hierárquica está perfeitamente integrada no *Windows Explorer* para permitir aos utilizadores navegar na estrutura hierárquica de diretórios.
- **Seleção direta** – Outra ideia de ligação é a seleção direta, em que apontar um ícone, uma palavra no texto ou o nome de uma variável num programa faz aparecer uma pequena janela com os detalhes do ícone, com a definição da palavra ou com a declaração da variável. Os balões de ajuda do *Macintosh* e os *tool-tips* do *Windows* são aplicações da seleção direta e os utilizadores devem ser capazes de definir esta coordenação facilmente.
- **Navegação bidimensional** – Este parente bidimensional da navegação hierárquica mostra uma vista geral de um mapa, gráfico ou fotografia numa janela e os detalhes numa segunda janela. Os utilizadores podem mover uma caixa de “ponto de vista” na vista geral para ajustar o conteúdo da “vista dos detalhes”.

- **Abertura de janelas dependentes** – Uma opção ao abrir uma janela pode consistir em abrir também janelas dependentes numa localização próxima e conveniente.
- **Fecho de janelas dependentes** – Uma opção ao fechar uma janela poderá consistir em fechar todas as janelas dependentes. Esta opção pode ser aplicada quando se fecham as janelas de diálogo, de mensagem ou de ajuda com uma única ação.
- **Gravação ou abertura do estado de janela** – Uma extensão natural do processo de gravação de um documento ou de um conjunto de preferências pode ser a gravação do estado corrente do ecrã, com todas as janelas e seu conteúdo. Esta funcionalidade pode ser implementada pela simples adição do item “Gravar ecrã como...” às ações do menu “Ficheiro”. Esta ação poderia criar um novo ícone representando o estado corrente; poderia ser aberto para reproduzir esse estado.

No contexto deste trabalho, é adequada a utilização de um conjunto de vistas simultâneas e sincronizadas, podendo corresponder cada vista a um dos métodos de visualização anteriormente apresentados. A Figura III-39 ilustra um exemplo da integração de quatro métodos de visualização numa única vista / ecrã. No caso ilustrado, a seleção de uma página num dos quadrantes provocará a seleção automática da mesma página nos restantes quadrantes, o que no caso do quadrante superior direito significa a representação da página selecionada.

Qualquer interação num dos quadrantes, com o seu esquema de visualização, além das alterações esperadas dessa interação nesse esquema de visualização, provoca alterações, de forma sincronizada, em todos os outros quadrantes, contendo outros métodos de visualização.

## 6. Algumas conclusões

Este capítulo descreve com algum detalhe o modelo proposto e utiliza para o efeito uma abordagem com diferentes níveis de complexidade. Começa pela proposta conceptual e holística do modelo, onde é possível compreender as características interativas, suportadas nos desígnios e definições do analista (objetivos), mas sem menosprezar a realidade do contexto empírico sobre análise. O capítulo dedica algum detalhe à caracterização dos diversos elementos do modelo e suas funcionalidades. Algumas das funcionalidades, pela sua capacidade de reutilização mais universal, são apresentadas com recurso a uma nomenclatura contextual muito específica, caso das funções e atributos de visualização.

As diferentes técnicas de visualização propostas e descritas adquirem nesta tese um carácter inovador, ainda hoje, de enorme utilidade para representar e explorar informação que resulte de comportamentos de interação (*user-driven analytics*) em sistemas computacionais (plataformas digitais). O capítulo termina com propostas de relação das diversas funções considerando que estas relações devem ser consistentes com os objetivos de análise ativos, devem estar em sincronia no tempo e consistentes com as alterações dos atributos visuais e portanto proporcionar uma correlação visual das diversas funções ou perspectivas de análise num único, e mesmo, campo visual (ecrã).

O contributo deste capítulo traduz-se numa proposta de modelo, seus elementos constituintes, técnicas de visualização e exploração de dados infocomunicacionais relacionados com o comportamento de interação de utilizadores em repositórios organizacionais. Termina com a proposta de estratégias de correlação visual das diferentes técnicas propostas com vista a construir cenários holísticos, síncronos e visuais de análise de interação infocomunicacional.

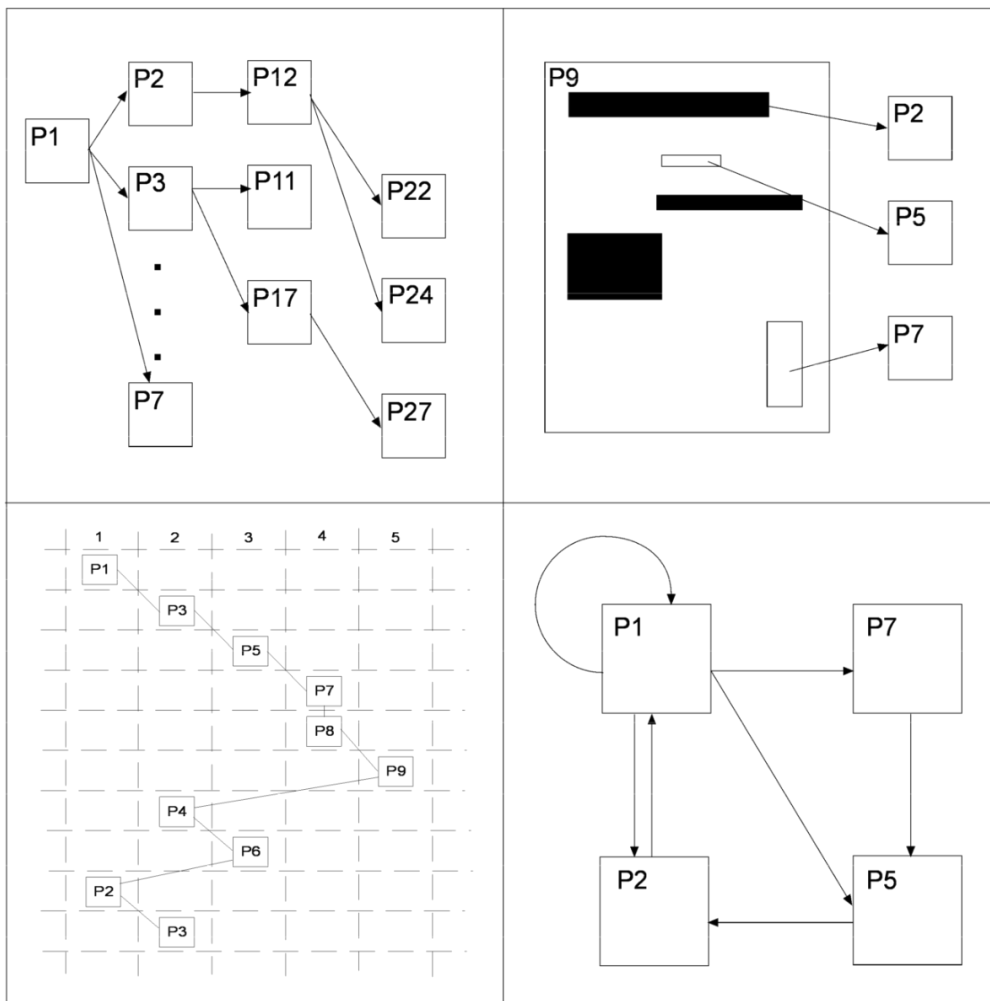


Figura III-39: Exemplo de integração de quatro métodos de visualização.

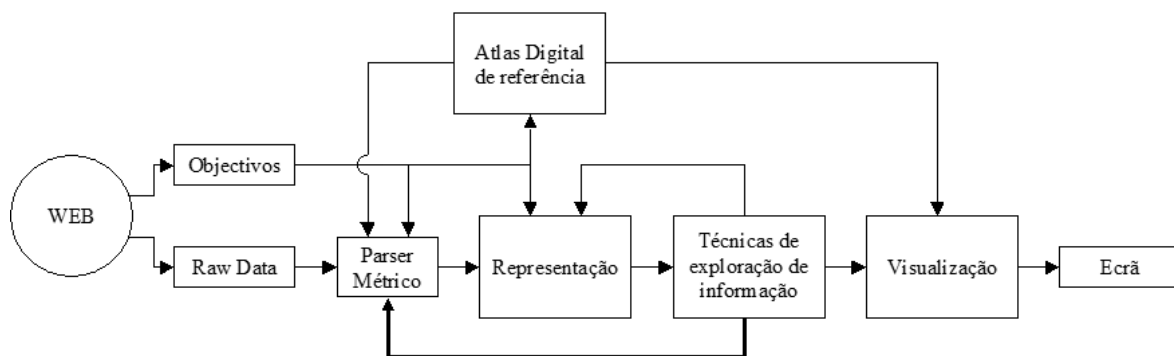
# **Capítulo IV.**

## **PROTÓTIPO**

# 1. Introdução

No capítulo anterior foram propostos e apresentados métodos de visualização e análise de informação. Neste capítulo, estes métodos são integrados num protótipo funcional e aplicados em contexto de repositórios reais. No Capítulo V, abaixo, estes métodos são sujeitos a avaliação empírica, por parte de utilizadores especialistas, durante a avaliação do protótipo.

O modelo conceptual do processo de visualização apresentado no início do capítulo anterior e ilustrado pelo esquema da Figura III-2, repetido abaixo na Figura IV-1, apresenta um conjunto de blocos que representam as várias partes constituintes da aplicação a desenvolver. Contudo, antes de se conseguir legitimar e promover uma aplicação final, convém validar se responde aos objetivos e requisitos colocados à partida.



**Figura IV-1:** Modelo do processo de visualização de um *site web*, usando uma arquitetura de representação dinâmica.

A avaliação da usabilidade de uma aplicação multimédia num contexto *web* é uma tarefa pouco trivial. Esta tarefa é ainda dificultada quando se pretende que seja tão pouco intrusiva para o utilizador quanto possível.

Neste capítulo, e para efeitos de avaliação do protótipo em causa, foi adotado uma metodologia de avaliação híbrida, contemplando avaliação analítica e empírica. Justifica-se esta abordagem por se considerar que as ferramentas analíticas disponíveis têm algumas limitações (Newman & Lamming, 1995) passíveis de serem minimizadas, no impacto da avaliação final, quando complementadas com instrumentos empíricos.

## 2. Tipos de protótipos

Normalmente, os utilizadores finais têm dificuldade em antecipar como irão utilizar novos sistemas de software para suportar o seu trabalho do dia a dia. Se esses sistemas forem grandes e complexos, é provável que seja impossível fazer essa avaliação antes do sistema estar construído e colocado a uso.

Uma maneira de lidar com esta dificuldade consiste na utilização de uma abordagem “evolutiva” do desenvolvimento do sistema. Isto significa dar ao utilizador um sistema que está incompleto e ir modificando e aumentando-o à medida que as necessidades do utilizador forem ficando mais claras. Alternativamente, pode decidir-se construir um protótipo “descartável” (*throw-away*) para ajudar na análise e validação da especificação (Sommerville, 2001).

Deborah Mayhew em (Mayhew, 1999) refere que pedir aos utilizadores para avaliar as ideias de alto nível iniciais ajuda a evoluir, no processo altamente interativo de perceber e resolver as necessidades do utilizador. Um processo iterativo, entre designers e utilizadores é a melhor estratégia para a análise das necessidades. Refere ainda que na maior parte das situações, um protótipo de “baixa-fidelidade” (*low-fidelity*) pode ser tão eficiente com um de “alta-fidelidade” (*high-fidelity*) e é mais barato e rápido de produzir.

Jenny Preece em (Preece et al., 1994; Preece et al., 2011) classifica os protótipos em quatro categorias:

1. Animação de requisitos – Permite que os requisitos possíveis (normalmente funcionais) sejam demonstrados num protótipo e possam ser avaliados pelo utilizador.
2. Protótipo rápido descartável – Usado para recolher informação sobre os requisitos e adequação dos designs possíveis. Este tipo de protótipo é deitado fora depois de utilizado no sentido em que não evolui para o produto final, ainda que seja um instrumento importante durante o desenvolvimento do projeto.
3. Protótipo evolutivo – Compromisso entre produção e protótipo. O sistema pode lidar com alterações durante e depois do desenvolvimento. Ajuda a colmatar a falha tradicional entre a especificação e a implementação.
4. Protótipo incremental – O sistema é construído de forma incremental, uma secção de cada vez. Este tipo de protótipo é usado sobretudo em sistemas de grandes dimensões em que a instalação se processa em fases para evitar atrasos entre especificação e entrega do produto.

Esta autora refere ainda a existência de outros tipos de protótipos na literatura de HCI<sup>79</sup>. Um **protótipo completo**, como o nome indica, contém a funcionalidade completa, ainda que com uma *performance* inferior. Um **protótipo horizontal** mostra a interface de utilizador mas não tem funcionalidade associada. Um **protótipo vertical** contém toda a funcionalidade de alto e baixo nível de uma parte restrita do sistema. **Protótipo de alta-fidelidade** refere-se à utilização de um meio que apresente/represente um conjunto completo das funcionalidades finais e que apresente, tão próximo quanto possível, a interface final. **Protótipos de baixa-fidelidade** envolvem o uso de materiais que estão longe da versão final e cujo desenvolvimento tende a ser mais barato e rápido. **Prototipagem conduzida** (*chouffeed*) envolve o utilizador a observar enquanto outra pessoa utiliza o sistema. Esta é uma maneira de testar se a interface satisfaz as necessidades do utilizador sem que este tenha que levar a cabo ações de baixo nível com o sistema. O protótipo tipo **mágico de Oz** (*wizard of Oz*) também envolve uma terceira pessoa, ainda que o utilizador não saiba disso. O utilizador interage com um sistema, mas em vez de ser um software a responder diretamente aos pedidos do utilizador, está outra pessoa a introduzir as respostas.

Neste trabalho vai usar-se a abordagem de (Sommerville, 2001). Seguidamente vai passar-se a descrever os dois tipos de protótipos: evolutivos e descartáveis.

## 2.1. Protótipo evolutivo

A prototipagem evolutiva é baseada na ideia de desenvolver uma implementação inicial, expondo-a aos comentários do utilizador e refinando-a iterativamente até que um sistema adequado esteja desenvolvido.

Existem duas grandes vantagens em adotar esta abordagem no desenvolvimento de software:

1. Entrega do sistema acelerada. Em alguns casos, a entrega rápida e a usabilidade são mais importantes que os detalhes da funcionalidade ou a manutenção a longo prazo.
2. Envolvimento do utilizador com o sistema. O envolvimento dos utilizadores com o processo de desenvolvimento implica que o sistema tenha uma maior probabilidade de satisfazer as suas necessidades porque os utilizadores estão comprometidos com as soluções implementadas.

Existem também alguns problemas associados à prototipagem evolutiva, nomeadamente quando se pretende desenvolver sistemas complexos e com um longo ciclo de vida:

---

<sup>79</sup> HCI: Interação humano-computador (Human Computer Interaction)



1. Problemas de gestão – Os protótipos evoluem tão rapidamente que pode não ser eficiente, do ponto de vista financeiro, produzir documentação suficientemente detalhada. Ainda mais grave, o desenvolvimento rápido de protótipos pode implicar o uso de técnicas pouco familiares.
2. Problemas de manutenção – A alteração contínua tende a corromper a estrutura do protótipo. Isto quer dizer que alguém fora da equipa de desenvolvimento original terá, provavelmente, grandes dificuldades em entendê-la. Acresce que se tiver sido utilizada tecnologia específica para prototipagem rápida, esta possa ter ficado obsoleta. Como consequência, pode ser difícil encontrar pessoas com os conhecimentos necessários para manter o sistema.
3. Problemas contratuais – O modelo contratual mais habitual entre um cliente e alguém que desenvolve software costuma ser baseado nas especificações do sistema. O cliente pode não ficar satisfeito com um contrato em que simplesmente paga pelo tempo despendido para o desenvolvimento. Por outro lado, quem desenvolve o software dificilmente aceitará um contrato com valor fixo uma vez que não pode controlar as alterações requeridas pelo utilizador final.

Uma forma de limitar os problemas do desenvolvimento evolutivo poderá ser o desenvolvimento incremental. É estabelecida uma arquitetura global do sistema logo no início do processo e que irá funcionar como *framework* (plataforma de trabalho). Os componentes do sistema são desenvolvidos de forma incremental e colocados no *framework*. Logo que o componente esteja validado e integrado, nem o *framework* nem os componentes são alterados a menos que sejam detetados erros. O feedback dos utilizadores aos componentes integrados pode, contudo, influenciar o desenvolvimento de componentes agendados para mais tarde (Sommerville, 2001).

## 2.2. Protótipo descartável

Esta abordagem estende o processo de análise de especificações com a intenção de reduzir os custos globais do ciclo de vida do produto. A principal função deste tipo de protótipos é clarificar as especificações e fornecer informação adicional para os gestores avaliarem os riscos do processo.

Esta abordagem à prototipagem de sistemas é usada frequentemente em sistemas de hardware. O protótipo é usado para verificar o design antes de se fazerem investimentos avultados associados ao fabrico do sistema.

Os protótipos descartáveis não são usados com muita frequência no design de sistemas de software, contudo podem ajudar a desenvolver as especificações do sistema. O design do protótipo é frequentemente muito diferente do sistema final. Pode ser removida funcionalidade do protótipo, os padrões de qualidade podem ser relaxados e os critérios de *performance* podem ser ignorados. A linguagem de programação usada no

desenvolvimento do protótipo também é, muitas vezes, diferente da utilizada no sistema final.

Muitas vezes, o protótipo é desenvolvido a partir do esboço das especificações, produzido para ser experimentado e modificado até que o cliente esteja satisfeito com a sua funcionalidade. Neste ponto, é iniciado um processo de desenvolvimento faseado, as especificações são obtidas do protótipo e o sistema é reimplementado na sua versão de produção final. Algumas componentes do protótipo podem ser reutilizadas no sistema final para que os custos de produção possam ser reduzidos.

Em vez de obter as especificações a partir do protótipo, é sugerido por vezes que as especificações do sistema devem ser a própria implementação do protótipo. As instruções de quem contrata o software podem ser simplesmente “escreva um sistema como este”. Contudo existem vários problemas nesta abordagem:

1. Características importantes podem ser deixadas de fora do protótipo para simplificar a rápida implementação. De facto, pode não ser possível prototipar alguma das partes mais importantes do sistema como funções de segurança críticas.
2. Uma implementação (prototipada) não tem suporte legal sob a forma de contrato entre cliente e contratado.
3. Requisitos não funcionais, como os relacionados com a fiabilidade, robustez e segurança não podem ser testados adequadamente num protótipo.

Um problema geral no desenvolvimento de protótipos descartáveis está no facto de o uso dado ao protótipo poder não corresponder ao do sistema final.

Os programadores são, por vezes, pressionados pelos gestores para produzirem protótipos descartáveis e mais tarde a usá-los na versão final, especialmente quando há atrasos no desenvolvimento da versão final. Contudo, isto não é sensato pelas seguintes razões (Sommerville, 2001):

1. Pode ser impossível afinar o protótipo de forma a satisfazer os requisitos não funcionais como *performance*, segurança, robustez e estabilidade que foram ignorados durante o desenvolvimento do protótipo.
2. Alterações rápidas durante o desenvolvimento significa, inevitavelmente, que o protótipo não é documentado. A única especificação do design é o código do protótipo. Isto não é suficiente para a manutenção a longo prazo.
3. As alterações feitas durante o desenvolvimento do protótipo irão, provavelmente, degradar a estrutura do sistema. O sistema será difícil e caro de manter.
4. Os *standards* de qualidade são normalmente relaxados durante o desenvolvimento de um protótipo.

Os protótipos descartáveis, que muitas vezes são também protótipos de baixa-fidelidade, não necessitam de ser software executável para serem úteis no processo de

especificação de requisitos. Esboços (*mock-ups*) em papel da interface de utilizador do sistema têm-se mostrado eficientes a ajudar programadores, muitas vezes em colaboração com os utilizadores, a refinar um design de interface. O seu desenvolvimento é muito barato e pode ser construído em pouco tempo.

## 3. Trabalho prévio exploratório

Antes de iniciar o desenvolvimento de uma aplicação é necessário escolher uma plataforma de desenvolvimento, ou seja, é necessário escolher uma linguagem de programação e respetivo ambiente de desenvolvimento. Estes princípios também se aplicam ao desenvolvimento de um protótipo ainda que a escolha possa ser diferente, o que acontece na maior parte dos casos.

### 3.1. Plataformas de prototipagem rápida em visualização

O mercado oferece algumas plataformas de prototipagem rápida de sistemas, nomeadamente de sistemas de visualização. A escolha depende do tipo de aplicação que se pretende prototipar bem como do tipo de protótipo pretendido.

Um dos métodos mais rápidos de construção de um protótipo, por exemplo dum sistema de visualização, consiste na utilização de pacotes de software “construtores de aplicações” (*application builders*), como é o caso do Iris Explorer (NAG)<sup>80</sup>. Este tipo de pacotes oferece uma série de módulos que comunicam entre si por meio de interfaces que são ligadas de forma interativa em *runtime*. Combinam propriedades de duas filosofias, fornecendo soluções prontas a usar (*turnkey*) em partes individuais do programa e a flexibilidade para adaptar a solução adotada. Os módulos fornecidos podem ser substituídos por outros construídos pelo utilizador à medida da necessidade, desde que estes mantenham a conformidade com as especificações da interface entre módulos, dando, no entanto, uma grande extensibilidade.

Nestes sistemas, virtualmente tudo o que o utilizador necessita é fornecido pelo programa. O utilizador só tem que direcionar o caminho de execução do programa, fornecer os dados e opcionalmente fornecer algum módulo construído por si se

---

<sup>80</sup> Iris Explorer: Criado pela Silicon Graphics, foi posteriormente adquirido pela NAG que, entretanto, descontinuou o produto [http://ftp.nag.co.uk/welcome\\_iec.asp](http://ftp.nag.co.uk/welcome_iec.asp).

necessário. A construção de novos módulos é auxiliada por um conjunto de ferramentas que facilitam o seu desenvolvimento e que, em geral, acompanham o pacote.

A aplicação é construída, na maioria dos casos, por intermédio de uma interface baseada no rato, manipulando ícones no ecrã e ligando-os com canais de dados e/ou por intermédio de *scripts* que podem ser editados com um editor de texto. Assim que os módulos necessários estiverem construídos e ligados, o programa pode ser executado. Protótipos de novas aplicações podem ser construídos muito rapidamente ligando, por exemplo, os módulos de forma diferente, no entanto o utilizador necessita de saber como manipular o fluxo de dados através da rede de ligações e pode necessitar de construir novos módulos para estender o conjunto de módulos disponíveis.

Exemplos deste tipo de pacotes são: Amira (FEI)<sup>81</sup>, AVS (Avs Inc.)<sup>82</sup>, D3<sup>83</sup>, OpenDX (IBM)<sup>84</sup>, PV-WAVE<sup>85</sup>, ou VTK<sup>86</sup>.

Alguns pacotes foram marcantes, mas foram descontinuados. Alguns exemplos: apE (Ohio Supercomputer Centre)<sup>87</sup>, Iris Explorer (NAG)<sup>88</sup>, Khoros (University of New Mexico)<sup>89</sup>, ou ManyEyes (IBM)<sup>90</sup>.

---

<sup>81</sup> Amira: <https://www.fei.com/software/amira-3d-for-life-sciences/>.

<sup>82</sup> AVS (*Application Visualization System*): O AVS foi criado pela empresa Stardent que desapareceu em meados da década de 1990. Atualmente é detido pela empresa Avs Inc. <http://www.avs.com/>.

<sup>83</sup> D3.js ou só D3 (*Data-Driven Documents*) é uma biblioteca JavaScript que permite produzir visualizações de dados de forma dinâmica e interativa num *browser web*. <https://d3js.org/>.

<sup>84</sup> OpenDX (*Open Data Explorer*): O OpenDX é um software de visualização científica da IBM e começou em 1991 como *Visualization Data Explorer*. <http://www.opendx.org/>.

<sup>85</sup> PV-WAVE (*Precision Visuals - Workstation Analysis and Visualization Environment*): <http://www.roguewave.com/products-services/pv-wave>. O PV-WAVE foi originalmente desenvolvido por uma empresa chamada Precision Visuals. Em 1992, depois de uma fusão com outra empresa, passa a chamar-se Visual Numerics. Em 2009 é comprada pela empresa Rogue Wave Software.

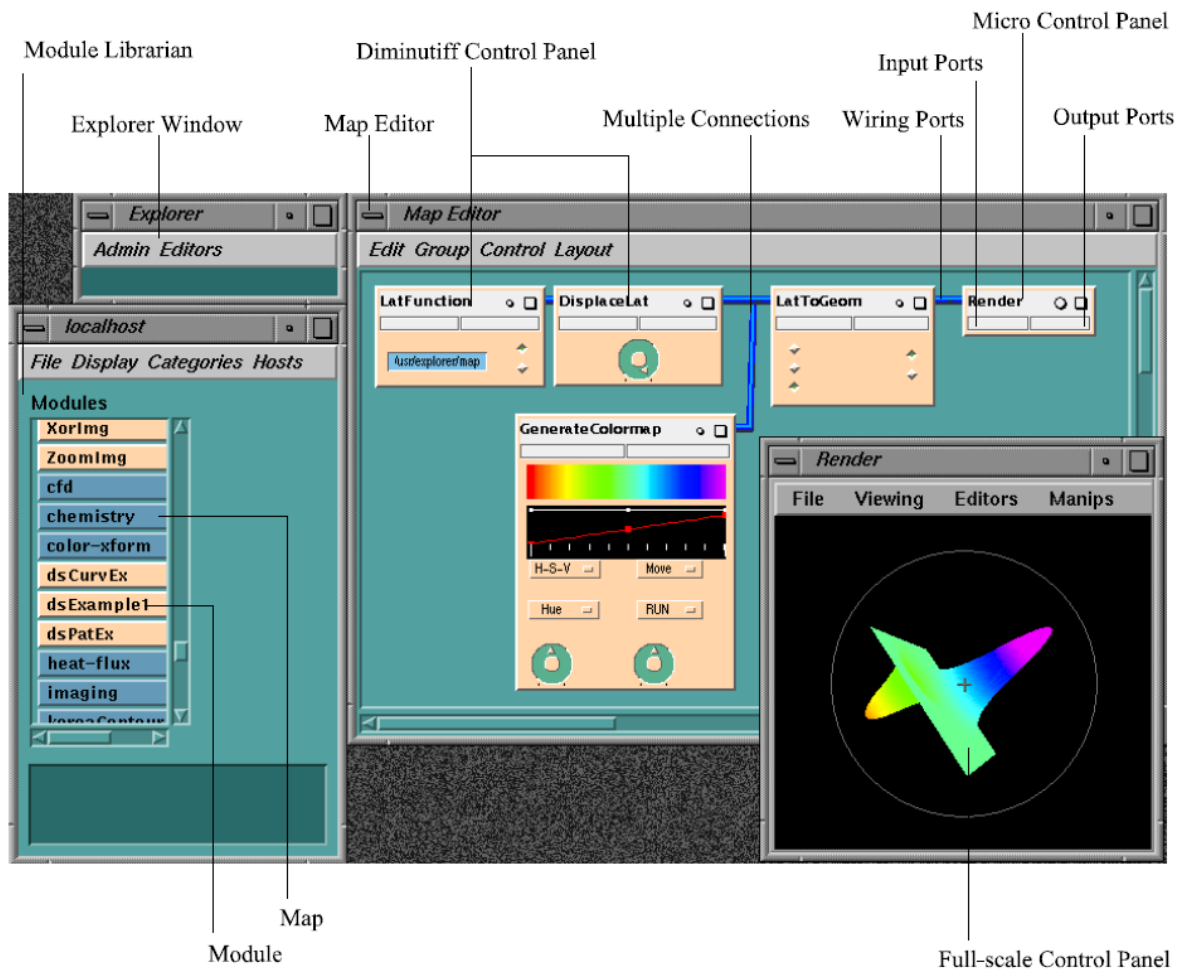
<sup>86</sup> VTK (*Visualization Toolkit*): <http://www.vtk.org/>. O VTK é um software livre, de código aberto, para computação gráfica 3D, processamento de imagem e visualização. Foi criado originalmente em 1993 para acompanhar um livro escrito por 3 investigadores da GE Corporate R&D.

<sup>87</sup> apE, abreviatura para *animation production Environment*, foi desenvolvido no final dos anos 1980 e início dos anos 1990 pelo *Ohio Supercomputer Graphics Project* (OSGP) na Universidade do Ohio (<https://www.ohio.edu/>). Este produto foi, entretanto, descontinuado.

<sup>88</sup> Iris Explorer: Criado pela Silicon Graphics, foi posteriormente adquirido pela NAG que, entretanto, descontinuou o produto [http://ftp.nag.co.uk/welcome\\_iec.asp](http://ftp.nag.co.uk/welcome_iec.asp).

<sup>89</sup> O Khoros nasceu na Universidade do Novo-México, tendo depois passado pela empresa Khoral Research, Inc. (<http://www.khoral.com/>), em 2004 foi vendido à AccuSoft Corporation (<https://www.accusoft.com/>) que, entretanto, descontinua o produto.

Como exemplo do que pode ser encontrado num destes pacotes, a Figura IV-2 apresenta o aspeto geral da interface de utilizador do Iris Explorer com uma pequena descrição de alguns dos componentes constituintes da aplicação.



**Figura IV-2:** Aspeto geral da interface do *Iris Explorer*. Estão representadas várias componentes da aplicação.

Depois de feitas algumas experiências com o Iris Explorer e de estudar o funcionamento de outras plataformas do mesmo tipo, pôde observar-se que este tipo de pacote era:

- Poderoso, flexível e suficientemente genérico para poder ser utilizado nos mais variados ramos de visualização.
- Que podia ser usado com vantagens significativas como plataforma de grande produtividade e que permitia a fácil reutilização de código em aplicações de processamento e visualização de dados.

<sup>90</sup> ManyEyes: Lançado pela IBM em 2010, era um serviço colaborativo de visualização de dados que permitia que qualquer pessoa facilmente fizesse o *upload*, visualizasse e partilhasse os seus dados, tudo baseado num serviço *web*. Foi descontinuado em 2015.

- Permitia testar a viabilidade das soluções de uma maneira mais rápida e fácil do que usando os métodos tradicionais (por exemplo uma biblioteca gráfica).

Tinha, no entanto, limitações inerentes à estratégia de implementação adotada:

- Tratava-se de um pacote com requisitos computacionais muito elevados, só podendo ser utilizado em máquinas com um certo porte, em termos de capacidade de cálculo e memória.
- Não apresentava um desempenho ótimo para todos os tipos de dados, nomeadamente para o tipo de dados utilizados ao longo deste trabalho: onde manifestava alguns problemas de gestão de memória. Estava, no entanto, especialmente bem-adaptado a dados vetoriais (como objetos definidos pelas suas superfícies).
- Não permitia a utilização, de forma fácil, de bibliotecas gráficas em módulos desenvolvidos pelo utilizador, impedindo assim de tirar o melhor proveito do hardware gráfico disponível.
- Não permitia ainda a geração de um ficheiro executável, situação que limitava a utilização de aplicações desenvolvidas sobre esta plataforma: para poder utilizar a aplicação era necessária a aquisição do Explorer, que podia ser demasiado dispendioso e supérfluo para a maioria dos possíveis utilizadores. Esta situação é comum à maioria dos outros pacotes.

Pelas limitações enunciadas, importantes para o trabalho desta tese, e pela impossibilidade de criação de uma aplicação final (ficheiro executável) nesta plataforma, decidiu-se abandonar esta solução.

## 3.2. Linguagens, ambientes ou plataformas de desenvolvimento

Um programador, ao iniciar um novo projeto de desenvolvimento de uma aplicação, tem que tomar um conjunto de opções. As opções dependem do tipo de projeto em causa, das suas necessidades e objetivos, nomeadamente a escolha da linguagem ou linguagens de programação a utilizar. Uma vez como causa, outras como consequência, irá escolher também um ambiente ou plataforma de desenvolvimento.

Está disponível uma diversidade de linguagens de programação bastante grande. Desde linguagens de muito baixo nível, como é o caso do *Assembler*, até às linguagens de mais elevado nível, como C++, C#, Java ou ADA.

As linguagens de mais baixo nível permitem desenvolver aplicações com um nível de otimização mais elevado. Contudo, apresentam alguns problemas quando é necessário lidar com projetos de maiores dimensões.

As linguagens de alto nível apresentam uma estrutura mais perto da linguagem natural, o que aumenta a produtividade. Apresentam ainda um leque cada vez mais alargado de bibliotecas, uma vez que dispõem de mecanismos eficientes de reutilização de código. Contudo, também têm alguns inconvenientes. A reutilização de código, que pode permitir poupar muito tempo e trabalho, tende a produzir código menos eficiente. Por vezes, para se utilizar uma função específica, é necessário incluir toda uma biblioteca, o que pode aumentar o tamanho da aplicação final com resultados negativos na utilização de recursos, no desempenho e estabilidade.

Outro fator condicionante na escolha de uma plataforma de desenvolvimento está nas bibliotecas disponíveis para essa plataforma. Por exemplo, para o presente trabalho era essencial a disponibilidade de uma biblioteca gráfica.

Naturalmente, há outros fatores que condicionam a escolha, por vezes tão ou mais importantes que os enunciados, tais como a familiaridade do programador com a plataforma ou linguagem, ou ainda a existência de algum trabalho prévio.

No início deste projeto considerou-se a possibilidade da utilização de Java. Tinha a vantagem de ser independente da plataforma, ser de alto nível e de já existirem algumas bibliotecas que complementam a funcionalidade de base. Contudo, alguns problemas como os relacionados com o desempenho (são produzidos programas que correm sobre uma máquina virtual) e a impossibilidade de adquirir uma biblioteca gráfica (como o Java 3D) ditaram o abandono desta plataforma.

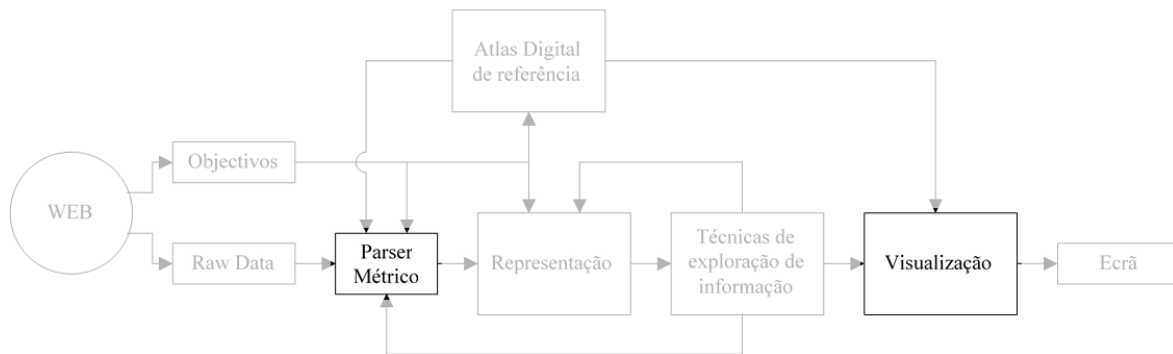
A alternativa seguinte seria a utilização de C / C++. Trata-se de uma linguagem que permite níveis de desempenho e otimização muito bons (note-se que a maioria dos sistemas operativos é hoje escrita com recurso a estas linguagens) e com um leque de bibliotecas disponíveis muito abrangente. Algumas das melhores bibliotecas gráficas disponíveis (como o Open GL e o Open Inventor) estão também disponíveis com *bindings* para C / C++, o que veio tornar óbvia a escolha desta opção.

## 4. Primeira proposta de protótipo – versão 0

Antes de se dedicar mais esforço a este projeto era necessário verificar a viabilidade de algumas ideias e especificar algumas partes do sistema a desenvolver. Como se pretendia, antes de mais, validar ideias a melhor forma encontrada passou pela utilização de alguns protótipos descartáveis. Estes permitiram refinar as especificações iniciais do sistema.

Havia dois aspetos que era importante estudar com algum detalhe. Em primeiro lugar verificar o processo idealizado para a filtragem e extração (*Parser*) de informação relevante dos ficheiros de *log* de um *site web*. Em segundo, aferir a aplicabilidade de alguns métodos de visualização, nomeadamente usando técnicas 3D, à informação extraída.

Estes dois aspetos correspondem a dois importantes blocos identificados no modelo conceptual do processo de visualização apresentado no início do Capítulo III e ilustrado pelo esquema da Figura III-2 que serve de base ao sistema que se pretende desenvolver. A Figura IV-3, abaixo, apresenta esse mesmo modelo com os dois blocos indicados enaltecidos.

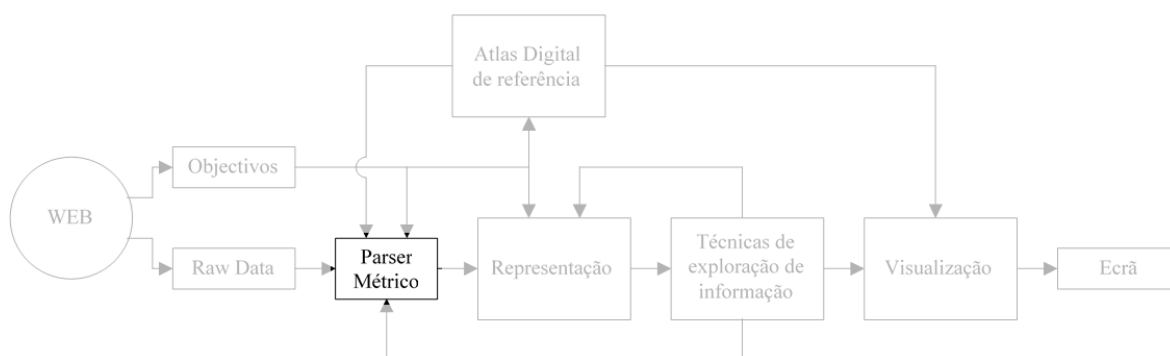


**Figura IV-3:** Modelo do processo de visualização de um *site web*, usando uma arquitetura de representação dinâmica (com os blocos “*Parser Métrico*” e “*Visualização*” enaltecidos).

### 4.1. *Parser Métrico* de ficheiro de *log*

O *Parser Métrico* é um dos blocos mais críticos apresentados no modelo conceptual (Figura IV-4). Como apresentado na secção 2.1 (Elementos constituintes) do Capítulo III, este módulo é o responsável pela filtragem e extração dos elementos relevantes a partir da *raw data* proveniente de, entre outras fontes, dos ficheiros de *log* dos *sites web*.





**Figura IV-4:** Modelo do processo de visualização de um *site web*, usando uma arquitetura de representação dinâmica (bloco *Parser Métrico*).

Neste protótipo foram implementados e testados alguns algoritmos de filtragem de informação que tinham por objetivo reduzir a quantidade de informação a processar.

Um servidor *web* regista nos seus ficheiros de *log* todos os pedidos recebidos. Uma página pode causar múltiplos pedidos. Um pedido para a página HTML e mais um pedido por cada objeto que a acompanha, como por exemplo as folhas de estilo, imagens, *applets* JAVA ou FLASH, etc. Uma forma de simplificar consiste em identificar o pedido referente à página em si e remover todos os referentes aos objetos que dela fazem parte.

Outra tarefa importante consiste na identificação do conjunto de páginas visitadas durante uma sessão de um utilizador. É necessário identificar a página inicial de entrada na qual um dado utilizador inicia a sua sessão, bem como todo o percurso efetuado durante a sua visita. Como resultado, obtém-se uma lista de páginas ordenadas cronologicamente que pode permitir identificar vários detalhes como: páginas visitadas, hora de início e fim da sessão, duração da visita, ou o tempo de visualização de cada página. Em alguns casos, pode ainda permitir a identificação do utilizador e da forma como chegou a este *site* (se foi usado um *link* numa página de outro *site* e qual).

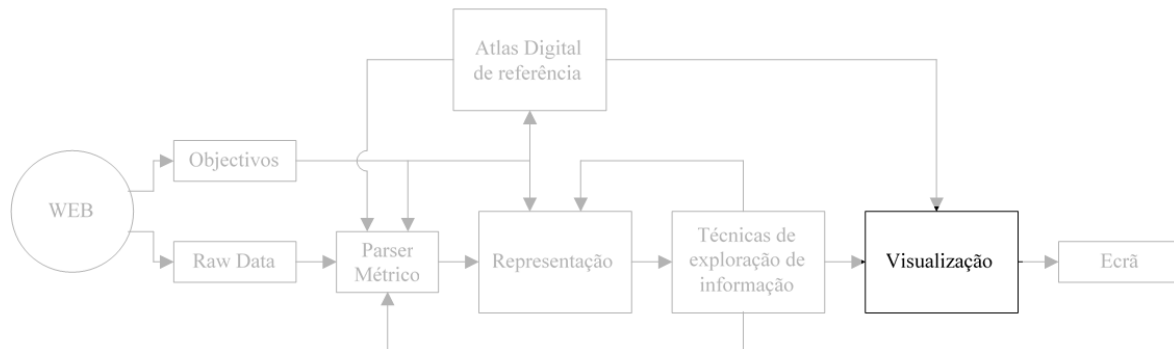
O protótipo foi sofrendo alterações à medida que se foram obtendo resultados e identificando melhorias a incorporar. Uma das últimas funcionalidades introduzidas nesta fase permitia especificar um utilizador e uma janela temporal (limites de início e fim da sessão) para limitar o espaço de pesquisa.

Nesta primeira fase, não chegou a ser desenvolvida uma interface que permitisse a um utilizador interagir com este protótipo. Todos os parâmetros necessários faziam parte integrante do código, sendo necessário recompilar o programa quando se pretendia alterar esses parâmetros. Basicamente, tratou-se de uma simples aplicação de linha de comando que lia um ficheiro de *log* e que apresentava o resultado diretamente no ecrã (*stdout*) em modo de texto.

## 4.2. Visualização

Na sequência do *Parser Métrico*, era necessário ter algum meio de visualizar os resultados obtidos. O modelo conceptual prevê um bloco com exatamente essa função.

Um dos objetivos deste segundo protótipo consistia em testar e validar algumas ideias referentes à visualização, outro bloco do modelo conceptual (Figura IV-5) descrito na secção 2.1 (Elementos constituintes) do Capítulo III.

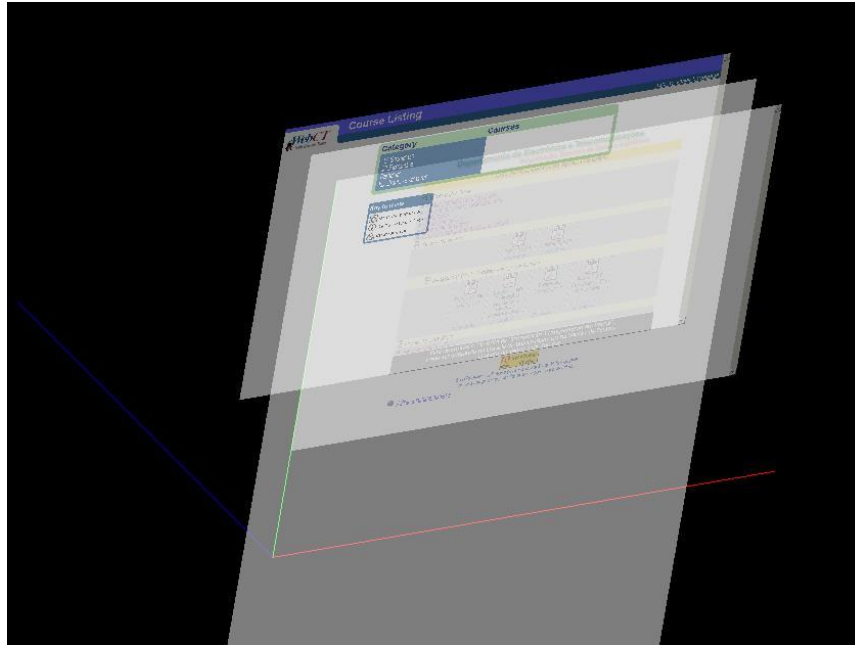


**Figura IV-5:** Modelo do processo de visualização de um *site web*, usando uma arquitetura de representação dinâmica (bloco “Visualização”).

Como referido na secção 6 (Ferramentas de análise de *log's*) do Capítulo II, o mercado oferece algumas aplicações de análise da atividade de um *site web*. Os resultados são apresentados, na maior parte dos casos, sob a forma de tabelas ou gráficos.

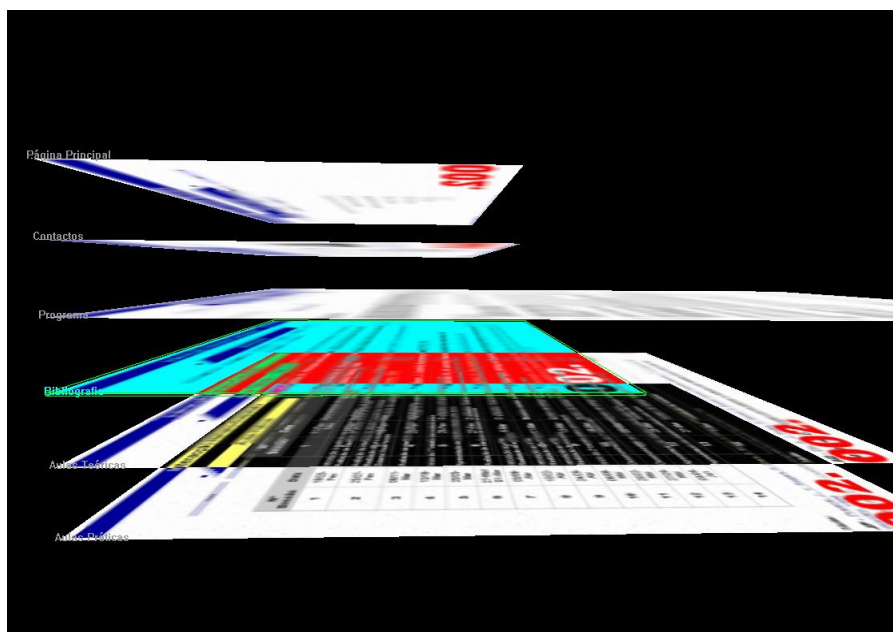
Pretendia-se estudar a viabilidade da aplicação de métodos de visualização mais complexas que os simples gráficos representado um conjunto de valores. Era importante estudar a utilização de técnicas 3D com recurso a transparências para, por exemplo, colocar numa única imagem informação respeitante a uma sessão de um utilizador, dando uma noção do tempo gasto com cada página e com o contexto do conteúdo dessa mesma página. As Figura IV-6 e Figura IV-7 mostram duas vistas com ângulos de visualização diferentes de duas sessões de utilização, obtidas por este protótipo.

A Figura IV-6, abaixo, representa parte de uma sessão de um utilizador enquanto este utilizava um *site* de ensino à distância (<http://webct.ua.pt>). As páginas aparecem empilhadas, sendo a distância entre cada página proporcional ao tempo decorrido. Para se ter uma melhor ideia de contexto, o conteúdo de cada página é também representado. O uso de transparência permite que a página do topo não oculte as que estão abaixo dela. Desta maneira, tem-se uma vista geral, que oferece contexto, e ao mesmo tempo algum detalhe.



**Figura IV-6:** Vista 3D das páginas de uma sessão de um utilizador. As páginas estão empilhadas recorrendo a semitransparência. Imagem obtida com a versão 0 do protótipo.

A Figura IV-7, abaixo, apresenta uma vista diferente. Está representada uma parte de um *site* (<http://www2.ca.ua.pt/ihc>) preparado, no âmbito deste trabalho, para uma unidade curricular lecionada no Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro. Tal como no caso anterior, as páginas são empilhadas em 3D, sendo o conteúdo de cada página também representado, com a diferença que em vez de transparência, foi utilizado um operador XOR.



**Figura IV-7:** Vista 3D das páginas de uma sessão de um utilizador. As páginas estão empilhadas recorrendo a um operador do tipo XOR. Imagem obtida com a versão 0 do protótipo.

### 4.3. Correlação e integração de módulos

Apesar dos protótipos dos dois blocos desenvolvidos terem permitido testar algumas soluções, a estratégia usada na sua implementação obrigava a alterações no código fonte e respetiva compilação sempre que pretendia realizar uma nova experiência. Isto implicava uma complexidade e dispêndio de tempo consideráveis.

Era necessária uma evolução dos dois protótipos. Na verdade, mais que uma evolução, o que pretendia era a combinação dos dois.

Em traços gerais, tratou-se de fazer as adaptações necessárias para que pudessem funcionar em conjunto, como uma única aplicação. O maior desafio foi conseguir que o resultado do módulo *Parser Métrico* pudesse ser diretamente utilizado no módulo *Visualização*.

Esta nova abordagem veio tornar mais rápido e eficaz o teste de novos métodos de filtragem e extração de informação, uma vez que mais facilmente se podia observar o resultado. Simultaneamente, permitiu otimizar o processo de teste de novos métodos de visualização, na medida em que passou a ser mais fácil conseguir dados para esses testes.

### 4.4. Interface gráfica de utilizador

Depois do teste com dados de utilização de dois *sites* e de se ter refinado a especificação do sistema a desenvolver, estava na altura de se pensar em adicionar numa interface gráfica de utilizador.

Os blocos desenvolvidos até ao momento eram protótipos do tipo descartável e pretendiam testar algumas questões da funcionalidade da aplicação a implementar. Adicionar uma interface gráfica a esses protótipos não se mostrava tecnicamente adequado, pelo que se optou pela criação de um novo protótipo para testar e refinar as especificações da interface de utilizador.

Começou-se por fazer uns esboços em papel das várias partes da interface. De forma a ter uma noção mais exata do resultado, criou-se um protótipo rápido e também descartável, da interface a desenvolver. Para o efeito usou-se o VB (*Visual Basic*) da *Microsoft*, que permite colocar os vários elementos e objetos numa interface usando a técnica de arrastar e largar (*drag and drop*).

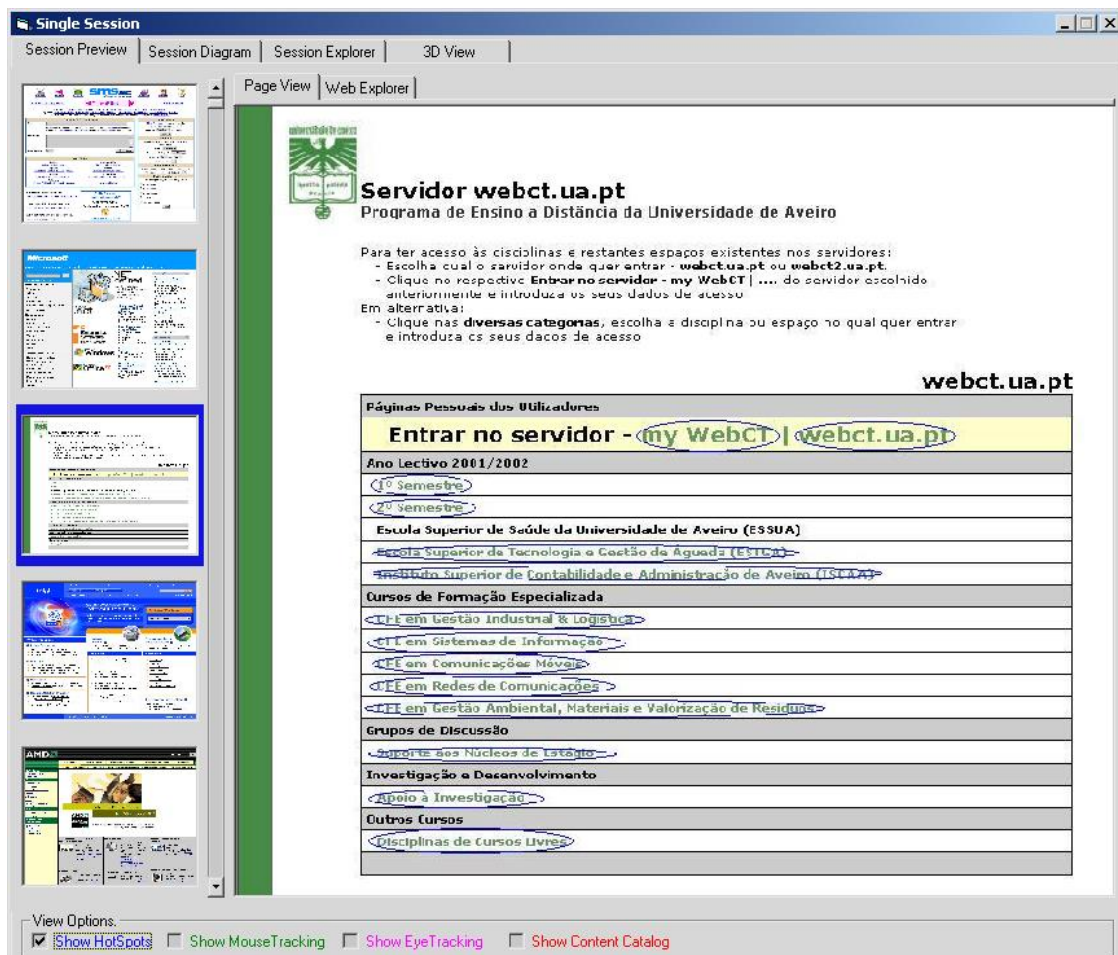
O resultado foi um protótipo da interface de utilizador sem quaisquer funcionalidades implementadas. Apresentava, contudo, um conjunto de ecrãs ilustrativos do aspeto das várias partes da aplicação a desenvolver, nomeadamente os elementos de interação e o

aspecto de alguns métodos de visualização que posteriormente foram sistematizados na secção 4 (Métodos de visualização e análise de informação) do Capítulo III.

Seguidamente são apresentadas algumas figuras ilustrativas do trabalho desenvolvido.

Na Figura IV-8 é apresentado o aspeto da interface de uma janela que permite a visualização de uma sessão de um utilizador. Na parte superior aparece um conjunto de tabuladores que permitem explorar mudar para umas das partes apresentadas. Do lado esquerdo representadas miniaturas das várias páginas visitadas. Do lado direito está representada a página selecionada com a identificação dos *Hotspots*. Finalmente, na parte inferior, existe um conjunto de opções de visualização que permitem sobrepor à página os seus *hotspots*, por exemplo.

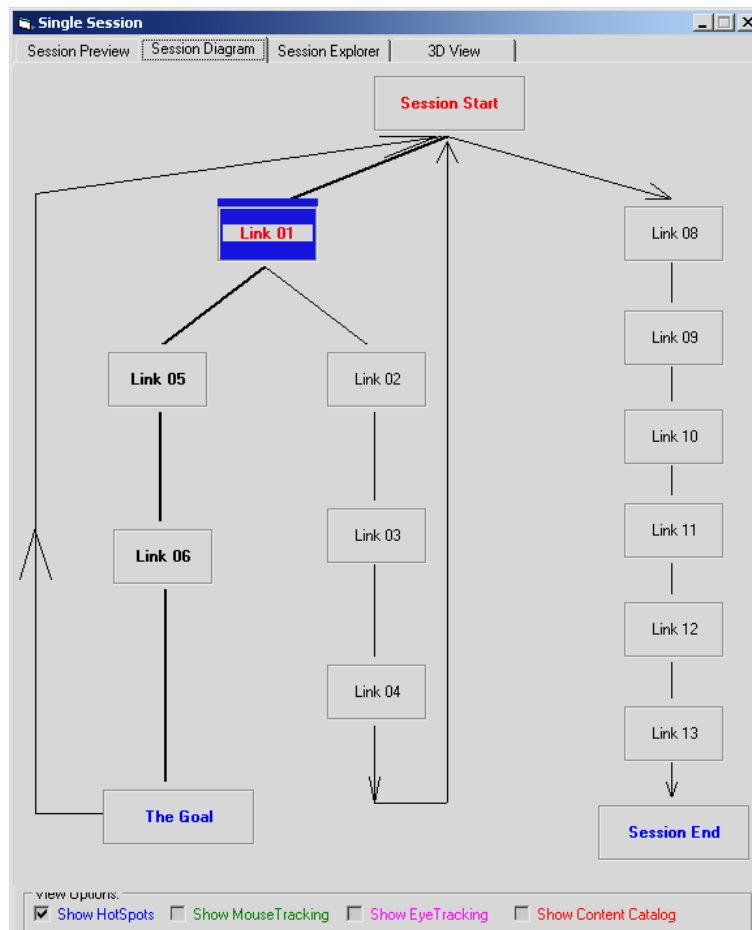
Este ecrã foi uma primeira abordagem às técnicas de visualização descritas na secção 4.5.12 (Visualização das páginas de uma sessão de um utilizador) em conjugação com o descrito na secção 4.4.1 (Visualização das áreas (blocos) de uma página), ambas do Capítulo III.



**Figura IV-8:** Visualização de uma sessão de um utilizador. Do lado esquerdo estão representadas miniaturas das várias páginas visitadas. Do lado direito está representada a página selecionada com a identificação dos *Hotspots*. Imagem obtida com a versão 0 do protótipo.

A Figura IV-9 apresenta um diagrama com a representação de uma sessão de um utilizador. É possível ver a página inicial, o percurso utilizado, o objetivo da sessão e a página de saída (última página visitada na sessão).

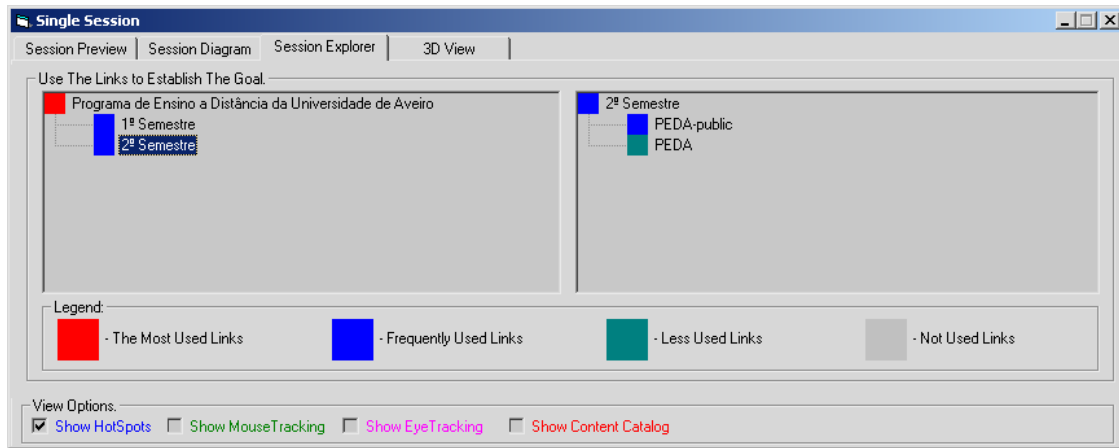
Este ecrã continha uma vista semelhante á descrita na secção 4.5.10 (Visualização dos trajetos possíveis) do Capítulo III.



**Figura IV-9:** Diagrama com a representação de uma sessão de um utilizador. Imagem obtida com a versão 0 do protótipo.

A Figura IV-10 apresenta uma proposta para a interface de navegação pelas páginas de um *site* ou de uma sessão de um utilizador. As páginas são representadas por blocos codificados com uma cor correspondente à sua popularidade. A metade esquerda contém a página selecionada e a que lhe deu origem (*referer*). Do lado direito, aparece a lista de *hotspots* da página selecionada.

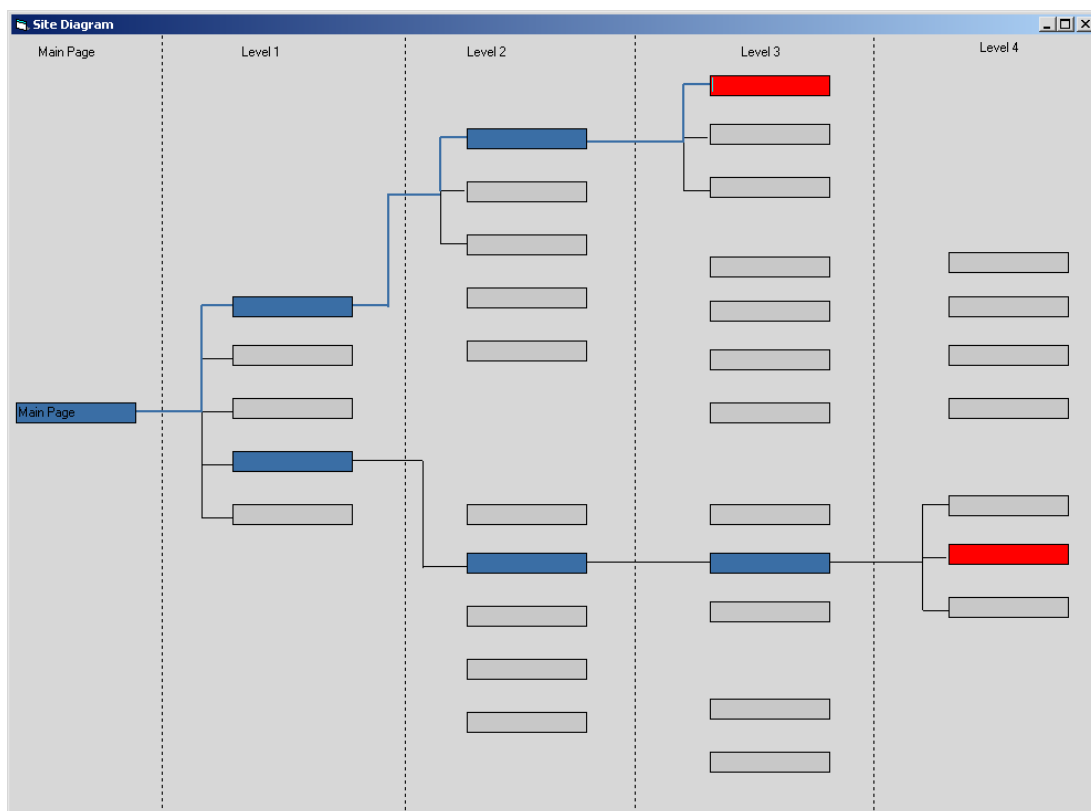
Este ecrã mostrava uma primeira tentativa para representar as técnicas de visualização descritas nas secções 4.5.9 (Visualização / enaltecimento de informação relacional de uma ligação ou nodo) e 4.5.11 (Visualização da taxa de utilização de cada página de um *site*), ambas do Capítulo III.



**Figura IV-10:** Interface de navegação de um *site*. Imagem obtida com a versão 0 do protótipo.

A Figura IV-11 apresenta um diagrama com uma representação de um *site* organizado de forma hierárquica. É possível identificar o percurso para chegar da página principal até uma dada página. Estão representadas duas possibilidades, codificadas a vermelho com os respetivos percursos, codificados a azul.

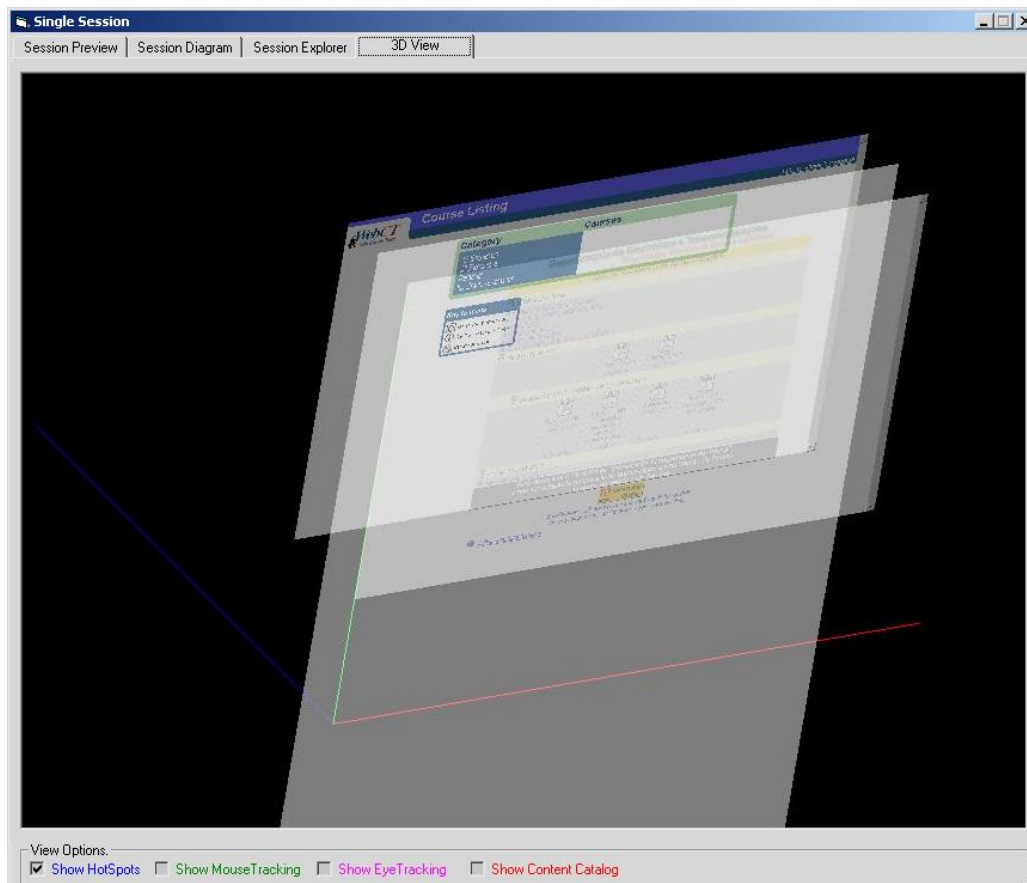
Este ecrã apresenta uma representação muito próxima da técnica descrita na secção 4.5.14 (Visualização das páginas de uma sessão de um utilizador (por níveis)) do Capítulo III.



**Figura IV-11:** Diagrama com uma representação de um *site*. São representados os percursos de duas pequenas sessões de utilizador. Imagem obtida com a versão 0 do protótipo.

A Figura IV-12 apresenta uma proposta para uma visualização 3D de uma sessão de um utilizador. Estava previsto ser possível mudar o ponto de vista (ângulo de visualização) por manipulação direta com o rato (clicando e arrastando).

A imagem usada neste ecrã tinha sido obtida no protótipo do módulo Visualização, descrito na secção 4.2 acima, neste capítulo, e implementava uma visualização muito próxima da descrita na secção 4.5.13 (Visualização em 3D das páginas de uma sessão de um utilizador) do Capítulo III.



**Figura IV-12:** Vista 3D das páginas usadas numa sessão de um utilizador. Imagem obtida com a versão 0 do protótipo.

## 5. Segundo protótipo – versão 1

Os protótipos apresentados na secção anterior ajudaram a testar e especificar algumas partes da funcionalidade e da interface da aplicação a desenvolver tais como a conceção e desenvolvimento de alguns métodos e técnicas de visualização para estes contextos infocomunicacionais. A utilização destes protótipos deixou também patente a falta de



uma solução de interface de utilização flexível e que permitisse o desenvolvimento de um ambiente experimental mais robusto destes métodos e técnicas.

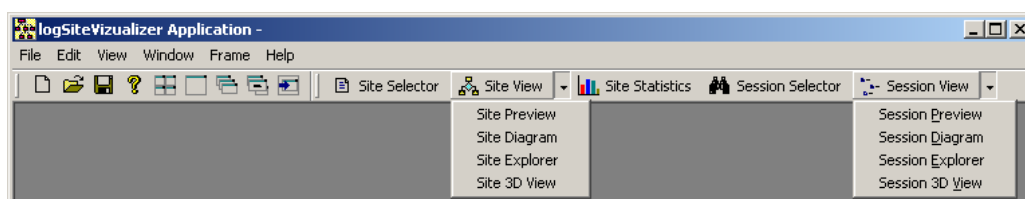
Por terem sido utilizados protótipos descartáveis não foi dedicado muito tempo nem recursos, contudo era necessário começar a desenvolver a aplicação final. Para o efeito, foi necessário recorrer a um novo protótipo, escrito de raiz, que mais facilmente pudesse evoluir e ser melhorado tendo como objetivo a aplicação final.

Tratando-se de um protótipo evolutivo, era necessário ter mais cuidados no que respeita às práticas de programação e qualidade do código produzido. Ainda assim, muito do código produzido para os protótipos anteriores pôde ser reaproveitado, permitindo reduzir um pouco o ciclo de produção.

Para desenvolver este protótipo recorreu-se ao pacote *Microsoft Visual Studio C/C++*, uma das plataformas de desenvolvimento de aplicações para ambiente *Windows* mais usadas do mercado.

## 5.1. Interface base

A Figura IV-13 apresenta uma vista geral da funcionalidade disponibilizada pela aplicação prototipada. Estas funcionalidades foram o resultado da evolução do protótipo, com muitas alterações e acrescentos, durante o seu uso no âmbito desta tese. Para que se possa ter uma ideia global das opções disponíveis, a figura foi manipulada de modo a mostrar o conteúdo de todos os menus.



**Figura IV-13:** Vista manipulada dos vários elementos da interface base do protótipo, com os menus visíveis. Imagem obtida com a versão 1 do protótipo, publicada em (Nunes et al., 2002).

As Figura IV-14 e Figura IV-15 apresentam duas situações em que a aplicação expunha todas as vistas disponíveis. No primeiro caso, as várias vistas eram apresentadas em cascata, sendo só necessário escolher a vista pretendida para esta fique no topo da pilha. No segundo caso, era possível ter uma visão geral sobre todas as vistas, ainda que a área disponível para cada janela seja reduzida. Em qualquer dos casos, podia-se maximizar uma janela, para se dispor de uma área maior de ecrã. Podia ainda fazer-se um arranjo com um conjunto de janelas correspondente aos métodos de visualização pretendidos, como o ilustrado na Figura IV-16. Este tipo de arranjos era o mais apropriado para tirar partido da sincronização entre as várias vistas, em que as ações do

utilizador numa dada vista tinham impacto, não só na vista corrente, mas também em todas as outras vistas, de forma sincronizada.

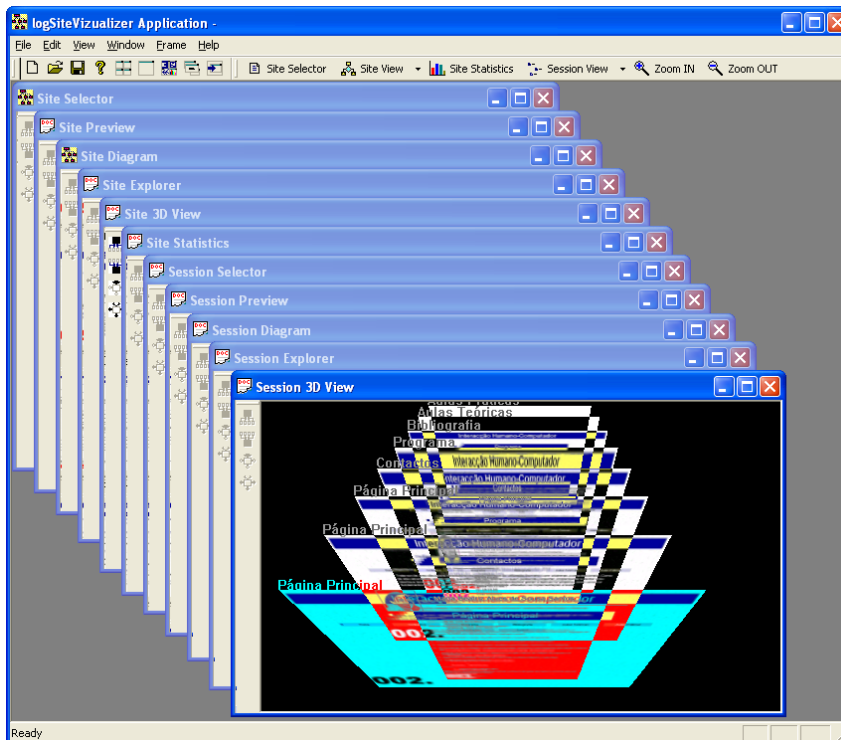


Figura IV-14: Aplicação com todas as vistas ativas. Note-se que as vistas estão dispostas em cascata, ficando a maior parte das janelas ocultas pelas que estão no topo. Imagem obtida com a versão 1 do protótipo.

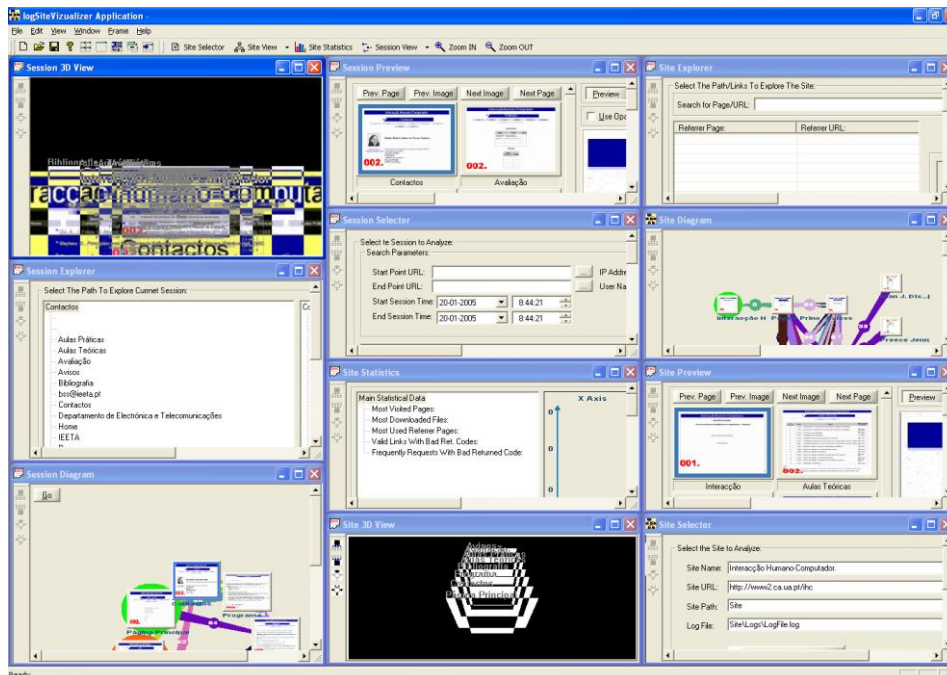
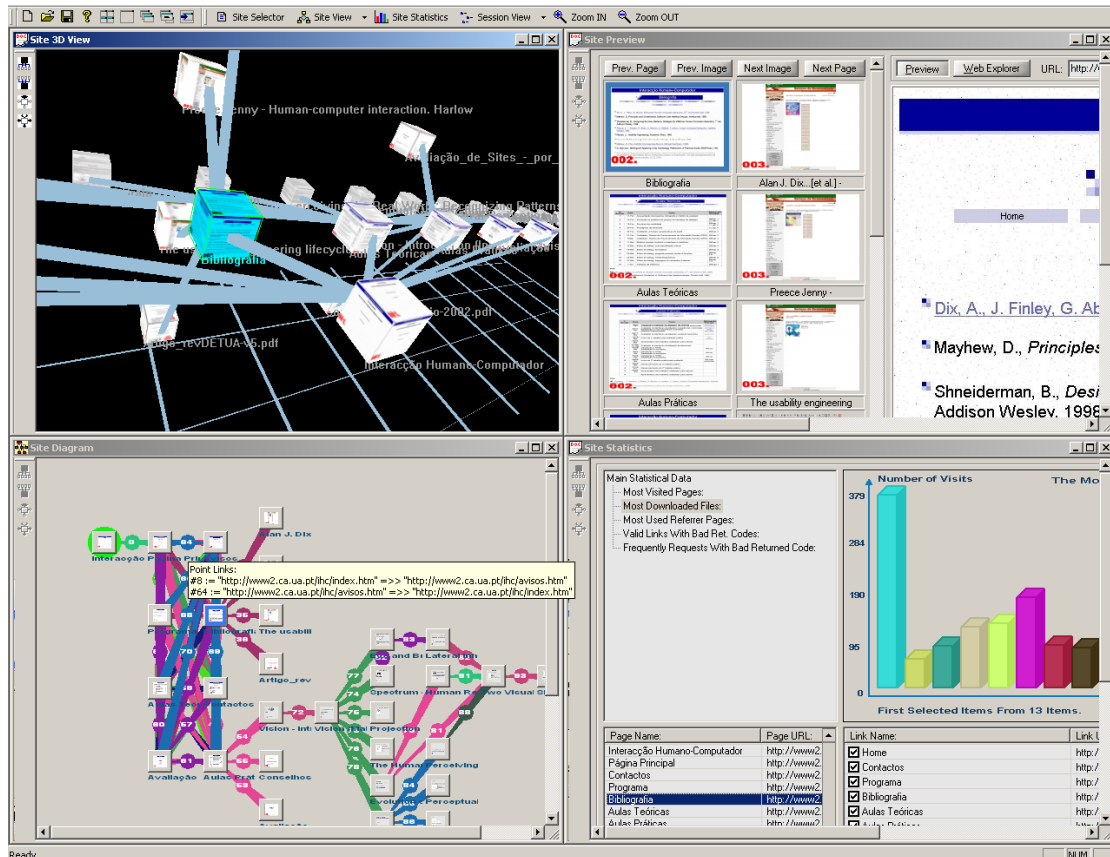


Figura IV-15: Aplicação com todas as vistas ativas. Note-se que estão dispostas de maneira a que todas sejam visíveis em simultâneo. Imagem obtida com a versão 1 do protótipo.



**Figura IV-16:** Múltiplas vistas sincronizadas. Ao seleccionar uma página num vista, são todas atualizadas de forma sincronizada em conformidade. Imagem obtida com a versão 1 do protótipo.

A possibilidade de se ter vistas múltiplas sincronizadas foi uma das mais-valias desta aplicação. Olhando para a Figura IV-16, pode observar-se que do lado esquerdo aparecem duas vistas que permitiam analisar a estrutura do site. A representação 3D apresentada em cima permitia uma visão geral, sobretudo quando se tratava de um site de grandes dimensões. Já o diagrama apresentado em baixo permitia uma vista de maior detalhe numa dada zona do site. Como estas vistas se complementavam, podia ser útil tê-las juntas para mais rapidamente localizar uma dada página no site. Usando, por exemplo, uma vista para encontrar a zona onde se encontrava, passando depois para a outra vista para analisar os detalhes relativos à página pretendida. Note-se que em ambas as vistas era apresentada informação adicional, uma, comum às duas vistas, outra, só presente numa delas, o que aumentava a sua complementaridade. Podia encontrar-se informação como a referente à posição da página na estrutura hierárquica do site, a lista de ligações de uma dada página (*hotspots* e *referers*), a identificação gráfica das ligações com outras páginas, etc.

A vista que aparece no quadrante superior direito permitia obter a página selecionada em tamanho real, para que se pudessem observar todos os detalhes pretendidos. Esta vista, além da página selecionada, apresentava ainda as restantes páginas do site representadas em miniatura, como se pode ver na secção mais esquerda desta vista. A

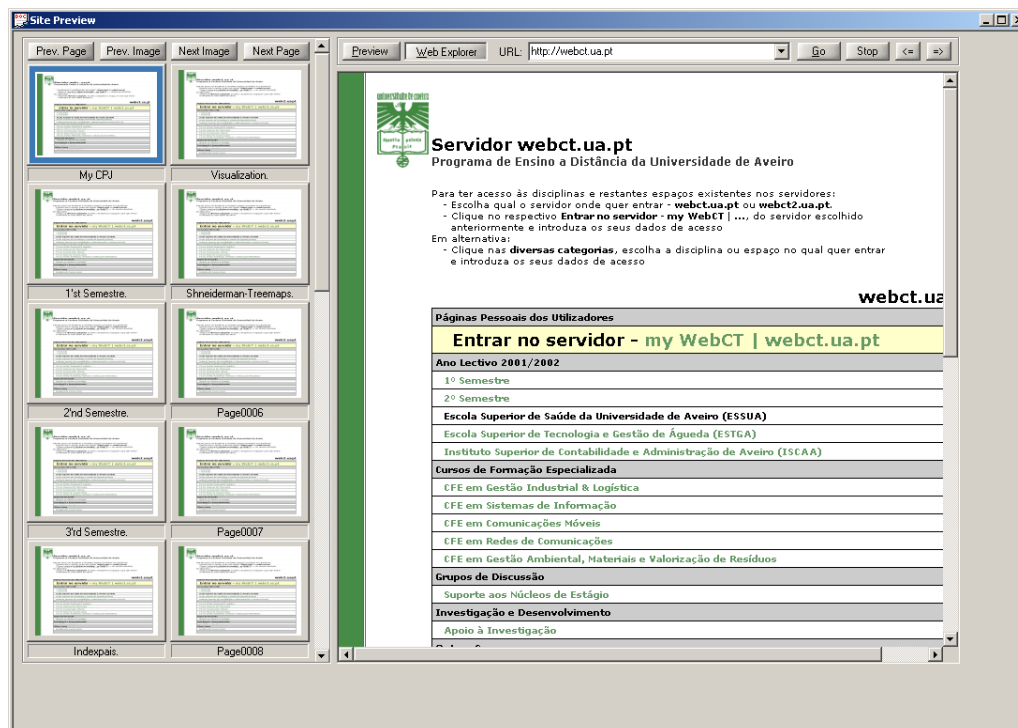
seleção de uma dessas páginas, mudava a página corrente, com a respetiva atualização sincronizada das restantes vistas.

Por último, a vista presente no quadrante inferior direito, permitia observar informação estatística relativa ao *site* e à página selecionada. No caso ilustrado, apresentava as páginas mais descarregadas, de entre os *links* presentes na página selecionada.

Seguidamente vão passar-se em revista alguns esquemas de visualização implementados na aplicação, com uma breve descrição dos mesmos.

## 5.2. Métodos de visualização

A Figura IV-17 apresenta as páginas de um *site*. Na parte esquerda era apresentada uma lista de todas as páginas do *site*, com uma pequena representação em miniatura de cada uma, bem como o nome da mesma, se esta o tiver. Do lado direito era apresentada uma representação em tamanho real da página selecionada. Esta tanto podia ser uma imagem (captura de ecrã) da página produzida no instante em que se procedeu à recolha e compilação da informação do *site* para o Atlas Digital de Referência que dá suporte à aplicação, como podia ser uma visualização da página real nesse momento, recorrendo ao navegador embebido na aplicação. Era ainda possível especificar manualmente o *URL* de uma página e navegar a partir daí.

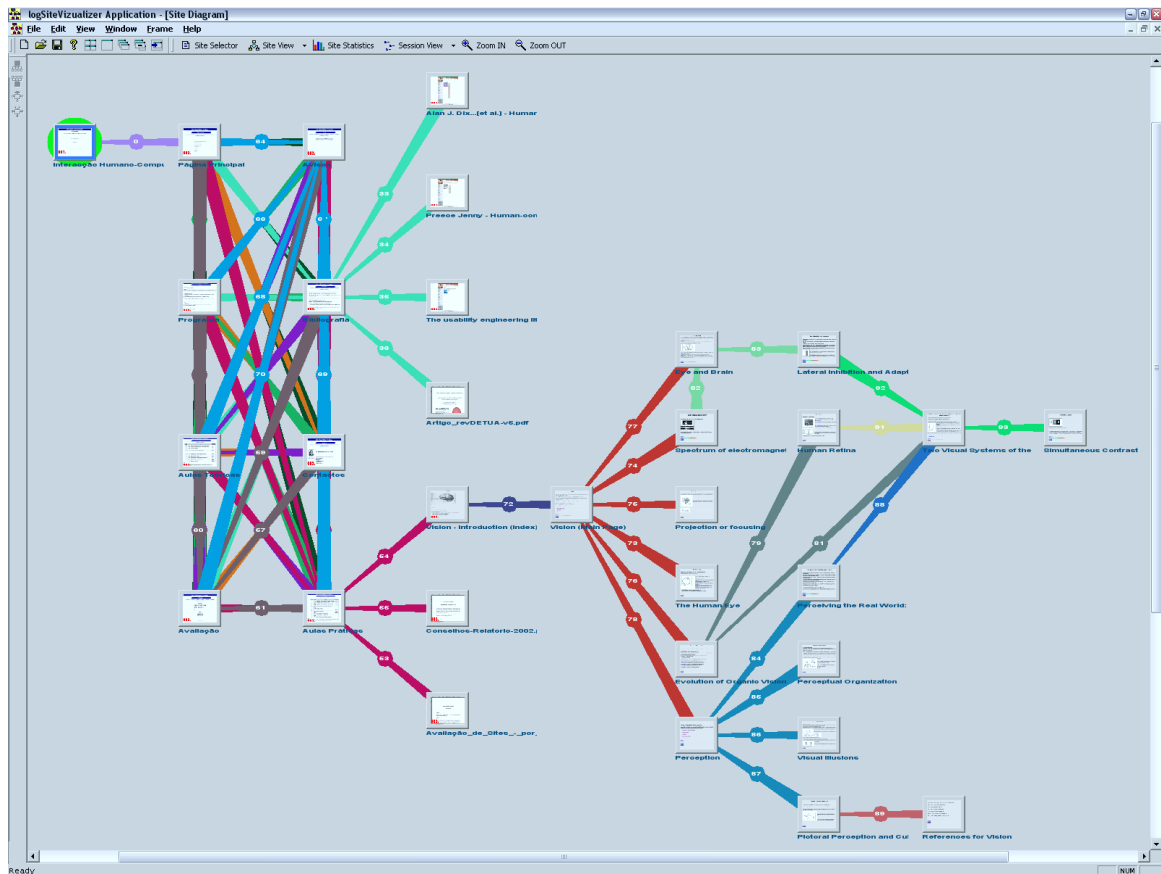


**Figura IV-17:** Vista das páginas de um *site*. Na parte esquerda está a lista das páginas do *site* e do lado direito está uma representação real da página selecionada. Imagem obtida com a versão 1 do protótipo, publicada em (Nunes et al., 2002).



Como se pode constatar, este esquema não apresentava todas as ligações entre as páginas representadas. Só eram apresentadas as ligações de uma página de um nível para as páginas do nível seguinte. Se a página tivesse ligações para outra página de outros níveis, estas não eram representadas de maneira a tentar manter a legibilidade do esquema. Para resolver esta e outras questões foi desenvolvido o esquema apresentado na Figura IV-19.

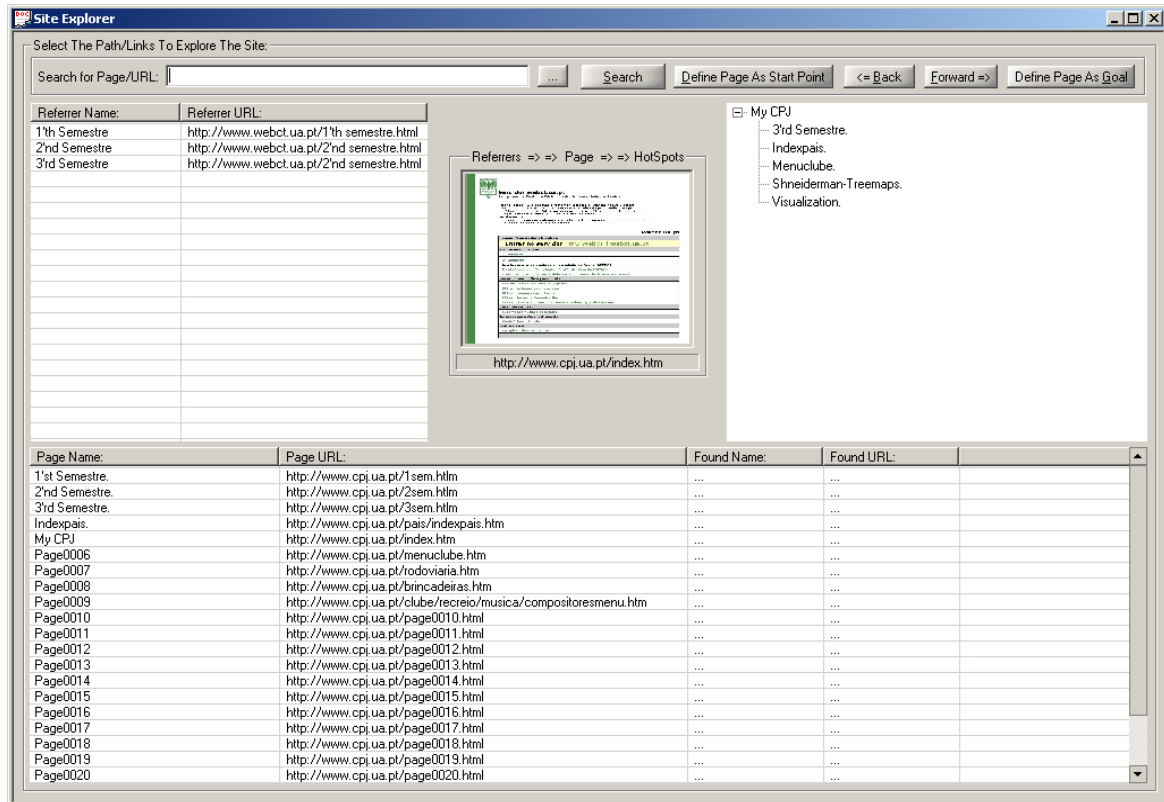
Neste esquema cada página era representada sob a forma de um pequeno retângulo com uma miniatura da página. Cada página era representada uma única vez, mas todas as ligações com outras páginas eram representadas. Isto, por vezes, provocava alguma confusão, na medida que as ligações acabavam por ficar todas sobrepostas, não sendo possível distinguir muitas delas.



**Figura IV-19:** Visualização em 2D da estrutura de um *site*. Cada página é representada por uma miniatura e as ligações entre cada página estão também representadas. Imagem obtida com a versão 1 do protótipo.

A janela ilustrada na Figura IV-20 permitia a exploração de um *site*. A parte inferior apresentava uma lista com todas as páginas constituintes do *site*. Na parte superior, ao centro era apresentada uma representação miniatura da página seleccionada, do lado direito uma lista com as páginas para as quais se podia ir a partir da página corrente (*links* associados aos *hotspots*) e do lado esquerdo uma lista com as páginas a partir das quais se podia chegar à página seleccionada (*referes*). A seleção de uma nova página

levava a que a página corrente fosse alterada e consequentemente também a lista de *referers* e *hotspots*. Para ser mais fácil encontrar a página pretendida, existia no topo uma ferramenta de pesquisa, bastando para tal digitar o *URL* da página pretendida.



**Figura IV-20:** Interface de exploração de um *site*. Em baixo aparece a lista das páginas do *site*. Na parte superior, no centro está a página corrente, do lado direito uma lista com os *links* e do lado esquerdo uma lista com os seus *referers*. Imagem obtida com a versão 1 do protótipo, publicada em (Nunes et al., 2002).

Na Figura IV-21 ilustra-se uma vista com informação estatística referente à utilização de um *site*. No canto superior esquerdo era possível selecionar o tipo de informação estatística que se pretendesse analisar. À sua direita aparecia um gráfico de barras apresentando a informação solicitada, também presente na lista em baixo à direita. Nesta lista era possível selecionar os elementos ou valores que se pretendesse que fizessem parte do gráfico, recorrendo às *check-boxes* à esquerda de cada linha da tabela.

A tabela em baixo à esquerda, apresentava uma lista com todas as páginas do *site*. A seleção de uma página aqui modificava a informação estatística apresentada que deixava de dizer respeito ao *site* em geral para passar a referir-se somente aos seus *hotspots*. Era uma maneira, por exemplo, de analisar quais os *hotspots* mais usados numa dada página.

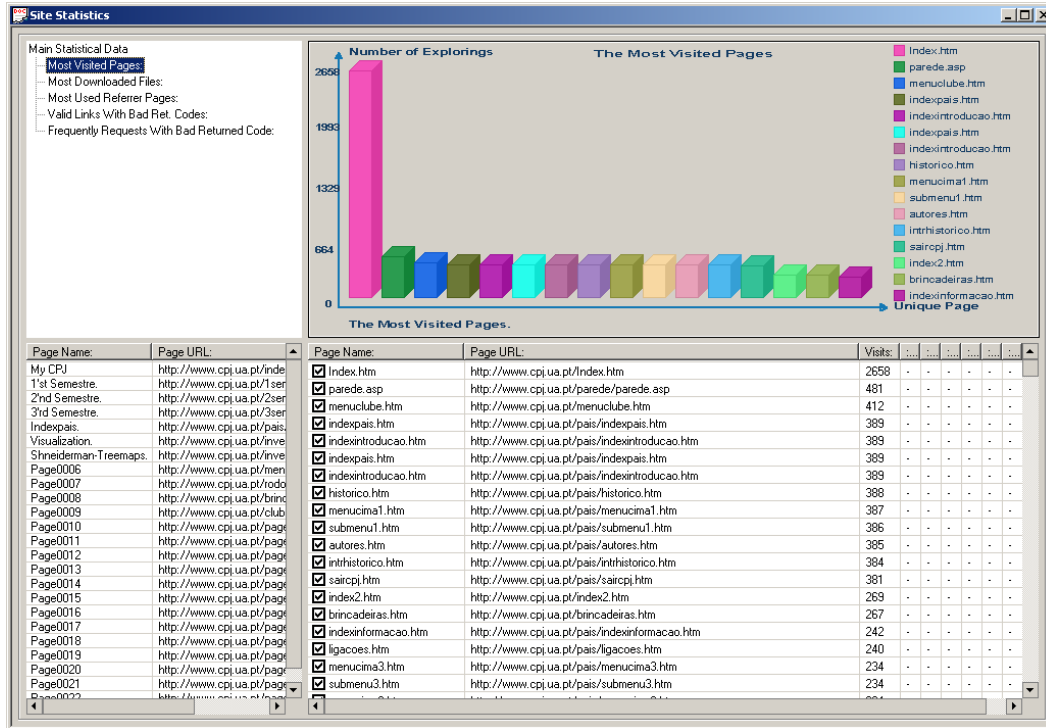


Figura IV-21: Apresentação da informação estatística. No quadrante superior esquerdo seleciona-se o tipo de informação pretendida e à direita a sua representação num gráfico de barras. Na parte inferior é possível selecionar os elementos constituintes do gráfico. Imagem obtida com a versão 1 do protótipo, publicada em (Nunes et al., 2002).

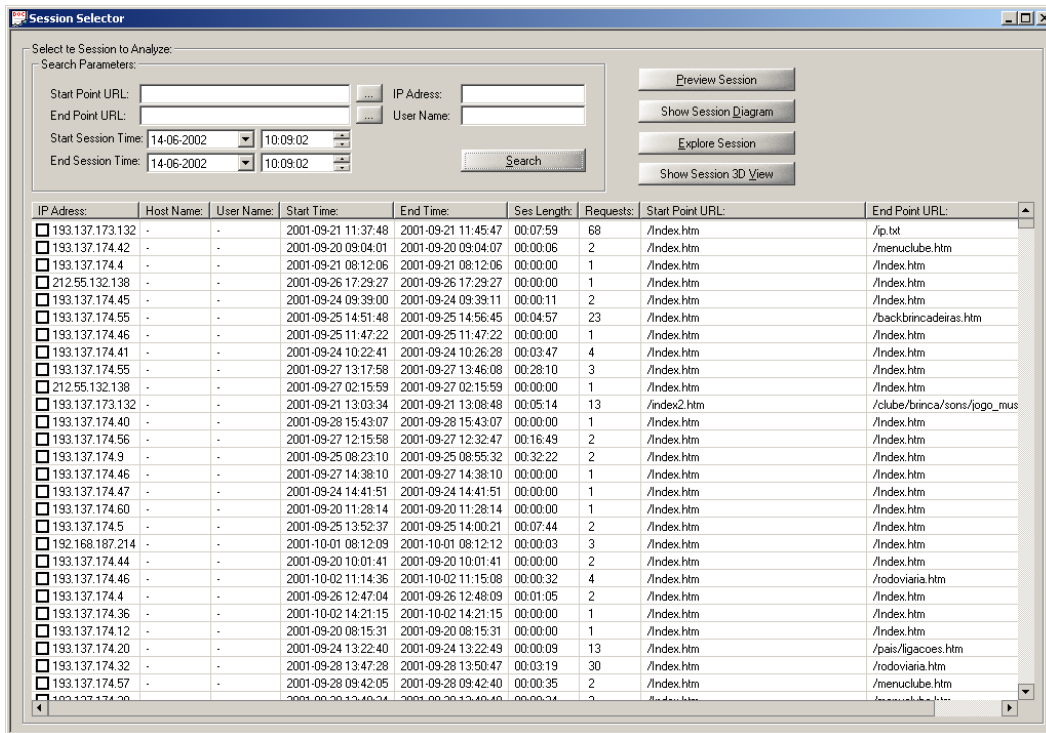


Figura IV-22: Interface que permite selecionar uma ou várias sessões. Permite ainda pesquisar e/ou ordenar segundo os critérios pretendidos. Imagem obtida com a versão 1 do protótipo, publicada em (Nunes et al., 2002).

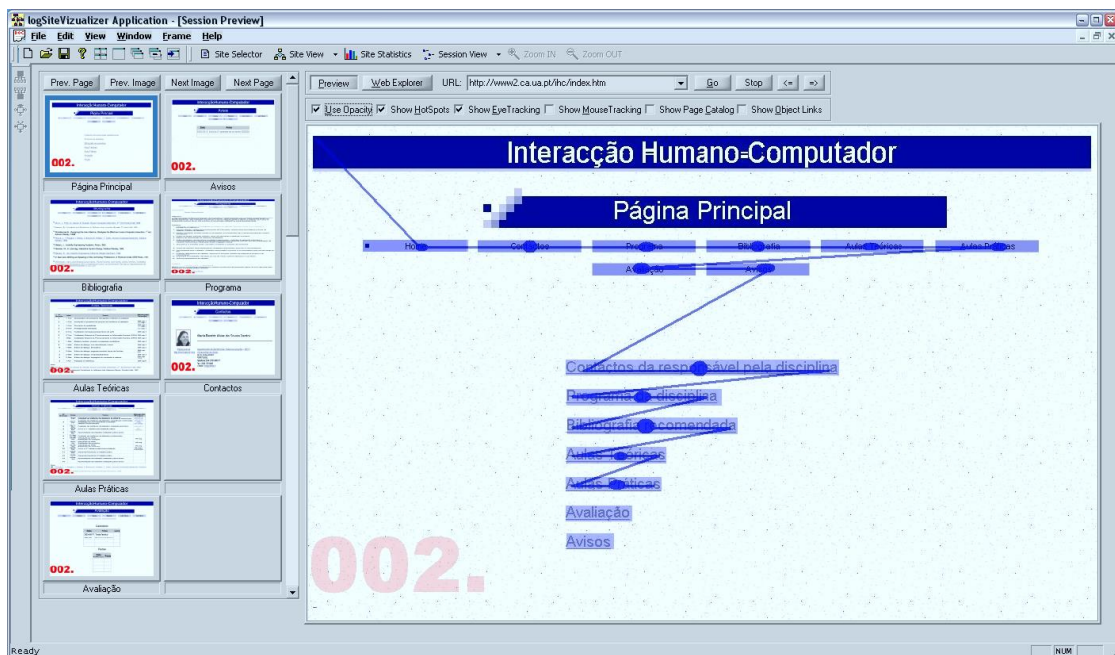


Se bem que a estrutura do *site* seja importante, o modo como os utilizadores navegam no *site* é, talvez, ainda mais importante. A Figura IV-22 mostra a interface disponível para seleccionar a, ou as, sessões que se pretendessem analisar. Para o efeito era suficiente marcar a *check-box* correspondente a cada sessão pretendida. A lista apresentada podia ser ordenada segundo qualquer das colunas apresentadas, bastando usar o rato para “cliquear” no identificador no topo da coluna respetiva.

A lista podia ser reduzida usando um ou vários critérios de pesquisa presentes na parte superior da janela. Nesse caso, só as sessões que respeitassem os critérios seleccionados eram apresentadas na lista.

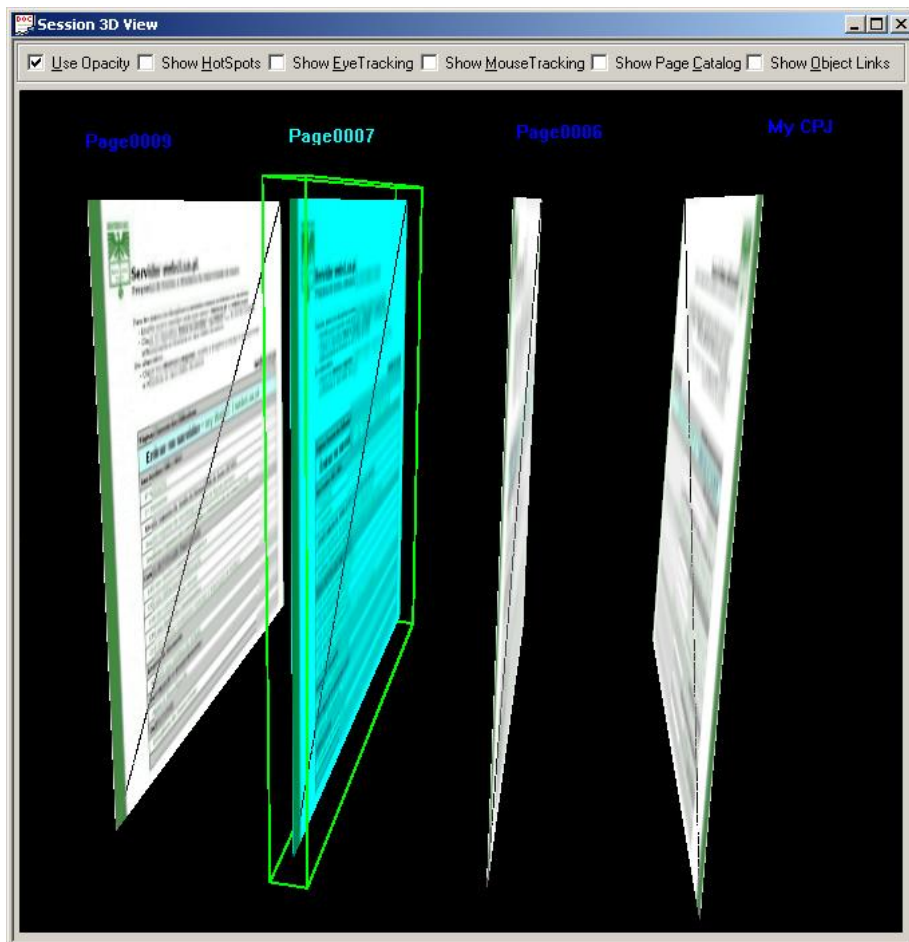
A visualização das páginas visitadas por um utilizador durante uma sessão tem muitas coisas em comum com a visualização das páginas de um *site*. Assim, não será estranho que o esquema apresentado na Figura IV-23 tenha tantas semelhanças com esquema apresentado na Figura IV-17, mais acima.

A visualização de uma sessão herdou toda a funcionalidade da visualização das páginas de um *site*, com algumas diferenças. A lista, presente do lado esquerdo, apresentava só as páginas visitadas durante a sessão, sendo apenas possível seleccionar uma dessas páginas. Por outro lado, foi adicionada funcionalidade, sendo possível observar informação adicional relacionada com a sessão, tal como a posição dos *hotspots*, percurso obtido com um *eyetracking*, percurso do rato, entre outras, sempre que essa informação estivesse disponível. Esta informação era representada sobreposta à página, usando semitransparência, ou não, dependendo da opção pretendida.



**Figura IV-23:** Visualização de uma sessão de utilizador. Na parte esquerda está a lista de páginas visitadas. No lado direito pode observar-se a página seleccionada, neste caso com informação de *EyeTracking* sobreposta. Imagem obtida com a versão 1 do protótipo.

Contudo, esta não era a única técnica disponível para observar uma sessão. Se bem que alguns métodos utilizados para observar a estrutura ou páginas de um *site* fossem adequados para o estudo de uma sessão, podia pensar-se em métodos desenhados especificamente para o efeito. O método ilustrado na Figura IV-24 aparece como um exemplo nessa linha. Apresentava uma representação 3D de uma sessão, em que as páginas eram colocadas numa pilha, utilizando a distância entre cada página para codificar o tempo. Podia seleccionar-se uma página, usando o rato para “clicar” sobre ela, ação que era assinalada com uma mudança na cor da página (esverdeado) e desenhando os contornos de uma caixa envolvente.



**Figura IV-24:** Representação 3D das páginas de uma sessão. As páginas são apresentadas por ordem cronológica. É ainda possível seleccionar uma página. Imagem obtida com a versão 1 do protótipo, publicada em (Nunes et al., 2002).

### 5.3. Evolução

Com o passar do tempo e à medida que se ia introduzindo mais funcionalidade no protótipo, foram-se tornando claras algumas limitações. Cada vez era mais complicado integrar novas formas de análise e visualização, implicado um crescendo de tempo e

trabalho necessários sempre que pretendia testar uma nova abordagem. Além disso, tornava-se urgente uma reformulação na estrutura de dados da aplicação para ser possível lidar com *sítes* que vão sendo modificados ou atualizados durante um período de estudo ou análise, entre outra informação importante. A própria forma de armazenar a informação recolhida começava a mostrar-se desadequada para as necessidades, sendo urgente uma reformulação da base de dados. Para se poder estudar *sítes* de maiores dimensões, que evoluem e são atualizados, e para se poder investigar e testar novos métodos de análise e visualização era necessário proceder a reformulações profundas do protótipo desenvolvido.

Nesta altura colocou-se um problema: fazer uma reformulação ou começar de raiz uma nova aplicação.

As alterações necessárias envolviam aspetos tão profundos, tanto do ponto de vista técnico como do ponto de vista da interface e respetiva interação, que obrigariam a reimplementar partes da estrutura base do programa, alterando radicalmente a filosofia de funcionamento.

Face a este cenário, o desenvolvimento de uma nova aplicação de raiz apareceu como a melhor solução. Iria implicar algum investimento adicional na fase inicial do seu desenvolvimento, mas com vantagens a médio e longo prazo. Além disso, a maioria dos algoritmos, e mesmo algum código, seria sempre possível ser reutilizado.

Nesta sequência decidiu-se começar a especificar uma nova aplicação e desenvolver um novo protótipo para a mesma.

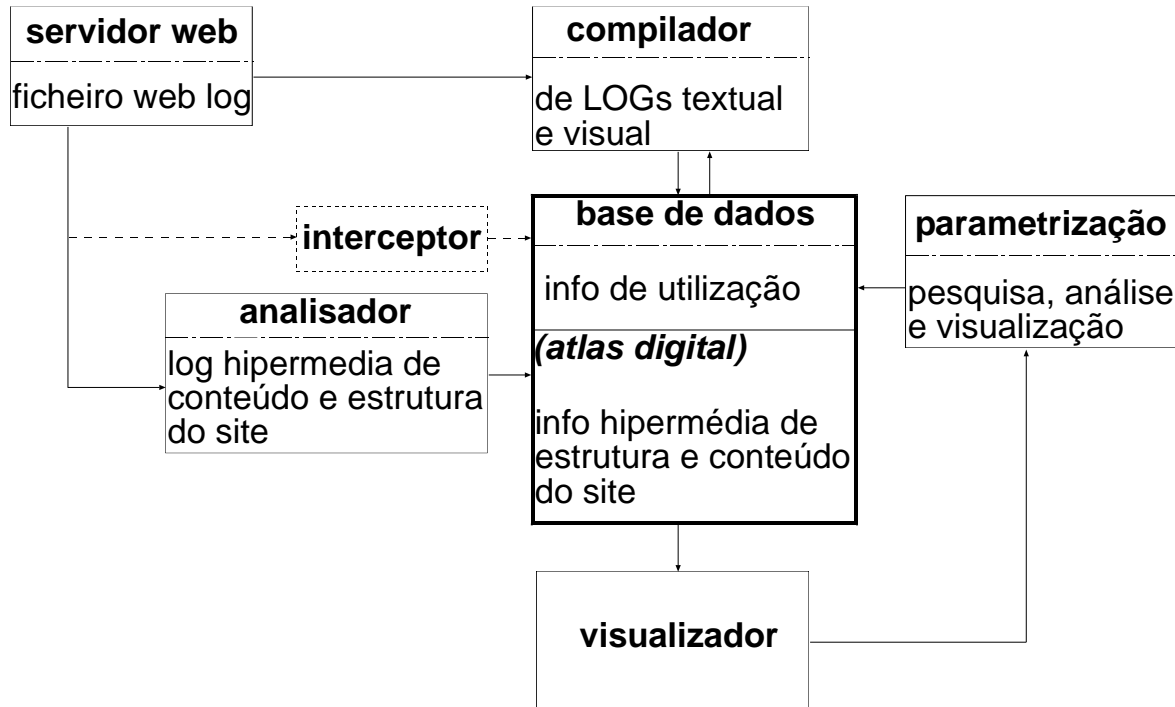
O primeiro passo nesse sentido passou por testar este protótipo (versão 1), identificar problemas e encontrar soluções. O processo está descrito na secção Capítulo V.2 abaixo (Avaliação das várias versões do protótipo). A partir daqui, foi necessário elaborar as especificações para o novo sistema.

## 6. Protótipo final – versão 2

Como descrito no final da secção anterior, o protótipo anterior, apesar das muitas alterações e melhorias que foi recebendo, acabou por manifestar que não estava preparado para evoluir de forma conveniente para fazer frente às necessidades. Na sequência da informação obtida nos testes e de tudo o que foi possível aprender com este protótipo definiram-se as especificações para a aplicação final. É neste contexto que surge esta nova versão do protótipo.

O novo protótipo foi desenvolvido tendo como base o modelo ilustrado, de forma simplificada, na Figura IV-25. Neste modelo podem identificar-se quatro blocos principais

suportados por uma base de dados. Estes quatro blocos: *Intercetor*, *Analizador*, *Compilador* e *Visualizador* foram desenvolvidos sob a forma de aplicações separadas que estão congregadas em torno de uma base de dados.



**Figura IV-25:** Modelo simplificado do sistema. Adaptada de (F. L. Zamfir, 2007).

O objetivo final deste sistema (constituído por um conjunto de blocos intersincronizados) é sublimado pelo bloco *Visualizador*, onde se faz o estudo e análise, de forma visual, do *site*. Este bloco usa a informação disponível na base de dados, contudo, é necessário “alimentar” previamente esta base de dados com toda a informação necessária. É aí que entram os restantes três blocos. O *Analizador* faz a catalogação e registo na base de dados, construindo um *Atlas Digital*, da estrutura do *site* em análise. O *Intercetor* é responsável pela captura, processamento e registo da interação do utilizador com o seu sistema enquanto utiliza o *site*. O *Compilador* processa a informação contida nos ficheiros de *log* e, depois de devidamente filtrada e processada, coloca-a também na base de dados. Finalmente o módulo *Visualizador* vai aplicar métodos de pesquisa, análise e visualização (recorrendo ao módulo *Parametrização*) para o estudo detalhado do *site* em causa.

Em termos temporais, o processo decorre como ilustrado na Figura IV-26. Começa-se por criar uma imagem do *site* em análise na base de dados (com o *Analizador*), captura-se a informação de utilização (com o *Intercetor*) e compilam-se os ficheiros de *log* (com o *Compilador*). Depois de tudo registado na base de dados está-se em condições de se proceder á análise e visualização do *site* (com o *Visualizador*, coadjuvado pelo módulo *Parametrização*).

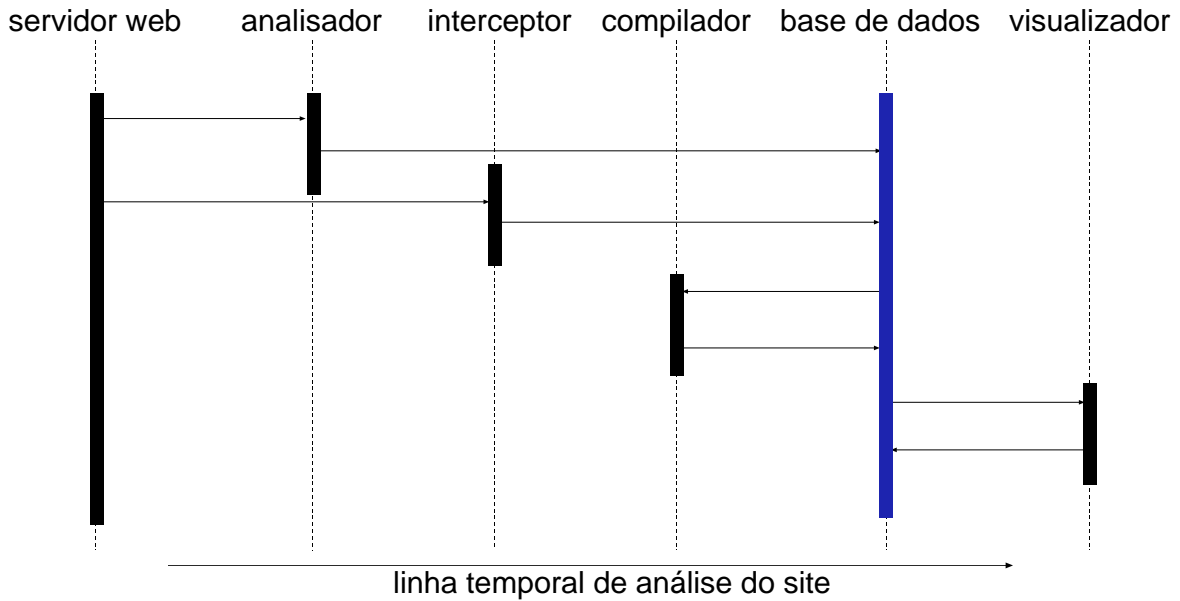


Figura IV-26: Linha temporal da análise a um *site*. Adaptada de (F. L. Zamfir, 2007).

A informação referente ao *site* registada na base de dados tem uma estrutura hierárquica como ilustrado na Figura IV-27. Um *site* tem um conjunto de páginas ordenadas segundo uma estrutura do tipo árvore. Uma página de nível *n* tem um conjunto de *links* a que correspondem páginas que irão aparecer no nível *n+1*, tendo, no entanto, cuidado para que cada página só apareça uma vez, no primeiro nível em que for detetada.

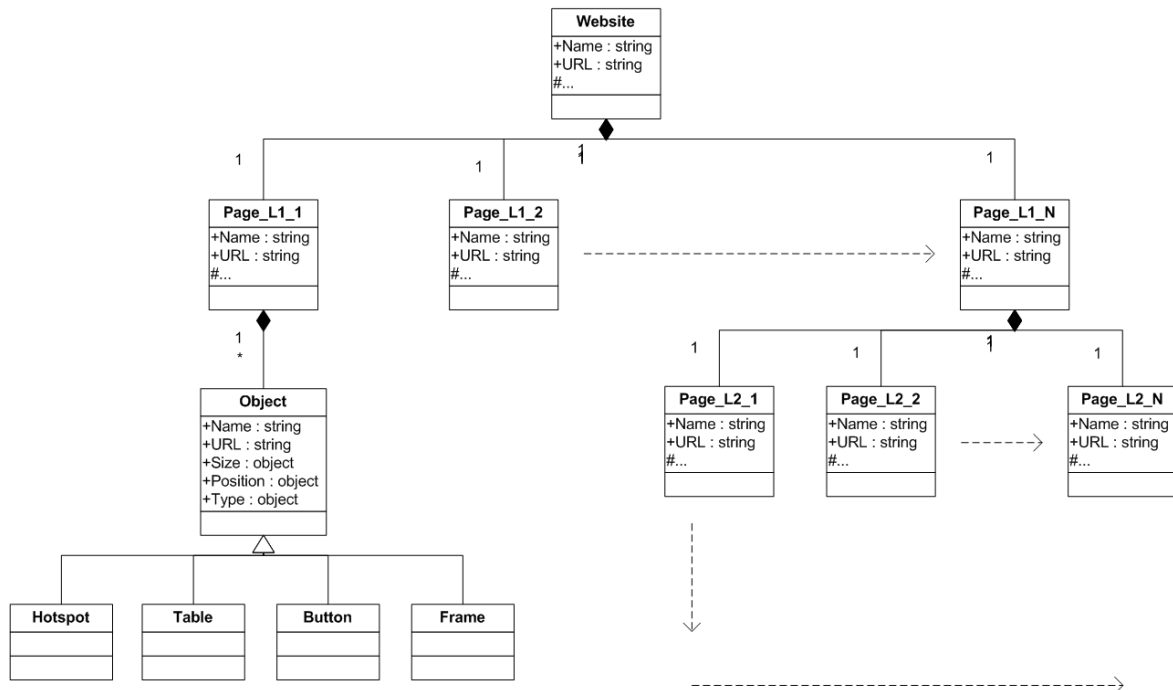


Figura IV-27: Representação hierárquica de um *site web* na base de dados. Retirado de (F. L. Zamfir, 2007).

Se o *site* é constituído por páginas, cada página é constituída por objetos como imagens, *hotspots*, botões, etc.. Também esta informação é igualmente registada na base de dados segundo uma estrutura hierárquica.

Os blocos principais da aplicação, naturalmente, têm características comuns. Estas características comuns têm por base uma estrutura de suporte (*framework*) que define uma plataforma de trabalho, a ligação entre os vários blocos e destes com a base de dados. A Figura IV-30 contém uma ilustração desta estrutura.

Os blocos identificados serão apresentados com mais detalhe nas secções seguintes.

## 6.1. Base de dados

A estrutura hierárquica ilustrada na Figura IV-27 tem correspondência com as estruturas da base dados ilustradas pelas Figura IV-28 e Figura IV-29 que apresentam, respetivamente, um resumo da informação relacionada com um *site web* e com cada página armazenada na base de dados. De notar que a informação foi previamente processada e registada por um, ou mais, dos módulos, *Intercetor*, *Analizador* ou *Compilador*, identificados nas Figura IV-25 e Figura IV-30. Estes módulos receberam a informação, no seu estado bruto das fontes originais, processaram e converteram para o formato adequado, antes de efetuarem o seu registo na base de dados.

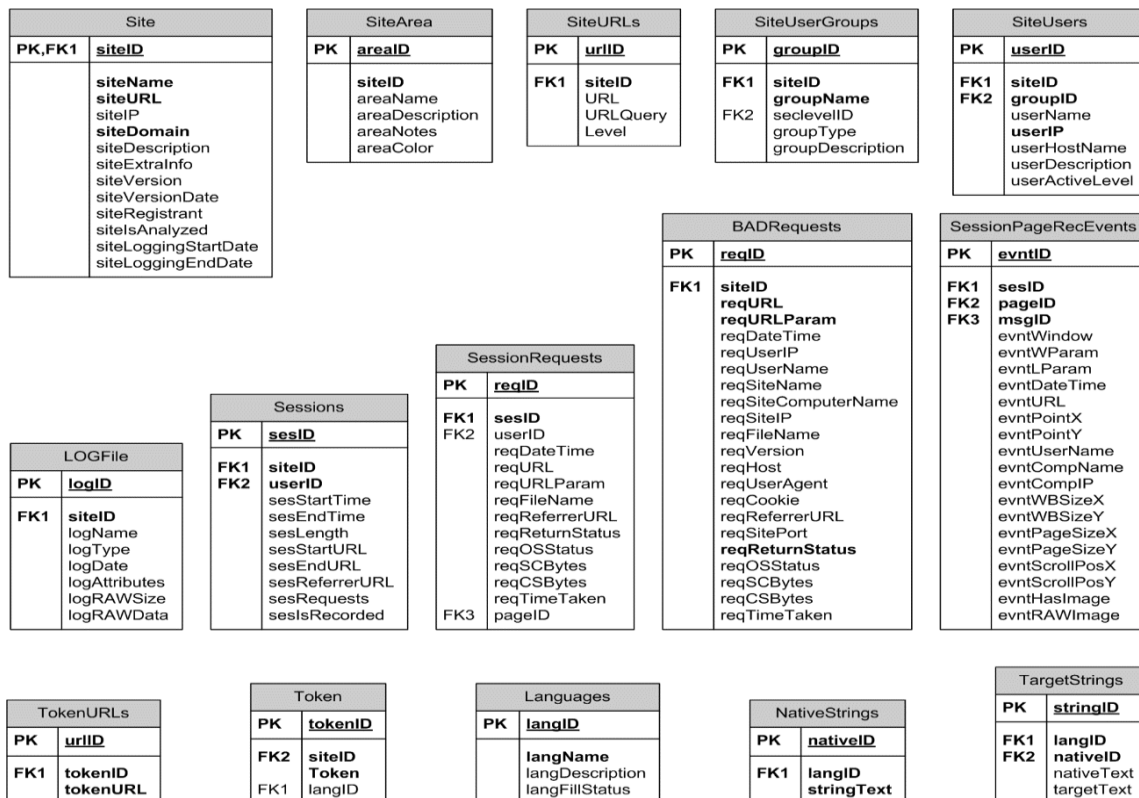


Figura IV-28: Informação de um *site web* na base de dados. Retirado de (F. L. Zamfir, 2007).

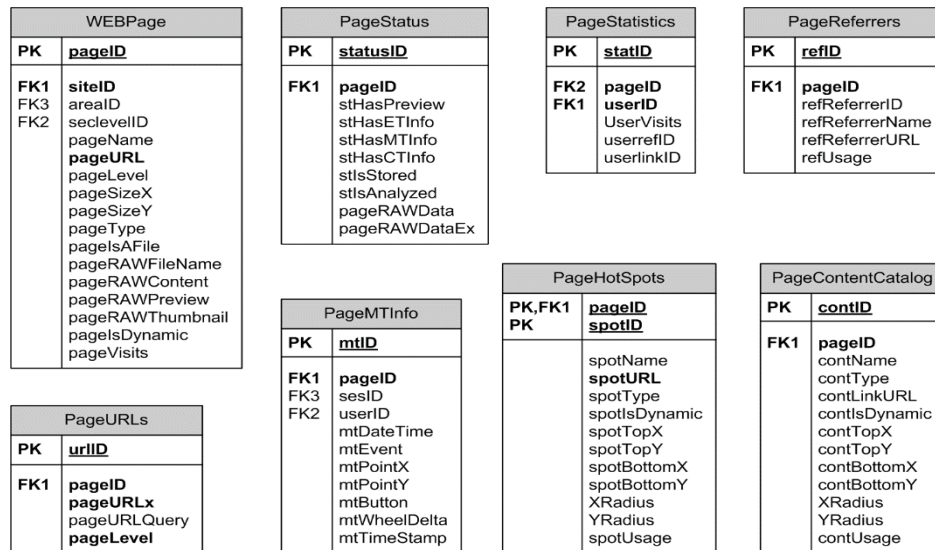


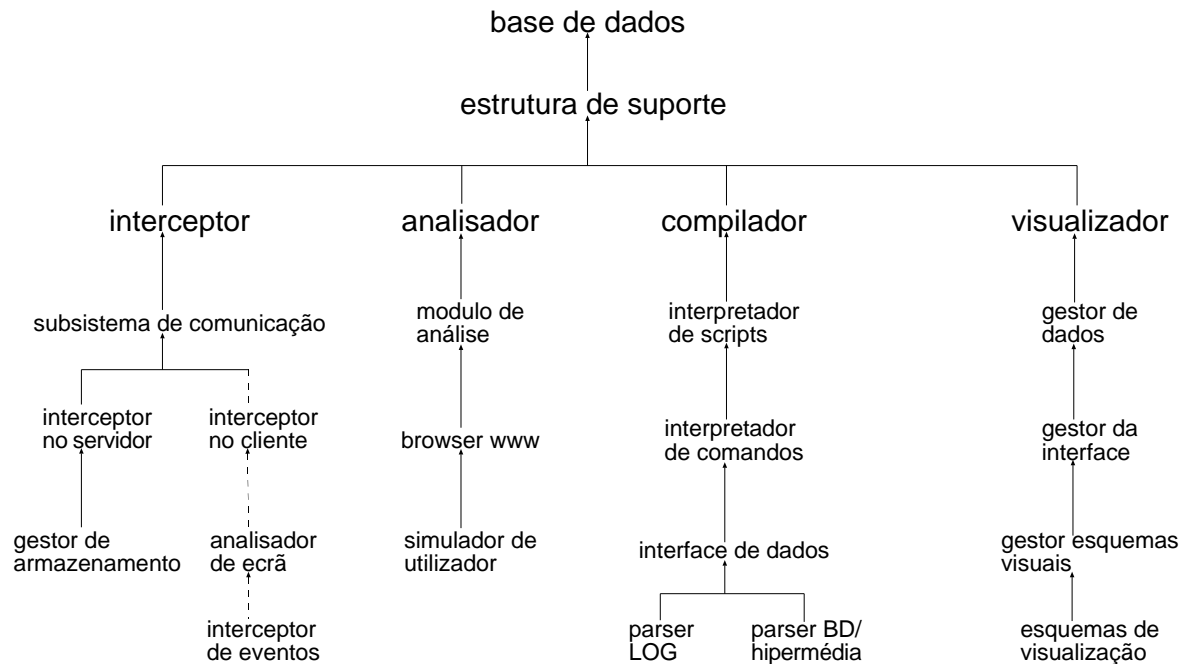
Figura IV-29: Informação de uma página web na base de dados. Retirado de (F. L. Zamfir, 2007).

Como visto acima, toda a informação relacionada com o site, páginas, utilização, etc., está congregada na base de dados. Os detalhes de implementação, profundamente técnicos, que deram lugar ao desenvolvimento de uma tese de mestrado por parte de um dos elementos da equipa de investigação (F. L. Zamfir, 2007), por se afastarem do objetivo desta tese, não serão aqui apresentados. Alguns detalhes podem ser encontrados no relatório (F. Zamfir, Nunes, Santos, et al., 2004), tendo sido também publicados em (F. Zamfir, Nunes, Teixeira, Mealha, et al., 2004; F. Zamfir, Nunes, Teixeira, Santos, et al., 2004).

## 6.2. Estrutura de suporte (*Framework*)

Durante o desenvolvimento da aplicação, a *Estrutura de Suporte* do sistema (*Framework*) foi conceptualizada e implementada para funcionar como uma interface comum e integrada para a manipulação da informação da base de dados, as comunicações de rede, a manipulação de imagens e um conjunto de controlos usados para a manipulação de diferentes tipos de informação. Todos os componentes da aplicação usam as entidades da *Estrutura de Suporte* permitindo uma mais fácil manipulação da informação e uma melhor integração.

A Figura IV-30 ilustra, de forma simplificada, as várias partes constituintes da aplicação. Note-se a relevância da posição central do bloco referente à *Estrutura de Suporte*. Todos os blocos principais estão relacionados com a *Estrutura de Suporte* e todos dependem dos seus serviços para acesso à base de dados.



**Figura IV-30:** Modelo simplificado das várias partes constituintes da aplicação. Adaptado de (F. Zamfir, Nunes, Teixeira, Santos, et al., 2004).

### 6.3. Intercetor

Este bloco tem como função capturar informação de utilização em duas vertentes distintas:

- Diretamente no servidor, interceptando e registando todos os pedidos efetuados ao servidor por um ou mais utilizadores.
- Na máquina cliente, que o utilizador está a usar durante a sua navegação no *site*, interceptando e registando todas as ações do utilizador. A informação recolhida inclui o conteúdo do ecrã, os movimentos do rato, a atividade do teclado, etc.. Enfim, todos os eventos que ocorrerem durante o período de monitorização.

Toda a informação recolhida é enviada para a base dados onde será convenientemente armazenada.

Estas aplicações foram só parcialmente implementadas durante o decorrer deste trabalho, sendo uma das áreas consideradas para trabalho futuro. Depois de terminada a parte experimental desta tese, surgiram soluções técnicas (como as usadas pela Hotjar<sup>91</sup>) que permitem, mesmo só com acesso ao conteúdo do *site* alojado do servidor, interceptar e registar tudo o que é enviado para o navegador do cliente, resolvendo o problema dos *sites* com conteúdo dinâmico.

<sup>91</sup> Hotjar (*All-in-one Analytics & Feedback*): <https://www.hotjar.com/>

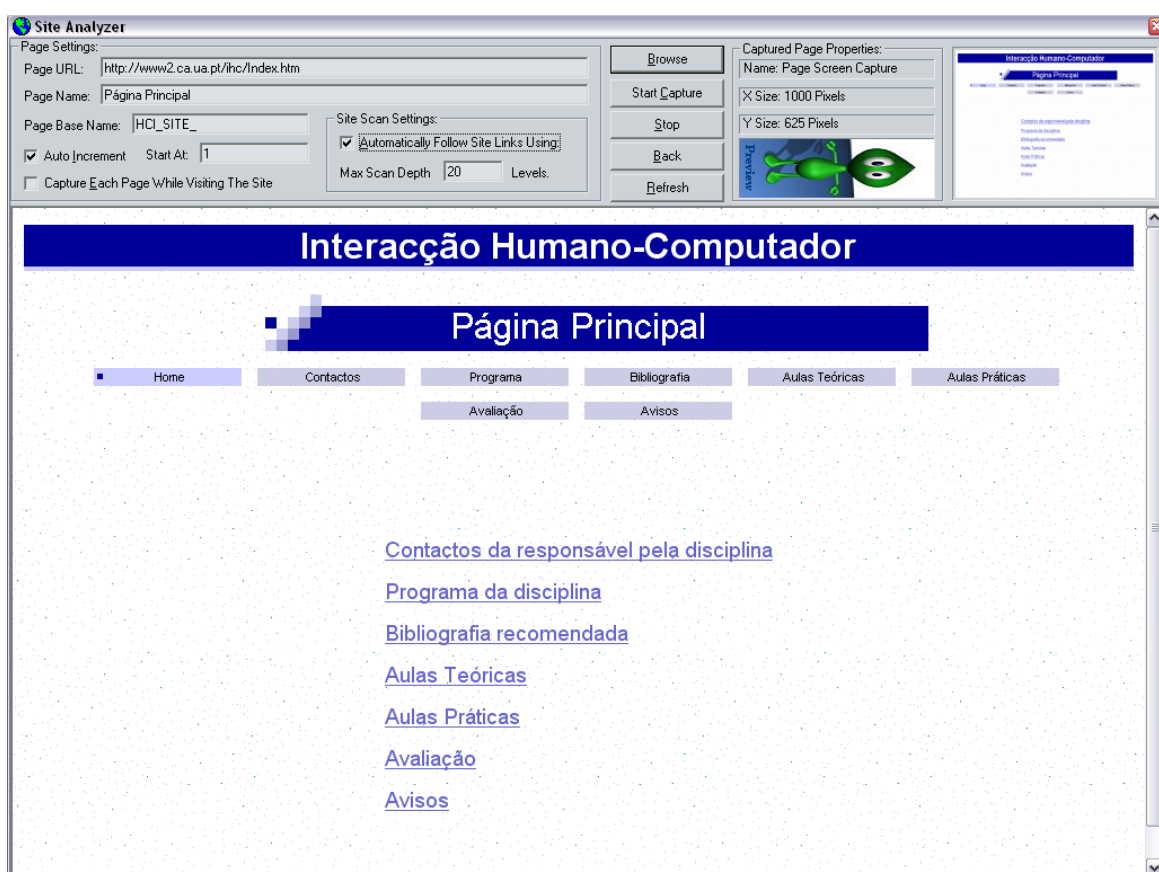


## 6.4. Analisador

Este bloco é responsável pela construção de uma parte significativa do *Atlas de Referência* descrito no modelo conceptual (Figura III-2) apresentado na secção 2.1 (Atlas Digital de referência) do Capítulo III.

A Figura IV-31 ilustra uma parte da interface da aplicação *Analisador*. Esta aplicação permite a captura da informação referente a um *site*. Depois de especificar a página inicial, a aplicação permite explorar, de forma automática ou manual, a estrutura do *site*, seguindo os *links* existentes em cada página. É possível navegar pelo *site* como se estivesse a utilizar um navegador normal, podendo capturar as páginas pretendidas. Alternativamente, o programa pode percorrer todos os *links* existentes em cada página, de forma recursiva, capturando todas as páginas encontradas.

A informação capturada consiste na identificação da página, no código HTML, imagens, etc. constituinte da mesma e numa imagem do resultado apresentado ao utilizador pelo navegador. Para maior flexibilidade, a aplicação permite especificar o tamanho da janela do navegador.



**Figura IV-31:** Interface da aplicação *SiteAnalyzer*. Na parte superior podem controlar-se os detalhes do que se pretende analisar. Na parte inferior é possível utilizar a página, como se de um navegador normal se tratasse. Imagem obtida com a versão 2 do protótipo, publicada em (F. L. Zamfir, 2007).

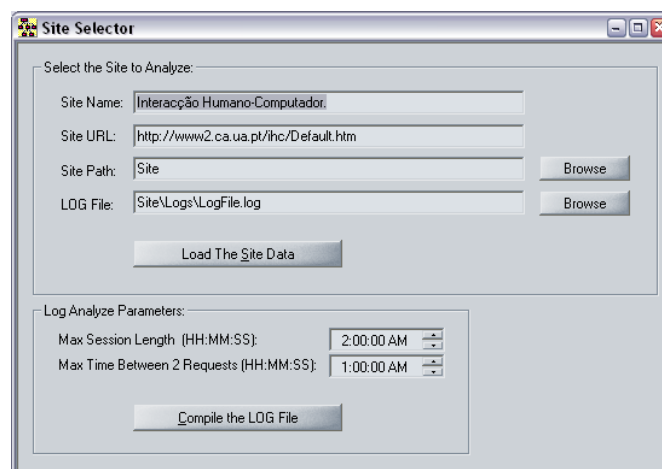
## 6.5. Compilador

Os ficheiros de *log* produzidos pelos servidores de *web* contêm informação, em estado bruto, sobre os pedidos efetuados a esses servidores. A partir destes ficheiros é possível extrair informação referente ao comportamento de interação infocomunicacional que decorreu num determinado *site* para cada uma das sessões de acesso. A tarefa do bloco *Compilador* consiste, exatamente, em filtrar e extrair a informação de interação considerada relevante destes ficheiros de *log*.

A Figura IV-32 mostra a interface da aplicação *Compilador*. Aqui especificam-se alguns detalhes do *site* em causa e do ficheiro de *log* a processar.

Uma das tarefas mais importantes deste bloco tem a ver com a identificação de sessões de utilizador. Um utilizador liga-se a um *site*, interage com o *site*, percorrendo várias páginas, porventura recolhendo e inserindo informação. Informação sobre quem, quando e como interage com um *site* é de importância capital quando se pretende estudar um *site*.

Neste sentido, a interface da aplicação permite especificar alguns parâmetros (como a localização do *site* e do ficheiro de *log*, bem como a indicação das restrições temporais para a deteção das sessões) usados pelo algoritmo de deteção dos limites das sessões.



**Figura IV-32:** Interface da aplicação *Compilador*. Imagem obtida com a versão 2 do protótipo, publicada em (F. L. Zamfir, 2007).

Do ponto de vista da aplicação desenvolvida, a informação de uma sessão é compilada numa entidade de sessão, onde são guardados vários detalhes como a página de entrada e de saída, a hora de início e fim e respetiva duração, entre outros detalhes, como a ilustrada na Figura IV-33.

Session	
+ SessionID	: int
+ SiteID	: Object
+ UserID	: Object
+ Cookie	: Text
+ StartURL	: Text
+ EndURL	: Text
+ StartTime	: DateTime
+ EndTime	: DateTime
+ Length	: TimeStamp
+ Requests	: Object
+ ReferrerURL	: Text

Figura IV-33: Entidade de sessão.

Finalmente, toda a informação é enviada para a base de dados, onde será armazenada de forma apropriada.

## 6.6. Visualizador

Este bloco é o culminar de todo o sistema e representa um dos instrumentos de *feedback* mais relevantes para um qualquer processo de análise da interação infocomunicacional. Toda a informação recolhida, processada e enviada para a base de dados será usada neste módulo que permite aplicar métodos de análise e visualização da informação.

Este bloco será abordado da seguinte forma: em primeiro lugar será apresentado o modelo utilizado para o seu desenvolvimento, seguidamente abordar-se-ão os aspetos da interface base e no final serão descritos vários métodos de análise e visualização implementados.

### 6.6.1. Modelo

A aplicação *Visualizador* foi desenvolvida tendo em mente o modelo de camadas ilustrado na Figura IV-34.

Em traços gerais, toda a estrutura assenta na *Estrutura de Suporte*, descrita atrás, e tem uma ligação estreita com a base de dados.

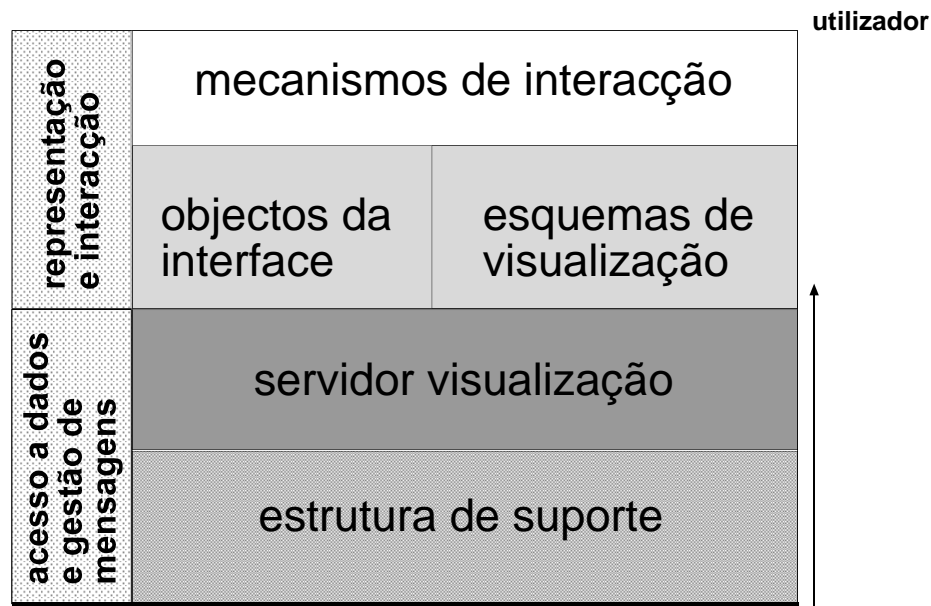


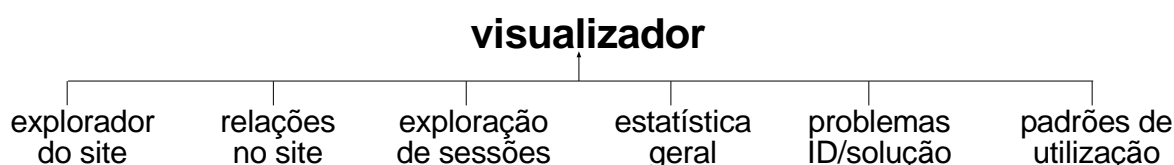
Figura IV-34: Modelo das camadas da aplicação Visualizador. Adaptado de (F. L. Zamfir, 2007).

Faz-se, a seguir, uma breve apresentação dos principais blocos presentes no modelo descrito na Figura IV-34:

- *Estrutura de Suporte* representa a ligação entre a plataforma de desenvolvimento e os componentes funcionais da aplicação. A *Estrutura de Suporte* facilita o acesso à informação guardada na base de dados, fornecendo um conjunto de ferramentas que permitem pesquisar e manipular a informação.
- *Servidor de visualização* representa uma componente cuja funcionalidade básica consiste na criação e remoção de esquemas de visualização, para gerir a sua necessidade de acesso à informação, comunicação de dados e sincronizações. Esta componente alimenta os esquemas de visualização com informação, obtida da base de dados pela *Estrutura de Suporte* e zela pelo envio de mensagens entre todos os métodos de visualização ligados. O nome *Servidor* tem a ver com o facto de, basicamente, as funções deste módulo estarem associadas a uma funcionalidade com uma topologia do tipo cliente-servidor.
- *Objetos da Interface* representam um conjunto de controlos de interface de utilizador funcionais, constituindo uma interface básica para a aplicação. Estes objetos podem ser janelas, menus, barras de ferramentas, controlos *ActiveX*, navegadores *web*, etc.. Todos estes objetos estão intimamente ligados e relacionados em termos de funcionalidade.
- *Esquemas de Visualização* representam um conjunto de componentes, em que cada componente combina um conjunto de controlos de interface de utilizador e de mecanismos de interação, implementando, em conjunto, um método de visualização ou uma combinação de métodos. A existência deste bloco justifica-

se devido aos seus aspetos funcionais e lógicos. Do ponto de vista da implementação, um esquema de visualização é representado como um objeto da interface de utilizador que pode ser adicionado e manipulado onde e quando se pretender, desde que o *Servidor de Visualização* esteja funcional. Os métodos de visualização são o objetivo central desta aplicação. Devido à sua importância, ser-lhes-á dedicada uma secção, mais abaixo.

- *Mecanismos de Interação* representam todos os meios com que se pode interagir com a aplicação, dando relevância à informação daí obtida. Cada interação com o sistema provoca alterações ao estado da aplicação bem como alterações de contexto, no aspeto ou conteúdo da representação atual, etc..



**Figura IV-35:** Subcomponentes da aplicação Visualizador. Adaptado de (F. L. Zamfir, 2007).

A aplicação *Visualizador* pode ser dividida em seis subcomponentes, como ilustrado na Figura IV-35, e descritas sumariamente a seguir:

- **Explorador do Site** – Local onde o utilizador tem a oportunidade de explorar *offline* a estrutura e conteúdo do *site* selecionado. Cada página é armazenada na base de dados e cada vez que o utilizador seleciona um *link* de uma página, a página correspondente é carregada da base de dados assim como todas as suas relações com outros documentos e páginas do *site*.
- **Relações no Site** – Apresenta a informação relacionada com a estrutura do *site*. Podem ser mostradas simultaneamente várias representações da informação, selecionadas a partir do conjunto de esquemas de visualização disponíveis. Pode ser apresentada, não só informação estrutural, mas também informação estatística, com padrões de utilização. São disponibilizados vários tipos de interação com a informação que permitem revelar um detalhe da informação ou enaltecê-la (como por exemplo, mostrar os caminhos possíveis entre duas páginas selecionadas).
- **Exploração de Sessões** – Local onde a informação de sessão pode ser observada usando várias representações ao mesmo tempo. Adicionalmente, podem ser observadas várias sessões simultaneamente. Informação estatística da utilização pode também ser representada aqui. Os mecanismos de interação apresentados permitem a utilização de filtros sobre a informação representada.
- **Estatística Geral** – Permite observar a informação estatística relacionada com a utilização de um *site*. Estão também disponíveis diferentes métodos de

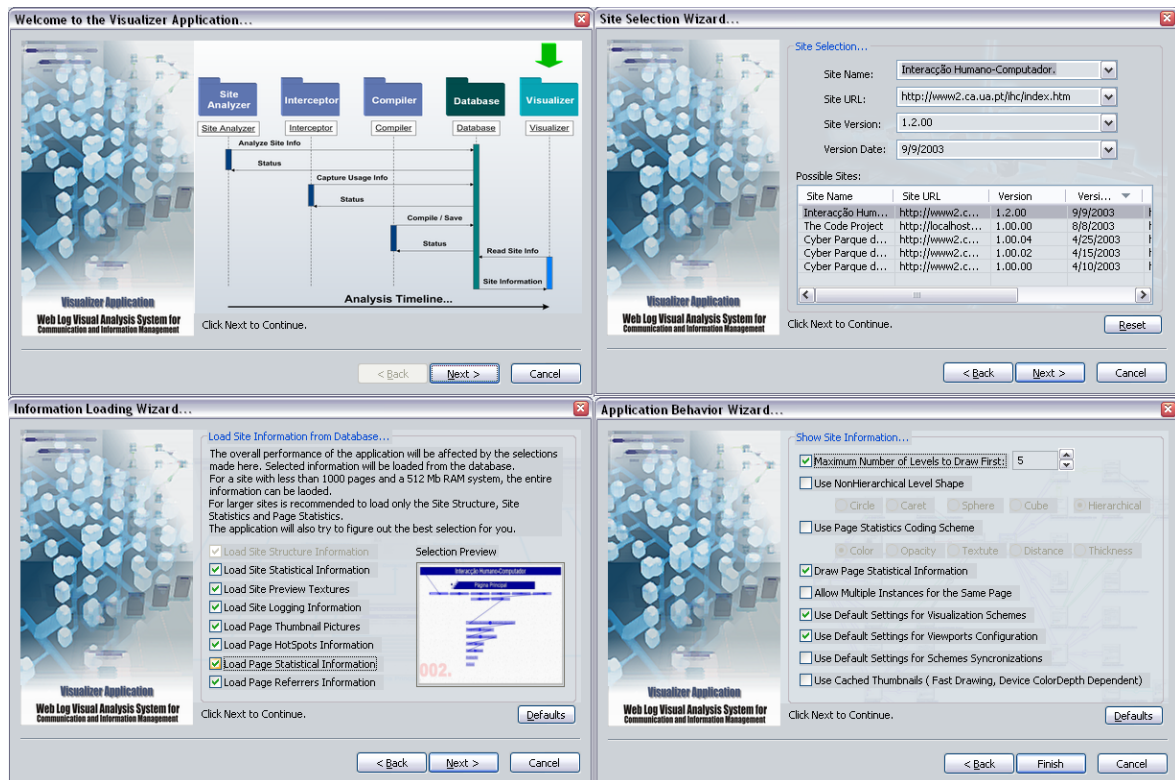
visualização para representar a informação. Adicionalmente, esta componente tem um alto nível de integração com a base de dados.

- **Problemas / Identificação de Solução** – Tem como objetivo mostrar informação sobre problemas de usabilidade relacionados com a estrutura do *site* (como *links* quebrados) e o resultado da aplicação de métricas para a avaliação da conformidade com linhas mestras de usabilidade específicas (como a utilização de cores, tamanho de fontes, organização da página, etc.). Comparando a informação atual com o paradigma associada à secção do *site*, a ferramenta poderá ser capaz (ou não) de fazer sugestões de soluções para o(s) problema(s) identificado(s).
- **Padrões de Utilização** – Permite avaliar a informação relacionada com a utilização de um *site* por um utilizador ou grupo de utilizadores, onde padrões podem ser enaltecidos, relacionada com o fluxo de informação ou tráfego de comunicação.

## 6.6.2. Interface de utilizador base

Nesta secção apresentam-se os elementos base da interface de utilizador do Visualizador. Há uma característica importante que merece destaque: a coordenação entre as múltiplas vistas (*multiple coordinated views*) apresentadas em cada momento.

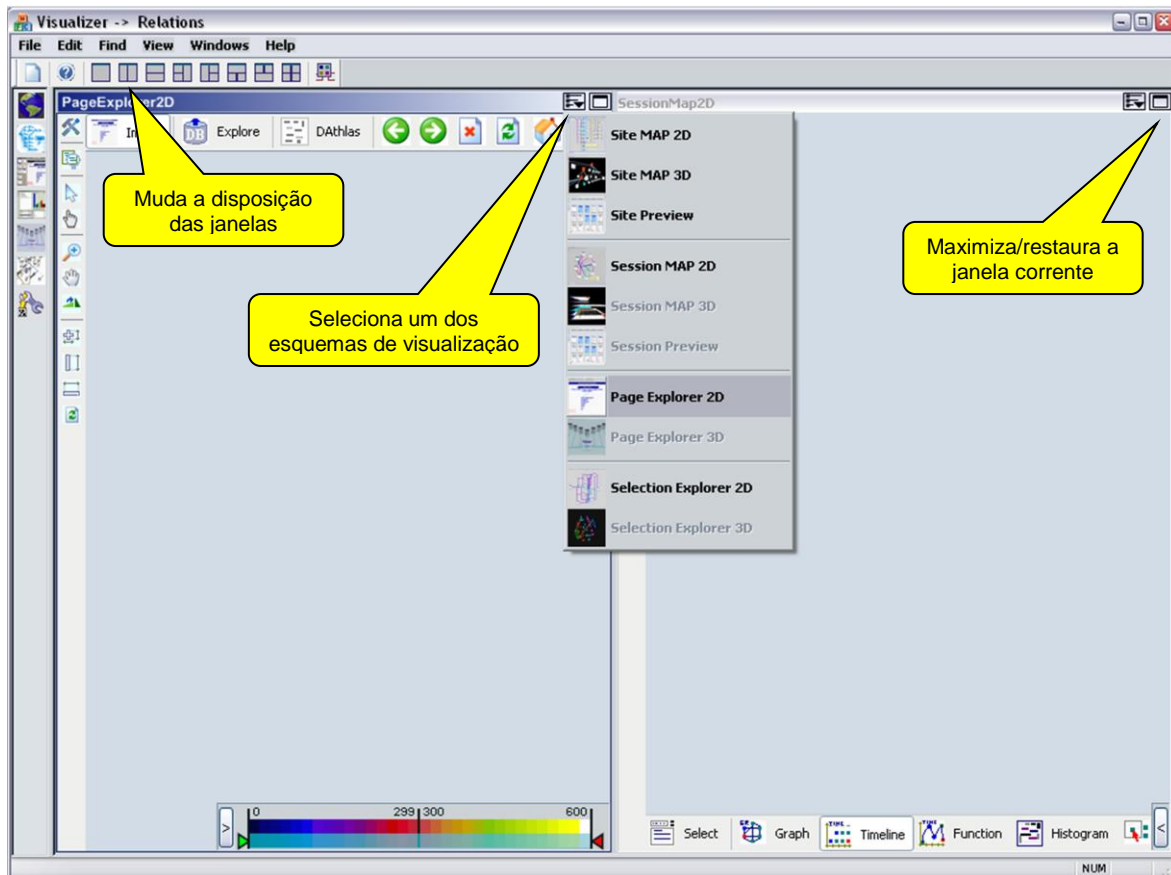
Quando se inicia a aplicação é apresentado ao utilizador um conjunto de janelas de diálogo iniciais (*Wizard*), como as presentes na Figura IV-36. Começa por apresentar uma janela com alguma informação sobre a aplicação. Na janela seguinte é selecionado o *site* que se pretende estudar, sendo apresentados alguns detalhes como o nome do *site*, o *URL*, a versão e data de cada umas das possibilidades. Na terceira janela pode seleccionar-se qual a informação a ser carregada, desde os vários tipos de informação relacionada com o *site* até aos detalhes de cada página. Finalmente, na última janela podem ajustar-se alguns detalhes relativos ao comportamento da aplicação no momento de apresentar informação.



**Figura IV-36:** Passos do processo de configuração inicial do Wizard (da esquerda para a direita e de cima para baixo). Imagem obtida com a versão 2 do protótipo, publicada em (F. L. Zamfir, 2007).

Depois das janelas de diálogo iniciais aparece a janela principal que tem o aspeto que se pode ver nas Figura IV-37, Figura IV-38 e Figura IV-39.

A Figura IV-37 apresenta algumas possibilidades de manipulação da interface de utilizador do Visualizador. É possível mudar a disposição das janelas com os esquemas de visualização seleccionando um ícone. É ainda possível, para uma dada janela, seleccionar o esquema de visualização. As janelas podem, em qualquer momento, ser maximizadas para tirar partido da área do ecrã e restauradas para o seu tamanho original.

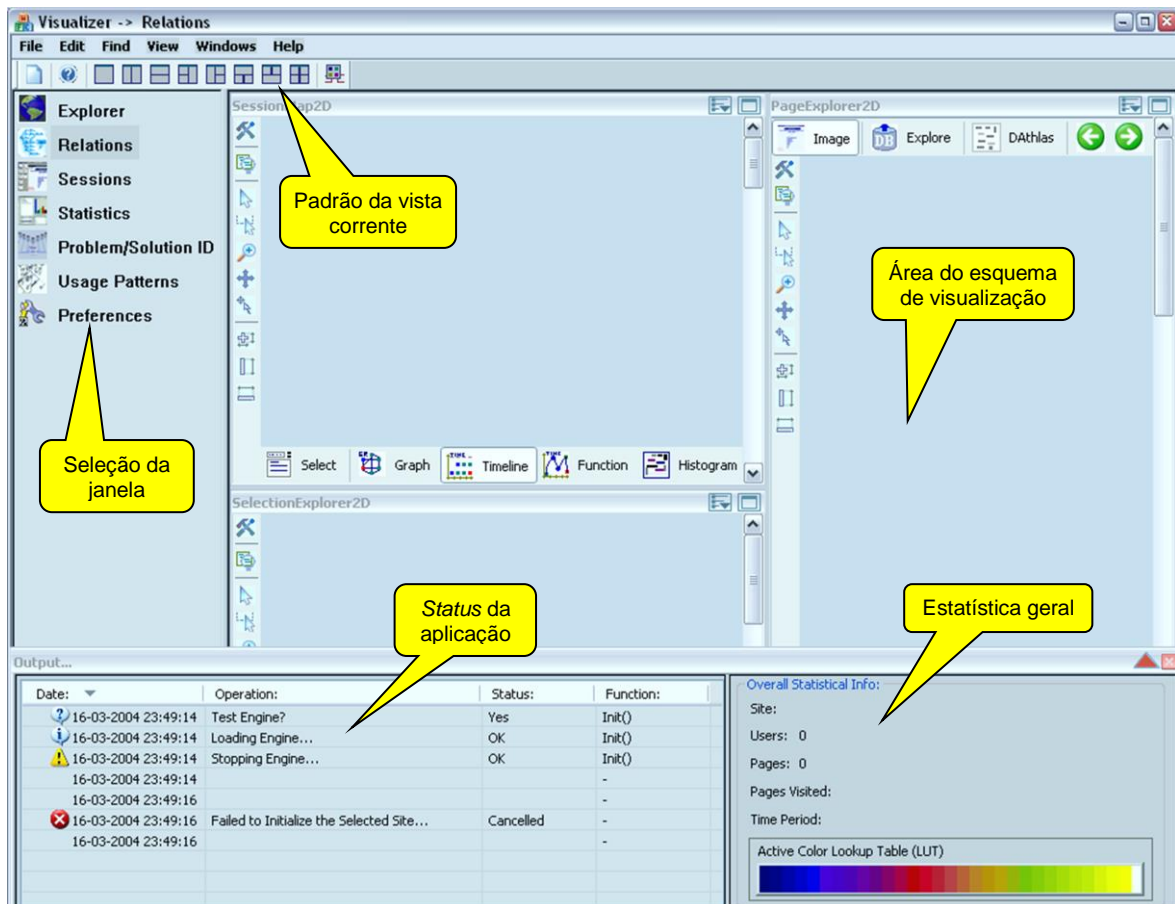


**Figura IV-37:** Possibilidades de manipulação da interface de utilizador do Visualizador. Imagem obtida com a versão 2 do protótipo, adaptado de (F. L. Zamfir, 2007).

A Figura IV-38 descreve o modelo da interface gráfica da aplicação com o utilizador. Podem ser identificadas algumas áreas que se passam a enumerar:

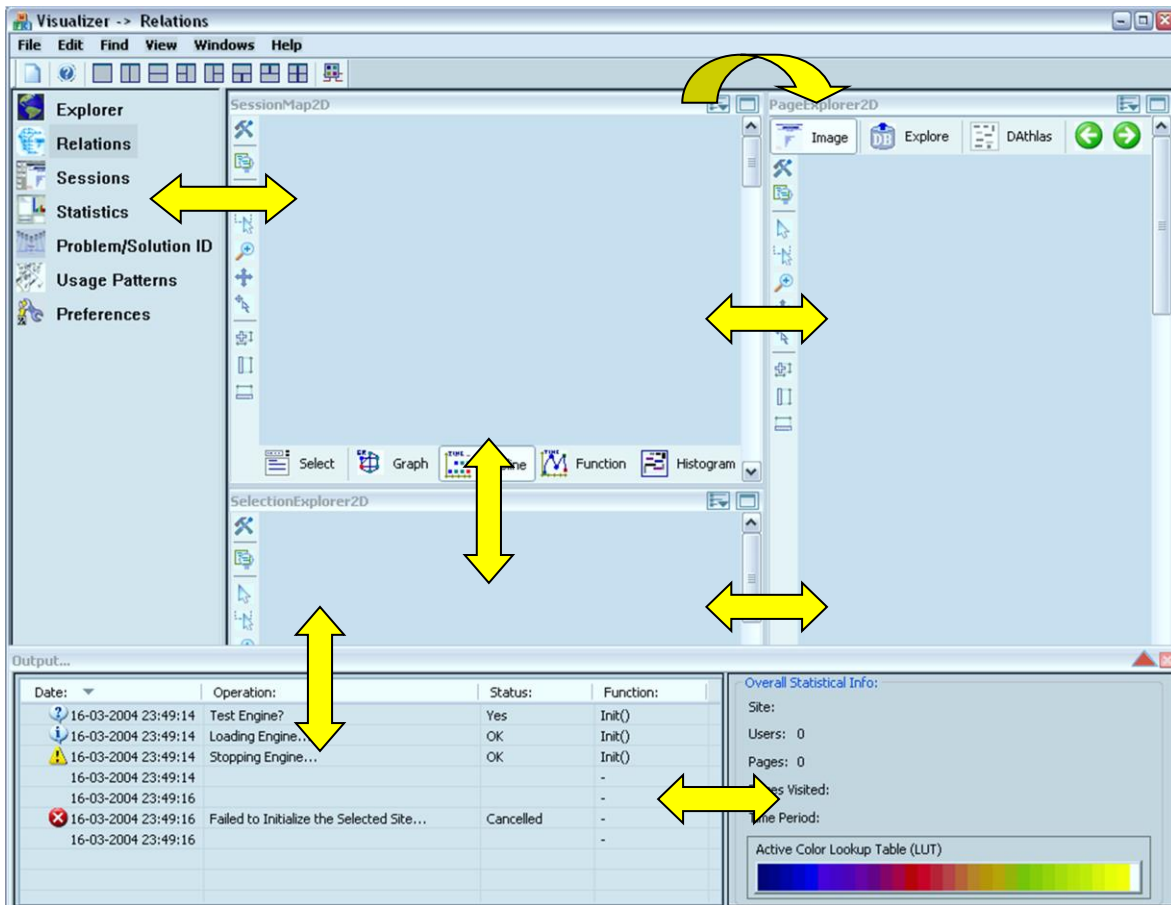
- Ícones que permitem a escolha do padrão da disposição das janelas para a vista corrente;
- Uma secção onde se pode seleccionar a janela com o conjunto de visualizações relacionadas com a escolha;
- Podem observar-se três áreas com outros tantos esquemas de visualização;
- Na parte inferior pode ainda ver-se uma área com o estado (*status*) da aplicação onde é indicado o sucesso, ou não, de cada operação realizada;
- E uma área com alguma informação estatística geral do *site* em estudo.





**Figura IV-38:** Modelo da interface de utilizador do Visualizador. Imagem obtida com a versão 2 do protótipo, adaptado de (F. L. Zamfir, 2007).

Na Figura IV-39 são ilustradas algumas das possibilidades de manipulação dos vários objetos presentes na interface da aplicação. A maioria dos objetos podem ser redimensionados. De notar que, na maior parte dos casos, ao redimensionar-se um objeto (uma subjanela, por exemplo) está a alterar-se o tamanho do ou dos objetos adjacentes. Pode dizer-se que, nesses casos, se está a manipular as linhas divisórias entre os objetos.



**Figura IV-39:** Ilustração da flexibilidade dos objetos constituintes da interface de utilizador do Visualizador. Imagem obtida com a versão 2 do protótipo, adaptado de (F. L. Zamfir, 2007).

Seguidamente serão descritos com mais detalhe alguns pormenores da interface apresentada acima.

Na Figura IV-40 é possível escolher o padrão da disposição dos esquemas de visualização no interior da janela principal. Para alterar o padrão basta escolher um dos ícones correspondentes ao número de disposição pretendidos.



**Figura IV-40:** Padrão da vista corrente.



**Figura IV-41:** Seleção tipo visualização.

Na Figura IV-41, para uma dada categoria de esquema de visualização, pode alterar-se o tipo de visualização. Pode escolher-se um dos tipos:

- *Select* – Este tipo de visualização permite uma mais fácil seleção dos vários objetos constituintes de uma página *web*;
- *Graph* – Apresenta as várias páginas sob a forma de um gráfico 2D / 3D, dependendo do esquema de visualização;
- *Timeline* – Usa uma linha temporal, como a usada na Figura IV-52.
- *Function* – Usa um gráfico em que num eixo é representada uma linha temporal e no outro eixo a posição na estrutura hierárquica do *site*. Pode ver-se um exemplo na Figura IV-56.
- *VisualSpace* – Representação de informação visual como a ilustrada na Figura IV-57, abaixo.



**Figura IV-42:** Barra ferramentas.

Os esquemas de visualização são acompanhados de uma barra de ferramentas ilustrada na Figura IV-42. Os ícones desta barra de ferramentas permitem (de cima para baixo):

- Abertura de janela de diálogo que permite alterar alguns detalhes sobre o esquema de visualização atual, como a apresentada na parte direita da Figura IV-49.
- Permite ver detalhes sobre um objeto.
- Escolha do cursor normal.
- Escolha do cursor de seleção.
- Ferramenta de ampliação / redução (*zoom*) da vista atual.
- Permite deslocar (translação – *drag*) da vista atual.
- Roda 90° a vista atual.
- Permite redimensionar na vertical e horizontal.
- Permite redimensionar na vertical.
- Permite redimensionar na horizontal.
- Redesenha a vista corrente.

Nos esquemas de visualização relacionados com a representação de uma página, pode ser encontrada a barra de ferramentas da Figura IV-43.



**Figura IV-43:** Barra ferramentas *web*.

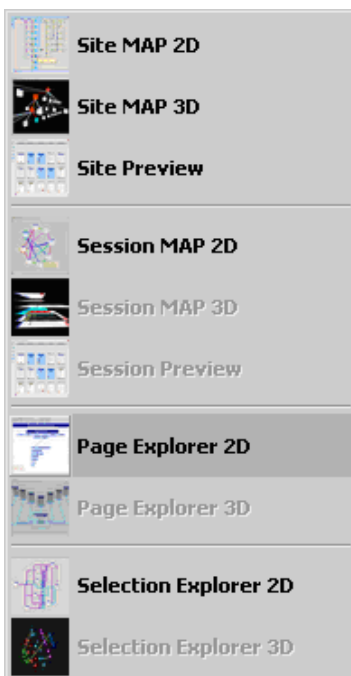
Os ícones desta barra têm as seguintes funções (da esquerda para a direita):

- *Image* – Mostra uma imagem da página armazenada na base de dados.

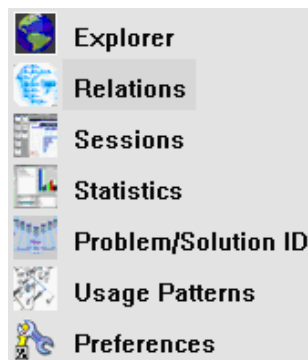
- *Explore* – A página é construída a partir do código armazenado na base de dados (como se de um servidor *web* se tratasse).
- *DAtlas* – Mostra informação adicional proveniente do atlas digital armazenado na base de dados.

Os ícones seguintes são comuns aos programas de navegação na *internet*:

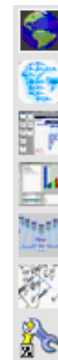
- Muda para a página anterior.
- Muda para a página seguinte.
- Interrompe o carregamento da página.
- Recarrega a página corrente.
- Vai para a página de entrada.



**Figura IV-44:** Menu seleção método visualização.



**Figura IV-45:** Seleção da janela com um conjunto de esquemas de visualização relacionados.



**Figura IV-46:** Seleção da janela com um conjunto de esquemas de visualização relacionados – vista só com os ícones.

A Figura IV-44 ilustra o menu que permite selecionar o método de visualização pretendido. Note-se que alguns métodos não são selecionáveis porque não estavam ativos no momento em que a figura foi capturada.

Os métodos de visualização que é possível selecionar serão descritos na secção seguinte (Métodos de visualização).

As Figura IV-45 e Figura IV-46 permitem a seleção da janela com um conjunto de visualizações relacionadas. A aplicação pode apresentar esta seleção, como aparece na Figura IV-45, ou só os ícones correspondentes, como ilustrado na Figura IV-46.

Cada janela contém um conjunto de métodos de visualização, relacionados com:

- *Explorer* – Exploração do *site*.
- *Relations* – Relações entre as páginas do *site*.
- *Sessions* – Representação de uma ou mais sessões de utilizador.
- *Statistics* – Estatísticas de utilização.
- *Problem/Solution ID* – Identificação e solução de problemas.
- *Usage Patterns* – Padrões de utilização.

As Figura IV-47 e Figura IV-48 ilustram a utilização de tabelas de cor presentes em alguns esquemas de visualização.

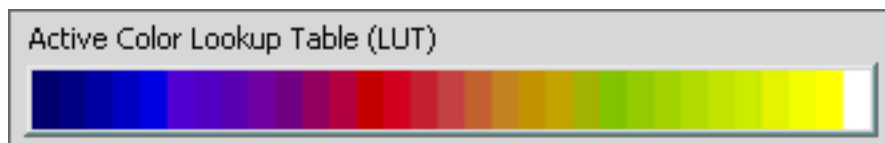


Figura IV-47: Vista da tabela de cor ativa.

A Figura IV-47 mostra a tabela de cor ativa, que em alguns casos pode ser manipulada como ilustrado na Figura IV-48, deslocando a seta verde para ajustar o limite inferior e a seta vermelha para ajustar o limite superior.

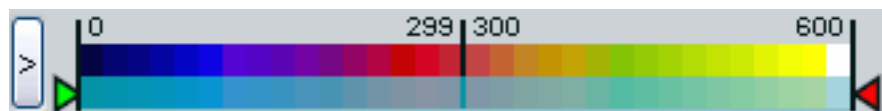


Figura IV-48: Manipulação da tabela de cor ativa.

Para concluir, é apresentado o aspeto geral da subjanela com um método de visualização parte esquerda da Figura IV-49 e do lado direito a janela de diálogo com as propriedades que permite ajustar alguns detalhes de um método de visualização.

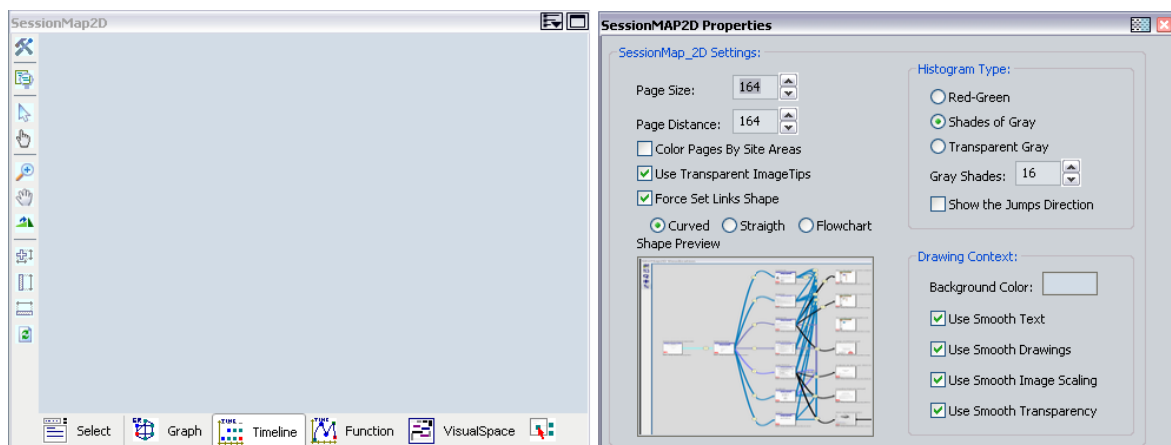


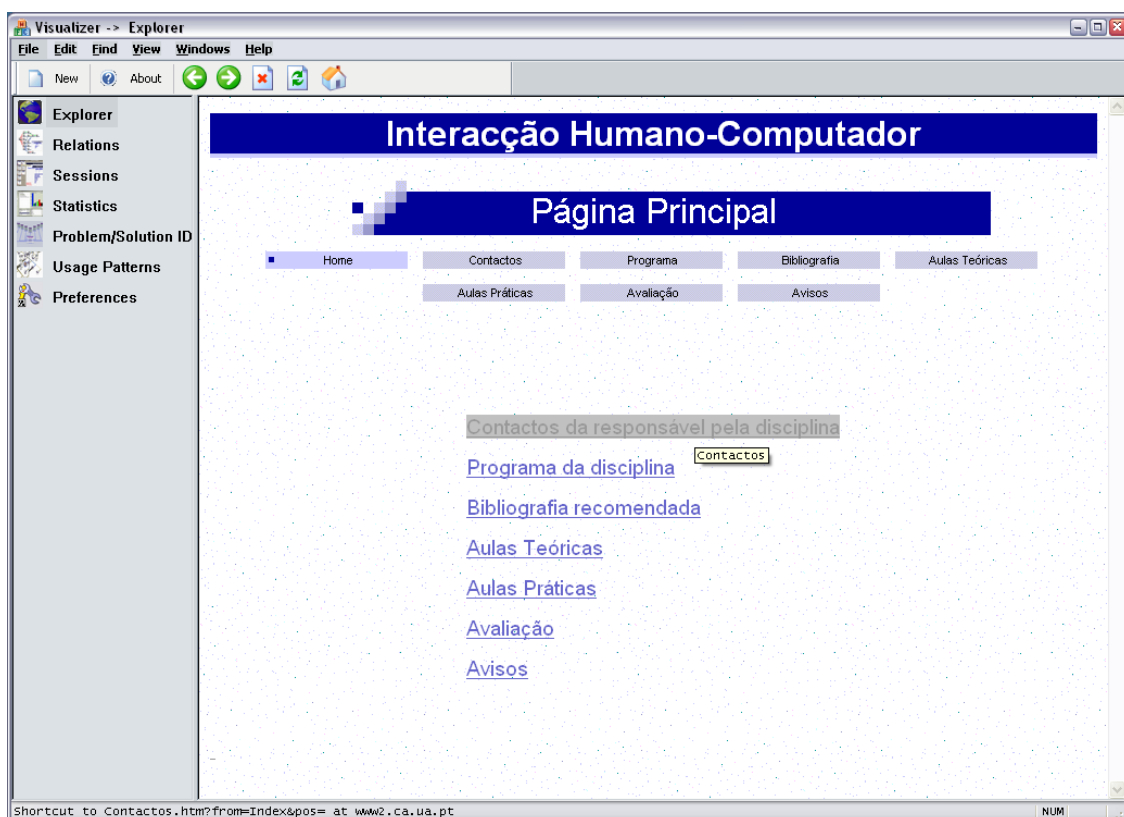
Figura IV-49: Detalhes de alguns aspetos da interface de utilizador. Imagem obtida com a versão 2 do protótipo, retirado de (F. L. Zamfir, 2007).

### 6.6.3. Métodos de visualização

Durante o desenvolvimento deste protótipo da aplicação foram implementados alguns métodos de visualização para teste. A lista de métodos implementados está longe de esgotar os métodos possíveis de integrar um modelo infocomunicacional desta natureza. Os métodos escolhidos apresentam-se como alguns dos mais adequados para demonstrar a relevância do modelo inerente a este projeto.

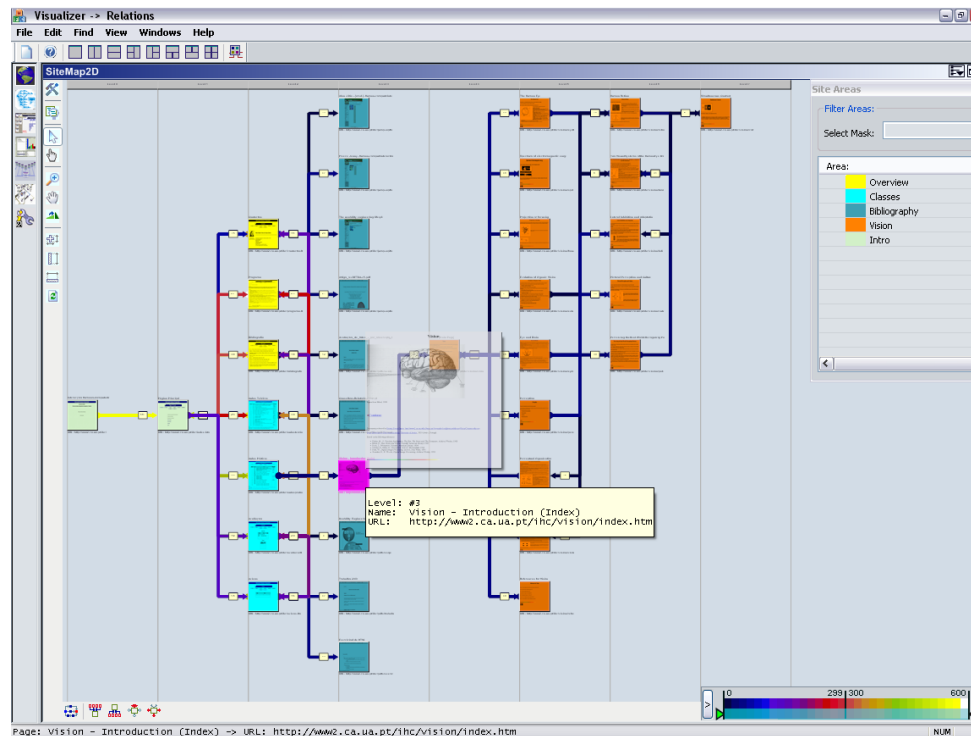
O método mais simples será, talvez, a possibilidade de explorar um *site* da mesma maneira que um utilizador normal o faz quando usa um navegador *web* no seu computador. O método ilustrado na Figura IV-50 reproduz essa funcionalidade, podendo a informação mostrada estar contida na base de dados ou descarregada em tempo real (*live*) do servidor do *site* em estudo.

Note-se que, para todos os efeitos, se está a usar o navegador *Internet Explorer*, embutido na aplicação, com a maioria das suas funcionalidades ativas. É possível mudar de página seguindo um *link*, observar o conteúdo dinâmico e até observar as *tooltips* presentes em algumas páginas.



**Figura IV-50:** Vista da componente *Explorer* (Exploração do *Site*). Imagem obtida com a versão 2 do protótipo, retirado de (F. L. Zamfir, 2007).

A estrutura do *site* é uma informação importante quando se pretende estudá-lo. A Figura IV-51 mostra uma vista dessa estrutura.



**Figura IV-51:** Representação da informação relacional das páginas de um *site web*. Imagem obtida com a versão 2 do protótipo

As páginas do *site* estão organizadas de forma hierárquica, sendo cada página ilustrada com uma pequena representação em miniatura do seu conteúdo. As páginas estão codificadas com um código de cores, correspondendo cada cor a uma secção do *site*.

As ligações entre as páginas são representadas por linhas respeitando uma tabela de cor que codifica a intensidade com que cada ligação foi utilizada. De forma a melhor explorar essa vertente é possível manipular a tabela de cor, escolhendo uma diferente ou simplesmente manipulando os seus limites inferior e superior.

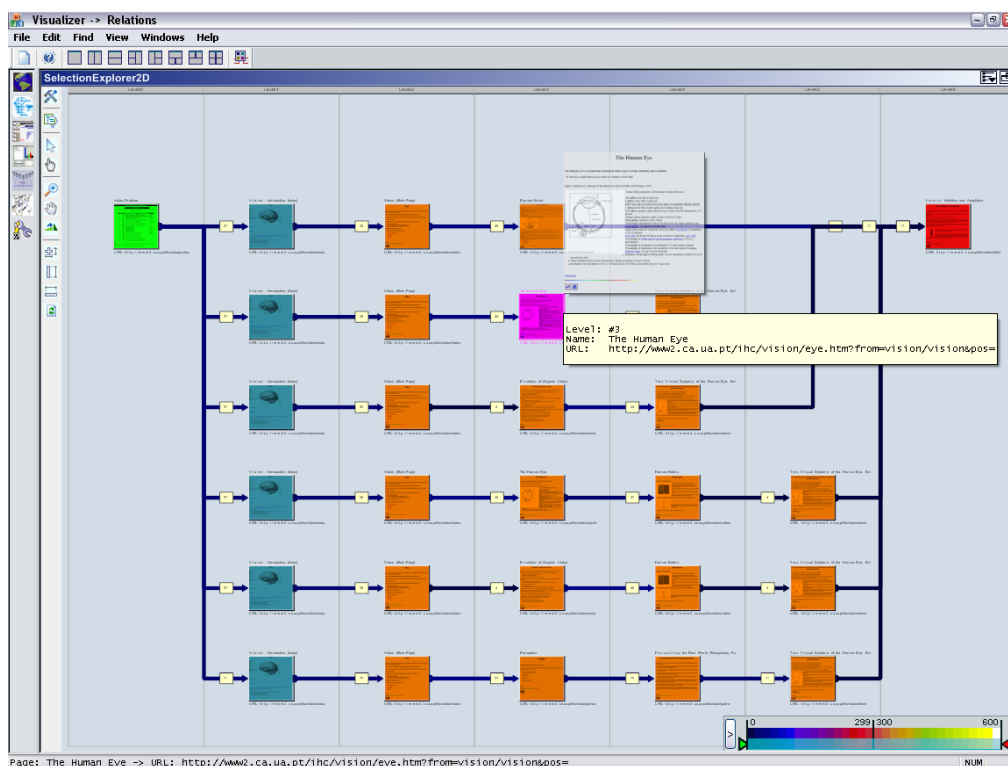
A forma de representação das ligações entre cada página pode ser alterada, do estilo diagrama de fluxo, para linhas retas que ligam diretamente as páginas envolvidas ou para linhas curvas. A interface de utilizador da janela de diálogo onde é possível fazer essa escolha está ilustrada na parte direita da Figura IV-49.

De forma a aumentar a quantidade de informação que se pode obter a partir da exploração do *site*, há alguns detalhes que são mostrados quando se interage com os objetos apresentados no ecrã. Quando se coloca o cursor do rato sobre a miniatura de uma página, é mostrada uma representação em ponto maior da página (recorrendo a semitransparência para não se perder o contexto do *site*) e alguns detalhes da página como o nível, o nome e o *URL* da página numa *tooltip*.

Quando o cursor está sobre uma linha de ligação também é mostrada uma *tooltip* contendo a lista de ligações (no caso de haver mais que uma linha sobreposta) com a indicação de detalhes, como o *URL*, da página de origem e de destino de cada ligação.

A Figura IV-52 mostra uma representação dos percursos possíveis quando se pretende navegar de uma dada página de partida até uma dada página destino. Dependendo da complexidade do *site* e das suas ligações poderão haver mais ou menos possibilidades. No exemplo ilustrado há seis percursos possíveis, tendo o mais curto, três páginas e os mais longos, 5 páginas intermédias.

No entanto, poderia acontecer que o percurso mais curto fosse uma ligação direta entre a página de partida e a página de destino (bastando para tal que a página de partida tivesse um *link* para página de destino) e a mais longa poderia ter uma dimensão infinita (situação muito comum quando há várias páginas com *links* entre si, levando a situações de referência circular).

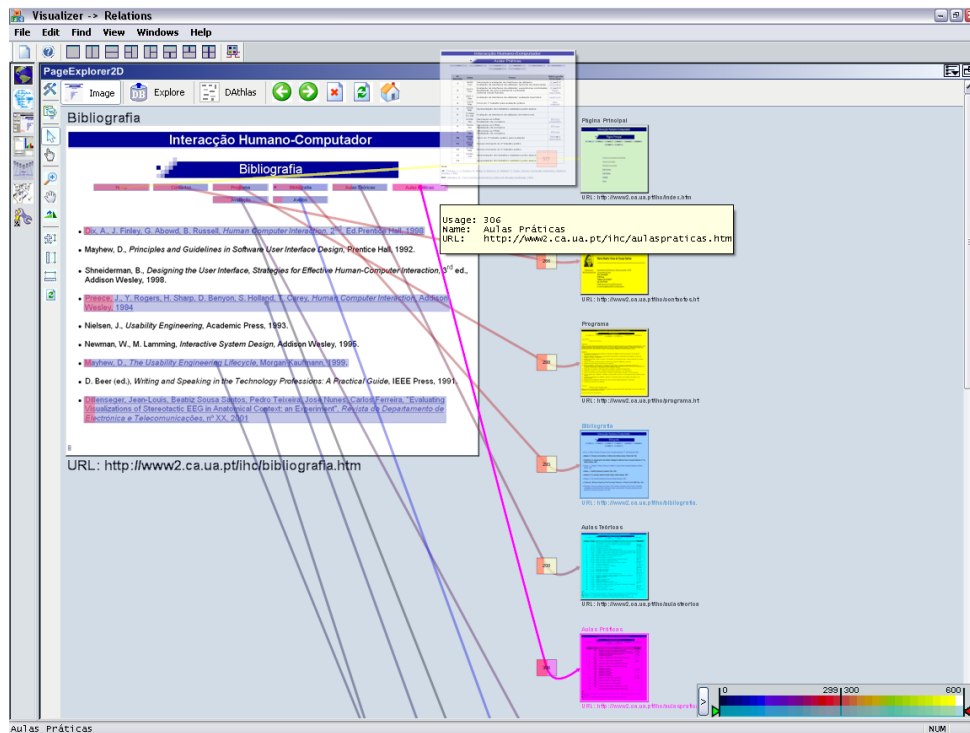


**Figura IV-52:** Representação dos percursos possíveis para se chegar a uma página destino a partir de uma página de partida. Imagem obtida com a versão 2 do protótipo, retirado de (F. L. Zamfir, 2007).

Este esquema de visualização tem aspetos comuns com o ilustrado acima na Figura IV-51. A classificação das páginas usando um esquema de cores, a utilização de uma tabela de cor para as ligações e a informação adicional apresentada quando se coloca o cursor sobre uma página ou ligação são semelhantes aos apresentados atrás.

O método apresentado acima e ilustrado pela Figura IV-50 pode ser complementado por uma quantidade de informação adicional. Os *hotspots* (áreas contendo *links*) são dos detalhes mais importantes quando se consulta uma página. De maneira a poder estudar a sua distribuição e utilização, foi desenvolvido o esquema de visualização ilustrado na Figura IV-53.





**Figura IV-53:** Localização do *Hotspots* numa página e informação hipermédia relacionada. Imagem obtida com a versão 2 do protótipo.

Neste esquema pode ver-se uma representação da página em estudo com a identificação dos seus *hotspots*. A área ocupada por cada *hotspot* está identificada por um retângulo semitransparente. Cada *hotspot* está ligado a uma representação miniatura da página a que diz respeito, na parte direita do ecrã.

A taxa de utilização de cada *hotspot* pode ser observada olhando para proporção relativa entre as cores rosa e azul do retângulo que o identifica na página. É também apresentada num pequeno quadrado associado à linha que liga a área do *hotspot* à miniatura da página associada. Neste quadrado além da proporção entre as cores, é ainda indicado um valor numérico com o número de vezes que esse *hotspot* foi utilizado.

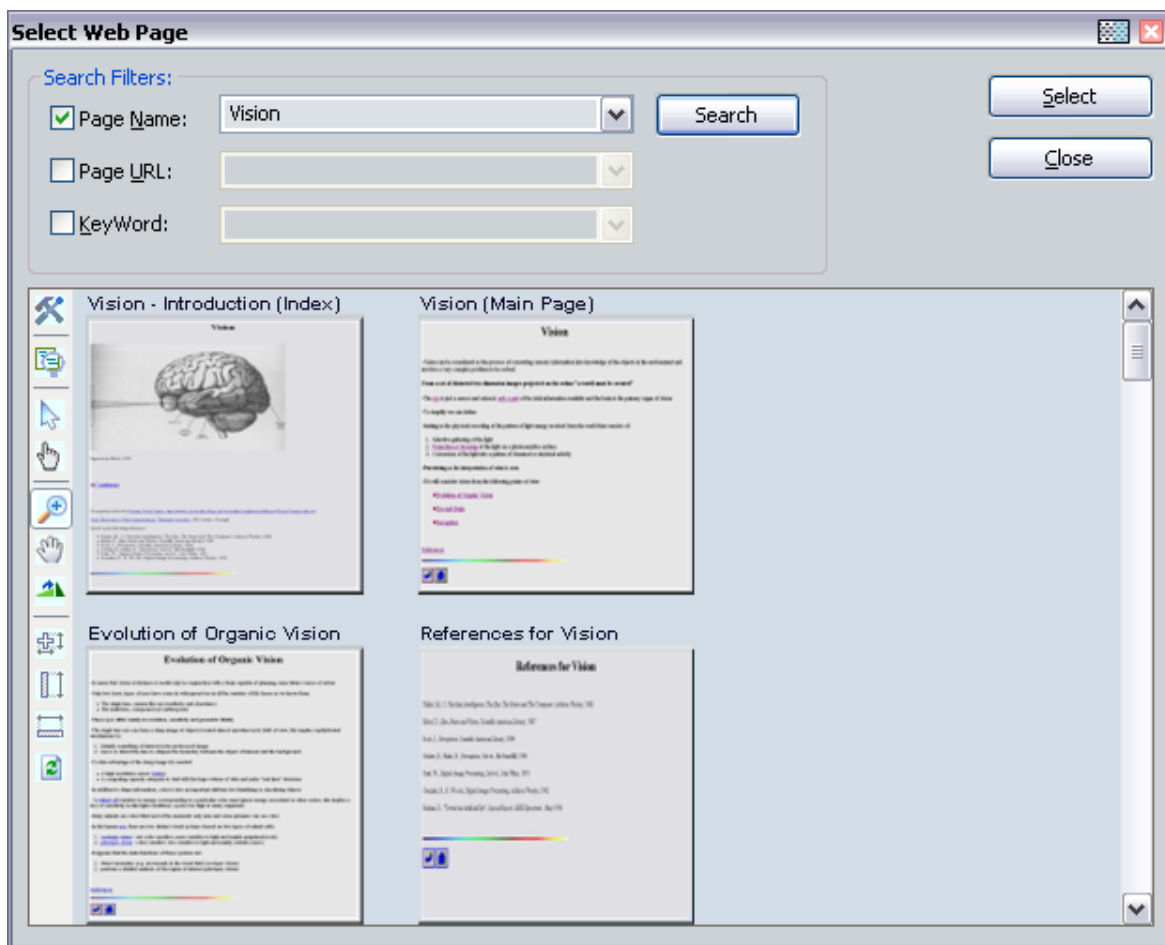
As cores das linhas, que ligam as áreas dos *hotspots* às miniaturas das páginas, respeitam uma tabela de cor que identifica o número de vezes que cada *hotspot* foi utilizado.

Tal como nos casos descritos cima, também aqui é possível manipular a tabela de cor utilizada. De igual modo, quando se coloca o cursor do rato sobre um *hotspot*, linha ou miniatura de página, obtêm-se detalhes sobre o *hotspot* em causa, como o número de vezes que foi utilizado, o nome e o URL da página para o qual aponta.

De maneira a ser possível representar uma miniatura de cada página associada a cada *hotspot* da página, e tendo em consideração que o espaço do ecrã poderá não ser suficiente, recorreu-se à utilização de uma *scroll bar*, aumentando dessa maneira a área útil disponível.

Para terminar, quando se seleciona um *hotspot* ou a miniatura da página correspondente, clicando com o rato sobre os mesmos, a página atual é atualizada para essa página, com a respetiva atualização da restante informação apresentada, não só neste esquema como em todos os outros que estejam a ser apresentados nesse instante.

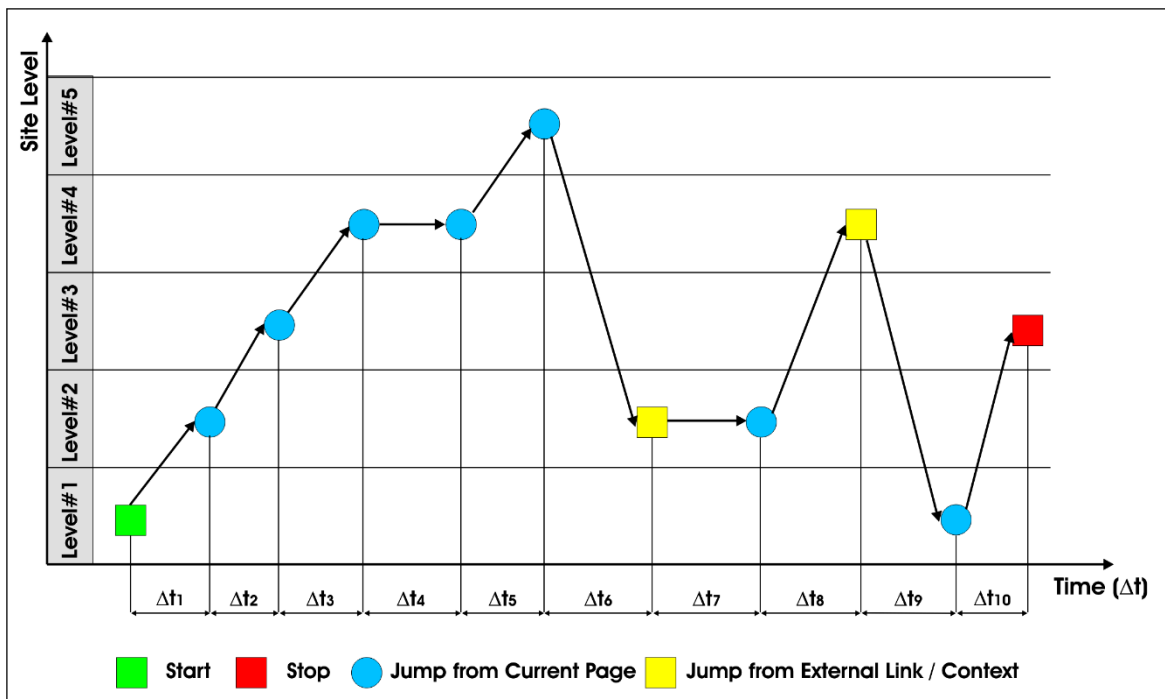
Num *site* de maiores dimensões, há situações em que é útil existir uma ferramenta que permita encontrar rapidamente a página pretendida. Para o efeito foi criada a caixa de diálogo apresentada na Figura IV-54, onde é possível especificar alguns detalhes dos critérios a utilizar no processo de pesquisa, e obter os resultados que satisfazem esses critérios. Se necessário, podem alterar-se os critérios de pesquisa para refinar os resultados obtidos. Os resultados são apresentados sob a forma de uma miniatura da representação de cada página identificada com o seu título. Depois de identificada a página pretendida, basta selecioná-la com o rato, clicar no botão *Select*, a janela fechará e a página escolhida estará selecionada no esquema de visualização, passando a ser a página ativa.



**Figura IV-54:** Representação dos resultados de uma pesquisa. As páginas aparecem desenhadas em miniatura (*thumbnail*). Imagem obtida com a versão 2 do protótipo.

Outro aspeto importante do estudo de um *site* tem a ver com as sessões de utilização dos utilizadores. A forma como um utilizador navega no *site* durante uma visita pode fornecer pistas importantes sobre potenciais problemas na estrutura do *site* ou de uma página.

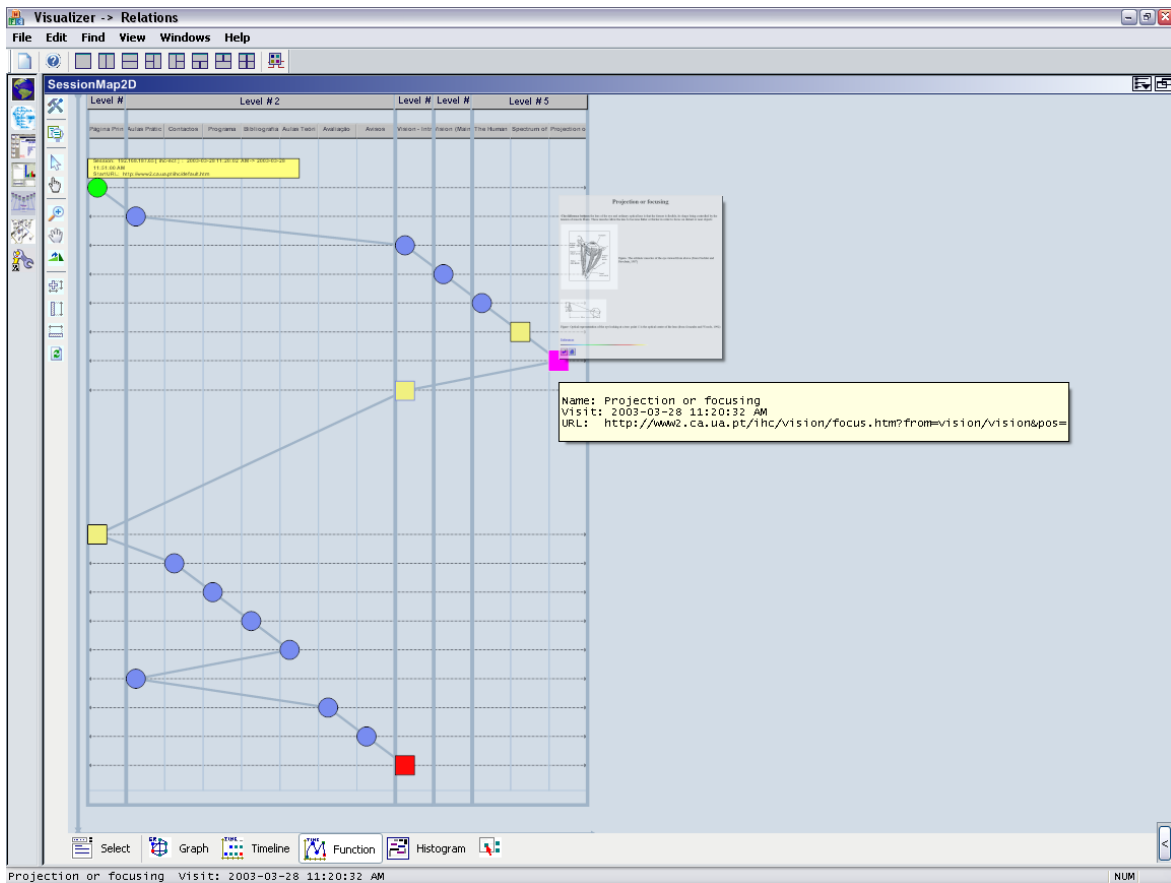
Dado que a maioria dos *sites* estão organizados de uma forma hierárquica, é importante dar visibilidade ao nível (representando a profundidade na estrutura do *site*) de cada página visitada, bem como ao tempo despendido entre cada página. A vista presente na Figura IV-55 ilustra uma possibilidade de representação dessa informação.



**Figura IV-55:** Representação dos tempos envolvidos no percurso pela estrutura em forma de árvore de um *site* por um utilizador durante uma sessão. Imagem obtida com a versão 2 do protótipo, retirado de (F. L. Zamfir, 2007).

Na Figura IV-55 cada página de uma sessão é representada por um polígono colorido, em que a forma e cor representam situações concretas. O eixo vertical identifica o nível dentro da estrutura em árvore do *site* e no eixo horizontal, a distância entre cada página e proporcional ao tempo que o utilizador demorou entre cada página.

A solução implementada acabou por ser ligeiramente diferente, como se pode ver na Figura IV-56. Esta ilustra a implementação efetivamente utilizada, que não é muito diferente da representação utilizada como ponto de partida (Figura IV-55). Aqui é apresentada uma sessão de um utilizador real. O tempo decorrido entre cada página está presente no eixo vertical e o nível (na estrutura do *site*) está patente no eixo horizontal.



**Figura IV-56:** Representação de uma sessão de utilizador (linha temporal vs. posição da página na estrutura em árvore do site). Imagem obtida com a versão 2 do protótipo.

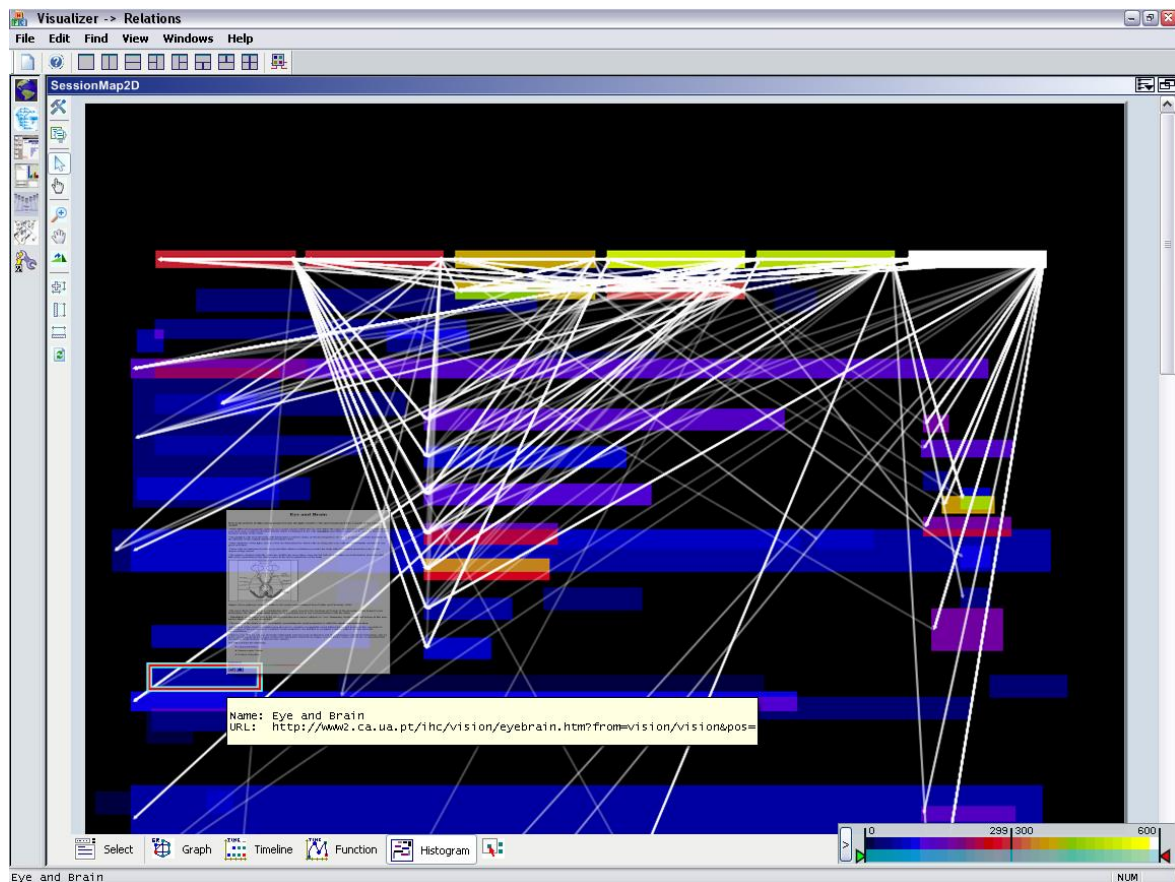
No eixo horizontal não está só o nível, mas também cada página de cada nível. Contudo, veio a verificar-se que esta não foi uma boa opção, na medida em que a distância neste eixo deixou de ser proporcional à profundidade no site, uma vez que os vários níveis podem ter um número distinto de páginas, correspondendo a uma largura também distinta. Uma possibilidade de alteração poderia passar por atribuir uma largura constante a cada nível, simplificando dessa maneira a representação cartesiana da navegação.

Este esquema de visualização apresenta ainda alguma informação adicional, à semelhança dos esquemas descritos anteriormente. Quando se coloca o rato sobre um dos polígonos representado uma página da sessão, é apresentada informação relativa a essa página como uma representação da página (usando semitransparência) e uma tooltip com o nome, a hora da visita e o URL da página em causa.

A seleção de uma página torna-a a página ativa, com a consequente atualização nos restantes métodos de visualização ativos.

Neste esquema de visualização existe a preocupação de construir uma imagem representativa do campo visual de interação original e efetivo. Cada ponto desta imagem tem propriedades cumulativas consequência dos padrões de utilização sob estudo. Desta maneira é possível analisar a coerência da distribuição das áreas dos *hotspots* nas várias páginas que constituem um *site* ou sessão(ões) específica(s) de um *site*.

Este esquema, ilustrado pela Figura IV-57, procura dar resposta a perguntas como: Será que existe uma estrutura de interação bem definida no *site*? Com consequência para uma interação preponderante numa determinada área?



**Figura IV-57:** Representação sobreposta dos *Hotspots* e da utilização das ligações. Imagem obtida com a versão 2 do protótipo.

Este esquema usa um *buffer* de acumulação, em que em cada ponto contém contagem do número páginas com *hotspot* nessa área. Seguidamente o valor é convertido numa cor de acordo com a tabela escolhida. Desta maneira é relativamente fácil identificar as zonas com uma maior densidade de *hotspots* que correspondem às zonas de interação. A tabela de cor pode ser seleccionada e manipulada da mesma maneira que nos métodos anteriores.

Adicionalmente, pode ativar-se a visualização da relação entre as áreas utilizadas, sendo que uma maior densidade das linhas indica que as áreas ligadas são utilizadas com muita frequência em conjunto na mesma sessão.

Quando se coloca o cursor do rato sobre uma área é mostrada uma representação da página relacionada com o *hotspot* em causa bem como um *tooltip* contendo informação adicional como o nome e o *URL* dessa página.

É ainda possível seleccionar uma área, o que identifica a página associada a esse *hotspot* como sendo a página ativa, com a consequente atualização nos restantes métodos de visualização ativos.

## 7. Algumas conclusões

Neste capítulo foi descrito o desenvolvimento do protótipo. Começou-se por analisar os vários tipos de protótipo, tendo-se optado por um protótipo evolutivo. De seguida foram analisadas algumas soluções de prototipagem, tendo-se concluído que as plataformas de prototipagem rápida não conseguiam responder a todas as necessidades do trabalho desta tese. Neste cenário, seria necessário escolher um plataforma de desenvolvimento e uma linguagem de programação, tendo-se começado por C / C++ em ambiente MS Visual Studio.

A opção por um protótipo evolutivo desenvolvido em C/C++ acabou por demonstrar as suas vantagens em alguns momentos. Contudo, simultaneamente, também deixou transparecer as suas limitações. A partir de determinado momento, tornou-se obvio que a simples evolução já não era suficiente para responder às necessidades, pelo que seria necessário uma alteração profunda na arquitetura utilizada, obrigando à criação de uma nova versão. Apesar disto foi possível reutilizar muito código já produzido.

Acabaram por ser desenvolvidas 3 versões (0, 1 e 2), sendo que a primeira (versão 0) foi, sobretudo, exploratória. Para cada versão desenvolvida, foram descritos os seus módulos constituintes, com especial destaque para a última versão (2) do protótipo, onde se recorreu a uma base de dados e se desenvolveu um *framework* para dar um maior suporte e facilidade a evoluções futuras e, dessa maneira, assegurar uma maior longevidade do protótipo.

As várias versões do protótipo desenvolvido foram sujeitas a um processo de avaliação que está descrito na secção 2 do Capítulo V, abaixo.

# **Capítulo V.**

## **ESTUDO DE APLICABILIDADE DO PROTÓTIPO**

# 1. Introdução

Neste capítulo será apresentada a aplicação do protótipo, durante os seus vários estágios de desenvolvimento iterativo, em contextos de registo infocomunicacionais reais numa configuração de 3 versões (0, 1 e 2). Em primeiro lugar será apresentada a avaliação das várias versões do protótipo, depois serão descritos os repositórios de informação e a forma como foram utilizados durante as três fases principais do desenvolvimento do protótipo. De seguida serão apresentados alguns resultados da utilização do protótipo em cada estágio, bem como a contribuição que teve para o seu desenvolvimento. Na parte final, será feita uma pequena reflexão crítica, tendo em mente os resultados obtidos.

Nos capítulos anteriores foram apresentados métodos de análise e visualização de informação, e a sua integração num protótipo funcional. Foram ainda descritos os vários estágios do desenvolvimento desse protótipo, utilizando repositórios de informação específicos que permitissem a validação dos conceitos e procedimentos referentes a cada estágio. A prova definitiva para qualquer aplicação é, sem dúvida, a sua utilização em contexto real. Assim, foi com naturalidade que se utilizaram repositórios de informação (ficheiros de *log*, páginas *web*, etc.) obtidos de sistemas a funcionar em contextos reais.

De maneira a testar um protótipo recorre-se com frequência a estudos que podem ser divididos em duas categorias:

- Estudo de caso – situação onde se recorre a um sistema que está já em funcionamento, sem qualquer intervenção sobre o mesmo, observando e inquirindo sujeitos em contexto de uso real;
- Estudo de laboratório – em que se cria um sistema, especialmente elaborado, de forma a permitir um controlo mais profundo das condições da experiência, contudo descontextualizado do ambiente de uso real, ainda que este possa ser simulado.

Regra geral, a utilização de protótipos não está livre de limitações. Em contexto de investigação, estes usam-se para validar conceitos, procedimentos ou funcionalidades propostos e desta forma legitimar ou aferir os pressupostos colocados no início do trabalho. O risco da utilização de um protótipo no domínio da investigação prende-se com o facto de se poder concentrar no estudo de um aspeto particular com um investimento reduzido, contudo perdendo a perspetiva holística / global do problema.



Neste capítulo, depois de se abordar a avaliação das várias versões do protótipo, são realçados dois aspetos fundamentais que se prendem com a sua utilização, com consequência direta para os resultados da investigação a saber:

- Identificação e utilização de repositórios de informação que levem à aferição para um sistema que se consolide como plataforma de investigação.
- Que a aferição e desenvolvimento do protótipo leve à acreditação de um sistema capaz de analisar problemas de usabilidade em domínio WWW.

Apesar das suas limitações, a utilização de protótipos foi de importância crucial durante o desenvolvimento deste trabalho. A utilização de repositórios de informação real, requereu um esforço adicional, mas permitiu detetar falhas e limitações mais precocemente, apresentando um balanço custo vs. benefício francamente positivo.

## **2. Avaliação das várias versões do protótipo**

Como referido no Capítulo IV, acima, houve necessidade de testar os protótipos para, entre outras coisas, identificar problemas e encontrar soluções. Só desta maneira seria possível evoluir de forma segura no desenvolvimento iterativo de várias versões de protótipos com o objetivo de convergir para uma aplicação beta final.

Com este propósito foram conduzidas avaliações às várias evoluções do protótipo (versão 0, 1, e 2) relatadas com bastante detalhe no artigo (Santos et al., 2004) publicado pelos elementos da equipa envolvida na elaboração do trabalho desta tese. A seguir descrevem-se estas avaliações com mais algum detalhe.

### **2.1. Introdução**

Apesar de existir um elevado número de métodos de visualização de informação – alguns bons exemplos estão incluídos em (Card et al., 1999) e (Bederson & Shneiderman, 2003) – e muitos sistemas que os utilizam para visualizar grandes quantidades de informação, há, comparativamente, poucos estudos que façam a avaliação destes métodos e sistemas (S. Carpendale, 2008). Isto deve-se, em grande medida, à complexidade inerente a este tipo de avaliação. No entanto, com o amadurecimento do campo da visualização de informação, um grande número de métodos tem vindo a ser desenvolvidos e aplicados na resolução de problemas reais, tornando-se verdadeiramente importante saber se estes métodos realmente são eficazes.

Adicionalmente, é necessário saber em que circunstâncias estes devem ser utilizados, como são comparáveis e para que tarefas melhor se adequam.

Apesar de nos últimos anos se ter evoluído bastante, a avaliação da visualização de informação ainda tem vários desafios por resolver. Na literatura podem encontrar-se alguns trabalhos que utilizam explicitamente a avaliação no processo de *design* de sistemas de visualização, avaliando sistemas específicos e métodos de visualização, bem como comparando formas de visualização alternativas. Alguns exemplos podem ser encontrados em (Hix et al., 1999), (North & Schneiderman, 2000), (Sebrechts et al., 1999), (Wiss et al., 1998), (Kobsa, 2001) ou (Barlow & Neville, 2001).

Existem mesmo alguns autores que reconhecem a sua importância e que desenvolvem esforços para desenvolver abordagens mais sistemáticas ao complexo problema da avaliação em visualização de informação (C. S. Freitas et al., 2002), (Brath, 1999), (Grinstein et al., 2004) ou o BELIV – Workshop sobre avaliação de Visualização de Informação<sup>92</sup>, que se realiza de dois em dois anos desde 2006. Apesar de muito interessantes, este tipo de trabalhos deparam-se frequentemente com algumas dificuldades (S. Carpendale, 2008): todos os que pretendem avaliar métodos e sistemas de visualização de informação ainda se debatem com muitas limitações no que respeita a técnicas e métodos específicos para conduzirem uma avaliação.

Uma abordagem utilizada com alguma frequência consiste em adaptar os mais conhecidos e amplamente utilizados métodos de engenharia da usabilidade (Jakob Nielsen, 1993), tendo em consideração as especificidades das técnicas e sistemas que se pretendem avaliar.

No presente trabalho pretende-se descrever uma forma de avaliação, com o objetivo de contribuir para o desenvolvimento de uma aplicação baseada na visualização, que pretende ajudar os gestores de sistemas de informação a entender como o sistema de informação da sua instituição/organização está a ser utilizado.

Esta secção foi desenvolvida tendo por base o trabalho publicado em (Nunes et al., 2003). Foi tido em consideração o trabalho desenvolvido por Freitas e tal (C. S. Freitas et al., 2002) para estruturar a avaliação, tal como se adaptaram algumas técnicas de teste de usabilidade para assim se obter respostas por parte dos utilizadores e usar essa informação para redesenhar a aplicação.

Nesta avaliação decidiu-se utilizar técnicas empíricas e experimentais como meio de sugerir, clarificar, refinar e gerar novas ideias que possam vir a ser futuramente exploradas em condições mais controladas (Carlsmith et al., 1976).

---

<sup>92</sup> BELIV – Workshop sobre avaliação de Visualização de Informação: <http://beliv.cs.univie.ac.at/>

## 2.2. Características gerais da aplicação

Como referido anteriormente, o objetivo deste trabalho consiste em ajudar os gestores de sistemas de informação das instituições/organizações a entender como os seus sistemas infocomunicacionais internos estão a ser utilizados. A presente proposta relaciona-se com a visualização de grandes quantidades de informação recolhidas dentro das instituições/organizações a partir, principalmente, da análise da estrutura do *site web* e da informação de utilização (*user-driven analytics*) armazenada (obtida, quer durante a normal utilização do *site*, quer através de experiências controladas).

Algumas perguntas podem ser colocadas, cujas respostas poderão trazer luz à problemática:

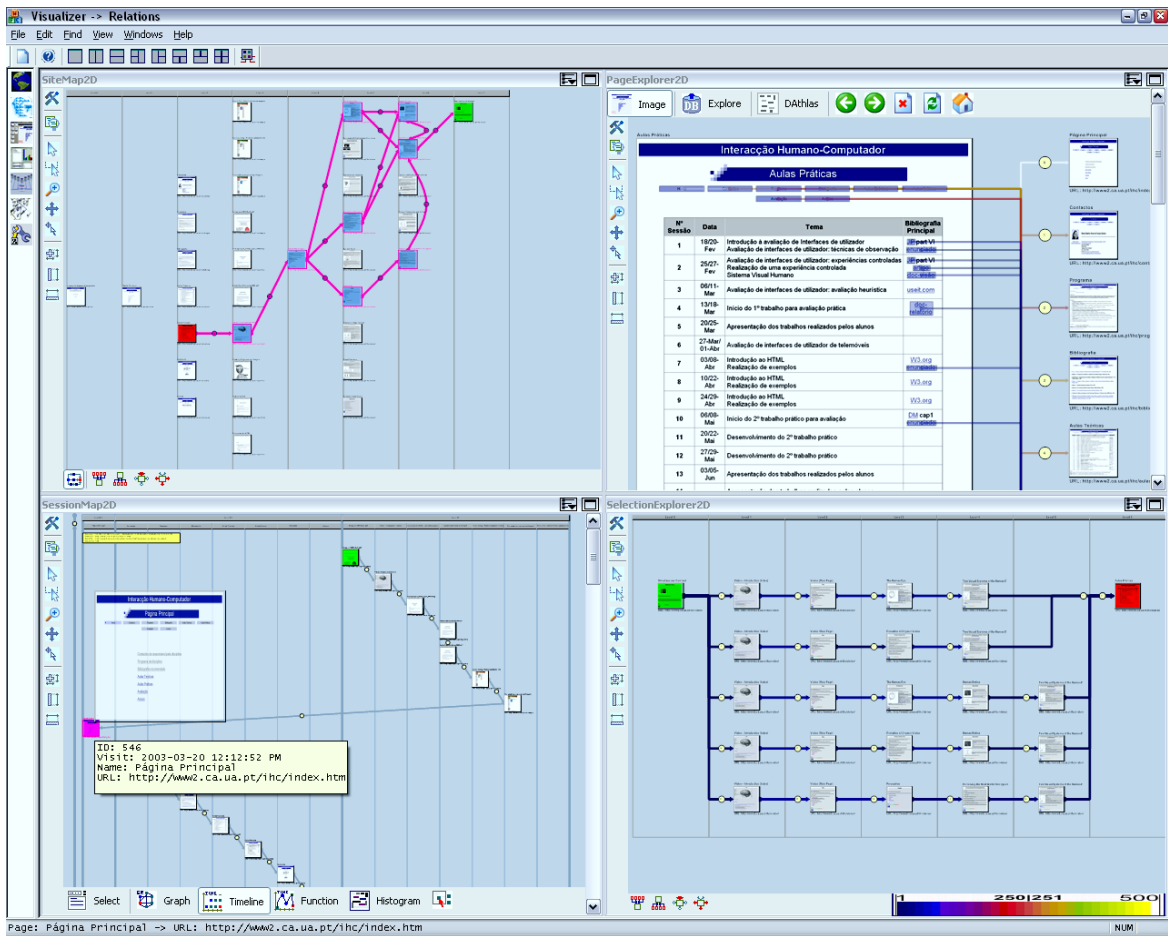
- Como é utilizado o *site* (páginas visitadas, seu tempo de visionamento, caminho utilizado, etc.)?
- Quem utiliza o *site*?
- Quais são os interesses dos utilizadores?
- Que informação estatística pode ser obtida através dos ficheiros de *log*?
- Quais são as áreas com problemas?
- Que problemas de usabilidade podem ser identificados?

Tenta-se responder a algumas destas questões proporcionando uma aplicação de visualização capaz de captar, compilar e apresentar informação relativa à estrutura de um *site web* e aos padrões de utilização (poderão ser encontrados mais pormenores em (Nunes et al., 2003) (F. Zamfir, Nunes, Teixeira, Santos, et al., 2004)).

Alguns esquemas de visualização, que podem ser utilizados interactivamente pelos utilizadores, foram desenvolvidos com o objetivo de representar a informação de uma forma compreensível. Foi possível identificar esquemas de visualização que, em princípio, pudessem ser úteis no contexto da aplicação para o apoio à análise visual da estrutura do *site* e da coerência do design da interface de utilizador, tal como, a localização e estatísticas de utilização das hiperligações da página. Estes esquemas de visualização são complementares, permitindo, cada um, uma forma diferente de representar a informação, o que deverá torná-la mais adequada para servir alguns dos objetivos do utilizador. Assim, estes esquemas de visualização deverão ser usados em combinação uns com os outros em vez de isoladamente.

Para que o utilizador possa usufruir das vantagens destes esquemas de visualização, existe a possibilidade de optar por:

- quantos esquemas de visualização pretende utilizar simultaneamente;
- quais os múltiplos padrões de visualização que pretende utilizar;
- onde no ecrã é mostrada a representação obtida usando cada esquema;



**Figura V-1:** Utilização simultânea, em sincronia, de quatro esquemas de visualização diferentes. Imagem obtida com a versão 2 do protótipo.

As diferentes vistas mostradas em simultâneo ao utilizador são síncronas, ou seja, qualquer interação com os objetos em qualquer dos esquemas de visualização reflete-se nas restantes vistas (exemplo: um objeto que seja selecionado será interpretado por todos os esquemas de visualização, resultando numa alteração do seu conteúdo ou aspeto). Esta sincronização ajuda o utilizador a explorar e a entender a informação. A Figura V-1 mostra um exemplo de uma imagem com um ecrã com quatro esquemas de visualização diferentes em utilização.

### 2.3. Metodologia de avaliação

Durante o processo de desenvolvimento dos esquemas de visualização e da interface do utilizador foram realizadas sessões de avaliação informais com várias pessoas, utilizando protótipos simples. Este grupo de pessoas era constituído por voluntários, sobretudo estudantes de engenharia informática que, regra geral, eram já detentores de alguma experiência no desenvolvimento de *sites web*. Estas sessões revelaram-se de grande utilidade, tendo permitido que algumas ideias fossem descartadas e abandonadas e que

outras fossem redefinidas. Dois dos membros da equipa, os menos envolvidos no processo de implementação do protótipo, também efetuaram alguns testes heurísticos à interface de utilizador e alguma avaliação informal aos esquemas de visualização. Concluído que foi este primeiro ciclo de avaliação, a equipa desenvolveu um protótipo mais sofisticado, incluindo seis esquemas diferentes e os mecanismos de visualização para os usar de forma síncrona. Depois, deu-se início a um segundo ciclo de testes. Neste ciclo, utilizou-se uma abordagem diferente, mais estruturada, que irá ser apresentada sucintamente nas próximas secções deste capítulo.

O segundo ciclo de testes tinha dois grandes objetivos: avaliar os aspetos principais da interface do utilizador e avaliar os esquemas de visualização. Estes esquemas de visualização incluem representações de visualização e os mecanismos de interação fornecidos aos utilizadores, a fim de estes poderem interagir com a informação através de uma representação visual. De acordo com (C. S. Freitas et al., 2002), existem dois conjuntos diferentes de critérios de avaliação para cada um destes últimos aspetos: para avaliar as representações visuais pode ser usada a complexidade cognitiva, organização espacial, codificação da informação e transição de estados; para avaliar os mecanismos de interação pode ser usada a orientação e a ajuda, navegação e interrogação e redução da informação.

Realizaram-se várias sessões de avaliação com dois tipos de utilizadores, tendo sido usadas principalmente técnicas de observação e inquérito para testar, tanto as funcionalidades da interface, como alguns aspetos dos esquemas de visualização. Estas técnicas são amplamente usadas em testes de usabilidade a interfaces de utilizador, sendo igualmente consideradas apropriadas para a avaliação de alguns aspetos das visualizações (Ware, 2000). A informação recolhida foi analisada utilizando análise de dados exploratória (*Exploratory Data Analysis*).

### 2.3.1. Caracterização de utilizadores

Com o objetivo de se realizarem duas avaliações distintas, ou seja, uma à interface do utilizador e outra à eficácia dos esquemas de visualização, foi pedida a colaboração de dois tipos diferentes de utilizadores: 32 estudantes de Engenharia Informática, inscritos numa disciplina introdutória de Interação Humano-Computador, e cinco profissionais colocados no Centro de Informática e Comunicações da Universidade de Aveiro (CICUA)<sup>93</sup>. Estes profissionais têm vários anos de experiência em desenvolvimento/programação para a *web*, gestores de *web*/redes, tendo três destes

---

<sup>93</sup> CICUA – Centro de Informática e Comunicações da Universidade de Aveiro. Entretanto, esta unidade orgânica sofreu uma alteração na sua designação, tendo passado a chamar-se sTIC – Serviços de Tecnologias da Informação e Comunicação.

profissionais frequentado alguma disciplina de Interação Humano-Computador durante a sua licenciatura. O perfil destes profissionais fez deles, não só bons representantes do público-alvo da aplicação prototipada, como também permitiu que entendessem melhor o tipo de contributo que esta equipa necessitava deles, no âmbito desta avaliação. Não sendo os alunos, regra geral, gestores/produtores *web*, detêm, no entanto, um perfil suficientemente apropriado para a realização desta avaliação, já que todos eles têm experiência como utilizadores da *web* e programação de computadores. Neste sentido, a equipa concentrou-se na avaliação da interface do utilizador com a ajuda dos alunos, tendo a avaliação da eficácia dos esquemas de visualização a colaboração dos cinco profissionais.

Uma vez que os alunos tinham vindo a praticar a avaliação de interfaces de utilizador utilizando diferentes métodos, no decorrer das suas aulas, decidiu-se capitalizar nas suas capacidades, pedindo-lhes para agir não só como utilizadores, mas também como observadores. Este procedimento permitiu à equipa obter informação de um maior número de utilizadores, considerando que era pedido aos observadores que registassem informação suficientemente simples (uma vez que não eram muito experientes). Considerou-se que assim se poderia proporcionar um bom treino a estes alunos, e que seria uma boa estratégia de motivação para que todos participassem, quer como utilizadores, quer como observadores.

Assim, enquanto metade dos alunos realizava as tarefas previamente definidas, a outra metade ia observando os colegas, registando tempos, tarefas completas e outras informações relevantes. Depois de decorrido um período de tempo pré-determinado, os alunos trocariam de tarefas. Obviamente que os alunos que agiram primeiramente como observadores manifestavam um conhecimento maior da interface do utilizador, um maior nível de consciência/conhecimento, podendo, portanto, ser considerados como utilizadores mais experientes. Daí que, para efeitos de análise da informação, os alunos foram divididos em duas classes distintas: utilizadores com menos experiência e utilizadores com um pouco mais de experiência. Há que sublinhar, no entanto, que apesar da existência destes dois grupos de utilizadores, a avaliação realizada é do tipo não-experimental, no sentido em que não existe um grupo de controlo, nem foi formulada qualquer hipótese. O objetivo principal era recolher ideias para depois serem exploradas nas fases posteriores desta avaliação.

A informação demográfica simples de cada um dos participantes foi recolhida antes do início das tarefas definidas, através de realização de um inquérito, que incluía também questões que pretendiam avaliar a sua experiência com aplicações de visualização de informação e como produtores ou gestores *web*. Depois de analisada esta informação, conclui-se que este grupo se situava entre os 19 e os 31 anos de idade (valor médio = 21 anos, um aluno com 30 anos e outro com 31), três do sexo feminino e 29 do sexo masculino, sem quaisquer problemas detetados a nível da perceção da cor. No que toca

à experiência com produção e avaliação *web*, a grande maioria era capaz de produzir páginas *web* de moderada complexidade e eram conhecedores dos métodos básicos de avaliação da *web*. Finalmente, revelou-se que, quanto à sua experiência na utilização de aplicações de visualização 2D, a maioria usa frequentemente vários pacotes de software, tais como os da Macromedia, Adobe e MatLab.

Depois de concluídas todas as tarefas, foi pedido aos alunos que preenchessem um questionário, a fim de se poder avaliar os seus níveis de satisfação e recolher a sua opinião sobre os vários assuntos abordados.

### 2.3.2. Instrumentos utilizados na experiência

A experiência foi conduzida recorrendo a um *site*, alunos e equipamentos da Universidade de Aveiro:

#### Base de dados

Foi desenvolvido um *site* interno contendo a informação (tal como o programa, material de estudo, trabalhos práticos, etc.), da disciplina de Interação Humano-Computador, lecionada no ano letivo 2002/2003 a um grupo de aproximadamente 40 estudantes de Engenharia Informática do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro. Este *site* foi especificamente desenvolvido para permitir uma recolha de dados tão completa quanto possível. A base de dados usada continha a informação recolhida ao longo do semestre, quer em contexto de utilização normal, quer em sessões controladas.

#### Equipamento

As sessões de avaliação foram realizadas num laboratório de aulas equipado com computadores pessoais, onde se instalou o protótipo e a aplicação SQL Server para a base de dados.

### 2.3.3. Objetivos e tarefas

Definiram-se um conjunto de tarefas relativamente simples que os utilizadores teriam que realizar durante as sessões de avaliação, consideradas representativas das operações típicas realizadas com a interface e com visualizações pelos utilizadores da aplicação.

Estabelecer tarefas simples facilita a análise da *performance* do utilizador. No entanto, essas tarefas não devem ser tão simples que ponham em causa a sua relevância (isto é,

tem que se questionar a frequência com que elas realmente ocorrem em situações do mundo real, assim como o quão significantes são na tarefa de solução/objetivo geral).

Cada utilizador tinha que completar dez tarefas dentro de uma determinada janela temporal. Essas tarefas relacionavam-se com a avaliação, quer da interface de utilizador, quer dos esquemas de visualização.

A seguir discriminam-se alguns exemplos de tarefas:

- manipulação e navegação nas várias janelas de representação;
- utilização de uma opção de menu ou botão para obter uma dada funcionalidade;
- seleccionar um determinado *site*;
- seleccionar uma determinada sessão;
- encontrar e seleccionar uma determinada página desse *site*;
- descobrir o número de vezes que um dado *link*, entre duas páginas, foi utilizado;
- mostrar os caminhos possíveis entre duas determinadas páginas;
- encontrar o número de páginas correspondentes ao caminho mais curto entre duas determinadas páginas;
- encontrar quantos “saltos” externos é que o utilizador efetuou durante a sessão.

As primeiras quatro tarefas da lista acima apresentada são muito simples e direccionadas para a avaliação de algumas funcionalidades da interface de utilizador (alteração das condições de visualização ou utilização de funcionalidades através dos botões ou opções de menu); as restantes tarefas foram combinadas, com algumas das anteriores, em tarefas mais complexas orientadas para a avaliação da *performance* do utilizador ao usar os esquemas de visualização para extrair informação quantitativa ou qualitativa através da interação com a informação. “Encontre e selecione uma determinada página do *site*” e “descubra quantas vezes um dado *link*, entre duas páginas, foi usado” são tarefas orientadas para avaliar funcionalidades “interrogativas” do esquema de visualização. “Encontre o número de páginas que correspondem ao caminho mais curto entre duas páginas” e “mostre os caminhos possíveis entre duas páginas” são tarefas que pretendem avaliar a funcionalidade “redução do espaço de dados”, de acordo com os critérios de avaliação propostos em (C. S. Freitas et al., 2002) para avaliar os mecanismos de interação dos esquemas de visualização.

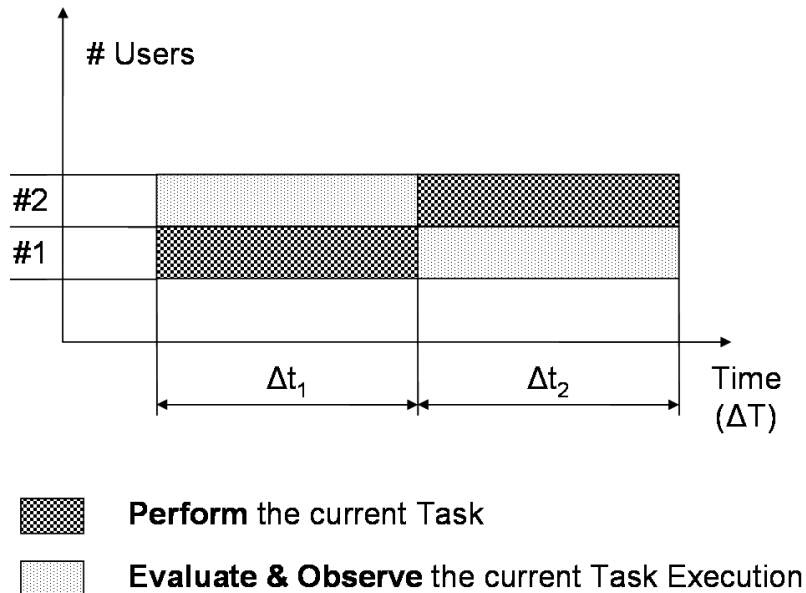
## 2.3.4. Procedimentos

Como referido acima, as sessões de testes realizaram-se com estudantes e com profissionais. Os procedimentos estão descritos nesta secção.



## Procedimentos com estudantes

Para se entender o grau de dificuldade das tarefas, a sua duração e clareza das questões colocadas procedeu-se a uma avaliação piloto com três estudantes, que não tinham participado em nenhuma das sessões abaixo descritas. Em consequência, foram alteradas algumas tarefas e questões.



**Figura V-2:** O papel dos estudantes. Retirado de (Santos et al., 2004).

Foram realizadas quatro sessões com estudantes: uma (TS1) com 18 alunos e noutra (TS2) com 14. Depois destas duas sessões de testagem, realizaram-se outras duas sessões com os mesmos alunos (TS3 e TS4). Assim, cada estudante participou em duas sessões de 1h45. Em cada sessão, os 32 estudantes foram organizados em grupos de dois alunos, nomeados de Utilizador#1 e Utilizador#2. Inicialmente, o Utilizador#1 funcionou como utilizador, passando depois a avaliador; o Utilizador#2 teve uma atuação inversa. Estes papéis foram atribuídos aleatoriamente. A Figura V-2 mostra-nos a sequência de papéis atribuídas a cada grupo de alunos.

Antes de se solicitar a cada participante que realizasse a tarefa, foi-lhes dada uma descrição genérica da aplicação e dos esquemas de visualização, tal como alguns detalhes da interface de utilizador. De seguida, foram informados acerca da avaliação: quais as tarefas que deveriam realizar e observar os inquéritos que teriam que responder, bem como das escalas utilizadas. Na primeira sessão, com cada grupo de alunos a explicação e resposta ao inquérito, durou aproximadamente 45 minutos. Depois deste período de tempo, todos os estudantes admitiram ter entendido o que lhes era proposto e manifestaram vontade de participar. Subsequentemente, foi pedido aos participantes que dessem início ao teste, realizando ou observando as tarefas durante 30 minutos, trocando posteriormente de papéis.

Depois das sessões TS1 e TS2 ficou a ideia que alguns estudantes não tinham entendido inteiramente as tarefas solicitadas. Assim, decidiu-se realizar uma segunda ronda de sessões (TS3 e TS4) duas semanas depois, recorrendo aos mesmos alunos e seguindo as mesmas orientações. Desta vez, não foram dadas quaisquer explicações acerca da aplicação, a equipa concentrou-se na explicação das tarefas e foi permitido aos alunos que praticassem durante 10 minutos antes de darem início à realização das tarefas. Considerou-se que esta seria uma abordagem razoável, já que os utilizadores-alvo eram pessoas experientes. Em consequência, nesta segunda ronda de sessões, esperava-se que a diferença na experiência entre os dois grupos de utilizadores, Utilizador#1 e #2, fosse menor, já que todos eles tinham já participado numa sessão de avaliação duas semanas antes.

## **Procedimentos com profissionais**

Na avaliação realizada com os cinco profissionais dos sTIC<sup>94</sup> da UA (à data da realização do estudo empírico este serviço da UA designava-se CICUA<sup>95</sup>) foi utilizada uma abordagem diferente e, de forma geral, menos estruturada. Numa sessão inicial de três horas com todos os profissionais envolvidos, foi dada uma descrição mais detalhada da aplicação e dos esquemas de visualização. Foram igualmente mostrados alguns detalhes da interface de utilizador. De seguida, houve lugar a um período de perguntas e respostas para que os participantes pudessem entender melhor o objetivo geral da aplicação e do tipo de avaliação que se pretendia realizar. Depois, foi instalado o protótipo nos seus computadores e foi-lhes dado algum tempo para o experimentarem livremente.

Nesta sessão, foi pedido aos utilizadores que pensassem acerca de como utilizariam a aplicação no seu trabalho diário e que tipo de pessoas beneficiariam também com a sua utilização. Depois, foram convidados a demonstrá-lo usando cenários simples.

Uma semana depois, em sessões individuais de 90 minutos, foi pedido a cada utilizador que realizasse um conjunto de tarefas com descrição em voz alta, ao mesmo tempo que eram observados por dois elementos da equipa de desenvolvimento. Estas tarefas eram basicamente as mesmas que tinham sido pedidas aos estudantes, adicionadas a outras que pretendiam avaliar principalmente a forma como os esquemas de visualização eram utilizados.

---

<sup>94</sup> sTIC: Serviços de Tecnologia de Informação e Comunicação da Universidade de Aveiro – <http://www.ua.pt/stic/>

<sup>95</sup> CICUA: Centro de Informática e Comunicações da Universidade de Aveiro. Este serviço sofreu, entretanto, uma alteração na sua designação para sTIC.

### 2.3.5. Informação recolhida

Nas sessões de avaliação realizadas com os alunos, cada observador tinha que registar os seguintes dados referentes à *performance* do utilizador:

- tempo despendido na realização de cada tarefa;
- se a tarefa foi ou não concluída;
- a resposta dada em cada tarefa (em algumas das tarefas);
- satisfação do utilizador;
- algum comentário que o observador considerasse relevante.

Depois de concluir o conjunto de tarefas definidas, solicitou-se aos utilizadores que preenchessem um inquérito, onde dariam as suas opiniões quanto aos esquemas de visualização e algumas das funcionalidades da interface.

Relativamente a cada esquema de visualização, foram colocadas as seguintes questões:

- é-lhe familiar e intuitivo?
- é de fácil utilização?
- é de fácil aprendizagem?
- fornece *feedback* adequado?
- utiliza um código de cores adequado?
- fornece mecanismos de interação adequados?

Algumas destas questões foram igualmente colocadas relativamente à iconografia utilizada em cada funcionalidade. Todas estas questões, incluindo aquelas referentes à satisfação do utilizador em cada tarefa, foram respondidas através da utilização de uma escala quantitativa:

1 | 2 | 3 | 4 | 5 | N

Onde **1** corresponde a “discordo totalmente”, **5** refere-se a “concordo totalmente” e **N** corresponde a não ter opinião ou não querer responder à questão. No final do inquérito, foi incluída uma questão de resposta aberta, dando assim a oportunidade de os utilizadores darem sugestões e/ou de fazer algum comentário.

Finalmente, foram recolhidas opiniões e sugestões adicionais através de uma conversa informal.

## 2.4. Resultados

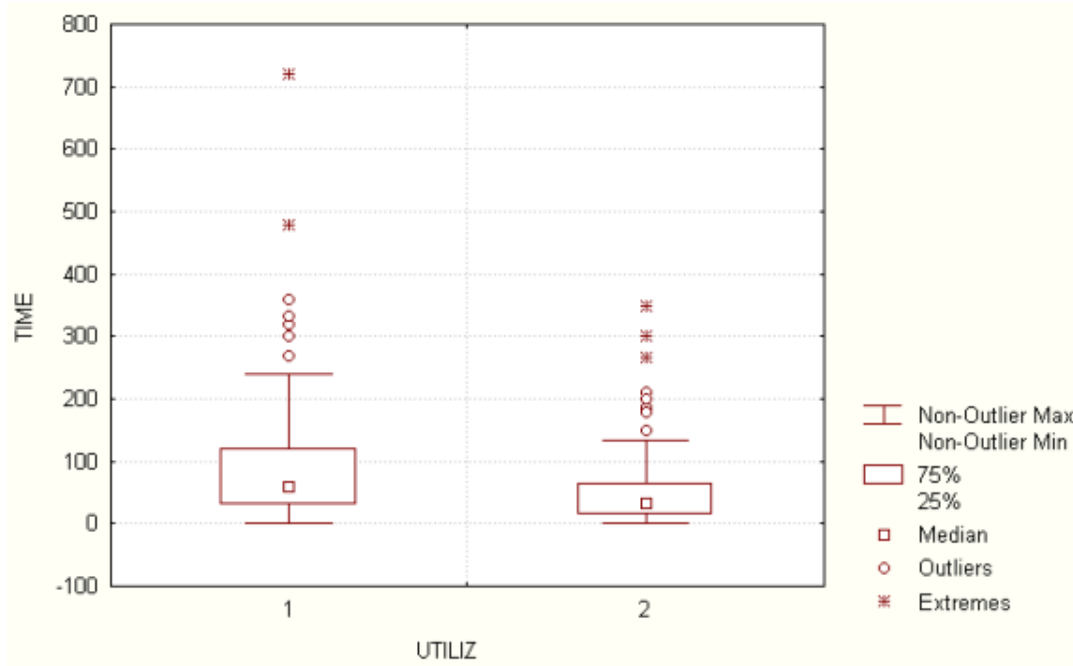
Nesta secção serão descritos os resultados obtidos com os alunos, mais de teor quantitativo e mais centralizados na interface do utilizador, tal como os resultados obtidos com os profissionais, que têm um teor mais qualitativo e focado na avaliação dos esquemas de visualização e no interesse geral da aplicação.

### 2.4.1. Resultados obtidos com estudantes

Como referido anteriormente, depois da primeira ronda de testes (TS1 e TS2) ficou a sensação que alguns alunos não tinham entendido completamente o que se pretendia que fizessem. Assim, simplificou-se o preenchimento dos formulários e da redação das tarefas aos observadores. Depois, numa primeira abordagem, decidiu-se analisar os seguintes itens recolhidos apenas nas sessões de avaliação TS3 e TS4:

- o tempo despendido em cada tarefa;
- se a tarefa foi concluída;
- se a resposta foi a correta (em algumas tarefas);
- a satisfação do utilizador.

A Figura V-3 mostra-nos o diagrama de caixas obtida através do *software* Statistica (STATISTICA, 1999), para os tempos que correspondem a todas as tarefas realizadas pelos utilizadores#1 (começaram como utilizadores) e utilizadores#2 (começaram como avaliadores). A mediana dos tempo situam-se nos 60 segundos e nos 32 segundos, respetivamente. Utilizando o teste de Wilcoxon, encontrou-se a diferença entre o significativo destes valores ( $p < 0,00001$ ). Enquanto os utilizadores#1, tal como os utilizadores#2, tinham mais ou menos o mesmo nível de experiência com a aplicação antes do início desta ronda de sessões de avaliação, e não se esperava observar diferenças significativas na realização das tarefas, existiu uma diferença na mediana do tempo. Isto parece indicar que os utilizadores #2 aprenderam a realizar as tarefas mais rapidamente pela simples observação dos colegas. No entanto, não foi registada uma diferença significativa no número de resposta correta entre os dois grupos de utilizadores.



**Figura V-3:** Diagrama de caixas correspondente aos tempos gastos pelos utilizadores #1 e #2. Retirado de (Santos et al., 2004).

No que se refere à satisfação do utilizador, a mediana para os utilizadores #1 e #2 são, respetivamente, de 3 e 4. Analisando, tarefa a tarefa, a satisfação do utilizador e o tempo (Figura V-4), observou-se de forma informal que os utilizadores têm níveis de satisfação inferiores sempre que demoram mais tempo na conclusão das tarefas. Este resultado foi depois confirmado como significativo nas tarefas 2, 4, 5, 6, e 7, de acordo com o coeficiente de correlação de Spearman.

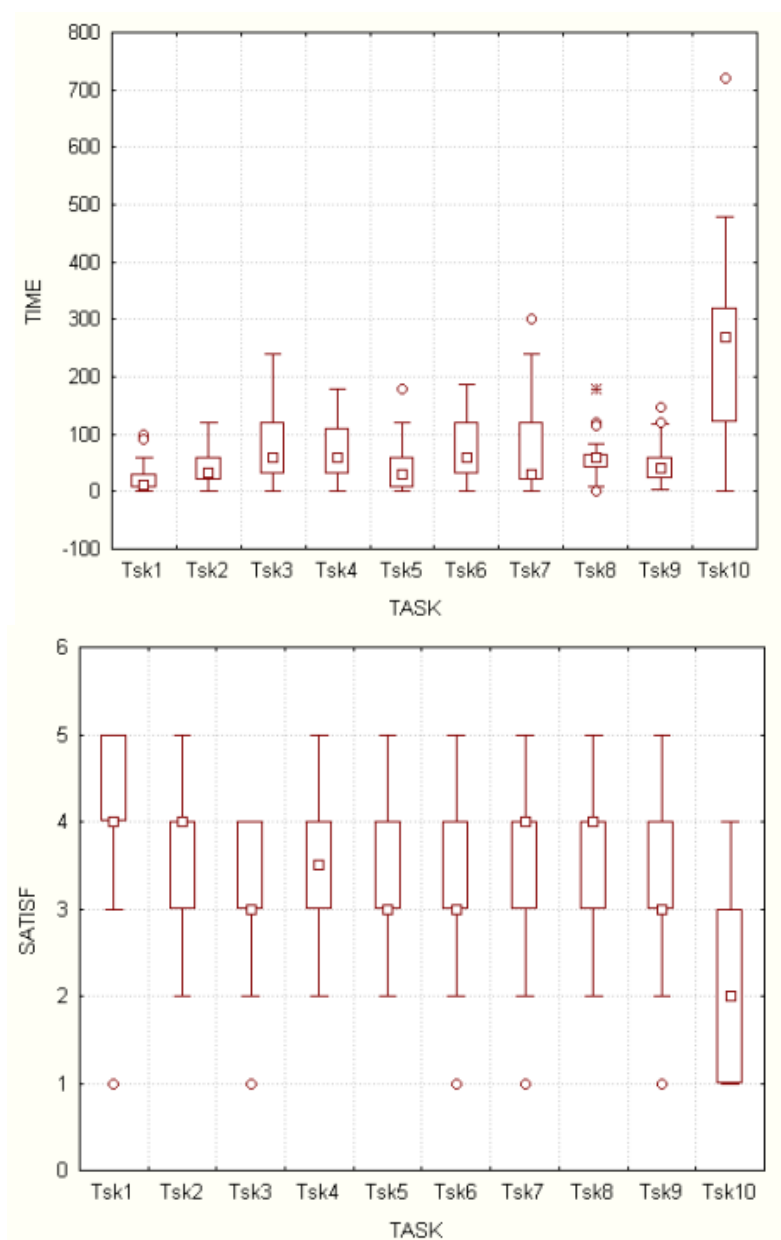
**Quadro V-1:** Número de utilizadores que completou, ou não, tarefas sem perguntas. Retirado de (Santos et al., 2004).

Tarefa	Completo	Não fez
1	32	0
2	32	0
4	31	1
6	31	1
8	29	3

O Quadro V-1 mostra-nos, para cada tarefa que não colocava qualquer questão para responder, o número de utilizadores que completou ou não completou as tarefas. Este quadro mostra-nos que a grande maioria dos utilizadores foi capaz de concluir as tarefas. Na realidade, 97% das tarefas foi terminada. Enquanto que para essas tarefas não há possibilidade de se saber se os utilizadores as completaram de forma correta, para as tarefas onde era colocada uma questão para responder, essa informação existe.

**Quadro V-2:** Número de utilizadores que completou corretamente, erradamente ou não fez tarefas que incluíam uma questão. Adaptado de (Santos et al., 2004).

Tarefa	Correta	Errada	Não fez
3	24	4	4
5	26	5	1
7	17	12	3
9	29	0	3
10	16	2	14



**Figura V-4:** Diagrama de caixas correspondente ao tempo e satisfação de cada tarefa. Retirado de (Santos et al., 2004).

O Quadro V-2 mostra, para cada uma destas tarefas, o número de utilizadores que as completaram corretamente, incorretamente ou que não as concluíram. Contando o número total de respostas corretas às questões, observou-se que uma grande percentagem de utilizadores foi capaz de encontrar corretamente a informação nos esquemas de visualização: 70% das tarefas que incluíam uma questão foram completadas corretamente.

Nas tarefas 3 e 7, os utilizadores deveriam descobrir quantas vezes um determinado *link* tinha sido utilizado, usando dois esquemas de visualização diferentes (SiteMap2D e PageExplorer2D). Dos 32 utilizadores, 24 completaram a tarefa 3 corretamente, quatro utilizadores concluíram a tarefa mas obtiveram uma resposta errada e quatro não completaram a tarefa. No que toca à tarefa 7, 17 utilizadores terminaram a tarefa corretamente, 12 completaram-na mas com resultados errados e três não realizaram a tarefa. Esta diferença no comportamento dos utilizadores foi confirmada como estatisticamente significativa, utilizando um teste de sinais não-paramétrico (*non-parametric sign test*). Isto sugere que o SiteMap2D suporta melhor este tipo de tarefa do que o PageExplorer2D. É interessante registar que, enquanto os utilizadores demonstraram um desempenho superior na tarefa 3, a sua satisfação foi superior na tarefa 7. Este resultado vem confirmar que a noção de que uma maior satisfação do utilizador não está necessariamente relacionada com um melhor desempenho. Pode, contudo, estar relacionada com o tempo despendido para a conclusão da tarefa, que é menor na tarefa 7 (mediana = 30 segundos) que na tarefa 3 (mediana = 60 segundos).

Na tarefa 5, os utilizadores tinham que encontrar o número de páginas correspondente ao caminho mais curto entre duas páginas *web* determinadas. Para concluir esta tarefa, os utilizadores tinham que selecionar as páginas usando o SiteMap2D e depois observar o resultado utilizando outro esquema de visualização apresentado noutra janela sincronizada. Dos 32 utilizadores, 26 obtiveram o número correto de páginas, cinco obtiveram um número incorreto (o que significa que provavelmente não foram capazes de selecionar as páginas corretas no SiteMap2D) e apenas um utilizador não foi capaz de realizar a tarefa. Igualmente relevante foi o facto de a mediana do tempo para a realização da tarefa ter sido o segundo mais baixo (30 segundos) de todas as tarefas. Isto parece demonstrar que a utilização síncrona destes dois esquemas de visualização é razoavelmente óbvia para os utilizadores.

A tarefa 9 implicava a utilização de outro esquema de visualização (SessionMap2D), com o objetivo de encontrar o número de “saltos” para páginas externas. Nesta tarefa, 29 dos 32 utilizadores completaram a tarefa e obtiveram o número correto de “saltos”. Apenas três utilizadores não foram capazes de a concluir. Este resultado parece demonstrar que o SessionMap2D suporta adequadamente esta tarefa.

A tarefa 10 era muito mais difícil, intencionalmente planeada para descobrir se os alunos eram capazes de realizar uma tarefa complexa usando esta aplicação. A mediana do

tempo foi de 268 segundos e a mediana dos níveis de satisfação foi de 2, o valor mais baixo no conjunto de todas as tarefas. Mesmo assim, 16 dos 32 utilizadores conseguiram realizá-la corretamente e obter a resposta certa, dois completaram a tarefa, mas obtiveram uma resposta errada e 14 não foram capazes de concluir a tarefa. Apesar de ter sido necessário fornecer ajuda a alguns estudantes sobre como realizar a tarefa, esperavam-se resultados piores nesta tarefa.

Apesar de a maioria das tarefas serem de relativa simplicidade, estes resultados revelam-se encorajadores, já que os estudantes não eram utilizadores experientes e foram capazes de completar corretamente uma percentagem elevada das tarefas. Igualmente, a mediana da satisfação global do utilizador foi de 3 para o grupo de utilizadores#1 e de 4 para o grupo de utilizadores#2.

### **2.4.2. Resultados obtidos com profissionais**

Durante as sessões realizadas com este grupo de utilizadores, foi recolhida informação de baixo nível interessante no que toca a funcionalidades específicas da interface de utilizador e dos esquemas de visualização incluídos no protótipo, assim como informação mais geral, de alto nível, relacionada com o interesse e utilidade da aplicação.

No que se refere à interface de utilizador, alguns ícones, as escalas de cor e as caixas de diálogo foram considerados os pontos mais fracos. De acordo com alguns utilizadores, a possibilidade de se terem múltiplas representações síncronas é a funcionalidade mais interessante da aplicação, tornando-a, no entanto, também mais complexa.

Em geral, estes utilizadores consideraram que a aplicação tem um grande potencial e demonstraram interesse em utilizá-la assim que se pudesse criar uma primeira versão consolidada.

## **2.5. Algumas conclusões**

Começou por realizar-se uma avaliação informal, passando-se a uma mais estruturada. Esta fase de testagem utiliza apenas métodos não-experimentais, de forma a gerar novas ideias e informação que, eventualmente, pudesse permitir uma mais completa e formal avaliação da aplicação.

Utilizaram-se técnicas de observação e inquéritos e foi pedido a dois tipos de utilizadores que realizassem um conjunto de tarefas, que pretendiam avaliar a usabilidade de algumas das funcionalidades da interface de utilizador e dos mecanismos de interação dos esquemas de visualização.



Enquanto 97% das tarefas que não incluíam qualquer questão para responder foram completadas com sucesso, 70% das tarefas que incluíam questões foram concluídas corretamente. Adicionalmente, as medianas do tempo de conclusão das tarefas situam-se entre os 10 e os 60 segundos (com exceção da tarefa 10), valores estes que se apresentam como razoáveis, considerando o tipo de tarefa e experiência dos utilizadores. Finalmente, a mediana da satisfação dos utilizadores situa-se entre o moderado e o elevado.

O *feedback* obtido forneceu muitas ideias para a melhoria da aplicação o que muito reforçou a motivação da equipa.

No que respeita à representação visual dos esquemas de visualização, realizou-se ainda uma avaliação informal da complexidade da representação e do tempo que demorará a sua reconstrução após a interação do utilizador.

A ajuda de um designer gráfico foi crucial para auxiliar a avaliação e adequação de alguns aspetos, como por exemplo, cores, ícones, organização espacial e coerência.

### 3. Repositórios web de estudo

O desenvolvimento do protótipo, ao longo deste trabalho, passou por três fases importantes, em que foram utilizados repositórios de informação diferentes, consoante o que estava em teste.

Numa primeira fase, pretendia-se testar métodos de extração de informação de um ficheiro de *log*. Para o efeito foram utilizados os dados do portal do CPJ (“Ciber-Parque de Jogos”), implementado no âmbito de um projeto para crianças com idades dos 6 aos 10 anos, sobretudo em atividades de ATL (Atividades de Tempos Livres). Este portal estava a ser utilizado por uma comunidade de uma dezena de escolas, cada uma com pelo menos dois PC's. Alguns autores deste trabalho integravam a equipa do projeto em causa e eram responsáveis pela gestão e manutenção do portal e respetivo servidor que se encontrava fisicamente localizado nas instalações do DeCA (Departamento de comunicação e Arte) da UA (Universidade de Aveiro). Este acesso privilegiado permitiu, numa primeira instância, a obtenção de dados de utilização real de um *site*.

Numa segunda fase, pretendia-se aferir a estrutura de dados de informação e os algoritmos de leitura dos ficheiros de *log*, bem como a definição de filtros de informação e teste desses filtros sobre os ficheiros de *log*. Nesta altura, fazia falta informação proveniente de um *site* com um nível da atividade mais elevada que o registado no portal

do CPJ. Para o efeito recorreu-se ao *site* (plataforma WebCT<sup>96</sup>) de suporte ao ensino à distância e de apoio a um número considerável de disciplinas lecionadas na UA. Os destinatários deste *site* eram os alunos das várias disciplinas. Este *site* foi, entretanto, descontinuado, sendo substituído por outro a funcionar sobre uma nova plataforma (BlackBoard<sup>97</sup>). Mais recentemente, já depois de terminada a fase de desenvolvimento desta tese, também esta plataforma foi substituída por uma nova (Moodle<sup>98</sup>).

Na última fase de desenvolvimento do protótipo, importava controlar / conhecer melhor todos os detalhes relativos ao *site*. Com isto em mente foi conceptualizado e implementado um *site* com um repositório de informação on-line para as disciplinas Interação Humano Computador e Interfaces Humano Computador (irá ser usada a designação IHC para ambas) do DETI (Departamento de Eletrónica, Telecomunicações e Informática) da UA. Este *site*, destinado aos alunos das disciplinas em questão, pretendia conseguir um controle total sobre a captação de informação de utilização do mesmo.

### 3.1. Portal CPJ [1ª fase]

Na primeira fase do desenvolvimento do protótipo, a validação e aferição da estrutura de informação e dos algoritmos de leitura de ficheiros de *log* era uma das tarefas principais tarefas.

Era necessário ter acesso a ficheiros de *log* com informação coerente, e nada melhor que informação de um sistema a funcionar em ambiente real para fornecer tal informação. É neste contexto que surge o portal do CPJ (“Ciber-Parque de Jogos”) (<http://www.cpj.ua.pt>) (Mealha et al., 1999) (Mealha, 1999), implementado no âmbito de um projeto para crianças com idades dos 6 aos 10 anos, sobretudo em atividades de ATL (Atividades de Tempos Livres). Este portal estava a ser utilizado por uma comunidade de uma dezena de escolas, cada uma com pelo menos dois PC's. Alguns autores deste trabalho integravam a equipa do projeto em causa e eram responsáveis pela gestão e manutenção do portal e respetivo servidor que se encontrava fisicamente localizado nas instalações do DeCA (Departamento de comunicação e Arte) da UA (Universidade de Aveiro) e, portanto, acessível. A possibilidade de aceder e alterar alguns parâmetros da configuração do servidor permitiu que os ficheiros de *log* contivessem o máximo de

---

<sup>96</sup> WebCT: Plataforma de apoio ao ensino criada na Universidade da Columbia Britânica. Em 2006 foi adquirida pela rival Blackboard Inc..

<sup>97</sup> BlackBoard: <http://www.blackboard.com/> - Plataforma de apoio ao ensino detida pela empresa Blackboard Inc..

<sup>98</sup> Moodle: <https://moodle.com/> - Plataforma de apoio ao ensino gratuita e de código aberto.

informação disponível no sistema, bem como o conhecimento detalhado sobre o formato / estrutura dos mesmos.

Este acesso privilegiado permitiu, numa primeira instância, a obtenção de dados de utilização real de um *site*.

Este portal estava alojado num servidor IIS 5.0 (*Microsoft Internet Information Services* versão 5.0) a correr numa plataforma *Windows 2000 Server*. Este servidor, no momento da instalação, por definição, não fazia qualquer registo de atividade. Naturalmente, a funcionalidade de registo (*log*) foi ativada imediatamente após a instalação, pelo administrador. O IIS 5.0 podia ser configurado para fazer o registo num de 4 formatos possíveis:

- Microsoft IIS Log File Format;
- NCSA Common Log File Format;
- ODBC Logging;
- W3C Extended Log File Format.

Foi escolhido o formato “W3C Extended” por ser o mais completo e flexível, permitindo configurar quais os campos de informação a registar, escolhidos a partir de uma lista predefinida. Toda a informação é registada em formato de texto em ficheiros que são criados segundo uma periodicidade configurável.

De maneira a maximizar a informação recolhida, foi necessário alterar a configuração do servidor para ativar o registo de todos os campos disponíveis.

O formato dos ficheiros resultantes foi discutido no Capítulo III, secções 3.2.1 e 3.3.1, bem como também foi discutida a sua utilização no protótipo no Capítulo IV, secção 4.1. Na secção 2 dos Anexos estão descritos todos os detalhes da informação registada.

Era importante criar uma estrutura de dados que acomodasse toda a informação pertinente presente nos ficheiros de *log*, bem como a sua validação e aferição. Era ainda importante testar os algoritmos que faziam a leitura e permitiam a extração da informação a colocar na estrutura de dados criada.

Esta primeira fase coincidiu com os primeiros passos do desenvolvimento do protótipo descritos no Capítulo IV., secção 4: Primeira proposta de protótipo – versão 0, com especial destaque para a secção com o título *Parser Métrico* de ficheiro de *log*.

## **3.2. Disciplina do WebCT (*e-learning*) [2ª fase]**

Na segunda fase do desenvolvimento do protótipo, importava validar e aferir a estrutura de informação e os algoritmos de leitura e extração de informação dos ficheiros de *log*, bem como a definição de filtros de informação e teste desses filtros sobre os ficheiros de

*log*. Para este fim, era conveniente estudar ficheiros de *log* de um *site* com um índice de atividade mais elevado e com uma estrutura mais extensa e complexa.

Ainda que existissem inúmeros *sites* com nível de utilização pretendidas, um detalhe importante limitou o leque de escolha: os ficheiros de *log*. A informação registada era importante mas sem acesso aos ficheiros nada se podia fazer. Aqui residiu um dos principais problemas nesta fase: muitos responsáveis pelos *sites* não se sentiam confortáveis com a ideia de fornecer este tipo de informação...

Depois de alguns contactos, o CEMED (Centro Multimédia e de Ensino a Distância), uma unidade funcional da UA e a entidade responsável pela gestão e manutenção do serviço de eLearning da UA, permitiu o acesso a um conjunto de ficheiros de *log* do sistema. Sem qualquer acesso à configuração do servidor, não foi possível intervir sobre a quantidade e/ou formato da informação recolhida.

O serviço de eLearning era assegurado por uma plataforma WebCT (<http://webct.ua.pt>) instalada num servidor a correr Linux com servidor *web* Apache.

Seguindo diretivas da reitoria, que incentivava a adoção do eLearning, um número considerável de disciplinas, lecionadas por vários departamentos da UA, usava a plataforma como suporte às suas atividades. Para algumas disciplinas resumia-se um conjunto de ficheiros em formato PDF enquanto noutras eram utilizadas mais funcionalidades, incluindo nos momentos de avaliação dos alunos. O nível de utilização era variável, dependendo sobretudo da estratégia adotada pelo docente de cada disciplina que em alguns casos passaram a usar a plataforma como principal canal de comunicação com os alunos.

Durante o decorrer deste trabalho, o *site* de suporte da plataforma WebCT foi descontinuado, tendo sido substituído por outro a correr sobre uma nova plataforma BlackBoard (<http://elearning.ua.pt>), entretanto adotada. Os conteúdos foram adaptados e migrados do WebCT para o BlackBoard, contudo as alterações na interface com o utilizador, bem como ao nível do funcionamento interno da plataforma, limitaram a utilização dos ficheiros de *log* fornecidos para o trabalho em curso.

O formato dos ficheiros foi discutido no Capítulo III, secções 3.2.1 e 3.3.1, bem como também foi discutida a sua utilização no protótipo no Capítulo IV, secção 4.1. Na secção 2 dos Anexos estão descritos todos os detalhes da informação registada.

Dado que o formato dos ficheiros de *log* era diferente dos obtidos no portal do CPJ, foi necessário, em primeiro lugar, alterar os algoritmos de leitura e extração de informação, bem como aferir a estrutura de informação recolhida.

O passo seguinte consistiu na definição e teste de filtros que permitissem, em primeiro lugar, isolar a informação referente a uma disciplina em concreto. Depois, que permitisse a extração da informação relativa à atividade de um utilizador durante uma sessão de

utilização. Para tal foi necessário definir os limites, página de entrada e de saída, de uma sessão de utilização por parte de um determinado utilizador.

### 3.3. Site da disciplina IHC [3ª fase]

Na última fase do desenvolvimento do protótipo, o objetivo era mais ambicioso: pretendia-se extrair não só a informação dos ficheiros de *log* como também registar toda a estrutura do *site* e o conteúdo de cada página do mesmo. Só assim seria possível ter uma ideia do que o utilizador estaria a ver durante a sua sessão de navegação no *site*, bem como perceber como ele interagiu com cada página que lhe foi apresentada.

Conseguir este tipo de informação não é um processo trivial. Ao navegar num *site*, o servidor só regista os pedidos e mais alguns detalhes fornecidos pelo navegador. Ainda que possa ficar registado qual a página pedida, a página onde se encontra o *link* usado (*referer*), alguns detalhes sobre o navegador e sobre o sistema, não se consegue muito mais. Por exemplo, não há maneira direta de saber qual o aspeto da página no ecrã do utilizador, ou em que posição do ecrã se encontra algum elemento da página como um *link*.

Uma maneira de obter este tipo de informação consistiu em proceder à conceptualização e implementação de um *site* para ter um controle total sobre a captação de informação da utilização, pelo que se decidiu conceptualizar e implementar um repositório de informação on-line para as disciplinas Interação Humano Computador e Interfaces Humano Computador (irá ser usada a designação IHC para ambas) do DETI (Departamento de Eletrónica, Telecomunicações e Informática) da UA.

O *site* (<http://www2.ca.ua.pt/ihc> entretanto mudado para <http://www3.ca.ua.pt/ihc>) foi implementado de maneira a permitir uma noção exata do aspeto de cada página no ecrã dos utilizadores, tendo sido programado de maneira a ser possível conhecer a posição exata de cada elemento no ecrã, bem como identificar de forma inequívoca os objetos usados pelo utilizador no processo de interação com as diversas páginas.

Para atingir estes objetivos, recorreu-se a alguns artifícios em *JavaScript* para garantir que a janela do navegador no ecrã do utilizador tivesse um tamanho fixo predefinido e que não podia ser alterado. Adicionalmente, o código de cada página foi modificado para que todos os *links* ficassem especialmente marcados de modo a ser possível identificar inequivocamente qual o *link* usado pelo utilizador, particularmente importante quando mais que um objeto (imagem, texto, etc.) da página têm o mesmo *link*. Para ajudar no processo de identificação dos limites de uma sessão de utilizador, foi adicionado um *cookie* de sessão, que consiste num identificador atribuído pelo servidor no início de cada sessão e que o navegador envia junto com todos os pedidos ao servidor.

Dado que este *site* foi alojado no mesmo servidor que o *site* do CPJ, as questões relacionadas com os ficheiros de *log* já foram abordadas anteriormente.

Era necessário proceder à captura da estrutura do *site*, conteúdo de cada página e a sua relação com a informação encontrada nos ficheiros de *log*. O protótipo utilizado nesta fase do trabalho socorria-se de uma base de dados para proceder ao armazenamento de uma imagem da totalidade do conteúdo do *site*, para referência futura (os *sites* tendem a evoluir e a ser alterados ao longo do tempo), de maneira a ser mantida a coerência. Foi desenvolvida uma aplicação destinada a proceder à captura e armazenamento do conteúdo do *site* na base dados.

O desafio desta fase do trabalho consistia em conseguir relacionar a informação de utilização do *site* proveniente dos ficheiros de *log* com a estrutura real do *site* e o conteúdo de cada página de maneira a poder-se representar visualmente a atividade do utilizador numa representação gráfica da estrutura do *site*, em que cada página tem uma representação realista do que o utilizador observou no momento em que a visualizou no seu ecrã.

A informação recolhida e armazenada foi discutida nas secções 3.2.2 e 3.3.2 do Capítulo III, a sua utilização e registo na base de dados do protótipo foi discutida na secção 6.1 do Capítulo IV.

## 4. Resultados

Cada fase de desenvolvimento tinha um objetivo, com a correspondente captura e análise de dados. Adicionalmente, cada repositório de dados tinha as suas características, qualidades e limitações inerentes à especificidade desse repositório.

De seguida serão apresentados alguns resultados obtidos com cada um dos repositórios de informação utilizados.

### 4.1. Dados do CPJ [1ª fase]

Os dados do portal do CPJ foram especialmente importantes durante a primeira fase do desenvolvimento do protótipo.

Eis alguns resultados obtidos com a utilização deste repositório de informação:

- Identificação do formato e particularidades dos ficheiros de *log* produzidos pelo servidor *web* do portal que serviram de base para todo o trabalho posterior;
- Extração de toda a informação disponível num ficheiro de *log*;

- Filtragem da informação pertinente obtida dum ficheiro de *log*;
- Identificação de sessões de utilizador, nomeadamente ponto de entrada e saída do *site*, bem como todo o percurso no interior do *site*, data / hora de início e duração da sessão, etc.;
- Tratamento estatístico da informação obtida, tanto da utilização do *site* em geral como de sessões individuais ou conjunto de sessões;
- Seleção de uma sessão ou sessões para estudo mais detalhado.

Contudo este repositório de dados tinha algumas limitações, entre as quais se pode destacar:

- O nível de utilização era relativamente baixo, não só pela dimensão da comunidade de utilizadores como pelas características da mesma;
- A dimensão do *site* acabou por se manifestar relativamente reduzida, pois acabou por não ter o crescimento previsto;
- A comunidade de utilizadores demonstrou algumas especificidades muito próprias, acabando por resultar numa utilização atípica do *site*. Como o próprio protótipo acabou por ajudar a verificar, a maioria das sessões de utilização limitavam-se à utilização de uma a três páginas.

Ainda assim, a informação recolhida, como a identificação de uma ou múltiplas sessões e respetiva análise detalhada, permitiu passar ao passo seguinte: a elaboração de métodos de visualização da informação extraída.

As limitações encontradas tornaram urgente encontrar um repositório de informação mais abrangente, diversificado.

## 4.2. Dados do WebCT [2ª fase]

Os dados da plataforma de *eLearning*, WebCT, vieram impulsionar o desenvolvimento do protótipo para uma nova fase. Foram usados sobretudo na versão 1, descrita na secção 5 do Capítulo IV.

Nesta fase, o protótipo tinha dado um salto significativo. De blocos soltos e fechados tinha evoluído para uma aplicação integrada, com uma interface gráfica e muito para evoluir. Havia muito trabalho para fazer, muita coisa para testar e validar. Muita coisa não seria possível testar e por em prática com os dados do CPJ. Os dados obtidos do WebCT vieram ajudar, e muito, em todo o processo de teste e desenvolvimento.

Este repositório de informação permitiu atingir vários objetivos dos quais se destacam os seguintes resultados:

- Adaptação e teste dos processos de extração de informação dos ficheiros de *log* (o formato dos ficheiros do WebCT eram diferentes do provenientes do CPJ);

- Refinamento dos algoritmos de filtragem da informação pertinente obtida dum ficheiro de *log*;
- Refinamento do processo de identificação de sessões de utilizador;
- Tratamento estatístico complementar à informação obtida para se obter ainda mais detalhe sobre a utilização do *site*;
- Melhoramentos no processo de seleção de uma sessão ou sessões para estudo mais detalhado, nomeadamente pelo uso de uma interface gráfica.
- Desenvolvimento e teste de vários métodos de visualização:
  - Da estrutura do *site*;
  - De uma sessão de um utilizador;
  - De múltiplas sessões;
  - De visualização de múltiplas vistas, em simultâneo.

Contudo este repositório de dados veio a mostrar que tinha algumas limitações e problemas, entre as quais se pode destacar:

- A atividade do *site* foi interrompida antes de todo o trabalho estar terminado, o que impediu a realização de uma parte significativa das tarefas previstas;
- Os ficheiros de *log* não continham alguma informação que poderia tornar alguns algoritmos utilizados mais eficientes, nomeadamente na identificação de sessões de utilização;
- A equipa só dispunha de um acesso de teste a uma disciplina, pelo que não era possível registar o aspeto de todas as páginas, nomeadamente pelo descrito no ponto seguinte;
- Uma parte do *site* era gerado de forma dinâmica. O conteúdo de algumas páginas era produzido em função de fatores como a identificação do utilizador, do dia e de outros fatores contextuais. Isto levava a que o conteúdo apresentado, por exemplo, a dois utilizadores diferentes pudesse ser profundamente diverso.

O facto de o WebCT ter sido interrompido, para além das restantes limitações, levou à procura de uma solução alternativa onde se tivesse um pouco mais de controlo sobre todo o processo.

### **4.3. Dados da disciplina IHC [3ª fase]**

O desenvolvimento de um *site* de raiz surgiu na sequência das limitações identificadas nas fontes de informação usadas anteriormente. Desta maneira os autores deste trabalho poderiam ter um maior controlo sobre todo o processo de recolha de informação. O esforço mostrou-se proveitoso pois permitiu testar e evoluir o protótipo para um novo nível a que correspondeu uma nova versão do mesmo.



Os dados provenientes do *site* da disciplina IHC foram determinantes para o desenvolvimento do protótipo. Começou a ser usado ainda na versão 1 e toda a versão 2, descritas nas secções 5 e 6 do Capítulo IV, respetivamente.

Este repositório de informação, em conjugação com as melhorias da plataforma tecnológica do protótipo, permitiu testar e desenvolver muitos aspetos, nomeadamente:

- Verificar o trabalho desenvolvido nas fases anteriores do protótipo;
- Criar bases de dados com a informação recolhida, não só dos ficheiros de *log* como da estrutura do próprio *site*, código HTML de cada página e imagem com a representação da forma como era apresentada no ecrã do utilizador;
- Desenvolver vários métodos de visualização com representações realistas das páginas visitadas pelos utilizadores, da estrutura do *site*, de sessões, entre outras;
- Identificar e representar na imagem da página a localização do *link* utilizado, pois estes estavam identificados de forma inequívoca;
- Criar esquemas de visualização que permitiram identificar as zonas de interação nas várias páginas do *site*;
- Desenvolvimento, refinamento e teste de vários métodos de visualização:
  - Da estrutura do *site*;
  - De uma sessão de um utilizador;
  - De múltiplas sessões;
  - De apresentação de informação estatística;
  - De visualização de múltiplas vistas, em simultâneo, de forma sincronizada.

A possibilidade de intervenção direta no servidor e nos conteúdos publicados permitiu fazer algumas adaptações à medida das necessidades da evolução do trabalho. Um exemplo disso foi a inclusão de código que permitisse identificar de forma individual todos os *links* presentes numa dada página.

## 5. Algumas conclusões

Neste capítulo foi descrita a utilização de vários repositórios de informação ao longo do desenvolvimento do protótipo, nas suas várias etapas. A ligação desses repositórios com as várias etapas do desenvolvimento do protótipo, nomeadamente nas suas várias versões, justifica uma breve reflexão.

No início do processo de desenvolvimento do protótipo apenas se precisava de uns ficheiros de *log* para testar e validar alguns algoritmos. Essa situação ficou profundamente diferente na fase final do desenvolvimento, onde se está muito mais próximo do que poderá ser uma aplicação final. Passou a ser importante dispor de mais informação, de maneira a poder demonstrar todo o potencial do protótipo desenvolvido.

Quase que se poderia dizer que foram feitos um estudo de caso e um estudo de laboratório. Ao utilizar-se um *site* (WebCT) a funcionar no seu estado normal, sem qualquer intervenção dos autores deste trabalho, como fonte de informação, poderia considerar-se estar perante um estudo de caso. Já o *site* de IHC pode ser considerado um estudo de laboratório sem reservas, na medida em que as condições da experiência estiveram sempre sobre algum tipo de controlo.

Os repositórios usados influenciaram de forma considerável a direção do desenvolvimento dos trabalhos. Cada repositório apresentou os seus desafios e motivou o desenvolvimento de soluções adequadas, tendo permitido aferir inúmeras propostas.

Estes dados tiveram uma utilidade inegável, mostrando a sua relevância ao permitirem validar a estratégia metodológica de investigação adotada. Um bom exemplo está no facto do protótipo ter identificado utilizações atípicas do *site* do CPJ.

Estes repositórios, nomeadamente o *site* de IHC, acabaram por se mostrar suficientemente robustos e adequados ao estudo levado a cabo. Foram identificadas e apresentadas algumas limitações dos *sites* do CPJ e WebCT. O *site* de IHC também não é perfeito: poderia ter mais conteúdo, um nível de atividade maior (com mais utilizadores), etc. Contudo, a maior limitação identificada acabou por ter a ver com o protótipo ao não ser especialmente eficiente a lidar com *sites* dinâmicos. Aliás, esta limitação não é exclusiva deste protótipo pois trata-se de um problema intrínseco a todas as ferramentas de análise de *sites web*.

Pode concluir-se que os resultados obtidos demonstram que a metodologia adotada (experiência empírica) foi adequada.

# **Capítulo VI.**

## **CONCLUSÕES E TRABALHO FUTURO**

# 1. Conclusões

Esta tese descreve um modelo conceptual, de base visual, para o estudo e monitorização da interação de utilizadores com um *site web*. Foi desenvolvida (prototipada) uma aplicação que implementa os conceitos essenciais deste modelo com o objetivo de o validar. Este protótipo foi sujeito a muitas alterações, adições e afinações durante as várias iterações do seu desenvolvimento com consequentes influências na conceptualização e detalhe dos elementos e funções inerentes ao modelo.

Usando o protótipo descrito foi possível identificar, por um lado, que alguns aspetos do modelo se tornaram mais simples, comparativamente à abordagem tradicional, por outro lado, que existem algumas áreas que ainda precisam de ser melhoradas.

Os cenários de infocomunicação mediados tecnologicamente podem ser mais facilmente analisados e percebidos com a ajuda de ferramentas como as descritas nesta tese, uma vez que permitem uma nova perspetiva (mais inovadora), geradora de informação nova, sobre os dados obtidos a partir dos ficheiros de *log* dum servidor *web*, dados que também traduzem comportamentos de interação holísticos dos utilizadores. A possibilidade de conjugar esta informação com a estrutura do *site*, com o *layout* da interface gráfica e com os dados da monitorização da interação humano-computador prometem um passo significativo no entendimento mais profundo do processo infocomunicacional, processo significativamente mais difícil ou impossível de se correlacionar na maioria das aplicações disponíveis.

Talvez uma das estratégias mais relevantes desta tese passe por analisar dados de fluxo infocomunicacional “projetados” ou “mapeados”, segundo técnicas de representação ou visualização adequadas, sobre o modelo de interface gráfica e sobre o modelo de estrutura dos repositórios de informação. Técnicas estas que, sem prejuízo de poderem ser exploradas ou utilizadas individualmente, são prioritariamente utilizadas de forma integrada, síncrona e sempre numa perspetiva holística de análise dos dados em determinado repositório.

# 2. Reflexão crítica

A tecnologia, em particular a tecnologia de informação e comunicação, tem vindo e, tudo indica, continuará a progredir de forma galopante. Este progresso tecnológico tem tido um forte impacto nos processo infocomunicacionais, começando nos dispositivos de

comunicação e informação usados pelas pessoas, como é o caso dos *smartphones*, passando pelos novos métodos de comunicação usados, como nas redes sociais, sem esquecer tantos outros detalhes como as redes de comunicação sem fios, cada vez com maior capacidade e facilidade de acesso. Mais importante que a evolução da tecnologia, em si, é o efeito que esta tem na vida das pessoas, que, no geral, se tem saldado como muito positiva.

Contudo, este progresso tecnológico galopante que tem vindo a ocorrer acabou por comprometer a estabilidade de uma solução tecnológica a adotar para o protótipo. Mesmo no curto período do desenvolvimento do protótipo foi perceptível a evolução em duas frentes: por um lado, nas tecnologias usadas na *web*, como novos protocolos, funcionalidades e serviços de comunicação e informação, por outro, na forma como estas tecnologias são usadas, com o surgimento de várias redes sociais e dos paradigmas infocomunicacionais associados. Se juntarmos a isto a evolução que também foi acontecendo nas plataformas de desenvolvimento e respetiva obsolescência das mais antigas, não será difícil perceber a dificuldade inerente ao desenvolvimento de um protótipo que tem como objetivo estudar estes processos infocomunicacionais.

Uma das áreas que foi objeto de grande evolução nos últimos anos e que permitiu um salto significativo nos serviços disponibilizados na *web*, como fica patente quando se visita, por exemplo, uma qualquer rede social, geralmente não é visível ou perceptível quando se usa um desses serviços porque está a funcionar de forma silenciosa nas entranhas dos servidores que lhe dão suporte. Trata-se de tecnologia (como PHP, ASP, etc.) que permite a criação de páginas HTML de forma dinâmica que são apresentadas ao visitante sem que este se aperceba deste facto. Esta tecnologia está hoje presente na grande maioria dos sites, permitindo funcionalidades que, de outra maneira, seriam simplesmente impossíveis. Contudo, esta tecnologia expõe uma das limitações do protótipo desenvolvido que teria de ser revisitada, com impacto mesmo ao nível de alguns elementos do modelo proposto.

Ao longo deste trabalho foram abordadas algumas técnicas de visualização que acabaram por não ser exploradas. O domínio da visualização 3D foi dos mais sacrificados e valeria a pena retomar esse assunto, tendo, contudo, em mente que não poderiam ser esquecidos estudos de psicologia aplicada que nas últimas décadas têm apontado para algumas diferenças, relacionadas com o género, na perceção e orientação espacial e na perceção de profundidade (Velez et al., 2015).

Pode-se ainda considerar que esta tese, fruto do grande investimento nas área tradicionalmente chamadas de "*Big Data*" e "*Data Warehousing*", que vieram alimentar o surgimento da disciplina de "*Analytics*", tem algum desajuste terminológico em relação aos conceitos fundamentais que aborda. O surgimento e domínio da Google na sua área de atividade é um dos principais responsáveis pela valorização dada ao "*Analytics*", pela sua capacidade de analisar o uso dos sites com base em dados de interação com as

diversas páginas *web*. As linhas de investigação relacionadas com “Inteligência Artificial” e “*Machine Learning*” hoje apontadas como umas das áreas com mais projeção e impacto futuro no desenvolvimento de toda a atividade na *web* (online), são também elas responsáveis pela pertinência dada ao “*Big Data*” e pelo despertar para a importância do “*Open Data*”.

Importa, contudo, mencionar que esta evolução nas áreas científicas – que, entretanto, se desenvolveram e ganharam relevância no tema deste trabalho – em nada compromete o contributo desta tese quanto aos métodos e técnicas inovadoras que propõe. A título de exemplo, mesmo o tremendo investimento da Google no seu *analytics*, nomeadamente na dimensão *user-driven analytics*, continua a não disponibilizar instrumentos de base visual que suportem a integração de múltiplas perspetivas de análise, para um uso mais universal. Os processos infocomunicacionais nas plataformas *web* podem ser analisados em duas perspetivas: por um lado, a interação com a interface gráfica da página e, por outro, a estrutura do site. Nesta tese, estas duas perspetivas são abordadas, com uma base visual, de forma integrada, síncrona e suportada em dados de interação dos utilizadores, num detalhe individual ou holístico e padronizado.

### 3. Contribuição original

Esta tese comporta um conjunto de contributos com carga inovadora que se podem organizar em 3 grandes dimensões: i) métodos e técnicas; ii) processo de avaliação de usabilidade; iii) contributo para estudos de *User Experience* (UX).

Desenvolvimento de um modelo de monitorização e análise, de base visual, de sites infocomunicacionais, em que os elementos constituintes e fundamentais deste modelo traduzem métodos e técnicas muito particulares de representação e análise da atividade de interação dos utilizadores em contexto de interface gráfica e em contexto de fluxo infocomunicacional na estrutura da plataforma *web* sob análise. Uma abordagem que acautela a melhor forma de representação e análise visual dos dados em questão mas garante também a sua integração e relação holística, síncrona e com a mesma métrica de registo. Estas condições tornam este modelo único, quando comparado com outras abordagens e paradigmas de análise, dentro desta problemática, que adotam técnicas ou métodos específicos não integrados de análise, deixando o trabalho de reflexão e impacto holístico e síncrono das ocorrências, para um exercício técnico do investigador ou analista. Tradicionalmente, o paradigma de representação visual não é prioritário e por vezes os dados são apresentados em lista unidimensionais, como folhas de Excel, ao contrário do que é adotado neste modelo, que considera a dimensão visual como paradigma de representação fundamental.

O processo de avaliação do modelo, na forma de protótipo, mereceu uma abordagem inovadora e empiricamente iterativa, na medida em que relacionou a dimensão subjetiva (qualitativa) dos sujeitos (utilizadores especialistas que usaram a aplicação prototipada) com a abordagem tradicional orientada pela avaliação de usabilidade ditada pela norma ISO 9241, parte 11, mais quantitativa (eficácia e eficiência). Num dos contextos de uso e estudo recorreu-se aos administradores de plataformas *web* e noutro contexto aos professores e estudantes, tentando sempre esgotar a recolha de opinião de todos os *stakeholders* (diferentes perfis de "sujeito" ou "parte interessada" no repositório sob estudo) de determinado contexto de uso em estudo. Apesar da natureza abstrata do modelo, desta forma, através de um protótipo em contexto de uso real (com os sujeitos do estudo), consegue-se avaliar a capacidade do modelo responder às necessidades de monitorização e análise porventura necessárias, ou perspetivadas como necessárias, num cenário de abordagem inovadora e geradora de conhecimento novo. No decurso deste processo, com base na opinião dos sujeitos e nos dados recolhidos em contexto de uso real, foi possível aferir o protótipo com consequências no refinamento conceptual do modelo. Um modelo que, desta forma pode ser caracterizado como "modelo heurístico" pelo recurso à observação e validação empírica, está subjacente ao método de avaliação adotado.

Conforme já foi referido, este estudo é suportado por técnicas de monitorização e análise inovadoras, conjugadas com métodos suportados em orientações de visualização holística e síncrona. Importa sublinhar com mais algum detalhe que estas técnicas processam dados, com base na estrutura das plataformas *web* sobre a qual mapeiam dados de contexto real da atividade de interação de cada utilizador, o que hoje se designa por *user-driven analytics* e que representam uma fonte analítica e técnica capaz de nutrir uma abordagem metodológica de "*User Experience (UX)*". Técnicas que se preocupam primeiro em compreender qual a estrutura da plataforma *web*, da respetiva interface gráfica, e sobre estas representam a interatividade ocorrida e sob estudo.

## 4. Trabalho futuro

Existem sinais socioeconómicos muito fortes que apontam para uma possível estratégia de envolvimento ativo do público-alvo (cliente) no processo e missão organizacional. Esta situação tem um carácter inovador na medida em que o cliente é convidado a opinar e porventura interferir nos processos de decisão organizacionais (ex. de um novo produto, ou alteração da marca, etc.). Neste contexto, assumido como "*social business*" por alguns (ex: IBM<sup>99</sup>, Opower<sup>100</sup>, Unanimous<sup>101</sup>) e onde o envolvimento do utilizador ou

---

<sup>99</sup> IBM: <https://www.ibm.com/>

cliente (*client/user-engagement*) é absolutamente fundamental, o contributo desta tese representa um instrumento de monitorização e aferição da participação do utilizador (*user-driven analytics*) na dinâmica das organizações, fulcral para acompanhar e consolidar o processo infocomunicacional e de decisão desta participação. Não é estranho que isto aconteça no advento da afirmação do pós-modernismo no setor ocidental do planeta, sobretudo com a enorme ênfase que é dada ao indivíduo e à sua participação singular em diversas plataformas sociais online. Começa-se agora a perceber que muitas destas plataformas traduzem organizações privadas com interesses muito específicos, e com consequências para diversas atividades, quase sempre com intuítos (interesses e intenções) socioeconómicos. Registam-se também iniciativas sociais, lúdicas, culturais e momentos até de revolução política, contudo, são as oportunidades económicas que mais se têm feito notar.

Com estes novos contextos de atuação em vista, sugerem-se algumas melhorias ao modelo proposto e à ferramenta (protótipo) desenvolvida:

- Perspetivas diagramáticas do *site*: a representação estrutural de *sites*, como *tree maps*, não são adequadas para todos os *sites*, pelo que será necessário aprofundar um pouco mais a investigação sobre este detalhe. Poderiam ser desenvolvidas alternativas 2D e 3D com o objetivo de obter um entendimento mais profundo do conteúdo do *site*, respetivo fluxo de interação (comportamento dos utilizadores) e o contributo de cada página nesta representação.
- Análise estatística (*analytics*): poderia ser interessante obter informação mais detalhada sobre as características da sessão e o comportamento do utilizador durante a navegação no *site*. Poderiam ser desenvolvidas representações 3D para analisar os dados da navegação do utilizador, como os vários percursos adotados pelos utilizadores e respetivo tempo despendido para atingir um objetivo predeterminado. Igualmente pertinente será a representação do tipo de atividade de interação que se deu numa determinada página e/ou no seu conteúdo, quer se trate de um serviço ou documento.
- Análise de multi-sessão: poderia ser interessante produzir uma vista integrada de mais que uma sessão de utilizador para comparar utilizadores distintos ou o mesmo utilizador em ocasiões distintas. Mais uma vez, vistas 2D ou 3D poderiam ser interessantes para obter um melhor entendimento da relação existente entre a distribuição da informação presente numa página e o comportamento do utilizador. Esta representação aproxima-se e responde às tendências de mercado para compreender padrões comportamentais. As áreas de marketing são pioneiras no desenvolvimento de métodos suportados em informação analítica

---

<sup>100</sup> Opower/Oracle: <https://opower.com/>

<sup>101</sup> Unanimous: <http://unanimous.ai/>



para tomar decisões sobre estratégias de promoção e venda de produtos e serviços. Contudo, uma técnica consolidada de representação e análise de padrões comportamentais, de base visual, sobre repositórios infocomunicacionais online seria uma mais valia que permitiria ir ainda mais longe. Uma técnica destas constituiria um instrumento de trabalho muito valioso para algumas organizações.

- Equacionar este desafio para *sites* com páginas dinâmicas: tendo em conta que a maioria dos *sites* atuais utilizam alguma tecnologia de páginas dinâmicas, como PHP, ASP, etc., seria interessante evoluir o protótipo desenvolvido para que pudesse suportar este tipo de *sites*.



# BIBLIOGRAFIA

- Abello, J., & Korn, J. (2000). Visualizing massive multi-digraphs. *IEEE Symposium on Information Visualization, 2000*, 39-47. doi: 10.1109/INFVIS.2000.885089
- Abello, J., & Korn, J. (2002). MGV: A System for Visualizing Massive Multidigraphs. *IEEE Transactions on Visualization and Computer Graphics*, 8(1), 21-38. doi: 10.1109/2945.981849
- Agnello, A., & Levkowitz, H. (2015). Quick Vis: A Web-Based Visualization Delivering Flexible Exploration of User-Driven Analytics. *19th International Conference on Information Visualisation*, 468-473. doi: 10.1109/iV.2015.84
- Andrews, K. (2002, July 5th, 2002). Information Visualisation - Lecture Tutorial Notes Retrieved 2012-11-22, from <http://www.iicm.edu/ivis/>
- Andriole, S. (2004, April 2nd, 2004). The Real Truth... Retrieved 2012-11-22, 2012, from <http://www.datamation.com/columns/bizalign/article.php/3335081>
- Barlow, T., & Neville, P. (2001). A Comparison of 2D Visualizations of Hierarchies. *IEEE Symposium on Information Visualization, InfoVis01*, 131-138. 0-7695-1342-5
- Bederson, B. B. (2000). Fisheye menus. *ACM symposium on User interface software and technology*, 2(2), 217-225. doi: 10.1145/354401.354782
- Bederson, B. B. (2001). PhotoMesa: a zoomable image browser using quantum treemaps and bubblemaps. *ACM symposium on User interface software and technology*, 3(2), 71-80. doi: 10.1145/502348.502359
- Bederson, B. B., & Shneiderman, B. (2003). *The Craft of Information Visualization: Readings and Reflections*. San Francisco: Morgan Kaufman.

- Benford, S., Taylor, I., Brailsford, D., Koleva, B., Craven, M., Fraser, M., Reynard, G., & Greenhalgh, C. (1999). Three dimensional visualization of the World Wide Web. *ACM Computing Surveys*, 31(4es), 1-16. doi: 10.1145/345966.346021
- Berners-Lee, T. (1989). Information Management: A Proposal Retrieved 2016-08-01, from <https://www.w3.org/History/1989/proposal.html>
- Berners-Lee, T. (2016). Biography Retrieved 2016-08-01, from <https://www.w3.org/People/Berners-Lee/>
- Bieber, M., Vitali, F., Ashman, H., Balasubramanian, V., & Oinas-Kukkonen, H. (1997). Fourth generation hypermedia: some missing links for the World Wide Web. *International Journal Human-Computer Studies*(47), 31-66.
- Bier, E. A., Stone, M. C., Pier, K., Buxton, W., & DeRose, T. D. (1993). Toolglass and magic lenses: the see-through interface. *Proceedings of the 20th annual conference on Computer graphics and interactive techniques, SIGGRAPH '93*, 73-80. doi: 10.1145/166117.166126
- Brath, R. (1999). Concept Demonstration Metrics for Effective Information Visualization. *IEEE Symposium on Information Visualization, InfoVis97*, 108-111.
- Callaghan, J. (2002). *Inside Intranets & Extranets: Knowledge Management and the Struggle for Power*. Palgrave Macmillan.
- Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings in information visualization: using vision to think*. San Francisco (CA): Morgan Kaufman.
- Carlsmith, J. M., Ellsworth, P., & Aronson, E. (1976). *Methods of Research in Social Psychology*. Addison-Wesley.
- Carpendale, M. S. T., & Montagnese, C. (2001). A framework for unifying presentation space. *ACM symposium on User interface software and technology*, 3(2), 61-70. doi: 10.1145/502348.502358
- Carpendale, S. (2008). Evaluating information visualizations. *Information Visualization: Human-Centered Issues and Perspectives, LNCS 4950*, 19-45. doi: 10.1.1.456.3403
- Ceruzzi, P. E. (2003). *A History of Modern Computing* (2nd. Edition ed.): The MIT Press.
- Chi, E. H., Pirolli, P., & Pitkow, J. (2000). The Scent of a Site: A System for Analysing and Predicting Information Scent, Usage, and Usability of Web Site. *Proceedings on Human Factors in Computing Systems*, 2(1), 161-168.

- Chi, E. H., Pitkow, J., Mackinlay, J., Pirolli, P., Gossweiler, R., & Card, S. K. (1998). Visualizing the evolution of Web ecologies. *SIGCHI Conference on Human Factors in Computing Systems*, 400-407. doi: 10.1145/274644.274699
- Cockburn, A., Karlson, A., & Bederson, B. B. (2008). A review of overview+detail, zooming, and focus+context interfaces. *ACM Computing Surveys (CSUR)*, 41(1), 2:1-31. doi: 10.1145/1456650.1456652
- Coulston, C., & Vitolo, T. M. (2001). A hypertext metric based on huffman coding. *ACM conference on Hypertext and Hypermedia*, 243-244. doi: 10.1145/504216.504275
- Cugini, J., & Scholtz, J. (1999). VISVIP: 3D visualization of paths through web sites. *International Workshop on Web-based Information Visualization*, 259-263. doi: 10.1109/DEXA.1999.795175
- Cunha, M. P. e., Rego, A., Cunha, R. C. e., & Cabral-Cardoso, C. (2003). *Manual de comportamento organizacional e gestão*. Lisboa: RH.
- Dieberger, A., & Lönnqvist, P. (2000). Visualizing interaction history on a collaborative web server. *ACM on Hypertext and hypermedia*, 220-221. doi: 10.1145/336296.336374
- Eick, S. G. (2001). Online Activity Visualizing. *Communications of the ACM*, 44(8), 45-50.
- Faraday, P., & Sutcliffe, A. (1998). Making contact points between text and images. *ACM international conference on Multimedia*, 29-37. doi: 10.1145/290747.290752
- Fielding, R. T., Gettys, J., Mogul, J. C., Nielsen, H. F., Masinter, L., Leach, P. J., & Berners-Lee, T. (1999). RFC 2616: Hypertext Transfer Protocol -- HTTP/1.1.
- Fisher, D. (2000). Visualizing Social Newsgroup Interaction. *CHI 2000 - Workshop The Future is Here*.
- Fogg, B. J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., Paul, J., Rangnekar, A., Shon, J., Swani, P., & Treinen, M. (2001). What makes Web sites credible?: a report on a large quantitative study. *SIGCHI Conference on Human Factors in Computing Systems*, 61-68. doi: 10.1145/365024.365037
- Freitas, C. S., Luzziardi, P., Cava, R., Winckler, M., Pimenta, M., & Nedel, L. (2002). Evaluating Usability of Information Visualization Techniques. *Proceedings 5th Symposium on Human Factors in Computer Systems IHC2002*.
- Freitas, L. d., Morin, E., & Nicolescu, B. (1994). Carta de Transdisciplinaridade. *Primeiro Congresso Mundial da Transdisciplinaridade*, 3.

- Furnas, G. W. (1986). Generalized fisheye views. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '86*, 16-23. doi: 10.1145/22627.22342
- Furnas, G. W. (2006). A fisheye follow-up: further reflections on focus + context. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '06*, 999-1008. doi: 10.1145/1124772.1124921
- Grinstein, G., Hoffman, P., Laskowski, S., & Pickett, R. (2004). Benchmark Development for the Evaluation of Visualization for Data Mining Retrieved 3/2004
- Gutwin, C. (2002). Improving focus targeting in interactive fisheye views. *Conference on Human Factors in Computing Systems*, 4(1), 267-274. doi: 10.1145/503376.503424
- Gutwin, C., & Skopik, A. (2003). Fisheyes are good for large steering tasks. *SIGCHI Conference on Human Factors in Computing Systems*, 5(1), 201-208. doi: 10.1145/642611.642648
- Healey, C. G., Amant, R. S., & Chang, J. (2001). Assisted Visualization of E-Commerce Auction Agents. *Graphics Interface*, 201-208.
- Hearn, D., & Baker, M. P. (2004). *Computer graphics with OpenGL*: Pearson Prentice Hall.
- Hendley, R. J., Drew, N. S., Wood, A. M., & Beale, R. (1996). Case study. Narcissus: Visualising Information. *Information Visualization*, 1995, 90-96. doi: 10.1109/INFVIS.1995.528691
- Hinrichs, R. J. (1997). *Intranets: what's the bottom line?* Mountain View (CA): Sun Microsystems Press.
- Hix, D., II, J. E. S., Gabbard, J., McGee, M., Durbin, J., & King, T. (1999). User-Centered Design and Evaluation of a Real-Time Battlefield Visualization Virtual Environment. *Proceedings IEEE Virtual Reality 99*, 96-103.
- Hochheiser, H., & Shneiderman, B. (1999). Understanding Pattern of User Visits to Web Sites: Interactive Starfield Visualizations of WWW Log Data. *ASIS'99 Proceedings of the 62nd Annual Meeting of the American Society for Information Science*, 36, 331-344.
- Igarashi, T., & Hinckley, K. (2000). Speed-dependent Automatic Zooming for Browsing Large Documents. *ACM Symposium on User Interface Software and Technology*, 2(2), 139-148. doi: 10.1145/354401.354435

- INE-Família. (2012). Utilização das Tecnologias de Informação e Comunicação pelas Famílias *Inquérito do Instituto Nacional de Estatística*: INE.
- Inselberg, A., & Dimsdale, B. (1990). Parallel coordinates: a tool for visualizing multi-dimensional geometry. *IEEE Conference on Visualization*, 361-378. doi: 10.1109/VISUAL.1990.146402
- ISO. (1998). Ergonomic requirements for office work with visual display terminals (VDTs) *Part 11 : Guidance on usability*. Geneve, Switzerland: International Organization for Standardization.
- Ivory, M. Y., & Hearst, M. A. (2002). Statistical profiles of highly-rated web sites. *SIGCHI Conference on Human Factors in Computing Systems*, 1(1), 367-374. doi: 10.1145/503376.503442
- Ivory, M. Y., Sinha, R. R., & Hearst, M. A. (2001). Empirically validated web page design metrics. *SIGCHI Conference on Human Factors in Computing Systems*, 3(1), 53-60. doi: 10.1145/365024.365035
- Jordan, P. W. (2001). *An Introduction to Usability*. London: Taylor & Francis.
- Keim, D. A. (2000). Designing Pixel-Oriented Visualization Techniques: Theory and Applications. *IEEE Transactions on Visualization and Computer Graphics*, 6(1), 59-78.
- Keim, D. A. (2001). Visual Exploration of Large Data Sets. *Communications of the ACM*, 44(8), 38-44.
- Keim, D. A. (2002). Information Visualization and Visual Data Mining. *IEEE Transactions on Visualization and Computer Graphics*, 8(1), 1-8. doi: 10.1109/2945.981847
- Kelly, S. U., Sung, C., & Farnham, S. (2002). Designing for improved social responsibility, user participation and content in on-line communities. *SIGCHI Conference on Human Factors in Computing Systems*, 1(1), 391-398. doi: 10.1145/503376.503446
- Kobsa, A. (2001). An Empirical Comparison of Three Commercial Information Visualization Systems. *IEEE Symposium on Information Visualization, InfoVis01*, 123-130.
- Koutsofios, E. E., North, S. C., & Keim, D. A. (1999). Visualizing large telecommunication data sets. *IEEE Computer Graphics and Applications*, 19(3), 16-19. doi: 10.1109/38.761543

- Kreuseler, M., Lopez, N., & Schumann, H. (2000). A scalable framework for information visualization. *IEEE Symposium on Information Visualization, 2000*, 27 - 36. doi: 10.1109/INFVIS.2000.885088
- Lamping, J., & Rao, R. (1994). Laying out and visualizing large trees using a hyperbolic space. *7th annual ACM symposium on User interface software and technology*, 13-14. doi: 10.1145/192426.192430
- Lamping, J., & Rao, R. (1996). The Hyperbolic Browser: A Focus+Context Technique for Visualizing Large Hierarchies. *Journal of Visual Languages & Computing*, 7(1), 33-55. doi: 10.1006/jvlc.1996.0003
- Lawton, S. (1995). Intranets Fuel Growth of Internet Access Tools. *Digital News & Review*, 1995-04-24, 20-21.
- Lueg, C., & Fisher, D. (2003). *From Usenet to CoWebs: interacting with social information spaces*: Springer.
- Mackinlay, J. D., Robertson, G. G., & Card, S. K. (1991). The perspective wall: detail and context smoothly integrated. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '91*, 173-176. doi: 10.1145/108844.108870
- Mayhew, D. J. (1999). *The Usability Engineering Lifecycle: a practitioner's handbook for user interface design*. San Francisco, USA: Morgan Kaufman.
- Mazza, R. (2009). *Introduction to Information Visualization* (1 ed.): Springer.
- McCormick, B. H., DeFanti, T. A., Brown, M. D., & (Ed.s). (1987). Visualization in scientific computing. *ACM SIGGRAPH Computer Graphics*, 21(6). 0097-8930
- McPhee, R. D., & Zaug, P. (2009). The Communicative Constitution of Organizations. In L. L. Putnam & A. M. Nicotera (Eds.), *Building Theories of Organization: The constitutive role of communication* (1 edition ed., pp. 21-47): Routledge.
- Mealha, Ó. (1999). Ciber Parque de Jogos. *Jornada Digital – Aveiro Cidade Digital*.
- Mealha, Ó. (2012). Ciências da Comunicação, Informação e Computação: Conhecimento Transdisciplinar. [Conference Proceedings]. *IX LUSOCOM - Congresso da Federação Lusófona de Ciências da Comunicação: Lusofonia e interculturalidade*, 957 - 976. doi: 2175-4683
- Mealha, Ó., Raposo, R., Almeida, P., Nunes, J., Antunes, M. J., & Monteiro, P. (1999). Ciber Parque de Jogos: Uma nova forma de aprender brincando. *Actas do VIII*



- Encontro de Ludotecas e Espaços Lúdicos ao ar Livre, Realidades e Virtualidades da Ludicidade na Viragem do Milénio*, 21-25.
- Mealha, Ó., Sousa-Santos, B., Nunes, J., & Zamfir, F. (2004). Integrated Visualization Schemes for an Information and Communication Web Log Based Management System. *Proceedings of International Conference on Information Visualization*, 295-301.
- Munzner, T. (1998). Exploring large graphs in 3D hyperbolic space. *IEEE Computer Graphics and Applications*, 18(4), 18 - 23. doi: 10.1109/38.689657
- Munzner, T. (2014). *Visualization Analysis and Design*: AK Peters/CRC Press.
- Newman, W. M., & Lamming, M. G. (1995). *Interactive System Design*. Harlow: Addison-Wesley.
- Nielsen, J. (1993). *Usability Engineering*: Academic Press.
- Nielsen, J., & Mack, R. L. (1994). *Usability inspection methods*. New York: John Wiley.
- Nielsen, J., & Sano, D. (1994). SunWeb: User interface design for Sun Microsystem's intranet Retrieved 2012-12-12, 2012, from <http://www.useit.com/papers/sunweb/>
- Norman, D. (2002). *The Design of Everyday Things* (1st Basic Paperback edition (29 Aug 2002) ed.): Basic Books.
- North, C., & Schneiderman, B. (2000). Snap-Together: Can Users Construct and Operate Coordinated Views? *International Journal on Human-Computer Studies*, 53(3), 715-739.
- Nunes, J., Zamfir, F., Mealha, Ó., & Santos, B. S. (2002). Web LogVisualizer for Communication and Information Management. *Actas da 3ª Conferência da Associação Portuguesa de Sistemas de Informação – CAPSI2002*, Artigo nº 77.
- Nunes, J., Zamfir, F., Mealha, Ó., & Santos, B. S. (2003). Web LogVisualizaer: a Tool for Communication and Information Management. *Proceedings of the 10th International Conference on Human-Computer Interaction, 3 (Human-Centered Computing: Cognitive, Social and Ergonomic Aspects)*, 824-828.
- Paganelli, L., & Paternò, F. (2002). Intelligent analysis of user interactions with web applications. *International conference on Intelligent user interfaces*, 111-118. doi: 10.1145/502716.502735

- Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., & Carey, T. (1994). *Human-Computer Interaction*. Harlow, England: Addison-Wesley.
- Preece, J., Sharp, H., & Rogers, Y. (2011). *Interaction Design: Beyond Human-Computer Interaction*. Harlow, England: John Wiley & Sons.
- Rego, A. (2003). Comunicação *Manual de Comportamento Organizacional e Gestão* (pp. 353-399): RH.
- Rhyne, T.-M. (2000). Scientific Visualization in the Next Millennium. *IEEE Computer Graphics and Applications*, 20(1), 20-21. doi: 10.1109/38.814538
- Rhyne, T.-M. (2003). Does the Difference between Information and Scientific Visualization Really Matter? *IEEE Computer Graphics and Applications*, 23(3), 6-8. doi: 10.1109/MCG.2003.1198256
- Riedl, M. O., & Amant, R. S. (2002). Toward Automated Exploration of Interactive Systems. *International conference on Intelligent user interfaces*, 135-142. doi: 10.1145/502716.502738
- Robertson, G., Cameron, K., Czerwinski, M., & Robbins, D. (2002). Polyarchy visualization: visualizing multiple intersecting hierarchies. *Conference on Human Factors in Computing Systems*, 4(1), 423-430. doi: 10.1145/503376.503452
- Santos, B. S., Zamfir, F., Ferreira, C., Mealha, Ó., & Nunes, J. (2004). Visual Application for the Analysis of Web-Based Information Systems Usage: A Preliminary Usability Evaluation. *Proceedings of International Conference on Information Visualization – IV04*, 812-818.
- Sarkar, M., & Brown, M. H. (1994). Graphical fisheye views. *Communications of the ACM*, 37(12), 73-83. doi: 10.1145/198366.198384
- Sebrechts, M., Vasilakis, J., Miller, M., Cugini, J., & Laskowski, S. (1999). Visualization of Search Results: A Comparative Evaluation of Text, 2D, and 3D Interfaces. *proceedings of the ACM Conf. Research and Development in Information Retrieval, ACM SIGIR 99*, 3-10.
- Shneiderman, B. (1996). The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. *IEEE Symposium on Visual Languages*, 336-343.
- Shneiderman, B. (1998). *Designing the User Interface: Strategies for effective Human-Computer Interaction* (3rd ed.): Addison-Wesley.

- Slack, J., & Munzner, T. (2006). Composite Rectilinear Deformation for Stretch and Squish Navigation. *Transactions on Visualization and Computer Graphics*, 12(5), 901-908. doi: 10.1109/TVCG.2006.127
- Sommerville, I. (2001). *Software Engineering* (6th ed.): Addison-Wesley.
- Spence, R. (2000). *Information Visualization* (1 ed.): Addison Wesley.
- Spence, R. (2007). *Information Visualization: Design for Interaction* (2 ed.): Pearson.
- Spence, R. (2014). *Information Visualization: An Introduction* (3 ed.): Springer.
- STATISTICA. (1999). *STATISTICA for Windows, version 5.5*: StatSoft Inc.
- Tan, D. S., Czerwinski, M., & Robertson, G. (2003). Women Go With the (Optical) Flow. *Proceedings on Human Factors in Computing Systems*, 5(1), 209-215.
- Tan, D. S., Gergle, D., Scupelli, P. G., & Pausch, R. (2003). With Similar Visual Angles, Larger Displays Improve Spatial Performance. *Proceedings on Human Factors in Computing Systems*, 5(1), 217-224.
- Telleen, S. L. (1995). IntraNet Methodology - Concepts and Rationale Retrieved 2016-08-01, from <http://www.iorg.com/papers/amdahl/concepts1.html>
- Tufte, E. R. (1994). *Envisioning information*. Cheshire (CT): Graphics Press.
- Tufte, E. R. (1998). *Visual explanations: images and quantities, evidence and narrative*. Cheshire (CT): Graphics Press.
- Tweedie, L. (1999). Characterizing Interactive Externalizations. In J. D. M. a. B. S. Stuart K. Card (Ed.), *Readings in Information Visualization: Using Vision to Think* (pp. 616-623). San Diego - USA: Morgan Kaufmann.
- vBulletin. (2013). vBulletin Community Forum - FAQ: What is a bulletin board? Retrieved 2013-10-07, 2013
- Velez, M. C., Silver, D., & Tremaine, M. (2015). Understanding Visualization through Spatial Ability Differences. *VIS 05 - IEEE Visualization, 2005*, 511-518. doi: 10.1109/VISUAL.2005.1532836
- Ware, C. (2000). *Information Visualization : Design for Perception*. London: Morgan Kaufmann.
- Ware, C. (2003). *Information Visualization : Design for Perception* (2nd ed.). London: Morgan Kaufmann.

- Wiss, U., Carr, D., & Jonsson, H. (1998). Evaluating Three-Dimensional Information Visualization Designs: a case Study of Three Designs. *Proceedings of the IEEE Information Visualization 98*, 137-144.
- Wittenburg, K., Lanning, T., Heinrichs, M., & Stanton, M. (2001). Parallel bargrams for consumer-based information exploration and choice. *ACM symposium on User interface software and technology*, 3(2), 51-60. doi: 10.1145/502348.502357
- Zamfir, F., Nunes, J., Santos, B. S., & Mealha, Ó. (2004). Web Usability Tools in Organizational Communication and Information Management: Models and Visualization Methods (pp. 1-113). Aveiro: IEETA / DeCA - Universidade de Aveiro.
- Zamfir, F., Nunes, J., Teixeira, L., Mealha, Ó., & Santos, B. S. (2004). Aplicação Visual para a Análise de Usabilidade de Sistemas de Informação e Comunicação Sustentados em Tecnologias Web. *Actas da 1ª Conferência Nacional em Interação Pessoa-Máquina - Interação 2004*, 247-249.
- Zamfir, F., Nunes, J., Teixeira, L., Santos, B. S., & Mealha, Ó. (2004). Visual Application for Management of Web-based Communication and Information Systems. *Proceedings of IADIS - International Conference Applied Computing 2004, II*, 119-125.
- Zamfir, F. L. (2007). *Website visualizer: a tool for the visual analysis of website usage*. (Master of Science Dissertation), Universidade de Aveiro Aveiro, Portugal. Retrieved from <http://biblioteca.sinbad.ua.pt/Teses/2008000996>
- Zhou, B., Chen, J., Shi, J., Zhang, H., & Wu, Q. (2001). Website Link Structure Evaluation and Improvement Based on User Visiting Patterns. *Proceedings of the 12th ACM Conference on Hypertext and Hypermedia*, 241-242.

# **ANEXOS**

# 1. Organização/arquitetura da informação recolhida

## 1.1. Representação do Site

### Estrutura de dados

#### Lista Páginas Site

N itens // Dimensão da lista  
Página [N] // Array de páginas

#### Página

N URLs // Dimensão da lista de URLs  
URL [N] // Lista de URLs (inclui parâmetros)  
RAW // Conteúdo da página  
Nível // Na estrutura em árvore  
Tipo documento // HTML, PDF, etc  
Dinâmico // Dinâmico ou não  
Dimensões // Altura, Largura  
Imagem // Representação no ecrã da página  
Lista Links // Lista de *links* da página  
Lista Referers // Lista de páginas com *links* para esta página  
Lista Objectos // Lista de objectos/blocos classificados  
Lista HotSpots // Lista das zonas clicáveis

#### Lista *Link*

N itens // Dimensão da lista  
URL [N] // Array de URLs (inclui parâmetros)

#### Lista Referers

N itens // Dimensão da lista  
URL [N] // Array de URLs (inclui parâmetros)

#### Lista Objectos

N itens // Dimensão da lista  
Objecto [N] // Array de objectos/blocos

#### Objecto

Vértices (x,y) // Coordenadas dos vértices sup-esq e inf-dir  
Categoria // Tipo de objecto  
Dinâmico // Dinâmico ou não

#### Lista HotSpots

N itens // Dimensão da lista  
HotSpot [N] // Array de *HotSpots*

#### HotSpot

URL // URL (inclui parâmetros)  
Dinâmico // Dinâmico ou não

---

 Objecto // Objecto/bloco

## Descrição da estrutura de dados

Lista Páginas <i>Site</i>	Lista contendo todas as páginas existentes no <i>site</i> no momento do estudo
N itens	Dimensão da lista – número de páginas que constituem o <i>site</i>
Página [N]	<i>Array</i> contendo cada página do <i>site</i>
Página	Informação referente a cada página do <i>site</i>
N URLs	Dimensão da lista – número de URLs que podem identificar a página
URL [N]	<i>Array</i> contendo cada URL que pode identificar a página – note-se que o URL inclui os parâmetros
RAW	Conteúdo da página – independentemente do tipo de conteúdo (HTML e/ou imagens, <i>flash</i> , PDF, etc.)
Nível	Na estrutura em árvore – correspondente ao percurso mais curto para chegar à página
Tipo documento	HTML, PDF, etc.
Dinâmico	Página dinâmica ou estática
Dimensões	Altura e Largura da representação no navegador / ecrã do utilizador
Imagem	Representação da página no navegador / ecrã que o utilizador obtém
Lista <i>Links</i>	Lista de todos os <i>links</i> contidos na página
Lista Referers	Lista das páginas, em todo o <i>site</i> , que contêm <i>links</i> para esta página
Lista Objectos	Lista de objetos / blocos classificados da página – por exemplo: área de texto, imagem, zona com lista de <i>links</i> , etc.
Lista <i>Hotspots</i>	Lista das zonas / objetos / blocos clicáveis (que contêm <i>links</i> )
Lista <i>Links</i>	Lista de todos os <i>links</i> contidos na página
N itens	Dimensão da lista – número de <i>links</i> / URLs que constituem a lista
URL [N]	<i>Array</i> contendo todos os <i>links</i> / URLs – note-se que dois URLs que só diferem nos seus parâmetros são considerados distintos
Lista Referers	Lista das páginas, em todo o <i>site</i> , que contêm <i>links</i> para esta página
N itens	Dimensão da lista – número de <i>links</i> / URLs que constituem a lista

URL [N]	<i>Array</i> contendo todos os <i>links</i> / URLs – note-se que dois URLs que só diferem nos seus parâmetros são considerados distintos
Lista Objectos	Lista de objetos / blocos classificados da página – por exemplo: área de texto, imagem, zona com lista de <i>links</i> , etc.
N itens	Dimensão da lista – número de objetos/blocos que constituem a lista
Objecto [N]	<i>Array</i> contendo cada objeto/bloco classificado na página
Objecto	Objeto / bloco, de forma retangular, classificado na página – correspondente a uma categoria de conteúdo
Vértices (x,y)	Coordenadas dos vértices superior-esquerdo e inferior-direito do retângulo que delimita a área em causa
Categoria	Tipo de objeto – por exemplo: texto, imagem, área de <i>links</i> , etc.
Dinâmico	Dinâmico ou estático – por exemplo, há objetos que se alteram à passagem do cursor do rato por cima.
Lista <i>Hotspots</i>	Lista das zonas / objetos / blocos clicáveis (correspondente a um <i>links</i> )
N itens	Dimensão da lista – número de <i>Hotspots</i> que constituem a lista
<i>Hotspot</i> [N]	<i>Array</i> contendo cada <i>Hotspot</i> da página
<i>Hotspot</i>	Zonas / objeto / bloco clicável (correspondente a um <i>link</i> )
URL	<i>Link</i> associado ao objeto – identificação da página de destino
Dinâmico	Dinâmico ou estático – por exemplo, há objetos que se alteram à passagem do cursor do rato por cima.
Objecto	Objeto / bloco, de forma retangular, classificado na página – correspondente a uma categoria de conteúdo

## 1.2. Representação de uma sessão de um utilizador

### Estrutura de dados

```

Lista Páginas Sessão
  ID utilizador      // Identificação do utilizador
  N itens           // Dimensão da lista
  Página Sessão [N] // Array páginas da sessão
    
```

```

Página Sessão
    
```



```

Página // Página visitada
Hora // Data/hora da ocorrência (visita à página)
Lista Eventos Teclado // Monitorização do teclado
Lista Eventos Rato // Monitorização do rato
Lista Eventos Oculares // Monitorização ocular

Lista Eventos Teclado
N itens // Dimensão da lista
Evento Teclado [N] // Array eventos teclado

Lista Eventos Rato
N itens // Dimensão da lista
Evento Rato [N] // Array eventos rato

Lista Eventos Oculares
N itens // Dimensão da lista
Evento Ocular [N] // Array eventos oculares

Evento Teclado
Hora // Data/hora da ocorrência
N itens // Dimensão da lista
Tecla [N] // Array contendo as teclas usadas (em simultâneo)

Evento Rato
Hora // Data/hora da ocorrência
Evento // Mouse down, up, move, drag, etc
Coordenadas (x,y) // Posição do cursor

Evento Ocular
Hora // Data/hora da ocorrência
Evento // hold, move, blink, etc
Coordenadas (x,y) // Posição do olhar no ecrã

```

## Descrição da estrutura de dados

Lista Páginas Sessão	Lista contendo todas as páginas visitadas por um utilizador durante uma dada sessão
ID utilizador	Identificação do utilizador
N itens	Dimensão da lista – número de páginas visitadas pelo utilizador – há a possibilidade de uma página ser visitada mais que uma vez
Página Sessão [N]	Array contendo cada página visitada durante a sessão
Página Sessão	Página visitada durante uma sessão de um utilizador
Página	Informação referente à página do <i>site</i> visitada
Hora	Data / hora da ocorrência (visita à página)
Lista Eventos Teclado	Lista de eventos resultantes da monitorização do teclado – teclas premidas
Lista Eventos Rato	Lista de eventos resultantes da monitorização do rato – movimento, telas, etc.

Lista Eventos Oculares	Lista de eventos resultantes da monitorização do globo ocular do utilizador – direção do olhar, etc.
Lista Eventos Teclado	Lista de eventos resultantes da monitorização do teclado
N itens	Dimensão da lista – número de eventos da lista
Evento Teclado [N]	Array contendo cada evento produzido pelo teclado
Lista Eventos Rato	Lista de eventos resultantes da monitorização do rato
N itens	Dimensão da lista – número de eventos da lista
Evento Rato [N]	Array contendo cada evento produzido pelo rato
Lista Eventos Oculares	Lista de eventos resultantes da monitorização do globo ocular do utilizador
N itens	Dimensão da lista – número de eventos da lista
Evento Ocular [N]	Array contendo cada evento desencadeado pelo globo ocular do utilizador em estudo
Evento Teclado	Evento produzido pelo teclado
Hora	Data / hora da ocorrência do evento
Tecla [N]	Conjunto de teclas usadas (em simultâneo) – por exemplo para identificar a utilização de uma tecla e modificadores (Alt, Ctrl, Shift, etc)
Evento Rato	Evento produzido pelo rato
Hora	Data / hora da ocorrência do evento
Evento	Identificação do evento: premir / soltar a tecla esquerda / direita / central do rato, mover o rato, arrastar (mover enquanto se prime uma tecla), etc.
Coordenadas (x,y)	Posição do cursor – coordenadas X e Y no ecrã
Evento Ocular	Evento desencadeado pelo globo ocular do utilizador em estudo
Hora	Data / hora da ocorrência do evento
Evento	Identificação do evento: mover o olhar, deter o olhar num local, pestanejar, etc.
Coordenadas (x,y)	Posição do olhar no ecrã – coordenadas X e Y no ecrã

## 1.3. Representação de múltiplas sessões

### Estrutura de dados

```

Lista Sessões           //
  N itens               // Dimensão da lista
  Sessão [N]           // Array de sessões
  Critério              // Critério de selecção das sessões
  Método integração    // Para o processo de visualização

```

### Descrição da estrutura de dados

Lista Sessões	Lista com o conjunto de sessões em análise – de um ou mais utilizadores
N itens	Dimensão da lista – número de sessões em análise
Sessão [N]	Array contendo cada sessão da análise (descrita na secção 1.2 deste capítulo)
Critério	Critério de seleção das sessões – por exemplo: por utilizador, data de início, duração, percurso utilizado, etc.

## 1.4. Informação estatística

### Estrutura de dados

```

Página                 // Informação referente a cada página
  Visitas              // Quanto às visitas a cada página
    Número            // N° de visitas
                     // Mais visitada, menos, média, etc.
  Datas [N]           // Data/hora de cada visita
                     // Hora, dia mais populares, etc.
  Duração [N]         // Duração de cada visita
                     // Max, min, média, etc.
  Utilizadores [N]    // ID do utilizador em cada visita
                     // Utilizador mais frequente, ocasional, ...

Objectos
  Localização         // Área do ecrã
                     // Áreas mais comuns, densidade espacial, ...
  Enquadramento no Campo Visual // Distribuição no campo visual
  Conformidade com design específico // Respeitam layout?
                                     // Coerência da interface
  Objectos mais visitados/observados // Da monitorização
                                     // Ocular e/ou rato
  Categorias          // Categorias mais presentes, distribuição, ...

Referers              // De onde vêm os visitantes (páginas)
  URLs                // Lista de URLs
  Datas               // Data/hora da visita à página anterior
  Tempos              // Tempo entre a página anterior e a corrente

Links usados         //
  URLs                // Lista dos URLs usados
  Datas               // Data/hora da utilização

```

```

    Tempos           // Tempo entre a apresentação e o uso
HotSpots usados    //
HotSpots           // Lista dos HotSpots
Quantas vezes     // Lista de utilizações de cada HotSpot
Datas             // Data/hora da utilização
Tempos           // Tempo entre a apresentação e o uso

Sessão            //
    Duração         // Duração Max, min, media, ...
    Páginas [N]     // Array com as páginas visitadas em cada sessão
    Percursos [M]   // Array com os percursos mais usados, ...
    Áreas mais visitadas // Zonas do site mais visitadas, ...

Rato              //
    Eventos         // Cliks, movimento, arrastar, ...
    Movimentos     // Percursos mais comuns, km de rato, ...
    Tempos         // Tempo entre clicks, média, ...

Ocular            //
    Zonas observadas // Zona do ecrã mais observada, menos, ...
    Tempos         // Tempo de observação médio por área, ...

```

## Descrição da estrutura de dados

<b>Página</b>	Informação estatística referente a cada página
<b>Visitas</b>	Informação referente às visitas de cada página
<b>Objectos</b>	Informação referente aos objetos / blocos identificados em cada página
<b>Referers</b>	De onde vêm os visitantes (páginas) para chegar a uma página
<b>Links usados</b>	Informação referente aos <i>links</i> usados na página
<b>Hotspots usados</b>	Informação referente aos <i>Hotspots</i> usados na página
<b>Visitas</b>	Informação referente às visitas de cada página
<b>Número</b>	Nº de visitas à página – página mais visitada, menos, média, etc.
<b>Datas [N]</b>	Data / hora de cada visita – dia/hora mais populares, etc.
<b>Duração [N]</b>	Duração de cada visita – duração máxima, mínima, média, etc.
<b>Utilizadores [N]</b>	ID do utilizador em cada visita – utilizador mais frequente, ocasional, ...
<b>Objectos</b>	Informação referente aos objetos / blocos identificados em cada página
<b>Localização</b>	Área do ecrã – áreas mais comuns contendo um dado tipo de objeto, densidade espacial, ...
<b>Enquadramento Visual</b>	Distribuição dos objetos no campo visual do utilizador
<b>Conformidade design</b>	Respeitam <i>layout?</i> – Coerência da interface ao longo das várias páginas do <i>site</i>

Hits	Objetos mais visitados / observados – Da monitorização ocular e/ou rato
Categorias	Categorias mais presentes, distribuição das categorias, etc.
Referers	De onde vêm os visitantes (páginas) para chegar a uma página
URLs	Lista de URLs das páginas de onde vêm os visitantes (dentro ou fora do <i>site</i> em análise)
Datas	Data / hora da visita à página anterior
Tempos	Tempo decorrido entre a visita à página anterior e a visita à corrente
<i>Links</i> usados	Informação referente aos <i>links</i> usados na página
URLs	Lista dos URLs usados
Datas	Data / hora da utilização
Tempos	Tempo decorrido entre a apresentação da página e o uso de um <i>link</i>
<i>Hotspots</i> usados	Informação referente aos <i>Hotspots</i> usados na página
<i>Hotspots</i>	Lista dos <i>Hotspots</i> usados
Quantas vezes	Lista de utilizações de cada <i>Hotspot</i>
Datas	Data / hora da utilização
Tempos	Tempo decorrido entre a apresentação da página e o uso de um <i>Hotspot</i>
Sessão	Informação referente às sessões de utilização de um <i>site</i>
Duração	Duração máxima, mínima, média, etc. de uma sessão
Páginas [N]	<i>Array</i> com as páginas mais visitadas, ...
Percursos [M]	<i>Array</i> com os percursos mais usados, ...
Áreas mais visitadas	Zonas do <i>site</i> mais visitadas, ...
Rato	Informação relacionada com utilização do rato durante a navegação no <i>site</i>
Eventos	Cliques, movimento, arrastar, ...
Movimentos	Percursos mais comuns, km de rato, ...
Tempos	Tempo entre cliques, média, ...
Ocular	Informação relacionada com monitorização do movimento ocular dos utilizadores durante a navegação no <i>site</i>

Zonas observadas Zona do ecrã mais observada, menos, ...

Tempos Tempo de observação médio por área, ...

## 2. Extração de informação de um ficheiro de *log* de um servidor *web*

Nesta secção vai ser apresentada a informação registada numa entrada de um ficheiro de *log* de dois servidores distintos. Para maior facilidade de comparação, o pedido efetuado a ambos os servidores são exatamente iguais, permitindo assim apreciar as diferenças entre os registos de ambos.

Como visto na secção 3.4.1 do Capítulo III, é importante referir que muitas vezes o servidor não obtém informação para registar em todos os campos, sendo essa ausência de informação gerida de forma dispare de servidor para servidor. A questão não se coloca só em servidores de fabricantes diferentes, mas também na forma como cada administrador configura o seu servidor.

### 2.1. Exemplo para um servidor Microsoft IIS 5.0 (aderente ao formato W3C)

Os campos presentes num ficheiro de *log* com o formato W3C proveniente de um servidor *web* Microsoft IIS 5.0 são os seguintes:

date	Date
time	Time
c-ip	Client IP Address
cs-username	User Name
s-sitename	Site Name
s-computername	Server Name
s-ip	Server IP
s-port	Server Port
cs-method	Method
cs-uri-stem	URL Requested
cs-uri-query	URL Query
sc-status	Protocol Status
sc-win32-status	Win32 Status

---

sc-bytes	Bytes Sent
cs-bytes	Bytes Received
time-taken	Time Taken
cs-version	Protocol Version
cs-Host	Server Host
cs(User-Agent)	User Agent
cs(Cookie)	Cookie
cs(Referer)	Referrer

Uma linha de um ficheiro de *log* com o formato W3C pode ter o seguinte aspecto:

```
2003-02-18 10:12:38 192.168.187.211 ihc-ect W3SVC4 CPJ-SRV 193.137.85.3
80 GET /ihc/Contactos.htm param=top 200 0 8225 511 16 HTTP/1.1
www2.ca.ua.pt
Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+5.0;+.NET+CLR+1.0.3705)
ASPSESSIONIDSAQDDADD=BEEJIDABBOOMKMBGHPIIINDC
http://www2.ca.ua.pt/ihc/Index.htm
```

A que corresponde a obtenção da seguinte informação:

Date	2003-02-18
Time	10:12:38
Client IP Address	192.168.187.211
User Name	ihc-ect
Site Name	W3SVC4
Server Name	CPJ-SRV
Server IP	193.137.85.3
Server Port	80
Method	GET
URL Requested	/ihc/Contactos.htm
URL Query	param=top
Protocol Status	200
Win32 Status	0
Bytes Sent	8225
Bytes Received	511
Time Taken	16
Protocol Version	HTTP/1.1
Server Host	www2.ca.ua.pt
User Agent	Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+5.0)

Cookie	ASPSESSIONIDSAQDDADD= BEEJIDABBOOMKMBGHPIIINDC
Referrer	http://www2.ca.ua.pt/ihc/Index.htm

## 2.2. Exemplo para um servidor Unix Apache genérico

Os campos presentes num ficheiro de *log* proveniente de um servidor *web* Unix Apache genérico são os seguintes:

ip	Client IP Address
ident	Ident Client Info
username	User Name
date	Date
time	Time
method	Method
url	URL Requested
url-param	URL Query/Params
protocol	Protocol Version
status	Protocol Status
bytes	Bytes Sent
ref	Referer
agent	User Agent
cookie	Cookie

Uma linha de um ficheiro de *log* proveniente de um servidor *web* Unix Apache genérico pode ter o seguinte aspecto:

```
192.168.187.211 jnunes@cpj1.ca.nat.ua.pt ihc-ect [18/Fev/2003:10:12:38  
+0000] "GET /ihc/Contactos.htm?param=top HTTP/1.1" 200 8225  
"http://www2.ca.ua.pt/ihc/Index.htm" "Mozilla/4.0 (compatible; MSIE 6.0;  
Windows NT 5.0; .NET CLR 1.0.3705)"  
ASPSESSIONIDSAQDDADD=BEEJIDABBOOMKMBGHPIIINDC
```

A linha é constituída pelos seguintes campos:

Client IP Address	192.168.187.211
Ident Client Info	jnunes@cpj1.ca.nat.ua.pt
User Name	ihc-ect



---

Date	2003-02-18
Time	10:12:38
Method	GET
URL Requested	/ihc/Contactos.htm
URL Query	param=top
Protocol Version	HTTP/1.1
Protocol Status	200
Bytes Sent	8225
Referer	http://www2.ca.ua.pt/ihc/Index.htm
User Agent	Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+5.0)
Cookie	ASPSESSIONIDSAQDDADD= BEEJIDABBOOMKMBGHPIIINDC