



Universidade do Minho

Escola de Engenharia

Teresa Susana Mendes Pereira Bernardino

**Perspectiva sobre a Utilização da
Tecnologia RSS no Contexto da
Comunicação Científica**

Dezembro 2006



Universidade do Minho

Escola de Engenharia

Teresa Susana Mendes Pereira Bernardino

**Perspectiva sobre a Utilização da
Tecnologia RSS no Contexto da
Comunicação Científica**

Tese de Mestrado em Sistemas de Informação

Trabalho efectuado sob orientação da

**Professora Doutora Ana Alice Rodrigues Pereira
Baptista**

Dezembro 2006

Aos meus pais,

Carlos Pereira e Emília Pinto

Às minhas irmãs,

Isabel e Ana Raquel

Ao meu marido,

Luís Bernardino

À minha princesa

Isabel

AGRADECIMENTOS

A realização deste trabalho beneficiou da cooperação, assistência, saber e generosidade de diversas pessoas. A todos agradeço, não podendo deixar de expressar a minha particular gratidão:

À Professora Ana Alice Baptista, minha orientadora neste trabalho, os meus sinceros agradecimentos por ter partilhado o seu saber, perspicácia, inteligência e rigor que sempre me incutiu nas suas análises e reflexões, e por todos os momentos em que foi mestre, conselheira e amiga: que assim se mantenha.

À minha irmã Isabel, pelo incentivo, apoio, dedicação e ajuda constantes, por nunca ter deixado de acreditar. Maninha, muito obrigada.

Ao Luís por todo o apoio e ajuda, principalmente nesta última fase, que permitiu o desenvolvimento deste trabalho.

João obrigada pela tua amizade, por aquilo que aprendi contigo e pela constante paciência nos meus desesperos e desânimos. Obrigada pela ajuda a vencer as dificuldades que foram surgindo durante a implementação do protótipo.

Ao Hugo Ravara, pela tua simpatia, generosidade e pela revisão do texto.

Ao Professor Pedro Henriques por poder sempre contar com a sua ajuda principalmente nos momentos mais difíceis. Um obrigada muito grande para si Professor.

Ao Professor Carlos Ramalho por estar sempre disponível em me ajudar. Muito obrigada Professor por poder contar consigo.

Ao Rodrigo Baptista pela ajuda na *Stylesheet*.

À Diana Pereira e à Maria João pelo ânimo, encorajamento e pela ajuda na resolução de todo o processo burocrático e administrativo que surgiu durante o desenvolvimento deste projecto de mestrado.

Por último, gostaria de dedicar este trabalho aos meus pais, às minhas irmãs, ao Luís e à minha princesa. Eles são a razão da minha existência. Eles são a grande fonte de estímulo e motivação ao crescimento e desenvolvimento pessoais. Eles são o segredo do meu progresso. Obrigada.

RESUMO

Actualmente a Internet é uma importante fonte de informação na descoberta de recursos de informação na *Web*. De facto a Internet tornou-se num instrumento de aplicação por parte dos investigadores e cientistas, no desenvolvimento diário do seu trabalho. O seu crescimento tem contribuído para a transformação dos processos de distribuição e disseminação do conhecimento científico produzido no seio das comunidades científicas e consequentemente a reestruturação do sistema de comunicação científica.

Esta dissertação de mestrado tem como objectivo instanciar o sistema desenvolvido no projecto OmniPaper no que se refere aos mecanismos de recuperação de informação distribuída, desenvolvidos no âmbito das notícias publicadas em jornais, e instanciá-los no contexto da publicação científica. Outro objectivo prende-se com a utilização de uma camada central de metadados, desenvolvida com o intuito de proceder à sindicância de conteúdos científicos, seguindo a abordagem tecnológica RSS.

O RSS é um formato normalizado para agregação e distribuição de conteúdos da *Web*, facilitando o processo de consulta e partilha de informação proveniente de diversas fontes de informação, que periodicamente está sujeita a alterações ou actualizações.

A concepção do protótipo pretende alcançar os objectivos propostos nesta dissertação de mestrado. Assim, foi definida uma camada de metadados que suporta as funcionalidades de pesquisa e de navegação desenvolvidas, e permite proceder à sindicância de conteúdos. Na definição camada de metadados estiveram envolvidas diversas tarefas, nomeadamente o levantamento e análise de diversos vocabulários normalizados de metadados e extensamente utilizados no domínio da literatura científica, selecção dos elementos de metadados adequados à descrição dos artigos científicos, seguido da definição do perfil de aplicação e do *template* RSS. As descrições RSS das publicações científicas foram armazenadas numa base de metadados e a sua gestão e manipulação é executada por um sistema de gestão de base de dados nativa RDF. A implementação dos processos de pesquisa e de navegação foram desenvolvidos no protótipo utilizando a versão RDFS do *WordNet* e do sistema de classificação da *Association for Computing Machinery Computing Classification System (ACM CCS)*.

O desenvolvimento destas tarefas resultou num protótipo que é uma instância do sistema desenvolvido no projecto OmniPaper no contexto da publicação científica, e agrega os metadados dos artigos científicos provenientes do repositório da APSI, facilitando a sua pesquisa.

Como trabalho futuro sugere-se, entre outras coisas, a implementação do processo de recolha dos metadados de vários repositórios para a base de metadados aqui implementada possibilitando, assim, o fornecimento de um serviço mais completo e com mais informação.

Palavras-Chave: Revistas científicas, metadados e RSS.

ABSTRACT

At present, the Internet is an important source, which enables the finding and access of information on the Web. In fact, the Internet has become an important tool used daily by researchers and scientists in the development of their work. Its growth has been promoting the transformation of distribution and dissemination processes of knowledge produced by scientific communities and as a result the restructuring of the scientific communication system.

This dissertation, intends to instance the system developed in the OmniPaper project, regarding the mechanisms of distributed information retrieval, developed in the scope of the news published in newspapers, instancing them in the context of the scientific publication. Another goal, concerns the use of a central metadatabase developed to accomplish the syndication of contents, through the RSS approach.

The RSS is a format used for the syndicating of content of news published on the Web, facilitating the access and sharing of information from several sources which are constantly changing.

The implementation of the prototype intends to reach the goals proposed in this master dissertation. In this sense, a metadata layer supporting the research and developed navigation functionalities was defined, and it allows the contents syndication. In the definition of the metadata layer several tasks were involved, namely the analysis of several metadata standard vocabularies widely used in the domain of scientific literature, selection of the metadata elements which best describe the features of the scientific contents. These elements will be included in the application profile and in the RSS template defined. The RSS descriptions of the scientific publications were stored in a metadatabase, through the use of a native RDF database management system. The implementation of the search and navigation processes was developed in the prototype through the use of the RDFS version of the WordNet and the RDFS version of classification system of Association for Computing Machinery Computing Classification System (ACM CCS).

The development of these tasks resulted in a prototype that is an instance of the system developed in the OmniPaper project, in the context of the scientific publication, and aggregates the metadata of the scientific articles provided by the APSI repository, facilitating its research.

As future work, among other things, the implement of metadata harvesting from several repositories to the metadata database implemented in this system is suggested, thus providing a more complete service and one with more information.

Keywords: Journals, metadata and RSS.

ÍNDICE

Agradecimentos.....	iii
Resumo	v
Abstract.....	vi
Índice de Figuras	x
Índice de Tabelas.....	xi
1. Introdução	1
1.1. Objectivos da Dissertação.....	5
1.2. Organização da Dissertação	6
2. Contextualização do Trabalho.....	8
2.1. Introdução	8
2.2. Objectivos do OmniPaper.....	9
2.3. Arquitectura do OmniPaper	10
2.4. Protótipos Desenvolvidos e Testes	14
2.4.1 <i>Local Knowledge Layer</i> (WP2)	14
2.4.2 RDF	16
2.4.2.1 Definição da Estrutura de Metadados.....	18
2.4.2.2 Esquema RDF: Vocabulário omni.....	19
2.4.2.3 Perfil de Aplicação	21
2.4.2.4 Implementação e Manipulação do Código RDF/XML	23
2.4.3 <i>Automatic Keyword Extraction</i> (AKE).....	24
2.4.4 Testes.....	26
2.5. <i>Overall Knowledge Layer</i> (WP3)	27
2.6. Protótipo Final (WP5)	31
2.7. Resumo	32
3. Comunicação Científica	34
3.1. Revistas Científicas Electrónicas.....	35
3.2. Características dos Artigos de Revistas Científicas Electrónicas.....	37
3.3. Resumo	43

4.	RSS.....	45
4.1.	O que é o RSS?	46
4.2.	Evolução do RSS	49
4.3.	Estrutura de um <i>Feed</i> RSS	53
4.4.	A Pesquisa de <i>Feeds</i> RSS	58
4.5.	Como Utilizar o RSS.....	59
4.6.	A Tecnologia RSS na Publicação Científica	62
4.7.	Resumo	68
5.	Descrição Geral do Trabalho Realizado	69
5.1.	Objectivos do Protótipo	70
5.2.	Colecção de Teste.....	71
5.3.	Especificação Funcional do Sistema	73
5.4.	Desenho do Sistema	74
5.5.	Resumo	80
6.	Descrição Detalhada do Trabalho Realizado	81
6.1.	Análise de Vocabulários Normalizados de Metadados.....	81
6.2.	Perfil de Aplicação.....	87
6.3.	Estrutura dos Elementos de Metadados.....	91
6.4.	Representação RSS dos Elementos de Metadados.....	92
6.5.	Processo de Transformação Dos Metadados.....	105
6.6.	Mecanismos de Pesquisa.....	109
6.6.1.	Subscrição dos <i>Feeds</i> RSS.....	111
6.7.	Integração dos Dados	111
6.8.	Resumo	112
7.	Conclusões	113
7.1.	Trabalho Futuro	115
	Bibliografia	116
	Apêndices	123
	Apêndice A.....	124

Apêndice B	127
Apêndice C	133
Apêndice D	144
Apêndice E	147
Apêndice F	156

ÍNDICE DE FIGURAS

Figura 1: Visão geral do Sistema OmniPaper	11
Figura 2: Versão final do sistema OmniPaper.....	13
Figura 3: Propriedades do Esquema omni.....	20
Figura 4: Relação definida entre as classes do esquema omni	21
Figura 5: Vocabulários de metadados utilizados no OmniPaper.....	23
Figura 6: <i>Print Screen</i> da interface do protótipo RDF implementado na WP2	24
Figura 7: Relação dos protótipos desenvolvidos com o sistema AKE	25
Figura 8: Exemplo da ramificação do assunto “ <i>Politics</i> ” representado na estrutura hierárquica dos IPTC-SC modelado em RDF-S.	29
Figura 9: <i>Print Screen</i> da interface do protótipo RDF implementado na WP3	30
Figura 10: Proporção do crescimento da utilização do RSS entre o ano 2001-2006.....	51
Figura 11: Exemplo de um <i>feed</i> RSS.....	57
Figura 12: <i>Print screen</i> do leitor RSS Wizz RSS 2.1.4.....	60
Figura 13: <i>Print screen</i> do leitor RSS AmphedaDesk	61
Figura 14: Camada semântica do sistema.....	74
Figura 15: Desenho do sistema	76
Figura 16: Correspondência dos elementos de metadados descritos segundo o XML <i>schema</i> OAI-PMH para o formato RSS de acordo com o perfil de aplicação definido	106
Figura 17: <i>Print Screen</i> da interface do protótipo desenvolvido.....	110

ÍNDICE DE TABELAS

Tabela 1: Propriedade uniqueID	20
Tabela 2: Propriedade KindOfText.....	20
Tabela 3: Propriedade Supplier	20
Tabela 4: Lista dos principais formatos de agregação de conteúdos	53
Tabela 5: Editoras científicas que utilizam os elementos de metadados do vocabulário do DC e PRISM na criação dos <i>feeds</i> RSS	64
Tabela 6: Editoras científicas que utilizam os elementos de metadados do vocabulário do DC na criação dos <i>feeds</i> RSS	65
Tabela 7: Editoras científicas que não utilizam elementos de metadados na descrição das publicações.....	65
Tabela 8: Identificação do recurso.....	88
Tabela 9: Identificação de quem tem a posse do recurso.....	89
Tabela 10: Relevância do recurso	89
Tabela 11: Classificação do recurso	89
Tabela 12: Informação sobre as ligações do recurso	89

1. INTRODUÇÃO

Actualmente as revistas científicas são um importante meio utilizado na disseminação do conhecimento produzido pelas diversas comunidades científicas. A evolução das tecnologias de informação e comunicação têm contribuído para uma reestruturação dos tradicionais processos de comunicação, em particular das revistas científicas. A generalidade das revistas científicas apenas procede à reprodução electrónica do seu conteúdo já impresso, outras há que já começam a tirar partido das potencialidades que o meio electrónico proporciona, em particular as revistas científicas puramente electrónicas que só existem na Internet [Baptista, 2002].

Efectivamente o meio electrónico está a conquistar cada vez mais adeptos no seio das comunidades científicas. A evolução tecnológica tem contribuído para o desenvolvimento de sofisticados serviços de informação que facilitam o processo de publicação científica e respondem com eficácia às necessidades de conhecimento por parte dos investigadores. Por outro lado, são eliminadas algumas das limitações a que estão sujeitos os tradicionais meios de comunicação em suporte de papel, e promovida a disseminação da informação e a comunicação entre utilizadores.

Os repositórios institucionais que se inserem no movimento de Acesso Livre¹ ao conhecimento científico estão a introduzir uma nova dimensão no processo de acesso aos resultados científicos investigados. Efectivamente estes sistemas de informação têm despertado a atenção das instituições académicas, na medida em que representam “coleções digitais que armazenam, divulgam, preservam e dão acesso à produção intelectual das comunidades académicas”. Os repositórios institucionais ao promoverem o Acesso Livre à literatura científica e académica, permitem dar uma maior projecção e visibilidade do trabalho desenvolvido pelos investigadores e instituições, e conseqüentemente contribui para a redução do monopólio das revistas científicas comerciais, resultando na reestruturação do processo de comunicação científica [Rodrigues *et al.*, 2004].

De facto, a evolução das tecnologias de informação e comunicação e conseqüentemente o desenvolvimento de sofisticados serviços de informação têm tido um papel fundamental na reforma do sistema de comunicação científica. A Internet e a sua crescente utilização têm

¹ do inglês *Open Access*

provocado transformações ao nível dos processos de distribuição e disseminação dos resultados investigados, disponibilizando para além disso, uma grande quantidade de informação e mecanismos de pesquisa e recuperação da mesma que a tornaram numa poderosa fonte de informação.

No entanto, o elevado crescimento da Internet, provocado pelo desenvolvimento descontrolado de páginas *Web* contribui para o aumento do seu volume de informação e consequentemente dificulta o processo de descoberta de recursos de informação e a identificação de actualizações que ocorrem dentro das próprias fontes de informação. Com efeito, a identificação de alterações aos conteúdos disponibilizados por uma determinada página *Web*, implica que o utilizador tenha uma percepção do seu conteúdo anterior para que possa detectar as últimas alterações desenvolvidas sobre essa mesma fonte de informação. Quanto maior for o número de fontes de informação usadas pelo utilizador, mais difícil será para ele manter um registo actualizado sobre o estado de cada uma dessas fontes de informação.

É neste contexto que a utilização do RSS² tem vindo a conquistar alguma popularidade junto dos consumidores de informação on-line, dado que esta tecnologia obedece a um formato normalizado para sindicância³ e distribuição de conteúdos da *Web*, facilitando o processo de consulta e partilha de informação proveniente de diversas fontes de informação, que regularmente estão sujeitas a alterações [Pilgrim, 2002].

No caso das revistas científicas electrónicas que periodicamente publicam artigos, a utilização do RSS trás mais valias aos investigadores e cientistas, já que lhes permite consultar as actualidades e desenvolvimentos científicos que vão evoluindo numa determinada área. Assim, as revistas científicas que disponibilizam as suas publicações sob a forma de *feed* RSS (descrito na secção 4.1), permitindo a sua subscrição, estão a facilitar o seu acesso e a identificação de novos conteúdos que vão sendo publicados. Por outro lado, a agregação de todos os conteúdos devidamente categorizados num único local, permitida pelos leitores RSS, facilita a pesquisa dos diversos conteúdos científicos que se encontram distribuídos por

² O acrónimo RSS pode ser interpretado como: *RDF Site Summary* ou *Rich Site Summary* ou ainda como *Really Simple Syndication*. Os diferentes significados atribuídos ao acrónimo RSS serão descritos no capítulo 4.

³ Esta palavra é a tradução do termo em inglês *syndication*. No dicionário da Porto Editora a palavra sindicância está traduzida como "inspecção que tem por fim inquirir de certos actos; inquérito". No contexto desta dissertação de mestrado o termo sindicância é utilizado para indicar a troca de informação actualizada entre diferentes páginas *Web*.

diferentes revistas científicas, evitando assim, a consulta periódica de diversas fontes de informação à procura de novas publicações.

De facto, já se observam algumas revistas científicas a disponibilizar os seus conteúdos sob a forma de *feed* RSS, apesar de actualmente esta tecnologia estar tendencialmente associada aos *blogs* e às notícias, dado que são unidades de informação que estão sujeitas a constantes alterações.

Este projecto de mestrado foi desenvolvido na sequência do projecto OmniPaper (*Smart Access to European Newspapers*, IST-2001-32174)⁴. Neste trabalho procedeu-se à implementação de uma instância do sistema OmniPaper no contexto das publicações científicas, seguindo a abordagem tecnológica RSS de modo a proceder à sindicância de conteúdos. O enquadramento do projecto OmniPaper é desenvolvido no âmbito das notícias de jornais, enquanto o trabalho de mestrado apresentado nesta dissertação está contextualizado no âmbito das publicações científicas. Embora as notícias publicadas em jornais e as publicações científicas sejam duas unidades de informação com características distintas, também têm semelhanças suficientes para proceder à sua instanciação, com o sistema desenvolvido no projecto OmniPaper.

Com efeito, hoje em dia verifica-se que grande parte dos jornais que publicam as notícias em documentos impressos também procedem à sua reprodução electrónica, disponibilizando-a na Internet. Por outro lado, atendendo à instantaneidade que caracteriza as notícias publicadas em jornais, a sua disponibilização na *Web* traz também significativas vantagens relativamente à versão impressa do jornal. De facto as notícias disponibilizadas na respectiva página *Web* do jornal são permanentemente actualizadas ao longo do dia, para além da própria notícia poder ser complementada com vídeos, áudio, hiperligações, etc. Nas publicações científicas não se verifica a mesma instantaneidade dos conteúdos. No entanto estas estão sujeitas a alterações periódicas, como resultado do trabalho desenvolvido pelos investigadores e cientistas que integram uma determinada comunidade científica. Por outro lado, enquanto na publicação de notícias a comunidade de autores e de leitores é distinta, o mesmo não se verifica na publicação de conteúdos científicos, já que os autores e leitores de conhecimento científico são normalmente os mesmos. Além disso, na publicação de notícias, os seus autores auferem uma remuneração enquanto que na publicação científica não só não são

⁴ <http://www.omnipaper.org>.

pagos, como até pode acontecer os autores terem de pagar para poderem publicar. Por exemplo, quando um autor pretende que o seu artigo fique publicado nas actas de uma conferência, muitas vezes precisa de pagar a taxa da conferência, mesmo que não vá apresentá-lo.

Apesar de existirem diferenças entre a publicação de notícias e a publicação de artigos científicos, ambas têm em comum o facto de tornarem pública uma determinada comunicação, e o desejo de que essa mesma comunicação chegue rapidamente aos seus consumidores, respondendo com eficácia às suas necessidades.

Assim, tirando partido das funcionalidades que a tecnologia RSS disponibiliza com o intuito de adaptá-las no contexto das publicações científicas, a estratégia seguida no desenvolvimento deste trabalho de mestrado passou pelo levantamento de um conjunto normalizado de vocabulários de metadados resultando na definição de uma estrutura de metadados adequada à descrição de publicações científicas. Esta estrutura de metadados é distinta da estrutura de metadados definida no processo de descrição de artigos de notícias publicados em jornais, utilizados no desenvolvimento do projecto OmniPaper, uma vez que se tratam de recursos de informação distintos.

A utilização de modelos de metadados na troca de informação, quer sejam notícias quer sejam conteúdos científicos, ocorre como uma estratégia, no sentido de permitir estruturar e indexar a informação e consequentemente facilitar a descoberta de recursos na *Web*.

As funcionalidades de pesquisa e de navegação semântica implementadas no protótipo desenvolvido neste projecto de mestrado foram instanciadas do sistema OmniPaper, embora no OmniPaper a navegação semântica seja efectuada sobre a estrutura hierárquica representada pelos *International Press Telecommunications Council Subject Codes* (IPTC-SC) [IPTC-SC, 2003], enquanto no âmbito deste projecto de mestrado a navegação semântica é efectuada sobre o sistema de classificação da ACM CCS⁵.

⁵ <http://www.acm.org/class/1998/>.

A validação da solução final é efectuada através da prova de conceito: foi desenvolvido um protótipo utilizando a ferramenta proprietária *RDF Gateway*⁶ que combina um sistema de gestão de base de dados nativa RDF e um servidor HTTP.

1.1. OBJECTIVOS DA DISSERTAÇÃO

Este trabalho de mestrado tem por objectivos verificar a possibilidade de instanciação do sistema OmniPaper no contexto da publicação científica e proceder à sindicância de conteúdos utilizando a base de metadados.

Esta proposta tenta tirar partido das funcionalidades da tecnologia RSS, no que se refere ao acesso e notificação do utilizador acerca de novos conteúdos que surgem todos os dias e a toda hora em diversas fontes de informação, e adapta-las às publicações científicas.

A implementação destes objectivos seguiu o desenvolvimento das seguintes tarefas:

1. Revisão de literatura;
2. Estudo das tecnologias;
3. Definição dos requisitos do protótipo;
4. Desenvolvimento do protótipo
 - 4.1 Levantamento e análise dos vocabulários normalizados de metadados, específicos no domínio da literatura científica;
 - 4.2 Definição e implementação em RDF/XML do perfil de aplicação utilizando o vocabulário *smes*;
 - 4.3 Definição de um *template* em RSS que compreende a estrutura de metadados utilizada na descrição das publicações científicas;
 - 4.4 Codificação em RSS dos artigos armazenados no repositório da APSI utilizando uma *stylesheet* definida em XSLT;
 - 4.5 Criação dos *feeds* RSS para cada categoria do sistema de classificação da ACM CCS;
 - 4.6 Definição e desenvolvimento de uma camada conceptual (através da estrutura hierárquica do sistema de classificação da ACM CCS e a utilização do *WordNet* [WordNet, 2003]);
 - 4.7 Integração e processamento completo dos dados.

⁶ <http://www.intellidimension.com/>.

5. Escrita da dissertação.

A validação é realizada através da prova de conceito consubstanciada na implementação de um protótipo que tem como *data set*⁷ a colecção de documentos existentes no repositório da Associação Portuguesa de Sistemas de Informação (ver secção 5.2) (APSI) [APSI, 2005].

1.2. ORGANIZAÇÃO DA DISSERTAÇÃO

Esta dissertação está organizada em seis capítulos, cujo conteúdo será apresentado de seguida de uma forma resumida.

No capítulo 2 pretende-se fazer uma descrição do projecto OmniPaper, seguido da contextualização e enquadramento da problemática da recolha da informação na Internet, em particular as encontradas pelo projecto OmniPaper.

No capítulo 3 pretende-se efectuar a contextualização do trabalho no âmbito da comunicação científica e das revistas científicas electrónicas.

No capítulo 4 é efectuada uma abordagem ao estado-da-arte da tecnologia RSS. Seguido da apresentação do RSS e a forma de o aplicar no contexto das publicações científicas. Por fim apresentam-se algumas iniciativas por parte de algumas revistas científicas electrónicas que já disponibilizam os seus conteúdos sob a forma de *feed* RSS, permitindo a sua subscrição por parte dos utilizadores.

No capítulo 5 é apresentada uma descrição geral do trabalho realizado. Neste capítulo são apresentados os objectivos que se pretendem alcançar com a implementação do protótipo, acompanhado da especificação funcional do sistema, bem como o desenho do sistema, com a descrição das componentes instanciadas do sistema desenvolvido no projecto OmniPaper, as componentes que foram instanciadas mas que foram alteradas atendendo às especificidades associadas à publicações científicas e as componentes definidas.

⁷ Entende-se por *Data Set* um conjunto de dados, que no âmbito deste trabalho se refere ao conjunto de publicações científicas. É utilizada a expressão em inglês porque é a mais utilizada na área e para não correr o risco de perder semântica ao proceder à sua tradução.

A descrição detalhada do trabalho realizado é apresentada no capítulo 6. Este capítulo é iniciado com a todo o trabalho desenvolvido na definição dos elementos de metadados a serem utilizados na descrição dos diferentes géneros de literatura científica que fazem parte do sistema, acompanhado do perfil de aplicação codificado em RDF/XML. De seguida é apresentada toda a informação relativa à descrição RSS dos artigos publicados pela APSI e termina com a descrição relativa à concepção do protótipo no ambiente de base de dados nativa RDF.

Por fim, descrevem-se as conclusões do trabalho realizado e dão-se algumas pistas sobre trabalho futuro a desenvolver

A dissertação tem seis apêndices.

No apêndice A apresenta-se o *template* do *feed* RSS utilizado para a descrição dos artigos científicos publicados pela APSI.

No apêndice B apresenta-se o perfil de aplicação.

O apêndice C inclui alguns *feeds* RSS definidos para cada categoria do sistema de classificação da ACM CCS.

No apêndice D encontra-se o código XSL relativo à *stylesheet* definida para a transformação da metainformação dos artigos científicos para o formato RSS.

No apêndice E, contém o código RSP do protótipo desenvolvido utilizando o *RDF Gateway*.

Por fim, no apêndice F é apresentada a tradução dos elementos do RSS definidos na especificação do RSS 1.0.

2. CONTEXTUALIZAÇÃO DO TRABALHO

2.1. INTRODUÇÃO

Actualmente a Internet disponibiliza e permite o acesso a uma grande quantidade de informação, tendo-se tornado numa poderosa fonte de informação de importância e relevância inquestionável. No entanto este crescimento em dimensão e diversidade assim como a disponibilização crescente da quantidade da informação dificulta a descoberta de recursos de informação relevantes, executada pelos actuais motores de pesquisa e de indexação. Apesar dos sistemas de indexação e do processo de recuperação de informação estarem cada vez mais desenvolvidos, os resultados devolvidos pelos sistemas aos utilizadores nem sempre atingem os níveis de *recall* e precisão desejados. Exemplo disso é o *Google*, que devolve um elevado universo de resultados, e grande parte deles não correspondem à informação relevante esperada pelo utilizador.

É neste contexto, que se enquadra o projecto *OmniPaper (Smart Access to European Newspapers)*, desenvolvido no âmbito do programa IST (*Information Society Technologies*) da Comissão Europeia, que pretendeu investigar mecanismos que proporcionem aos utilizadores o acesso estruturado, personalizado e multilingue a todo o conjunto diversificado de artigos de notícias digitais de diversos jornais Europeus.

No desenvolvimento do projecto *OmniPaper* estiveram envolvidos um grupo multidisciplinar de especialistas de diferentes países europeus (Bélgica, Portugal, Espanha e Áustria) que trabalharam juntos no desenvolvimento deste projecto, nas seguintes áreas específicas:

- Os fornecedores locais de notícias responsáveis por disponibilizar os conteúdos de notícias de jornais para o desenvolvimento dos protótipos eram:
 - My News, S.L.⁸
 - *pte - presstext.austria Nachrichtenagentur AG*⁹
 - Mediargus¹⁰
- Especialistas na estruturação da informação:

⁸ <http://www.mynews.es/>

⁹ <http://www.presstext.at/>

¹⁰ <http://www.mediargus.be/NL/index.asp>

- *Katholieke Universiteit Leuven*¹¹
- Universidade do Minho¹²
- Inteligência Artificial e especialistas em idiomas:
 - Daedalus - *Data, Decisions and Languages*, S.A.¹³
 - *Universidad Politécnica de Madrid*¹⁴
- Especialistas na análise da usabilidade:
 - CURE - *Center for Usability Research and Engineering*¹⁵

Este grupo constitui o consórcio do projecto OmniPaper, cada um responsável pela implementação das *derivables* especificadas no plano de desenvolvimento do projecto OmniPaper.

2.2. OBJECTIVOS DO OMNIPAPER

O principal objectivo do projecto OmniPaper consistiu na definição de um nível de navegação multilingue e de uma camada lógica sobre as fontes distribuídas de informação. Este objectivo foi alcançado através do desenvolvimento e implementação das seguintes fases que constituíram este processo:

1. Definição e teste de mecanismos de procura eficientes, na recuperação de informação a partir de fontes distribuídas.
2. Definição e teste da criação de um ponto de acesso uniforme às várias fontes distribuídas de informação.
3. Tornar estes pontos de acesso num ambiente mais utilizável e amigável possível.
4. Utilização de colecções digitais distribuídas.

No final deste projecto, os objectivos enunciados foram alcançados através da:

- Criação de um guia de referência (*BluePrint*). Este documento relata todo o trabalho de investigação desenvolvido no âmbito do projecto OmniPaper, assim como os resultados obtidos pelo consórcio durante os três anos de duração do projecto. Este guia de

¹¹ <http://www.kuleuven.be/kuleuven/>

¹² <http://www.uminho.dsi.uminho.pt>

¹³ www.daedalus.es/

¹⁴ <http://www.upm.es/>

¹⁵ <http://www.cure.at/>

referência (*BluePrint*) inclui o estudo do problema, as soluções estudadas e a elaboração de comparações com outras soluções desenvolvidas no acesso a diferentes fontes distribuídas de informação. Este documento dará continuidade e suporte ao desenvolvimento do conhecimento científico que se produza na área da recuperação da informação distribuída.

- Implementação de um protótipo, que permite aos utilizadores (quer aos utilizadores ocasionais, quer aos profissionais) um acesso estruturado e simultâneo a um elevado número de artigos que compõem os jornais digitais Europeus. Um sistema que proporciona aos utilizadores mecanismos estruturados, personalizados e multilingue de acesso a um conjunto diversificado de notícias.

2.3. ARQUITECTURA DO OMNIPAPER

O projecto OmniPaper teve a duração de 3 anos e era composto por sete *workpackages* (WPs): WP1- Gestão do Projecto¹⁶; WP2 – Recuperação de Informação Distribuída¹⁷; WP3 - Camada *Overall knowledge*; WP4 – *Blueprint*; WP5 – Camada de Utilização e Apresentação Amigável do Conhecimento¹⁸; WP6 - Avaliação e Demonstração¹⁹ e a WP7 - Exploração e Disseminação²⁰.

De acordo com a visão geral do sistema OmniPaper, ilustrado na figura 1, e com as especificações exigidas, as *workpackage* WP2, WP3 e WP5 foram implementadas individualmente, resultando na concepção de vários protótipos distintos. As restantes WP's foram desenvolvidas à medida que os protótipos evoluíam e de acordo com os resultados obtidos dos processos de avaliação e testes realizados sobre os protótipos. Por outro lado, estes resultados deram contributos para a implementação do protótipo final.

¹⁶ do inglês *Project Management*

¹⁷ do inglês *Distributed information retrieval*

¹⁸ do inglês *User-friendly presentation of knowledge layer*

¹⁹ do inglês *Evaluation and Demonstration*

²⁰ do inglês *Exploitation and Dissemination*

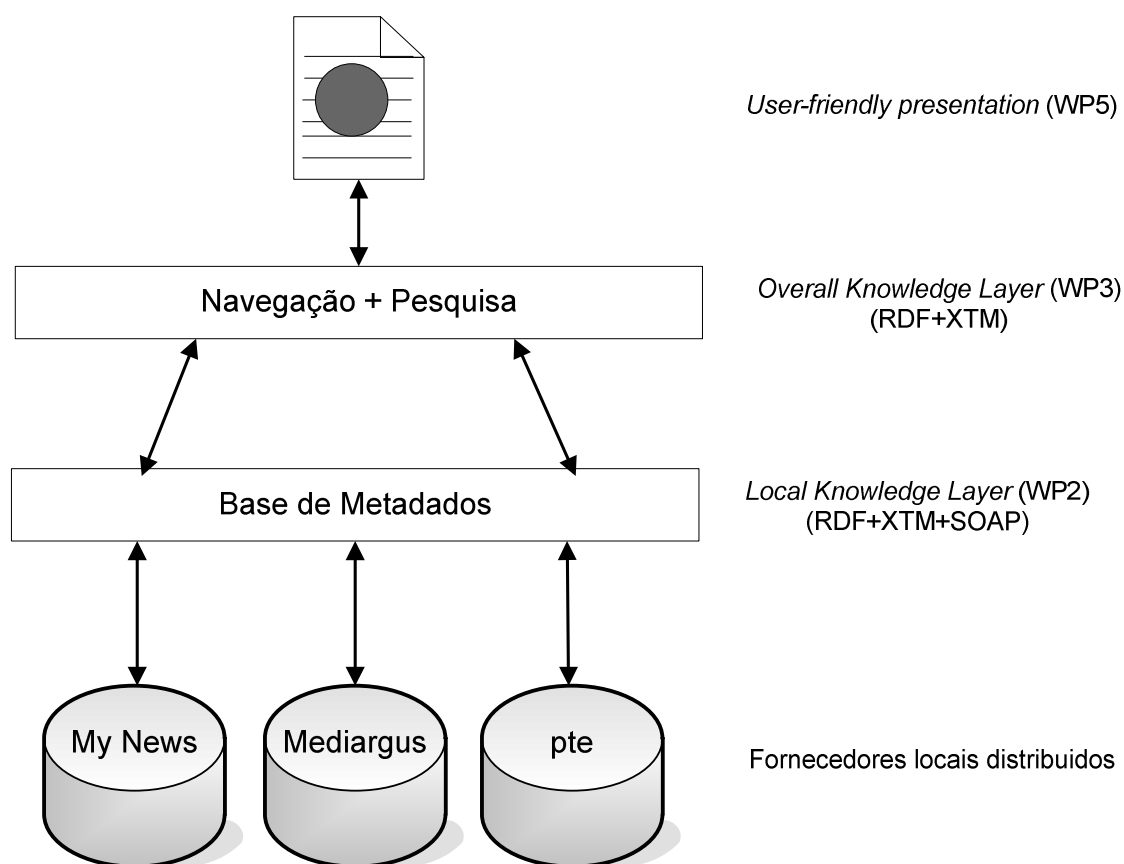


Figura 1: Visão geral do Sistema OmniPaper

Na base do sistema OmniPaper estão as fontes distribuídas, ou seja os fornecedores locais responsáveis pela disponibilização de conteúdos de notícias, que contêm aproximadamente 8,7 milhões de documentos, em diferentes formatos de base de dados e diferentes mecanismos de indexação [Paepen, 2005]. Estas diferenças eram justificadas pelo facto de os conteúdos de notícias serem disponibilizadas pelos diferentes grupos do consórcio responsáveis por esta tarefa, designadamente a *pte*, a *Mediargus* e a *My News*.

A WP2 localizada no primeiro nível da camada local teve início em Março de 2002 e foi a primeira camada a ser implementada. Foram analisadas e desenvolvidas novas tecnologias para melhorar os actuais processos de recuperação de informação distribuída. A utilização de modelos de metadados na troca de informação foi a estratégia seguida, no sentido de facilitar a descoberta de recursos na *Web* e consequentemente melhorar os sistemas de recuperação de informação.

Assim, foi definida uma camada central constituída por uma base de metadados dos artigos descritos, que permitia ao utilizador pesquisar sobre a camada de metadados e apenas

aceder ao conteúdo dos artigos que seleccionou, que permaneciam armazenados na origem de cada fornecedor local de conteúdos de notícias.

A definição da estrutura de metadados seguiu duas abordagens tecnológicas distintas, uma em *Topic Maps* (TM)²¹ e outra em *Resource Description Framework* (RDF)²², resultando na implementação de dois protótipos. Em paralelo ao desenvolvimento destes dois protótipos foi implementado um terceiro protótipo utilizando a tecnologia SOAP. Neste protótipo a recuperação da informação era efectuada directamente sobre as fontes de informação.

A implementação destes três protótipos permitiu realizar o seu teste cruzado contribuindo para a análise e obtenção de conclusões acerca dos desempenhos obtidos no processo de pesquisa e recuperação de informação das tecnologias envolvidas. Os resultados obtidos da realização dos testes efectuados aos protótipos serão na secção 2.4.4.

Os resultados das *queries* efectuadas na camada local constituíram o *input* da gestão dos dados da camada *Overall Knowledge Layer*. Esta camada coordena as características do processo de integração da informação distribuída, com a capacidade de acoplar conteúdos semânticos correspondentes. O processo multilingue é suportado pelo método de extracção de palavras-chave e dos metadados a partir dos arquivos heterogéneos de informação, procedendo à sua associação através da utilização do *WordNet*.

O protótipo desenvolvido nesta camada permitia a pesquisa sobre três arquivos distintos (*My News*, *Mediargus* e *pressetext*) em sete línguas diferentes (Inglês, Francês, Alemão, Espanhol, *Dutch*, Catalão, Português), permitindo aos utilizadores efectuarem as suas pesquisas no seu próprio idioma, obtendo resultados em sete línguas [Paepen, 2005]. Para além do mecanismo multilingue, outras funcionalidades foram implementadas neste protótipo, designadamente a pesquisa simples, a pesquisa avançada e a definição de uma vista semântica de conceitos relacionados [Paepen, 2005]. Estas funcionalidades serão descritas nas secções que se seguem.

Por fim, na WP5 foi desenvolvido o protótipo final que resultou da integração dos dois protótipos implementados na camada *Overall Knowledge Layer*.

²¹ <http://www.topicmaps.org/>

²² <http://www.w3.org/RDF/>

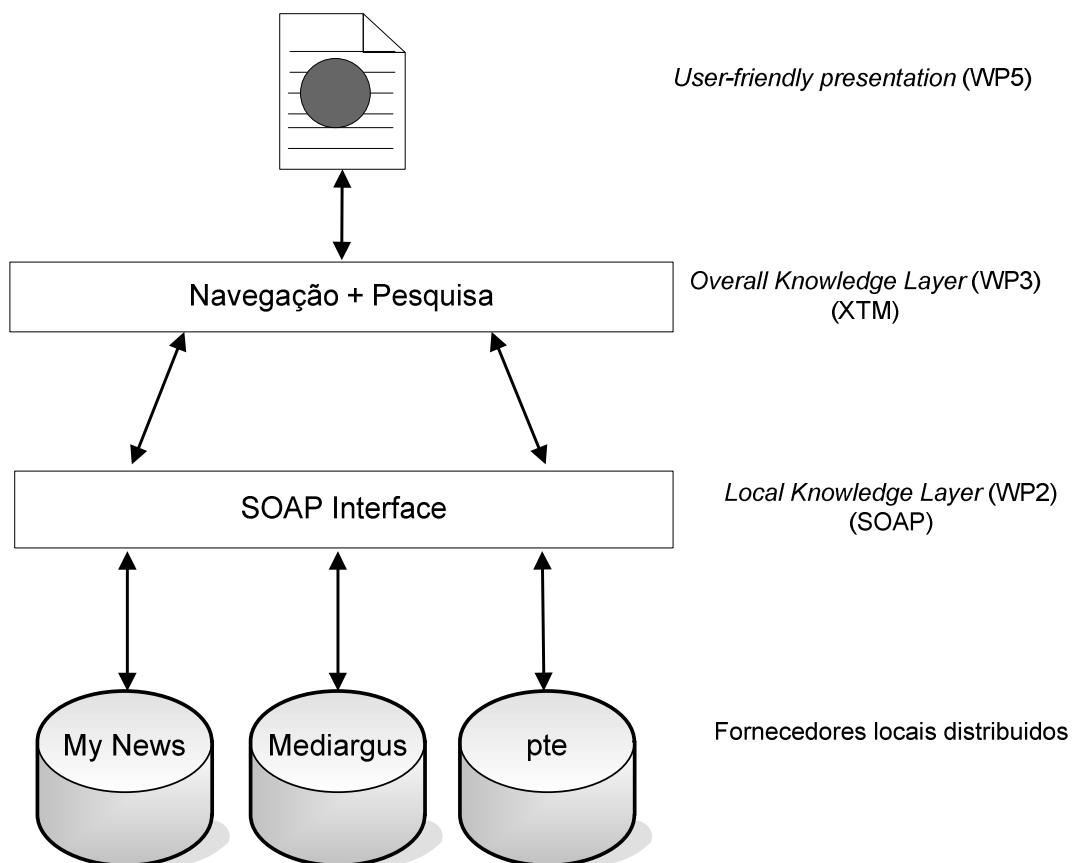


Figura 2: Versão final do sistema OmniPaper

A visão geral do sistema OmniPaper apresentada na figura 1 sofreu algumas alterações relativamente ao plano original do projecto, conforme é ilustrado na figura 2. Inicialmente estava previsto a implementação do protótipo RDF na camada *Local Knowledge Layer* enquanto o protótipo TM era implementado na camada *Overall Knowledge Layer* e posteriormente proceder à integração dos dois protótipos. No entanto os parceiros do consórcio acordaram implementar em simultâneo os dois protótipos e proceder à execução de testes para analisar o desempenho de cada um deles. As componentes com melhor desempenho seriam integradas no protótipo final da WP5.

A base de metadados inicialmente proposta na arquitectura geral do sistema e implementada na camada central não é utilizada no protótipo final, passando o utilizador a efectuar a sua pesquisa directamente sobre as fontes que disponibilizam os conteúdos de notícias através da utilização de *Web Services*. Esta alteração é justificada pelos resultados obtidos nos testes efectuados ao desempenho dos protótipos implementados na *Local knowledge Layer* da WP2. A descrição dos testes efectuados bem como os resultados obtidos que justificam a selecção do protótipo SOAP relativamente ao protótipo RDF e ao protótipo XTM é apresentado na secção 2.4.4.

Deste modo, no protótipo final do OmniPaper, o sistema central passou a ser responsável, pela disponibilização das funcionalidades do processo de refinamento da pesquisa, pela linguagem, pelo controlo de acesso às fontes distribuídas de informação e pela execução da extracção automática de palavras-chave.

Esta opção foi justificada pelo consórcio, tendo em conta a grande volatilidade que caracteriza as notícias, e portanto a utilização de uma camada central num sistema distribuído de recuperação de informação, implicaria a criação de mecanismos adequados de modo a permitir manter actualizada a base de metadados, o que iria contribuir para grandes *overheads* na manutenção diária do sistema. Por outro lado a pressão temporal que os jornalistas estão sujeitos condiciona a utilização de metadados na descrição dos seus artigos [Paepen, 2005]²³.

2.4. PROTÓTIPOS DESENVOLVIDOS E TESTES

O projecto OmniPaper seguiu uma abordagem *bottom-up* na implementação de diferentes protótipos do sistema, através da utilização de tecnologias que se encontravam em estado-da-arte, contribuindo para a sua análise e comparação. Nos protótipos desenvolvidos nas WP2, WP3 e WP5 foram implementados e comparados diferentes métodos de pesquisa.

2.4.1 LOCAL KNOWLEDGE LAYER (WP2)

Como se disse anteriormente, na camada *local knowledge layer* da WP2 foram implementados três protótipos distintos utilizando diferentes tecnologias.

O protótipo SOAP foi desenvolvido nesta camada com o objectivo de analisar e testar a tecnologia SOAP no processo de recuperação directa de informação a partir de arquivos de notícias geograficamente dispersos. Este protótipo é denominado como *Direct Retrieval Approach* uma vez que o método de pesquisa e navegação dos conteúdos de notícias era realizado directamente sobre as fontes de informação.

A utilização da tecnologia SOAP permitiu solucionar problemas provenientes de incompatibilidades entre sistemas computacionais remotos no processo de recuperação de informação, já que o SOAP é um protocolo de comunicação, que permite a troca de dados entre

²³ Esta opção foi do consórcio e não individual, uma vez que os motores de pesquisa como por exemplo o Google já utilizam base de dados centralizadas no processo de recuperação de informação.

diferentes redes e aplicações. Os SOAP *handlers* em cada base de dados do consórcio asseguraram a recuperação uniforme dos artigos. Se uma *query* é efectuada, o protótipo *Direct Retrieval Approach* desenvolvido utilizando o SOAP, remete-a para um ou mais arquivos. Assim tanto o pedido como a resposta estavam em formato XML e a *query* era executada ao nível de cada base de dados local.

Este protótipo continha uma interface simples baseada na *Web* permitindo aos utilizadores navegar pelas diferentes categorias de notícias e executar pesquisas simples ou avançadas. Através do protocolo de comunicação HTTP/SOAP, o sistema trabalhava com o servidor SOAP remoto dos arquivos locais.

Nesta camada, para além da implementação do protótipo SOAP foram também desenvolvidos em simultâneo mais dois protótipos seguindo abordagens tecnológicas distintas, com o intuito de definir mecanismos inteligentes de pesquisa sobre uma camada de navegação multi-arquivo e analisar os desempenhos de cada uma das tecnologias. A implementação destes protótipos passou pela combinação das actuais tecnologias de metadados como o RDF/XML e da tecnologia de administração do conhecimento *Topic Maps*, com o objectivo de ir mais além do que as abordagens de texto integral seguidas pela generalidade dos actuais sistemas de recuperação de informação.

O protótipo implementado em *Topic Maps* foi desenvolvido pelos parceiros do consórcio da *Katholieke Universiteit Leuven*. O objectivo deste protótipo consistia em avaliar de que forma a tecnologia *Topic Maps* poderia ser usada na abordagem *Local Knowledge Layer* [Paepen, 2005].

O protótipo XTM desenvolvido na camada *Local Knowledge Layer* compreende duas funcionalidades [Paepen, 2005]:

1. Pesquisa simples: permite aos utilizadores a pesquisa de palavras-chave dentro dos conceitos identificados, devolvendo os artigos relacionados com esses conceitos. É permitido ao utilizador redefinir a sua pesquisa;
2. Pesquisa avançada: permite aos utilizadores efectuar a pesquisa por palavras-chave, com a possibilidade de aplicar um conjunto extra de constantes aos metadados.

Independentemente do mecanismo de pesquisa seleccionado pelo utilizador, o resultado da pesquisa efectuada, consiste na apresentação dos metadados dos artigos. A recuperação completa dos artigos, aos arquivos de notícias era efectuada através do protocolo de comunicação SOAP [Paepen, 2005].

De seguida é apresentada uma descrição mais detalhada do protótipo implementado em RDF, uma vez que este protótipo foi desenvolvido pelo Departamento de Sistemas de Informação da Universidade do Minho que integrou o consórcio deste projecto e do qual a autora desta dissertação de mestrado esteve envolvida.

2.4.2 RDF

O *Resource Description Framework*²⁴ contém, antes de tudo, um modelo para expressar semântica.

Uma asserção RDF faz declarações sobre recursos, usando uma propriedade e tendo como resultado da aplicação dessa propriedade ao recurso, um valor. Uma asserção pode ser vista como um triplo composto por três elementos: propriedade (predicado), recurso (sujeito) e valor (objecto). Um recurso pode ser qualquer coisa identificável por um URI [Baptista, 2002].

O modelo RDF é simplesmente um modelo de triplos, o que o torna muito poderoso, mas difícil de implementar. Por definição, a descrição usando os triplos, usando o grafo ou usando a sintaxe RDF/XML é equivalente. O *parser* RDF/XML é responsável por ler, verificar a sintaxe RDF/XML, e transformar o código escrito na sintaxe RDF/XML num conjunto de triplos e, eventualmente, num grafo RDF [Baptista, 2002].

O RDF está dividido em duas partes, contendo duas especificações distintas:

1. A *RDF Model and Syntax Specification* (RDFMSS) [Lassila e Swick, 1999] é uma recomendação do W3C que contém um modelo para representar metadados RDF, bem como uma sintaxe para codificar e transportar metadados de forma a maximizar a interoperabilidade de servidores e clientes *Web* desenvolvidos independentemente;
2. A *RDF Schema Specification* [Brickley e Guha, 2000] é uma especificação de esquemas. Com o Esquema RDF podem-se desenhar e implementar de uma forma consistente, vocabulários de metadados específicos. Estes podem ainda ser

²⁴ Tendo em conta que o projecto OmniPaper foi desenvolvido entre 2002 e 2004, tudo o que é apresentado neste capítulo acerca do RDF é sobre a especificação de 1999.

desenvolvidos no seio de outros projectos gerando, assim uma rede de esquemas de metadados.

No âmbito do projecto OmniPaper, o principal objectivo do protótipo RDF desenvolvido na camada *Local Knowledge Layer* consistia em disponibilizar uma descrição semântica normalizada de todos os artigos, de modo a permitir aos utilizadores um acesso estruturado e uniforme aos arquivos de notícias. Consequentemente, foram investigados modelos eficientes na descrição e armazenamento de metainformação disponibilizada pelos fornecedores de notícias, através da utilização do RDF e de tecnologias relacionadas.

O desenvolvimento do protótipo RDF compreendeu a realização dos seguintes passos [Baptista, 2003a]:

1. Definição e desenvolvimento de uma base de metadados;
2. Criação do vocabulário omni usando o RDF-S;
3. Definição do perfil de aplicação;
4. Definição e desenvolvimento de uma camada conceptual (através da estrutura hierárquica de assunto e a utilização de um *thesaurus* léxico);
5. Integração dos protótipos desenvolvidos num protótipo final.

As funcionalidades implementadas no protótipo RDF desenvolvido na WP2 foram as seguintes:

1. Armazenamento da informação: O protótipo desenvolvido permitia o armazenamento de metadados descritos em RDF/XML.
2. Pesquisa avançada: Este mecanismo de pesquisa permitia aos utilizadores pesquisar em todos os campos de metadados. Os resultados são apresentados segundo a relevância do artigo com o título, data e resumo. Numa fase inicial, a informação sobre todos os artigos eram armazenados localmente, passando mais tarde a ser utilizado o SOAP na transferência de informação, no caso de o utilizador pretender consultar o conteúdo completo do artigo.

Este protótipo contém uma interface baseada na *Web* que permite aos utilizadores executar as *queries* e navegar sobre a camada de metadados definida. Todos os metadados são mantidos numa base de dados nativa RDF local, da ferramenta *RDF Gateway*.

2.4.2.1 DEFINIÇÃO DA ESTRUTURA DE METADADOS

De uma forma simples, o termo metadados é definido como dados sobre dados ou informação sobre informação (recursos). O glossário do Dublin Core [Woodley, 2003] define metadados como informação que expressa conteúdo intelectual, propriedade intelectual e/ou características instanciadas de um recurso de informação. No âmbito do projecto OmniPaper, os metadados descrevem recursos que são artigos de notícias e portanto facilitam a sua descoberta e o seu acesso. Por exemplo, para um determinado artigo de um jornal, os metadados consistem em descrever o autor da notícia, identificar o assunto abordado pela notícia, a data da publicação, etc.

Numa primeira fase de implementação do protótipo RDF, no âmbito do projecto OmniPaper, procedeu-se à definição da estrutura de metadados utilizados na descrição dos recursos baseada na análise e selecção dos elementos de vários vocabulários normalizados, nomeadamente do *News Industry Text Format* (NITF)²⁵, *News Agency Implementation Guidelines* (NewsML)²⁶, *Dublin Core Qualifiers* (DCQ) [DCQ, 2002], *Dublin Core Metadata Element Set* (DCMES) [DCMES, 2003] e XMLNews²⁷. A estrutura de metadados descreve vinte e três elementos básicos, agrupados nas seguintes categorias: Identificação, Fornecedor Local, Localização, Relevância, Classificação e Informação das ligações.

Depois de seleccionados os metadados, procedeu-se à sua codificação através da tecnologia RDF. Para além do RDF existe um conjunto diversificado de tecnologias que permite a codificação de metadados, designadamente o HTML ou XML. A *Dublin Core Metadata Initiative* (DCMI) [DCMI, 2002] disponibiliza um conjunto de documentos que auxiliam a codificação de metadados em XML ou HTML. Do mesmo modo, as bases de dados relacionais podem ser utilizadas com o WSDL para disponibilizar a metainformação para o exterior [Paepen, 2005].

A tecnologia RDF foi seleccionada para a codificação dos metadados tomando em consideração os seguintes aspectos [Baptista, 2003b]:

- O RDF é uma recomendação da *World Wide Web Consortium* (W3C)²⁸ para a descrição de recursos, desde 1999;

²⁵ <http://www.nitf.org>.

²⁶ <http://www.newsml.org>.

²⁷ <http://www.xmlnews.org/>.

²⁸ <http://www.w3c.org>.

- É rica para expressar semântica. Isto significa que acrescenta valor às aplicações de metadados, uma vez que não só acrescenta valor à descrição dos recursos como também ao relacionamento entre recursos;
- O RDF é uma infra-estrutura que permite a codificação, troca e reutilização de metadados estruturados;
- A infra-estrutura RDF permite a interoperabilidade de metadados;
- O RDF utiliza o XML (*eXtensible Markup Language*) como sintaxe para codificar e transportar metadados de forma a maximizar a interoperabilidade de servidores e clientes *Web* desenvolvidos independentemente;
- O RDF/XML é uma aplicação XML que contém métodos para expressar semântica, permite a codificação consistente, a troca e o processamento automático de metadados normalizados;
- O RDF permite o uso de convenções que facilitam a interoperabilidade modular entre diferentes conjuntos de elementos de metadados;
- O W3C *Semantic Web Activity* (SWA) teve a sua origem dentro da comunidade de desenvolvimento RDF e ainda se encontra fortemente ligado a esta comunidade. Portanto, todos os desenvolvimentos efectuados no âmbito da tecnologia RDF estão devidamente contextualizados na SWA.

2.4.2.2 ESQUEMA RDF: VOCABULÁRIO OMNI

Durante o processo de definição e estrutura de metadados a serem utilizados na descrição dos recursos, o consórcio OmniPaper acordou na necessidade de definir elementos de metadados que não constavam em vocabulários normalizados e vocabulários largamente utilizados. De modo a alcançar este objectivo, foi criado um vocabulário específico do OmniPaper, denominado omni, que contém estes elementos de metadados, através da utilização do esquema RDF.

Deste modo, e no contexto do projecto OmniPaper são ilustrados, na Figura 3 que se segue, os elementos definidos no esquema do *namespace*²⁹:

²⁹ Um *namespace* XML "é uma colecção de nomes, identificados por uma referência de *Uniform Resource Identifier* (URI), que são usados em documentos XML como tipos de elementos e nomes de atributos" [Bray *et al.*, 1999].

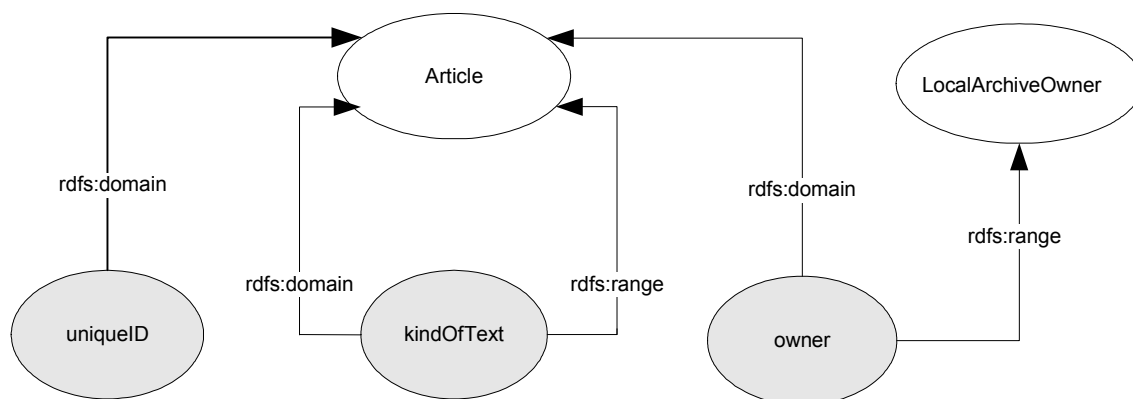


Figura 3: Propriedades do Esquema omni [Baptista, 2003a]

As tabelas que se seguem apresentam a informação associada às propriedades definidas no esquema omni.

Name	uniqueID
Schema	http://www.dsi.uminho.pt/omn/schemas/omn-schema
Encoding Schema	-
subPropertyOf	-
Comment	Identificador do recurso. Como por exemplo o identificador de um artigo de noticias especifico.

Tabela 1: Propriedade uniqueID

Name	KindOfText
Schema	http://www.dsi.uminho.pt/omn/schemas/omn-schema
Encoding Schema	-
subPropertyOf	-
Comment	Identificação do tipo do artigo

Tabela 2: Propriedade KindOfText

Name	Supplier
Schema	http://www.dsi.uminho.pt/omn/schemas/omn-schema
Encoding Schema	-
subPropertyOf	-
Comment	Identificação de todos os fornecedores dos artigos (<i>Local Archives Suppliers</i>).

Tabela 3: Propriedade Supplier

A relação entre as classes é ilustrada na Figura 4.

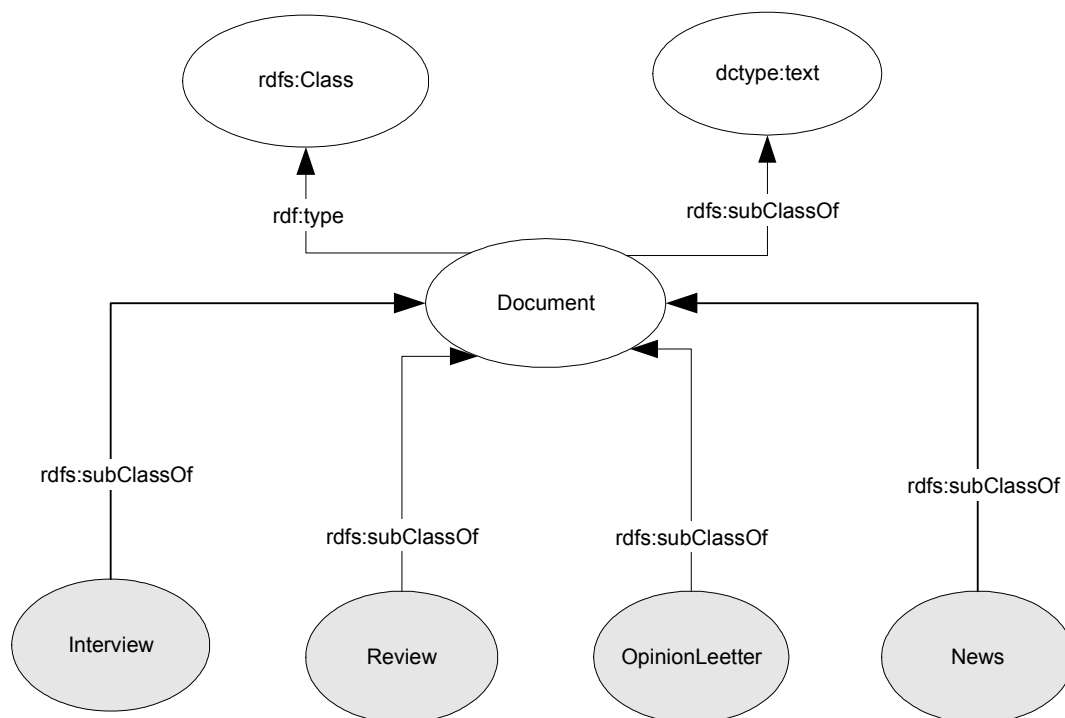


Figura 4: Relação definida entre as classes do esquema omni [Baptista, 2003a]

2.4.2.3 PERFIL DE APLICAÇÃO

As aplicações e Esquemas RDF estão directamente relacionadas com a noção de perfis de aplicações. Enquanto que num esquema RDF, se define um vocabulário, que pode ser usado no contexto de uma ou mais aplicações RDF, com os elementos de cada perfil de aplicação é possível identificar os esquemas RDF, os elementos de cada vocabulário e o seu contexto de aplicação num determinado documento RDF [Heery *et al.*, 2000].

O conceito de perfil de aplicação surgiu nas discussões sobre esquemas de metadados, relacionados com o trabalho desenvolvido no registo de metadados, nomeadamente no *Dublin Core Metadata Initiative* e teve origem na UKOLN, no projecto DESIRE³⁰.

E é neste sentido que é motivado todo o trabalho desenvolvido nos perfis de aplicações, numa tentativa de alcançar um contexto para cada vocabulário.

³⁰ O projecto DESIRE reforçou o “conhecimento comum”, isto é, os especialistas utilizam os esquemas de metadados normalizados de forma pragmática, ou seja, utilizam esquemas de metadados conforme lhes é mais conveniente, uma vez que não existem políticas de utilização de metadados [UKOLN, 1999].

Heery e Patel definem perfis de aplicações como “esquemas que contêm elementos provenientes de um ou mais *namespaces*, combinados por especialistas e otimizados para uma aplicação local em particular” [Heery *et al.*, 2002]. Considerando a distinção entre esquema do *namespace* (contém todos os elementos definidos num determinado *namespace*) e esquema do perfil de aplicação (que contém a combinação de um ou mais subconjuntos de esquemas de *namespaces*).

Efectivamente, os perfis de aplicações consistem na declaração de elementos de metadados cujo *namespace* diz respeito a um determinado esquema, utilizado numa aplicação ou projecto específico, contribuindo, para a reutilização semântica de um ou vários *namespaces*. Mais concretamente, pretendeu-se com a definição do perfil de aplicação identificar os esquemas RDF, os elementos utilizados de cada vocabulário e o contexto de aplicação segundo o género dos documentos que pretendemos descrever.

No âmbito do projecto OmniPaper, foi definido um perfil de aplicação, de forma a descrever detalhadamente os elementos retirados de diferentes vocabulários, definindo os elementos de metadados mais adequados à descrição do género dos artigos de notícias. O perfil de aplicação definido no âmbito do projecto OmniPaper inclui os seguintes seis vocabulários:

- Dublin Core Metadata Element Set (DCMES) – <http://purl.org/dc/elements/1.1/>;
- Dublin Core Qualifiers [DCQ] - <http://purl.org/dc/elements/1.1/>;
- News Industry Text Format (NITF) – <urn:nitf:iptc.org:20010419:NITF>;
- News Markup Language Format (NewsML): <urn:newsml:iptc.org:20010421:NEWSML>;
- Omnipaper RDF Schema [Omnipaper] - <http://www.dsi.uminho.pt/omn/schemas/omn-schema#>;
- vCard - <http://www.w3.org/2001/vcard-rdf/3.0#>.

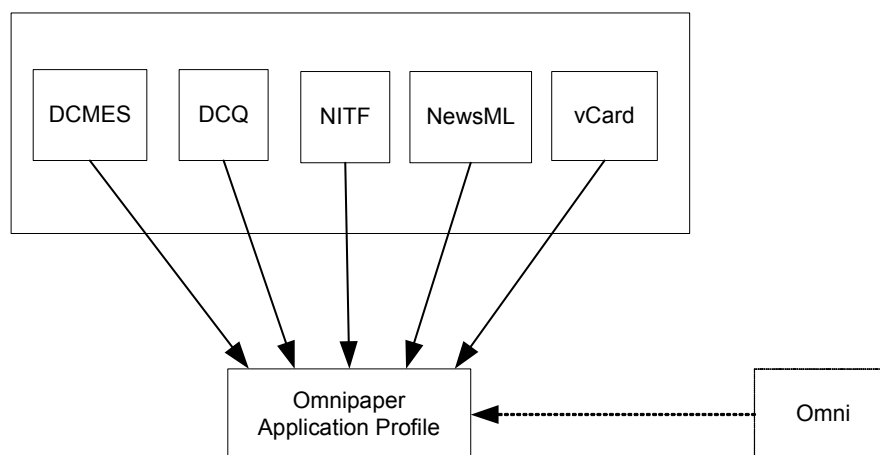


Figura 5: Vocabulários de metadados utilizados no Omnipaper [Yaginuma *et al.*, 2004]

Após a definição do perfil de aplicação com todos os elementos de metadados, procedeu-se ao estabelecimento de regras para a codificação dos metadados e a construção de um *template* RDF/XML. Nas descrições efectuadas sobre os artigos de notícias de jornais europeus foram consideradas as recomendações feitas no documento *Expressing Qualified Dublin Core in RDF/XML*, pelo Kokklink e Schwänzl [Kokklink e Schwänzl, 2002], apesar de este documento ter sido uma recomendação candidata da *Dublin Core Metadata Initiative* (DCMI).

2.4.2.4 IMPLEMENTAÇÃO E MANIPULAÇÃO DO CÓDIGO RDF/XML

Após a selecção dos elementos de metadados adequados à descrição dos artigos de notícias, da definição dos documentos RDF/XML e da sua respectiva validação, procedeu-se à criação de uma base de metadados nativa RDF, que armazenava toda a meta-informação que descrevia os artigos de notícias. A plataforma da base de metadados utilizada foi o *RDF Gateway*.

O *RDF Gateway* é uma ferramenta que conjuga os poderes de um servidor HTTP com o sistema de Gestão de bases de dados nativas RDF. O conteúdo do *RDF Gateway* pode ser acedido através de um *Web browser* especificando o URL da aplicação que faz parte do conteúdo de uma *package* definida no *RDF Gateway*.

As aplicações são desenvolvidas num ambiente *script* denominada *RDF Server Pages (RSP)* semelhante às *ASP (Active Server Pages)* e as *scripts* são implementadas utilizando o *RDF*

Query Language (RDFQL). Como resultado das *queries* efectuadas pelo utilizador, é exibido um conjunto de metadados (título, data e autor) permitindo, caso seja pedido, visualizar o conteúdo do artigo.

Na fase inicial, foi desenvolvido um programa que transforma os artigos XML provenientes dos arquivos locais em ficheiros RDF e efectua o seu *upload* para uma base de metadados. Este programa transforma os artigos e os ficheiros que contêm as *keywords* dos artigos, em ficheiros RDF, procedendo posteriormente ao seu carregamento.



Figura 6: *Print Screen* da interface do protótipo RDF implementado na WP2

2.4.3 AUTOMATIC KEYWORD EXTRACTION (AKE)

Este protótipo foi desenvolvido com o intuito de alcançar dois objectivos [Paepen, 2005]:

1. Criar um sistema de Extracção Automática de Palavras-chave (AKE *Automatic Keyword Extraction*) que permitisse extrair as palavras mais representativas do conteúdo do texto de um artigo, para serem utilizadas pelos restantes protótipos desenvolvidos (XTM e RDF);
2. Desempenhar a pesquisa AKE, para os protótipos que usam o modelo *Vector Space* no processo de recuperação das notícias.

Tendo em conta que os protótipos XTM e RDF utilizam o *WordNet*, no processo de pesquisa e navegação, os artigos tinham de conter algum mecanismo que lhes permitisse estabelecer a ligação com o *WordNet*. É o sistema de Extracção Automática de Palavras-chave

(AKE *Automatic Keyword Extraction*) que estabelece essa ligação através da extracção das palavras-chave mais relevantes do conteúdo do artigo e procede à indexação da base de dados [Paepen, 2005].

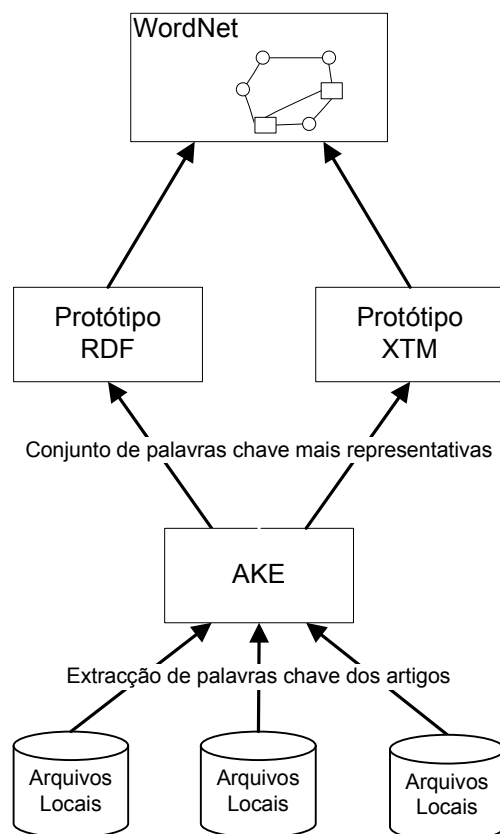


Figura 7: Relação dos protótipos desenvolvidos com o sistema AKE

Num sistema de recuperação da informação, a questão mais importante para o utilizador é obter como resultado da sua pesquisa ou navegação documentos sobre um determinado tópico. Esses tópicos são descritos a partir de um conjunto de palavras-chave do documento. Um problema que surge de imediato, associado a este processo, está em definir um mecanismo automático que permita extrair do documento os termos considerados significativos. Assim, o principal objectivo do sistema AKE consiste em extrair as palavras-chave mais representativas dos documentos, com o objectivo de melhorar a informação relativa aos metadados, e promover

melhorias dos sistemas de recuperação da informação desenvolvidos [Paepen, 2005]. No processo de extracção automática de palavras-chave, foram utilizadas técnicas de *data mining*³¹.

2.4.4 TESTES

No âmbito do projecto OmniPaper foram realizados testes aos diferentes protótipos implementados (SOAP, RDF e XTM), sobre todos os aspectos distintos de cada um deles. Estes testes tiveram como objectivo, por um lado a detecção de erros e de inconsistências e por outro lado analisar as “fraquezas” e a “força” de cada protótipo de modo a obter conclusões estratégicas que contribuíssem na implementação do protótipo final. Os protótipos têm desempenhos diferentes de acordo com a *queries* efectuadas. Esta informação foi importante durante o processo de integração dos protótipos.

Assim o consórcio do projecto realizou um importante esforço na definição de um programa de testes, resultando no desenvolvimento dos seguintes passos: (1) Definição dos critérios de teste; (2) Criação de um conjunto de testes elaborados com base nos critérios definidos; (3) Criação de um Programa Automático de Testes que permitiu o teste rápido e automático dos protótipos desenvolvidos. Finalmente, os diferentes protótipos foram testados, e é com base nos resultados obtidos, que os protótipos foram integrados na melhor “combinação de partes”. Os critérios definidos no desenvolvimento dos testes foram a precisão e o *recall*, o tempo de resposta e o tamanho dos dados.

Relativamente ao critério relevância, os resultados obtidos dos testes efectuados nos protótipos desenvolvidos na WP2 foram os seguintes [Baptista, 2003b]:

- Relativamente ao *recall* e precisão o protótipo RDF obteve melhores resultados que o protótipo XTM, em todos os testes;
- Relativamente à precisão, o protótipo RDF obteve melhores resultados que o primeiro protótipo SOAP implementado e pior que os outros dois;
- Relativamente ao *recall*, o protótipo RDF não obteve resultados tão bons comparativamente com os protótipos SOAP implementados;
- Relativamente ao tempo de resposta, o protótipo RDF obteve melhores resultados que os protótipos XTM e SOAP;

³¹ *Data Mining* consiste no “processo de varrer grandes bases de dados à procura de padrões como regras de associação e sequências temporais, para proceder à classificação ou agrupamento (*clustering*) de itens” [Wikipedia, 2006a].

- Relativamente ao tamanho dos dados, o protótipo RDF obteve melhores resultados que o protótipo XTM.

O resultado obtido nos testes efectuados para o critério *recall* resulta do facto de o protótipo devolver vários resultados que não são os mais relevantes relativamente à *query* executada. De facto, tendo em conta que o protótipo procedeu à procura das palavras-chave no elemento de metadados *omni:keyList*, isto significa que as palavras-chave de cada documento não foram correctamente identificadas pelo sistema AKE. Caso as palavras-chave tivessem sido devidamente determinadas, o protótipo RDF obteria resultados significativamente melhores.

2.5. OVERALL KNOWLEDGE LAYER (WP3)

Os protótipos implementados na camada Overall Knowledge Layer da WP3 resultaram da integração dos protótipos RDF e XTM desenvolvidos na camada *Local Knowledge Layer* da WP2. Os objectivos que se pretenderam alcançar com a implementação destes protótipos foram semelhantes aos objectivos propostos pelos protótipos RDF e XTM implementados, na camada de recuperação de informação distribuída (WP2). Essencialmente, foram adicionadas novas funcionalidades, que facilitaram a navegação semântica sobre uma estrutura hierárquica de assuntos e facilitaram o processo de integração multilígue. No final, o protótipo seleccionado foi o XTM.

A principal alteração efectuada ao protótipo XTM apresentado nesta camada foi na criação de uma vista conceptual relacional que permitia aos utilizadores navegar pelos *topic maps* [Paepen, 2005].

O protótipo XTM definido continha uma interface baseada na *Web* permitindo aos utilizadores executarem *queries* ou refinarem as mesmas *queries* interagindo com o sistema e navegar pelos *Topic Maps*. Toda a informação, designadamente os *Topic Maps* e os metadados é mantida numa base de dados local ao protótipo [Paepen, 2005].

As *queries* efectuadas pelos utilizadores são executadas por um sistema central que procede à pesquisa das palavras-chave e dos conceitos sobre o *Topic Maps*. Este mapeamento contém uma ligação entre as palavras-chave, os conceitos e as relações semânticas entre os conceitos derivado do *WordNet*. A base de dados *WordNet* foi convertida na sua totalidade para o formato XTM [Paepen, 2005].

Relativamente ao protótipo RDF definido na WP3, este também resulta da integração do protótipo implementado na WP2. Ou seja, ao protótipo RDF implementado na WP2 foram desenvolvidas novas funcionalidades de pesquisa e navegação, com o intuito de melhorar o sistema inicialmente desenvolvido.

Assim, ao protótipo RDF implementado na WP3 foi adicionado o mecanismo de navegação semântica sobre uma árvore de conceitos, representada pela estrutura hierárquica dos IPTC-SC e o mecanismo de expansão da *query* através da utilização de uma versão baseada em RDF do *WordNet*.

O vocabulário controlado que compõe os IPTC-SC é constituído por uma estrutura hierárquica de três níveis organizada por um conjunto de termos classificados pelas diferentes categorias. Os tópicos apresentados ao nível dos termos *Subject* contém termos relacionados com a descrição editorial do conteúdo das notícias; ao nível do *SubjectMatter* contém termos relacionados com a descrição a um nível semântico mais preciso, e finalmente o *SubjectDetail* contém termos relacionados com um nível semântico mais específico do conteúdo das notícias.

Para representar os IPTC-SC, várias linguagens foram analisadas e estudadas de forma a seleccionar a que melhor se adaptava à sua representação hierárquica de conceitos. No entanto os IPTC-SC sob o ponto de vista de relacionamento semântico não são assim tão ricos. Deste modo, atendendo à sua simplicidade, uma vez que apenas era necessário definir os seus conceitos hierárquicos, verificou-se que o RDF-S era a linguagem de representação suficiente para descrever a estrutura hierárquica representada nos IPTC-SC.

Após a descrição dos IPTC-SC através da linguagem de representação RDF-S, procedeu-se ao seu armazenamento numa base de metadados. A ligação aos elementos apresentados na árvore hierárquica dos IPTC-SC é efectuada através do elemento de metadados "*dc:subject*". Na definição do perfil de aplicação, é indicado que o "*rdfs:range*" do elemento de metadados "*dc:subject*" são os IPTC-SC [Pereira e Baptista, 2004]. Isto significa que, para cada descrição armazenada na base de metadados, o elemento de metadado "*dc:subject*" apenas pode conter valores dos IPTC-SC.

A figura 11 ilustra uma parte do ramo da árvore de conceitos dos IPTC-SC para o exemplo *dc:subject* “Politics” modelado em RDF-S. No primeiro nível o elemento *dc:subject* é constituído pelo valor “Politics”, que por sua vez contém o valor “Government” como *subject matter*, e os *subject details* associados são o valor “Safety Citizens” e “Civil & Public Services”.

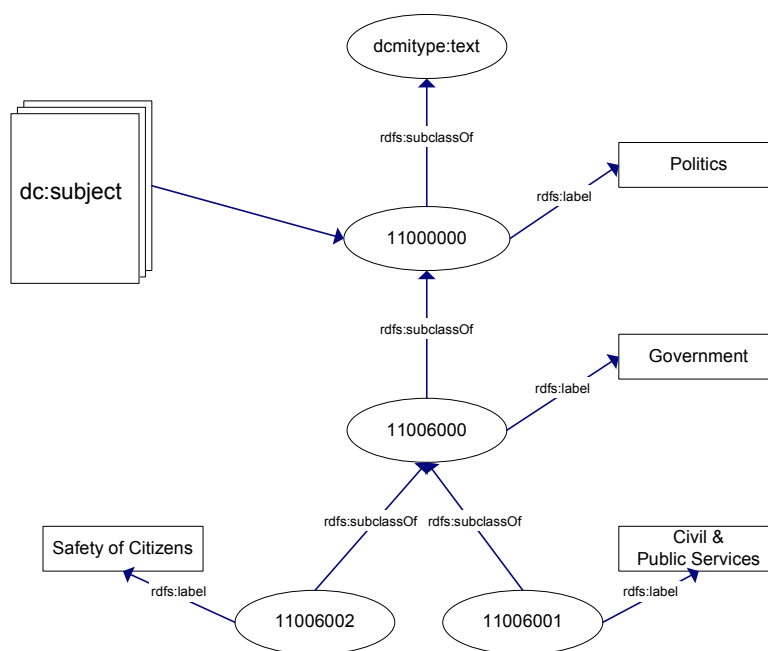


Figura 8: Exemplo da ramificação do assunto “Politics” representado na estrutura hierárquica dos IPTC-SC modelado em RDF-S [Pereira e Baptista, 2004].

O *WordNet* foi criado no laboratório da Ciência Cognitiva da Universidade de *Princeton*. O *WordNet* é um sistema lexical constituído por um extenso conjunto de nomes, verbos, adjectivos e advérbios ingleses, organizados em conjuntos de sinónimos onde cada um representa um conceito lexical.

A versão 1.6 baseada em RDF do *WordNet* foi descarregada e incluída numa base de metadados local. A sua ligação aos artigos foi efectuada através do elemento de metadados “*omni:key_list*”.

Não foi implementada qualquer relação entre o *WordNet* e os IPTC-SC. A única relação existente entre ambos está no facto de quando é executada a pesquisa de um conceito sobre os IPTC-SC, a mesma pesquisa é realizada no *WordNet*, para aquela palavra em particular. De

facto, quando uma pesquisa é efectuada sobre os IPTC-SC, o *WordNet* devolve palavras relacionadas com esse conceito, que podem ser (sinónimos, antónimos, advérbios, etc). Posteriormente, o utilizador ao clicar sobre essas palavras devolvidas pelo *WordNet*, permite-lhe aceder a resultados que não foram previamente devolvidos pelo sistema. Ou seja, o utilizador ao clicar sobre as palavras devolvidas pelo *WordNet*, o sistema simplesmente expande a *query* e desempenha a pesquisa sobre o conteúdo do elemento de metadado *omni:keyList*.

Efectivamente, a pesquisa e a navegação são considerados métodos alternativos e complementares para encontrar informações relevantes. Os dois métodos de pesquisa interagem um com o outro e juntos produzem uma combinação de experiências do utilizador que pode ser exprimida como “*find what you were looking for and then browse away from it*” [Paepen, 2005]. Na verdade, no protótipo é considerado quer a navegação, quer a pesquisa efectuada através da introdução de uma palavra-chave. A única diferença é que na navegação o utilizador segue caminhos predefinidos, enquanto que na pesquisa, o utilizador é totalmente livre de apresentar a pesquisa que pretende.

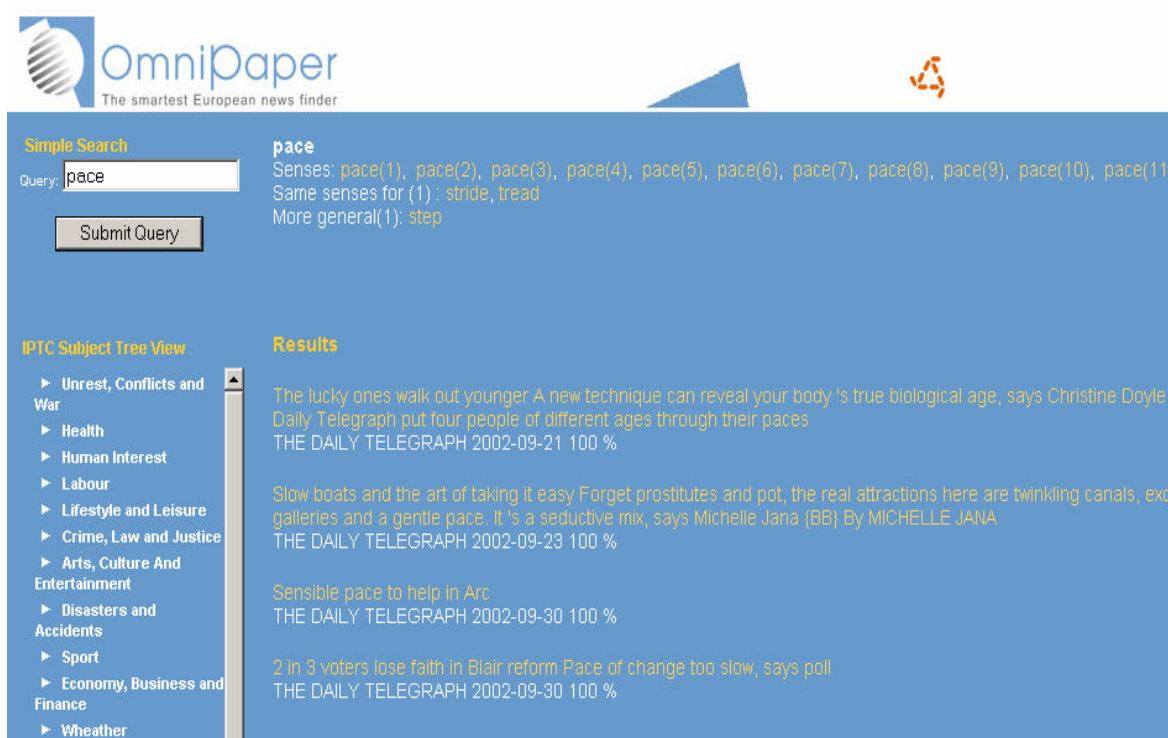


Figura 9: Print Screen da interface do protótipo RDF implementado na WP3

2.6. PROTÓTIPO FINAL (WP5)

O protótipo final implementado na WP5 pretendia ser um demonstrador de um futuro serviço comercial do OmniPaper. Este protótipo resultou da integração do protótipo SOAP implementado na camada *Local Knowledge Layer* da WP2 e do protótipo XTM desenvolvido na camada *Overall Knowledge Layer* da WP3. A este protótipo foram adicionadas novas funcionalidades e à remoção de outras [Paepen, 2005].

Neste protótipo os utilizadores podem efectuar a pesquisa na sua própria língua, permitindo obter resultados em 5 línguas presentes nos arquivos locais. A interface do utilizador é disponibilizada em 7 idiomas (5+ Catalão e Português) [Paepen, 2005].

As funcionalidades básicas estão limitadas à: pesquisa simples, pesquisa avançada, *Web of Concepts* e *Query Tool*.

O modo de funcionamento da *Web of Concepts*, consiste na tradução automática da *query* e da sua expansão em todos os idiomas. Este método de pesquisa permite ao utilizador desambiguar os termos da *query* e proceder ao seu refinamento, através da alteração dos termos da *query* por termos semanticamente relacionados. De seguida, a *query* expandida é enviada para os arquivos locais de notícias, através da utilização do SOAP.

O modo de funcionamento da *Query Tool* consiste em proceder à tradução automática da *query*, mantendo o seu estado semântico original, e permitir ao utilizador refinar a *query* através da adição de termos relacionados ou da remoção de termos da *query*. A *query* também é enviada aos arquivos locais de notícias, através da utilização do SOAP. Este método de pesquisa é diferente do anterior, na medida em que permite ao utilizador controlar a expansão semântica da sua *query*.

Assim, as funcionalidades desenvolvidas que integraram o protótipo final foram as seguintes:

- Pesquisa inteligente de artigos de notícias:
 - Pesquisa simples e avançada de notícias, sobre todos os arquivos locais, usando o SOAP. A pesquisa avançada utiliza campos de metadados, que permitiu limitar a pesquisa. Os elementos de metadados utilizados foram: *Title*, *Subject*, *Key-List*, *Abstract*, *Publisher*, *Creator*, *Issued-From*, *Issued-To*. Por defeito estes campos são combinados através do operador lógico AND.

- Suporte à pesquisa multilingue
 - Pesquisa de notícias em vários idiomas, independentemente do idioma introduzido na *query*;
 - Utilização de um identificador automático de idiomas, permitindo fazer a distinção entre idiomas;
 - Tradução automática dos artigos.
- Métodos de pesquisa
 - Desambiguação da *query* através da *Web of Concepts*;
 - Refinamento da *query* através da *Web of Concepts* e da *Query Tool*;
 - Resultados relevantes através da utilização do sistema AKE;
 - Expansão semântica da *query* através da utilização do EWN.
- Recuperação do artigo: é apresentado o texto completo do artigo, através da execução de um pedido SOAP.

Este protótipo está disponível *on-line*³² aos utilizadores que pretendam consultar estas funcionalidades. No entanto, tendo em conta que o sistema contém artigos com restrições de *copyright*³³ apenas é permitido o acesso e utilização do sistema durante um período experimental de 14 dias.

2.7. RESUMO

No projecto OmniPaper foram investigados mecanismos inteligentes de pesquisa e navegação de modo a proporcionar ao utilizador um acesso estruturado aos artigos de notícias em formato digital de diversos jornais Europeus. Estas funcionalidades foram implementadas e testadas em vários protótipos que foram desenvolvidos ao longo do projecto utilizando diferentes tecnologias na sua implementação, e que permitiram obter conclusões acerca dos desempenhos das diferentes tecnologias.

³² <http://www.omnipaper.org/>

³³ Copyright é definido como “Um direito exclusivo conferido por um governo ao criador de obras literárias ou artísticas originais, como livros, artigos, desenhos, fotografias, composições musicais, gravações, filmes, e programas de computador. O copyright tem alcance internacional e garante, ao criador, os direitos de reprodução, derivação, distribuição, execução e exibição. A Convenção de Berna determina que o período de proteção de copyright cubra a vida do autor mais 50 anos.” [USIA, 1998].

O trabalho desenvolvido no projecto OmniPaper, em particular os desenvolvimentos envolvidos na implementação do protótipo RDF foi o suporte ao desenvolvimento do sistema implementado neste trabalho de mestrado. No entanto, tendo em conta que as publicações científicas e os artigos de notícias são dois recursos de informação com estruturas e conteúdos distintos, é apresentado no capítulo seguinte, as propriedades que caracterizam as publicações científicas disponibilizadas em formato digital, contextualizadas no processo de comunicação científica.

3. COMUNICAÇÃO CIENTÍFICA

A comunicação científica pode ser definida como a partilha de conhecimentos entre membros de uma determinada comunidade científica [Ziman, 1984]. A evolução e desenvolvimento de qualquer área do saber são traduzidos maioritariamente através da literatura científica produzida e divulgada pelos investigadores e cientistas. A maturidade do conhecimento científico é consolidado através da divulgação dos resultados da investigação efectuados pela comunidade científica, e das críticas dos membros dessa comunidade, pois vão contribuir para a validação dos resultados no processo de disseminação do conhecimento. Por outro lado, o reconhecimento dos resultados investigados pelos membros da comunidade científica poderá contribuir para a continuidade do desenvolvimento do conhecimento científico já divulgado e validado, estimulando outros investigadores e cientistas a promover avanços científicos, através da identificação e estabelecimento de novas perspectivas, proporcionando o desenvolvimento de novos trabalhos na área de interesse [Ziman, 1984].

Neste contexto, a comunicação científica tem um papel fundamental, uma vez que promove a cooperação entre investigadores e cientistas, contribui para o reconhecimento de resultados, confirmação de competências profissionais e o estabelecimento de credibilidade e aceitação do investigador e cientista dentro da comunidade científica [Oliveira *et al.*, 2005].

No processo de comunicação científica estão normalmente envolvidas duas actividades essenciais: o de produção, e de disseminação do conhecimento, que pode ser expressa utilizando os tradicionais canais formais e informais de comunicação. A comunicação formal está normalmente associada à comunicação escrita, como por exemplo livros, artigos de revistas científicas, monografias etc. A sua principal vantagem assenta no facto de poder ser armazenada permanentemente, facilitando a sua recuperação e localização. Além disso, como passa pela avaliação de instâncias superiores, contribui para uma maior credibilidade do seu conteúdo. A comunicação informal de comunicação entre pares na comunidade científica está normalmente associada à troca de informação através de canais de carácter mais pessoal ou destituídos de formalismos, como por exemplo conversas informais entre investigadores (pessoalmente, por telefone, ou via correio electrónico), relatos de reuniões científicas e

“colégios invisíveis”³⁴ [Moreira, 2005]. A sua principal vantagem compreende a rapidez do processo de divulgação das informações, facilitando a obtenção de informações e o contacto com outros investigadores e cientistas. Enquanto a sua principal desvantagem está associada ao facto de a informação e conhecimento partilhado estar vinculado a um conjunto restrito de pessoas que integram uma determinada comunidade científica.

A forma como o desenvolvimento da pesquisa é apresentada à comunidade científica, constitui uma parte tão importante da ciência como o embrião da ideia a que deu origem [Ziman, 1984].

3.1. REVISTAS CIENTÍFICAS ELECTRÓNICAS

O crescimento e evolução das tecnologias de informação e de comunicação têm contribuído para a transformação profunda dos tradicionais processos de comunicação formal e informal, estabelecendo um novo canal de comunicação científica: a comunicação científica electrónica.

O crescimento da Internet, e conseqüentemente a sua crescente utilização, proporcionaram um melhor aproveitamento da tecnologia, disponibilizando novos meios e funcionalidades, de suporte ao processo de publicação e comunicação científica, quando comparadas com os tradicionais meios utilizados na comunicação científica baseados em papel [Baptista, 2002]. Os actuais serviços de informação, em particular o sistema de comunicação científica encontram-se num processo de reestruturação profunda, devido à crescente utilização das tecnologias de informação, na criação, organização, armazenamento, preservação e disseminação do conhecimento científico, promovendo a partilha e a troca de conhecimento entre diversas comunidades científicas, e minorando as limitações físicas de armazenamento e disponibilização de conteúdos, dos tradicionais meios de comunicação científica.

No entanto, no seio das comunidades científicas têm surgido alguns obstáculos ao processo de publicação e comunicação científica electrónica, nomeadamente a aceitação de serviços de informação desenvolvidos através da utilização de meios tecnológicos, e

³⁴ do inglês *invisible colleges*. É constituído por investigadores e cientistas dedicados a uma mesma área de pesquisa, com vínculos não formais, que partilham informação entre si. Os elos que unem estes profissionais são as conferências e congressos, as mensagens trocadas via correio electrónico baseadas em listas de discussão [Dias, 1999].

disponibilizados à comunidade científica. Aliado a este facto, são levantadas várias questões, como por exemplo a integridade, e fiabilidade da informação difundida através de meios de comunicação científica electrónica, a propriedade intelectual e direitos de autor [Baptista, 2002]. Efectivamente, são as grandes editoras que ainda detêm o domínio deste mercado, mas atendendo à quantidade de mecanismos tecnológicos desenvolvidos e disponibilizados aos consumidores de conhecimento científico, facilitando-lhes o rápido acesso e a rápida transferência de conteúdos científicos, têm certamente despertado as editoras para a revisão dos seus modelos económicos e consequentemente as suas práticas de mercado.

A publicação científica é o recurso vital para as editoras cuja lógica consiste na maximização do lucro. Por outro lado os investigadores e cientistas, os seus principais clientes, são cada vez mais exigentes e portanto as editoras têm necessidade de reformar os seus tradicionais meios de difusão do conhecimento científico produzido pelas diversas comunidades científicas.

Actualmente, são já reconhecidas importantes vantagens ao canal de comunicação científica electrónica em relação à publicação tradicional baseada em papel, tanto para o editor como para o consumidor final da informação.

No que se refere aos editores, as vantagens da publicação científica electrónica identificadas foram as seguintes:

1. Rápida difusão do conhecimento científico produzido pelas diversas comunidades científicas, permitindo obter uma visão mais geral do estado actual das pesquisas desenvolvidas numa determinada área científica;
2. Acessibilidade, eliminando as limitações de acesso, condicionadas pela tradicional comunicação científica baseada em papel;

Segundo a perspectiva do utilizador, enquanto leitor, produtor, membro de um corpo editorial, revisor, etc., as vantagens da publicação científica electrónica identificadas foram:

1. O rápido e o baixo custo de acesso à informação, eliminando os custos associados à reprodução e transporte, sujeitos nos tradicionais meios de comunicação científica baseada em papel;
2. Facilita a realização de uma cópia e/ou a impressão;
3. A informação encontra-se mais actualizada e fácil de localizar, através de mecanismos de procura sofisticados;

4. Possibilidade de diálogo directo com os autores.

No contexto dos periódicos, são identificadas duas categorias distintas: a reprodução electrónica do conteúdo de uma revista já impressa e as revistas puramente electrónicas. A primeira é um modelo mais conservador, enquanto a segunda apresenta-se com uma atitude mais arrojada, tirando partido do potencial que os meios electrónicos disponibilizam, seguindo processos igualmente rigorosos de revisão e crítica dos conteúdos dos artigos. Efectivamente, o processo de avaliação dos pares enquanto forma de certificação da qualidade científica é um ponto crucial na aceitação do modelo de publicação científica electrónica [Sabbatini, 1999].

Actualmente, já existem várias revistas científicas puramente electrónicas em várias áreas científicas com reconhecida qualidade e prestígio, que têm os contributos dos melhores investigadores e cientistas na área, são por exemplo, a D-Lib³⁵, *Journal of Electronic Publishing* (JEP)³⁶, E-Lis³⁷ e a Ariadne³⁸.

O impacto das novas tecnologias de informação e comunicação estão a revolucionar os modelos formais e informais de comunicação científica. A crescente utilização da Internet tem contribuído de forma significativa para o desenvolvimento do processo de comunicação científica electrónica promovida pelos diversos grupos das áreas do saber.

Na secção seguinte serão apresentadas as principais características analisadas dos artigos das revistas científicas electrónicas, no âmbito deste trabalho de mestrado.

3.2. CARACTERÍSTICAS DOS ARTIGOS DE REVISTAS CIENTÍFICAS ELECTRÓNICAS

O reconhecimento da importância das revistas científicas como canal de comunicação científica e conseqüentemente a transferência desta visão para a revista científica electrónica levanta um conjunto de questões, no que diz respeito à identificação deste recurso e a

³⁵ <http://www.dlib.org/>.

³⁶ <http://www.press.umich.edu/jep/>.

³⁷ <http://eprints.rclis.org/>.

³⁸ <http://www.ariadne.ac.uk/>.

possibilidade de ser igualado à publicação tradicional baseada em papel, a qual é largamente utilizada e que já conquistou a sua credibilidade.

Na generalidade espera-se que as revistas científicas electrónicas sejam o reflexo das tradicionais revistas científicas, desde que as primeiras assegurem os mesmos padrões de qualidade, e garantam o aumento de valor acrescentado, que se traduz na rápida disseminação dos resultados, facilidade de acesso, informação mais actualizada e fácil de localizar, bem como as características que lhes estão associadas, já legitimadas pelas tradicionais revistas científicas, nomeadamente o reconhecimento da credibilidade, fiabilidade e qualidade.

As tradicionais revistas científicas em suporte de papel, para além das limitações de acessibilidade apresentadas na secção anterior, contêm também um conjunto de limitações para expressar conteúdos, nomeadamente a integração de novas formas de apresentação, que compreendem a integração de imagens, vídeos, sons, hiperligações, etc., que não podem ser expressos nas tradicionais revistas científicas [Baptista, 2002].

Neste contexto, os meios electrónicos e digitais em relação aos meios físicos ganham cada vez mais adeptos no que diz respeito à publicação e comunicação científica. Além disso, no meio digital, o artigo científico pode ser hiperligado a outros documentos e submetidos à discussão da respectiva comunidade científica, assegurando deste modo as preocupações associadas à qualidade, fiabilidade e credibilidade do conteúdo que é divulgado [Baptista, 2002].

Os principais canais de disseminação do conhecimento científico produzido pelas várias comunidades científicas, em Portugal são normalmente as revistas científicas, as actas de conferências, as monografias, as teses de mestrado e de doutoramento. No entanto, “as revistas científicas são consideradas, em geral, a forma mais importante de realizar comunicação científica” [Costa 1999, *in* Baptista 2002], pelo que são seleccionadas para objecto no âmbito deste trabalho. De modo a restringir o domínio deste trabalho, considerou-se a selecção de documentos que compõem as revistas científicas da área da informática.

Neste trabalho, procurou-se identificar um conjunto de características normalmente encontradas nos artigos científicos disponibilizados e publicados em revistas científicas electrónicas. Deste modo, procedeu-se numa primeira fase ao levantamento e análise de alguns artigos científicos de várias revistas científicas, da área da informática, nomeadamente a D-Lib, E-Lis e Ariadne. Todas estas revistas apresentam conteúdos lexicais, estilos e formatos distintos.

As plataformas utilizadas na consulta destas revistas foram a b-on e a *ISI Web of Knowledge* (WOK).

Este trabalho de pesquisa, permitiu comparar as diferentes características encontradas nos artigos científicos de variadas revistas científicas, disponíveis em formato digital. As observações efectuadas prenderam-se com a análise estrutural e organizacional dos artigos, o formato com que são apresentados, os metadados utilizados e os vocabulários associados, e não com aspectos relacionados com a definição dos artigos de revistas científicas nem com os seus conteúdos.

Após o levantamento e análise de alguns artigos da área da informática que compõem as diferentes revistas científicas enunciadas acima, procedeu-se à identificação e definição das características que melhor se adequam à descrição dos artigos científicos. Este estudo contribuiu posteriormente para uma melhor selecção dos elementos de metadados que melhor se adequam à descrição dos artigos científicos, e que será apresentado no próximo capítulo.

O resultado obtido da análise efectuada aos vários artigos científicos permitiu identificar as características mais adequadas à descrição dos artigos científicos no âmbito deste trabalho, de acordo com um conjunto de categorias, designadamente: a identificação dos artigos científicos, a identificação de quem tem a posse do artigo científico, relevância do artigo científico, a classificação do artigo científico e por fim a informação sobre as ligações do artigo científico. Estas características foram consideradas como potencialmente influentes na utilização das revistas científicas electrónicas como canal de comunicação, e espera-se que o projecto final implementado ajude a avaliar as reais possibilidades que os artigos científicos disponibilizados em formato electrónico oferecem como veículos de disseminação do conhecimento científico.

a. Identificação dos artigos científicos

As revistas científicas contêm na sua composição um conjunto de artigos científicos “com hiperligações no seu corpo e nas suas referências” [Baptista, 2002]. Nos artigos científicos consultados nas revistas científicas enunciados acima verificou-se que a estrutura organizacional dos artigos científicos entre as revistas, não é muito distinta. É utilizado um conjunto de metadados que são comuns a todas as revistas científicas, designadamente o título do artigo, o autor, a data de publicação e a utilização de um identificador do artigo. Estes elementos de metadados facilitam o processo de pesquisa sobre os artigos científicos. As revistas científicas disponibilizavam vários mecanismos

para efectuar pesquisas sobre os artigos científicos, nomeadamente através da utilização do elemento de metadados autor, ou do elemento de metadados título, ou então através dos metadados, que contêm a data de publicação, permitindo ao utilizador consultar as edições mais recentes de uma determinada revista. Neste contexto, o principal objectivo será definir uma estrutura uniforme de organizar os metadados que permitem identificar os artigos científicos e conseqüentemente facilitar e promover o seu rápido acesso. No âmbito deste trabalho, para além dos elementos de metadados nucleares normalmente utilizados na descrição de publicações científicas, designadamente o autor, o título e a data da publicação do artigo, foram identificados e seleccionados um conjunto variado de elementos de metadados que contribuem para o enriquecimento da descrição semântica dos artigos científicos.

A identificação dos artigos científicos compreende os seguintes atributos: Identificador, Título, Autor, Resumo, Língua, Data de Criação, Data de Publicação, Formato, Dimensão e É-Parte-De.

O significado dos atributos é apresentado de seguida:

- **Identificador** – Este atributo servirá para identificar um determinado artigo científico.
- **Título** – Este atributo irá permitir conter o título do artigo científico.
- **Autor** – Este atributo será utilizado para guardar os dados relativos ao autor ou autores do artigo científico.
- **Resumo** - Este atributo irá conter o resumo do artigo científico.
- **Língua** – Este atributo será utilizado para identificar o idioma do artigo científico que está a ser descrito.
- **Data de Criação** – Serve para indicar a data de criação do artigo científico.
- **Data de Publicação** – Serve para indicar a data de publicação do artigo científico.
- **Formato** - Este atributo é usado para indicar o formato do artigo. Por exemplo o artigo pode estar em formato PDF (*Portable Document Format*), em HTML (*Hypertext Markup Language*), em XML (*eXtensible Markup Language*), etc
- **Dimensão** – Este atributo servirá para indicar o tamanho do artigo.

- **É-Parte-De** – Este atributo permite identificar a revista em que o artigo esta inserido. Este atributo irá conter o URL da revista.

b. Identificação de quem tem a posse do artigo científico a ser descrito

Como foi referido anteriormente, a questão dos direitos de autor é muito “sensível” no seio das comunidades científicas, em particular no que se refere à publicação electrónica.

O recente movimento denominado de Acesso Livre têm intensificado a sua actividade no sentido de estimular os autores a disponibilizar, sempre que possível, o Acesso Livre aos artigos científicos produzidos no seio das comunidades científicas, conservando a propriedade intelectual do artigo científico. Normalmente, no caso das revistas científicas, sempre que é publicado um artigo, os direitos da propriedade intelectual são quase sempre transferidos do autor para a editora da revista. No entanto, graças à pressão do movimento Acesso Livre os autores podem solicitar à editora da revista, o direito de colocar um *postprint*³⁹ do artigo num repositório de artigos científicos de Acesso Livre ou inclusivamente coloca-lo na página pessoal do autor.

Esta questão tem impactos a nível internacional, intensificada com o surgimento do movimento de Acesso Livre, de qualquer forma, para mais informações acerca deste assunto, é sugerida a consulta do Projecto RoMEo (*Rights METadata for Open archiving*)⁴⁰ que incentiva os autores a conservar os seus direitos e a procederem ao auto-arquivo das suas publicações, permitindo-lhes disponibilizar os seus artigos sem estarem a violar as restrições de *copyright* normalmente impostas pelas revistas.

Efectivamente esta é uma das principais questões que distingue a publicação de notícias em jornais e a publicação de artigos científicos, isto é, a entidade que tem a posse do artigo científico é distinta da entidade que normalmente tem a posse de um artigo publicado num jornal. De facto, a posse de um artigo publicado num jornal é do

³⁹ “O texto digital de um artigo que foi avaliado e revisto (*peer-reviewed*) e que foi aceite para publicação por uma revista científica. Isto inclui:

1. O draft digital final do autor revisto e aceite;
2. A versão revista e corrigida do editor, possivelmente em PDF;
3. Qualquer revisão subsequente, com correcções do draft final *peer-reviewed*. ” [LusoDSpace, 2005]

⁴⁰ <http://www.lboro.ac.uk/departments/ls/disresearch/romeo/> .

autor, de acordo com o Código do Direito de Autor e dos Direitos Conexos (CDADC)⁴¹, enquanto que na publicação de um artigo científico numa revista, os direitos sobre a posse do artigo são normalmente transferidos para a editora.

Deste modo, a identificação da entidade que tem a posse do artigo científico a ser descrito compreende os atributos editora e direitos.

O significado destes atributos é apresentado de seguida:

- **Editora** – Este atributo destina-se a identificar a editora da revista científica responsável pela edição, divulgação e preservação dos artigos científicos.
- **Direitos** – Este atributo guarda a informação correspondente aos direitos sobre o artigo científico.

c. Relevância do artigo científico

A relevância do artigo científico é determinada de acordo com as necessidades de informação de uma determinada audiência, que o artigo científico pretende focar. A gestão deste processo é efectuada pela entidade responsável pela preservação do artigo científico.

A relevância do artigo científico a ser descrito compreende o seguinte atributo: Mediador. O significado deste atributo é apresentado de seguida:

- **Mediador** – Este atributo define uma classe ou uma entidade responsável por mediar o acesso do artigo científico ao público-alvo do artigo.

d. Classificação do artigo científico

Normalmente as revistas científicas procedem à classificação ou categorização dos artigos científicos, em geral de acordo com o seu assunto, ou podem eventualmente proceder à criação de um sistema de classificação próprio. A definição de um sistema de classificação permite a utilização de uma linguagem controlada no que se refere à classificação do artigo científico, o que contribui para a interoperabilidade e uniformidade entre classificações e consequentemente facilitar o processo de pesquisa.

⁴¹ <http://www.spautores.pt/page.aspx?contentId=559&idMasterCat=39> e http://www.gda.pt/codigo/lei_03.html .

No âmbito deste trabalho, será utilizado o sistema de classificação da ACM CCS versão 1998, para indicar o assunto do artigo científico.

A classificação do artigo científico a ser descrito compreende o atributo assunto.

O significado deste atributo é apresentado de seguida:

- **Assunto** – Este atributo irá permitir indicar o assunto do artigo científico. No contexto deste trabalho os valores armazenados neste atributo serão retirados do sistema de classificação da ACM.

e. Ligações do artigo científico

Uma das vantagens da publicação científica electrónica está no facto de os artigos científicos conterem hiperligações no seu corpo e nas suas referências facilitando o acesso das mesmas ao utilizador final do conteúdo do artigo. O registo e a indexação das hiperligações e referências bibliográficas dos artigos científicos representam um mecanismo de controlo das citações bibliográficas efectuadas nos artigos científicos e consequentemente a visibilidade das próprias revistas científicas.

As ligações do artigo científico a ser descrito compreende os seguintes atributos: Citação Bibliográfica e Referências.

O significado destes atributos é apresentado de seguida:

- **Citação Bibliográfica** - Este atributo irá armazenar a referência bibliográfica do recurso que está a ser descrito. Normalmente é aconselhada a escrita das referências bibliográficas segundo uma norma. Assim sugere-se a norma NP 405 - 4. 2003, Informação e Documentação - Referências Bibliográficas.
- **Referências** – Este atributo irá guardar as referências bibliográficas utilizadas pelo autor na elaboração do conteúdo do artigo científico.

3.3. RESUMO

A evolução da comunicação electrónica associada às necessidades de conhecimento por parte dos consumidores de informação tem contribuído para a reestruturação dos meios tradicionalmente utilizados quer na publicação de artigos científicos quer também na publicação

de notícias. Os jornais e as revistas científicas deixam de publicar exclusivamente em documentos impressos, passando também a disponibilizar os seus conteúdos na *Web*, tirando partido das funcionalidades que meio o electrónico oferece.

Contudo, verifica-se que grande parte das revistas científicas electrónicas apenas procede à reprodução electrónica da versão já impressa mantendo-se ainda fortemente vinculada aos tradicionais processos de publicação e comunicação científica. A crescente utilização da *Web* como principal fonte de informação por parte dos cientistas e investigadores tem contribuído para o aparecimento de revistas científicas puramente electrónicas com reconhecida qualidade e prestígio. A consulta destas revistas permitiu identificar um conjunto de propriedades que caracterizam os artigos científicos publicados nessas revistas e deste modo auxiliar o processo de descrição dos seus conteúdos científicos.

4. RSS

Hoje em dia, a Internet é uma importante fonte de informação. Tem-se tornado num instrumento de aplicação constante por parte dos investigadores e cientistas, no desenvolvimento diário do seu trabalho.

No entanto verifica-se um crescimento descontrolado e desordenado da Internet. Por um lado assiste-se a um crescimento incontrolável do número de páginas *Web* e ao desenvolvimento de sofisticados motores de pesquisa, com o objectivo de facilitar aos consumidores de informação o processo de pesquisa. Por outro lado, assiste-se ao emergir de um conjunto diversificado de tecnologias, que contribuem para a implementação de mecanismos inteligentes de pesquisa e de navegação, desenvolvidos e disponibilizados aos consumidores de informação, com objectivo de facilitar o rápido acesso à mesma disponível na *Web*.

A Internet tornou-se no principal recurso no processo de pesquisa e de acesso à informação mais recente e actualizada, relativamente a um determinado tópico. Mas, tendo em conta o volume de informação disponível na *Web*, o processo de pesquisa e localização de conteúdos específicos e consequentemente a identificação de eventuais actualizações ou alterações que possam ocorrer dentro de um conjunto diversificado de temas, torna-se difícil e complexo para o utilizador, principalmente quando existem inúmeras fontes de informação.

A generalidade dos utilizadores gere este processo adicionando aos seus favoritos um conjunto de *links* de páginas *Web*, seleccionados de acordo com os seus interesses pessoais, procedendo posteriormente à regular consulta dessas mesmas páginas, para verificar se estas sofreram alterações. No entanto, cada vez que se visita uma determinada página *Web* à procura de possíveis actualizações é necessário ter presente o conteúdo anterior da mesma, para ter a efectiva percepção de todas as alterações desenvolvidas nessa página *Web*.

Neste contexto, verifica-se o emergir de sofisticados serviços que notificam os utilizadores acerca de novos conteúdos ou actualizações mais recentes, que vão surgindo nas páginas *Web* previamente seleccionadas, facilitando ao utilizador a sua leitura sem que este tenha que aceder directamente à página *Web* para verificar se foi adicionada nova informação ou se um determinado conteúdo sofreu alguma alteração. Estes serviços permitem ao utilizador assegurar uma gestão mais eficiente do seu tempo.

O *RDF Site Summary* (RSS) ou *Rich Site Summary* (RSS) [Begeed *et al.*, 2000] ou ainda *Really Simple Syndication* (RSS) [HarvardLaw, 2006] surge como uma solução tecnológica, que permite aos utilizadores tomar conhecimento acerca de novos conteúdos provenientes de uma determinada fonte de informação, sem terem de aceder directamente à respectiva página *Web*. O RSS é um formato normalizado para agregação e distribuição de conteúdos da *Web* facilitando o processo de consulta e partilha de informação proveniente de diversas fontes de informação, que periodicamente estão sujeitas a alterações ou actualizações [Pilgrim, 2002].

A crescente utilização da Internet tem intensificado o uso da tecnologia RSS, na medida em que esta vem inovar os actuais mecanismos de consulta e de acesso à informação mais actual, disponibilizando um conjunto de serviços de alerta, para novos conteúdos que são disponibilizados nas páginas *Web* ou então através do envio de notificações ao utilizador via e-mail sobre novos conteúdos.

Esta tecnologia começa a desafiar a ortodoxia das tradicionais páginas *Web* reformando e redefinindo os princípios que foram definidos e mantidos nos últimos 10 anos sobre a *Web*. [Berners-Lee, 1990]. De facto a concepção original de *Tim Berners-Lee* sobre a *Web* estabelecia a partilha estruturada de informação, ao invés de um caleidoscópio de leitura, como acabou por suceder [Hammond *et al.*, 2004].

4.1. O QUE É O RSS?

Tendo sido originalmente desenhado para permitir a distribuição e divulgação de notícias agrupadas de um conjunto diversificado de fontes de informação, o RSS acabou por ser utilizado não só no contexto das notícias de jornais mas também na disponibilização de qualquer tipo de informação que normalmente está sujeita a frequentes alterações.

Actualmente a tecnologia RSS tem sido amplamente utilizada no seio da comunidade dos *blogs*, uma vez que proporciona a partilha e acompanhamento das últimas novidades, ou textos completos assim como a distribuição de ficheiros multimédia, através do método *Podcasting*⁴². No ano 2000, a utilização do RSS difundiu-se para grandes empresas de notícias

⁴² *Podcasting* é uma forma de publicação de programas de áudio, vídeo e/ou fotografias pela Internet, permitindo aos utilizadores acompanhar a sua actualização. Os programas ou arquivos, gravados em qualquer formato digital (MP3, AAC e OGG são os mais utilizados nos *podcasts* de áudio), ficam armazenados num servidor na Internet. Através de um *feed* RSS, que funciona como um índice actualizável dos arquivos

como a *Reuters*, CNN e a BBC [Wikipédia, 2006d]. Estas empresas permitiam que outras páginas *Web* incorporassem as suas notícias e resumos, através de vários acordos de utilização, com o objectivo de abranger um maior número de leitores. Actualmente o RSS é utilizado em diversas áreas, nomeadamente no marketing, *bug-reports*, previsão do tempo, informações sobre o trânsito, informações da área económica, lista de empregos disponíveis e qualquer outra actividade que envolva actualização dinâmica de conteúdos.

Por trás do conceito existe a tecnologia que o implementa. O RSS é um formato baseado em XML normalizado criado para agrupar conteúdos. Este processo é denominado por “sindicância de conteúdos da *Web*”⁴³ [Hammond, 2003].

A sindicância de conteúdos da *Web*, consiste no termo técnico utilizado para a troca regular de informação actualizada entre diferentes páginas *Web* [Wittenbrink, 2005]. A especificação do RSS define sindicância como o processo de “disponibilizar dados on-line de modo a permitir a sua recuperação, transmissão, agregação ou publicação on-line” [Begeg *et al.*, 2000].

Os jornais, revistas e as tradicionais formas de publicação têm progressivamente disponibilizado os seus conteúdos na *Web* [Powers, 2003], e conseqüentemente assiste-se a um crescimento exponencial de informação digital. Com o intuito de auxiliar os utilizadores no processo de acesso à informação, grande parte dos fornecedores de informação pública documentos RSS denominados como *feeds* ou *Web feeds*.

O termo *feed* vem do verbo em inglês “alimentar”. Na Internet, os “*feeds*” constituem listas actualizadas de conteúdos sobre uma determinada página *Web* [Wikipedia, 2006b].

Um *feed* RSS está organizado por um conjunto de *itens* onde cada *item* contém informação dos conteúdos a serem publicados. Efectivamente um *feed* RSS não só faz referência a um recurso, como contém a informação original [Wittenbrink, 2005].

A identificação das páginas *Web* que disponibilizam os *feeds* RSS é efectuada através da utilização de um ícone (geralmente na cor laranja) com os acrónimos “RSS” ou “XML”. Assim que o fornecedor de informação disponibiliza o *feed* RSS na respectiva página *Web*, os

disponíveis, novos programas de áudio, vídeo ou fotos são automaticamente reunidos para permitir ao leitor através de um agregador, identificar os novos arquivos e proceder automaticamente à sua agregação na máquina [Wikipédia, 2006c].

⁴³ do inglês *Web Syndication*.

utilizadores subscrevem os *feeds* e procedem à sua leitura através da utilização de programas específicos denominados por agregadores RSS⁴⁴ ou leitores RSS⁴⁵. Estes programas agrupam e apresentam os *feeds* RSS, disponibilizados pelas diversas fontes de informação, permitindo a distribuição dos seus conteúdos facilitando a sua rápida consulta e análise e contribuindo para a partilha de informação.

Actualmente já existe um conjunto diversificado de leitores RSS que permitem a subscrição de *feeds* RSS. De facto, já existe uma lista destes programas, que permitem a leitura de *feeds* RSS. No entanto o sítio *RSSfeeds.com* disponibiliza uma lista actualizada por ordem alfabética, com todos os leitores RSS gratuitos bem como os comerciais.

De seguida são apresentados alguns exemplos:

- Aplicações para o ambiente de trabalho, como por exemplo:
 - *Feedreader* - <http://www.feedreader.com/>
 - *Amphedadesk* - <http://www.disobey.com/amphetadesk/>
 - *FeedDemon* – <http://www.bradsoft.com/feeddemon/>
 - *RSS Bandit* - <http://www.rssbandit.org/>
 - *NetNewsWire*, um agregador RSS para ser utilizado em MAC ou Macintosh – <http://ranchero.com/netnewswire/>
- Aplicações baseados na *Web*, como por exemplo:
 - *Bloglines* - <http://www.bloglines.com/>
- *Plug-ins* para *Web browsers* e clientes de e-mail, como por exemplo:
 - *Newsgator* permite a sua utilização no Microsoft Outlook - <http://www.newsgator.com/home.aspx>
 - Mozilla Firefox disponibiliza uma barra com o leitor RSS
 - *Browsers* com leitores RSS embutidos, como por exemplo o Opera 7.50 e Safari.

O grupo de trabalho da *Nature Publishing Group* (NPG)⁴⁶ esteve também envolvido no desenvolvimento de um agregador RSS denominado *Urchin*⁴⁷. Esta aplicação foi desenvolvida

⁴⁴ do inglês *RSS aggregators*.

⁴⁵ do inglês *RSS reader*.

⁴⁶ <http://www.nature.com/index.html>

⁴⁷ <http://urchin.sourceforge.net/>.

no âmbito do projecto ROSA⁴⁸ e consiste numa aplicação de código fonte aberto desenhada para agregar e filtrar *feeds* RSS e outras fontes de dados. O *Urchin* foi inicialmente financiado pelo *Joint Information Systems Committee* (JISC) no Reino Unido e implementado pelo grupo de trabalho denominado *Publishers and Library/Learning Solutions (PALS) Metadata and Interoperability Group*⁴⁹ [Hammond *et al.*, 2004].

O funcionamento básico da aplicação *Urchin* consiste em agrupar informação proveniente de um conjunto diversificado de fontes de dados (incluindo todas as versões do RSS, páginas HTML e base de dados) e internamente proceder ao seu armazenamento. Assim que é efectuado um pedido, a informação é filtrada e emitida no formato seleccionado. A NPG utiliza a aplicação *Urchin* para disponibilizar aos seus colaboradores palavras-chave filtradas dos *feeds* RSS de modo a alimentar um portal de notícias relacionadas com ciência e tecnologia. Ou seja, esta aplicação permite seleccionar artigos, de *feeds* RSS em função de uma palavra-chave específica [Hammond *et al.*, 2004].

4.2. EVOLUÇÃO DO RSS

Antes de surgir o RSS, existiram outros formatos semelhantes que permitiam efectuar a agregação de conteúdos, nomeadamente o formato *Meta Content Framework* (MCF) desenvolvido pela *Apple Computer* no âmbito do projecto experimental *HotSauce*, cuja origem remonta ao ano de 1995 [Hammond, 2003].

Em 1997 a Microsoft junto com a *Pointcast* e outras empresas criou o formato *Channel Definition Format* (CDF) baseado em XML para a descrição de páginas *Web*. Este formato permitia a descrição de conteúdos, a publicação de planos e a utilização de metadados no processo de descrição de páginas *Web* [Wittenbrink, 2005]. Foi incorporado no Internet Explorer 4.0 para operar como suporte tecnológico no recurso denominado pela Microsoft como *Active Channel*. Este formato é mais tarde adaptado seguindo um perfil baseado em RDF, para ser utilizado no serviço “*My Netscape Network*” disponibilizado pelo portal da *Netscape* [Wittenbrink, 2005]. Uma revisão deste perfil RDF resultou no aparecimento de uma versão *draft* do RSS 0.90 [Hammond, 2003]. Em Março de 1999 surge o *Rich Site Summary* (RSS) 0.90, a primeira versão oficial do RSS criada por *Dan Libby*. No entanto foi considerado um formato demasiado complexo para os objectivos que se propunha alcançar [Pilgrim, 2002]. De seguida é proposta

⁴⁸ http://www.jisc.ac.uk/whatwedo/programmes/programme_pals/project_rosa.aspx.

⁴⁹ http://www.jisc.ac.uk/index.cfm?name=programme_pals.

uma versão simplificada, 0.91, mas é rapidamente abandonada uma vez que não estava de acordo com os planos de negócio da *Netscape*.

A versão RSS 0.91 baseada em XML passa a ser propriedade da *UserLand Software* sob a direcção do CEO *Dave Winer* que utilizou esta tecnologia como suporte base para o desenvolvimento dos seus produtos de software associados ao *Weblogging* [Hammond, 2003].

Entretanto surge um terceiro grupo não comercial, denominado *RSS-DEV Working Group* (<http://groups.yahoo.com/group/rss-dev/>) que dividiu e projectou um novo formato baseado nos mesmos princípios que deram origem ao RSS versão 0.90 (antes de este ser simplificado para a versão 0.91). Em Dezembro do ano 2000 surge o formato RSS 1.0 baseado na tecnologia RDF, tirando partido da extensibilidade do RDF e da utilização dos *namespaces* que asseguravam a não colisão entre elementos [Begeed *et al.*, 2000]. A *UserLand Software* como não estava incluída neste grupo de trabalho não ficou satisfeita quando foi anunciada a versão RSS 1.0 e ao invés de aceitar esta nova versão procedeu à simplificação da versão RSS 0.90 dando origem às versões RSS 0.92, RSS 0.93, RSS 0.94 e em 2002 é lançada a actual versão RSS 2.0 [Pilgrim, 2002]. A 15 de Julho de 2003 a *UserLand Software* transferiu a propriedade da especificação do RSS 2.0 para o *Berkman Center for Internet & Society at Harvard Law School* [Hammond, 2003].

É neste contexto que são justificados os diferentes significados do acrónimo RSS. Ou seja "*Rich Site Summary*", ou "*RDF Site Summary*" está associado às versões RSS 0.90 e RSS 1.0 enquanto "*Really Simple Syndication*" faz referência às versões RSS 0.91, RSS 0.92, RSS 0.93, RSS 0.94 e RSS 2.0. A diferença entre estas versões assenta no facto de as versões RSS 0.90 e RSS 1.0 serem ambas uma aplicação XML, em conformidade com a especificação RDF do W3C e extensível via *XML-namespace* ou através da modularização baseada em RDF, enquanto as versões RSS 0.91, RSS 0.92, RSS 0.93, RSS 0.94 e RSS 2.0 seguem a especificação XML do W3C. Esta divisão resulta numa família de especificações que estão ligeiramente relacionadas mas que foram desenvolvidas por diferentes grupos de trabalho [Hammond, 2003]. Com efeito, as duas variantes do RSS continuam a evoluir paralelamente permanecendo o mesmo envolvimento por parte dos diferentes grupos de trabalho. Além disso, verifica-se uma maior utilização da versão do RSS 1.0 seguido da versão RSS 2.0, na criação de *feeds*, como é publicado pelo directório rssfeeds.com.

Na figura 13 é ilustrado a proporção do crescimento total de *feeds* RSS desde Setembro de 2001 até Maio de 2006 publicado pela Syndic8⁵⁰, um importante directório RSS na Internet.

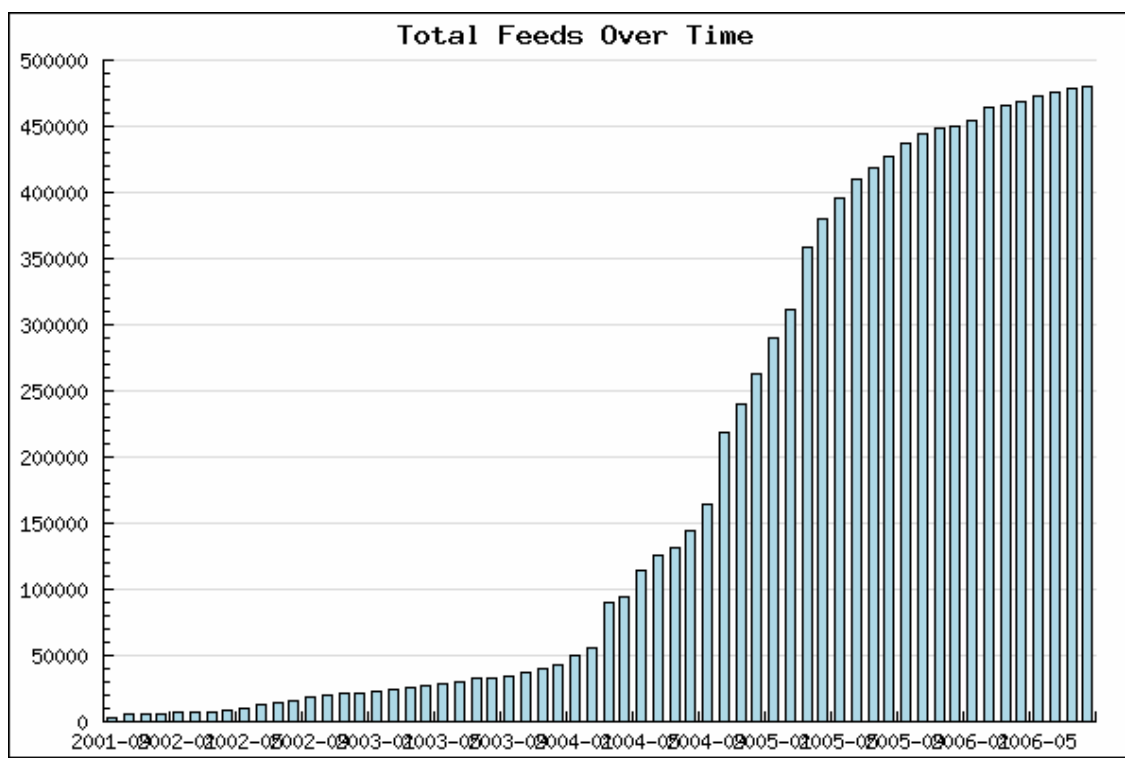


Figura 10: Proporção do crescimento da utilização do RSS entre o ano 2001-2006

(Retirado da página estatística do Syndic8 disponibilizada no sítio: <http://www.syndic8.com/stats.php#Actions> acedido em Junho de 2006)

De seguida é apresentada na tabela 5 informação sobre os formatos de agregação de conteúdos mais importantes que surgiram até aos nossos dias.

Nome	Data de Publicação	Autor	URI da especificação
MCF (<i>Metadata Content Format</i>)	1995	Ramanathan V. Guha/ Apple Computer	http://www.xspace.net/hotsauce/mcf.html
CDF (<i>Channel</i>)	9 de Março de 1997	Castedo Ellerman/	http://msdn.microsoft.com/workshop/delivery/cdf/reference/CDF.asp

⁵⁰ Syndic8 representa a maior directoria de *channels* RSS e permite a pesquisa dos mesmos em áreas específicas.

<i>Definition Format)</i>		Microsoft	
<i>Meta Content Format</i> Através da utilização do XML	6 de Junho de 1997	R. V. Guha/ Netscape Tim Bray/ Textuality	http://www.w3.org/TR/NOTE-MCF-XML http://www.textuality.com/1997/12/15/scriptingNewsInXML
<i>Scripting News</i>	27 de Dezembro de 1998	Dave Winer/ UserLand	http://davenet.scripting.com/1997/12/15/scriptingNewsInXML
RSS 0.90	15 Março 1999	Netscape	http://www.purplepages.ie/RSS/netscape/rss0.90.html
RSS 0.91	10 Jul 1999	Dan Libby/Netscape	http://my.netscape.com/publish/formats/rss-spec-0.91.html
RSS 0.91 (versão UserLand)	6 Apr 2000	Dan Libby/ Netscape Dave Winer/ UserLand	http://backend.userland.com/rss091
RSS 1.0	14 Aug 2000	Rael Dornfest/ O'Reilly <i>et al.</i>	http://Web.resource.org/rss/1.0/
OPML (<i>Outline Processor Markup Language</i>)	15 Sep 2000	Dave Winer/ UserLand	http://www.opml.org/spec
RSS 0.92	25 Dec 2000	Dave Winer/ UserLand	http://backend.userland.com/rss092
RSS 0.93 (draft)	20 Apr 2001	Dave Winer/ UserLand	http://backend.userland.com/rss093
RSS 2.0	18 Sep 2002	Dave Winer	http://blogs.law.harvard.edu/tech/rss
RSS 1.1 (draft)	23 Jan 2005	Sean B. Palmer,	http://inamidst.com/rss1.1/

		Christopher Schmidt	
Atom 0.4 (<i>draft</i>)	18 Apr 2005	Mark Nottingham, Richard Sayre <i>et al.</i>	http://www.ietf.org/internet-drafts/draft-ietf-atompub-format-08.txt
Atom 1.1 (<i>spec.</i>)	15 Aug 2005	Mark Nottingham, Richard Sayre <i>et al.</i>	http://www.ietf.org/internet-drafts/draft-ietf-atompubformat-11.txt

Tabela 4: Lista dos principais formatos de agregação de conteúdos, adaptado de [Wittenbrink, 2005]

4.3. ESTRUTURA DE UM *FEED* RSS

Um *feed* RSS descreve um recurso identificado por um URI. Os dados introduzidos num *feed* compreendem a descrição dos conteúdos dinâmicos⁵¹ do recurso, através dos elementos básicos do RSS designadamente o *title*, *link* e opcionalmente o elemento *description*.

Um *feed* RSS expressa a informação mais recente relativamente a um determinado recurso. Está associado à actualização da informação, independentemente da estrutura interna dos dados, assim como os tópicos a que se refere. É tão universal, que é sempre possível criar um *feed* RSS de qualquer tipo de informação [Wittenbrink, 2005]. Um *feed* pode referir-se a um *wiki* como também a um *Weblog*, um portal de informação ou uma compilação de actualizações de software. Qualquer colecção de informação que sofre alterações temporais quer por períodos curtos ou longos é candidata a um *feed* RSS.

A estrutura de um *feed* RSS está organizada por uma hierarquia de dois níveis de informação, designadamente, colecções de *itens* de informação e *itens* individuais de informação [Wittenbrink, 2005]. As colecções correspondem ao elemento denominado “*channel*”⁵² enquanto os itens individuais de informação, dentro da colecção, correspondem ao elemento designado

⁵¹ Neste contexto, entende-se por conteúdos dinâmicos toda a informação que é periodicamente actualizada ou esta sujeita a alterações.

⁵² Por uma questão de coerência serão mantidos os nomes dos elementos XML que compõem o formato de um documento *feed*. Ou seja “*channel*”, “*item*” e “*items*”.

“*item*”. Ou seja, um *channel* consiste na descrição de um recurso, que pode ser constituído por um ou vários *items*, enquanto um *item* consiste num objecto individual de informação que compõe o recurso que está a ser descrito. Os elementos RSS que estão presentes na descrição da informação destes dois níveis são os elementos *title*, *link* e *description*, podendo opcionalmente ser adicionada metainformação no processo de descrição do recurso.

Todos os formatos RSS têm um modelo básico em comum [Çelikbas, 2005]: todas as versões do RSS são baseadas em XML e a sua estrutura geral é muito semelhante. Independentemente da versão, um documento RSS segue as seguintes linhas de desenvolvimento:

1. Um documento RSS é baseado em XML, então este deve ser um documento *well-formed*.
2. O primeiro elemento de um documento RSS é o elemento <channel>. Este elemento contém metadados que descrevem o próprio canal, designadamente um título, uma breve descrição e um URL do recurso descrito. Este URL deve ser único. Normalmente o URL definido pode ser o da página *Web* que está a ser descrita, ou então, o URL onde o *feed* RSS é disponibilizado [Powers, 2003].
3. O elemento <title> contém a informação acerca do título do recurso descrito. Se este elemento está a ser utilizado dentro de um elemento <item>, então o elemento <title> refere-se ao título de um conteúdo específico.
4. O elemento <link> indica o URL da página *Web* que corresponde ao *feed* RSS. No caso de este elemento estar a ser utilizado dentro do elemento <item>, então o elemento <link> refere-se ao URL de um conteúdo específico.
5. O elemento <description> descreve o *feed* RSS ou um determinado item.
6. O elemento <item> especifica cada artigo ou conteúdo dentro do documento RSS. Os sub-elementos necessários para este elemento são: <title>, <description> e <link>, podendo ser adicionadas opcionalmente mais metainformação. As especificações das duas variantes do RSS exigem que exista pelo menos um item.

Independentemente da versão RSS que o utilizador pretenda utilizar, estas são algumas considerações a serem seguidas na criação de um documento RSS. No entanto existem diferenças estruturais entre a versão RSS 1.0 e a versão RSS 2.0, assim como nas suas versões

descendentes, como é demonstrado pelas diferentes especificações definidas para cada uma destas versões.

Efectivamente, as especificações diferem não só na filosofia mas também na implementação [Ayers, 2003]. A criação manual de conteúdos no formato RSS 2.0 é extremamente fácil para qualquer utilizador. Enquanto o RSS 1.0 não é tão fácil, quando comparado com o RSS 2.0, justificado pelo facto de utilizar a tecnologia RDF na sua codificação. No entanto, o que torna o RSS 1.0 notavelmente interessante é o facto de permitir a interoperabilidade com outras linguagens RDF/XML, facilitar a sua leitura e o seu processamento por outras máquinas e permitir a extensibilidade com outros vocabulários, promovendo a descrição semanticamente rica de recursos *Web*. Por outro lado, o RSS 2.0 é caracterizado por Danny Ayers como sendo vazio em termos semânticos [Ayers, 2003].

De modo a ilustrar a estrutura de um documento RSS é apresentada de seguida um exemplo de um *feed* RSS, implementado no formato RSS 1.0, aplicado à descrição de conteúdos provenientes do repositório da APSI. Seguirá no apêndice 1 o *template* do *feed* RSS definido no âmbito desta dissertação de mestrado.

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns="http://purl.org/rss/1.0/"
  xmlns:dcterms="http://purl.org/dc/terms">
  <channel rdf:about="http://repositorio.apsi.pt:8080/index.jsp">
    <title>Repositório da APSI </title>
    <link>http://repositorio.apsi.pt:8080/index.jsp</link>
    <description>Repositório institucional da Associação Portuguesa de Sistemas de Informação</description>
    <dc:publisher>Associação Portuguesa de Sistemas de Informação [APSI] </dc:publisher>
    <dc:creator>Rui Dinis de Sousa </dc:creator>
    <dc:rights>APSI. Todos os direitos reservados</dc:rights>
    <dc:date>2006-05-23T09:47:57Z</dc:date>
    <!-- Um item é um recurso (um artigo) -->
    <items>
      <rdf:Seq>
        <rdf:li rdf:resource="http://hdl.handle.net/2287/20"/>
        <rdf:li rdf:resource="http://hdl.handle.net/2287/56"/>
      </rdf:Seq>
    </items>
  </channel>
  <item rdf:about="http://hdl.handle.net/2287/20">
    <title>Reconhecimento de Voz - Voice Car System (VCS)</title>
    <link>http://hdl.handle.net/2287/20</link>
```

<description>Os avanços tecnológicos possibilitam que os computadores reconheçam a voz humana e a interpretem, de forma a executarem determinadas tarefas previamente definidas. De modo a demonstrar a aplicabilidade destas novas tecnologias desenvolvemos uma aplicação que tem por objectivo a criação de uma interface de reconhecimento de voz, recorrendo para isso ao módulo Sensory Voice Extreme™ Toolkit. A aplicação em causa simula o controlo de algumas funções de um automóvel, activadas através do reconhecimento da voz humana. Numa utilização a nível real, a interacção do condutor perante a placa de reconhecimento da voz deverá ser efectuada através de um módulo que poderá ser integrado, por exemplo, no computador de bordo do automóvel.

```

</description>
<!--PUBLISHER-->
<dc:publisher>APSI</dc:publisher>
<!--AUTOR-->
<dc:creator>
  <rdf:Seq>
    <rdf:li>HugoNeiva</rdf:li>
    <rdf:li>Paulino</rdf:li>
    <rdf:li>Bruno Silva</rdf:li>
    <rdf:li>Pedro Silva</rdf:li>
  </rdf:Seq>
</dc:creator>
<!--RIGHTS -->
<dc:rights>APSI. Todos os direitos reservados</dc:rights>
<!--SUBJECT -->
<dc:subject rdf:datatype="http://www.acm.org/class/1998/acmccs98-1.2.3.xml"> General Literature
  </dc:subject>
<!--FORMATO -->
<dcterms:extent>868596</dcterms:extent>
<dcterms:medium>
  <dcterms:IMT>
    <rdf:value>pdf</rdf:value>
  </dcterms:IMT>
</dcterms:medium>
<!--IDIOMA seguindo a recomendação de codificação ISO639-1 -->
<dc:language>pt</dc:language>
<!--DATA DE CRIAÇÃO DO ARTIGO -->
<dcterms:created>
  <dcterms:W3CDTF>
    <rdf:value> 2004-11-03</rdf:value>
  </dcterms:W3CDTF>
</dcterms:created>
<!--DATA DE PUBLICAÇÃO DO ARTIGO -->
<dcterms:issued>
  <dcterms:W3CDTF>
    <rdf:value>2005-05-26T20:09:02Z </rdf:value>
  </dcterms:W3CDTF>
</dcterms:issued>
</item>
<!--DESCRIÇÃO DO 2º ITEM -->
<item rdf:about="http://hdl.handle.net/2287/56">
  <title>O novo ciclo de desenvolvimento de sistemas de informação – dos processos de negócio à operação em poucas horas</title>
  <link>http://hdl.handle.net/2287/56</link>
  <!-- Deixo de ter dc:abstract e passo a ter description-->

```

`<description>`O processo de desenvolvimento tradicional de Sistemas de Informação tem enormes problemas derivados do tempo que demora tanto a construir a primeira versão do sistema de informação (pelo menos vários meses mas para grandes projectos normalmente mais de um ano) como também, e principalmente, a alterar ou acrescentar requisitos. Estes problemas são cada vez mais graves porque estas alterações são cada vez mais frequentes e além disso os Sistemas de Informação têm ciclos de vida cada vez mais curtos. Neste artigo propomos a utilização de ferramentas de concepção e desenvolvimento do tipo RAD e que tiveram uma grande evolução nos últimos anos. Estas ferramentas de última geração – como aquelas produzidas pela OutSystems em Portugal – estão agora preparadas para suportar um novo ciclo completo de desenvolvimento de sistemas de informação que permite cobrir todas as etapas tradicionais em poucas horas. Esta tecnologia foi validada com um caso de estudo (Portfolios da LEIC) e essa experiência é relatada neste artigo.`</description>`

```

<!--PUBLISHER-->
<dc:publisher>APSI</dc:publisher>
<!--AUTOR-->
<dc:creator>
  <rdf:Seq>
    <rdf:li>Feliciano, Carlos</rdf:li>
    <rdf:li>Silva, Miguel Mira</rdf:li>
  </rdf:Seq>
</dc:creator>
<!--RIGHTS -->
<dc:rights>APSI. Todos os direitos reservados</dc:rights>
<!--SUBJECT -->
<dc:subject rdf:datatype="http://www.acm.org/class/1998/acmccs98-1.2.3.xml"> General Literature
  </dc:subject>
<!--FORMATO -->
<dcterms:extent>200599</dcterms:extent>
<dcterms:medium>
  <dcterms:IMT>
    <rdf:value>pdf</rdf:value>
  </dcterms:IMT>
</dcterms:medium>
<!--IDIOMA -->
<dc:language>pt</dc:language>
<!--DATA DE CRIAÇÃO DO ARTIGO -->
<dcterms:created>
  <dcterms:W3CDTF>
    <rdf:value> 2004-11-03T10:06:02Z</rdf:value>
  </dcterms:W3CDTF>
</dcterms:created>
<!--DATA DE PUBLICAÇÃO DO ARTIGO -->
<dcterms:issued>
  <dcterms:W3CDTF>
    <rdf:value>2005-06-20T10:06:02Z </rdf:value>
  </dcterms:W3CDTF>
</dcterms:issued>
</item>
</rdf:RDF>

```

Figura 11: Exemplo de um *feed* RSS

4.4. A PESQUISA DE FEEDS RSS

Hoje em dia, já existem muitas páginas *Web* que disponibilizam aos seus utilizadores *feeds* RSS. Para identificar um *feed* RSS numa determinada página *Web* basta procurar por um ícone (normalmente na cor laranja). Geralmente os *feeds* RSS estão conectados a esse mesmo ícone.

Também, já existem diversos directórios que auxiliam o utilizador a pesquisar os *feeds* RSS disponíveis assim como os conteúdos que cada um lhes pode oferecer.

De seguida serão apresentados alguns importantes directórios de *feeds* RSS:

- **RSSfeeds.com** (<<http://www.rssfeeds.com>>) – publica uma lista de *feeds* RSS organizada por diferentes categorias.
- **Syndic8** (<<http://www.syndic8.com>>) – representa a maior categoria de *channels* RSS. Apresenta uma lista de aproximadamente 480,112 *feeds* (fonte: <http://www.syndic8.com/stats.php?section=overview>). O Syndic8 disponibiliza mecanismos de pesquisa e navegação de *channels* organizados por áreas específicas.
- **NewsIsFree** (<<http://www.newsisfree.com>>) – é o directório de *feeds* RSS mais antigo e estabelecido. Os utilizadores podem navegar por assunto, pesquisar a base de dados por nome ou descrição, ou então pesquisar os últimos títulos das páginas *Web* que estão indexadas. O NewsIsFree disponibiliza uma secção denominada “*Latest Channels*” que disponibiliza o acesso aos últimos *feeds* que foram adicionados à base de dados. O NewsIsFree pode também ser utilizado como um agregador para a subscrição de listas de *feeds* na sua página *Web*.
- **LISFeeds.com** (<<http://www.lisfeeds.com>>) – não é apenas uma directoria mas também um agregador RSS. Reúne títulos provenientes de diferentes páginas *Web* e serviços especializados em notícias orientadas às bibliotecas e a sua apresentação é mais orientada para investigadores bibliotecários.

4.5. COMO UTILIZAR O RSS

Para utilizar RSS não são necessários conhecimentos na área da programação, basta apenas seleccionar um leitor RSS, subscrever os *feeds* RSS e por fim ler e apreciar.

Como já foi dito, existe uma variedade de leitores RSS gratuitos ou comerciais, basta seleccionar um, descarrega-lo e instalar. A função dos leitores RSS consistem em permitir ao utilizador subscrever uma variedade de conteúdos de um número ilimitado de fontes de informação, num só local. Estes programas irão coleccionar, alterar e apresentar os *feeds* RSS num local central, de onde o utilizador irá aceder.

A generalidade dos leitores RSS é portadora de uma colecção de *feeds* RSS que normalmente está organizado por Notícias, Saúde, Finanças, etc. Ou seja, quando o utilizador instala um leitor RSS, este já contém um conjunto de *feeds* RSS. O utilizador pode manter apenas os *feeds* do seu interesse e eliminar os restantes e/ou subscrever novos *feeds*.

De seguida, o utilizador dirige-se à página *Web* do seu interesse e caso essa página *Web* disponibilize *feeds* RSS, o utilizador deve procurar o botão que permite subscrever o *feed*.

A generalidade dos leitores RSS permite a subscrição de *feeds* RSS de várias formas:

- Subscrever através de um click – quando o utilizador selecciona um determinado *feed*, simplesmente carrega no botão ou no URL que permite a subscrição e segue o *workflow* do leitor RSS para completar a subscrição;
- Subscrever efectuando “*drag and drop*” no leitor RSS;
- Adicionar *feeds* manualmente no caso de o utilizador conhecer o URL do *feed* que pretende subscrever.

Depois de o utilizador subscrever o *feed* RSS o leitor RSS vai recolher informação e apresenta-la através dos elementos básicos de RSS que guardam o título, a identificação do URL de acesso e opcionalmente uma descrição do conteúdo recolhido pelo leitor RSS. Deste modo, o utilizador pode rapidamente analisar as novidades ou as últimas informações que foram publicadas nas páginas *Web* que foram subscritas. Os conteúdos que sejam do seu interesse, basta o utilizador efectuar um click sobre o título, para que este seja encaminhado para a origem do artigo e ler o seu conteúdo.

De seguida serão apresentados dois exemplos de leitores RSS. O primeiro leitor RSS apresentado é o Wizz RSS 2.1.4⁵³ que é uma extensão do browser Mozilla Firefox e o segundo leitor RSS o AmphedaDesk é um exemplo de uma aplicação para ser utilizada no *desktop*.

Leitor RSS

Channels
subscritos

Artigos não consultados

Artigo já consultado

Descrição do artigo seleccionado

Títulos dos artigos

A Foundation for Automatic Digital Preservation
Miguel Ferreira, Ana Alice Baptista and José Carlos Ramalho propose a Service-Oriented Architecture to help cultural heritage institutions to accomplish automatic digital preservation.

Introduction
Efforts to archive a large amount of digital material are being developed by many cultural heritage institutions. We have evidence of this in the numerous initiatives aiming to harvest the Web [1-5] together with the impressive burgeoning of institutional repositories [6]. However, getting the material inside the archive is just the beginning for any initiative concerned with the long-term preservation of digital materials.
Digital preservation can best be described as the activity or set of activities that enable digital information to be intelligible for long periods of time. In general, digital information kept in an archival environment is expected to be readable and interpretable for periods of time much longer than the expected lifetime of the individual hardware and software components that comprise the repository system, as well as the formats in which the items of information are encoded [7][8].
Over the past decade, a vast number of preservation strategies have emerged from the various preservation projects developed, literally, all over the world [9-12]. Nonetheless, the most cited and applied preservation strategy continues to be migration [13][14], especially in contexts where non-interactive digital objects, such as images, databases or text documents, are the focus of preservation.
Migration can be described as a (...) set of organized tasks designed to achieve the periodic transfer of digital materials from one hardware/software configuration to another or from one generation of computer technology to a subsequent generation. [15]

Figura 12: Print screen do leitor RSS Wizz RSS 2.1.4

⁵³ <http://www.wizzcomputers.com/Welcome.php>

Go to channel homepage.
 Email the channel owner.
 XML View the raw XML source.
 Remove channel.

Channels Home / My Channels / Add a Channel / My Settings [Give Thanks](#) / [Docs & Website](#) / [Email](#)

Below are the newest items from your 6 channels, ordered by channel modification date.
 A "channel" (also "feed") is a website offering its headlines, news or content in a simplistic format.

Repositório da APSI
 Last Downloaded: 2006-07-20 10:25:58. [XML](#) [Remove](#)

[Reconhecimento de Voz - Voice Car System \(VCS\)](#) Os avanços tecnológicos possibilitam que os computadores reconheçam a voz humana e a interpretem, de forma a executarem determinadas tarefas previamente definidas. De modo a demonstrar a aplicabilidade destas novas tecnologias desenvolvemos uma aplicação que tem por objectivo a criação de uma interface de reconhecimento de voz, recorrendo para isso ao módulo Sensory Voice Extreme™ Toolkit. A aplicação em causa simula o controlo de algumas funções de um automóvel, activadas através do reconhecimento da voz humana. Numa utilização a nível real, a interacção do condutor perante a placa de reconhecimento da voz deverá ser efectuada através de um módulo que poderá ser integrado, por exemplo, no computador de bordo do automóvel.

[O novo ciclo de desenvolvimento de sistemas de informação – dos processos de negócio à operação em poucas horas](#) O processo de desenvolvimento tradicional de Sistemas de Informação tem enormes problemas derivados do tempo que demora tanto a construir a primeira versão do sistema de informação (pelo menos vários meses mas para grandes projectos normalmente mais de um ano) como também, e principalmente, a alterar ou acrescentar requisitos. Estes problemas são cada vez mais graves porque estas alterações são cada vez mais frequentes e além disso os Sistemas de Informação têm ciclos de vida cada vez mais curtos. Neste artigo propomos a utilização de ferramentas de concepção e desenvolvimento do tipo RAD e que tiveram uma grande evolução nos últimos anos. Estas ferramentas de última geração – como aquelas produzidas pela OutSystems em Portugal – estão agora preparadas para suportar um novo ciclo completo de desenvolvimento de sistemas de informação que permite cobrir todas as etapas tradicionais em poucas horas. Esta tecnologia foi validada com um caso de estudo (Portfolios da LEIC) e essa experiência é relatada neste artigo.

[Running and Debugging UML Models](#) Software development evolution is a history of permanent seeks for raising the abstraction level to new limits overcoming new frontiers. Executable UML (xUML) comes this way as the expectation to achieve the next level in abstraction, offering the capability of deploying a xUML model in a variety of software environments and platforms without any changes. This paper comes as a first expedition inside xUML, exploring the main aspects of its specification including the action languages support and the fundamental MDA compliance. We also explore the model debugging capabilities as a premature means of conceptual fail discovery. In this paper is presented a new xUML tool called XIS-xModels that gives Microsoft Visio new capabilities of running and debugging xUML models.

[Formalizing Object-Relational Structural Metrics](#) This paper describes a meta-modeling approach to formalize a set of metrics suited for object-relational databases. Those metrics are expected to help database designers in their activity, offering a mechanism for comparing and selecting, among possible schema designs, the one with more quality. Formalization of metrics is important for avoiding misinterpretation of

Publico.pt Local
 Last Downloaded: 2006-09-04 09:08:25. [XML](#) [Remove](#)

[Quercus pede fim da caça na Lagoa de Melides](#)

Chumbo contamina zona húmida alentejana

No dia em que começa a caça às aves aquáticas, a Quercus alertou hoje para a contaminação do chumbo nas zonas húmidas e pediu o fim da caça na Lagoa de Melides, no concelho de Grândola.

[Normalização do Aeroporto Sá Carneiro vai demorar algumas horas](#)

Aviação

Figura 13: Print screen do leitor RSS AmphedaDesk

A subscrição de feeds RSS implica um conjunto diversificado de vantagens para o utilizador, nomeadamente [Çelikbas, 2005]:

1. **Fácil cancelamento** – No processo de cancelamento de um determinado feed o utilizador não precisa de enviar uma mensagem “unsubscribe” por e-mail como

acontece com as *listservs* ou seguir um complexo processo numa página *Web*. Para cancelar um *feed* o utilizador basta eliminar o *feed* da lista;

2. **Gestão dos conteúdos** – disponibiliza um serviço de alerta sobre novos conteúdos que são publicados, assim como os conteúdos que ainda não foram lidos. Por exemplo no leitor RSS Wizz RSS 2.1.4 os títulos dos artigos que ainda não foram consultados são identificados com uma bola verde, enquanto as que já foram lidas são identificadas com uma bola vermelha.

4.6. A TECNOLOGIA RSS NA PUBLICAÇÃO CIENTÍFICA

O RSS é um formato baseado na tecnologia XML que permite listar o conteúdo de páginas *Web* facilitando a distribuição e disseminação dos seus conteúdos. É um formato particularmente prático para a consulta de informação que está em permanente actualização ou alteração. O RSS funciona como um sinal de que algures na *Web* uma página *Web* sofreu alterações. A sindicância e a anotação estão na ordem do dia e têm contribuído para a transformação dos actuais processos de comunicação e de recuperação de informação [Hammond *et al.*, 2004].

Actualmente o RSS é largamente utilizado nos *Weblogs* e pelos jornais, começando a dar os primeiros passos no contexto das publicações científicas periódicas de diversas áreas. No entanto, os *feeds* RSS utilizados no contexto das revistas científicas electrónicas diferem dos *feeds* regularmente utilizados nos jornais de notícias ou nos *Weblogs*, num requisito chave para os leitores. De facto, é essencial para os consumidores de publicações científicas, que os *feeds* contenham informação suficiente de modo a que estes possam citar ou produzir uma citação para um determinado artigo, dentro do editorial de uma revista científica. Consequentemente surge a necessidade de adicionar informação para além dos principais elementos RSS utilizados na descrição dos recursos, nomeadamente os elementos *title*, *link* e (opcionalmente) *description*. Com efeito, a informação contida nos elementos básicos do RSS não são suficientes que permitam aos autores de publicações científicas produzir citações, resultando na necessidade de se proceder à utilização de metadados na descrição semântica dos artigos científicos [Hammond, 2003].

Com efeito, os objectivos originais do RSS focavam a descrição de metadados na criação de *feeds*, mas uma recente tendência associou a tecnologia RSS à sindicância de *blogs* e assim limitar os seus poderes descritivos [Hammond, 2003].

Deste modo, uma característica que torna interessante a utilização do RSS no âmbito das publicações científicas está no facto de permitir incluir metadados adicionais na descrição de recursos. O facto de o RSS 1.0 estar em conformidade com a especificação RDF/XML torna-o ideal para a inclusão suplementar de metainformação, promovendo a troca estruturada de metadados. O *Dublin Core* é o vocabulário candidato para introduzir elementos de metadados dentro de um *feed* RSS, tendo em conta que o *Dublin Core* é um vocabulário de 15 elementos desenvolvido no âmbito da *Dublin Core Metadata Initiative* (DCMI) operando como língua franca no processo de descrição de metadados na *Web*. No seio da comunidade RSS já foram definidas um conjunto de instruções a serem seguidas para a utilização do vocabulário do *Dublin Core* no âmbito do RSS 1.0 [Beged *et al.*, 2000]. Da mesma forma, o vocabulário *Publisher Requirements for Industry Standard Metadata*⁵⁴ é também utilizado no seio da comunidade RSS uma vez que permite a extensão da utilização dos elementos do *Dublin Core* no que se refere à informação bibliográfica dos artigos, designadamente: *issn*, *volume*, *number*, *startingPage*, etc. De facto, os RSS contêm um mecanismo de extensão modular simples que permite organizar novos vocabulários [Hammond *et al.*, 2004].

O facto de os editores científicos disponibilizarem os seus *feeds* RSS e consequentemente procederem à sindicância dos seus metadados traz vantagens significativas tanto para os autores dos artigos científicos como para quem os publica, no que se refere à divulgação e disseminação dos conteúdos científicos desenvolvidos. O RSS, na sua essência, permite abranger uma área significativamente extensa da *Web* e consequentemente permitir uma maior projecção e visibilidade dos trabalhos científicos desenvolvidos pelos investigadores [Hammond *et al.*, 2004]. Por outro lado, “quantos mais dados disponibilizarmos acerca dos nossos conteúdos, mais caminhos vão dar a eles” [Hammond *et al.*, 2004].

Uma aplicação imediata da tecnologia RSS no contexto da publicação científica consiste no serviço de alerta para novos artigos que são disponibilizados em tabelas de conteúdos, ou no caso particular da *IngentaConnect*⁵⁵ que procede ao envio de notificações para os seus subscritores de *feeds*, a informar sobre novos conteúdos que foram publicados [Hammond *et al.*, 2004]. Deste modo o utilizador é notificado acerca de novos conteúdos científicos que são

⁵⁴ <http://www.prismstandard.org/>

⁵⁵ <http://www.ingentaconnect.com/>.

disponibilizados por uma determinada revista científica ou grupo de trabalho, acompanhada de informação básica do artigo, designadamente o título, a identificação do URL de acesso e opcionalmente uma breve descrição do artigo. Se esse artigo for do interesse do utilizador, este pode de seguida proceder ao acesso do conteúdo completo do artigo, de acordo com as normas de utilização de cada revista científica, através da informação disponibilizada pelo elemento *link* do RSS. Alguns editores de publicações científicas, nomeadamente a *Nature Publishing Group*, *International Union of Crystallography*, *IngentaConnect*, adicionaram aos seus *feeds* RSS, metadados através da utilização do conjunto de elementos definidos no *Dublin Core* e no *PRISM*, enquanto outros editores, como por exemplo a *BioMed Central*, *Institute of Physics*, *Oxford University Press*, *Extenza* optaram apenas por utilizar os elementos básicos do *Dublin Core* [Hammond *et al.*, 2004].

De seguida é apresentada um conjunto de tabelas que indicam algumas editoras científicas que disponibilizam os seus conteúdos em formato RSS, adaptado de [Hammond *et al.*, 2004].

Editora Científica	Versão RSS	Conteúdos disponibilizados
NPG (Nature Publishing Group)	RSS 1.0	Tabela de conteúdos, artigos, notícias e lista de empregos
IUCr (Int. Union of Crystallography)	RSS 1.0	Tabela de conteúdos e artigos de Acesso Livre
IngentaConnect	RSS 1.0	Publicações mais recentes

Tabela 5: Editoras científicas que utilizam os elementos de metadados do vocabulário do DC e PRISM na criação dos *feeds* RSS

Editora Científica	Versão RSS	Conteúdos disponibilizados
D-Lib Magazine	RSS 1.0	Tabela de conteúdos e artigos publicados do mês
Ariadne	RSS 1.0	Artigos
BMC (BioMed Central)	RSS 1.0	Tabela de conteúdos e artigos mais visitados
IOPP (Institute of Physics)	RSS 1.0, RSS 0.91	Tabela de conteúdos,

Publishing)		notícias, lista de empregos, revisões, eventos, lançamentos de novos produtos
OUP (Oxford University Press)	RSS 1.0	Tabela de conteúdos
Extenza	RSS 1.0	Tabela de conteúdos
American Journal of Neuroradiology	RSS 1.0	Tabela de conteúdos
National Geographic News	RSS 1.0	Notícias

Tabela 6: Editoras científicas que utilizam os elementos de metadados do vocabulário do DC na criação dos feeds RSS

Editora Científica	Versão RSS	Conteúdos disponibilizados
Montague Institute Review	RSS 0.91	Tabela de conteúdos
AIP (American Institute of Physics)	RSS 2.0	Tabela de conteúdos
Medscape	RSS 2.0	Tabela de conteúdos
Blackwell Publishing	RSS 1.0	Notícias de imprensa
BMJ (British Medical Journal)	RSS 0.91	Tabela de conteúdos
ACS (American Chemical Society)	RSS 0.91	Notícias

Tabela 7: Editoras científicas que não utilizam elementos de metadados na descrição das publicações

No directório *eFeeds* pode ser consultada informação sobre outros grupos de trabalho e revistas que publicam artigos científicos e disponibilizam os seus conteúdos em formato RSS para além dos que foram enunciados acima [McKiernan, 2005a]. A *eFeeds:Web Feeds from Electronic Journals* é responsável pela publicação de uma lista actualizada de todas as revistas científicas electrónicas, organizadas por diferentes áreas científicas que disponibilizam os seus conteúdos em formato RSS ou ATOM. O directório RSS: *Rich Site Services* [McKiernan, 2005b]

disponibiliza uma lista actualizada de bibliotecas que disponibilizam *Web feeds* em formato RSS ou ATOM.

No contexto da publicação científica, verifica-se que o RSS para além de ser utilizado como um serviço de alerta, as editoras também disponibilizam *feeds* RSS para distribuir e disseminar um conjunto diversificado de serviços, designadamente lista de empregos na área da investigação científica, informação de produtos, eventos, etc. Mas não é exclusivamente sobre informação mais recente que a tecnologia RSS pode ser utilizada. Efectivamente uma utilização importante do RSS está envolvida na construção e manutenção de *feeds* RSS que representam repositórios de dados estruturados. Assim, tendo em conta que o RSS compreende uma estrutura aberta de metadados, contribui para que as bibliotecas procedam à recolha e análise transparente dos dados provenientes de diversos *feeds* RSS. Isto permite, por um lado, às bibliotecas cobrir um conjunto mais diversificado de áreas, com conteúdos sempre actualizados e por outro lado proceder à sindicância dos seus conteúdos para o exterior [Hammond *et al.*, 2004].

Um exemplo que demonstra este facto é o Instituto Nacional de Saúde (*National Institutes of Health* - NIH) de Bethesda, em particular a biblioteca do Instituto Nacional de Cancro (*National Cancer Institute* -NCI) que estão a criar uma base de dados, denominada LION (*Library ONline*) que contém *feeds* RSS agrupados da Internet [Çelikbas, 2005]. Utilizaram a tecnologia RSS para agregar os conteúdos recolhidos da Internet no sistema da biblioteca do Instituto Nacional de Cancro e proceder à distribuição desses mesmos conteúdos pelo sistema de bibliotecas do Instituto, facilitando aos seus utilizadores o acesso a novos conteúdos através da disponibilização dos *feeds* RSS. Este sistema estabelece a ligação com um conjunto diversificado de fontes de informação que disponibilizam os seus conteúdos sob a forma de *feeds* RSS, nomeadamente: *BBC News: Health; Moreover: Breast Cancer News; Moreover: Cancer News; News York Times: Health; e a Reuters Health eLine* [Çelikbas, 2005].

Outra propriedade importante do RSS está no facto desta tecnologia não estar apenas restrita à sindicância de informação textual. Com efeito, tem vindo a ser utilizada na transmissão completa de *data sets* científicos. Exemplo disso é o trabalho desenvolvido por *Peter Murray-Rust* e *Henry Rzepa* que utilizam o RSS para distribuir dados químicos codificados em *Chemical Markup Language* (CML) [Hammond *et al.*, 2004] efectuando também estudos experimentais com o *Mathematical Markup Language* (MathML) e com o *Scalable Vector Graphics* (SVG) [Hammond *et al.*, 2004].

Outra utilização do RSS em informação não textual é o já anteriormente apresentado *podcasting*, que independentemente dos dados serem em formato áudio ou imagem podem ser descarregados, por exemplo, para um *iPod* ou para qualquer dispositivo portátil semelhante. Neste caso, o *feed* RSS não contém o conteúdo mas sim uma referência para esse conteúdo, através da utilização de uma aplicação adequada que permita descarregar os dados para o respectivo dispositivo.

Actualmente, para além do RSS já existem outras ferramentas igualmente sofisticadas que permitem a sindicância de metadados. A comunidade de bibliotecas digitais têm utilizado o protocolo *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH) [Lagoze, Van de Sompel *et al.*, 2002] como ferramenta para a sindicância de metadados. Tanto o RSS como o protocolo OAI são um meio utilizado para a troca de dados. Efectivamente o protocolo OAI-PMH é utilizado para “a disseminação de metainformação. É uma forma de os repositórios (denominados *data providers*) partilharem (exporem) os seus metadados para serem recolhidos (*harvested*) por serviços (*service providers*) que permitem a pesquisa por entre vários repositórios *OAI-Compliant*” [LusoDSpace, 2005].

No entanto, apesar de o RSS e o OAI-PMH terem a mesma estrutura tecnológica na base da sua implementação, têm objectivos distintos. Ambos utilizam documentos XML, que são transportados através do protocolo HTTP, para além disso, ambos permitem múltiplos vocabulários. Embora o RSS seja predominantemente utilizado no processo de sindicância de conteúdos (normalmente através da indicação de referências para os conteúdos), enquanto o OAI-PMH foca o seu trabalho na recolha (*harvesting*) de metadados [Hammond *et al.*, 2004].

O RSS define uma metodologia simples de encapsulamento que pode ser usada por um conjunto diversificado de classes de aplicações, designadamente os agregadores ou leitores RSS, enquanto o OAI-PMH define um esquema e um protocolo ao nível aplicacional. O RSS serve particularmente a transferência de dados para o ambiente de trabalho do utilizador, enquanto o OAI-PMH foi desenvolvido para gerir processos entre sistemas tipicamente sincronizados de repositórios institucionais [Hammond *et al.*, 2004].

A característica chave que estabelece a distinção entre o RSS e outros protocolos de sindicância está no facto de o RSS estar mais orientada para uma solução *Business to consumer* (B2C), uma vez que permite uma maior visibilidade dos conteúdos disponibilizados nas páginas

Web através da agregação e distribuição dos mesmos conteúdos pelos subscritores dos *feeds* RSS.

4.7. RESUMO

A tecnologia RSS tem sido amplamente utilizada no contexto dos *blogs* e das notícias, uma vez que os seus conteúdos estão sujeitos a frequentes alterações. No entanto a consulta de diversas revistas científicas electrónicas permitiu verificar que estas já começam a utilizar a tecnologia RSS no processo de distribuição e disseminação das publicações científicas. De facto a revisão de literatura efectuada no âmbito deste trabalho de mestrado, refere casos verdadeiramente inovadores que utilizam a tecnologia RSS para agregar conteúdos e proceder à sua distribuição por diversos sistemas. No âmbito desta dissertação de mestrado, a tecnologia RSS foi seleccionada no sentido de tirar partido das suas potencialidades no que se refere à sindicância de conteúdos, e adaptá-las no contexto da publicação científica.

5. DESCRIÇÃO GERAL DO TRABALHO REALIZADO

Nas últimas décadas tem-se verificado um crescimento exponencial da informação em formato digital disponível aos utilizadores da *Web*. Esta percepção, contribui para a necessidade de organizar e agrupar a informação a nível semântico, e de desenvolver esforços na implementação de mecanismos inteligentes de pesquisa e de navegação, de modo a facilitar e a promover o rápido acesso à informação digital disponível na *Web*. Efectivamente, o crescimento da informação digital disponível na *Web* e conseqüentemente o aumento do número de utilizadores que usam a *Web* para efectuar pesquisas e navegarem na rede, têm contribuído para a necessidade de organizar o imensurável número de páginas *Web* que surgem todos os dias a todas as horas na Internet. Por outro lado, na Internet, a informação encontra-se geograficamente distribuída por todo o mundo, com diversos métodos de acesso, formatos e estruturas de armazenamento, existindo dezenas de variações nos sistemas operativos e nas aplicações de acesso à informação.

Neste contexto, este projecto prevê a implementação de uma instância do protótipo RDF desenvolvido no âmbito do projecto OmniPaper (ver secção 2.4.2 e 2.5), com a perspectiva de instanciar o sistema no contexto da literatura científica, e utilizar a camada de metadados para proceder à sindicância de conteúdos.

O projecto OmniPaper foi desenvolvido no âmbito das notícias de jornais, enquanto o trabalho apresentado nesta dissertação de mestrado está contextualizado no âmbito das publicações científicas. Efectivamente as notícias e as publicações científicas electrónicas são duas unidades de informação distintas, no que se refere à estrutura semântica, aos seus consumidores, etc. No entanto ambas tentam tirar partido das funcionalidades que o meio tecnológico oferece. O desenvolvimento das tecnologias de informação e comunicação, em particular da Internet e conseqüentemente a sua crescente utilização associada as exigências dos consumidores de informação têm contribuído para a reestruturação profunda dos meios tradicionalmente utilizados na publicação de informação em geral, e em particular das notícias de jornais e da publicação científica.

De facto o meio electrónico disponibiliza um conjunto de meios e funcionalidades de suporte ao processo de publicação de informação, que os tradicionais meios utilizados na publicação baseada em documentos impressos não oferecem.

Hoje em dia as publicações digitais vão além da simples reprodução electrónica do conteúdo editado das versões já impressas, disponibilizando dados e informações complementares que ficam de fora da edição em papel, designadamente hiperligações, excertos de entrevistas, imagens de coberturas de eventos, etc., em diversos formatos, como por exemplo vídeos e áudio.

Por outro lado as notícias publicadas na *Web* chegam ao utilizador em tempo real, ou seja as notícias chegam em intervalos de 10 a 15 minutos. Esta instantaneidade da notícia acontece através de parecerias entre jornais que publicam notícias provenientes de várias agências jornalísticas. Foi neste contexto que a tecnologia RSS ganhou popularidade no meio jornalístico, pois permite a agregação de notícias provenientes de diversas empresas de notícias geograficamente dispersas. No ano 2000, a utilização da tecnologia RSS difundiu-se para grandes empresas de notícias como a *Reuters*, CNN e a BBC. Estas empresas permitiam que outras agências de informação incorporassem as suas notícias e resumos, através de vários acordos de utilização, com o objectivo de abranger um maior número de leitores.

Actualmente, apesar da tecnologia RSS ser amplamente utilizada no contexto das notícias de jornais, já começa a dar sinais de utilização no contexto das publicações científicas periódicas de diversas áreas, promovendo a distribuição e disseminação dos seus conteúdos. No âmbito deste trabalho de mestrado foi utilizada a tecnologia RSS no processo de descrição da metainformação dos artigos da APSI, seguido da criação de uma base de metadados, de modo a permitir proceder à sindicância de conteúdos e suportar as funcionalidades de navegação e pesquisa desenvolvidas no sistema.

Neste capítulo pretende-se apresentar uma descrição geral do trabalho realizado, de acordo com os objectivos propostos no âmbito desta dissertação de mestrado.

5.1. OBJECTIVOS DO PROTÓTIPO

O protótipo desenvolvido neste trabalho de mestrado é uma instância do protótipo RDF implementado no projecto OmniPaper. A sua implementação compreende vários objectivos, designadamente: (1) o desenvolvimento de mecanismos inteligentes de pesquisa e de navegação semântica sobre conteúdos científicos que permita aos utilizadores um acesso estruturado e simultâneo a um conjunto de publicações científicas e (2) disponibilizar os mesmos

recursos sob a forma de *feed* RSS, aos utilizadores que usam a tecnologia RSS como um mecanismo de consulta e de acesso à informação disponibilizada na *Web*.

A implementação do protótipo compreendeu a realização dos seguintes passos:

1. Levantamento e análise dos vocabulários normalizados de metadados, específicos no domínio da literatura científica;
2. Definição e implementação em RDF/XML do perfil de aplicação utilizando o vocabulário smes;
3. Definição de um *template* que compreende a estrutura de metadados utilizada na descrição das publicações científicas;
4. Codificação em RSS dos artigos armazenados no repositório da APSI utilizando uma *stylesheet* definida em XSLT;
5. Criação dos *feeds* RSS para cada categoria do sistema de classificação da ACM CCS;
6. Definição e desenvolvimento de uma camada conceptual (através da estrutura hierárquica do sistema de classificação da ACM CCS e a utilização de um *thesaurus* léxico);
7. Integração e processamento completo dos dados.

O desenvolvimento da estrutura de metadados que compreende a selecção dos vários elementos de metadados a utilizar, o perfil de aplicação criado, e os processos de suporte aos mecanismos de navegação, pesquisa e sindicância de conteúdos desenvolvidos no protótipo serão apresentados no capítulo 6.

5.2. COLECÇÃO DE TESTE

Conforme se disse na Introdução desta dissertação de mestrado, o *data set* utilizado na concepção do protótipo compreende as publicações científicas provenientes do repositório da APSI.

O Repositório da APSI é o repositório institucional da Associação Portuguesa de Sistemas de Informação, constituído com o objectivo de armazenar, preservar, divulgar e dar acesso aos artigos publicados na revista "Sistemas de Informação", aos artigos aceites na Conferência da Associação Portuguesa de Sistemas de Informação (CAPSI) e à produção intelectual em formato

digital. O Repositório da APSI pretende reunir, num único sítio, o conjunto das publicações científicas da APSI contribuindo desse modo para o aumento da sua visibilidade e impacto e garantindo a preservação da sua memória intelectual [APSI, 2005].

O Repositório da APSI está inserido no crescente movimento de constituição de repositórios e arquivos de Acesso Livre, que utiliza o protocolo OAI-PMH para a troca de dados, desenvolvido no âmbito do *Open Archives Initiative* [OAI, 2002]. Esta iniciativa surgiu no seio da comunidade dos “*e-prints*”⁵⁶ e partiu de uma abordagem essencialmente técnica (de que resultou o protocolo OAI-PMH), sem grande preocupação “filosófica”. Mas ao fornecer uma base estável para a interoperabilidade de “arquivos” abertos, e face ao número crescente de servidores que o implementam, contribui para dar maior visibilidade e encorajamento ao movimento de Acesso Livre ao Conhecimento” [LusoDspace, 2005].

O Repositório da APSI foi implementado na plataforma *DSpace*⁵⁷, desenvolvida pelas bibliotecas do *Massachusetts Institute of Technology* (MIT) em conjunto com a *Hewlett-Packard* (HP). A plataforma foi disponibilizada publicamente em Novembro de 2002 de acordo com os termos da *Berkeley Standard Distribution license* (BSD) [LusoDspace, 2005].

O DSpace é uma plataforma de software *Open Source* que permite recolher, descrever, distribuir e preservar a longo prazo documentos digitais. Esta plataforma foi concebida de modo a permitir a cada comunidade adaptá-la de acordo com as suas necessidades específicas, permitindo-lhes deste modo definir e controlar o processo de *workflow*.

Para a pesquisa e recuperação dos documentos, o processo de submissão de documentos no DSpace permite a sua descrição usando uma versão qualificada do vocabulário normalizado de metadados do *Dublin Core* baseado no formato de registo sugerido pela *Libraries Working Group Application Profile*⁵⁸.

⁵⁶ Um termo genérico utilizado para versões electrónicas de artigos científicos ou qualquer outro documento científico semelhante. Estes podem incluir artigos de revistas científicas antes de serem revistos, a versão final de uma publicação, artigos apresentados em conferências, etc. [FAIR Synthesis: Glossary, 2006].

⁵⁷ <http://www.dspace.org/>

⁵⁸ <http://dublincore.org/documents/library-application-profile/>.

5.3. ESPECIFICAÇÃO FUNCIONAL DO SISTEMA

Como se disse anteriormente, um dos objectivos que se pretende alcançar com a implementação do protótipo consiste em disponibilizar mecanismos inteligentes de pesquisa e de navegação sobre os conteúdos científicos que estão armazenados no repositório institucional da APSI. Assim, conforme é ilustrado na figura 17, a concepção do protótipo passou pela implementação de mecanismos de pesquisa e de navegação que facilita aos consumidores de conteúdos científicos o acesso estruturado ao conhecimento científico produzido na área dos Sistemas de Informação.

O protótipo desenvolvido compreende as seguintes funcionalidades:

- Armazenamento de informação: O protótipo permite o armazenamento dos *feeds* RSS criados.
- Pesquisa simples: Esta pesquisa permite aos utilizadores a pesquisa de termos no elemento de metadado *description*. No desenvolvimento deste método de pesquisa foi utilizado um *thesaurus* léxico: o *WordNet*, com características orientadas à organização da informação resultando na melhoria dos processos de navegação e pesquisa, nomeadamente o mecanismo de expansão da *query*. Os resultados apresentados são o título e a descrição do artigo. Cada título contém uma hiperligação para a origem do artigo, designadamente o repositório da APSI.
- Pesquisa sobre uma árvore de conceitos: permite aos utilizadores navegar sobre uma estrutura hierárquica, baseada no sistema de classificação da ACM CCS e subscrever os *feeds* RSS dentro de cada categoria específica da estrutura hierárquica do sistema de classificação da ACM. A pesquisa de termos é executada sobre o elemento de metadados *dc:subject*.

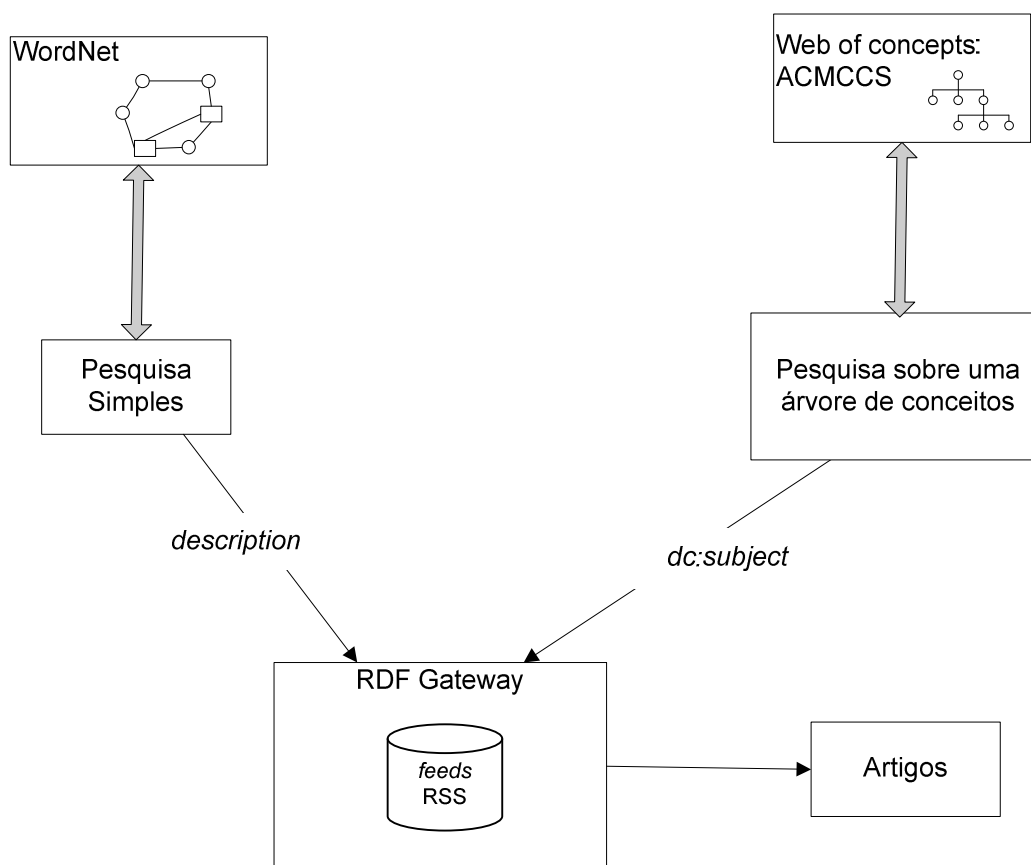


Figura 14: Camada semântica do sistema

5.4. DESENHO DO SISTEMA

O sistema desenvolvido no âmbito desta dissertação de mestrado segue o conceito do sistema implementado no projecto OmniPaper. No entanto, como o sistema OmniPaper foi desenvolvido no âmbito das notícias publicadas em jornais, a sua instanciação no contexto das publicações científicas implicou a execução de algumas alterações ao nível dos dados e dos processos.

De facto, a estrutura do sistema ilustrada na figura 18, e que foi seguida na implementação do sistema desenvolvido no âmbito das publicações científicas é muito semelhante à estrutura do sistema desenvolvido no projecto OmniPaper. Essencialmente, as semelhanças traduzem-se na instanciação dos mecanismos de pesquisa e de navegação desenvolvidos no sistema OmniPaper e na definição de uma camada de metadados que suporta esses mesmos mecanismos. No entanto atendendo às especificidades das publicações científicas houve necessidade de proceder a alterações de alguns desses processos. As cores ilustradas na figura 18 pretendem precisamente representar as componentes que foram

instanciadas na íntegra, e as que foram parcialmente instanciadas a partir do sistema OmniPaper.

Assim, a componente ilustrada a rosa, designadamente o mecanismo de pesquisa, foi instanciado do sistema OmniPaper, enquanto que as componentes representadas a azul foram igualmente instanciadas do sistema OmniPaper, mas no contexto das publicações científicas tiveram que ser alteradas. As restantes componentes foram definidas de acordo com os requisitos do sistema.

As letras B e C apresentadas no desenho do sistema pretendem indicar as fases envolvidas no desenvolvimento do protótipo. Enquanto que a fase A, ilustrada num rectângulo a tracejado pretende indicar que não esteve compreendida na concepção deste sistema, uma vez que não foram utilizados os artigos armazenados no repositório da APSI, mas sim os seus metadados descritos segundo o XML *schema* OAI-PMH.

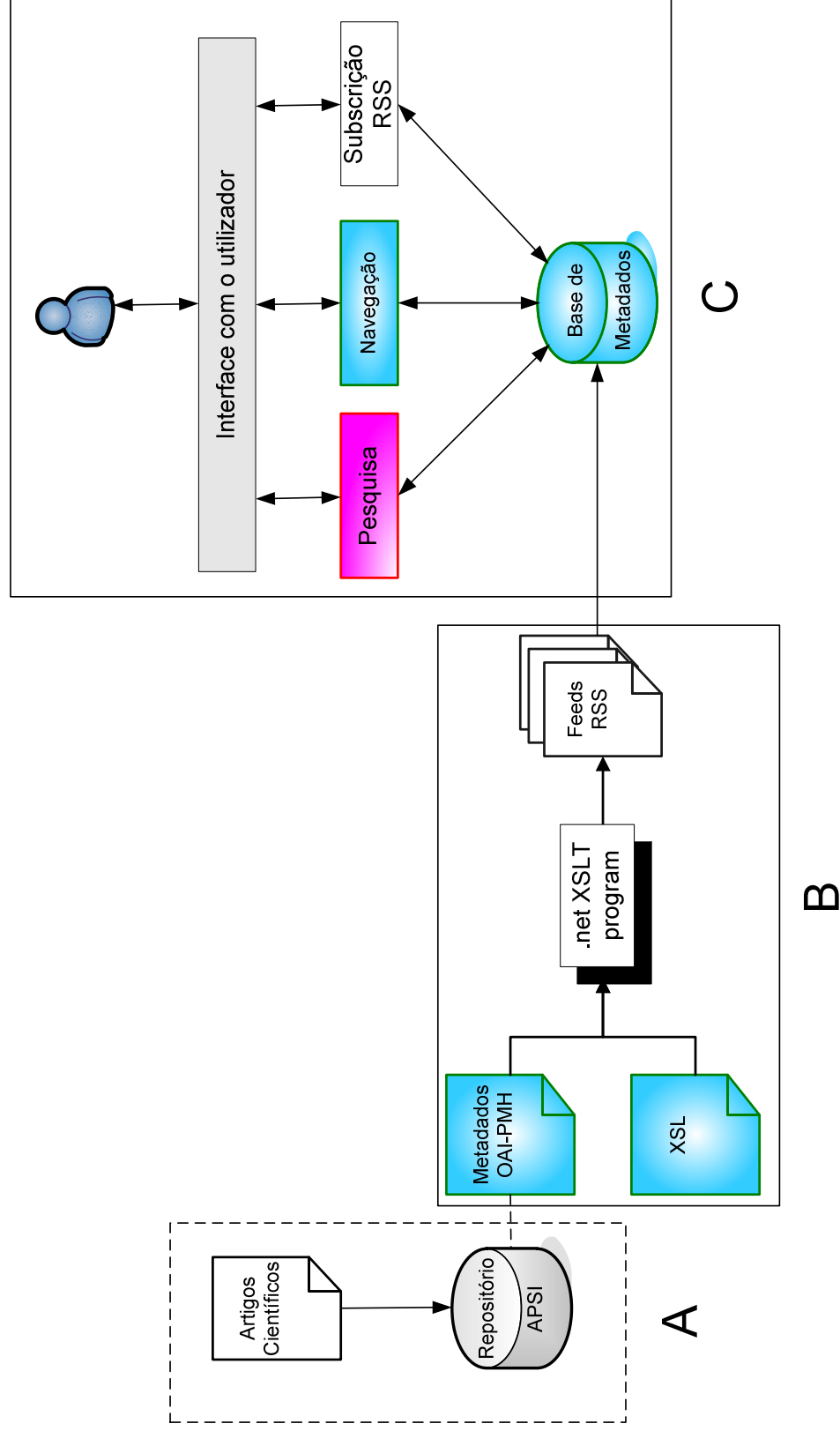


Figura 15: Desenho do sistema

O desenvolvimento da fase B, ilustrada na figura 18, compreendeu a implementação de um conjunto de tarefas que permitiram obter os *feeds* RSS.

A primeira tarefa consistiu no levantamento e análise de um conjunto de vocabulários normalizados de metadados no domínio da literatura científica. Esta análise permitiu seleccionar um conjunto de elementos de metadados adequados à descrição semântica dos artigos científicos da APSI. O resultado desta tarefa deu origem à definição do perfil de aplicação (ver secção 6.2) e ao *template* RSS (ver apêndice A).

De seguida e tendo em conta que a metainformação dos artigos científicos disponibilizados no *data set* utilizado estavam descritos segundo o XML *schema* OAI-PMH, houve necessidade de se proceder à sua transformação de acordo com a estrutura compreendida no *template* RSS definido. Deste modo, procedeu-se à criação de uma *stylesheet* em XSLT para executar a transformação dos metadados dos artigos científicos da APSI disponibilizados segundo o XML *schema* OAI-PMH, para a estrutura compreendida no *template* RSS definido.

No sistema OmniPaper foi também definida uma estrutura de metadados, constituída por um leque variado de elementos, provenientes de diversos vocabulários normalizados de metadados. No entanto, tendo em conta que as notícias publicadas em jornais e os artigos científicos são dois recursos de informação distintos, originaram estruturas de metadados distintas.

Por outro lado, no sistema OmniPaper a metainformação das notícias disponibilizadas pelos fornecedores locais distribuídos estavam em formato XML, e portanto procedeu-se à sua transformação para o formato RDF/XML. De facto no projecto OmniPaper, o RDF foi a abordagem tecnológica seguida no processo de descrição dos metadados dos artigos de notícias. Assim, no projecto OmniPaper foi também definida uma *stylesheet* em XSLT para executar a transformação dos metadados disponibilizados pelos fornecedores de notícias em XML, para o formato RDF. A descrição de uma única notícia correspondia a um documento RDF, enquanto que no sistema desenvolvido no contexto das publicações científicas, os metadados disponibilizados em OAI-PMH foram transformados para o formato RSS, dando origem ao *feed* RSS. Este *feed* é constituído por um conjunto de *items* que compreendem a descrição da metainformação de todos os artigos científicos disponibilizados pelo *data set* utilizado.

O *feed* RSS criado foi armazenado numa base de metadados, conforme está ilustrado no desenho do sistema apresentado na figura 18. É sobre esta base de metadados que os utilizadores desenvolvem as suas pesquisas e efectuam a navegação sobre a estrutura de conceitos representada pelo sistema de classificação da ACM CCS.

A ferramenta utilizada na criação da base de metadados, tanto no OmniPaper como no sistema desenvolvido no âmbito desta dissertação de mestrado foi o *RDF Gateway*. Esta é uma ferramenta proprietária que combina um sistema de gestão de base de dados nativa RDF e um servidor HTTP.

O processo de pesquisa implementado no sistema desenvolvido, no contexto desta dissertação de mestrado tem associado o *WordNet*, uma ferramenta com propriedades orientadas para a organização da informação, que permite acrescentar valor aos mecanismos de pesquisa, nomeadamente desenvolver a expansão da *query*. Este processo de pesquisa é instanciado do sistema OmniPaper, sem se efectuar qualquer alteração, já que quando é efectuada uma pesquisa de um conceito, o *WordNet* devolve palavras relacionadas com esse conceito, que podem ser sinónimos, antónimos, advérbios, etc. Com efeito, o *WordNet* é um sistema constituído por um conjunto de nomes, verbos, adjectivos e advérbios ingleses, organizados em conjuntos de sinónimos onde cada um representa um conceito lexical, e portanto a sua utilização não está restrita a nenhum recurso de informação específico.

Relativamente ao processo de navegação desenvolvido no sistema implementado no contexto desta dissertação de mestrado, foi também instanciado do sistema OmniPaper, no entanto foi necessário realizar alterações. De facto no sistema OmniPaper o processo de navegação era efectuado sobre a estrutura hierárquica de conceitos, representada pelos IPTC-SC (ver secção 2.5), específicos para a descrição editorial do conteúdo das notícias. No entanto no âmbito desta dissertação de mestrado contextualizada nas publicações científicas, não seria adequado classificar as publicações científicas da APSI com termos específicos para a classificação de notícias de jornais. Deste modo foi utilizado o sistema de classificação da ACM CCS, mais indicado para a classificação de artigos científicos. Tendo em conta, que já existe uma versão RDFS disponibilizada pela ACM desta estrutura hierárquica, não houve necessidade de efectuar a sua codificação RDF, ao contrário do que sucedeu com os IPTC-SC, utilizados no sistema OmniPaper.

Para além das funcionalidades de pesquisa e navegação foi implementado no sistema, o processo denominado “subscrição RSS”, como é apresentado na figura 18. Este processo permite ao utilizador a subscrição dos diferentes *feeds* RSS, definidos para cada tópico compreendido no primeiro nível da estrutura hierárquica do sistema de classificação da ACM CCS. No entanto, tendo em conta que as publicações da APSI não estavam classificadas segundo o sistema de classificação da ACM CCS, houve necessidade de alterar o *data set* de modo a permitir realizar a prova de conceito. Deste modo procedeu-se à criação manual de *feeds* RSS para cada categoria do primeiro nível do sistema de classificação da ACM CCS, e permitir aos utilizadores proceder à sua subscrição, de acordo com as áreas específicas de interesse de cada utilizador. Este processo não foi implementado no sistema desenvolvido no projecto OmniPaper, uma vez que este não tinha como objectivo proceder à sindicância de conteúdos.

Estas funcionalidades foram disponibilizadas ao utilizador num ambiente mais utilizável e amigável possível através da definição da *interface* do protótipo desenvolvido. A *interface* definida neste sistema é diferente da *interface* definida no sistema desenvolvido no projecto OmniPaper, apesar de serem mantidos os processos de pesquisa e navegação. Os mecanismos de pesquisa e navegação desenvolvidos são traduzidos na *interface* através das seguintes funcionalidades:

- 1 Apresentação da opção de pesquisa: a interface apresentada aos utilizadores disponibiliza um rectângulo branco, onde o utilizador insere um termo de pesquisa, ou seja uma palavra-chave. De seguida a *query* é enviado ao sistema através do click de um botão. O resultado devolvido pelo sistema ao utilizador consiste na apresentação dos artigos científicos associados à palavra-chave introduzida, bem como um conjunto de significados que estão associados à mesma palavra-chave, permitindo ao utilizador desenvolver a expansão da *query*. Esta funcionalidade é permitida pela utilização do *WordNet*.
- 2 Apresentação da árvore de conceitos: Para além da opção de pesquisa mencionada anteriormente, é apresentado num painel à esquerda da *interface*, a árvore de conceitos que representa o sistema de classificação da ACM CCS. Inicialmente a árvore apresenta o primeiro nível da estrutura hierárquica. Quando o utilizador clica num determinado termo são apresentados todos os sub-termos associados. No painel principal são apresentados os títulos e uma descrição dos artigos que pertencem ao *subject* seleccionado.

- 3 Apresentação dos resultados: No painel principal, os resultados apresentados são o título e a descrição do artigo. Quando o utilizador clica no título, este é direccionado para a origem do artigo, que no âmbito deste trabalho é o repositório da APSI, procedendo à consulta do respectivo conteúdo do artigo.
- 4 Subscrição dos diferentes *feeds* RSS pode ser efectuada através dos ícones em cor laranja, disponibilizados na estrutura hierárquica representada pelo sistema de classificação da ACM CCS. O utilizador pode subscrever apenas os *feeds* RSS de acordo com a sua área de interesse.

O protótipo desenvolvido no âmbito desta dissertação de mestrado é uma aplicação orientada para a *Web*, em que o utilizador vai enviar e receber informação através do protocolo HTTP. As páginas HTML são geradas a partir do RDF *Server Pages* (RSP) uma linguagem *script* muito semelhante ao ASP.

O código RSP permite efectuar a interacção entre o utilizador e o motor de base de metadados, através da execução de *queries* sobre a base de metadados e procedendo ao envio de resultados para o utilizador.

5.5. RESUMO

Neste capítulo foram apresentadas as várias fases que compreendem o desenvolvimento do sistema implementado no âmbito desta dissertação de mestrado. Na descrição destas fases foram indicadas as componentes que foram instanciadas do sistema desenvolvido no projecto OmniPaper e as que foram implementadas de acordo com os requisitos do sistema.

A instanciação do sistema OmniPaper no contexto da publicação científica implicou alterações no sistema ao nível dos dados e dos processos. De facto os elementos de metadados utilizados na descrição das notícias publicadas em jornais são distintos dos elementos utilizados na descrição das publicações científicas, originando estruturas de metadados distintas. Por outro lado, ao nível dos processos as alterações foram menos significativas, mantendo-se o mecanismo de pesquisa e navegação semântica sobre os recursos, embora se tenha procedido à alteração da estrutura hierárquica de conceitos implementada no sistema OmniPaper.

6. DESCRIÇÃO DETALHADA DO TRABALHO REALIZADO

Neste capítulo são descritas as várias tarefas que estiveram envolvidas no desenvolvimento do protótipo, de acordo com a descrição geral do trabalho realizado, apresentado no capítulo 5.

6.1. ANÁLISE DE VOCABULÁRIOS NORMALIZADOS DE METADADOS

Na generalidade dos actuais sistemas de pesquisa, os termos de pesquisa adoptados consistem na pesquisa por palavras em texto integral. Nestes sistemas é disponibilizado ao utilizador um mecanismo de pesquisa, que consiste na procura de uma ou mais palavras-chave. De seguida é enviada uma instrução para o servidor, no qual executa um programa que através de um OR ou de um AND lógico procura todas as instâncias das palavras-chave introduzidas pelo utilizador.

Este processo ainda é bastante utilizado, mas tendo em conta o elevado crescimento da Internet e conseqüentemente a quantidade incontrolável de informação disponibilizada diariamente na Internet, os sistemas de recuperação de informação apresentam alguns problemas no que se refere aos resultados devolvidos aos utilizadores, uma vez que, nem sempre atingem os níveis de relevância e precisão desejados.

Neste contexto a aplicação dos metadados na descrição de recursos de informação da *Web* está a ganhar cada vez mais terreno e cada vez mais adeptos, uma vez que estes vêm impor uma ordem predefinida nos conteúdos semânticos da *Web* [Lagoze, 2001], e a sua utilização na descrição de recursos contribui significativamente para melhorar a eficácia e a eficiência dos actuais serviços de informação.

Os metadados estão presentes desde que os bibliotecários iniciaram o processo de elaboração de uma lista de itens para classificar e organizar os recursos das bibliotecas. Actualmente, o termo metadados é definido como dados sobre dados ou informação sobre informação (recursos). No âmbito do projecto OmniPaper, os metadados descrevem recursos que são artigos de notícias de jornais e portanto facilitam a sua descoberta e o seu acesso. No contexto deste trabalho, um recurso compreende a literatura científica publicada em revistas científicas e os metadados descrevem a informação específica do domínio da literatura científica, como por exemplo o autor da publicação, a data, o título, etc. De facto, a definição da camada de

metadados é a tarefa chave na implementação e desenvolvimento de todo este trabalho de mestrado.

A camada de metadados vai auxiliar o funcionamento dos mecanismos de pesquisa e de navegação, uma vez que estes processos serão efectuados sobre a metainformação dos documentos científicos e não sobre o texto integral dos documentos científicos. Deste modo, espera-se reduzir o universo de resultados, quando comparados com a generalidade dos actuais motores de pesquisa por palavras-chave em texto integral.

No âmbito deste trabalho, o processo de selecção e descrição dos metadados é fundamental no que se refere à definição e gestão de toda a estrutura organizacional dos conteúdos semânticos que compreendem a literatura científica. O processo de selecção dos elementos de metadados mais adequados à descrição dos diversos géneros de literatura científica precedeu o levantamento de um conjunto normalizado de vocabulários de metadados específicos no domínio da literatura científica. Verificou-se que têm sido desenvolvidas várias iniciativas por parte de diversas organizações, que trabalham no sentido de identificar e manter um conjunto normalizado de elementos de metadados. Foi com base na análise do conjunto de elementos de metadados mantidos por estas iniciativas que se procedeu à definição da estrutura de metadados que compõe este trabalho.

As iniciativas analisadas foram as seguintes:

- O guia de referência de metadados proposto e utilizado pelos departamentos académicos do MIT que estão envolvidos com a criação de recursos digitais e pelo grupo de trabalho das bibliotecas do MIT - *Metadata Advisory Group of the MIT Libraries* [MITLibraries, 2005]. Os vocabulários de metadados seleccionados deste guia, que melhor se enquadram no contexto da literatura científica foram os seguintes:
 - DDI (*Data Documentation Initiative*)
 - DC (*Dublin Core*)
- Organizações envolvidas com o desenvolvimento, adopção e promoção de metadados normalizados:
 - ADL (*Advanced Distributed Learning Initiative, developers*), promovido pelo SCORM (*Sharable Content Object Reference Model*)

- <http://www.adlnet.org>
- ALIC (*Advanced Learning Infrastructure Consortium*)
<http://www.alic.gr.jp/eng/>
- CETIS (*Centre for Educational Technology Interoperability Standards (UK)*)
<http://metadata.cetis.ac.uk/>
- CEN (*Comité Européen de Normalisation*)
<http://www.cenorm.be/cenorm/index.htm>
- DLF (*Digital Library Federation*) promovido pelo METS
<http://www.diglib.org/>
- ECC (*E-learning Competency Centre*)
<http://www.ecc.org.sg/>
- PRISM (*Publishing Requirements for Industry Metadata*)
<http://www.prismstandard.org/>
- EdNA (*Educational Network Australia*)
<http://www.edna.edu.au/>
- *European SchoolNet*
<http://www.educat.hu-berlin.de/~kluck/datahandbook.htm>
- GEM – *The Gateway to Educational Materials*
<http://www.geminfo.org/index.html>
- MEG – *The Metadata for Education Group*
<http://www.ukoln.ac.uk/metadata/education/>
- MedBiquitous – *Enabling medical education*
<http://www.medbig.org/>
- SC36 – *Standardization of IT for Learning, Education, and Training.*
<http://jtc1sc36.org/>
- EDitEUR – *European group for electronic commerce in the book and serials sectors.*
- IEEE – *Institute of Electrical and Electronics Engineering*
- IETF – *Internet Engineering Task Force*
- ISO – *International Standards Organization*
- NISO – *National Information Standards Organization*
- W3C – *World Wide Web Consortium*

Actualmente a *Dublin Core Metadata Initiative* (DCMI) é a principal iniciativa na área dos metadados utilizados na descrição de recursos de informação da *Web*. Começou em 1995 com um objectivo e um âmbito muito específicos: “desenvolver um formato descritivo fácil de criar e manter para facilitar a descoberta de recursos na *Web* através de diferentes domínios” [Lagoze, 2001].

O principal objectivo do DCMI consiste em criar um amplo e interdisciplinar consenso sobre o conjunto de elementos de metadados adequados à descrição de um conjunto de recursos que se encontram na Internet, e que são úteis às diversas comunidades ou domínios de interesse, no processo de descoberta e recuperação de informação [Sutton e Mason, 2001].

Apesar da reconhecida necessidade da utilização dos metadados em diversas organizações que se encontram a crescer e que se estão a direccionar cada vez mais para a *Web* e para a comunicação electrónica, uma vez que lhes facilita o processo de disseminação e acesso à informação [Liddy, 2005], a DCMI apenas endereça os seus interesses para com o processo de descoberta e recuperação de informação. A *DCMI Usage Board* é um órgão da DCMI cuja missão consiste em assegurar uma evolução ordenada de vocabulários de metadados. É responsável pela revisão e análise de propostas de alterações a elementos de metadados ou propostas sobre a apresentação de novos elementos, submetidas à apreciação do DCMI. Os resultados da sua deliberação podem ser uma recomendação DCMI ou a rejeição de uma recomendação, acompanhada de uma justificação. Uma recomendação aceite pela DCMI atribui um estado que pode ser “*Cross-Domain*” ou “*Domain-Specific*” reflectindo deste modo o nível de interoperabilidade [Sutton e Mason, 2001].

O *Dublin Core Metadata Element Set* (DCMES) [DCMES, 2003] é um vocabulário constituído por 15 elementos desenvolvido no âmbito da DCMI. Este vocabulário foi criado com o objectivo de disponibilizar um conjunto de elementos, partilhados por diversas disciplinas ou por organizações que tenham necessidade de classificar a informação. A definição de cada um dos seus elementos é suficientemente abrangente, para permitir adaptar-se a um diversificado número de situações e disciplinas de estudo. Actualmente a especificação do DCMES encontra-se na segunda versão e é uma norma ANSI/NISO (Z39.85-2001)⁵⁹ desde Setembro 2001 e uma norma ISO (15836-2003)⁶⁰ desde Fevereiro de 2003. O DCMES é constituído por um conjunto

⁵⁹ <http://www.niso.org/standards/resources/Z39-85.pdf>

⁶⁰ <http://www.niso.org/international/SC4/n515.pdf>

de elementos de metadados simples de compreender e extensível a outros modelos de descrição semânticos. Uma das suas principais características reside no facto de o conjunto de elementos de metadados que constituem o DCMES ser bastante abrangente e flexível permitindo a sua adaptabilidade a um diverso conjunto de domínios de aplicação. Um dos benefícios na utilização do DCMES na descrição de metadados está na sua interoperabilidade semântica, ou seja, a descrição semântica de recursos através do DCMES contribui para uma maior possibilidade dos mesmos recursos serem descobertos por outras aplicações externas de diferentes áreas. A interoperabilidade semântica promove a fácil utilização e rápida disseminação. Este é o objectivo que *Tim Berners-Lee* pretende alcançar quando se refere ao conceito da *Web Semântica*.

Aos elementos do DCMES podem ser associados outros termos que permitem especificar com mais detalhe o seu significado. Estes termos são denominados por elementos de refinamento de elemento. Os elementos do *Dublin Core* e os elementos de refinamento de elemento são propriedades [Johnston, 2005]. Uma propriedade é “um aspecto, característica, atributo ou relação específica, utilizada para descrever recursos” [Johnston, 2005]. Por exemplo, o elemento *dc:date* tem como elementos de refinamento de elemento os termos *dcterms:issued*, *dcterms:created*, *dcterms:available*, *dcterms:modified* e o *dcterms:valid*.

Por outro lado, aos elementos utilizados no refinamento do valor de um determinado elemento estão também associados esquemas de codificação⁶¹. O glossário do *Dublin Core* define um esquema de codificação como esquemas que disponibilizam informação contextual ou regras de análise gramatical que contribuem para uma melhor interpretação do valor de um elemento. Tal informação contextual pode ter a forma de vocabulários controlados que obedecem a uma notação formal, ou análise gramatical de regras [Woodley, 2003]. Existem dois tipos de esquemas de codificação: Vocabulário de Esquemas de Codificação⁶² e Sintaxe de Esquemas de Codificação⁶³ [Woodley, 2003].

A Sintaxe de Esquemas de Codificação consiste na atribuição de um tipo específico a um determinado valor, obedecendo a uma notação formal. Por exemplo, o elemento *dc:date* - "2000-01-01" - é escrita como uma expressão normalizada de uma data [Woodley, 2003]. Enquanto Vocabulário de Esquemas de Codificação indica que o valor de um elemento tem

⁶¹ do inglês *encoding schema*

⁶² do inglês *Vocabulary Encoding Schemes*

⁶³ do inglês *Syntax Encoding Schemes*.

como qualificadores de esquema de codificação outros esquemas que ajudam na interpretação do valor de um elemento [Woodley, 2003]. Por exemplo, o mesmo elemento *dc:date*, tem como qualificadores de esquema de codificação os esquemas DCMI *Period*⁶⁴ e W3C-DTF⁶⁵.

No âmbito deste trabalho de mestrado, os elementos de metadados identificados no vocabulário DCMES, seguem as recomendações feitas no documento *Expressing Qualified Dublin Core in RDF/XML* [Kokkelink e Schwänzl, 2002], apesar de esta ainda ser uma recomendação candidata da DCMI.

Em Agosto de 1999, o *Dublin Core Usage Board* (DCUB) formou um grupo de trabalho no domínio da educação DCMI *Education Working Group* (DC-Ed) [Hillmann e Sutton, 1999] responsável pela “discussão e desenvolvimento de propostas, para a utilização dos elementos de metadados do *Dublin Core* na descrição de recursos do domínio da educação” [Sutton e Mason, 2001]. O principal objectivo do grupo de trabalho DC-Ed baseia-se no desenvolvimento de um conjunto comum de elementos e qualificadores recomendados pela DCMI, para serem utilizados no âmbito dos materiais desenvolvidos no domínio da Educação e consequentemente promover a interoperabilidade entre projectos.

Entre várias aplicações do *Dublin Core* podem-se citar como exemplos o *Gateway to Educational Materials* (GEM)⁶⁶, *Educational Network Austrália* (EdNA)⁶⁷, a *Networked Digital Library of Theses and Dissertation* (NDLTD)⁶⁸, *Art, Design, Architecture & Media Information Gateway and Visual Arts Data Service* (ADM)⁶⁹ e *Consortium Interchange of Museum Information* (CIMI)⁷⁰.

Resumindo, as razões que justificam a selecção do DCMES como o vocabulário normalizado a ser utilizado no âmbito deste trabalho de mestrado excluindo os elementos desenvolvidos pelo grupo de trabalho DC-Ed, é justificado pelo facto destes serem mais

⁶⁴ Especificação dos limites de um intervalo de tempo.

Disponível em <http://www.dublincore.org/documents/dcmi-terms/#H3>.

⁶⁵ Regras de codificação da W3C que compreendem períodos de datas e tempos - um perfil baseado na norma ISO 8601. Disponível em <http://www.dublincore.org/documents/dcmi-terms/#H3>.

⁶⁶ <http://www.learningcommons.org/educators/library/gem.php>.

⁶⁷ <http://www.edna.edu.au/edna/page1.html>.

⁶⁸ <http://www.NDLTD.org>.

⁶⁹ <http://adam.ac.uk>

⁷⁰ <http://www.cimi.org>.

orientados para a educação e, portanto, não abrangendo os artigos científicos que são orientados aos pares, ou seja a outros cientistas.

No âmbito deste trabalho de mestrado, optou-se por seleccionar apenas elementos de metadados do vocabulário DCMES tendo em conta que estes cobriam todas as necessidades de descrição das publicações científicas, pelo facto de o DCMES ser uma norma NISO desde 2001 e uma norma ISO desde 2003, e também atendendo ao facto de ser extensamente utilizado em diversas situações.

6.2. PERFIL DE APLICAÇÃO

Um perfil de aplicação tem como objectivo definir um conjunto de elementos de metadados estruturados, que uma determinada comunidade ou organização utiliza na descrição semântica de recursos de informação, no âmbito de um determinado contexto aplicacional [Woodley, 2003]. Os elementos de metadados podem pertencer a um ou mais vocabulários de metadados, no âmbito de uma dada aplicação.

No âmbito deste projecto, foi definido um perfil de aplicação, de forma a descrever o mais pormenorizadamente possível todos os elementos retirados de diferentes vocabulários e utilizados no contexto particular desta aplicação, definindo quais os elementos usados no género dos artigos científicos publicados em revistas científicas.

O perfil de aplicação definido no âmbito deste projecto foi desenvolvido tendo em conta os recursos a serem descritos e o facto de os elementos do RSS (*title*, *link* e *description*) não serem suficientes na descrição semântica das publicações científicas. Assim, completou-se o processo de descrição com a introdução de elementos de metadados provenientes do vocabulário de metadados DCMES, já que este vocabulário de metadados contém uma grande variedade de elementos de metadados que compreendem na sua totalidade todos os requisitos necessários à descrição dos recursos a serem descritos no âmbito deste projecto.

Neste sistema é definida uma camada de metadados onde serão organizados e armazenados os elementos de metadados, separadamente do conteúdo original, contribuindo para melhorar as funcionalidades de pesquisa e de navegação sobre os artigos publicados em revistas científicas.

Para além dos elementos de metadados definidos na sintaxe do RSS, os restantes elementos de metadados utilizados são seleccionados do vocabulário normalizado DCMES, com o principal objectivo de maximizar a interoperabilidade semântica entre os metadados. Todos os elementos têm em comum o facto de poderem ser repetidos e opcionais com a excepção do elemento de metadado *link* do RSS.

O perfil de aplicação definido integra um conjunto de 16 elementos de metadados. Nas tabelas que se seguem são apresentados todos os elementos ordenados segundo as 5 categorias já apresentadas na secção 3.2, nomeadamente: Identificação do recurso; Identificação de quem tem a posse do recurso; Relevância do recurso; Classificação do recurso; Informação sobre as ligações do recurso.

Para cada elemento é apresentado o seu nome, a definição do elemento, o vocabulário a que pertence e o esquema de codificação recomendado.

Vocabulário	Elemento	Definição	Esquema de codificação
RSS	Link	Contém o URL do recurso.	
RSS	Title	Título do recurso.	
RSS	Description	Uma descrição do conteúdo do recurso.	
DCMES	Creator	Autor do recurso.	
DCMES	Language	O idioma do conteúdo do recurso.	ISO 1766 & 639
DCMES	Created	Data de criação do recurso.	W3C-DTF
DCMES	Issued	Data formal da emissão (i.e. publicação) do artigo.	W3C-DTF
DCMES	Medium	O suporte material ou físico do recurso.	IMT
DCMES	Extent	Indicar o tamanho do recurso.	
DCMES	isPartOf	O recurso descrito é uma parte física ou lógica do recurso referenciado.	

Tabela 8: Identificação do recurso

Vocabulário	Elemento	Definição	Encoding Schema(s)
DCMES	Rights	Informação sobre os direitos de acesso e de utilização do recurso.	
DCMES	Publisher	A entidade responsável por disponibilizar o recurso.	

Tabela 9: Identificação de quem tem a posse do recurso

Vocabulário	Elemento	Definição	Encoding Schema(s)
DCMES	Mediator	A classe ou uma entidade responsável por mediar o acesso ao recurso e para quem o recurso é dirigido ou útil.	

Tabela 10: Relevância do recurso

Vocabulário	Elemento	Definição	Encoding Schema(s)
DCMES	Subject	Tópico sobre o conteúdo de um recurso, especificado de acordo com o Sistema de Classificação da ACM CCS.	ACM CCS

Tabela 11: Classificação do recurso

Vocabulário	Elemento	Definição	Encoding Schema(s)
DCMES	BibliographicCitations	Referência bibliográfica para o recurso	
DCMES	References	Referências bibliográficas utilizadas pelo autor na elaboração do conteúdo do recurso científico.	

Tabela 12: Informação sobre as ligações do recurso

Após a conclusão do processo de definição da estrutura dos elementos de metadados a serem utilizados na descrição dos diferentes géneros de literatura científica que compõe o sistema a implementar, segue-se a implementação do perfil de aplicação seguindo a abordagem RDF/XML baseada no vocabulário smes⁷¹.

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:vCard="http://www.w3.org/TR/2001/vcard-rdf/3.0#"
  xmlns:rss="http://purl.org/rss/1.0/"
  xmlns:smes="http://www.schemas-forum.org/registry/SCHEMAS/1.0/smes#">
  <!-- RDF file description-->
  <smes:ApSchema rdf:about="http://www.dsi.uminho.pt/metadata/appProfiles/apsi/APSIApApplicationProfile.rdf">
    <smes:isProfileOf rdf:resource="http://www.dsi.uminho.pt/metadata/appProfiles/apsi/APSII-ap.html"/>
    <dc:title>Perfil de Aplicação dos Artigos da APSI v1.0</dc:title>
    <dc:date>2005-02-21</dc:date>
    <dc:description>Perfil de Aplicação definido no âmbito dos artigos publicados pela Associação Portuguesa
de Sistemas de Informação</dc:description>
    <dc:identifier rdf:resource="http://www.dsi.uminho.pt/apsi/schemas/APSII-ap.rdf"/>
    <dc:creator>Teresa Pereira</dc:creator>
    <dc:publisher>Universidade do Minho</dc:publisher>
    <dc:type>Perfil de Aplicação</dc:type>
    <dc:format>RDF/XML</dc:format>
    <dc:language>pt</dc:language>
    <!-- Semantics from RSS-->
    <smes:uses rdf:resource="http://purl.org/rss/1.0/title"/>
    <smes:uses rdf:resource="http://purl.org/rss/1.0/description"/>
    <smes:uses rdf:resource="http://purl.org/rss/1.0/link"/>
    <!-- Semantics from the DCMES-->
    <smes:uses rdf:resource="http://purl.org/dc/elements/1.1/creator"/>
    <smes:uses rdf:resource="http://purl.org/dc/elements/1.1/subject"/>
    <!-- O rdfs:range do dc:subject é o Sistema de Classificação ACM CCS-->
    <smes:uses rdf:resource="http://purl.org/dc/elements/1.1/publisher"/>
    <smes:uses rdf:resource="http://purl.org/dc/elements/1.1/language"/>
    <smes:uses rdf:resource="http://purl.org/dc/elements/1.1/rights"/>
```

⁷¹ O vocabulário smes foi desenvolvido no âmbito do projecto SCHEMAS.

```
<!-- Semantics from DCQ-->
<smes:uses rdf:resource="http://purl.org/dc/terms/created"/>
<smes:uses rdf:resource="http://purl.org/dc/terms/issued"/>
<smes:uses rdf:resource="http://purl.org/dc/terms/medium"/>
<smes:uses rdf:resource="http://purl.org/dc/terms/extent"/>
<smes:uses rdf:resource="http://purl.org/dc/terms/isPartOf"/>
<smes:uses rdf:resource="http://purl.org/dc/terms/mediator"/>
<smes:uses rdf:resource="http://purl.org/dc/terms/references"/>
<smes:uses rdf:resource="http://purl.org/dc/terms/bibliographicCitations"/>
</smes:ApSchema>
</rdf:RDF>
```

6.3. ESTRUTURA DOS ELEMENTOS DE METADADOS

Os elementos de metadados definidos no perfil de aplicação foram integrados no *template* da estrutura de metadados codificado segundo o formato RSS.

A criação dos documentos RSS que foram desenvolvidos no âmbito deste projecto de mestrado foi realizada através da tecnologia RSS 1.0. Esta selecção, conforme é justificada no capítulo 4, é fundamentada pelo facto de permitir a interoperabilidade com outras linguagens RDF/XML, facilitar a sua leitura e o seu processamento por outras máquinas, e permitir a extensibilidade com outros vocabulários de metadados, promovendo a descrição semanticamente rica de recursos da *Web*.

A especificação do RSS 1.0 define o RDF *Site Summary* como um formato de descrição que está associado à sindicância de conteúdos da *Web* e tem como objectivo a descrição de metadados [Begeed *et al.*, 2000]. O RSS 1.0 é uma aplicação XML que está em conformidade com a especificação RDF do W3C e é extensível através dos XML-*namespace* ou através do mecanismo de extensão modular baseados no RDF, que permite a utilização de metadados provenientes de diferentes vocabulários na criação dos *feeds* RSS [Begeed *et al.*, 2000].

A estrutura de metadados foi definida utilizando elementos de metadados que fazem parte da sintaxe do RSS (alguns dos elementos são obrigatórios), e a utilização de elementos de metadados provenientes de outros vocabulários normalizados, tendo em conta a extensibilidade de metadados que a tecnologia RSS 1.0 proporciona.

Na secção que se segue será apresentada a codificação RSS de todos os elementos de metadados utilizados na criação dos *feeds* RSS e que integram o *template* da estrutura de metadados definido no âmbito deste trabalho de mestrado, em conformidade com a especificação do RSS 1.0.

6.4. REPRESENTAÇÃO RSS DOS ELEMENTOS DE METADADOS

No âmbito deste trabalho, os elementos de metadados utilizados no processo de descrição dos conteúdos científicos da APSI, compreenderam a utilização de alguns dos elementos do RSS (*title*, *link*, e *description*) e também a utilização de elementos provenientes do vocabulário de metadados do DCMES. Como foi justificado na secção 6.2 os elementos deste vocabulário compreendiam na sua totalidade todas as necessidades associadas ao processo de descrição das publicações científicas, em particular as disponibilizadas no repositório da APSI.

Neste capítulo será apresentada a codificação RSS dos elementos de metadados utilizados na criação dos *feeds* RSS, desenvolvidos de acordo com a estrutura do *template* definido, que compreende elementos do RSS, bem como elementos provenientes de outros vocabulários de metadados, codificados de acordo com a especificação do RSS 1.0.

Passa-se, de seguida à descrição da codificação RSS da estrutura dos *feeds* RSS criados bem como os elementos que os compõem.

1. `<?xml version="1.0"?>`

Um documento RSS é uma aplicação XML, no entanto não é exigido começar com uma declaração XML. A especificação do RSS 1.0 recomenda a sua utilização como boa prática e também para assegurar a compatibilidade com a versão do RSS 0,9 (que o exige).

Sintaxe: `<?xml version="1.0"?>`

Requisito: Opcional

2. <rdf:RDF>

O nível mais externo da estrutura hierárquica de um documento RSS é constituído pelo elemento RDF. A abertura da etiqueta⁷² RDF associa o prefixo do *namespace* “rdf:” (permite identificar o *namespace* que está a ser utilizado) com a sintaxe do esquema RDF e estabelece o esquema RSS 1.0 como o *namespace* definido por defeito para o documento. A especificação do RSS 1.0 indica que qualquer prefixo de *namespace* válido pode ser usado, no entanto é recomendada a utilização normativa do prefixo “rdf:”. Todos os utilizadores que pretendam assegurar a compatibilidade com a versão RSS 0.9, a especificação do RSS 1.0 recomenda a utilização do prefixo “rdf:”.

A especificação do RSS 1.0 define as seguintes propriedades e sintaxe para este elemento:

Sintaxe: <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns="http://purl.org/rss/1.0/">

Requisito: Obrigatório exactamente como é apresentado na sintaxe, permitindo a declaração de *namespace* adicionais.

Modelo: (*channel*, *image*?⁷³, *item*+⁷⁴, *textInput*?)

Exemplo:

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns="http://purl.org/rss/1.0/"
xmlns:dcterms="http://purl.org/dc/terms">
```

3. <channel>

O elemento *channel* é um dos elementos nucleares de um documento RSS, contém elementos de metadados que descrevem o próprio *channel*, designadamente um título, uma breve descrição, e o URL do recurso que permite efectuar a ligação à origem do recurso descrito. O URL do recurso é

⁷² do inglês *tag*.

⁷³ Na descrição do modelo, o símbolo “?” significa que o elemento ou o atributo é opcional.

⁷⁴ Na descrição do modelo, o símbolo “+” significa que é permitido “uma ou mais” instâncias deste elemento ou atributo.

descrito no atributo `rdf:about` e este tem de ser único. Deste modo, um *channel* é sempre identificado por uma referência de *Uniform Resource Identifier* (URI). Normalmente, este URI corresponde ao URL da *homepage* do recurso que está a ser descrito ou então o URL onde o documento RSS é disponibilizado. No âmbito deste trabalho um recurso compreende as publicações científicas armazenadas no repositório da APSI.

A especificação do RSS 1.0 define as seguintes propriedades para este elemento:

Sintaxe: `<channel rdf:about="{recurso}">`

Requisito: Obrigatório

Atributo exigido: *rdf:about*

Modelo: (*title, link, description, image?, items, textinput?*)

Exemplo:

```
<channel rdf:about="http://repositorio.apsi.pt:8080/index.jsp">
  <title>Repositório da APSI </title>
  <link>http://repositorio.apsi.pt:8080/index.jsp</link>
  <description>Repositório institucional da Associação Portuguesa de Sistemas de
Informação</description>
  <dc:publisher>Associação Portuguesa de Sistemas de Informação (APSI) </dc:publisher>
  <dc:creator>Rui Dinis Sousa </dc:creator>
  <dc:rights>APSI. Todos os direitos reservados</dc:rights>
  <dc:date>2006-05-23T09:47:57Z</dc:date>
  <!-- Um item é um recurso (um artigo) -->
  <items>
    <rdf:Seq>
      <rdf:li rdf:resource="http://hdl.handle.net/123456789/20"/>
      <rdf:li rdf:resource="http://hdl.handle.net/123456789/56"/>
      <rdf:li rdf:resource="http://hdl.handle.net/123456789/90"/>
      <rdf:li rdf:resource="http://hdl.handle.net/123456789/97"/>
    </rdf:Seq>
  </items>
</channel>
```

De seguida são apresentados os atributos utilizados na descrição do elemento *channel*, de acordo com o exemplo apresentado acima e que compreende o *feed*

RSS criado para descrever as publicações científicas incluídas na categoria *General Literature* pertencente ao sistema de classificação da ACM CCS, utilizado no âmbito deste trabalho.

a. <title>

Este elemento descreve o título do elemento *channel*.

As propriedades deste elemento são definidas na especificação do RSS 1.0 da seguinte forma:

Sintaxe: <title>{channel_title}</title>

Requisito: Obrigatório

Modelo: (#PCDATA)

Comprimento máximo (sugerido): 40 (caracteres)

b. <link>

Este elemento descreve o URL do elemento *channel*. É o elemento *link* que permite executar a ligação do título do elemento *channel* à origem da informação que está a ser descrita.

As propriedades deste elemento são definidas na especificação do RSS 1.0 da seguinte forma:

Sintaxe: <link>{channel_link}</link>

Requisito: Obrigatório

Modelo: (#PCDATA)

Comprimento máximo (sugerido): 500

c. <description>

Este elemento contém um breve sumário ou resumo do conteúdo do elemento *channel*, designadamente funções, fonte, etc.

As propriedades deste elemento são definidas na especificação do RSS 1.0 da seguinte forma:

Sintaxe: <description>{channel_description}</description>

Requisito: Obrigatório

Modelo: (#PCDATA)

Comprimento máximo (sugerido): 500

d. **<dc:publisher>**

O elemento *dc:publisher* contém o prefixo “dc” indicando que o vocabulário a que este elemento pertence é o *Dublin Core*. Para estar sintacticamente correcto, foi declarado o *namespace* deste vocabulário como um atributo XML no elemento *rdf:RDF*. Esta declaração foi efectuada através da seguinte linha de codificação:

```
xmlns:dc="http://purl.org/dc/elements/1.1/"
```

O elemento *dc:publisher* é utilizado para descrever a entidade responsável por disponibilizar as publicações, que no âmbito deste trabalho é a Associação Portuguesa de Sistemas de Informação (APSI).

e. **<dc:creator>**

O elemento *dc:creator* também pertence ao conjunto de elementos de metadados do *Dublin Core*, isto é, o DCMES (*Dublin Core Metadata Element Set*). Este elemento é utilizado para descrever o autor que representa a página *Web* da APSI e é responsável pela informação disponibilizada no elemento *channel* que está a ser descrito.

f. **<dc:rights>**

O elemento *dc:rights* é outro elemento pertencente ao conjunto de elementos de metadados do *Dublin Core*, utilizado para descrever o elemento *channel* definido no *feed* RSS. Este elemento descreve os direitos de acesso e de utilização dos conteúdos disponibilizados no *channel* criado.

g. **<dc:date>**

O elemento *dc:date* pertence igualmente ao conjunto de elementos do *Dublin Core* e descreve a data em que o *feed* foi criado.

h. <items>

O elemento *channel* é constituído por um ou vários *items* e cada *item* compreende a descrição de uma publicação científica.

O *rdf:resource* de cada conteúdo {item_uri} deve ser igual ao *rdf:about* do respectivo item do conteúdo científico {item_uri}.

A especificação do RSS 1.0 sugere que o elemento *items* utilize na codificação RSS o *container*⁷⁵ RDF Seq (sequência) invés do RDF Bag, uma vez que o RDF Seq permite definir uma ordem dos itens dos artigos que compõem o elemento *channel*.

As propriedades deste elemento são definidas na especificação do RSS 1.0 da seguinte forma:

Sintaxe: <items><rdf:Seq><rdf:li resource="{item_uri}" /> ... </rdf:Seq></items>

Requisito: Obrigatório

Exemplo:

```
<items>
  <rdf:Seq>
    <rdf:li rdf:resource="http://hdl.handle.net/123456789/20"/>
    <rdf:li rdf:resource="http://hdl.handle.net/123456789/56"/>
    <rdf:li rdf:resource="http://hdl.handle.net/123456789/90"/>
    <rdf:li rdf:resource="http://hdl.handle.net/123456789/97"/>
  </rdf:Seq>
</items>
```

⁷⁵ Os *containers* RDF são recursos utilizados para representar coleções. A descrição do RDF indica três tipos de *containers*: Bag, Seq e Alt. O Bag é sub-classe da classe *rdfs:container* e pretende indicar ao leitor que o *container* não tem uma ordem definida [Brickley e Guha, 2004]. O Seq é sub-classe da classe *rdfs:container* e pretende indicar ao leitor que a ordem numérica do *container* tem associado um significado [Brickley e Guha, 2004]. Por fim o Alt é igualmente sub-classe da classe *rdfs:container* e pretende indicar ao leitor que o processamento típico deste *container* consiste em seleccionar um dos membros do *container* [Brickley e Guha, 2004].

Depois de definido e descrito o elemento *channel* na criação de um *feed* RSS o passo seguinte consiste na descrição de cada um dos *items* que foram definidos no elemento *channel*.

4. <item>

O elemento *item* está normalmente associado a um artigo de uma notícia ou de um *blog*, no entanto devido à extensibilidade modular que caracteriza o RSS 1.0, um *item* refere-se a qualquer objecto que seja identificado através de um URI [Beged *et al.*, 2000].

No contexto deste trabalho, um elemento *item* refere-se a um documento científico disponibilizado no repositório da APSI e que está incluído no *data set* utilizado na implementação do protótipo, desenvolvido no âmbito deste trabalho de mestrado.

A especificação do RSS 1.0 indica que um documento RSS deve conter no mínimo um elemento *item*. Apesar da especificação do RSS 1.0 não impor um limite máximo, esta recomenda o número máximo de 15 *items* por documento RSS de modo a assegurar a compatibilidade com as versões RSS 0,9 e o RSS 0,91.

A identificação do atributo {*item_uri*} é efectuada através de um URI, e este deve ser único relativamente a qualquer atributo *rdf:about* definido no documento RSS. A especificação do RSS 1.0 recomenda que o elemento {*item_uri*} deve ser igual ao valor definido no sub-elemento <*link*> do elemento <*item*>.

As propriedades deste elemento são definidas na especificação do RSS 1.0 da seguinte forma:

Sintaxe: <item rdf:about="{*item_uri*}">

Requisito: A cardinalidade deste elemento deve ser maior ou igual a um.

Recomendação para garantir a compatibilidade com as versões 0.9x: De 1 a 15

Atributo exigido: *rdf:about*

Modelo: (*title, link, description?*)

Exemplo:

```

<item rdf:about="http://hdl.handle.net/123456789/20">
  <title>Reconhecimento de Voz - Voice Car System (VCS)</title>
  <link>http://hdl.handle.net/123456789/20</link>
  <description>Os avanços tecnológicos possibilitam que os computadores reconheçam a
voz humana e a interpretem, de forma a executarem determinadas tarefas previamente
definidas. De modo a demonstrar a aplicabilidade destas novas tecnologias desenvolvemos uma
aplicação que tem por objectivo a criação de uma interface de reconhecimento de voz,
recorrendo para isso ao módulo Sensory Voice Extreme™ Toolkit. A aplicação em causa simula
o controlo de algumas funções de um automóvel, activadas através do reconhecimento da voz
humana. Numa utilização a nível real, a interacção do condutor perante a placa de
reconhecimento da voz deverá ser efectuada através de um módulo que poderá ser integrado,
por exemplo, no computador de bordo do automóvel.

  </description>
  <!--PUBLISHER-->
  <dc:publisher>APSI</dc:publisher>
  <!--AUTOR-->
  <dc:creator>
    <rdf:Seq>
      <rdf:li>HugoNeiva</rdf:li>
      <rdf:li>Paulino</rdf:li>
      <rdf:li>Bruno Silva</rdf:li>
      <rdf:li>Pedro Silva</rdf:li>
    </rdf:Seq>
  </dc:creator>
  <!--RIGHTS -->
  <dc:rights>APSI. Todos os direitos reservados</dc:rights>
  <!--SUBJECT -->
  <dc:subject rdf:datatype="http://www.acm.org/class/1998/acmccs98-1.2.3.xml"> General
Literature
  </dc:subject>
  <!--FORMATO -->
  <dcterms:extent>868596</dcterms:extent>
  <dcterms:medium>
    <dcterms:IMT>
      <rdf:value>pdf</rdf:value>
    </dcterms:IMT>
  </dcterms:medium>
  <!--IDIOMA -->

```

```
<dc:language>pt </dc:language>
<!--DATA DE CRIAÇÃO DO ARTIGO -->
<dcterms:created>
  <dcterms:W3CDTF>
    <rdf:value> 2004-11-03</rdf:value>
  </dcterms:W3CDTF>
</dcterms:created>
<!--DATA DE PUBLICAÇÃO DO ARTIGO -->
<dcterms:issued>
  <dcterms:W3CDTF>
    <rdf:value>2005-05-26T20:09:02Z </rdf:value>
  </dcterms:W3CDTF>
</dcterms:issued>
</item>
```

a. <title>

O elemento *title* utilizado dentro de um *item* pretende descrever o título do elemento *item* que está a ser descrito, ou seja o título da publicação científica que está a ser descrita.

As propriedades deste elemento são definidas na especificação do RSS 1.0 da seguinte forma:

Sintaxe: <title>{item_title}</title>

Requisito: Obrigatório

Modelo: (#PCDATA)

Comprimento máximo (sugerido): 100

b. <link>

Este elemento descreve o URL do elemento *item* que está a ser descrito, ou seja a publicação científica que está a ser descrita.

As propriedades deste elemento são definidas na especificação do RSS 1.0 da seguinte forma:

Sintaxe: <link>{item_link}</link>

Requisito: Obrigatório

Modelo: (#PCDATA)

Comprimento máximo (sugerido): 500

c. <description>

Este elemento contém uma breve descrição ou resumo do *item* que está a ser descrito, ou seja a publicação científica que está a ser descrita.

As propriedades deste elemento são definidas na especificação do RSS 1.0 da seguinte forma:

Sintaxe: <description>{item_description}</description>

Requisito: Opcional

Modelo: (#PCDATA)

Comprimento máximo (sugerido): 500

d. <dc:publisher>

O elemento *dc:publisher* incluído na descrição do elemento *item* pretende descrever a entidade que disponibiliza a publicação científica que está a ser descrita.

A representação RSS deste elemento resulta na seguinte codificação:

```
<dc:publisher>APSI</dc:publisher>
```

e. <dc:creator>

O elemento *dc:creator* é utilizado para descrever o autor ou autores da publicação científica que está a ser descrita. A representação RSS deste elemento resulta na seguinte codificação:

```
<dc:creator>  
<rdf:Seq>  
<rdf:li>HugoNeiva</rdf:li>
```

```
<rdf:li>Paulino</rdf:li>
<rdf:li>Bruno Silva</rdf:li>
<rdf:li>Pedro Silva</rdf:li>
</rdf:Seq>
</dc:creator>
```

f. **<dc:rights>**

O elemento *dc:rights* é utilizado para descrever os direitos de acesso e de utilização da publicação científica que está a ser descrita. A descrição deste elemento em RSS resulta na seguinte codificação:

```
<dc:rights>APSI. Todos os direitos reservados</dc:rights>
```

g. **<dc:subject>**

O elemento *dc:subject* pertence ao conjunto de elementos do *Dublin Core* e descreve o tópico do conteúdo da publicação, especificado de acordo com o Sistema de Classificação da ACM CCS. A representação RSS deste elemento resulta na seguinte codificação:

```
<dc:subject rdf:datatype="http://www.acm.org/class/1998/acmccs98-1.2.3.xml">
General Literature </dc:subject>
```

O elemento *dc:subject* está definido como um literal ou seja é constituído por uma string e por um datatype RDF que faz referência a um URI RDF. No exemplo apresentado (General Literature, <http://www.acm.org/class/1998/acmccs98-1.2.3.xml>) o atributo datatype com o valor “<http://www.acm.org/class/1998/acmccs98-1.2.3.xml>” deve ser interpretado como um esquema W3C XML.

h. **<dc:language>**

Este elemento descreve o idioma do conteúdo da publicação científica que está a ser descrita. A representação RSS deste elemento resulta na seguinte codificação:

```
<dc:language>pt </dc:language>
```

i. <dcterms:extent>

Este elemento descreve o tamanho da publicação científica que está a ser descrita. A representação RSS deste elemento resulta na seguinte codificação:

```
<dcterms:extent>868596</dcterms:extent>
```

j. <dcterms:medium>

O elemento *dcterms:medium* descreve o suporte material ou físico da publicação científica que está a ser descrita. Este elemento tem como qualificador de esquema de codificação o esquema IMT (*Internet Media Type*). A representação RSS deste elemento resulta na seguinte codificação:

```
<dcterms:medium>  
  <dcterms:IMT>  
    <rdf:value>pdf</rdf:value>  
  </dcterms:IMT>  
</dcterms:medium>
```

k. <dcterms:created>

O elemento *dcterms:created* descreve a data em que a publicação científica foi criada. Este elemento tem como qualificador de esquema de codificação o esquema W3C-DTF que corresponde às regras do W3C para a codificação de datas e horas – um perfil baseado na norma ISO 8601. A representação RSS deste elemento resulta na seguinte codificação:

```
<dcterms:created>  
  <dcterms:W3CDTF>  
    <rdf:value>2004-11-03</rdf:value>  
  </dcterms:W3CDTF>  
</dcterms:created>
```

l. <dcterms:issued>

O elemento *dcterms: issued* descreve a data formal da publicação do conteúdo científico. Este elemento, também utiliza o qualificador de esquema de codificação o esquema W3C-DTF, que corresponde às regras do W3C para a codificação de datas e horas – um perfil baseado no ISO 8601. A representação RSS deste elemento resulta na seguinte codificação:

```
<dcterms: issued>  
  <dcterms:W3CDTF>  
    <rdf:value> 2005-05-26T20:09:02Z</rdf:value>  
  </dcterms:W3CDTF>  
</dcterms: issued>
```

m. <dcterms:isPartOf>

O elemento *dcterms: isPartOf* indica que a publicação científica que está a ser descrita é uma parte física ou lógica da publicação científica referenciada. A representação RSS deste elemento resulta na seguinte codificação

```
<dcterms:isPartOf rdf:resource="http://....."/>
```

n. <dcterms:mediator>

O elemento *dcterms: mediator* descreve a classe ou entidade responsável por mediar o acesso à publicação científica e para quem a publicação científica é dirigida ou útil. A representação RSS deste elemento resulta na seguinte codificação:

```
<dcterms:mediator>Associação Portuguesa de Sistemas de Informação  
</dcterms:mediator>
```

o. <dcterms:bibliographicCitations>

O elemento *dcterms: bibliographicCitations* descreve a referência bibliográfica da publicação científica que está a ser descrita. A representação RSS deste elemento resulta na seguinte codificação:

```
<dcterms:BibliographiCitations>Proceedings of the International Conference on Dublin  
Core and metadata for e-communities, 2002; DC-2002: Metadata for e-Communities:  
Supporting Diversity and Convergence, Florence, Italy, 13-17 October 2002, pp 71-  
80</dcterms:BibliographiCitations>
```

p. <dcterms:references>

O elemento *dcterms: references* descreve as referências bibliográficas utilizadas pelo autor da publicação científica na elaboração do conteúdo científico produzido. A representação RSS deste elemento resulta na seguinte codificação:

```
<dcterms:references>  
  <rdf:Seq>  
    <rdf:li rdf:resource="http://....."/>  
    <rdf:li rdf:resource="http://....."/>  
  </rdf:Seq>  
</dcterms:references>
```

6.5. PROCESSO DE TRANSFORMAÇÃO DOS METADADOS

Os *feeds* RSS definidos contêm os elementos de metadados propostos no perfil de aplicação (apresentado na secção 6.2) e seguem a estrutura definida no *template* RSS.

Conforme foi apresentado no desenho do sistema (ver secção 5.4), os elementos de metadados disponíveis no *data set* utilizado, estavam descritos segundo o XML *schema* OAI-PMH e foram transformados para o formato RSS. Este processo de transformação foi efectuado através da implementação de uma *stylesheet* em XSLT (o código XSL é apresentado no apêndice D).

O desenvolvimento desta *stylesheet* passou pela análise dos elementos de metadados descritos segundo o XML *schema* OAI-PMH, e foram comparados com os elementos de metadados definidos no perfil de aplicação no âmbito desta dissertação de mestrado. Esta análise permitiu definir uma correspondência entre os elementos de metadados descritos, segundo o XML *schema* OAI-PMH e os elementos definidos no perfil de aplicação.

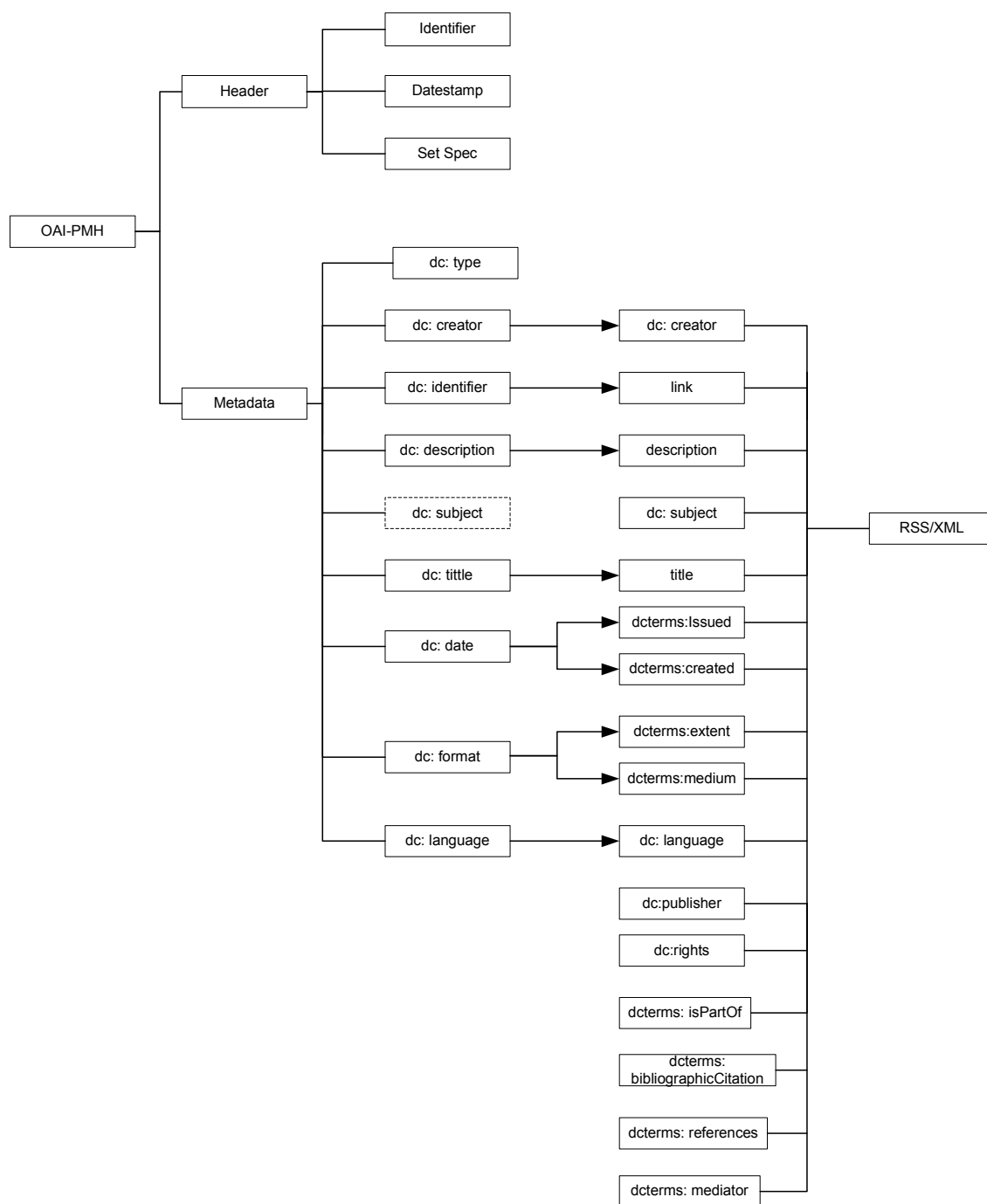


Figura 16: Correspondência dos elementos de metadados descritos segundo o XML *schema* OAI-PMH para o formato RSS de acordo com o perfil de aplicação definido

Como se pode verificar pela análise da figura 19, a estrutura de metadados definida no contexto desta dissertação de mestrado compreende mais elementos de metadados (identificados na secção 3.2) do que aqueles utilizados na *data set* descritos segundo o XML *schema* OAI-PMH.

Por outro lado a correspondência entre os elementos de metadados definidos segundo o XML *schema* do OAI-PMH e os elementos definidos no perfil de aplicação foi implementada apenas entre alguns elementos. De facto, houve elementos de metadados utilizados no *data set* que não foram instanciados ao sistema desenvolvido no âmbito desta dissertação de mestrado, como é o caso do elemento *dc:subject*. Com efeito, não foi estabelecida uma correspondência entre o elemento *dc:subject* utilizado no *data set* e o *dc:subject* definido no perfil de aplicação, tendo em conta que no âmbito desta dissertação de mestrado foi utilizada uma estrutura hierárquica de conceitos e portanto optou-se por indexar o elemento *dc:subject* ao sistema de classificação da ACM CCS.

De modo a demonstrar o processo de transformação efectuado entre os dois formatos é de seguida apresentado um exemplo de um recurso descrito segundo o XML *schema* OAI-PMH e o resultado da sua transformação para o formato RSS.

```

<record>
  <header>
    <identifier>oai:repositorio.apsi.pt:2287/20</identifier>
    <timestamp>2005-06-21T01:14:53Z</timestamp>
    <setSpec>hdl_2287_10</setSpec>
  </header>
  <metadata>
    <oai_dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
      xmlns:dc="http://purl.org/dc/elements/1.1/"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
        http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
      <dc:creator>Neiva, Hugo</dc:creator>
      <dc:creator>Paulino, Juvenal</dc:creator>
      <dc:creator>Santos, João</dc:creator>
      <dc:creator>Silva, Bruno</dc:creator>
      <dc:creator>Silva, Pedro</dc:creator>
      <dc:date>2005-05-26T20:09:02Z</dc:date>
      <dc:date>2005-05-26T20:09:02Z</dc:date>
      <dc:date>2004-11-03</dc:date>
      <dc:identifier>http://hdl.handle.net/123456789/20</dc:identifier>
      <dc:description>Os avanços tecnológicos possibilitam que os computadores reconheçam a voz humana e a interpretem, de forma a executarem determinadas tarefas previamente definidas. De modo a demonstrar a aplicabilidade destas novas tecnologias desenvolvemos uma aplicação que tem por objectivo a criação de uma interface de reconhecimento de voz, recorrendo para isso ao módulo Sensory Voice Extreme™ Toolkit. A aplicação em causa simula o controlo de algumas funções de um automóvel, activadas através do reconhecimento da voz humana. Numa utilização a nível real, a interacção do condutor perante a placa de reconhecimento da voz deverá ser efectuada através de um módulo que poderá ser integrado, por exemplo, no computador de bordo do automóvel.</dc:description>
      <dc:format>868596 bytes</dc:format>
    </oai_dc:dc>
  </metadata>
</record>

```

```

    <dc:format>application/pdf</dc:format>
    <dc:language>pt_PT</dc:language>
    <dc:subject>reconhecimento</dc:subject>
    <dc:subject>processamento</dc:subject>
    <dc:subject>controlo</dc:subject>
    <dc:subject>fala</dc:subject>
    <dc:title>Reconhecimento de Voz - Voice Car System (VCS)</dc:title>
    <dc:type>TEXTO</dc:type>
  </oai_dc:dc>
</metadata>
</record>

```

A tradução do exemplo apresentado acima segundo o XML *schema* do OAI-PMH resulta na seguinte representação em RSS:

```

<item>
  <title>Reconhecimento de Voz - Voice Car System (VCS)</title>
  <link>http://hdl.handle.net/123456789/20</link>
  <description>Os avanços tecnológicos possibilitam que os computadores reconheçam a voz humana e a interpretem, de forma a executarem determinadas tarefas previamente definidas. De modo a demonstrar a aplicabilidade destas novas tecnologias desenvolvemos uma aplicação que tem por objectivo a criação de uma interface de reconhecimento de voz, recorrendo para isso ao módulo Sensory Voice Extreme&#8482; Toolkit. A aplicação em causa simula o controlo de algumas funções de um automóvel, activadas através do reconhecimento da voz humana. Numa utilização a nível real, a interacção do condutor perante a placa de reconhecimento da voz deverá ser efectuada através de um módulo que poderá ser integrado, por exemplo, no computador de bordo do automóvel.</description>
  <dc:publisher>APSI</dc:publisher>
  <dc:creator>
    <rdf:Seq>
      <rdf:li>Neiva, Hugo</rdf:li>
      <rdf:li>Paulino, Juvenal</rdf:li>
      <rdf:li>Santos, João</rdf:li>
      <rdf:li>Silva, Bruno</rdf:li>
      <rdf:li>Silva, Pedro</rdf:li>
    </rdf:Seq>
  </dc:creator>
  <dc:rights>APSI. Todos os direitos reservados</dc:rights>
  <dc:subject rdf:datatype="http://www.acm.org/class/1998/acmccs98-1.2.3.xml"/>
  <dcterms:extent>868596</dcterms:extent>
  <dcterms:medium>
    <dcterms:IMT>
      <rdf:value>pdf</rdf:value>
    </dcterms:IMT>
  </dcterms:medium>
  <dc:language>pt </dc:language>
  <dcterms:created>
    <dcterms:W3CDTF>
      <rdf:value>2004-11-03</rdf:value>
    </dcterms:W3CDTF>
  </dcterms:created>

```



```
</dcterms:W3CDTF>  
</dcterms:created>  
<dcterms:issued>  
  <dcterms:W3CDTF>  
    <rdf:value>2005-05-26T20:09:02Z</rdf:value>  
  </dcterms:W3CDTF>  
</dcterms:issued>  
</item>
```

6.6. MECANISMOS DE PESQUISA

As aplicações de metadados estão normalmente associadas aos mecanismos de pesquisa e de navegação. Quando uma aplicação é desenhada para executar pesquisas, normalmente são criados catálogos de informação sobre os quais são efectuadas pesquisas ou então são indexados ao processo de pesquisa [Baptista, 2004]. A criação e utilização destes catálogos facilitam a descoberta de recursos sobre a *Web*.

Por outro lado, quando a aplicação é implementada com o objectivo de promover a navegação, normalmente tem associada uma rede de conceitos, baseada em conhecimento organizado e estruturado. Nesta abordagem o principal objectivo prende-se com a possibilidade de navegar sobre uma estrutura de conceitos ligadas aos recursos. Estas ligações contêm significados específicos [Baptista, 2004].

No âmbito desta dissertação de mestrado foi utilizada uma estrutura hierárquica de conceitos e utilizado um *thesaurus* léxico que estão directamente ligados à descrição dos *feeds* RSS criados. Assim, ao protótipo implementado foi adicionado o mecanismo de navegação semântica sobre uma árvore de conceitos, representada pela estrutura hierárquica do sistema de classificação da ACM CCS e o mecanismo de expansão da *query* através da utilização de uma versão baseada em RDF do *WordNet*.

O *feed* RSS gerado pela *stylesheet*, foi armazenado numa base de metadados nativa RDF, permitindo aos utilizadores efectuar a pesquisa por palavra-chave ou através da navegação sobre a árvore de conceitos representada pela estrutura hierárquica do sistema de classificação da ACM CCS. A versão RDFS da estrutura hierárquica do sistema de classificação da ACM CCS, disponibilizada pela ACM foi utilizada na implementação deste protótipo, com objectivo adaptar o protótipo com novas funcionalidades, disponibilizando ao utilizador mecanismos inteligentes de navegação semântica sobre a árvore de conceitos, para além dos

tradicionais métodos de pesquisa por palavra-chave. O facto de se utilizar linguagem controlada no processo de descrição de recursos permite eliminar algumas das ambiguidades associadas à linguagem natural, simplificando desta forma a pesquisa de conteúdos científicos no sistema.

A ligação aos elementos representados na estrutura hierárquica do sistema de classificação da ACM CCS é realizada através do elemento de metadados “*dc:subject*”. Na definição do perfil de aplicação, é indicado que o “*rdfs:range*” do elemento de metadados “*dc:subject*” são os elementos do ACM CCS (ver secção 6.2). Isto significa que, para cada descrição armazenada na base de metadados, o elemento de metadado “*dc:subject*” apenas pode conter valores do sistema de classificação da ACM CCS.

De facto, a utilização do sistema de classificação da ACM CCS permite disponibilizar ao utilizador mecanismos inteligentes de navegação semântica sobre uma estrutura hierárquica de conceitos, para além dos tradicionais métodos de pesquisa por palavra-chave.

Com o objectivo de acrescentar valor aos mecanismos de navegação e pesquisa, nomeadamente a expansão da *query*, foi utilizada outra ferramenta com propriedades orientadas para a organização da informação, designadamente o *WordNet*.

A versão 1.6 do *WordNet* codificada em RDF (ver secção 2.5) foi incluída na base de metadados e a sua conexão aos artigos da base de metadados é efectuada através do elemento *description*.

De seguida é apresentada na figura 20 um *print screen* do protótipo desenvolvido.

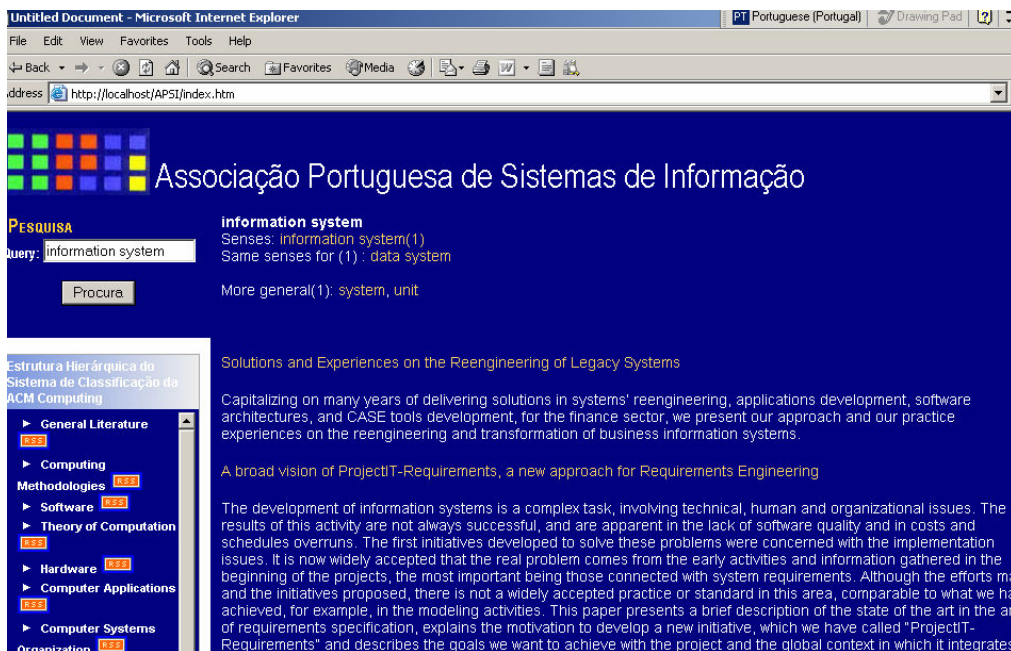


Figura 17: *Print Screen* da interface do protótipo desenvolvido

6.6.1. SUBSCRIÇÃO DOS FEEDS RSS

O facto de os conteúdos científicos da APSI serem disponibilizados sob a forma de *feeds* RSS, distribuídos pelos diferentes tópicos que compreende o sistema de classificação da ACM CCS, permite aos utilizadores efectuar a sua respectiva subscrição, facilitando-lhe por um lado o processo de consulta de publicações de uma área específica e por outro lado manterem-se actualizados acerca de novos conteúdos que vão sendo disponibilizados numa determinada área.

Na estrutura hierárquica da ACM CCS apresentada no painel esquerdo da *interface* ilustrada na figura 20, os termos apresentados no primeiro nível contêm um ícone laranja “RSS”, o que significa que a respectiva categoria representada na estrutura hierárquica do sistema de classificação da ACM CCS disponibiliza um *feed* RSS. Ou seja, este ícone indica ao utilizador que é disponibilizado um *feed* RSS e que este pode proceder à sua exclusiva subscrição. O utilizador pode subscrever apenas os *feeds* do seu interesse ou subscrever todos. Ao proceder à subscrição dos *feeds*, o utilizador através de um leitor RSS consulta as publicações disponibilizadas e toma conhecimento acerca de novos conteúdos que possam eventualmente ter sido publicados pela APSI, sem ter necessidade de aceder à interface do protótipo.

Com efeito a utilização do RSS na descrição de publicações científicas foi um projecto ambicioso, já que a tecnologia RSS está tendencialmente associada às notícias de jornais disponibilizadas na *Web* e aos *blogs*, no entanto a revisão de literatura efectuada permitiu verificar que algumas prestigiadas revistas científicas e algumas comunidades científicas, conforme foi apresentado no capítulo 4, já começam a disponibilizar os seus conteúdos sob a forma de *feed* RSS.

6.7. INTEGRAÇÃO DOS DADOS

O protótipo desenvolvido no âmbito desta dissertação de mestrado foi implementado utilizando a ferramenta comercial RDF *Gateway*. Esta é uma ferramenta que conjuga os poderes do servidor HTTP com o sistema de gestão de base de dados nativa RDF.

O RDF *Gateway* é simultaneamente um cliente e um servidor *Web* constituído por uma base de dados nativa RDF para administrar a informação. O acesso aos dados é feito via HTTP.

O RDF Gateway trabalha directamente sobre os triplos RDF, pelo que foi necessário proceder à extracção dos triplos dos feeds RSS criados, bem como da versão RDFS do sistema de classificação da ACM CCS e do *WordNet* seguido do seu armazenamento em três tabelas da base de dados. A sua manipulação é executada sobre as tabelas de triplos através da utilização do RDF Query Analyzer (RDFQA). O código RSP (com RDFQL embebido) definido no âmbito deste protótipo apresenta-se no anexo E.

6.8. RESUMO

Neste capítulo foram descritas as várias tarefas que estiveram envolvidas na implementação do protótipo desenvolvido no âmbito desta dissertação de mestrado, de acordo com a descrição geral do trabalho realizado.

Ao nível dos dados, estiveram envolvidas diversas tarefas, desde o levantamento e análise de diversos vocabulários normalizados de metadados e extensamente utilizados no domínio da literatura científica, selecção dos elementos de metadados adequados à descrição dos artigos científicos, seguido da definição do perfil de aplicação e do template RSS. Os elementos de metadados utilizados no processo de descrição dos artigos científicos foram os do DCMES, para além dos elementos do RSS. Os mecanismos de pesquisa e navegação implementados bem como a sindicância de conteúdos são efectuados sobre a camada de metadados definida.

A implementação dos processos de pesquisa e de navegação foram desenvolvidos no protótipo utilizando a versão RDFS do *WordNet* e do sistema de classificação da ACM CCS, de modo a permitir a pesquisa e navegação sobre a metainformação dos artigos científicos da APSI e não sobre o texto integral dos mesmos.

7. CONCLUSÕES

Neste capítulo descrevem-se as conclusões do trabalho realizado, bem como os resultados obtidos e as contribuições fundamentais daí decorrentes. Por fim apresentam-se algumas sugestões sobre o trabalho futuro a desenvolver.

Neste trabalho de mestrado propuseram-se dois objectivos: (1) Instanciar o sistema desenvolvido no projecto OmniPaper, no contexto das publicações científicas; (2) Proceder à sindicância de conteúdos utilizando a base de metadados.

Na implementação do protótipo o *data set* utilizado compreendeu a colecção de publicações científicas existentes no repositório da APSI.

Assim, tendo em conta que as notícias publicadas em jornais (utilizadas no sistema OmniPaper) e as publicações científicas são dois tipos de recursos distintos, a revisão de literatura realizada incidiu sobre a análise de um conjunto normalizado de vocabulários de metadados e extensamente utilizados no domínio da literatura científica. Esta análise permitiu seleccionar um conjunto de elementos de metadados adequados à descrição semântica de artigos científicos. O resultado desta tarefa deu origem à definição do perfil de aplicação e ao *template* RSS. De facto os elementos de metadados utilizados na descrição das notícias publicadas em jornais são distintos dos elementos utilizados na descrição das publicações científicas, originando estruturas de metadados distintas.

No âmbito desta dissertação de mestrado os elementos de metadados utilizados no processo de descrição dos artigos científicos foram os do DCMES, para além dos elementos do RSS.

De seguida, tendo em conta que a metainformação disponibilizada pelo *data set* utilizado, estava descrita segundo o XML *schema* OAI-PMH procedeu-se à sua transformação, para a estrutura RSS (*feed* RSS) anteriormente definida, através da utilização de uma *stylesheet* em XSLT. Este *feed* é constituído por um conjunto de *items* que compreendem a metainformação sobre todos os artigos científicos disponibilizados pela APSI, seguido do seu armazenamento numa base de metadados.

Os processos de pesquisa e navegação semântica desenvolvidos no protótipo bem como a sindicância de conteúdos são efectuados sobre esta camada de metadados.

Os mecanismos de navegação e pesquisa foram instanciados do sistema OmniPaper, mas o processo de navegação implementado no sistema desenvolvido nesta dissertação de

mestrado foi alterado. De facto, no sistema OmniPaper o processo de navegação era efectuado sobre a estrutura hierárquica de conceitos representada, pelos IPTC-SC, específicos para a descrição editorial do conteúdo das notícias e portanto a sua utilização não era a mais adequada para ser utilizada no contexto das publicações científicas. Deste modo, foi utilizado o sistema de classificação da ACM (o ACM CCS) ao invés dos IPTC-SC. O processo de pesquisa com a expansão de *queries* através da utilização do *WordNet* foi instanciado na íntegra a partir do sistema OmniPaper.

O desenvolvimento destas tarefas resultou num protótipo que é uma instância do sistema desenvolvido no projecto OmniPaper (excepto para algumas especificidades da área), no contexto da publicação científica e portanto é atingido o primeiro objectivo apresentado na Introdução desta dissertação.

Relativamente ao segundo objectivo este foi alcançado, uma vez que o sistema implementado agrega os metadados do repositório da APSI (embora o sistema esteja preparado para agregar metadados de outros repositórios), expressos segundo o formato RSS, que compreende o *feed* RSS, permitindo ao utilizador efectuar a sua subscrição e proceder à sindicância dos conteúdos científicos. Deste modo, o utilizador desenvolve a consulta sobre a metainformação dos conteúdos científicos agregados, ao invés de proceder à consulta dos conteúdos directamente no repositório da APSI.

Um contributo deste trabalho de mestrado prende-se com a oportunidade de tirar partido da camada central de metadados no sentido de estruturar e indexar a informação e consequentemente suportar as funcionalidades de pesquisa, e a navegação semântica de conteúdos científicos. E, por outro lado, efectuar a sindicância dos metadados de vários repositórios, e deste modo produzir melhorias no processo de pesquisa e distribuição dos conteúdos científicos provenientes de várias fontes de informação. Dado que o sistema OmniPaper utilizava o RDF/XML na sua base de metadados e que aqui se pretendia realizar a sindicância de conteúdos em RSS, faz todo o sentido em que não haja duplicação de esforços e que a base de metadados esteja já, nativamente, em RSS (cuja versão 1.0 é um formato RDF).

Assim, este trabalho de mestrado pretendeu dar um pequeno contributo para a melhoria dos mecanismos de pesquisa e distribuição do conhecimento científico produzido no seio das comunidades científicas.

7.1. TRABALHO FUTURO

A proposta de trabalho futuro relaciona-se com a implementação do processo de recolha dos metadados de vários repositórios para a base de metadados aqui implementada, de modo a que este sistema fique interoperável com os repositórios que implementam o protocolo OAI-PMH e deste modo disponibilizar um serviço mais completo e com mais informação.

Outros trabalhos poderão ter a ver com a manipulação de *tags* criadas socialmente para a descrição dos recursos e as suas relações tanto com os mecanismos de pesquisa e expansão de *queries* como com os mecanismos de navegação.

BIBLIOGRAFIA

- [APSI, 2005] Dinis, R. (editor), (2005). *Associação Portuguesa de Sistemas de Informação* [on-line].
Disponível em: <http://www.apsi.pt/>.
[Acedido em 2005].
- [Ayers, 2003] Ayers, D., (2003). *Extending RSS* [on-line]. XML.com.
Disponível em: <http://www.xml.com/pub/a/2003/07/23/extendingrss.html>.
[Acedido em Maio de 2006].
- [Baptista, 2002] Baptista, A. A., (2002). *Informática online: um enquadramento para a publicação em linha de revistas científicas electrónicas*. Tese de Doutoramento. Universidade do Minho.
Disponível em: <https://repositorium.sdum.uminho.pt/handle/1822/284>.
- [Baptista, 2003a] Baptista, A. A., (2003). *Prototype Design RDF – Resource Description Framework Prototype*.
Relatório técnico: Projecto OmniPaper. 08 Abril de 2003.
- [Baptista, 2003b] Baptista, A. A., (2003). *The metadata base on the WP3 OmniPaper prototype*. *Relatório técnico: Projecto OmniPaper*. 11 Junho de 2003.
- [Baptista, 2004] Baptista, A. A., (2004). Searching and browsing using RDF-Encoded Metadata: the case of OmniPaper. *Canadian Journal of Communication*, 29 (3), 317-328. Disponível em: <https://repositorium.sdum.uminho.pt/handle/1822/5080>.
- [Beged *et al.*, 2000] Beged-Dov, G., Brickley, D., Dornfest, R., Davis, I., Dodds, L., Eisenzopf, J., Galbraith, D., Guha, R. V., MacLeod, K., Miller, E., Swartz, A. e Vlist, E. (2000). *RDF Site Summary (RSS) 1.0* [on-line].
Disponível em: <http://web.resource.org/rss/1.0/spec#>.
[Acedido em Janeiro de 2006].
- [Berners-Lee, 1990] Berners-Lee, T., (1990). *Information Management: A Proposal* [on-line].
Disponível em: <http://www.w3.org/History/1989/proposal.html>.
[Acedido em Maio de 2006].
- [Bray *et al.*, 1999] Bray, T., Hollander, D., Layman, A., Tobin, R., (1999). *Namespace in XML 1.0 (Second Edition)* [on-line]. W3C.
Disponível em: <http://www.w3.org/TR/REC-xml-names/>.
[Acedido em Maio de 2005].

- [Brickley e Guha, 2000] Brickley, D., e Guha, R., (2000). *Resource Description Framework (RDF) Schema Specification 1.0 - W3C Candidate Recommendation 27 March 2000* [on-line]. W3C.
Disponível em: <http://www.w3.org/TR/rdf-schema/>.
[Acedido em 2003].
- [Brickley e Guha, 2004] Brickley, D., e Guha, R., (2004). *Resource Description Framework (RDF) Vocabulary Language 1.0: RDF Schema - W3C Recommendation 10 February 2004* [on-line]. W3C.
Disponível em: <http://www.w3.org/TR/rdf-schema/>.
[Acedido em 2005].
- [Çelikbas, 2005] Çelikbas, Z., 2004. What is RSS and how can it serve libraries. In: Yalvaç, M. e Gülseçen, S., ed. *Proceedings First International Conference on Innovations in Learning for the Future: e-Learning*, pp. 277-292, İstanbul, Turkey.
Disponível em: <http://eprints.rclis.org/archive/00002531/>.
[Acedido em Fevereiro de 2006].
- [DCMES, 2003] Dublin Core Metadata Element Set (2003). *Dublin Core Metadata Element Set, Version 1.1: Reference Description* [on-line]. Dublin Core Metadata Initiative.
Disponível em: <http://www.dublincore.org/documents/dces/>.
[Acedido em Janeiro de 2003].
- [DCMI, 2002] Dublin Core Metadata Initiative (2002). *Dublin Core Metadata Initiative Home Page* [on-line].
Disponível em: <http://www.dublincore.org/>.
[Acedido em Janeiro de 2002].
- [DCQ, 2002] Dublin Core Qualifiers (2002). *Dublin Core Qualifiers* [on-line]. Dublin Core Metadata Initiative.
Disponível em: <http://dublincore.org/documents/dcmes-qualifiers/>.
[Acedido em Janeiro de 2002].
- [Dias, 1999] Dias, C. A., (1999). *Comunicação Científica* [on-line].
Disponível em: <http://www.geocities.com/claudiaad/comunica.pdf>.
[Acedido em Janeiro de 2006].
- [FAIR Synthesis: Glossary, 2006] *FAIR Synthesis: Glossary* [on-line].
Disponível em:
http://www.jisc.ac.uk/whatwedo/programmes/programme_fair/fair_synthesisintro/fairsynthesis_glossary.aspx.
[Acedido em Novembro de 2006].
- [Hammond, 2003] Hammond, T., (2003). *Why Choose RSS 1.0?* [on-line]. XML.com.
Disponível em: <http://www.xml.com/pub/a/2003/07/23/rssone.html>.
[Acedido em Maio de 2006].

- [Hammond *et al.*, 2004] Hammond, T., Hannay, T. e Lund, B., (2004). The Role of RSS in Science Publishing Syndication and Annotation on the Web. *D-Lib Magazine* [on-line], 10 (12).
Disponível em: <http://www.dlib.org/dlib/december04/hammond/12hammond.html>.
[Acedido em Abril de 2006].
- [HarvardLaw, 2006] Technology at Harvard Law. *RSS 2.0 Specification* [on-line].
Disponível em: <http://blogs.law.harvard.edu/tech/rss>.
[Acedido em Maio de 2006].
- [Heery *et al.*, 2000] Heery, R., Patel, M., (2000). Application Profiles: Mixing and Matching Metadata Schemas. *Ariadne* [on-line], (25).
Disponível em: <http://www.ariadne.ac.uk/issue25/app-profiles/>.
[Acedido em Maio de 2002].
- [Hillmann e Sutton, 1999] Hillmann, D. e Sutton, S., (1999). *DCMI Education Working Group* [on-line]. Dublin Core Metadata Initiative.
Disponível em: <http://dublincore.org/groups/education/index.shtml>.
[Acedido em Setembro de 2005].
- [IPTC, 2003] International Press Telecommunications Council (2003). *International Press Telecommunications Council* [on-line].
Disponível em: <http://www.iptc.org/pages/index.php>.
[Acedido em Abril de 2004].
- [IPTC-SC, 2003] IPTC Subject Codes (2003). *IPTC Subject Codes* [on-line]. International Press Telecommunications Council.
Disponível em: <http://xml.coverpages.org/NITF30-subject-codes.html>.
[Acedido em Abril de 2004].
- [Johnston, 2005] Johnston, P., (2005). *Element Refinement in Dublin Core Metadata* [on-line]. Dublin Core Metadata Initiative.
Disponível em: <http://dublincore.org/documents/dc-elem-refine/>.
[Acedido em 2005].
- [Kokkink e Schwänzl, 2002] Kokkink, S. e Schwänzl, R., (2002). *Expressing Qualified Dublin Core in RDF/XML* [on-line]. Dublin Core Metadata Initiative.
Disponível em: <http://www.dublincore.org/documents/2002/04/14/dcq-rdf-xml/>.
[Acedido em 2003].

- [Lagoze, 2001] Lagoze, C., (2001). Keeping Dublin Core Simple. *D-Lib Magazine* [on-line], 7(1). Disponível em: <http://www.dlib.org/dlib/january01/lagoze/01lagoze.html>.
[Acedido em 2003].
- [Lagoze, Van de Sompel *et al.*, 2002] Lagoze, C., Van de Sompel, H., Nelson, M. e Warner, S., (2002). *The Open Archives Initiative Protocol for Metadata Harvesting, Protocol Version 2.0* [on-line]. Open Archives Initiative. Disponível em: <http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>.
[Acedido em 2003].
- [Lassila e Swick, 1999] Lassila, O., Swick, R., (1999). *Resource Description Framework (RDF) Model and Syntax Specification*. Fevereiro 1999. W3C. Disponível em: <http://www.w3.org/TR/REC-rdf-syntax/>.
[Acedido em 2003].
- [Liddy, 2005] Liddy, E., (2005). Metadata: A Promising Solution [on-line]. *EDUCAUSE review*, Maio/Junho de 2005. Disponível em: <http://www.educause.edu/ir/library/pdf/ERM0536.pdf>.
[Acedido em 2005].
- [LusoDSpace, 2005]. Luso DSpace (2005). *Portal sobre o DSpace para a comunidade de língua portuguesa (Glossário de termos)* [on-line]. Disponível em: <http://lusodspace.sdum.uminho.pt:8080/pt/glossary.jsp>.
[Acedido em 2005].
- [McKiernan, 2005a] McKiernan, G., (2005). *eFeeds(sm): Web Feeds from Electronic Journals* [on-line]. Disponível em: <http://www.public.iastate.edu/~CYBERSTACKS/eFeeds.htm>.
[Acedido em 2005].
- [McKiernan, 2005b] McKiernan, G., (2005). *RSS(sm): Rich Site Services* [on-line]. Disponível em: <http://www.public.iastate.edu/~CYBERSTACKS/RSS.htm>.
[Acedido em 2005].
- [MITLibraries, 2005] MITLibraries (2005). *Metadata Reference Guide* [on-line]. Disponível em: <http://libraries.mit.edu/guides/subjects/metadata/standards.html>.
[Acedido em 2005].
- [Moreira, 2005] Moreira, W., (2005). Os colégios virtuais e a nova configuração da comunicação científica. *IBICT Ciência da Informação* [on-line], 34 (1). Disponível em: <http://www.ibict.br/cienciadainformacao/viewarticle.php?id=699&layout=abstract>.
[Acedido em 2006].

- [Oliveira *et al.*, 2005] Oliveira, É. B. e Noronha, D., (2005). A comunicação científica e o meio digital. *Informação & Sociedade: Estudos* [on-line], 15 (1).
Disponível em: <http://www.informacoesociedade.ufpb.br/ojs2/index.php/ies/article/viewFile/53/51>.
[Acedido em 2006].
- [OAI, 2002] Open Archives Initiative (2002). Open Archives Initiative [on-line].
Disponível em: <http://www.openarchives.org/>.
[Acedido em 2005].
- [Paepen, 2005] Paepen, B., (2005). *Blueprint: a universal standard model for efficient information retrieval. Relatório técnico: Projecto OmniPaper*. 28 Fevereiro de 2005.
- [Pereira e Baptista, 2004] Pereira, T. e Baptista, A. A., (2004). Incorporating a Semantically Enriched Navigation Layer Onto an RDF Metadatabase. In: Engelen, J., Costa Sely., M. S., Moreira, Ana Cristina S., ed. *Building digital bridges: linking cultures, commerce and science : Proceedings of the ICCC International Conference on Electronic Publishing, ELPUB*, Julho 2004 Brasília, Brasil.
Disponível em: <https://repositorium.sdum.uminho.pt/handle/1822/604>.
- [Pilgrim, 2002] Pilgrim, M., (2002). *What is RSS* [on-line]. XML.com.
Disponível em: <http://xml.com/pub/a/2002/12/18/dive-into-xml.html>.
[Acedido em 2005].
- [Powers, 2003] Powers, S., 2003. *Practical RDF*. Sebastopol: O'Reilly & Associates.
- [Rodrigues *et al.*, 2004] Rodrigues, E., Almeida, M., Miranda, A., Guimarães, A., Castro, D., (2004). RepositórioUM : criação e desenvolvimento do Repositório Institucional da Universidade do Minho. In: CONGRESSO NACIONAL DE BIBLIOTECÁRIOS, ARQUIVISTAS E DOCUMENTALISTAS, ed. "Nas encruzilhadas da informação e da cultura : (re)inventar a profissão" : actas. 2004 Estoril, Lisboa.
Disponível em: <https://repositorium.sdum.uminho.pt/handle/1822/422>.
- [Sabbatini, 1999] Sabbatini, M., (1999). *As publicações electrónicas dentro da comunicação científica* [on-line].
Disponível em: <http://bocc.ubi.pt/pag/sabbatini-marcelo-publicacoes-electronicas.html>.
[Acedido em 2006].
- [Sutton e Mason, 2001] Sutton, S. A. e Mason, J., (2001). The Dublin Core and Metadata for Educational Resources. In: Oyama, K. e Gotoda, H., ed. *DC2001: Proceedings of the International Conference on Dublin Core and Metadata Applications, 24-26 Outubro 2001 Tokyo, Japan, 25-31*.
- [UKOLN, 1999] UKOLN (1999). *DESIRE Metadata Registry* [on-line].
Disponível em: <http://desire.ukoln.ac.uk/registry/>.
[Acedido em 2004].

[USIA, 1998] USIA Economic Perspectives (1998). *Glossário de Termos de Propriedade Intelectual. USIA* [on-line], 3 (3). Maio de 1998.

Disponível em: <http://usinfo.state.gov/journals/ites/0598/ijep/ie059811.htm>.

[Acedido em Outubro de 2006].

[Wikipedia, 2006a] Wikipedia (2006). *Data Mining* [on-line]. Wikipedia.

Disponível em: http://pt.wikipedia.org/wiki/Data_mining.

[Acedido em 2006].

[Wikipedia, 2006b] Wikipedia (2006). *Feed* [on-line]. Wikipedia.

Disponível em: <http://pt.wikipedia.org/wiki/Feed>.

[Acedido em 2005].

[Wikipedia, 2006c] Wikipedia (2006). *Podcasting* [on-line]. Wikipedia.

Disponível em: <http://pt.wikipedia.org/wiki/Podcasting>.

[Acedido em 2006].

[Wikipedia, 2006d] Wikipedia (2006). *RSS* [on-line]. Wikipedia

Disponível em: <http://pt.wikipedia.org/wiki/RSS>.

[Acedido em 2006].

[Wittenbrink, 2005] Wittenbrink, H., (2005). *RSS and ATOM Understanding and Implementing Content Feeds and Syndication* [on-line]. PACKT Publishing.

Disponível em: http://www.packtpub.com/files/RSS_and_Atom_Book_Chapter1_what_are_newsfeeds.pdf.

[Acedido em Maio de 2006].

[Woodley, 2003] Woodley, M. S., (2003). *DCMI Glossary* [on-line]. Dublin Core Metadata Initiative. Disponível em: <http://www.dublincore.org/documents/usageguide/glossary.shtml#M>.

[Acedido em 2003].

[WordNet, 2003] WordNet (2003). *WordNet, a lexical database for the English language* [on-line]. Princeton University.

Disponível em: <http://wordnet.princeton.edu/>.

[Acedido em 2004].

[Yaginuma, 2002] Yaginuma, T., (2002). *Comparison Criteria for the Local Knowledge Layer Prototypes. Relatório técnico: Projecto OmniPaper*. 28 Novembro de 2002.

[Yaginuma et al., 2004] Yaginuma, T., Pereira, T., Ariza, C. e Baptista, A. A., (2004). Implementation of Metadata for OmniPaper RDF Prototype. In: Sugimoto, S., ed. *International Symposium on Digital Libraries and*

Knowledge Communities in Networked Information Society DLKC'04, 2-5 Março 2004, Tsukuba, Ibaraki, Japan.

Disponível em: <https://repositorium.sdum.uminho.pt/handle/1822/606>.

[Ziman, 1984] Ziman, J., 1984. *An Introduction to Science Studies: the Philosophical and Social Aspects of Science and Techonology*. Cambridge: Cambridge University Press.

APÊNDICES

Apêndice A.....	124
Apêndice B.....	127
Apêndice C.....	133
Apêndice D.....	144
Apêndice E.....	147
Apêndice F.....	156

APÊNDICE A

Neste apêndice apresenta-se o template do *feed* RSS definido, para ser utilizado na descrição semântica dos artigos científicos publicados pela APSI.

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- edited with XML Spy v4.2 U (http://www.xmlspy.com) by Teresa Bernardino (IPVC) -->
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:dcterms="http://purl.org/dc/terms/"
xmlns="http://purl.org/rss/1.0/">
  <channel rdf:about="http://localhost/apsi/rss_artigos.rdf">
    <title>Repositório da APSI </title>
    <link>http://repositorio.apsi.pt:8080/index.jsp</link>
    <description>Repositório institucional da Associação Portuguesa de Sistemas de Informação</description>
    <dc:publisher>Associação Portuguesa de Sistemas de Informação (APSI) </dc:publisher>
    <dc:creator>Rui Dinis de Sousa </dc:creator>
    <dc:rights>APSI. Todos os direitos reservados</dc:rights>
    <dc:date>2006-05-23T09:47:57Z</dc:date>
    <!-- Um item é um recurso (um artigo) -->
    <items>
      <rdf:Seq>
        <rdf:li rdf:resource="http://"/>
        <rdf:li rdf:resource="http://"/>
        <rdf:li rdf:resource="http://"/>
      </rdf:Seq>
    </items>
  </channel>
  <!-- Descrição dos respectivos items que constituem este canal -->
  <item rdf:about="http://">
    <title>Introdução do título do item que esta a ser descrito</title>
    <link>Introdução do link do item que esta a ser descrito</link>
    <!-- Deixo de ter dc:abstract e passo a ter dc:description-->
    <description>Introdução de uma descrição do item que esta a ser descrito</description>
    <!--PUBLISHER-->
    <dc:publisher>APSI</dc:publisher>
    <!--CREATOR-->
    <dc:creator>
      <rdf:Seq>
        <rdf:li>Introdução dos autores do item que esta a ser descrito</rdf:li>
        <rdf:li>Introdução dos autores do item que esta a ser descrito</rdf:li>
      </rdf:Seq>
    </dc:creator>
  </item>
</rdf:RDF>
```



```

        </rdf:Seq>
    </dc:creator>
    <!--RIGHTS -->
    <dc:rights>APSI. Todos os direitos reservados</dc:rights>
    <!--SUBJECT -->
    <dc:subject rdf:datatype="http://www.acm.org/class/1998/acmccs98-1.2.3.xml">
        Introdução do assunto que esta relacionado com o item que está a ser descrito, de acordo com os tópicos definidos no
        sistema de classificação da ACM CCS</dc:subject>
    <!--FORMATO -->
    <dcterms:extent>Introdução do tamanho (número de bytes) do item </dcterms:extent>
    <dcterms:medium>
        <dcterms:IMT>
            <rdf:value>Introdução do formato da aplicação do item (pdf, doc, txt, etc) </rdf:value>
        </dcterms:IMT>
    </dcterms:medium>
    <!--IDIOMA -->
    <dc:language>Introdução do idioma do item </dc:language>
    <!--DATA DE CRIAÇÃO DO ARTIGO -->
    <dcterms:created>
        <dcterms:W3CDTF>
            <rdf:value> Introdução da data de criação do item (artigo) a ser descrito segundo o formato W3CDTF
        </rdf:value>
        </dcterms:W3CDTF>
    </dcterms:created>
    <!--DATA DE PUBLICAÇÃO DO ARTIGO -->
    <dcterms:issued>
        <dcterms:W3CDTF>
            <rdf:value> Introdução da data de publicação do item (artigo) a ser descrito segundo o formato W3CDTF
        </rdf:value>
        </dcterms:W3CDTF>
    </dcterms:issued>
    <!-- ISPARTOF -->
    <dcterms:isPartOf rdf:resource="http://....."/>
    <!-- MEDIATOR -->
    <dcterms:mediator> Introdução da classe ou uma entidade responsável pela mediação do acesso ao recurso
    </dcterms:mediator>
    <!--REFERENCES-->
    <dcterms:references>
        <rdf:Seq>
            <rdf:li rdf:resource="http://....."/>
            <rdf:li rdf:resource="http://....."/>
        </rdf:Seq>
    </dcterms:references>

```

<!-- BIBLIOGRAPHIC CITATIONS EX: "Proceedings of the International Conference on Dublin Core and metadata for e-communities, 2002; DC-2002: Metadata for e-Communities: Supporting Diversity and Convergence, Florence, Italy, 13-17 October 2002, pp 71-80"-->

<dcterms:BibliographicCitations>Introdução da referência bibliográfica do recurso que esta a ser descrito

</dcterms:BibliographicCitations>

</item>

</rdf:RDF>

APÊNDICE B

Neste apêndice apresenta-se o perfil de aplicação dos documentos científicos considerados neste trabalho.

O Perfil de Aplicação é definido com o objectivo de declarar os elementos de metadados e os correspondentes *namespaces*, utilizados na descrição das publicações editadas pela APSI. Assim, os *namespaces* utilizados foram:

- Dublin Core Metadata Element Set [<http://purl.org/dc/elements/1.1/>]
- Elementos de refinamento do Dublin Core [<http://purl.org/dc/terms/>]
- RSS [<http://purl.org/rss/1.0/>]

Tabela de Conteúdos:

Nome	Title
Schema	http://purl.org/rss/1.0/
Encoding Schema	-
subPropertyOf	-
Definição	Titulo do recurso.
Comentário	É considerado como recurso Poster ou Artigo

Nome	Description
Schema	http://purl.org/rss/1.0/
Encoding Schema	-
subPropertyOf	Description
Definição	Uma descrição do conteúdo do elemento que está a ser descrito, que pode ser a descrição do <i>channel</i> ou do <i>item</i> de um <i>channel</i> .
Comentário	-

Nome	Link
Schema	http://purl.org/rss/1.0/
Encoding Schema	URI – Uniform Resource Identifier
subPropertyOf	-
Definição	-
Comentário	Este elemento descreve o URL do elemento que está a ser descrito, ou seja o URL do <i>channel</i> ou do <i>item</i> de um <i>channel</i> .

Nome	Creator
Schema	http://purl.org/dc/elements/1.1/
Encoding Schema	-
subPropertyOf	-
Definição	Autor do recurso.
Comentário	Exemplos de um autor, inclui uma pessoa, uma organização ou um serviço. Normalmente, o nome de um autor deve ser usado para indicar uma entidade.

Nome	Subject
Schema	http://purl.org/dc/elements/1.1/
Encoding Schema	Sistema de Classificação da ACM CCS.
subPropertyOf	-
Definição	Tópico sobre o conteúdo de um artigo, especificado de acordo com o Sistema de Classificação da ACM CCS.
Comentário	Normalmente, o elemento <i>Subject</i> expressa palavras-chave, frases ou sistemas de classificação que descrevem tópicos de um recurso. A especificação da DCMES recomenda a selecção de um valor a partir de um vocabulário controlado. O <code>rdfs:range</code> do <code>dc:subject</code> é o sistema de classificação do ACM CCS.

Nome	Publisher
Schema	http://purl.org/dc/elements/1.1/
Encoding Schema	-
subPropertyOf	-
Definição	A entidade responsável por disponibilizar o recurso.
Comentário	Exemplos de um <i>Publisher</i> inclui uma pessoa, uma organização, ou um serviço. Normalmente o nome de um <i>Publisher</i> deve ser usado na indicação de uma entidade.

Nome	Language
Schema	http://purl.org/dc/elements/1.1/
Encoding Schema	-
subPropertyOf	-
Definição	O idioma do conteúdo científico do artigo.
Comentário	É recomendada a utilização do RFC 3066 [RFC3066], que combinado com a norma ISO 639 [ISO639], define através de duas ou três letras primárias do idioma, <i>tags</i> com <i>subtags</i> opcionais. Exemplo incluir "en" or "eng" para o idioma Inglês, "en-GB" para o Inglês utilizado no Reino Unido e "pt" para o idioma Português.

Nome	Rights
Schema	http://purl.org/dc/elements/1.1/rights
Encoding Schema	-
subPropertyOf	-
Definição	Informação sobre os direitos de acesso e utilização do recurso.
Comentário	-

Nome	Created
Schema	http://purl.org/dc/terms/
Encoding Schema	W3C-DTF
subPropertyOf	Date
Definição	Data da criação (e.g. publicação) do artigo.
Comentário	-

Nome	Issued
Schema	http://purl.org/dc/terms/
Encoding Schema	W3C-DTF
subPropertyOf	Date
Definição	Data formal da emissão (e.g. publicação) do artigo.
Comentário	-

Nome	Medium
Schema	http://purl.org/dc/terms/
Encoding Schema	IMT
subPropertyOf	Format
Definição	O suporte material ou físico do recurso.
Comentário	-

Nome	isPartOf
Schema	http://purl.org/dc/terms/
Encoding Schema	URI
subPropertyOf	Relation
Definição	O recurso descrito inclui o recurso referenciado fisicamente ou logicamente.
Comentário	O recurso descrito é uma parte física ou lógica do recurso referenciado

Nome	Mediator
Schema	http://purl.org/dc/terms/
Encoding Schema	-
subPropertyOf	Audience
Definição	A classe ou uma entidade responsável por mediar o acesso ao recurso e para quem o recurso é dirigido ou útil.
Comentário	A audiência de um recurso é de dois tipo básico de classes: (1) beneficiário final do recurso, e (2) normalmente, uma entidade responsável por mediar o acesso ao recurso. O elemento refinador do elemento <i>mediator</i> representa a segunda destas duas classes.

Nome	BibliographicCitations
Schema	http://purl.org/dc/terms/
Encoding Schema	-
subPropertyOf	Identifier
Definição	Referência bibliográfica para o recurso
Comentário	É recomendado como boa prática incluir detalhes bibliográficos suficientes para identificar o recurso de forma desambigua, independentemente da citação estar ou não num formato standard.

Nome	References
Schema	http://purl.org/dc/terms/
Encoding Schema	URI
subPropertyOf	Relation
Definição	Descrição das referências bibliográficas utilizadas, ou citadas no recurso, ou que apontam para o recurso referenciado.
Comentário	-

Nome	Extent
Schema	http://purl.org/dc/terms/
Encoding Schema	-
subPropertyOf	-
Definição	-
Comentário	O elemento <i>extent</i> inclui informação sobre o tamanho do artigo.

APÊNDICE C

Neste apêndice apresentam-se exemplos de dois *feeds* codificados em RSS, definidos para cada tópico do sistema de classificação da ACM CCS. Os exemplos apresentados referem-se aos tópicos *Computing Methodologies* e *Hardware*.

C.1 *Computing Methodologies*

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns="http://purl.org/rss/1.0/"
  xmlns:dcterms="http://purl.org/dc/terms/"
  <channel rdf:about="http://repositorio.apsi.pt:8080/index.jsp">
    <title>Repositório da APSI </title>
    <link>http://repositorio.apsi.pt:8080/index.jsp</link>
    <description>Repositório institucional da Associação Portuguesa de Sistemas de Informação</description>
    <dc:publisher>Associação Portuguesa de Sistemas de Informação (APSI) </dc:publisher>
    <dc:creator>Rui Dinis de Sousa </dc:creator>
    <!-- Esta na página da APSI como Editor. Posso assumir que é o Creator?-->
    <dc:rights>APSI. Todos os direitos reservados</dc:rights>
    <dc:date>2006-05-23T09:47:57Z</dc:date>
    <!-- Um item é um recurso (um artigo) -->
    <items>
      <rdf:Seq>
        <rdf:li rdf:resource="http://hdl.handle.net/123456789/57"/>
        <rdf:li rdf:resource="http://hdl.handle.net/123456789/81"/>
        <rdf:li rdf:resource="http://hdl.handle.net/123456789/82"/>
        <rdf:li rdf:resource="http://hdl.handle.net/123456789/141"/>
      </rdf:Seq>
    </items>
  </channel>
  <!-- DESCRIÇÃO DO 1º ITEM -->
  <item rdf:about="http://hdl.handle.net/123456789/57">
    <title>Teorias Cognitivas no Design Instrucional de Sistemas Hipermedia e Sites de Ensino a Distância</title>
    <link>http://hdl.handle.net/123456789/57</link>
    <description>Aquando do desenvolvimento da estrutura de um sistema hipermedia e/ou website, é
```

necessário organizar os conteúdos, estabelecendo o tipo de ligação entre os nós do sistema.

Tal deverá ser determinado pelos objectivos do sistema. As organizações mais conhecidas podem ser divididas em duas classes: as sequenciais e as relacionais, que, por sua vez, podem ser subdivididas em associativas, elaborativas e hierárquicas. Com

estes tipos de ligações entre conteúdos constroem-se as formas mais comuns de estruturas fornecidas pela informática. Quando devidamente utilizadas, resultam em formas bem organizadas de representar informação, pelo menos para os tradicionais fins comerciais e/ou corporativos. No entanto, quando em causa estão sistemas hipermédia de aprendizagem e/ou websites de e-learning, os resultados nem sempre são adequados aos fins perseguidos. O design instrucional é um processo de tradução dos princípios da aprendizagem e instrução para o uso em materiais didácticos e deve constituir o ponto central no processo de planeamento e desenvolvimento de um sistema hipermédia de aprendizagem, provendo métodos para traduzir os meios gerais ou específicos e contribuindo, assim, para o sucesso do processo ensino-aprendizagem. O design instrucional deve nortear todo o processo de construção do programa, adequando-o ao objectivo educacional, tendo em conta as características do conteúdo do sistema, o público-alvo e a experiência do designer. As teorias cognitivas e da aprendizagem podem influenciar positivamente o design instrucional de sistemas hipermédia de aprendizagem e, indirectamente, a sua construção. Este trabalho mostra a aplicação da Teoria da Aprendizagem Significativa de Ausubel e da Teoria da Flexibilidade Cognitiva no design instrucional do sistema hipermédia e do website Biomec. O principal objectivo é promover uma aprendizagem dos conceitos de Mecânica Básica que tenha significado na estrutura cognitiva dos utilizadores primários dos sistemas, os estudantes das licenciaturas em Educação Física. </description>

```
<!--PUBLISHER-->
<dc:publisher>APSI</dc:publisher>
<!--AUTOR-->
<dc:creator>
  <rdf:Seq>
    <rdf:li>Silva, Alexander M.</rdf:li>
    <rdf:li>Santana, Silvina</rdf:li>
  </rdf:Seq>
</dc:creator>
<!--RIGHTS -->
<dc:rights>APSI. Todos os direitos reservados</dc:rights>
<!--SUBJECT -->
<dc:subject rdf:datatype="http://www.acm.org/class/1998/acmccs98-1.2.3.xml"> Computing Methodologies
</dc:subject>
<!--FORMATO -->
<dcterms:extent>202033 </dcterms:extent>
<dcterms:medium>
  <dcterms:IMT>
    <rdf:value> pdf</rdf:value>
  </dcterms:IMT>
</dcterms:medium>
<!--IDIOMA -->
<dc:language>pt</dc:language>
<!--DATA DE CRIAÇÃO DO ARTIGO -->
<dcterms:created>
  <dcterms:W3CDTF>
    <rdf:value> 2004-11-03T14:25:53Z</rdf:value>
  </dcterms:W3CDTF>
</dcterms:created>
<!--DATA DE PUBLICAÇÃO DO ARTIGO -->
<dcterms:issued>
```

```

    <dcterms:W3CDTF>
      <rdf:value>2005-06-20T14:25:53Z </rdf:value>
    </dcterms:W3CDTF>
  </dcterms:issued>
</item>
<!--DESCRIÇÃO DO 2º ITEM -->
<item rdf:about="http://hdl.handle.net/123456789/81">
  <title>Design de Interface do BIOMECC</title>
  <link>http://hdl.handle.net/123456789/81</link>
  <description>A Biomecânica é uma área de estudo da Bioengenharia, adoptada pela Educação Física, que se propõe
entender as funções biológicas à luz de conceitos básicos de Mecânica. É sensocomum entre os professores de Biomecânica
que o aluno, em geral, ingressa no curso com reduzidos conhecimentos de Física. Os estudantes parecem ver pouca utilidade
neste conhecimento e consideram o estudo da Física uma tarefa árdua, apesar de os conceitos físicos serem necessários à
compreensão do conteúdo da disciplina.
No sentido de tentar remediar estes problemas, desenvolveram-se o sistema hipermedia e o website Biomec.
A qualidade dos sistemas hipermedia e dos websites é especialmente dependente do design de interface, que comporta factores
humanos e a comunicação do sistema com o utilizador.
Uma interface bem estruturada é capaz de permitir que utilizadores com variados níveis de conhecimento em navegação
hipertextual possam percorrer as páginas intuitivamente, possibilitando o atendimento das suas necessidades. Nesta perspectiva,
a interface é o próprio produto em si. Estas considerações são ainda mais pertinentes quando em causa estão sistemas
hipermedia e websites de aprendizagem. O objectivo do presente estudo é descrever o design das interfaces do sistema
hipermedia e do website de aprendizagem Biomec, que têm como objectivo interrelacionar os conceitos de Desporto e de
Mecânica Básica, numa área de conhecimento da Biomecânica denominada Física do Desporto.</description>
  <!--PUBLISHER-->
  <dc:publisher>APSI</dc:publisher>
  <!--AUTOR-->
  <dc:creator>
    <rdf:Seq>
      <rdf:li>Silva, Alexander M.</rdf:li>
      <rdf:li>Santana, Silvina</rdf:li>
    </rdf:Seq>
  </dc:creator>
  <!--RIGHTS -->
  <dc:rights>APSI. Todos os direitos reservados</dc:rights>
  <!--SUBJECT -->
  <dc:subject rdf:datatype="http://www.acm.org/class/1998/acmccs98-1.2.3.xml"> Computing Methodologies
  </dc:subject>
  <!--FORMATO -->
  <dcterms:extent>285443 </dcterms:extent>
  <dcterms:medium>
    <dcterms:IMT>
      <rdf:value> pdf</rdf:value>
    </dcterms:IMT>
  </dcterms:medium>
  <!--IDIOMA -->

```

```

<dc:language>pt</dc:language>
<!--DATA DE CRIAÇÃO DO ARTIGO -->
<dcterms:created>
  <dcterms:W3CDTF>
    <rdf:value> 2004-11-03T22:30:36Z</rdf:value>
  </dcterms:W3CDTF>
</dcterms:created>
<!--DATA DE PUBLICAÇÃO DO ARTIGO -->
<dcterms:issued>
  <dcterms:W3CDTF>
    <rdf:value>2005-06-21 </rdf:value>
  </dcterms:W3CDTF>
</dcterms:issued>
</item>
<!--DESCRIÇÃO DO 3º ITEM -->
<item rdf:about="http://hdl.handle.net/123456789/82">
  <title>Utilização do Linux Caixa Mágica nas escolas: o caso de estudo português.</title>
  <link>http://hdl.handle.net/123456789/82</link>
  <description>A utilização de software livre nas escolas pode ter impacto sobre a sociedade por duas vias: a) os alunos
(re)conhecerem novas tecnologias e b) permitir poupanças a nível de licenciamento e manutenção. O presente artigo propõe
apresentar o caso de estudo do sistema idealizado e posto em prática para a instalação do Linux Caixa Mágica em 15.000
computadores da rede escolar portuguesa, no âmbito do projecto das Salas TIC do Ministério da Educação.
O sistema tem várias particularidades: interacção com o servidor da sala a funcionar sobre 2003 Server, sistema próprio de
reposição das estações de trabalho e selecção criteriosa de aplicações - chave que vão de encontro às necessidades dos
alunos.
As soluções apresentadas reflectem os novos modelos de desenvolvimento de sistemas de informação baseados em software
livre. Em concreto, no decorrer do projecto foi necessário corrigir software de terceiras partes, tendo essas correcções sido
integradas mais tarde nesse software.</description>
  <!--PUBLISHER-->
  <dc:publisher>APSI</dc:publisher>
  <!--AUTOR-->
  <dc:creator>
    <rdf:Seq>
      <rdf:li>Nunes, Susana</rdf:li>
      <rdf:li>Moringa, Flávio</rdf:li>
      <rdf:li>Lourenço, Miguel</rdf:li>
      <rdf:li>Trezentos, Paulo</rdf:li>
    </rdf:Seq>
  </dc:creator>
  <!--RIGHTS -->
  <dc:rights>APSI. Todos os direitos reservados</dc:rights>
  <!--SUBJECT -->
  <dc:subject rdf:datatype="http://www.acm.org/class/1998/acmccs98-1.2.3.xml"> Computing Methodologies
  </dc:subject>
  <!--FORMATO -->

```

```

<dcterms:extent>250109 </dcterms:extent>
<dcterms:medium>
  <dcterms:IMT>
    <rdf:value> pdf</rdf:value>
  </dcterms:IMT>
</dcterms:medium>
<!--IDIOMA -->
<dc:language>en</dc:language>
<!--DATA DE CRIAÇÃO DO ARTIGO -->
<dcterms:created>
  <dcterms:W3CDTF>
    <rdf:value>2004-11-03T22:30:36Z</rdf:value>
  </dcterms:W3CDTF>
</dcterms:created>
<!--DATA DE PUBLICAÇÃO DO ARTIGO -->
<dcterms:issued>
  <dcterms:W3CDTF>
    <rdf:value>2005-06-21 </rdf:value>
  </dcterms:W3CDTF>
</dcterms:issued>
</item>
<!--DESCRIÇÃO DO 4º ITEM -->
<item rdf:about="http://hdl.handle.net/123456789/141">
  <title>The Holistic Approach to Information Systems Implementation</title>
  <link>http://hdl.handle.net/123456789/141</link>
  <description>In this paper, we present an argument in favour of an ontological/epistemological stance, labelled as Organizational Holism approach to IS/IT implementation. In order to frame the argument, firstly the literature on the existing classifications of theoretical perspectives on IS/IT implementation is reviewed (DeSanctis and Pool, 1994; Campbell, 1996; Markus and Robey, 1988; Symons, 1991) and the following re-arranged classification is suggested: Technological Optimism, Strategic Rationality, Socio-Technical Interactionism and Organizational Holism. According to the Organizational Holism perspective, IS/IT implementation should not be seen as a "one-off" event, which is finished when the information systems development cycle is complete. Rather than a single step in the methodological frameworks popularized by the technical or the strategic approaches, IS/IT implementation is seen as a process more akin to organizational growth, learning and change. The causes and consequences of applying information systems to organizations cannot be reduced to a series of single events or analytical snapshots, but have to be seen as an holistic phenomena with pervasive and continuous consequences which must be studied by means of longitudinal analytical tools. Secondly, Organizational Holism is presented as part of a broader movement in the managerial and the organizational worlds pointing in similar intellectual directions and founded upon the application of Complexity to these fields of study (Zimmerman and Hurst, 1993; von Krogh and Roos, 1995; Guedes, 1999; Wheatley, 1999; Pascale et al, 2000; Lewin and Regine, 2000; Stacey, 2001). The paper concludes with a recommendation to organizations in general but to institutions of higher education in particular, for serious thought to be given to the need of a much stronger organizational component in information systems thinking, education and management.</description>
  <!--PUBLISHER-->
  <dc:publisher>APSI</dc:publisher>
  <!--AUTOR-->
  <dc:creator>

```

```

    <rdf:Seq>
      <rdf:li>Magalhães, Rodrigo</rdf:li>
    </rdf:Seq>
  </dc:creator>
  <!--RIGHTS -->
  <dc:rights>APSI. Todos os direitos reservados</dc:rights>
  <!--SUBJECT -->
  <dc:subject rdf:datatype="http://www.acm.org/class/1998/acmccs98-1.2.3.xml"> Computing Methodologies
    </dc:subject>
  <!--FORMATO -->
  <dcterms:extent>126694</dcterms:extent>
  <dcterms:medium>
    <dcterms:IMT>
      <rdf:value>pdf</rdf:value>
    </dcterms:IMT>
  </dcterms:medium>
  <!--IDIOMA -->
  <dc:language>en</dc:language>
  <!--DATA DE CRIAÇÃO DO ARTIGO -->
  <dcterms:created>
    <dcterms:W3CDTF>
      <rdf:value> 2004-11-03</rdf:value>
    </dcterms:W3CDTF>
  </dcterms:created>
  <!--DATA DE PUBLICAÇÃO DO ARTIGO -->
  <dcterms:issued>
    <dcterms:W3CDTF>
      <rdf:value>2005-06-26T19:05:35Z</rdf:value>
    </dcterms:W3CDTF>
  </dcterms:issued>
</item>
</rdf:RDF>

```

C.2 Hardware

```

<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns="http://purl.org/rss/1.0/"
  xmlns:dcterms="http://purl.org/dc/terms/"
  <channel rdf:about="http://repositorio.apsi.pt:8080/index.jsp">
    <title>Repositório da APSI </title>
    <link>http://repositorio.apsi.pt:8080/index.jsp</link>
    <description>Repositório institucional da Associação Portuguesa de Sistemas de Informação</description>

```

```

<dc:publisher>Associação Portuguesa de Sistemas de Informação (APSI) </dc:publisher>
<dc:creator>Rui Dinis de Sousa </dc:creator>
<dc:rights>APSI. Todos os direitos reservados</dc:rights>
<dc:date>2006-05-23T09:47:57Z</dc:date>
<!-- Um item é um recurso (um artigo) -->
<items>
  <rdf:Seq>
    <rdf:li rdf:resource="http://hdl.handle.net/123456789/92"/>
    <rdf:li rdf:resource="http://hdl.handle.net/123456789/93"/>
    <rdf:li rdf:resource="http://hdl.handle.net/123456789/94"/>
    <rdf:li rdf:resource="http://hdl.handle.net/123456789/95"/>
  </rdf:Seq>
</items>
</channel>
<!-- DESCRIÇÃO DO 1º ITEM -->
<item rdf:about="http://hdl.handle.net/123456789/92">
  <title>Representação As-Is em Engenharia Organizacional</title>
  <link>http://hdl.handle.net/123456789/92</link>
  <description>O Modelo Organizacional As-Is visa representar as organizações nas suas várias dimensões. Apesar do reconhecimento da mais valia deste modelo como suporte a várias tarefas organizacionais, como a reengenharia de processos de negócio, a implementação de sistemas de gestão de qualidade e a captura de requisitos para o desenvolvimento de sistemas de informação, tem-se encarado a sua existência como algo descartável após a sua utilização. Este facto leva a que o modelo tenha que ser reconstruído sempre que seja necessário, pois as organizações são dinâmicas. Neste documento tenta demonstrar-se se não valerá a pena tornar o modelo organizacional numa componente dinâmica da organização, reflectindo a organização constantemente ao longo do tempo. É apresentando um meta-modelo do processo de actualização dinâmica do modelo organizacional baseado na Framework CEO.</description>
  <!-- PUBLISHER -->
  <dc:publisher>APSI</dc:publisher>
  <!-- AUTOR -->
  <dc:creator>
    <rdf:Seq>
      <rdf:li>Castela, Nuno</rdf:li>
      <rdf:li>Tribolet, José</rdf:li>
    </rdf:Seq>
  </dc:creator>
  <!-- RIGHTS -->
  <dc:rights>APSI. Todos os direitos reservados</dc:rights>
  <!-- SUBJECT -->
  <dc:subject rdf:datatype="http://www.acm.org/class/1998/acmccs98-1.2.3.xml"> Hardware</dc:subject>
  <!-- FORMATO -->
  <dcterms:extent>134727 </dcterms:extent>
  <dcterms:medium>
    <dcterms:IMT>
      <rdf:value> pdf</rdf:value>
    </dcterms:IMT>

```

```

</dcterms:medium>
<!--IDIOMA -->
<dc:language>pt </dc:language>
<!--DATA DE CRIAÇÃO DO ARTIGO -->
<dcterms:created>
  <dcterms:W3CDTF>
    <rdf:value> 2004-11-03</rdf:value>
  </dcterms:W3CDTF>
</dcterms:created>
<!--DATA DE PUBLICAÇÃO DO ARTIGO -->
<dcterms:issued>
  <dcterms:W3CDTF>
    <rdf:value>2005-06-21 </rdf:value>
  </dcterms:W3CDTF>
</dcterms:issued>
</item>
<!--DESCRIÇÃO DO 2º ITEM -->
<item rdf:about="http://hdl.handle.net/123456789/93">
  <title>Método de Auditoria de Sistemas de Informação Simplificado: Uma Proposta para PME's</title>
  <link>http://hdl.handle.net/123456789/93</link>
  <description>O papel da informação tem aumentado significativamente nas actividades desenvolvidas pelas
organizações. Idalberto Chiavenato [Chiavenato 2000] afirma que, na era da informação, o recurso mais importante deixou de ser
o capital financeiro para ser o capital intelectual, baseado no conhecimento. Neste sentido, a função de controlo ou Auditoria de
Sistemas de Informação é cada vez mais, uma função necessária e essencial para as organizações. Este artigo tem como
propósito apresentar uma solução que viabilize a execução de projectos de Auditoria de Sistemas de Informação em Pequenas e
Médias Empresas, para tal, é proposto um Método e uma Ferramenta de Auditoria de Sistemas de Informação cuja aplicação
minimiza os recursos necessários para a execução de um projecto deste tipo.</description>
  <!--PUBLISHER-->
  <dc:publisher>APSI</dc:publisher>
  <!--AUTOR-->
  <dc:creator>
    <rdf:Seq>
      <rdf:li>Mendonça, Vítor</rdf:li>
      <rdf:li>Amaral, Luís</rdf:li>
    </rdf:Seq>
  </dc:creator>
  <!--RIGHTS -->
  <dc:rights>APSI. Todos os direitos reservados</dc:rights>
  <!--SUBJECT -->
  <dc:subject rdf:datatype="http://www.acm.org/class/1998/acmccs98-1.2.3.xml"> Hardware</dc:subject>
  <!--FORMATO -->
  <dcterms:extent>222238 </dcterms:extent>
  <dcterms:medium>
    <dcterms:IMT>
      <rdf:value>pdf</rdf:value>

```



```

        </dcterms:IMT>
    </dcterms:medium>
    <!--IDIOMA -->
    <dc:language>pt</dc:language>
    <!--DATA DE CRIAÇÃO DO ARTIGO -->
    <dcterms:created>
        <dcterms:W3CDTF>
            <rdf:value>2004-11-03</rdf:value>
        </dcterms:W3CDTF>
    </dcterms:created>
    <!--DATA DE PUBLICAÇÃO DO ARTIGO -->
    <dcterms:issued>
        <dcterms:W3CDTF>
            <rdf:value>2005-06-21T01:08:21Z</rdf:value>
        </dcterms:W3CDTF>
    </dcterms:issued>
</item>
<!--DESCRIÇÃO DO 3º ITEM -->
<item rdf:about="http://hdl.handle.net/123456789/94">
    <title>Arquitectura de Sistemas de Informação do Instituto da Vinha e do Vinho (IVV)</title>
    <link>http://hdl.handle.net/123456789/94</link>
    <description>O IVV com a colaboração de consultoria externa, definiu a sua arquitectura de sistemas de informação,
com o objectivo de lançar um concurso publico internacional para a implementação da referida arquitectura. O presente
documento tem como objectivo descrever a metodologia utilizada para a definição da arquitectura, bem como descrever
sumariamente as diferentes actividades desenvolvidas.</description>
    <!--PUBLISHER-->
    <dc:publisher>APSI</dc:publisher>
    <!--AUTOR-->
    <dc:creator>
        <rdf:Seq>
            <rdf:li>Vieira, Armando</rdf:li>
            <rdf:li>Amaro, Pedro</rdf:li>
            <rdf:li>Gato, Paulo</rdf:li>
            <rdf:li>Sousa, Pedro</rdf:li>
            <rdf:li>Marques, Alexandra</rdf:li>
        </rdf:Seq>
    </dc:creator>
    <!--RIGHTS -->
    <dc:rights>APSI. Todos os direitos reservados</dc:rights>
    <!--SUBJECT -->
    <dc:subject rdf:datatype="http://www.acm.org/class/1998/acmccs98-1.2.3.xml"> Hardware</dc:subject>
    <!--FORMATO -->
    <dcterms:extent>488041 </dcterms:extent>
    <dcterms:medium>
        <dcterms:IMT>

```

```

        <rdf:value>pdf</rdf:value>
    </dcterms:IMT>
</dcterms:medium>
<!--IDIOMA -->
<dc:language>en</dc:language>
<!--DATA DE CRIAÇÃO DO ARTIGO -->
<dcterms:created>
    <dcterms:W3CDTF>
        <rdf:value>2004-11-03</rdf:value>
    </dcterms:W3CDTF>
</dcterms:created>
<!--DATA DE PUBLICAÇÃO DO ARTIGO -->
<dcterms:issued>
    <dcterms:W3CDTF>
        <rdf:value>2005-06-21T13:46:02Z </rdf:value>
    </dcterms:W3CDTF>
</dcterms:issued>
</item>
<!--DESCRIÇÃO DO 4º ITEM -->
<item rdf:about="http://hdl.handle.net/123456789/95">
    <title>O Losango em ambiente SPSS como suporte à Análise de Dados</title>
    <link>http://hdl.handle.net/123456789/95</link>
    <description>A nossa experiência no desenvolvimento de sistemas informáticos para apoio a projectos de investigação
na área das ciências sociais permitiu-nos, a partir das soluções concretas, generalizar e surgir, há algum tempo, com a proposta
de uma plataforma para montar sistemas desta família, o FDASE. Este framework descreve uma arquitetura multi-camada que
contempla todas as fases desde a aquisição, armazenamento, fusão e análise dos dados até à divulgação do conhecimento
extraído nas fases anteriores. Desta forma concentramos em cada camada as questões específicas de cada uma das tarefas,
concentrando-nos na sua resolução e garantindo independência em relação às restantes camadas. A comunicação entre
patamares do framework faz-se por partilha dos dados. No desenho do framework teve-se em atenção o facto de, nestas áreas,
existirem sistematicamente situações de: fontes de informação heterogéneas; grandes volumes de dados; e dados omissos ou
imprecisos. Da aplicação da referida solução a vários casos de estudo, concluímos que o esforço realizado no patamar de
análise era enorme e deixava o utilizador final pouco autónomo. Especialmente gritante é o facto de todas as ferramentas de
análise, que preconizamos para serem usadas alternativamente e em paralelo, requererem um processo de preparação de dados
que é complexo e consumidor de recursos. Daí termos concluído que seria premente criar um nível externo de preparação
comum, providenciando os dados tratados num formato legível pelas várias ferramentas, o qual designamos de Losango1 para
Análise de Dados (LAD), e garantindo a adaptação dessa operação às necessidades dos analistas. É nossa convicção que
contribuímos para tornar todo o processo de análise mais rápido e fácil, aumentando a autonomia dos seus utilizadores.
Atendendo a que o SPSS é a ferramenta mais usada pelos investigadores desta área e que fornece mecanismos de preparação
dos dados para seu próprio uso, é objectivo deste artigo mostrar que o dito SPSS pode actuar como produtor do LAD. Além disso
o SPSS oferece mecanismos de memorização de Workflows que podem ser reutilizados em diversas situações, sistematizando o
trabalho e poupando esforço.</description>
    <!--PUBLISHER-->
    <dc:publisher>APSI</dc:publisher>
    <!--AUTOR-->
    <dc:creator>

```

```
<rdf:Seq>
  <rdf:li>Faria, Fernanda</rdf:li>
  <rdf:li>Henriques, Pedro</rdf:li>
</rdf:Seq>
</dc:creator>
<!--RIGHTS -->
<dc:rights>APSI. Todos os direitos reservados</dc:rights>
<!--SUBJECT -->
<dc:subject rdf:datatype="http://www.acm.org/class/1998/acmccs98-1.2.3.xml"> Hardware</dc:subject>
<!--FORMATO -->
<dcterms:extent>445477 </dcterms:extent>
<dcterms:medium>
  <dcterms:IMT>
    <rdf:value>pdf</rdf:value>
  </dcterms:IMT>
</dcterms:medium>
<!--IDIOMA -->
<dc:language>en</dc:language>
<!--DATA DE CRIAÇÃO DO ARTIGO -->
<dcterms:created>
  <dcterms:W3CDTF>
    <rdf:value> 2004-11-03</rdf:value>
  </dcterms:W3CDTF>
</dcterms:created>
<!--DATA DE PUBLICAÇÃO DO ARTIGO -->
<dcterms:issued>
  <dcterms:W3CDTF>
    <rdf:value>2005-06-21T13:52:32Z</rdf:value>
  </dcterms:W3CDTF>
</dcterms:issued>
</item>
</rdf:RDF>
```

APÊNDICE D

Neste apêndice apresenta-se o código XSL relativo à *stylesheet* definida para efectuar a transformação da metainformação dos artigos científicos descritos segundo o XML *schema* OAI-PMH para o formato RSS.

```

<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet
xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns="http://purl.org/rss/1.0/"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:dcterms="http://purl.org/dc/terms#"
xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc"
xmlns:oai="http://www.openarchives.org/OAI/2.0"
version="1.0">

<xsl:output indent="yes" method="xml" encoding="ISO-8859-1"/>
<xsl:template match="oai:OAI-PMH">
<rdf:RDF>
<channel rdf:about="http://repositorio.apsi.pt:8080/index.jsp">
  <title>Repositório da APSI </title>
  <link>http://repositorio.apsi.pt:8080/index.jsp</link>
  <description>Repositório institucional da Associação Portuguesa de Sistemas de Informação</description>
  <dc:publisher>Associação Portuguesa de Sistemas de Informação (APSI) </dc:publisher>
  <dc:creator>Rui Dinis Sousa </dc:creator>
  <dc:rights>APSI. Todos os direitos reservados</dc:rights>
  <dc:date>2006-05-23T09:47:57Z</dc:date>
  <items>
    <rdf:Seq>
      <xsl:for-each select="//oai:record">
        <xsl:element name="rdf:li">
          <xsl:attribute name="rdf:resource">
            <xsl:value-of select="oai:metadata/oai_dc:dc/dc:identifier"/>
          </xsl:attribute>
        </xsl:element>
      </xsl:for-each>
    </rdf:Seq>
  </items>
</channel>
<xsl:for-each select="//oai:record">
  <xsl:element name="item">

```

```

<xsl:attribute name="rdf:about">
  <xsl:value-of select="oai:metadata/oai_dc/dc/dc:identifier"/>
</xsl:attribute>
<title><xsl:value-of select="oai:metadata/oai_dc/dc/dc:title"/></title>
<link><xsl:value-of select="oai:metadata/oai_dc/dc/dc:identifier"/></link>
<description>
  <xsl:value-of select="oai:metadata/oai_dc/dc/dc:description"/>
</description>
<dc:publisher>APSI</dc:publisher>
<dc:creator>
  <rdf:Seq>
    <xsl:for-each select="oai:metadata/oai_dc/dc/dc:creator">
      <rdf:li>
        <xsl:value-of select="."/>
      </rdf:li>
    </xsl:for-each>
  </rdf:Seq>
</dc:creator>
<dc:rights>APSI. Todos os direitos reservados</dc:rights>
<xsl:element name="dc:subject">
  <xsl:attribute name="rdf:datatype">http://www.acm.org/class/1998/acmccs98-1.2.3.xml</xsl:attribute>
</xsl:element>
<dcterms:extent>
  <xsl:value-of select="oai:metadata/oai_dc/dc/dc:format"/>
</dcterms:extent>
<dcterms:medium>
  <dcterms:IMT>
    <rdf:value>
      <xsl:value-of select="oai:metadata/oai_dc/dc/dc:format[2]"/>
    </rdf:value>
  </dcterms:IMT>
</dcterms:medium>
<dc:language>
  <xsl:value-of select="oai:metadata/oai_dc/dc/dc:language"/>
</dc:language>
<dcterms:created>
  <dcterms:W3CDTF>
    <rdf:value>
      <xsl:value-of select="oai:metadata/oai_dc/dc/dc:date[3]"/>
    </rdf:value>
  </dcterms:W3CDTF>
</dcterms:created>
<dcterms:issued>
  <dcterms:W3CDTF>
    <rdf:value>

```

```
<xsl:value-of select="oai:metadata/oai_dc:dc:date"/>
</rdf:value>
</dcterms:W3CDTF>
</dcterms:issued>
</xsl:element>
</xsl:for-each>
</rdf:RDF>
</xsl:template>
</xsl:stylesheet>
```

APÊNDICE E

Neste apêndice apresenta-se o código RSP relativo à parte do protótipo desenvolvida no RDF Gateway designadamente a pesquisa de informação sobre a base de metadados.

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN">
<html>
  <link href="http://localhost/apsi/subject.css" rel="stylesheet" type="text/css" />
<head>
<title>Estrutura hierárquica do sistema de classificação da ACM Computing versão 1998</title>
<!-- <meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1"> -->
</head>

<body bgcolor="white" text="white" onload="init()" onresize="sizePanels()">
<script>
var icons = new Array(2);
var g_loc = null;
var g_selectedTopic = null;

icons[0] = new Image();
icons[0].src = "plus.gif";
icons[1] = new Image();
icons[1].src = "minus.gif";
icons[2] = new Image();
icons[2].src = "rss_ico.gif";

function init()
{
  sizePanels();
  setLocation("/pages/default.rsp");
}
function setLocation(loc)
{
  g_loc = loc;
  selectTopic(getTopic(loc));
}

function loadTopic(loc)
{
  var loader = document.getElementById("loader");
  try
  {
```

```
loader.setParameter("path", loc);
var ancestors = loader.execute("tree.rsp").split(" ");
for (var a=ancestors.length - 3; a>=0; a--)
{
    var elDiv = document.getElementById("div_" + ancestors[a]);
    var elImg = document.getElementById("img_" + ancestors[a]);
    if (elDiv.className == "needsdata")
    {
        elDiv.innerHTML = getData(ancestors[a]);
        elDiv.className = "hasdata";
    }
    if (elImg != null)
    elImg.src = icons[1].src;
    elDiv.style.display = "";
}
return findTopic(loc);
}
catch(e)
{return null;}
}

function findTopic(loc)
{
    var anchors = document.getElementsByTagName("A");
    for (var i=0; i<anchors.length; i++)
    {
        if (anchors[i].href.search(loc) != -1)
            return anchors[i];
    }
    return null;
}

function getTopic(loc)
{
    var topic = findTopic(loc);
    if (topic != null)
        return topic;
    return loadTopic(loc);
}

function selectTopic(el)
{
    if (g_selectedTopic != null)
        g_selectedTopic.style.backgroundColor = g_selectedTopic.parentNode.style.backgroundColor;
    if (el != null)
    {
        var p = el;
```



```
while ((p=p.parentNode) != null && p.style != null)
{
    var pid = p.id;
    if (pid.search("div_") == 0)
    {
        var id = pid.substr(4);
        var eImg = document.getElementById("img_" + id);
        if (eImg != null)
            eImg.src = icons[1].src;
    }
    p.style.display = "";
}
try
{
    el.scrollIntoView(false);
}
catch(e)
{
}
el.style.backgroundColor = "blue";
}
g_selectedTopic = el;
}
function next()
{
    top.frames["main"].location = "/topic.rsp?a=next&t=" + escape(g_loc);
}

function prev()
{
    top.frames["main"].location = "/topic.rsp?a=prev&t=" + escape(g_loc);
}

function getData(topic)
{
    var loader = document.getElementById("loader");

    try
    {
        loader.setParameter("topic", topic);
        return loader.execute("tree.rsp");
    }
    catch(e)
    {
        return "";
    }
}
```

```
}  
}  
  
function toggle(topic)  
{  
  var elDiv = document.getElementById("div_" + topic);  
  var elImg = document.getElementById("img_" + topic);  
  
  if (elDiv.className == "needsdata")  
  {  
    elDiv.innerHTML = getData(topic);  
    elDiv.className = "hasdata";  
  }  
  
  if (elDiv.style.display == 'none')  
  {  
    if (elImg != null)  
      elImg.src = icons[1].src;  
    elDiv.style.display = "";  
  }  
  else  
  {  
    if (elImg != null)  
      elImg.src = icons[0].src;  
    elDiv.style.display = 'none';  
  }  
}  
  
function toggle2(topic)  
{  
  var elDiv = document.getElementById("div_" + topic);  
  var elImg = document.getElementById("img_" + topic);  
  
  if (elDiv.className == "needsdata")  
  {  
    elDiv.innerHTML = getData(topic);  
    elDiv.className = "hasdata";  
  }  
  
  if (elDiv.style.display == 'none')  
  {  
    if (elImg != null)  
      elImg.src = icons[1].src;  
    elDiv.style.display = "";  
  }  
}
```

```
else
{
    if (ellmg != null)
        ellmg.src = icons[0].src;
    elDiv.style.display = 'none';
}
}

function sizePanels()
{
    var extra = 10;

    var sidebars = document.getElementById("sidebars");
    var brbar = document.getElementById("browsebar");

    var pageHeight = window.innerHeight==null?document.body.clientHeight>window.innerHeight;

    var thh = getTop(brbar) - getTop(sidebars);
    var avail = pageHeight
        - getTop(sidebars)
        - thh
        - extra;
    brbar.style.height = avail;
}

function getTop(el)
{
    {
        var top = 0;
        var o = el;
        do{
            top += o.offsetTop;
        }
        while ((o=o.offsetParent) != null);
    }
    return top;
}

function newsearch(topic)
{
    var url = "http://localhost/apsi/prot11_01.asp?pnorm=3&query=" + topic;
    window.open(url, 'mainframe', 'width=780, height=300, scrollbars=1, resizable=1, toolbar=0, location=0,
    directories=0, status=0, menubar=0');
}
</script>
<div id="sidebars">
<table width="100%" cellspacing="0" cellpadding="3" class="outline">
<tr>
```

```

        <th id="toggle" align="left" style="background-color: #000080">Estrutura Hierárquica do Sistema de
Classificação da ACM Computing <applet id="loader" code="jod.class" width="1" height="1"></applet></th>
</tr>
<tr>
    <td bgcolor="#000080">
        <div id="browsebar" style="overflow:auto;height:100">
            <table width="100%" cellspacing="0">
                <tr>
                    </tr>
            </table>
        </div>
        <%
use apsi;
var id;
var categoria;
var subcategoria;
var subsubcategoria;
var subsubsubcategoria;
var subsubsubsubcategoria;

var rs= (select ?id ?name using acmccs where
{{[http://www.w3.org/2000/01/rdf-schema#subClassOf] ?id [http://purl.org/dc/dcmitype/Text]}
and {[http://www.w3.org/2000/01/rdf-schema#label] ?id ?name})
while (!rs.EOF)
{
    id=rs["id"];
    categoria=rs["name"];

    var rs1= (select ?id ?subclasse using acmccs where
                {[http://www.w3.org/2000/01/rdf-schema#subClassOf] ?id #id}
                and {[http://www.w3.org/2000/01/rdf-schema#label] ?id ?subclasse})
    if (rs1.EOF)
        response.write("src='empty.gif' ");
    else
        <%>
        <div class="category">
            <a href="javascript:toggle('asd<%=categoria%>')">
            <img <%=categoria%>" <%
                response.write("src='plus.gif' ");
            </a>
            <%> align="absbottom" border="0"></a>

            <a
                class="subcategory"
                target="mainframe"
href="http://localhost/apsi/prot11_v01.asp?pnorm=3&keylist=<%=categoria%>" ><%=categoria%></a>
            <a href="<%=categoria%>.rdf"><img src='rss_ico.gif' > </a>
            <div style="display:none" id="div_asd<%=categoria%>"
            <%

```

```

if (rs1.EOF)
    response.write("class='hasdata' >");
else
    response.write("class='needdata' >");
%>

<%
while (!rs1.EOF)
{
    id=rs1["id"];
    subcategoria=rs1["subclasse"];

    var rs2= (select ?id ?subsubclasse using acmccs where
    {[http://www.w3.org/2000/01/rdf-schema#subClassOf] ?id #id}
    and {[http://www.w3.org/2000/01/rdf-schema#label] ?id ?subsubclasse});

%>
<div class="category">
    <a href="javascript:toggle('asd<%=subcategoria%>')">
    <img <%=subcategoria%>" <%

    if (rs2.EOF)
        response.write("src='empty.gif' ");
    else
        response.write("src='plus.gif' ");

    %> align="absbottom" border="0"></a>
    <a
        class="subcategoria"
        target="mainframe"
href="http://localhost/apsi/prot11_v01.asp?pnorm=3&keylist=<%=subcategoria%>"><%=subcategoria%></a>
    <div style="display:none" id="div_asd<%=subcategoria%>"
    <%

        if (rs2.EOF)
            response.write("class='needdata' >");
        else
            response.write("class='hasdata' >");

    %>
    <%

    while (!rs2.EOF)
    {
        id=rs2["id"];
        subsubcategoria=rs2["subsubclasse"];
        var rs3= (select ?id ?subsubsubclasse using acmccs where
        {[http://www.w3.org/2000/01/rdf-schema#subClassOf] ?id #id}
        and {[http://www.w3.org/2000/01/rdf-schema#label] ?id ?subsubsubclasse});
        %>

```

```

<div class="category">
<a href="javascript:toggle('asd<%=subsubcategoria%>')">
<img <%=subsubcategoria%>" <%
if (rs3.EOF)
    response.write("src='empty.gif' ");
else
    response.write("src='plus.gif' ");
%>
    align="absbottom" border="0"></a>
<a
        class="subcategory"
        target="mainframe"
href="http://localhost/apsi/prot11_v01.asp?pnorm=3&keylist=<%=subsubcategoria%>"><%=subsubcategoria%></a>
<div style="display:none" id="div_asd<%=subsubcategoria%>"
<%
if (rs3.EOF)
    response.write("class='needdata' >");
else
    response.write("class='hasdata' >");
%>
<%
while (!rs3.EOF)
{
    id=rs3["id"];
    subsubsubcategoria=rs3["subsubsubclasse"];
    var rs4 =(select ?id ?subsubsubsubclasse using acmccs where
    {[http://www.w3.org/2000/01/rdf-schema#subClassOf] ?id #id}
    and {[http://www.w3.org/2000/01/rdf-schema#label] ?id ?subsubsubsubclasse})
%>
<div class="category">
<a href="javascript:toggle('asd<%=subsubsubcategoria%>')">
<img <%=subsubsubcategoria%>" <%
if (rs4.EOF)
    response.write("src='empty.gif' ");
else
    response.write("src='plus.gif' ");
%>
    align="absbottom" border="0"></a>
<a
        class="subcategory"
        target="mainframe"
href="http://localhost/apsi/prot11_v01.asp?pnorm=3&keylist=<%=subsubsubcategoria%>"><%=subsubsubcategoria%
></a>
<div style="display:none" id="div_asd<%=subsubsubcategoria%>"
<%
if (rs4.EOF)
    response.write("class='needdata' >");
else
    response.write("class='hasdata' >");

```

```

        %>
        <%
            while (!rs4.EOF)
            {
                subsusubsubcategoria=rs4["subsusubsubclasse"];
            }
        %>
        <div class="category">
            <img id="img_asd<%=subsusubsubcategoria%>" src="empty.gif" align="absbottom"
border="0"></a>
            <a class="subcategory" target="mainframe"
href="http://localhost/apsi/prot11_v01.asp?pnorm=3&keylist=<%=subsusubsubcategoria%>"><%=subsusubsubcate
goria%></a><br>
        </div>
    <%
        rs4.MoveNext();
    }
    %>
</div>
</div>
<%
    rs3.MoveNext();
}
%>
</div>
</div>
<%
    rs2.MoveNext();
}
%>
</div>
<%
    rs1.MoveNext();
}
%>
</div>
</div>
<%
    rs.MoveNext();
}
%>
</body>
</html>

```

APÊNDICE F

Este apêndice compreende a tradução dos elementos definidos na especificação do RSS 1.0. Tendo em conta, que a especificação do RSS 1.0 foi bastante referenciada ao longo desta dissertação e que os seus elementos são a base da estrutura de metadados proposta, decidiu-se incluir parte da sua tradução neste documento.

1. `<?xml version="1.0" ?>`

Um documento RSS é uma aplicação XML, no entanto não é exigido começar com uma declaração XML. A especificação do RSS 1.0 recomenda a sua utilização como boa prática e também para assegurar a compatibilidade com a versão do RSS 0,9 (que o exige).

Sintaxe: `<?xml version="1.0" ?>`

Requisito: Opcional

2. `<rdf:RDF>`

O nível mais externo da estrutura hierárquica de um documento RSS é constituído pelo elemento RDF. A abertura da etiqueta⁷⁶ RDF associa o prefixo do *namespace* "rdf:" (permite identificar o *namespace* que está a ser utilizado) com a sintaxe do esquema RDF e estabelece o esquema RSS 1.0 como o *namespace* definido por defeito para o documento. A especificação do RSS 1.0 indica que qualquer prefixo de *namespace* válido pode ser usado, no entanto é recomendada a utilização normativa do prefixo "rdf:". Todos os utilizadores que pretendem estabelecer compatibilidade com a versão RSS 0.9 têm de utilizar o prefixo "rdf:".

Sintaxe: `<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns="http://purl.org/rss/1.0/">`

Requisito: Obrigatório exactamente como é apresentado na sintaxe, permitindo a declaração de *namespace* adicionais.

Modelo: (*channel*, *image*?⁷⁷, *item*+⁷⁸, *textInput*?)

⁷⁶ do inglês *tag*.

⁷⁷ Na descrição do modelo, o símbolo "?" significa que o elemento ou o atributo é opcional.

⁷⁸ Na descrição do modelo, o símbolo "+" significa que é permitido "uma ou mais" instâncias deste elemento ou atributo.

3. <channel>

O elemento *channel* é o elemento nuclear de um documento RSS, contém elementos de metadados que descrevem o próprio canal, designadamente um título, uma breve descrição, e o URL do recurso que permite efectuar a ligação à origem do recurso descrito. O URL do recurso é descrito no atributo *rdf:about* e este tem de ser único. Deste modo, um canal é sempre identificado através de um URI. Normalmente, este URI corresponde ao URL da *homepage* do recurso que está a ser descrito ou então o URL onde o documento RSS é disponibilizado.

Sintaxe: <channel rdf:about="{recurso}">

Requisito: Obrigatório

Atributo exigido: *rdf:about*

Modelo: (*title, link, description, image?, items, textinput?*)

Exemplo:

```
<channel rdf:about=" http://repositorio.apsi.pt:8080/index.jsp">
  <title> Repositório da APSI</title>
  <link> http://repositorio.apsi.pt:8080/index.jsp</link>
  <description>
Repositório institucional da Associação Portuguesa de Sistemas de Informação.
  </description>

  <image rdf:resource="http://xml.com/universal/images/ reposi.jpg" />

  <items>
  <rdf:Seq>
    <rdf:li resource=" http://hdl.handle.net/2287/57" />
    <rdf:li resource=" http://hdl.handle.net/2287/81" />
  </rdf:Seq>
  </items>

  <textinput rdf:resource="http://search.repositorio.apsi.pt" />

</channel>
```

3.1 <title>

Este elemento permite descrever o título do elemento *channel*.

Sintaxe: <title>{channel_title}</title>

Requisito: Obrigatório

Modelo: (#PCDATA)

Comprimento máximo (sugerido): 40 (caracteres)

3.2 <link>

Este elemento descreve o URL do elemento *channel*. É o elemento *link* que permite a ligação do título do *channel* até à origem da informação que está a ser descrita.

Sintaxe: <link>{channel_link}</link>

Requisito: Obrigatório

Modelo: (#PCDATA)

Comprimento máximo (sugerido): 500

3.3 <description>

Este elemento contém um breve sumário do conteúdo do elemento *channel*, designadamente funções, fonte, etc.

Sintaxe: <description>{channel_description}</description>

Requisito: Obrigatório

Modelo: (#PCDATA)

Comprimento máximo (sugerido): 500

3.4 <image>

De acordo com a especificação do RSS 1.0, este elemento estabelece a associação RDF entre o elemento opcional *image* e o elemento *channel* RSS que está a ser descrito. O `rdf:resource` {image_uri} deve ser igual ao atributo `rdf:about` do elemento *image* {image_uri}.

As propriedades deste elemento são definidas na especificação do RSS 1.0 da seguinte forma:

Sintaxe: <image rdf:resource="{image_uri}" />

Requisito: Só é obrigatório se o elemento *image* for utilizado

Modelo: vazio

3.5 <items>

Este elemento contém uma tabela de conteúdos RDF, que associa os itens dos documentos com o elemento *channel* RSS que está a ser descrito. O *rdf:resource* de cada artigo {item_uri} deve ser igual ao *rdf:about* do respectivo *item* do artigo {item_uri}.

Na especificação do elemento *items* é utilizado o *container* RDF Seq (sequência) invés do RDF Bag, uma vez que o RDF Seq permite definir uma ordem dos itens dos artigos que compõem o elemento *channel*.

Sintaxe: <items><rdf:Seq><rdf:li resource="{item_uri}" /> ... </rdf:Seq></items>

Requisito: Obrigatório

3.6 <textinput>

Este elemento estabelece uma associação RDF entre o elemento opcional *textInput* e o *channel* RSS que está a ser descrito. No elemento {textInput_uri} o atributo *rdf:resource* deve ser igual ao *rdf:about* do elemento *textInput* {textInput_uri}.

Sintaxe: <textinput rdf:resource="{textInput_uri}" />

Requisito: Só é obrigatório se o elemento *textInput* for utilizado

Modelo: vazio

4. <image>

O elemento *image* está associado ao elemento *channel*. Esta imagem deve estar definida num formato que suporte a maioria dos *Web browsers*.

Sintaxe: <image rdf:about="{image_uri}">

Requisito: Este elemento é opcional, no entanto, no caso de ser utilizado, este deve estar presente no elemento *channel*.

Atributo exigido: *rdf:about*

Modelo: (*title, url, link*)

Exemplo:

```
<image rdf:about="http://xml.com/universal/images/xml_tiny.gif">
  <title>XML.com</title>
  <link>http://www.xml.com</link>
  <url>http://xml.com/universal/images/xml_tiny.gif</url>
</image>
```

4.1 <title>

Este elemento permite descrever o título que está associado à imagem do elemento *channel*.

Sintaxe: <title>{image_alt_text}</title>

Requisito: Só é obrigatório se o elemento *image* for utilizado

Modelo: (#PCDATA)

Comprimento máximo (sugerido): 40

4.2 <URL>

Este elemento permite descrever o URL da imagem que está associada ao elemento *channel*.

Sintaxe: <url>{image_url}</url>

Requisito: Só é obrigatório se o elemento *image* for utilizado

Modelo: (#PCDATA)

Comprimento máximo (sugerido): 500

4.3 <link>

Este elemento contém o URL que permite estabelecer a ligação à imagem do canal.

Sintaxe: <link>{image_link}</link>

Requisito: Só é obrigatório se o elemento *image* for utilizado

Modelo: (#PCDATA)

Membro de: *image*

Comprimento máximo (sugerido): 500

5. <item>

O elemento *item* está normalmente associado a um artigo de uma notícia ou de um *blog*, no entanto devido à extensibilidade modular que caracteriza o RSS 1.0, um *item* refere-se a qualquer objecto que seja identificado através de um URI [Begeg-Dov *et al.*, 2000].

Um documento RSS deve conter no mínimo um elemento *item*. Apesar da especificação do RSS 1.0 não impor um limite máximo é recomendado o número máximo de 15 artigos por documento RSS de modo a permitir a compatibilidade com as versões RSS 0,9 e o RSS 0,91.

É o URI que identifica o elemento *item*. A identificação do {*item_uri*} deve ser único relativamente a qualquer atributo *rdf:about* definido no documento RSS. A especificação do RSS 1.0 recomenda que o elemento {*item_uri*} deve ser igual ao valor definido no sub-elemento <*link*> do elemento <*item*>.

Sintaxe: <item rdf:about="{*item_uri*}">

Requisito: A cardinalidade deste elemento deve ser maior ou igual a um.

Recomendação para garantir a compatibilidade com as versões 0.9x: De 1 a 15

Atributo exigido: *rdf:about*

Modelo: (*title, link, description?*)

Exemplo:

```
<item rdf:about=" http://hdl.handle.net/2287/20">
  <title> Reconhecimento de Voz - Voice Car System (VCS)</title>
  <link> http://hdl.handle.net/2287/20</link>
  <description>Os avanços tecnológicos possibilitam que os computadores reconheçam a voz humana e a interpretem, de forma a executarem determinadas tarefas previamente definidas. De modo a demonstrar a aplicabilidade destas novas tecnologias desenvolvemos uma aplicação que tem por objectivo a criação de uma interface de reconhecimento de voz, recorrendo para isso ao módulo Sensory Voice Extreme™ Toolkit. A aplicação em causa simula o controlo de algumas funções de um automóvel, activadas através do reconhecimento da voz humana. Numa utilização a nível real, a interacção do condutor perante a placa de reconhecimento da voz deverá ser efectuada através de um módulo que poderá ser integrado, por exemplo, no computador de bordo do automóvel.
  </description>
</item>
```

5.1 <title>

Este elemento descreve o título do elemento *item* que está a ser descrito.

Sintaxe: <title>{item_title}</title>

Requisito: Obrigatório

Modelo: (#PCDATA)

Comprimento máximo (sugerido): 100

5.2 <link>

Este elemento descreve o URL do elemento *item* que está a ser descrito.

Sintaxe: <link>{item_link}</link>

Requisito: Obrigatório

Modelo: (#PCDATA)

Comprimento máximo (sugerido): 500

5.3 <description>

Este elemento contém uma breve descrição ou resumo do *item* que está a ser descrito.

Sintaxe: <description>{item_description}</description>

Requisito: Opcional

Modelo: (#PCDATA)

Comprimento máximo (sugerido): 500

6. <textinput>

O elemento *textinput* dispõe um método que permite submeter dados através de um *form* para um URL arbitrário. O processamento de recepção apenas é executado através do método HTTP GET.

O elemento {*textinput_uri*} é identificado através de um URI, que deve ser único relativamente a qualquer outro atributo *rdf:about* definido no documento RSS. O elemento

{*textInput_uri*} deve ser idêntico ao valor definido no sub-elemento *<link>* do elemento *<textInput>*.

Sintaxe: *<textInput rdf:about="{textInput_uri}">*

Requisito: Este elemento é opcional, no entanto no caso de ser utilizado, este deve estar presente no elemento *channel*.

Recomendação para garantir a compatibilidade com as versões 0.9x: De 1 a 15

Atributo exigido: *rdf:about*

Modelo: (*title, description, name, link*)

Exemplo:

```
<textInput rdf:about="http://search.xml.com">
  <title>Search XML.com</title>
  <description>Search XML.com's XML collection</description>
  <name>s</name>
  <link>http://search.xml.com</link>
</textInput>
```

6.1 *<title>*

Este elemento contém um título descritivo para o campo *textInput*. Por exemplo: "Subscreva" ou "Procure"

Sintaxe: *<title>{textInput_title}</title>*

Descrição: Título do *textInput*

Requisito: Obrigatório

Modelo: (#PCDATA)

Comprimento máximo (sugerido): 40

6.2 *<description>*

Este elemento contém uma breve descrição do objectivo do campo de *textInput*. Por exemplo: "Subscreva a nossa *newsletter* para... " ou "Pesquisa sobre o nosso arquivo local..."

Sintaxe: *<description>{textInput_description}</description>*

Requisito: Obrigatório

Modelo: (#PCDATA)

Comprimento máximo (sugerido): 100

6.3 <name>

Este elemento contém o nome do campo *textinput*.

Sintaxe: <name>{textinput_varname}</name>

Requisito: Obrigatório

Modelo: (#PCDATA)

Comprimento máximo (sugerido): 500

6.4 <link>

Este elemento contém o URL para o qual uma submissão do elemento *textinput* será dirigida (utilizando o método GET).

Sintaxe: <link>{textinput_action_url}</link>

Requisito: Obrigatório

Modelo: (#PCDATA)

Comprimento máximo (sugerido): 500