



Universidade do Minho
Escola de Engenharia

Helder Noel Monteiro Firmino

**Organização e Publicação dos Termos do
Website da ANACOM sob uma Perspetiva
*Linked Open Data***

Tese de Mestrado
Engenharia e Gestão de Sistemas de Informação

Trabalho efetuado sob a orientação da
Professora Doutora Ana Alice Rodrigues Baptista

Dezembro de 2013

Declaração RepositoriUM

Declaro que concedo à Universidade do Minho e aos seus agentes uma licença não-exclusiva para arquivar e tornar acessível, nomeadamente através do seu repositório institucional, nas condições abaixo indicadas, a minha dissertação, no todo ou em parte, em suporte digital.

Declaro que autorizo a Universidade do Minho a arquivar mais de uma cópia da dissertação e sem alterar o seu conteúdo, converter a tese ou dissertação entregue, para qualquer formato de ficheiro, meio ou suporte, para efeitos de preservação e acesso.

Retenho todos os direitos de autor relativos à dissertação, e o direito de a usar em trabalhos futuros (como artigos ou livros).

Concordo que a minha dissertação seja colocada no repositório da Universidade do Minho com o seguinte estatuto:

- ✓ Disponibilização imediata do conjunto do trabalho para acesso mundial.

Universidade do Minho, 31/10/2013

Helder Noel Monteiro Firmino



AGRADECIMENTOS

Gostaria de expressar o meu reconhecimento às pessoas e instituições que, direta ou indiretamente, contribuíram para a realização deste trabalho académico.

À minha família e amigos pelo apoio e confiança incondicionais.

À minha orientadora, Professora Doutora Ana Alice Baptista, pela sua disponibilidade e sapiência.

À Laura Henriques e Susana Pedro, da ANACOM, pela disponibilidade e receptividade.

À Universidade do Minho, pelo acolhimento durante esta caminhada.

RESUMO

O propósito da Web Semântica é conseguir uma Web de dados totalmente ligada, isto é, numa perspectiva *Linked Open Data*. A Web Semântica deve garantir (estabelecendo padrões tecnológicos, vocabulários, linguagens lógicas, etc.) que os conteúdos publicados na Web sejam inteligíveis quer por agentes humanos, quer por agentes máquina.

Esta dissertação tem como objetivo responder à um problema delimitado, propondo uma solução no quadro da Web Semântica e suas tecnologias. Partindo-se de uma lista de termos em linguagem natural utilizados no *Website* da ANACOM (Autoridade Nacional de Comunicações), propomos uma organização de acordo com metodologias de construção de ontologias e vocabulários. Inspirámo-nos em duas metodologias, o *Ontology Development 101* e o *Process and Methodology for Core Vocabularies*. O vocabulário controlado resultante, tem como base tecnológica o modelo de organização de conhecimento, recomendado pelo W3C (World Wide Web Consortium), o SKOS (*Simple Knowledge Organization System*). Trata-se de uma tecnologia standard da W3C desde 2009, utilizada na criação de tesouros, esquemas de classificação, taxonomias, glossários e outros tipos de vocabulários controlados.

Como resultado da nossa intervenção, conseguimos organizar e codificar em SKOS, cerca de cinco centenas de termos identificados no *Website* da ANACOM. Para além da proposta do vocabulário controlado, passámos em revista às tecnologias e teorias que sustentam a temática da Web Semântica.

Palavras-chave: Web Semântica, SKOS, Vocabulários Controlados, *Linked Open Data*

ABSTRACT

The purpose of the Semantic Web is to achieve a web of data and integrally connected, i.e. under a Linked Open Data perspective. The Semantic Web should ensure (setting technological standards, vocabularies, logical languages, etc.) that the contents published on the Web are intelligible either by human actors or machine agents.

This thesis aims to answer a delimited problem, proposing a solution within the scope of Semantic Web and related technologies. Starting from a list of terms in natural language and without any structure, belonging to the website of ANACOM (Portuguese Communications Authority), we propose an organization in accordance with ontology and vocabularies development methodologies. We inspired ourselves in two methodologies for building ontologies, the Ontology Development 101 and the Process and Methodology for Core Vocabularies. The resulting controlled vocabulary is based on SKOS (Simple Knowledge Organization System), a knowledge organization system model, recommended by the W3C (World Wide Web Consortium). SKOS, which became a W3C recommendation in 2009, provides a model for expressing the basic structure and content of concept schemes such as thesauri, classification schemes, subject heading lists, taxonomies, folksonomies, and other similar types of controlled vocabulary.

As a result of our intervention, we organized and encoded in SKOS nearby five hundred terms, identified at the website of ANACOM. In addition to the proposal of a controlled vocabulary, we spent review to the technologies and theories that support the Semantic Web.

Keywords: Semantic Web, SKOS, Controlled Vocabularies, Linked Open Data

ÍNDICE

| | |
|-------------------------------------|------|
| AGRADECIMENTOS | iii |
| RESUMO | v |
| ABSTRACT | vii |
| ÍNDICE | ix |
| LISTA DE TABELAS | xiii |
| LISTA DE FIGURAS..... | xv |
| LISTA DE ACRÓNIMOS..... | xvii |
| OPÇÕES DE GRAFIA | xix |
| 1. INTRODUÇÃO..... | 1 |
| 1.1. Contextualização..... | 1 |
| 1.2. Problema..... | 3 |
| 1.3. Motivação | 4 |
| 1.4. Objetivos..... | 6 |
| 1.5. Publicações | 7 |
| 1.6. Estrutura da Dissertação | 7 |
| 2. FUNDAMENTOS..... | 9 |
| 2.1. Web Semântica..... | 9 |
| 2.2. Linked “Open” Data..... | 23 |

| | |
|---|----|
| 2.3. Vocabulários Controlados..... | 25 |
| 2.3.1. Conceito de Ontologia | 25 |
| 2.3.2. Conceitos de Vocabulário e de Vocabulário Controlado..... | 30 |
| 2.4. SKOS..... | 32 |
| 2.4.1. Elementos Fundamentais do SKOS..... | 32 |
| 2.4.2. Mapeamento entre Vocabulários | 37 |
| 3. ABORDAGEM METODOLÓGICA..... | 41 |
| 3.1. Seleção das Fontes de Pesquisa | 42 |
| 3.2. Princípios de Desenho de Ontologias e de Vocabulários Controlados..... | 44 |
| 3.2.1. Ontology Development 101 | 44 |
| 3.2.2. Process and Methodology for Core Vocabularies | 48 |
| 3.2.3. Methontology..... | 52 |
| 3.2.4. On-To-Knowledge..... | 54 |
| 3.2.5. DILIGENT..... | 55 |
| 3.2.6. Metodologia NeOn | 57 |
| 3.2.7. Outras Metodologias..... | 58 |
| 4. RESULTADOS | 65 |
| 4.1. Caracterização da ANACOM..... | 65 |
| 4.2. Proposta de organização dos termos em SKOS | 68 |
| 4.3 Discussão | 75 |

| | |
|--|-----|
| 5. CONCLUSÃO E TRABALHO FUTURO | 79 |
| 6. REFERÊNCIAS..... | 83 |
| 7. APÊNDICES..... | 91 |
| Apêndice A – Trechos de Código do Vocabulário Controlado no formato RDF/Turtle | 91 |
| Apêndice B – Prefixos de <i>Namespaces</i> mais comuns | 98 |
| Apêndice C – Exemplos de Vocabulários SKOS | 99 |
| Apêndice D – Exemplos de Fontes de <i>Linked Data</i> | 100 |
| 8. ANEXOS..... | 107 |
| Anexo A – Lista inicial dos termos do <i>website</i> da ANACOM em inglês | 107 |
| Anexo B – Lista inicial dos termos do <i>website</i> da ANACOM em português..... | 111 |

LISTA DE TABELAS

| | |
|---|----|
| Tabela 1 - The Five Star Scheme (Berners-Lee, 2010) | 2 |
| Tabela 2 - Exemplos de triplos (Allemang & Hendler, 2011)..... | 14 |
| Tabela 3 - Especificação explícita de uma conceptualização (Moreira, 2003) | 27 |
| Tabela 4 - Propriedades dos Metadados <i>Dublin Core</i> (Baker et al., 2012) | 50 |
| Tabela 5 - Prefixos de <i>Namespaces</i> mais comuns..... | 98 |
| Tabela 6 - Vocabulários publicados em SKOS | 99 |

LISTA DE FIGURAS

| | |
|---|----|
| Figura 1 - Níveis de Interoperabilidade para Metadados Dublin Core (Nilsson et al., 2009) | 4 |
| Figura 2 - Web 3.0, web 2.0 e tecnologias web semânticas (James Hendler, 2009) | 10 |
| Figura 3 - Arquitetura da Web Semântica | 12 |
| Figura 4 - Um grafo RDF descrevendo Eric Miller (Manola & Miller, 2004) | 15 |
| Figura 5 -Arquitetura da Web Semântica, versão 2005 (Pereira, 2007) | 22 |
| Figura 6 - 5 Star Deployment Scheme (adaptação de Michael Hausenblas, 2012) | 24 |
| Figura 7 - Interoperabilidade entre vocabulários (Pastor-Sánchez et al., 2012) | 38 |
| Figura 8 - Methontology - Processo de Desenvolvimento de Ontologias (Cardoso, 2010) | 53 |
| Figura 9 - Ciclo de Vida da Ontologia On-To-Knowledge (Staab et al., 2001) | 55 |
| Figura 10 - Ciclo de Vida da Ontologia DILIGENT (Pinto et al., 2004) | 56 |
| Figura 11 - Cenários da Metodologia NeOn (Baonza, 2010) | 57 |
| Figura 12 - Hierarquia de Classes do VC Proposto | 69 |
| Figura 13 - Lista de Termos | 70 |
| Figura 14 - Etiquetagem do termo EUA | 71 |
| Figura 15 - Vista OntoGraf do termo EUA | 72 |
| Figura 16 - Relação semântica hierárquica (propriedade skos:broader) | 72 |
| Figura 17 - Relação semântica hierárquica (propriedade skos:narrower) | 73 |
| Figura 18 - Protégé SPARQL <i>Plug-in</i> | 73 |

| | |
|---|-----|
| Figura 19 - Interrogação SPARQL com a propriedade <code>skos:broader</code> | 74 |
| Figura 20 - Vista Ontograf de todos os termos..... | 76 |
| Figura 21 - Web of Data (Tarasova, 2012)..... | 100 |
| Figura 22 - Artigo no Wikipedia | 101 |
| Figura 23 - Extração de informação a partir da <i>infobox</i> de um artigo do Wikipédia | 101 |
| Figura 24 - Desambiguação de entidades..... | 102 |
| Figura 25 - Pesquisa no Linked MDB, via SPARQL Endpoint (Hassanzadeh, 2011)..... | 104 |
| Figura 26 - Mapa interativo de sismos mundiais (Hassanzadeh, 2011) | 105 |

LISTA DE ACRÓNIMOS

| | |
|---------------|--|
| ADMS | Asset Description Metadata Schema |
| ACM | Association for Computing Machinery |
| ANACOM | Autoridade Nacional de Comunicações |
| CCS | Computing Classification System |
| CERN | Organização Europeia de Pesquisa Nuclear |
| CCIO | Classificação Portuguesa do Consumo Individual por Objectivo |
| CSV | Comma Separated Values |
| DAML | DARPA Agent Markup Language |
| DCAP | Dublin Core Application Profiles |
| DCMI | Dublin Core Metadata Initiative |
| DIGIT | Directorate Generale for Informatics |
| DOAP | Description of a Project |
| ERCIM | European Research Consortium for Informatics |
| INE | Instituto Nacional de Estatística |
| ISA | Interoperability Solutions for European Public Administrations |
| JCR | Journal Citation Reports |
| JSON | JavaScript Object Notation |
| KEIO | Keio University (KEIO Gijuku Daigaku) |
| LCSH | Library of Congress Subject Headings |
| LOD | Linked Open Data |
| MIT | Massachussets Institute of Technology |
| NAICS | North American Industry Classification System |
| NAL | National Agriculture Library |
| NDLTD | Networked Digital Library of Theses and Dissertations |
| OIL | Ontology Interchange Language |
| OWL | Web Ontology Language |
| RDF | Resource Description Framework |
| RDFS | RDF Schema |
| RCAAP | Repositório Científico de Acesso Aberto de Portugal |
| SciELO | Scientific Electronic Library Online |
| SIOC | Semantically-Interlinked Online Communities Project |
| SKOS | Simple Knowledge Organization System |
| SMORE | Semantic Markup Ontology and Resource Description Framework Editor |

| | |
|---------------|--|
| SPARQL | SPARQL Protocol and RDF Query Language |
| SQL | Structured Query Language |
| SWAN | Semantic Web at ANACOM |
| SWOOP | Semantic Web Ontology Editor |
| Turtle | Terse RDF Triple Language |
| TOVE | Toronto Virtual Enterprise |
| URI | Uniform Resource Identifier |
| W3C | World Wide Web Consortium |
| WGS | World Geodetic System |
| WHATWG | Web Hypertext Application Technology Working Group |
| WWW | World Wide Web |
| XML | eXtensible Markup Language |

OPÇÕES DE GRAFIA

Durante a execução deste trabalho sentimos a necessidade de fazer algumas opções de escrita e de formatação, com o intuito de suprir algumas lacunas onde a norma institucional de formatação de teses era omissa.

Seguidamente listamos as opções tomadas:

- **Tipo de letra** - utilizada a fonte serifada, Garamond;
- **Itálicos** - usados para fazer referência a palavras ou locuções estrangeiras;
- **Avanço de parágrafos** - 0,63 cm;
- **Código em SKOS, HTML, RDF/XML** - é usada a fonte não serifada, NewsGoth, tamanho 10;
- **Aspas francesas (« »)** - usadas em transcrições e citações de outros autores;

1. INTRODUÇÃO

Neste capítulo introdutório é feita uma contextualização do tema a abordar. Será apresentado o problema que pretendemos resolver, serão enunciados os objetivos propostos. Daremos igualmente nota dos artigos, submetidos a conferências. E para finalizar o capítulo, é feita uma referência à estrutura da dissertação.

1.1. Contextualização

O surgimento da Web, em 1991, no seio do CERN (Organização Europeia de Pesquisa Nuclear) em Genebra, foi um marco na história da humanidade, alterando a forma de comunicação, democratizando o acesso à informação e ao conhecimento. Com a Web, a humanidade passa para um estágio de inteligência coletiva, onde todos, com exceção dos infoexcluídos, têm a possibilidade de partilha de informação (Lévy, 1999)¹. A Web foi inventada pelo visionário cientista inglês, Tim Berners-Lee.

A Web cresceu e presentemente estamos na era da Web 3.0 (ou simplesmente Web Semântica), sucedendo à Web Social (Web 2.0).

Tim Berners-Lee e o W3C (World Wide Web Consortium) estão agora envolvidos em outro projeto: a Web de dados. «O objetivo final da Web de dados é permitir que os computadores possam realizar um trabalho mais útil e desenvolver sistemas que possam apoiar interações confiáveis na rede»(Berners-Lee, 2010).

Baptista(2002), assevera que o W3C foi criado em Outubro de 1994, com o objetivo de «desenvolver protocolos comuns para promover a evolução e assegurar a interoperabilidade da Web». Segundo apuramos no sítio do W3C, a organização é promovida pelas seguintes instituições: MIT (Massachussets Institute of Technology), KEIO University, ERCIM (European Research Consortium for Informatics) e o Beihang University (Beijing University of Aeronautics & Astronautics). Presentemente o W3C

¹ Pierre Lévy, filósofo francês, defensor da Inteligência Coletiva e um dos maiores estudiosos da Internet e suas repercussões na sociedade atual

dispõe de 384 membros que estão diretamente envolvidos nos diversos projetos desta organização internacional.²

A missão do W3C é «levar a World Wide Web ao seu máximo potencial, desenvolvendo protocolos e diretrizes que garantam o crescimento a longo prazo da Web». Assenta em princípios *Open Standards* e *Design Principles*. Os seguintes princípios de *Design* guiam a atividade do W3C: 1) Web para Todos (*Web for All*); 2) Web em Tudo (*Web on Everything*). A visão que o W3C almeja para a Web é a **participação** de qualquer pessoa, promovendo a **partilha** de conhecimento e construindo assim a **confiança** à uma escala global.³

Tim Berners-Lee, sugeriu um esquema de implementação Cinco Estrelas para Dados Abertos (Tabela 1). De acordo com Berners-Lee (2010), a Web Semântica não é apenas colocar de dados na Web. Trata-se de fazer as ligações, para que uma pessoa ou máquina possa explorar a web de dados. Tecnologias *Linked Data* permitem a descrição e relação explícita de recursos por meio de URIs (*Uniform Resource Identifiers*), o que significa que as relações implícitas com outros recursos podem ser encontradas ou inferidos. Quando falamos em dados vinculados abertos, estamos então a referir à expressão *Linked Open Data* (LOD).

Tabela 1 - The Five Star Scheme (Berners-Lee, 2010)

| | |
|-------|---|
| ★ | 1 Available on the Web (whatever format) but with an open licence, to be Open Data |
| ★★ | 2 Available as machine-readable structured data (e.g. excel instead of image scan of a table) |
| ★★★ | 3 Use non-proprietary formats (e.g. CSV instead of excel) |
| ★★★★ | 4 All the above plus use open standards from W3C (RDF and SPARQL) to identify things (URIs), so that people can point at your stuff |
| ★★★★★ | 5 All the above, plus link your data to other people's data to provide context |

² <http://www.w3.org/Consortium/facts>

³ <http://www.w3.org/Consortium/mission.html>

As regras definidas para a Web de Dados são bem simples:

- 1) Usar URIs como nomes para as coisas (*Cool* URIs para a Web Semântica são aqueles que não se alteram);
- 2) Usar http URIs, para que as pessoas possam procurar esses nomes;
- 3) Quando alguém procura um URI, fornecer informações úteis, usando padrões adequados (RDF, SPARQL);
- 4) Incluir ligações para outros URIs, de modo a que as pessoas possam descobrir mais coisas (Méndez & Greenberg, 2012).

1.2. Problema

A ANACOM é a reguladora e supervisora das comunicações postais e das comunicações eletrónicas em Portugal. Como já foi anteriormente mencionado, este trabalho pretende dar continuidade à primeira fase do projeto SWAN onde foram analisados os requisitos da presença de web semântica na ANACOM.

Desde alguns anos a esta parte que a ANACOM tem vindo a desenvolver um tipo de vocabulário controlado que consiste numa lista de termos controlados, sem estrutura e sem relações específicas para as propriedades Dublin Core em uso.

A lista de termos que a ANACOM tem vindo a utilizar é uma lista que contém todos os termos que utilizam no seu website (aproximadamente cinco centenas de termos). Esta lista é atualizada de acordo com os conteúdos introduzidos no *website*. Se quisermos utilizar uma analogia, podemos afirmar que eles encontram-se todos amontoados num balde e sem relacionamentos, isto é, de forma isolada.

É mantida em linguagem natural, duas listas de termos, uma em inglês e outra em português. Podemos afirmar que não estão separados segundo a sua função descritiva; Não estão relacionados uns com os outros; Não estão relacionados com outros termos exteriores à ANACOM; Não estão disponíveis com um todo para outras entidades utilizarem; E não estão codificados numa linguagem normalizada.

Uma das exigências do projeto é organizar essa lista em um ou mais vocabulários controlados, relacioná-los com outros vocabulários LOD, codificá-los em SKOS e publicá-los em pelo menos duas línguas (Português e Inglês).

Relativamente à questão (ou questões) de investigação, por se tratar de um tipo de investigação aplicado (onde temos um problema a resolver), não se aplica ao nosso estudo.

1.3. Motivação

Esta dissertação de mestrado surge na sequência de um trabalho académico, efetuado no âmbito da unidade curricular Web Semântica, do 1º ano do plano de estudos do Mestrado em Engenharia e Gestão de Sistemas de Informação. E também para dar seguimento ao projeto SWAN (acrónimo em inglês, de Web Semântica na ANACOM). Na verdade, a ANACOM (Autoridade Nacional de Comunicações), já tem todos os recursos do seu website, descritos, utilizando o Dublin Core (DC) e codificados em HTML, o que corresponde ao nível 1 de Interoperabilidade para metadados Dublin Core (Nilsson, Baker, & Johnston, 2009) e ao nível 3 do Esquema de Cinco Estrelas (Berners-Lee, 2010), descrito na secção anterior.

| |
|---|
| <p>➤ 4: Description Set Profile Interoperability</p> <ul style="list-style-type: none">• Shared formal vocabularies and constraints in records |
| <p>➤ 3: Description Set syntactic interoperability</p> <ul style="list-style-type: none">• Shared formal vocabularies in exchangeable records |
| <p>➤ 2: Formal semantic interoperability</p> <ul style="list-style-type: none">• Shared formal vocabularies based on formal semantics |
| <p>➤ 1: Shared term definitions</p> <ul style="list-style-type: none">• Shared formal vocabularies defined in natural language |

Figura 1 - Níveis de Interoperabilidade para Metadados Dublin Core (Nilsson et al., 2009)

Seguidamente, e de acordo com o Dublin Core⁴, caracterizaremos sucintamente cada um dos níveis constantes na Figura 1.

Nível 1: Definições Partilhadas de Termos

Este nível corresponde à utilização de definições em linguagem natural dos termos Dublin Core.

- A utilização dos URIs do termo não é um requisito – os termos podem ser referenciados de qualquer forma.
- A conformidade com os domínios e contradomínios especificados dos termos não é um requisito.
- A conformidade com o Modelo Abstrato da DCMI não é um requisito.
- Historicamente falando, a interoperabilidade informal com o “Dublin Core” significou a utilização dos quinze elementos , embora o conceito de interoperabilidade informal seja, em princípio, aplicável a quaisquer termos de metadados definidos em linguagem natural.

Nível 2: Interoperabilidade Semântica Formal

Este nível corresponde à utilização implícita ou explícita da semântica RDF subjacente aos termos DCMI. Assim, qualquer utilização dos termos necessita de ser precisa na sua conformidade com o modelo RDF e com os domínios e contra-domínios dos termos.

- Enquanto a especificação/aplicação não necessita explicitamente de codificar os dados utilizando URIs, deve ser possível inferir os URIs dos termos.
- É um requisito a conformidade com os domínios e contra-domínios especificados dos termos.
- Conformidade com a semântica de sub-propriedades das propriedades utilizadas.

⁴ http://wiki.dublincore.org/index.php/PT_Traduções/Níveis_de_Interoperabilidade

- Conformidade com o Modelo Abstrato da DCMI não é um requisito.

Nível 3: Interoperabilidade Sintática de Conjuntos de Descrições

Este nível corresponde à utilização explícita do Modelo Abstrato da DCMI nos metadados.

- Os metadados devem ser estruturados utilizando as noções de Descrição e Conjuntos de Descrições do Modelo Abstrato da DCMI.
- As Declarações (*Statements*) numa Descrição devem utilizar a estrutura definida pelo Modelo Abstrato da DCMI.

Nível 4: Interoperabilidade do Perfil de Conjuntos de Descrições

A especificação "*Description Set Profiles: A constraint language for Dublin Core Application Profiles*" fornece um modelo de informação e expressão XML de restrições estruturais num conjunto de descrições. Uma aplicação tal como o perfil de aplicação da Scholarly Works (Eprints) pode ser enunciada como “Perfil de Conjuntos de Descrições Interoperável” se fornece restrições formais num Conjunto de Descrições que são compatíveis com os da especificação “*Description Set Profiles*”. Uma especificação relacionada, o Enquadramento de Singapura para Perfis de Aplicação Dublin Core (DCAP), delinea um pacote de elementos documentais necessários à apresentação de uma aplicação de metadados para interoperabilidade e reutilização máximas – elementos tais como Requisitos Funcionais, um Modelo de Domínio e um Perfil de Conjuntos de Descrições cobrindo a totalidade do conjunto de metadados.

1.4. Objetivos

Os objetivos que pretendemos alcançar com este projeto são os seguintes:

- Analisar e organizar os termos que estão sendo usados para descrever recursos no sítio da ANACOM;
- Propor um ou mais do que um vocabulário controlado como resultado da organização dos termos;

- Relacionar o vocabulário controlado resultante com outros vocabulários existentes;
- Codificar o vocabulário controlado resultante em SKOS (*Simple Knowledge Organization System*) e em pelo menos, dois idiomas (português e inglês), e sob uma perspectiva LOD cinco estrelas.

1.5. Publicações

Paralelamente ao trabalho de dissertação de mestrado, estivemos envolvidos no processo de escrita científica com um *extended abstract* e um poster.

O *extended abstract*, com o título “*Making sense of a flat list of terms into Linked Open Data SKOS Vocabularies⁵*”, foi submetido à 17th International Conference on Electronic Publishing, ELPUB 2013, e foi aceite e publicado nas atas da referida conferência. O seu conteúdo foi apresentado oralmente durante a conferência que teve lugar no Blekinge Tekniska Högskola (BTH), em Karlskrona, Suécia.

O Poster, cujo título é “*Upgrading a flat list of terms into a Linked Open Data Structure for the Portuguese National Authority on Communications (ANACOM)*”, foi submetido ao International Conference on Dublin Core and Metadata Applications, DC-2013, que se desenrolou em Setembro do corrente ano nas instalações do Instituto Superior Técnico em Lisboa. O poster foi aceite mas, infelizmente, não conseguimos marcar presença no evento para a sua apresentação.

1.6. Estrutura da Dissertação

Esta dissertação está organizada em cinco capítulos:

- No capítulo 1, **Introdução**, é feita uma contextualização do tema. É igualmente apresentado o problema e a motivação, seguem-se os objetivos, que precedem a secção das publicações, e por último é apresentada a estrutura da dissertação;

⁵ Disponível em acesso livre em <http://iospress.metapress.com/content/04227402576tp123/fulltext.pdf>

- No capítulo 2, **Fundamentos**, apresentamos uma revisão de literatura sobre os conceitos associados à temática da Web Semântica;
- No capítulo 3, centramos na **Abordagem Metodológica**, onde são descritas as metodologias de construção de ontologias e vocabulários.
- No capítulo 4, **Resultados** disponibilizamos o Vocabulário Controlado resultante, são discutidos e relacionados os resultados atingidos com outros vocabulários.
- No capítulo 5, **Conclusão e Trabalho Futuro**, é apresentada uma síntese da dissertação, a discussão do trabalho e são sugeridas ideias para trabalho futuro.

2. FUNDAMENTOS

Este capítulo tem como propósito discorrer sobre os temas-chave abarcados na presente dissertação. Está dividido em quatro secções: na primeira secção abordaremos o tema da Web Semântica; na segunda o enfoque vai para o *Linked “Open” Data*; na terceira secção centraremos nos vocabulários controlados; e por último descreveremos o SKOS (*Simple Knowledge Organization System*).

2.1. Web Semântica

Com a presente secção pretendemos apresentar os fundamentos teóricos da Web Semântica e a sua evolução.

A partir do nome Web Semântica pode-se facilmente depreender que se trata de algo relacionado com a *World Wide Web*, e que está relacionada com semântica. Semântica, por sua vez, tem a ver com a compreensão da natureza do significado, mas até mesmo a palavra semântica tem uma série de significados (Allemang & Hendler, 2011).

Segundo o Dicionário da Porto Editora⁶, Semântica é um nome feminino e etimologicamente vem do grego *semantiké* [tékhne], “a arte da significação”. No campo da Linguística, Semântica «é a disciplina da linguística que se ocupa da significação das palavras e expressões linguísticas bem como das relações de sentido que estas estabelecem entre si». No ramo da Filosofia e Lógica, representa «a parte da semiótica que estuda as relações dos signos com os objetos que eles representam».

O criador da *World Wide Web* (daqui em diante apenas referida como Web), Tim Berners-Lee, idealizou a evolução da Web com a estruturação e adição de semântica ao conteúdo da Web atual. Tal evolução não passa pela criação de uma nova Web, mas sim pela extensão da Web atual, fornecendo significado bem definido às informações, permitindo que as pessoas e computadores possam trabalhar em cooperação. A esse novo

⁶ Disponível na www <URL: <http://www.infopedia.pt/pesquisa-global/sem%C3%A2ntica>>, consultado a 29.12.2012.

paradigma e conceito de Web em que o conteúdo possui informação estruturada e com semântica associada, cunhou-se o nome de Web Semântica (Prazeres, 2009).

De acordo com Thuraisingham (2002), tem havido discussões sobre as diferenças entre Web e Web Semântica. Alguns dizem que a Web é um sistema em que os seres humanos conseguem visualizar e ler as páginas da Web, enquanto que a Web Semântica é um sistema onde a máquina lê e entende as páginas da Web. Outros afirmam que a Web de hoje é a Web Semântica de ontem e que a Web de amanhã é a Web Semântica de hoje.

No nosso entendimento a afirmação do autor Thuraisingham (2002), traduz a evolução natural da Web, uma vez que passamos da era da Web Sintática para a era da Web Semântica, onde é embutido aos computadores conhecimento e capacidade de decisão autónoma.

Num artigo publicado na revista IEEE Computer Society, Hendler (2009), fala da emergência da Web 3.0, onde afirma que a Web 3.0 é uma extensão da Web 2.0 adicionando tecnologias web semânticas e *linked data* (Figura 2).

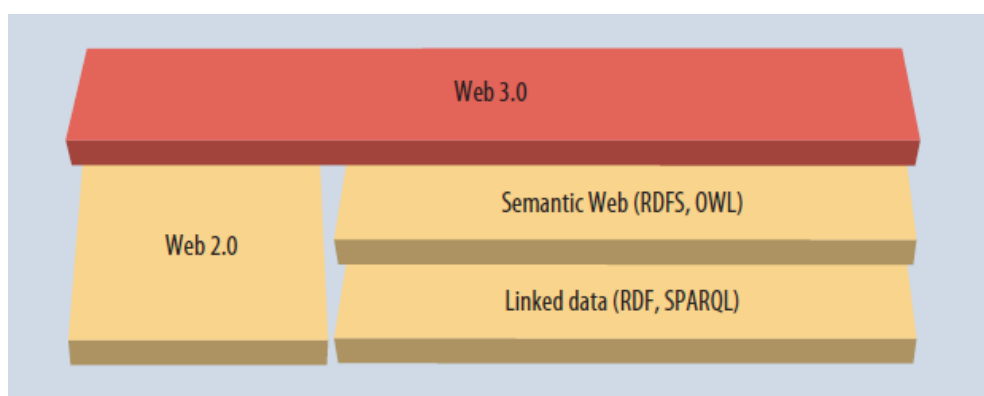


Figura 2 - Web 3.0, web 2.0 e tecnologias web semânticas (James Hendler, 2009)

A ideia de que a informação Web era acessível apenas a pessoas altamente qualificadas, profissionais de Tecnologias de Informação, é uma questão ultrapassada. Presentemente estamos na era da democratização da Web.

Essencial para a noção de Web, está associada a ideia de comunidade aberta, isto é, cada um pode contribuir com suas ideias para todos, para que qualquer um possa aceder. Essa ideia é consubstanciada através do slogan AAA “*Anyone can say Anything about Any topic*” (Allemang & Hendler, 2011).

Catarino (2009), menciona na sua tese de doutoramento que o «conceito da Web Semântica foi referido em 2001 quando da publicação de um artigo na revista *Scientific American* por Tim Berners-Lee, James Hendler e Ora Lassila. O intento é desenvolver tecnologias, linguagens, padrões e recomendações que tornem a informação legível pelas máquinas».

Ainda segundo Catarino (2009), a Web Semântica é o nome genérico para representar o projeto da W3C que «pretende embutir inteligência e contexto nos códigos XML utilizados para realização de páginas Web», de forma a facilitar o intercâmbio de informações. E para atingir tal propósito é necessária «uma padronização de tecnologias, de linguagens e de metadados» para determinar regras comuns que facilitem o uso das informações armazenadas de maneira automática.

A Web Semântica é uma coleção de estruturas que trabalham juntas (ou interoperam entre si). Essas estruturas limitam o alcance de possíveis interpretações no sentido de permitir a transferência de significado ao longo da Web (Tennis, 2007).

Fahmi, Ellermann, Scholing, & Zhang (2008), defendem que a Web Semântica pode ser utilizada para introduzir relações significativas e explícitas entre documentos, com base no seu conteúdo.

A noção de Web semântica pode ser resumida segundo Goñi et al. (2002), citado por Gonçalves (2007), quando responde à seguinte questão: «como fazer com que os computadores entendam o conteúdo da Web? O primeiro passo será organizar e estruturar as informações e o segundo será adicionar semântica às informações da Web, de tal forma que os agentes de software possam compreendê-las».

Outro contributo sobre a importância da Web Semântica é-nos dado por Davies, Studer, & Warren (2006), em que advogam que a Web Semântica e as tecnologias web semânticas oferece-nos uma nova abordagem na gestão da informação e dos processos, em que o princípio fundamental é a criação e utilização de metadados semânticos.

Para a informação, metadados podem existir em duas dimensões. Por um lado, metadados podem descrever um documento; por exemplo uma página web, ou parte de um documento, onde poderemos indicar como exemplo, um parágrafo. Por outro lado,

metadados podem descrever entidades dentro de um documento; por exemplo, uma pessoa ou empresa. Em qualquer dos casos, o que importa é a característica semântica dos metadados, pois este nos informa sobre o conteúdo de um documento (p.ex. qual o assunto, ou relacionamentos com outros documentos) ou sobre uma entidade dentro de um documento (Davies et al., 2006).

A arquitetura da Web semântica (ver Figura 3) proposta pela W3C, ilustra as várias camadas para a Web, com suas linguagens e padrões. A estrutura da Web semântica é igualmente designada por “*Layer Cake*” ou *Semantic Web Stack*.

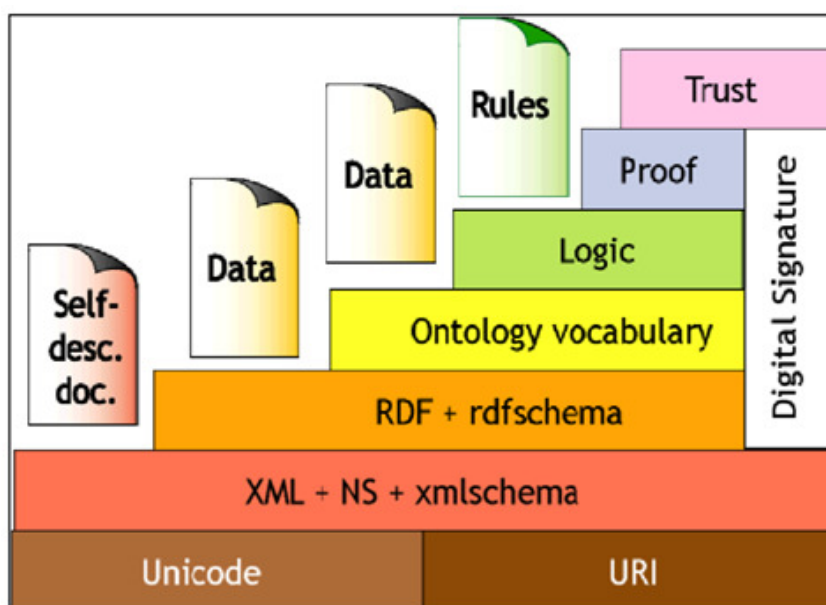


Figura 3 - Arquitetura da Web Semântica

Para Pereira (2009), o processo de criação de uma camada da web semântica em cima de outra camada, deve seguir dois princípios básicos:

- **Compatibilidade descendente** – Agentes com conhecimento total de uma camada, deverão ser capazes de interpretar e usar a informação disponibilizada pelas camadas inferiores. Por exemplo, agentes com conhecimento de OWL, poderão tirar total vantagem das informações escritas em RDF e RDF Schema;
- **Interpretação parcial ascendente** – Por outro lado, agentes cabalmente conscientes de uma camada, deverão no mínimo poder tirar vantagem parcial da informação dos níveis superiores. Por exemplo, um agente apenas ciente da semântica do RDF e RDF Schema, pode parcialmente interpretar conhecimento

escrito em OWL, ignorando os elementos que vão para além do RDF e RDF Schema.

Na base do “*cake*”, temos a camada Unicode e URI, que garantem o uso padronizado de caracteres (UNICODE) e uma forma exata para identificar as páginas da Web (URI – *Uniform Resource Identifier*). A camada XML com *namespace* e *schemas* permite a integração de *statements* da Web Semântica com outros padrões baseados em XML. Com RDF e RDFS é possível descrever recursos da Web que possuam identificadores URI e definir vocabulários que se relacionam com estes recursos. A camada Ontologia suporta a evolução de vocabulários que podem definir relações entre diferentes conceitos. A camada Assinatura Digital permite detectar alterações nos documentos. A camada Lógica possibilita a redação de regras enquanto a camada de Prova executa as regras e a camada de Confiança avalia se a prova está correta (Catarino, 2009).

A arquitetura da Web semântica tem sido alvo de atualizações constantes. Derivado à evolução do conhecimento e também devido à natureza volátil das tecnologias, torna-se necessário adequar a arquitetura aos novos padrões sintáticos e semânticos.

Em 2002, o W3C apresentou algumas mudanças que, apesar de pouco significativas, permitiram clarificar a arquitetura para a Web Semântica (Berners-Lee, 2002, citado por Gonçalves, 2007).

De acordo com Gonçalves (2007), «as alterações resumiram-se ao agrupamento de algumas tecnologias (as especificações XML e *xmleschema* foram agrupadas na camada XML que passou a incluir o *xmleschema*) e à subdivisão de outras (a camada RDF+RDF Schema foi subdividida na camada RDF Modelo & Sintaxe e na camada RDF Schema, enquanto que o *Ontology Vocabulary* foi subdividido em duas camadas: a camada ontologia e a camada regras). Foi também acrescentada transversalmente a criptografia para garantir, juntamente com a assinatura digital, a confidencialidade e a confiabilidade das informações na Web Semântica».

RDF (*Resource Description Framework*)

O *Resource Description Framework* é um *standard* da W3C, e conforme o documento RDF Primer, o *Resource Description Framework* (RDF) é uma linguagem para representação de

informação sobre recursos na Web. É particularmente indicada para representação de metadados relativos a recursos Web, tais como título, autor e data de modificação de uma página Web, ou ainda direitos de autor e licença de utilização de informações de um documento Web. (Manola & Miller, 2004).

RDF, RDFS e OWL são as linguagens de representação básicas da Web Semântica. O RDF serve como fundação, isto é, as linguagens da Web Semântica são construídas em cima do RDF. O RDF aborda uma questão fundamental da Web Semântica, que é a gestão de dados distribuídos. Todas as outras normas da Web Semântica, constroem sobre esta base de dados distribuída (Allemang & Hendler, 2011).

O RDF descreve recursos através de triplos. Em RDF um triplo é constituído por três partes: o recurso, a propriedade e o valor associado à propriedade. Em muita literatura consultada, alguns autores designam as partes constituintes de um triplo como sendo: Sujeito, Predicado e Objeto.

A Tabela 2 mostra-nos alguns triplos, onde o identificador para a linha é designado por sujeito (a ideia de sujeito é análoga ao da gramática). O identificador para a coluna é o predicado do triplo. O valor na célula é denominado de objeto do triplo.

Tabela 2 - Exemplos de triplos (Allemang & Hendler, 2011)

| Subject | Predicate | Object |
|----------------|------------------|---------------|
| Shakespeare | Wrote | King Lear |
| Shakespeare | Wrote | MacBeth |
| Anne Hathaway | Married | Shakespeare |
| Shakespeare | livedIn | Stratford |
| Stratford | IsIn | England |
| MacBeth | SetIn | Scotland |
| England | partOf | UK |
| Scotland | partOf | UK |

O RDF é baseado na ideia de identificação de recursos utilizando identificadores Web (os chamados *Uniform Resource Identifiers* ou URIs) e na descrição dos recursos através de propriedades simples e de valores associados a essas mesmas propriedades. Tudo o que se acabou de dizer, permite ao RDF representar *statements* (triplos) sobre recursos, como um grafo de nodos e arcos representando os recursos, suas propriedades e seus valores (Manola & Miller, 2004).



Figura 4 - Um grafo RDF descrevendo Eric Miller (Manola & Miller, 2004)

A Figura 4, representa um grafo, onde é possível verificar que existe uma pessoa (Person) identificada pelo URI, <http://www.w3.org/People/EM/contact#me>, cujo nome é Eric Miller, endereço eletrônico em@w3.org, e cujo título acadêmico é Dr.

Pode-se verificar ainda na Figura 4, que o RDF utiliza URIs para identificar:

- Literais, p. ex., *Eric Miller*, relacionado pela propriedade <http://www.w3.org/2000/10/swap/pim/contact#fullName>

- Tipo de coisas, p. ex., *Person*, relacionado pela propriedade
`http://www.w3.org/2000/10/swap/pim/contact#Person`
- Propriedades associadas às coisas, p.ex., *mailbox*, relacionado pela propriedade
`http://www.w3.org/2000/10/swap/pim/contact#mailbox`
- Valores associados às propriedades, p. ex. `mailto:em@w3.org` como valor da propriedade *mailbox* (RDF usa igualmente strings como "Eric Miller", e valores de outros tipos de dados como inteiros e data, como valores associados às propriedades)

Além dos grafos e dos triplos, o RDF fornece uma sintaxe baseada no XML, chamada de RDF/XML, para armazenar e exportar os grafos para outro formato. Podemos verificar no pequeno trecho de código em RDF/XML, correspondência ao grafo da Figura 4 (Manola & Miller, 2004).

```
<?xml version="1.0"?>
  <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
           xmlns:contact="http://www.w3.org/2000/10/swap/pim/contact#">
    <contact:Person rdf:about="http://www.w3.org/People/EM/contact#me">
      <contact:fullName>Eric Miller</contact:fullName>
      <contact:mailbox rdf:resource="mailto:em@w3.org"/>
      <contact:personalTitle>Dr.</contact:personalTitle>
    </contact:Person>
  </rdf:RDF>
```

Formatos RDF

Os autores Heath & Bizer (2011), salientam que é importante lembrar que o RDF não é um formato de dados, mas sim, um modelo de dados para descrever recursos sob a forma de triplos (sujeito, predicado e objeto). Antes de se publicar um grafo RDF na Web, a primeira coisa a fazer, é serializar o grafo utilizando uma sintaxe RDF. Dois formatos de serialização RDF, RDF/XML e RDFa, foram padronizados pelo W3C. Além disso, vários outros formatos de serialização não-padronizados são utilizados para satisfazer necessidades mais específicas.

RDF/XML

A Sintaxe XML/RDF é amplamente utilizada para publicar dados interligados na Web. No entanto, a sintaxe também é vista como sendo de difícil leitura e escrita para os seres humanos, daí a necessidade de considerar a utilização de outras serializações na gestão de dados e manutenção de *workflows* que envolvam a intervenção humana (Heath & Bizer, 2011).

RDFa

RDFa é um formato de serialização que incorpora triplos RDF em documentos HTML. Os dados RDF não estão incorporados nos comentários dentro do documento HTML, como foi o caso com algumas primeiras tentativas de misturar RDF e HTML, mas está entrelaçada no DOM (*Document Object Model*) HTML. Isso significa que o conteúdo existente em uma página pode ser marcado com RDFa modificando o código HTML (Heath & Bizer, 2011).

Exemplo de sintaxe RDFa:

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML+RDFa 1.0//EN"
    "http://www.w3.org/Markup/DTD/xhtml-rdfa-1.dtd">
<html xmlns="http://www.w3.org/1999/xhtml"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:foaf="http://xmlns.com/foaf/0.1/">

  <head>
    <meta http-equiv="Content-Type" content="application/xhtml+xml; charset=UTF-8"/>
    <title>Profile Page for Dave Smith </title>
  </head>

  <body>
    <div about="http://biglynx.co.uk/people#dave-smith" typeof="foaf:Person">
      <span property="foaf:name">Dave Smith
    </div>
  </body>

</html>
```

Microdados

Microdados (em inglês, Microdata), trata-se de uma especificação HTML, desenvolvida pela WHATWG, que é utilizada para misturar metadados com conteúdos existentes nas páginas Web⁷.

Microdados surgiu com o intento de proporcionar uma forma mais simples na rotulagem de elementos HTML com *tags* inteligíveis pelas máquinas, em vez da utilização das abordagens RDFa e Microformats⁸.

Segundo Hickson (2011), «microdados permitem que grupos agrupados em pares nome-valor, sejam adicionados aos documentos, em paralelo com o conteúdo existente. Os grupos são chamados de *items*, e cada par nome-valor é uma propriedade. *Items* e propriedades são representadas por elementos regulares. Para criar um *item*, é usado o atributo *itemscope*. Para adicionar uma propriedade a um *item*, o atributo *itemprop* é usado num dos descendentes do *item*».

Os vocabulários *microdata* fornecem a semântica ou o significado de um *item*. Os desenvolvedores Web poderão criar um vocabulário ou recorrer à vocabulários disponíveis na Web. Uma coleção de vocabulários de marcação usualmente utilizados são fornecidos por esquemas disponíveis no Schema.org⁹. Os principais operadores de motores de busca como o Bing, Google, e Yahoo! contam com o microdados para melhorar os resultados das pesquisas.

O exemplo a seguir, retirado da Wikipédia, ilustra a utilização de microdados.

```
<section itemscope itemtype="http://schema.org/Person">  
  Hello, my name is  
  <span itemprop="name">John Doe</span>,  
  I am a
```

⁷ *Microdata — HTML5 Draft Standard* (<http://www.whatwg.org/specs/web-apps/currentwork/multipage/microdata.html>)

⁸ [http://en.wikipedia.org/wiki/Microdata_\(HTML\)](http://en.wikipedia.org/wiki/Microdata_(HTML))

⁹ <http://www.schema.org/>

```

<span itemprop="jobTitle">graduate research assistant</span>
at the
<span itemprop="affiliation">University of Dreams</span>.
My friends call me
<span itemprop="additionalName">Johnny</span>.
You can visit my homepage at
<a href="http://www.JohnnyD.com" itemprop="url">www.JohnnyD.com</a>.
<section itemprop="address" itemscope itemtype="http://schema.org/PostalAddress">
I live at
<span itemprop="streetAddress">1234 Peach Drive</span>,
<span itemprop="addressLocality">Warner Robins</span>,
<span itemprop="addressRegion">Georgia</span>.
</section>

</section>

```

Turtle

Turtle é um formato *text plain*, que é mais compacto que o formato RDF/XML, utilizado para representação de dados RDF e que permite a abreviação com prefixos, que são declarados no início.

No âmbito desta dissertação, iremos utilizar a notação *Turtle* sempre que for necessário recorrer a exemplos que envolvam a representação de dados RDF ou SKOS. Seguimos a notação *Turtle* não só por ser mais compacta e inteligível para nós humanos, mas também por ser a notação adotada no livro *Semantic Web for Working Ontologist*, que é uma das nossas referências principais.

Exemplificando a sintaxe *Turtle* em RDF:

```

@prefix rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix ex:<http://www.example.com/>.
    ex:aResource ex:aProperty ex:anotherResource;
    ex:anotherProperty "An RDF Literal"@en.

```

O exemplo acima é equivalente a seguinte expressão na sintaxe RDF/XML:

```

<rdf:RDF xmlns:rdf=http://www.w3.org/1999/02/22-rdf-syntax-ns#
        xmlns:ex="http://www.example.com/">
  <rdf:Description rdf:about="http://www.example.com/aResource">
    <ex:aProperty rdf:resource="http://www.example.com/anotherResource"/>
    <ex:anotherProperty xml:lang="en">An RDF Literal</ex:anotherProperty>
  </rdf:Description>
</rdf:RDF>

```

De acordo com Allemang & Hendler (2011), existem no entanto outras formas de expressar RDF na forma textual.

O **N-Triples** é a forma mais simples. Referencia os recursos sem abreviações e utilizando a totalidade do seu URI. Cada URI é escrito entre sinais angulares (< e >). Três recursos são expressos na ordem Sujeito/Predicado/Objeto, seguidos por um ponto (.).

Exemplo de um *N-Triples*:

```

<http://www.workingOntologist.org/Examples/Chapter3Manufacture#
http://www.workingOntologist.org/Examples/Chapter3/Manufacture#Product1
http://www.workingOntologist.org/Examples/Chapter3/Manufacture#Product .

```

Como se pode verificar é difícil transpor *N-Triples* para uma página de um livro, pois não é permitido incluir novas linhas dentro de um triplo. Um *ntriple* deverá ter o todo o triplo numa única linha.

O exemplo acima em *Turtle* ficava (subentendendo a declaração do prefixo):

```
Product1 rdf:type Product .
```

RDF/JSON

Para os autores, Heath & Bizer (2011), o RDF/JSON refere-se ao esforço de fornecer uma serialização RDF para o JSON (*JavaScript Object Notation*). A disponibilização de uma serialização JSON para o RDF é altamente desejada, com o crescente número de linguagens de programação a fornecerem suporte nativo para o JSON, incluindo as linguagens Web, JavaScript e PHP. Além de que, publicar dados RDF em JSON, torna os

dados acessíveis aos desenvolvedores Web, sem a necessidade de instalação de bibliotecas adicionais para manipulação de dados RDF.

De seguida tecemos algumas considerações sobre o OWL.

OWL (Web Ontology Language)

Na sua tese de doutoramento, Gonçalves (2007), refere que na atualização de 2002 da arquitetura da Web Semântica, «os esforços concentraram-se essencialmente na implementação de linguagens e ferramentas para o desenvolvimento de ontologias. Consequentemente, em 2004, principalmente como resultado do grupo de trabalho *Web Ontology Working Group*, foi publicada a *Web Ontology Language* (OWL) como linguagem recomendada pelo W3C para o desenvolvimento de ontologias».

O OWL é uma linguagem da Web Semântica, projetada para representar conhecimento rico e complexo sobre recursos, grupo de recursos e relações entre recursos. «O OWL é uma linguagem computacional lógica, cujo conhecimento expresso em OWL pode ser avaliado logicamente por outros programas de computadores, verificando igualmente a consistência do conhecimento ou tornando implícito o conhecimento explícito. Documentos OWL, conhecidos como ontologias, podem ser publicados na web e podem referenciar a outras ou serem referenciados por outras ontologias OWL» (Hitzler, Krötzsch, Parsia, Patel-Schneider, & Rudolph, 2012).

De acordo com Gonçalves (2007), o *Web Ontology Language* (OWL) é uma linguagem de criação de ontologias para a Web baseada em RDF(S) e DAML+OIL que inclui mais algum vocabulário para descrever propriedades, classes e relações entre classes, entre outros elementos. DAML+OIL é uma linguagem semântica de marcação e de criação e integração de ontologias para os recursos da Web baseada nas características das linguagens DAML (*DARPA Agent Markup Language*) e OIL (*Ontology Inference Layer* ou *Ontology Interchange Language*).

Foi desenhada para ser utilizada por aplicações que necessitassem do processamento do conteúdo da informação em vez de somente a apresentar às pessoas. O OWL promove uma melhor interoperabilidade dos conteúdos da Web, suporta diversos formatos tais como XML, RDF e RDF Schema, disponibilizando um vocabulário estendido, com

formalismo semântico, permitindo assim o desenvolvimento de ontologias formais (Pereira, 2009).

Pereira (2009) afirma ainda que «algumas dificuldades de integração e de compatibilidade da camada ontologia com as restantes camadas da arquitectura motivaram novas alterações à arquitectura para a Web Semântica».

A Figura 5 corresponde à versão mais recente da arquitectura da Web Semântica.

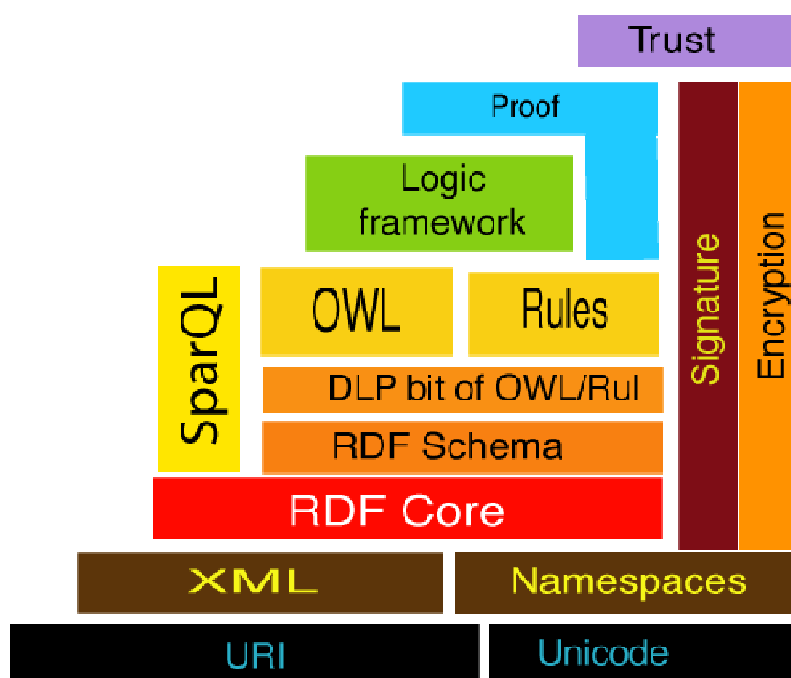


Figura 5 -Arquitetura da Web Semântica, versão 2005 (Pereira, 2007)

Uma das últimas alterações diz respeito à introdução de uma tecnologia candidata a recomendação na camada DLP (*Description Logic Programs*). Esta tecnologia baseia-se na Lógica Descritiva (OWL DL) e na Programação Lógica (*F-Logic*) e tem como objectivo integrar as duas principais tecnologias para a Web Semântica (RDF e OWL), proporcionando uma estrutura de conhecimento mais flexível (Calvanese et al., 2006, citado por Pereira, 2007). Outra dessas alterações refere-se à camada de inferência SPARQL (*RDF Query Language and Protocol*) que permite realizar consultas sobre documentos RDFS ou OWL (Pereira, 2009).

SPARQL

De acordo com Harris & Seaborne (2013), o SPARQL pode ser usado para expressar interrogações (*queries*) a diversas fontes de dados. Para recuperar os dados através de *queries*, os dados deverão estar guardados, nativamente, em RDF. O SPARQL suporta igualmente agregações, *subqueries*, negações, criação de valores por expressões e *queries* com restrições por intermédio de grafos RDF. Os resultados de *queries* SPARQL podem resultar em conjuntos ou grafos RDF.

O SPARQL é uma linguagem recomendada pela W3C, e a sua mais recente atualização ocorreu em Março do corrente ano. SPARQL 1.1 «é um conjunto de especificações que fornece linguagens e protocolos para interrogar e manipular grafos RDF encontrados na Web ou em repositórios RDF». (W3C SPARQL Working Group, 2013).

Segundo Tarasova (2012), o SPARQL possui uma sintaxe similar à sintaxe do SQL (*Structured Query Language*). Uma *querie* SPARQL dispõe de duas cláusulas mandatórias:

- SELECT: Especifica quais os subconjuntos de variáveis que se pretende;
- WHERE: Define os triplos onde se deseja fazer o *matching*.

Existem outras opções, não obrigatórias, que se podem utilizar, como é o caso do PREFIX, que pode ser utilizado para declaração de prefixos que poderão ser precisos numa *querie* SPARQL. Existem outras cláusulas opcionais que têm a mesmo significado no SQL, que são as cláusulas DISTINCT, ORDER BY, LIMIT, OFFSET, entre outros.

2.2. Linked “Open” Data

A publicação de dados na Web, permite acesso fácil a dados e suporta a reutilização de dados. O *Hypertext Transfer Protocol* (HTTP), é utilizado para aceder ao conteúdo apontado pelo *Uniform Resource Identifier* (URI) e para recuperar dados acerca de um recurso. Correspondente aos *links* nos documentos HTML, os dados podem fornecer *links* para outros recursos, formando deste modo a Web de Dados (Eckert, 2013).

De acordo com Méndez & Greenberg (2012) a ideia de *linked data* não emergiu da noite para o dia. Como se sabe, muitos dos desenvolvimentos nos campos da ciência da

computação, ciência da informação, são anteriores à própria Web. As autoras dão exemplos que nos levam a recuar no tempo, nomeadamente a noção de controlo universal bibliográfico ou a apresentação em 1945 do Memex de Vannevar Bush, um dispositivo hipotético que utilizava ligações associativas para sustentar a memória ao longo do tempo. Outro exemplo, o relatório anual do Instituto Belga de Bibliografia, faz-nos recuar até o ano de 1912.

Em 2010, Tim Berners-Lee, redefiniu a concepção de *linked data* acrescentando a filosofia *openness*, em que LOD são dados vinculados, que são publicados sob uma licença aberta, o que não impede a sua reutilização e de forma gratuita.

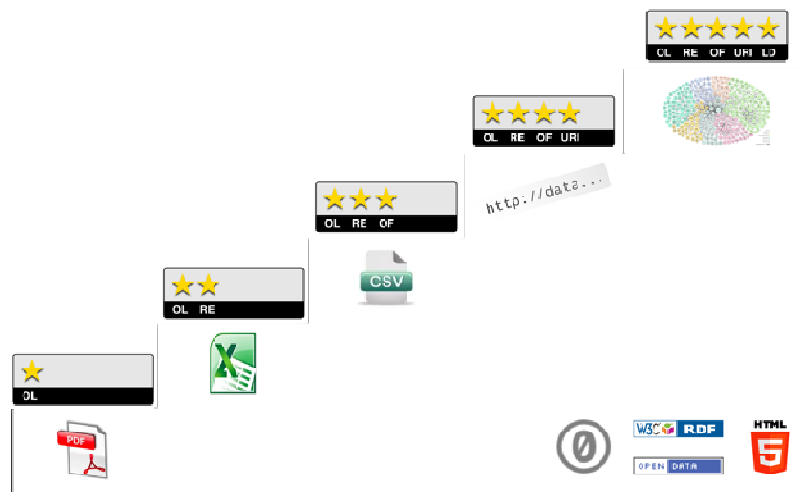


Figura 6 - 5 Star Deployment Scheme (adaptação de Michael Hausenblas, 2012)

Um exemplo típico de *Linked Dataset* é o DBPedia, que basicamente torna o conteúdo da Wikipedia disponível em RDF. A importância do DBPedia, segundo o W3C, não é apenas pelo facto de incluir dados do Wikipedia, mas por incorporar ligações para outros *datasets* na Web, como por exemplo para o *Geonames*. Ainda de acordo com o W3C, fornecendo estas ligações extra (em termos de triplos RDF), aplicações poderão, aquando da criação de suas aplicações, explorar conhecimentos de outros *datasets* e integrar factos de outros *datasets*.

2.3. Vocabulários Controlados

Pretendemos nesta secção ilustrar o que a literatura revela acerca do uso de ontologias, mais especificamente sobre os vocabulários controlados.

2.3.1. Conceito de Ontologia

O termo ontologia tem sido utilizado desde há muito tempo. Segundo McGuinness (2003), a primeira utilização do termo ontologia data de 1721 e fornece duas definições: A primeira como sendo, «*um ramo da metafísica preocupada com a natureza e as relações do ser*» e a segunda definição, «*uma teoria particular sobre a natureza do ser ou do tipo de coisas existentes*». Ainda segundo McGuinness (2003), desde há muitos anos que se conhecem noções mais formais e matemáticas sobre ontologia. Smith (1998) citado por McGuinness (2003), afirma que desde 1900, existe uma noção formal sobre ontologia.

Presentemente a ontologia não é apenas objeto de estudo na área da Filosofia, ela tem sido alvo de estudo em áreas como a Linguística, Ciência da Informação, Ciências da Computação, Engenharia do Conhecimento, Inteligência Artificial, só para citar alguns.

Conforme o Dicionário da Porto Editora¹⁰, ontologia é um nome feminino, e na área da Filosofia, «é a parte da metafísica que estuda o ser em si, as suas propriedades e os modos por que se manifesta»; «é a ciência que estuda os seres considerados em geral».

De acordo com Catarino & Baptista (2010), «a origem da palavra ontologia remonta ao grego *Ontos* (ser) + *Logos* (razão/palavra) e actualmente existem dois sentidos diferentes para a palavra. Ontologia como campo de estudo da Filosofia e Ontologia como uma tecnologia para cientistas da computação e da informação.»

Para Guarino, Oberle, & Staab (2009), o termo ontologia é utilizado com diferentes significados em diferentes campos. Pode-se distinguir entre um nome indefinido, Ontologia e a utilização de um nome definido, uma ontologia. No primeiro caso estamos perante um ramo da Filosofia, que se dedica à natureza e estrutura da realidade. Segundo

¹⁰ Disponível na www <URL: <http://www.infopedia.pt/lingua-portuguesa/ontologia>>, acedido a 29.12.2012.

esses mesmos autores, devemos ao Aristóteles¹¹ a definição de Ontologia como ciência. No segundo caso, que reflete uma maior prevalência de utilização no ramo da Ciência da Computação, referimos a uma ontologia, como sendo um artefacto computacional. Ainda segundo Guarino et al. (2009) e escudando-se no pragmatismo da Inteligência Artificial, pode-se afirmar que o que existe é o que se pode representar.

Segundo Swartout & Tate (1999), citados por Librelotto, Ramalho, & Henriques (2005), na área de sistemas de informação, ontologia é definida como um conjunto de conceitos e termos ligados entre si (numa rede) que podem ser usados para descrever alguma área do conhecimento ou construir uma representação para o conhecimento.

Contudo, Noy & McGuinness (2001), defendem que uma ontologia é uma descrição explícita e formal de conceitos num dado domínio (classes), exibindo propriedades de cada conceito, descrevendo várias características e atributos (*slots*) de um conceito, e restrições sobre esses *slots*. Os mesmos autores defendem ainda que uma ontologia define um vocabulário comum para investigadores que necessitam de partilhar informação num domínio específico.

As classes são o foco da maioria das ontologias, descrevem conceitos num dado domínio. Por exemplo, uma classe de vinhos representa todos os vinhos, e vinhos específicos são as instâncias dessa classe. As classes por sua vez podem ter subclasses, que representam conceitos que são mais específicos em relação à sua superclasse. Tomando ainda como exemplo os vinhos, pode-se dividir a classe de todos os vinhos em vinho tinto, branco e rosé. Em alternativa pode-se ainda dividir a classe de todos os vinhos em espumantes e não espumantes (Noy & McGuinness, 2001).

Noy & McGuinness (2001), apresentam os seguintes componentes de uma ontologia:

- a) **Classes ou conceitos** – descrevem conceitos do domínio (são muitas vezes o foco das ontologias);
- b) **Relações** – representam as diversas formas de associação entre conceitos de um domínio;

¹¹ Filósofo grego, do período da antiguidade clássica.

- c) **Atributos** – descrevem as características que compõem os conceitos e as instâncias;
- d) **Instâncias** – são usadas para representar os elementos de uma ontologia, podem ser as ocorrências dos conceitos até sua individualização. Uma instância é um conceito que pertence a uma classe e possui determinadas propriedades;
- e) **Axiomas** – são expressões verdadeiras que usam linguagens de lógica para descrever os *constructs* na ontologia;
- f) **Regras** – são geralmente usadas para inferir conhecimento na ontologia.

Por seu turno, Gruber (1993) citado por Librelotto et al. (2005), afirma que ontologia é uma especificação explícita de uma conceptualização¹². Trata-se da definição mais prevalente e também a mais citada na literatura consultada.

Podemos ilustrar com um exemplo, proposto por Moreira (2003), a noção de conceptualização e especificação. Suponhamos que queremos modelar um domínio académico. Uma conceptualização deste modelo poderia conter os conceitos de “aluno”, “professor”, “disciplina”, “nota”, “curso”, “ministra”, etc. Uma especificação explícita desta conceptualização em lógica de primeira ordem poderia ser algo como o ilustrado na Tabela 3.

Tabela 3 - Especificação explícita de uma conceptualização (Moreira, 2003)

| Sentença em Lógica | Linguagem Natural |
|---|--|
| $\forall x \text{ disciplina}(x) \Rightarrow \exists y(\text{professor}(y) \wedge \text{ministra}(x, y))$ | Para toda disciplina existe um professor que a ministra. |
| $\forall x \text{ disciplina}(x) \Rightarrow \exists y(\text{curso}(y) \wedge \text{parte-de}(x, y))$ | Toda disciplina é parte de um curso. |
| $\forall x \text{ aluno}(x) \Rightarrow \exists y(\text{curso}(y) \wedge \text{curso}(x, y))$ | Todo aluno cursa um curso. |
| $\forall x \text{ aluno}(x) \Rightarrow \text{Corpo-académico}(x)$ | Todo aluno é um elemento do corpo académico. |
| $\forall x \text{ professor}(x) \Rightarrow \text{Corpo-académico}(x)$ | Todo professor é um elemento do corpo académico. |

¹² Conceptualização é um nome feminino, que significa ação de formar conceitos ou ação de organizar em conceitos. Disponível na [www: <URL: http://www.infopedia.pt/lingua-portuguesa/conceptualiza%C3%A7%C3%A3o>](http://www.infopedia.pt/lingua-portuguesa/conceptualiza%C3%A7%C3%A3o).

Uma outra definição de ontologia e que vem explicitar o significado proposto por Gruber é a seguinte: Ontologia é uma representação formal de um modelo de uma determinada porção da realidade e pode ser chamada de Vocabulário Controlado, de Tesouro (ou Ontologia Informal), de Taxonomia, etc. (*Ontologia e a Informação*, 2011).

De acordo com Sewald et al. (2013), citando Guarino (1998), «em Inteligência Artificial, uma ontologia refere-se a um artefato de engenharia, constituído por um vocabulário específico usado para descrever certa realidade, mais um conjunto de hipóteses explícitas sobre o significado pretendido de um vocabulário palavras. Este conjunto de pressupostos geralmente tem a forma de lógica de primeira ordem, onde as palavras do vocabulário aparecem como predicados unário ou binário, respectivamente chamados conceitos e relações. No caso mais simples, uma ontologia descreve uma hierarquia de conceitos relacionados por relações de subordinação, em casos mais sofisticados, axiomas apropriados são adicionados a fim de expressar outras relações entre conceitos e para restringir a sua interpretação pretendida».

Como se pode constatar, existem várias definições acerca do conceito. Cada uma direcionada para o seu *locus* científico.

No âmbito deste trabalho adotaremos a definição proposta por Noy & McGuinness (2001) porque se trata de uma das fontes mais citadas. Além de ser uma boa fonte, enuncia com clareza as definições e conceitos, sem entrar em demasiados detalhes e formalismos desnecessários para o âmbito do trabalho em questão.

Em jeito de síntese podemos pedir emprestada a afirmação dos autores Codina & Pedraza-Jiménez (2011), em que assinalam que outra das virtudes das ontologias, como campo de estudo, é que estão a ajudar a revitalizar a semântica documental, tanto na vertente das linguagens documentais como nos esquemas de metadados, dotando a ambos de uma nova e inesperada potencialidade. Ainda os mesmos autores, afirmam que novas portas se abrem para os profissionais e estudiosos, pois é um campo novo e que ainda há muito para se explorar. E utilizando uma metáfora dos autores podemos dizer que, «graças às ontologias e ao projeto da web semântica, estamos de novo em uma situação em que é impossível focar o telescópio sem descobrir alguma coisa».

Tipos de Ontologia

Bôaventura (2009), citando Breitman (2005) e Chandrasekaran (1999), advoga que as ontologias podem ser classificadas de acordo com o seu espectro semântico e a sua generalidade.

Segundo o seu espectro semântico, ontologias podem ser divididas em:

- a) **Vocabulários controlados** que são listas finitas de termos, como, por exemplo, catálogo norte-americano NAICS, que lista os produtos e serviços oferecidos em diversas áreas;
- b) **Glossário** que é uma lista de termos com significados em linguagem natural, por exemplo, formato de um dicionário;
- c) **Tesauros** que é uma lista de termos que padroniza a utilização de palavras indexadas;
- d) **Hierarquias “tipo-de” informais** que utiliza o relacionamento de generalização (tipo-de) de maneira informal;
- e) **Hierarquias “tipo-de” formais** que incluem instâncias de um domínio em que os relacionamentos de generalização são respeitados integralmente, como, por exemplo, a classificação dos seres vivos.

Na classificação, segundo sua generalidade, as ontologias podem ser divididas em:

- a) **Ontologias de alto nível**, que descrevem conceitos muito genéricos, tais como: espaço, tempo e eventos;
- b) **Ontologias de domínio**, que descrevem o vocabulário relativo a um domínio específico, usando conceitos presentes na ontologia de alto nível, por exemplo, radiografias ou mamogramas são termos específicos no campo da radiologia;
- c) **Ontologias de tarefas**, que descrevem um vocabulário relativo a uma tarefa genérica ou atividade pela especificação de conceitos presentes na ontologia de alto nível;
- d) **Ontologias de aplicação** são mais específicas, contêm toda a informação que se precisa para modelar o conhecimento necessário a uma aplicação em particular.

2.3.2. Conceitos de Vocabulário e de Vocabulário Controlado

Definição de Vocabulário

No campo da Web Semântica, e segundo o W3C, vocabulários definem conceitos e relacionamentos usados para descrever e representar uma área de conhecimento. São usados para classificar termos que podem ser utilizados para uma aplicação particular, caracterizando possíveis relacionamentos e definindo possíveis limitações na utilização desses termos. Na prática, os vocabulários podem ser muito complexos (com vários milhares de termos) ou muito simples (descrevendo apenas um ou dois conceitos).

Segundo o W3C não existe uma clara divisão entre o que é referido como vocabulário e ontologia. A tendência é usar o termo ontologia quando se refere a coleções de termos mais complexos e formais, enquanto que se deve utilizar o termo vocabulário quando o formalismo não é estritamente necessário.

Por seu turno, Gray, Gray, & Ounis (2009), definem Vocabulários como sendo, mecanismos para formalizar a terminologia utilizada dentro de um domínio de discurso e as relações semânticas entre os termos. Tipicamente um vocabulário é desenvolvido para (i) limitar os termos de que se necessitam e que se encontram disponíveis e (ii) fornecer consenso sobre o significado dos termos.

Ainda de acordo com esses mesmos autores, Gray, Gray, & Ounis (2009), um vocabulário encerra um conjunto de conceitos, cada um dos quais captura um termo preferido, termos alternativos, definições e notas sobre uma ideia num determinado domínio. Os conceitos de um vocabulário podem ser relacionados a outros conceitos de diversas maneiras: Um conceito c1 pode ser relacionado a outro conceito c2 como um *broader term* (c2 é termo mais geral), *narrower term* (c2 é mais específico), ou um *related term* (c2 partilha uma associação). Os relacionamentos *broader/narrower* definem uma estrutura multi-hierárquica para os conceitos. Quer dizer que qualquer recurso recuperado via um dado termo, deve ser igualmente recuperado por intermédio do seu *broader term*.

Para van Assem (2010), um vocabulário centra-se usualmente em apenas poucos atributos e comporta terminologias originárias de um domínio específico, como por

exemplo artes, medicina ou geografia. Vocabulários incorporam alguns princípios que são úteis, nomeadamente para simplificar as procuras/pesquisas.

- O primeiro princípio nos diz que se deve criar um conjunto de termos de pesquisa finitos. Uma lista com todos os *subjects* que são precisos para indexação, ajudam o pesquisador, uma vez que ele ou ela não precisa adivinhar quais assuntos estão disponíveis e quais termos foram utilizados para denotar o tema.
- Um segundo princípio é agrupar sinónimos diferentes, para que haja diversas formas para encontrar um assunto oriundo de uma lista de *subjects*.
- Um terceiro princípio é fornecer um único identificador ao *subject*, para que a indexação possa ser feita com um identificador em vez de um termo. Isto previne problemas relacionados com homónimos (por exemplo, “banco” pode referir-se a um banco ou a uma instituição financeira).
- O quarto princípio é a generalização de assuntos: uma hierarquia é criada entre os assuntos. Como veremos mais adiante, na secção sobre o SKOS, todos esses princípios são acautelados.

Noção de Vocabulário Controlado

Vocabulário controlado é uma lista restrita de termos ou termos tipicamente utilizados para catalogações descritivas, *tagging* ou indexação. É controlada porque os utilizadores (catalogadores, etiquetadores, indexadores) poderão querer selecionar apenas termos de uma lista num determinado contexto. É igualmente controlada, porque apenas em circunstâncias específicas ou processos de revisão, poderão os termos dentro de um vocabulário controlado serem alvo de alteração ou aumento. E esta é uma tarefa da responsabilidade do editor do vocabulário controlado ou ontologista, e não dos utilizadores (Hedden, 2010).

Uma outra definição, mais sintética, de vocabulário controlado é-nos dada por Allemang & Hendler (2011), em que afirmam que vocabulário controlado é um conjunto de termos que fornecem referências e padrões comuns para sistemas de informação interligados.

2.4. SKOS

O SKOS (*Simple Knowledge Organization System*) propõe um modelo para expressar a estrutura básica e conteúdo para esquemas de conceitos tais como tesouros, esquemas de classificação, taxonomias, entre outros tipos similares de vocabulários controlados. O SKOS é baseado no RDF (*Resource Description Framework*), permite a criação de conceitos e sua publicação na Web, interligando com outros dados na Web e integrando em outros esquemas de conceitos (Isaac & Summers, 2009).

O desenvolvimento de SKOS começou no ano 2002 no seio do grupo de trabalho *SWAD-Europe* e difundiu-se publicamente através de um *draft* publicado em Novembro de 2005. A partir daquela data a proposta passou a designar-se de SKOS Core, etiqueta que ainda perdura em muitos trabalhos recentes. Em Agosto de 2009 o SKOS alcançou o *status* de Recomendação W3C (Pastor-Sanchez, Martinez-Mendez, & Rodriguez-Munoz, 2012).

O modelo de dados SKOS é formalmente definido como uma ontologia OWL-full, permitindo a representação de qualquer tipo de sistema de organização de conhecimento em RDF. Os dados SKOS são expressos através de triplos RDF, podendo ser codificados utilizando uma sintaxe RDF/XML ou uma sintaxe *Turtle*¹³ (Isaac & Summers, 2009).

O SKOS pode ser utilizado isoladamente ou em combinação com outras linguagens mais formais como o OWL (*Web Ontology Language*). O objectivo do SKOS não é o de substituir qualquer tecnologia, mas sim de contribuir para a interoperabilidade entre sistemas, baseado num modelo simplificado. O seu âmbito de aplicação se estende a praticamente todos os vocabulários controlados (Pastor-Sanchez et al., 2012).

2.4.1. Elementos Fundamentais do SKOS

Para a representação de sistemas de organização de conhecimento (em inglês, KOS - *Knowledge Organisation System*), o SKOS propõe um modelo que assenta em alguns princípios a que seguidamente faremos referência.

¹³ <http://www.w3.org/TeamSubmission/turtle/>

Segundo (Pastor-Sanchez et al., 2012) em SKOS os elementos de um vocabulário representam-se utilizando conceitos entre os quais se estabelecem relações semânticas hierárquicas (simples ou transitivas) e associativas.

Em SKOS os recursos conceptuais (conceitos) são identificados através dos seus URIs, rotulados em *strings* lexicais em um ou mais linguagens naturais, documentados com vários tipos de notas, semanticamente relacionados entre si e organizados em hierarquias informais e redes associativas; e agregados em esquemas de conceitos (Isaac & Summers, 2009).

Conceitos

O elemento fundamental de um vocabulário SKOS é o conceito. De acordo com Isaac & Summers (2009), conceitos são unidades de pensamento - ideias, significados, (ou categorias) de objetos e eventos, subjacentes aos sistemas de organização do conhecimento. Como tal, os conceitos existem na nossa mente como entidades abstratas, que são independentes dos termos utilizados para os rotular.

O SKOS apresenta a classe skos:Concept, que permite aos implementadores/desenvolvedores de ontologias afirmar que um determinado recurso é um conceito. Este é feito em dois passos:

1. Criando (ou reutilizando) um URI para identificar univocamente o conceito.
2. Afirmando em RDF, utilizando a propriedade `rdf:type`, que o recurso identificado pelo URI é do tipo `skos:Concept`.

Por exemplo¹⁴:

```
<http://www.example.com/animals> rdf:type skos:Concept.
```

Que poderá ser igualmente representado na sintaxe Turtle, mais compactamente, utilizando o prefixo *namespace* `ex`, abaixo definido:

¹⁴ Os exemplos desta secção foram retirados do documento SKOS Primer (Isaac & Summers, 2009).

ex:animals rdf:type skos:Concept.

Etiquetas (*Labels*)

São as expressões que são utilizadas para se referir aos conceitos em linguagem natural. SKOS define três tipos de rótulos/*labels* (em vários idiomas) para um conceito: preferidos, alternativos e ocultos.

Os rótulos preferidos - Um mesmo recurso conceptual, apenas pode ter uma etiqueta preferida em cada idioma;

Os rótulos alternativos - Permitem enriquecer semanticamente um vocabulário, definindo vários pontos de acesso a um conceito (sinónimos, abreviaturas, siglas, etc);

Os rótulos ocultos - Estão escondidos, isto é, não são visíveis diretamente aos utilizadores finais e são utilizados apenas para funções de processamento interno, antecipando alguns erros de ortografia mais comuns (como por exemplo, incluindo variantes com erros ortográficos, ou seja aceitando termos que faltem sinais gráficos).

Em notação SKOS, as três propriedades das etiquetas acima referidas são: skos:prefLabel, skos:altLabel e skos:hiddenLabel respetivamente.

O exemplo a seguir mostra um conceito com etiquetas preferidas em dois idiomas, uma em inglês e outra em francês:

```
ex:animals rdf:type skos:Concept;  
            skos:prefLabel "animals"@en;  
            skos:prefLabel "animaux"@fr.
```

O próximo exemplo ilustra um conceito com a utilização de etiquetas alternativas para além da etiqueta preferida:

```
ex:animals rdf:type skos:Concept;  
            skos:prefLabel "animals"@en;  
            skos:altLabel "creatures"@en;  
            skos:prefLabel "animaux"@fr;
```

```
skos:altLabel "créatures"@fr.
```

Exemplificando a utilização da etiqueta oculta, no caso um termo sem acentuação betes, para otimização de operações de pesquisa:

```
ex:animals rdf:type skos:Concept;  
            skos:prefLabel "animaux"@fr;  
            skos:altLabel "bêtes"@fr;  
            skos:hiddenLabel "betes"@fr.
```

Relações Semânticas

Nos sistemas de organização de conhecimento, as relações semânticas desempenham um papel crucial na definição dos conceitos. O significado de um conceito não se define apenas pelas palavras de linguagem natural nos seus rótulos, mas também pelas suas ligações com outros conceitos no vocabulário (Isaac & Summers, 2009).

SKOS comporta três propriedades *standard*:

- skos:broader e skos:narrower permitem a representação de relações hierárquicas, como por exemplo, a relação entre um género e suas espécies mais específicas, ou dependendo da interpretação, a relação entre o todo e suas partes. São propriedades inversas, não definidas como transitivas.
- skos:related permite a representação de relações associativas (não hierárquicas), como a relação entre um tipo de evento e uma categoria de entidades. Uma outra finalidade do skos:related é a associação de duas categorias em que nenhuma é mais genérica ou mais específica que a outra. Se por exemplo os recursos A e B estão relacionados via skos:related, não pode haver uma cadeia de relações skos:broader de A para B, situação análoga se verifica para relações skos:narrower. É uma propriedade simétrica não definida como transitiva.

Relações genéricas/específicas

De acordo com o documento SKOS Primer para indicar que um conceito tem um significado mais amplo (mais geral) do que outro, deve-se utilizar a propriedade `skos:broader`. A propriedade `skos:narrower` utiliza-se para realizar a afirmação inversa, que é como quem diz, quando um conceito tem um significado mais concreto (mais específico) do que outro.

Por exemplo:

```
ex:animais rdf:type skos:Concept;  
           skos:prefLabel "animais"@pt;  
           skos:narrower ex:mamíferos.
```

```
ex:mamíferos rdf:type skos:Concept;  
            skos:prefLabel "mamíferos"@pt;  
            skos:broader ex:animais.
```

Relações Associativas

Para denotar uma relação de associação entre dois conceitos pode-se utilizar o `skos:related`. Como se pode verificar no exemplo a seguir:

```
ex:pássaros rdf:type skos:Concept;  
           skos:prefLabel "pássaros"@pt;  
           skos:related ex:ornitologia
```

```
ex:ornitologia rdf:type skos:Concept;  
             skos:prefLabel "ornitologia"@pt.
```

Notas Documentais

Para documentação geral o SKOS dispõe do `skos:note`. Para especializações úteis para os utilizadores de um esquema de conceitos pode-se utilizar: `skos:scopeNote`, `skos:definition`,

skos:example e o skos:historyNote. Para especializações úteis para os gestores de Sistemas de Organização de Conhecimento tem-se:

- skos:editorialNote – fornece informação que é uma auxílio à arrumação administrativa, como lembrete de trabalho editorial ainda a ser feito, ou avisos no caso de ocorrerem alterações editoriais futuras;
- skos:changeNote – documenta pequenas alterações a um conceito para propósitos administrativos e de manutenção.

No âmbito do nosso projeto utilizamos outras propriedades não SKOS para documentar os conceitos. Tirando exemplos do nosso projeto, temos a propriedade rdfs:comment, que serve para documentar conceitos ou a propriedade dc:creator do Dublin Core, que serve para indicar a pessoa que criou o conceito.

Exemplos de propriedades não SKOS, em que temos uma propriedade rdfs:comment em dois idiomas, uma em inglês e outra em português:

rdfs:comment "Associação de Comércio Eletrónico de Portugal"@pt ;

rdfs:comment "Asymmetric Digital Subscriber Line"@en ;

2.4.2. Mapeamento entre Vocabulários

De acordo com Allemang & Hendler (2011), uma das principais vantagens que é apontada ao SKOS é a possibilidade de interligação de vários vocabulários. Os vocabulários controlados surgiram antes da Era dos computadores e muitos vocabulários foram desenvolvidos antes de se ter a ideia de os representar numa plataforma computacional, mormente em ter vocabulários interligados na Web.

O SKOS foi desenhado especialmente com o propósito de se interligar recursos informacionais. Sendo o SKOS representado em RDF, cada termo tem um único identificador (seu URI), que pode ser referenciado na web. Com isto, é perfeitamente possível relacionar termos de um vocabulário com termos de outro vocabulário (Allemang & Hendler, 2011).

Os mesmos autores, Allemang & Hendler (2011), declaram que o SKOS dispõe de um leque de propriedades de *matching*, especialmente concebidos para interligação com outros recursos. Estas propriedades são: *skos:exactMatch*, *skos:closeMatch*, *skos:broadMatch* e *skos:narrowMatch*. Com estas propriedades podemos relacionar termos do nosso vocabulário com termos de outros vocabulários.

Podemos ter num futuro a necessidade de efetuar *matching* de qualquer termo constante do nosso vocabulário com outro vocabulário. A Figura 7 (Pastor-Sánchez et al., 2012), mostra a interoperabilidade entre três vocabulários diferentes, utilizando as propriedades de *matching* supra mencionadas.

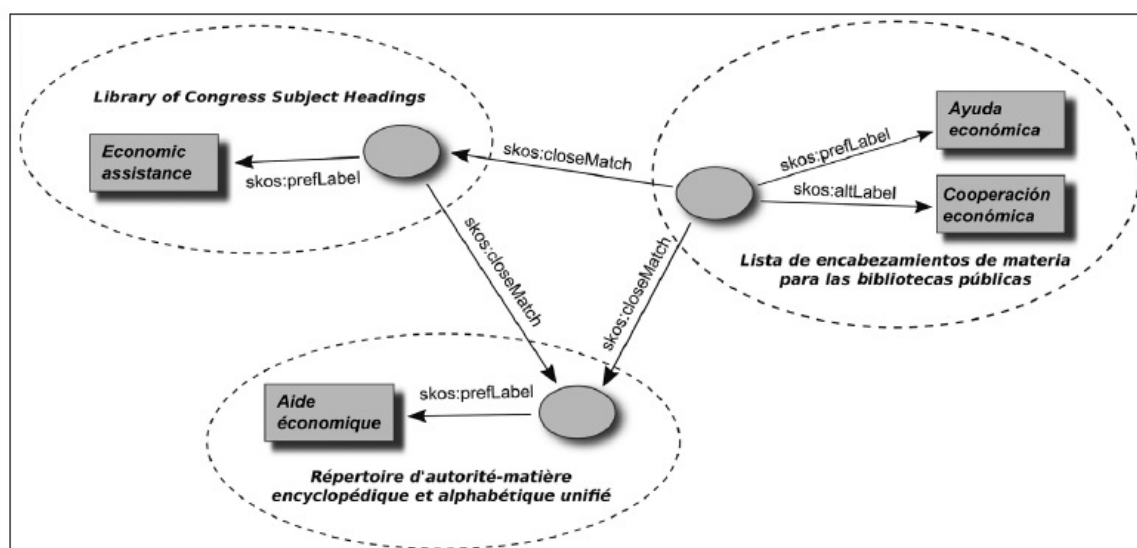


Figura 7 - Interoperabilidade entre vocabulários (Pastor-Sánchez et al., 2012)

Como se pode verificar na Figura 7, estamos perante três vocabulários diferentes e em três idiomas diferentes. Um em inglês (*Library of Congress Subject Headings*), um em francês (*Répertoire d'autorité-matière encyclopé et alphabétique unifié*) e outro em castelhano (*Lista de encabezamientos de materias para las bibliotecas públicas*). Podemos observar que cada vocabulário define *labels* (*skos:prefLabel*) para os seus literais, e recorrendo a propriedades de *matching* (*skos:exactMatch*) consegue-se uma interoperabilidade entre os três vocabulários.

A pensar neste desiderato, desenhamos o nosso vocabulário de maneira que, por exemplo, um recurso do nosso vocabulário que seja um *concept* do tipo “organização” pode vir a ser mapeado com termos de outros vocabulários. Para conceitos do tipo referido, poderemos seleccionar claramente o vocabulário RegOrg – *Registered Organization Vocabulary*¹⁵ (trata-se de uma recente recomendação da W3C).

¹⁵ <http://www.w3.org/ns/regorg#>

3. ABORDAGEM METODOLÓGICA

Este capítulo visa a apresentação do método de investigação adotado, para responder ao problema definido no capítulo propedêutico. Faremos igualmente referência às metodologias escolhidas para o processo de desenho e construção de vocabulários controlados.

Iremos considerar o método de investigação qualitativo, pois a nosso ver, é o que mais se adequa à realização deste trabalho. No que se refere às técnicas de pesquisa, utilizaremos igualmente técnicas de pesquisa qualitativa, nomeadamente com a realização de entrevistas, para validação dos termos do website da Autoridade Nacional de Comunicações (ANACOM).

O método de investigação qualitativo é amplamente utilizado em várias disciplinas e áreas, de onde poderemos destacar igualmente a área dos Sistemas de Informação.

De acordo com Myers (1997), os métodos de investigação podem ser classificados de várias formas, no entanto uma das classificações mais utilizadas é a distinção entre método de investigação qualitativo e método de investigação quantitativo. Ainda segundo o mesmo autor os métodos de investigação quantitativos foram originalmente desenvolvidos no campo das ciências naturais, onde estudavam fenómenos naturais. Exemplos de métodos quantitativos, presentemente bem aceites nas ciências sociais incluem *Surveys*, experiências de campo, métodos formais (p. ex. Econometria) e métodos numéricos como a modelação matemática.

Os métodos de investigação qualitativos, foram desenvolvidos nas ciências sociais para permitir aos investigadores o estudo de fenómenos sociais e culturais. São exemplos de métodos qualitativos o *Action Research*, o Estudo de Caso e a Etnografia. As fontes de dados qualitativas incluem a observação e trabalho de campo, entrevistas e questionários, documentos e textos e as impressões e conclusões do pesquisador (Myers, 1997).

Myers (1997), destaca a motivação de se fazer investigação qualitativa em oposição à investigação quantitativa, na constatação do fator que diferencia o humano do mundo natural, a sua capacidade de falar (caraterística inata do humano). Os métodos de

investigação são desenhados para ajudar o investigador a compreender as pessoas no contexto social e cultural que eles vivem.

Kaplan & Maxwell (1994) citados por Myers (1997), argumentaram que o objetivo da compreensão de um fenómeno, partindo-se do ponto de vista dos participantes e o seu contexto particular e institucional é em larga medida perdida quando os dados textuais são quantificados.

De acordo com Berndtsson, Hansson, Olsson, & Lundell (2008), «os métodos qualitativos possuem as suas raízes nas ciências sociais, e as suas preocupações originais tinham a ver com desenvolvimento e compreensão de uma área, em vez de produzir uma explanação sobre ela. A pesquisa qualitativa é usualmente utilizada em contextos sociais específicos. Ao longo dos anos, diferentes estilos e variações de métodos de pesquisa qualitativa têm sido propostos na literatura».

Ainda segundo Berndtsson et al. (2008), a pesquisa qualitativa é muitas vezes associada a trabalho de campo e análise a um reduzido número de cenários organizacionais. Para estes autores, um problema é frequentemente estudado segundo um único prisma, e o pesquisador empreende sua análise de uma posição próxima do seu objeto de estudo.

Para a validação de alguns aspectos específicos houve a necessidade de utilizar técnicas de pesquisa qualitativa, nomeadamente com a realização de entrevistas.

Entrevistas podem ser realizadas de diversas formas e de acordo com a finalidade pretendida. Existe uma variedade de aspectos que devem ser considerados aquando da decisão sobre a forma apropriada de entrevista. Diferentes estilos de entrevista possuem forças e fraquezas diferenciadas, e estão intimamente ligadas à própria natureza e competências do investigador na condução de entrevistas (Berndtsson et al., 2008).

3.1. Seleção das Fontes de Pesquisa

As fontes que utilizamos para recuperação dos artigos científicos para o nosso tema, foram variadas. As fontes de pesquisa bibliográfica foram escolhidas, tendo por base alguns fatores. Com efeito, destacámos as fontes de pesquisa que proporcionavam análises

estatísticas e métricas, como a métrica fator de impacto (referente ao número de citações), o *H Index* (métrica que avalia o autor), o JCR (*Journal Citation Reports*), entre outras métricas.

Para a pesquisa documental foi adotada a técnica de pesquisa por palavras-chave, tendo sido consultadas algumas bases de dados de documentos científicos, nomeadamente o Google Académico, Science Direct, Scopus, Web of Science, Ohister, Web of Knowledge, NDLTD, RepositoriUM, RCAAP, SciELO. O processo de análise e interpretação dos documentos foi levado a cabo com utilização de técnicas de análise de conteúdo.

A análise de conteúdo contempla um conjunto de técnicas de análise de comunicações visando a obtenção de indicadores (quantitativos ou não) que permitam a inferência de conhecimentos presentes nas mensagens (Bardin, 1977, citado por Silva et al., 2008).

Existe uma técnica de análise de conteúdo conhecida como “leitura flutuante”, na qual consiste em estabelecer contato com os documentos a serem analisados, obtendo impressões e orientações acerca do conteúdo (Bardin, 1977, citado por Silva et al., 2008).

Algumas das regras provenientes da análise de conteúdo foram consideradas na seleção de documentos. De acordo com Silva et al. (2008), temos:

- **Regra da exaustividade** - «não se pode deixar de fora qualquer elemento por esta ou por aquela razão que não possa ser justificável no plano do rigor».
- **Regra da representatividade** - «amostragem diz-se rigorosa se a amostra for parte representativa do universo inicial».

As palavras-chave utilizadas nas bases de dados, foram “Semantic Web”, “Vocabulários Controlados”, “SKOS”, “Ontologias”, “Ontologia”, “Linked Open Data”, “Metodologias Construção de Ontologias”. Muitas vezes optou-se por pesquisar por palavras-chave em língua inglesa para alargar as nossas possibilidades de êxito.

A nossa abordagem é híbrida, isto é, tentamos conjugar as melhores características de cada metodologia. Quando incidirmos na tarefa de construção do vocabulário controlado, adotaremos a Metodologia *Core Vocabulary* proposta pela ISA (PwC EU Services EESV, 2012). Trata-se de uma das propostas mais recentes na construção de vocabulários e

integraremos igualmente algumas etapas da metodologia *Ontology Development 101* (Noy & McGuinness, 2001).

3.2. Princípios de Desenho de Ontologias e de Vocabulários Controlados

De acordo com Noy & McGuinness (2001), não existe uma forma correta ou um método mais adequado para o desenvolvimento de ontologias. Depende da realidade que se pretende modelar, do espectro da ontologia e de possíveis integrações.

Segundo Sewald et al. (2013) para a concepção e/ou definição de ontologias pode-se (e deve-se) seguir uma metodologia de concepção. Ainda segundo os mesmos autores, o uso de metodologias tem como objetivo combater os seguintes problemas, mais prováveis de acontecer quando da não aplicação:

- A conceptualização da ontologia não fica muito clara no código da implementação;
- A falta de padronização acaba por dificultar a sua reutilização, pela dificuldade de compreensão;
- Gera dificuldade para a implementação de ontologias mais complexas, devido a falta do uso de uma metodologia, o que torna a transferência do conhecimento para a implementação mais difícil.

Constatamos que existem na literatura e das diferentes comunidades de prática, várias metodologias, métodos e guias que orientam o processo de construção e validação de ontologias.

Seguidamente vamos apresentar e discorrer sobre algumas propostas de metodologias de construção de ontologias.

3.2.1. *Ontology Development 101*

Noy & McGuinness (2001), sugerem uma abordagem iterativa (constituída por 7 etapas) para a construção de ontologias a que chamaram de *Ontology Development 101*.

ETAPA 1 - Determinação do domínio e âmbito da ontologia

É sugerido pelas autoras, que se inicie o desenvolvimento de uma ontologia, começando pela definição do domínio e do âmbito da ontologia (no nosso caso, vocabulário controlado), respondendo a um conjunto de questões:

- Qual o domínio que o vocabulário controlado deverá abarcar?
- Para que propósito irá ser utilizado o vocabulário controlado?
- Para que tipos de questões de informação deverá, o vocabulário controlado, fornecer respostas?
- Quem vai utilizar e manter o vocabulário controlado?

Para determinar o âmbito do vocabulário controlado, sugere-se a criação de uma lista de questões a que o vocabulário deverá ser capaz de responder. A estas questões designam-se por *Competency Questions* (Grüninger & Fox, 1995).

Essas questões servirão mais tarde para o teste decisivo: Será que o vocabulário contém informação suficiente para responder àquelas questões? Será que as respostas requerem um nível particular de detalhe ou representação numa área particular?

Tomando como exemplo o domínio dos vinhos, proposto por (Noy & McGuinness, 2001), podemos considerar como Questões de Competência as seguintes:

- Quais as características do vinho deverão ser consideradas na escolha de um vinho?
- Será que o *Bordeaux* é um vinho tinto ou um vinho branco?
- O *Cabernet Sauvignon* combina com os frutos do mar ou outro prato de peixe?
- Qual a melhor escolha de vinho para carnes grelhadas?
- Que características de um vinho afetam a sua adequação a um prato?
- Um vinho é mais encorpado dependendo do seu ano de colheita?
- Quais foram as boas safras (colheitas) para o *Napa Zinfandel*?

A julgar por esta lista de questões, a ontologia deverá incluir informação sobre várias características de vinhos, tipos de vinhos, anos de colheita (boas e más), classificação dos

pratos que são importantes na escolha do vinho apropriado, combinações recomendadas para vinho e prato (Noy & McGuinness, 2001).

ETAPA 2 - Considerar a reutilização de ontologias existentes

De acordo com as autoras existem muitas ontologias já criadas que poderão ser estendidas ou refinadas, dependendo do nosso domínio particular ou finalidade desejada. Noy & McGuinness (2001), enumeram algumas ontologias que poderão ser reutilizadas, dando como exemplos o Ontolingua ontology library¹⁶ ou o DAML ontology library¹⁷. No nosso caso específico iremos interagir com vocabulários variados, onde podemos destacar o Dublin Core, RDFS, etc.

ETAPA 3 - Enumeração dos termos da ontologia

Esta etapa consiste na enumeração dos termos que se consideram mais importantes para a ontologia. No nosso caso esta etapa não será levada a cabo, uma vez que o nosso objetivo é a organização de termos já existentes.

ETAPA 4 - Definição das classes e hierarquia de classes

Segundo Uschold & Grüninger (1996), citados por Noy & McGuinness (2001), existem várias abordagens possíveis no desenvolvimento de uma hierarquia de classes. Um processo de desenvolvimento *top-down*, que começa com a definição dos conceitos mais gerais do domínio e a subsequente especialização dos conceitos. Um processo de desenvolvimento *bottom-up*, que é iniciado com a definição das classes mais específicas, com subsequentes agrupamentos das classes em conceitos mais genéricos. Um processo de desenvolvimento híbrido, isto é, um processo que combina as abordagens *top-down* e *bottom-up*.

¹⁶ <http://www.ksl.stanford.edu/software/ontolingua/>

¹⁷ <http://www.daml.org/ontologies/>

ETAPA 5 - Definição das propriedades das classes (*slots*)

As classes sozinhas não fornecem informação suficiente para dar resposta às questões de competência da etapa 1. Uma vez definidos algumas das classes, devemos então passar à descrição da estrutura interna dos conceitos. Todas as subclasses herdam as propriedades da sua superclasse.

ETAPA 6 - Definição das facetas dos *slots*

Para esta etapa definem-se as diferentes facetas que os *slots* poderão assumir. Estamos a referir aos tipos de dados, a cardinalidade ou seja número de valores e outros valores que os *slots* poderão tomar. A cardinalidade dos *slots*, define quantos valores um slot poderá ter. Alguns sistemas distinguem apenas entre cardinalidade simples (permitindo no máximo um valor) e cardinalidade múltipla (permitindo qualquer número de valores).

Alguns sistemas permitem a especificação de uma cardinalidade mínima e máxima para descrever o número de valores de slots com mais precisão. Cardinalidade mínima de N, significa que um *slot* deverá possuir no mínimo N valores. Em algumas situações poderá ser necessário definir a cardinalidade máxima como sendo 0 (zero). Se tal ocorrer, então os *slots* não poderão ter qualquer valor para uma subclasse particular. Relativamente ao tipo de valor que poderá assumir, os tipos mais comuns são: *String*, *Number*, *Boolean* ou *Enumerated* (em Protégé o tipo *enumerated* é o tipo *Symbol*).

ETAPA 7 - Criação de instâncias

A etapa final é responsável pela criação das instâncias individuais das classes da hierarquia. Definir uma instância de uma classe requer: 1) Escolha da classe, 2) Criação de uma instância individual da classe e 3) Preenchimento com valores *slots*. Tomando ainda como exemplo os vinhos proposto por Noy & McGuinness (2001), podemos criar uma instância individual *Chateau-Morgon-Beaujolais* para representar um tipo específico de vinho *Beaujolais*. *Chateau-Morgon-Beaujolais* é uma instância da classe *Beaujolais* representando todos os vinhos *Beaujolais*.

De seguida damos a conhecer a Metodologia *Core Ontology*, proposta pelo programa Europeu ISA, onde são destacadas as etapas que consideramos mais adequados melhor se adaptavam ao projeto.

3.2.2. Process and Methodology for Core Vocabularies

A Comissão Europeia e o seu *Directorate Generale for Informatics (DIGIT)*, criaram um programa a que chamaram ISA (*Interoperability Solutions for European Public Administrations*), cuja finalidade é a promoção da interoperabilidade semântica entre os estados-membros da União Europeia para a coleção, avaliação, indexação e disponibilização de um grande número de recursos de interoperabilidade semântica de um único ponto de acesso. Desta forma os investigadores e equipas de desenvolvimento poderão facilmente descobrir e reutilizar ativos, tais como os modelos de dados, taxonomias, tesouros e vocabulários, desenvolvidos por outras pessoas que se confrontam com casos de uso similares. (PwC EU Services EESV, 2012).

A metodologia proposta pela ISA, designada por *Process and Methodology for Core Vocabularies*, consiste na identificação de um conjunto de 10 etapas técnicas, e em 24 métodos associados às etapas.

Etapas:

- 1) Identificação de um conjunto significativo de conceitos fundamentais;
- 2) Pesquisa e revisão de soluções existentes;
- 3) Pesquisa de dados e serviços existentes;
- 4) *Use Cases*;
- 5) Requisitos;
- 6) Terminologia e Modelo Conceptual de Dados;
- 7) Convenções de nomenclatura (*Naming Conventions*);
- 8) Convenções de Identificadores (*Identifiers Conventions*);
- 9) Documento *Namespace*;
- 10) Garantia de Qualidade / Critérios de Conformidade

Seguidamente ilustramos algumas das etapas que consideramos mais relevantes para o nosso trabalho.

Identificação de um conjunto significativo de conceitos fundamentais

Um primeiro passo necessário é a identificação de um conjunto de conceitos fundamentais, em que o vocabulário a ser desenvolvido deverá ser posicionado. Como qualquer vocabulário, vocabulários controlados estão relacionados um ao outro. É importante identificar outros trabalhos relevantes realizados antes de começar qualquer trabalho sobre Vocabulários.

Pesquisa e revisão de soluções existentes

Para este passo é preciso saber o que está a ser utilizado no momento e se já se identificaram vocabulários relevantes?

Para um vocabulário pode-se fazer três perguntas fundamentais:

- Abrange o âmbito requerido?
- É estável e /ou sujeito a um processo de mudança formal?
- Já é amplamente utilizado?

Se um vocabulário existente passa todos esses testes então pode-se alcançar a interoperabilidade simplesmente usando-o. De seguida mostra-se uma lista de vocabulários que cumprem todos os três testes e são relevantes para o nosso trabalho:

- O *Dublin Core Metadata Initiative* (DCMI), vocabulário de termos de metadados que define atributos de metadados gerais, tais como título, criador, data e assunto.
- O *Basic Geo* (WGS), vocabulário que define termos como latitude e longitude para descrever lugares geográficos.
- *GeoNames*, integra dados geográficos, como nome de lugares, países, cidades, etc. E em vários idiomas. Utiliza as coordenadas WGS84 – *World Geodetic System*.

De acordo com o DCMI, existem duas formas de descrição de recursos. O Simple e o Qualificado. Na sua forma simple o DCMI propõe uma lista de 15 termos (ver Tabela 4) para descrição de recursos.

Tabela 4 – Propriedades dos Metadados *Dublin Core* (Baker et al., 2012)

| | Termo | Definição |
|----|--------------|---|
| 1 | contributor | Entidade que contribuiu para o conteúdo do recurso. |
| 2 | coverage | Define a abrangência (espacial e temporal) do conteúdo do recurso |
| 3 | creator | A entidade responsável pela criação do recurso |
| 4 | date | Uma data associada à publicação do recurso |
| 5 | description | A descrição do recurso, tipicamente contém o resumo |
| 6 | format | Define o tipo de mídia ou dimensões do recurso |
| 7 | identifier | Contém um identificador único para o recurso |
| 8 | language | A linguagem usada no conteúdo do recurso |
| 9 | publisher | A entidade responsável pela publicação do recurso |
| 10 | relation | Uma referência a um recurso relacionado |
| 11 | rights | A entidade que possui direitos autorais sobre o recurso |
| 12 | source | Uma referência ao local onde o recurso se localiza |
| 13 | subject | Um assunto referente ao recurso |
| 14 | title | Título do recurso |
| 15 | type | Define a natureza ou gênero em que o recurso está inserido |

Terminologia e Modelo Conceptual de Dados

Um vocabulário deverá descrever conceitos, relacionamentos e propriedades.

Conceito

Um conceito pode ser entendido como uma entidade, como por exemplo, um veículo, uma pessoa, um conjunto de dados ou um documento. Em termos RDF/SKOS, um conceito será representado como sendo uma Classe.

Relacionamentos

Um relacionamento é algo que interliga dois conceitos. Em termos RDF/SKOS, um relacionamento é um predicado que liga duas classes.

Propriedades

Uma propriedade é genericamente um valor numa dimensão particular, tais como o peso de um objeto, o nome de uma organização.

Naming Conventions

O *Core Vocabulary* indica que deve ser utilizada dois tipos de nomes: Técnica e convencional.

Restrições técnicas

Os termos deverão começar com um letra minúscula ou maiúscula (A – Z, a - z) ou o *underscore* (_). Esta restrição nos diz que não devemos começar um termo com um número, hífen ou caracteres acentuados.

Restrições convencionais

Para além das restrições técnicas pode-se seguir uma via convencional ao nomear os termos.

- As classes deverão começar com uma letra maiúscula, como por exemplo, Pessoa, Veículo, Endereço;
- Relacionamentos e propriedades usar o *camel case*, como por exemplo: *creatorName*, *isPartOf*, *familyName*, etc.

Identifier Conventions

Esta metodologia sugere igualmente boas práticas para publicação, nomeadamente:

- a) Assinatura dos identificadores de cada termo (o seu URI).
- b) Utilizar o caractere especial (#) para a terminação dos *namespaces*. Neste *namespace*, <http://www.w3.org/ns/org#> o identificador do conceito “organization” é <http://www.w3.org/ns/org#Organization>.

Para a realização do nosso trabalho iremos optar pela combinação da metodologia proposta pela ISA e pelo *Ontology Development 101*, no entanto passaremos em revista outras metodologias de construção de vocabulários.

Baonza (2010), na sua tese de doutoramento efetuou um levantamento das principais metodologias para a criação de ontologias, tendo elencado igualmente vários métodos e guias. E ainda como resultado da sua investigação propõe uma nova metodologia, a que chamou de *NeOn Methodology*.

As principais metodologias identificadas pela autora são o *Methontology*, *On-To-Knowledge* e o *DILIGENT*.

3.2.3. Methontology

A metodologia *Methontology* foi desenvolvida no seio do *Ontology Engineering Group* da Universidade Politécnica de Madrid.

A metodologia *Methontology* tem uma grande preocupação com o desenvolvimento de ontologias, indicando atividades a serem executadas, bem como indica a sequência e o nível de detalhamento dessas atividades. Esta metodologia propõe um ciclo de vida baseado na evolução de protótipos para o desenvolvimento de ontologias porque permite adicionar, mudar ou remover termos em cada nova versão. A construção de ontologias segue as atividades: especificação, aquisição do conhecimento, conceptualização, formalização, integração, implementação, avaliação, documentação e manutenção (Sewald et al., 2013).

Esta metodologia inclui: A identificação do processo de desenvolvimento de ontologias (que é o conjunto de atividades a serem realizadas para a construção de ontologias); um ciclo de vida com base em prototipagem evolutiva; técnicas para a realização de cada atividade durante a fase de gestão, desenvolvimento orientado e suporte das atividades.



Figura 8 - Methontology - Processo de Desenvolvimento de Ontologias (Cardoso, 2010)

A Figura 8 ilustra as diversas atividades do processo de desenvolvimento de ontologias:

Atividades de gestão de ontologias (*Management*) incluem as atividades de iniciação, monitorização e controlo de um projeto de ontologia em todo o seu ciclo de vida. Essas atividades são:

Atividades orientadas ao desenvolvimento de ontologias (*Development Oriented*) são agrupadas, conforme apresentado na Figura 8, em atividades de pré-desenvolvimento, desenvolvimento e pós-desenvolvimento.

Atividades de Suporte de Ontologias (*Support*) incluem as atividades que são necessárias para assegurar a conclusão com sucesso de um projeto de ontologia. Este grupo inclui uma série de atividades realizadas ao mesmo tempo, como as atividades orientadas para o desenvolvimento, sem os quais a ontologia não poderia ser construída.

O processo de desenvolvimento de ontologias não identifica a ordem em que as atividades deverão ser executadas. Esta é a função do ciclo de vida de ontologias. O ciclo de vida de ontologia descreve as atividades que deverão ser realizadas em cada fase e como as fases deverão estar relacionadas (Baonza, 2010).

A metodologia *Methontology* propõe o uso de questões de competência ou representações intermediárias para descrever os requisitos que a ontologia deve cumprir. No entanto, esta metodologia não fornece orientações detalhadas para realização desta atividade.

3.2.4. On-To-Knowledge

Por seu turno a metodologia *On-To-Knowledge* teve como objetivo a aplicação de ontologias para informações disponíveis eletronicamente, para melhorar a qualidade da gestão do conhecimento em organizações de grande dimensão e distribuídas.

Esta metodologia preconiza a construção de ontologias tendo em conta a sua utilização futura em aplicações de gestão de conhecimento. Por conseguinte, as ontologias desenvolvidas com esta metodologia são altamente dependentes da aplicação.

Esta metodologia propõe igualmente as questões de competência para a atividade de especificação de requisitos da ontologia; no entanto não são fornecidas orientações detalhadas para essa atividade.

Segundo Baonza (2010), a metodologia *On-To-Knowledge* propõe o desenvolvimento incremental e cíclico de uma ontologia, com base no modelo evolutivo de ciclo de vida.

Esta ontologia de ciclo de vida, ilustrada na Figura 9, é composta por 5 fases: **Estudo de Viabilidade** – Fase onde são identificados os problemas e áreas de oportunidade; **Kickoff da Ontologia** – Nesta fase é feita uma especificação de requisitos e são analisados as fontes; **Clarificação** – Fase destinada a elucidação dos conceitos com os especialistas do domínio; **Avaliação** – Fase de revisão e expansão baseada em *feedback* e análise de questões de competência; **Manutenção** – Fase destinada à gestão e manutenção da ontologia.

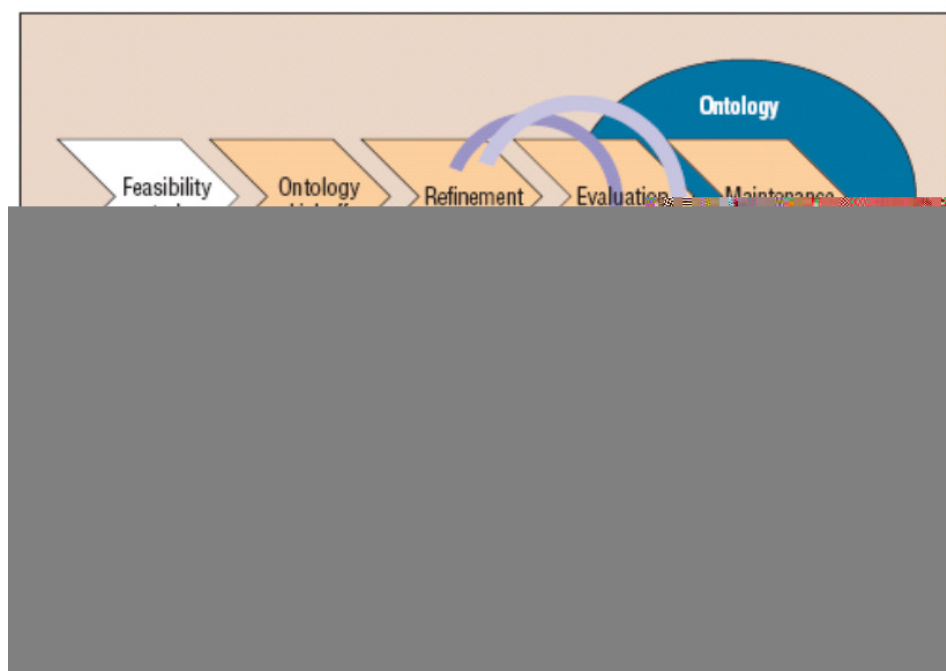


Figura 9 - Ciclo de Vida da Ontologia On-To-Knowledge (Staab et al., 2001)

3.2.5. DILIGENT

A metodologia DILIGENT, foi desenvolvida pelo Institute AIFB da Universidade de Karlsruhe (Alemanha) e pelo Instituto Superior Técnico de Lisboa (Portugal). Esta metodologia tem por objetivo o suporte a domínios especialistas num ambiente distribuído com o intuito de projetar e desenvolver ontologias. Esta metodologia está focada na engenharia de ontologias colaborativas, e a sua questão central é manter e acompanhar os argumentos de mudança (Baonza, 2010).

De acordo com Baonza (2010), o processo de desenvolvimento de ontologias proposto por esta metodologia inclui cinco fases principais: Construção, Adaptação Local, Análise, Revisão e Atualização Local.

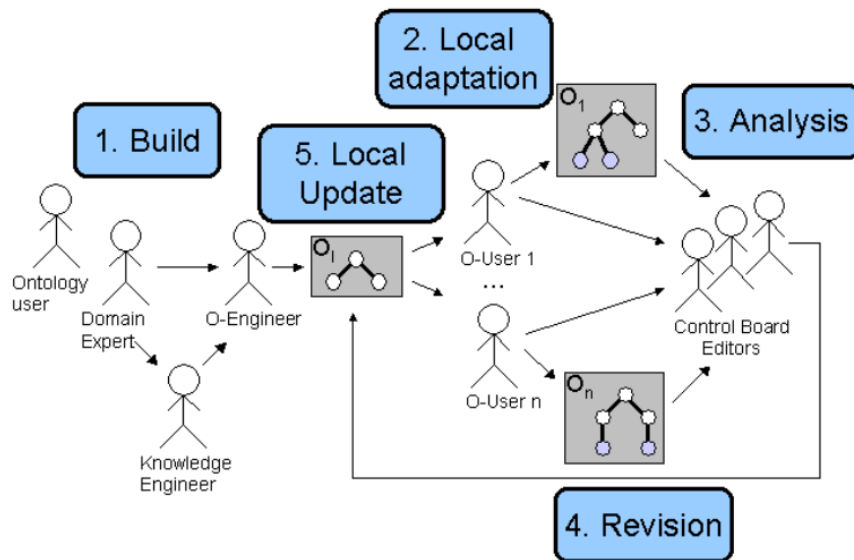


Figura 10 - Ciclo de Vida da Ontologia DILIGENT (Pinto et al., 2004)

Davies et al. (2006), descrevem com detalhe os cinco principais passos da metodologia.

Construção - O processo inicia com o domínio de especialistas, utilizadores, engenheiros de conhecimento e ontologistas que constroem uma ontologia inicial. A equipa envolvida na construção da ontologia inicial deverá ser relativamente pequena, no sentido de se chegar facilmente a um consenso para a primeira versão da ontologia;

Adaptações Local - Assim que a ontologia base estiver disponível, os utilizadores deverão trabalhar nela, adaptando-a localmente às suas próprias necessidades.

Análise - A equipa analisa as ontologias locais e os pedidos de alteração e tenta identificar similaridades nas ontologias dos utilizadores. Espera-se que as ferramentas consigam efetuar uma análise eficiente aos pedidos de alteração. Como todas as alterações introduzidas ou requisitadas pelos utilizadores não serão contempladas na ontologia, uma atividade crucial da equipa é decidir quais as alterações que deverão ser refletidas na próxima versão partilhada da ontologia.

Revisão - A equipa diretiva deve regularmente revisar a ontologia partilhada, para que as ontologias locais não desviem muito da ontologia partilhada. O objetivo da revisão é de realinhar com as necessidades óbvias dos utilizadores e deste modo ganhar aceitação

partilhada e com menos diferenças locais. Portanto, a equipa deverá ter uma balanceada e representativa participação de diferentes tipos de participantes envolvidos no processo: engenheiros de conhecimento, especialistas de domínio, engenheiros de ontologia e utilizadores.

3.2.6. Metodologia NeOn

A Metodologia NeOn vocacionada para a construção de ontologias é uma metodologia baseada em cenários, que fornece um conjunto de nove cenários que podem ser combinados entre si. Cada cenário é decomposto em diferentes processos e atividades (que estão incluídas no Glossário NeOn). Para cada processo ou atividade, linhas orientadoras são fornecidas (Baonza, 2010).

A Figura 11 apresenta o conjunto dos 9 cenários possíveis para a construção de ontologias. As setas com os números associados representam os diferentes cenários. Cada cenário é decomposto em diferentes processos e atividades. Os processos e atividades são representados pelos círculos coloridos. A figura em causa, mostra (caixas a tracejado) os recursos de conhecimento existentes para reutilização e os possíveis *outputs* resultantes da execução de alguns dos cenários apresentados.

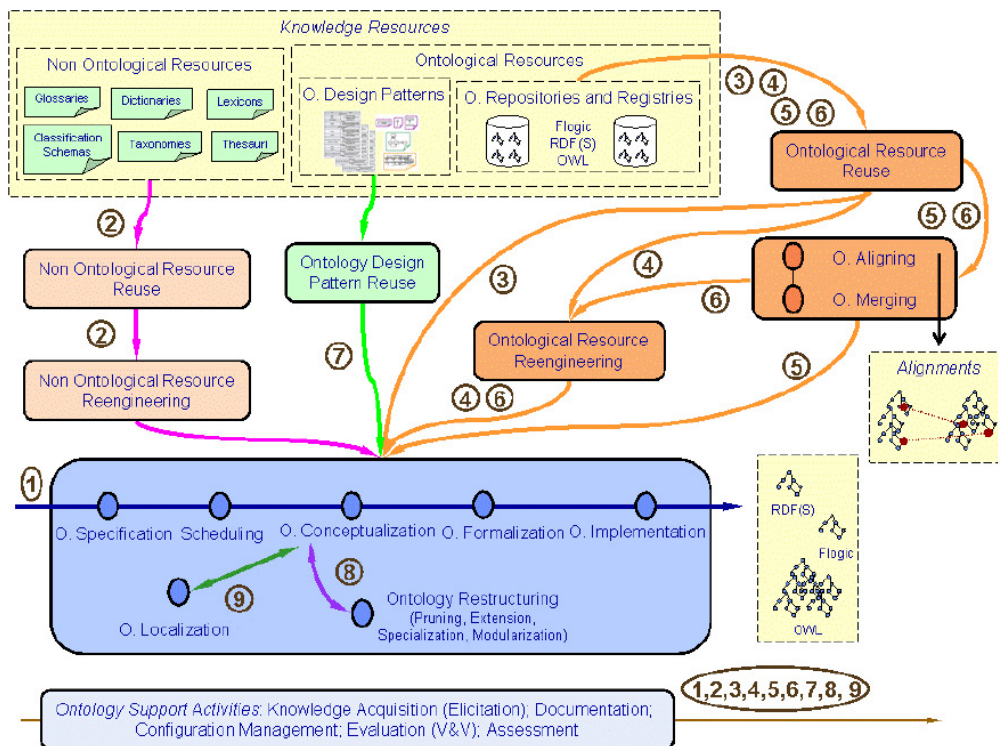


Figura 11 - Cenários da Metodologia NeOn (Baonza, 2010)

Cenário 1 - Da especificação à implementação. A Ontologia é rabuscada de início, sem a reutilização de recursos disponíveis.

Cenário 2 - Reutilização e Reengenharia de Recursos Não-Ontológicos.

Cenário 3 - Reutilização de Recursos ontológicos. Os desenvolvedores de ontologias utilizam recursos ontológicos (ontologias como um todo, módulos de ontologias, e/ou triplos ontológicos).

Cenário 4 - Reutilização e reengenharia de recursos ontológicos.

Cenário 5 - Reutilização e *Merging* de recursos ontológicos.

Cenário 6 - Reutilização, *Merging* e Reengenharia de recursos ontológicos.

Cenário 7 - Reutilização de *Ontology Design Patterns* (ODP). Os ontologistas acedem aos repositórios ODP para reutilização dos padrões de desenho.

Cenário 8 - Reestruturação dos recursos ontológicos.

Cenário 9 - Localização de recursos ontológicos. Os ontologistas adaptam a ontologia a outras linguagens e comunidades, obtendo uma ontologia multilingue.

Aquisição de Conhecimento, Documentação, Configuração de Gestão e Avaliação são as atividades de suporte que devem ser realizadas ao longo do desenvolvimento da ontologia (Baonza, 2010).

3.2.7. Outras Metodologias

Existem ainda referências a outras metodologias, tais como a Metodologia de Uschold & King, Metodologia de Gruninger e Fox, Método Kactus, Método Cyc, entre outros.

Método de Uschold & King

Este método é orientado para o desenvolvimento de ontologias e foi inicialmente proposto por Mike Uschold e Martin King. (Uschold & King, 1995). O método foi construído

baseando-se no desenvolvimento do projeto da *Enterprise Ontology* e em 1996 foi estendido por Mike Uschold e Michael Gruninger. Segundo Silva, Souza, & Almeida (2008), esta ontologia foi desenvolvida como parte do projeto *Enterprise* pelo Instituto de Aplicações em Inteligência Artificial da Universidade de Edinburgh e parceiros como IBM, Unilever e outros.

Uschold & King (1995), consideram os seguintes estágios como sendo necessários a uma metodologia abrangente: i) identificação do propósito da ontologia, onde é identificada a necessidade de construção, o grau de formalismo (desde o informal com uso de linguagem natural até o rigorosamente formal com uso de declarações lógicas) e as classes de utilizadores da ontologia, incluindo desenvolvedores e utilizadores das aplicações; ii) construção da ontologia, que se divide em: a) captura ou concepção da conceitualização da ontologia; b) codificação ou implementação através de uma linguagem de representação de ontologias e c) integração com ontologias já existentes; iii) avaliação da ontologia através dos requisitos especificados; iv) documentação acerca das pretensões da ontologia e das primitivas usadas para expressar as definições na ontologia (Silva et al., 2008).

Método de Uschold & Gruninger

Como foi referido no método anterior, este método é uma extensão do método Uschold & King. Segundo o autor, Stuckenschmidt (2003), Uschold & Gruninger definem 4 fases principais para o processo de desenvolvimento de ontologias, que são as seguintes:

1) Identificação do propósito e âmbito: Especialização, intenção de utilização, cenários, conjunto de termos incluindo características e granularidade.

2) Construção da ontologia

- a. Captura da ontologia: Aquisição de conhecimento – fase que interage com os requisitos da fase 1.
- b. Codificação da ontologia: Estruturação do domínio de conhecimento através de um modelo conceptual.

- c. Integração com ontologias existentes: Reutilização de ontologias existentes para acelerar o processo de desenvolvimento de ontologias no futuro.

3) **Avaliação**: Verificação e validação.

4) **Guidelines** para cada fase.

Metodologia TOVE

Esta metodologia foi desenvolvida “baseada na experiência de desenvolvimento da ontologia do projeto TOVE (*Toronto Virtual Enterprise*)”. Verificamos em alguma literatura que esta metodologia é muitas vezes designada por metodologia de Gruninger & Fox. As etapas da metodologia TOVE permitem descrever um domínio de processos de negócio e a modelação de atividades com o uso de ontologias, possibilitando a construção de um modelo lógico a partir de cenários informais expressos em linguagem natural (Gruninger & Fox, 1995).

De acordo com Silva et al. (2008), a metodologia de Gruninger e Fox foi usada no *Enterprise Integration Laboratory* (Laboratório de Integração de Empresas) da Universidade de Toronto para o projeto e avaliação de ontologias integradas, incluindo propostas de construção de novas ontologias e extensões de ontologias já existentes.

Foram propostos seis procedimentos no âmbito desta metodologia, que são os seguintes:

- 1) Elaboração de cenários de motivação, que auxiliam a identificar problemas no ambiente atual;
- 2) Especificação de questões de competência informal, que objetivam a especificar em linguagem natural os requisitos que a ontologia deverá ser capaz de atender;
- 3) Concepção da terminologia formal, em que, mediante declarações em lógica de primeira ordem, os conceitos e suas propriedades são organizados numa taxonomia;
- 4) Especificação de questões de competência formal, em que problemas são definidos de modo consistente perante os axiomas na ontologia;

- 5) Especificação de axiomas formais, que restringem a interpretação dos termos envolvidos nas questões de competência formal;
- 6) Verificação de teoremas completos, que determinam as condições sobre as quais as soluções das questões são completas.

Ontologia STAP

Trata-se de uma ontologia genérica, proposta por Catarino & Baptista (2010), mais vocacionada para a descrição de termos de metadados, tendo por base fundamentalmente o Dublin Core. O objetivo desta metodologia, segundo os autores é «compor uma especificação de todos os termos (propriedades, classes e esquemas de codificação), necessários para descrição de recursos etiquetados em repositórios, para que sejam compartilhados em aplicações RDF».

Compreende 6 etapas:

- 1) **Identificação de Propósito e Especificação de Requisitos** - etapa inicial e de planeamento da ontologia a construir;
- 2) **Capturas da Ontologia** - com base nos propósitos, identificam-se os conceitos e as suas relações;
- 3) **Formalização da Ontologia** - a conceitualização capturada é representada explicitamente numa linguagem formal;
- 4) **Integração com ontologias existentes;**
- 5) **Avaliação** - que deve ser feita concorrentemente com as etapas de Captura e de Formalização;
- 6) **Documentação** - consiste na documentação de todas as etapas.

Método Kactus

No âmbito do projeto europeu *Esprit Kactus* é enfatizado a questão de organização de bases de conhecimento que podem ser compartilhadas e reutilizadas em diferentes sistemas baseados em conhecimento. Para tal, utiliza ontologias de domínio para organizar o conhecimento independente da aplicação de *software* que será construída.

Segundo Silva et al. (2008), alguns investigadores, tendo por base o projeto Kactus investigaram a viabilidade da reutilização do conhecimento em sistemas de complexidade técnica, como o domínio de redes elétricas, e o papel das ontologias como suporte a tais sistemas. Tal investigação resultou num método de construção de ontologias, cujos processos envolvidos estariam condicionados ao desenvolvimento da aplicação, ou seja, toda vez que uma aplicação fosse construída, a ontologia, que representa o conhecimento necessário para a aplicação, seria refinada. Tais processos seriam os seguintes: i) desenvolvimento de uma lista de necessidades ou requisitos que precisam ser atendidos pela aplicação; ii) identificação de termos relevantes para o domínio da aplicação a partir de tais requisitos, construindo, assim, um modelo preliminar; iii) refinar e estruturar a ontologia a fim de obter um modelo definitivo; iv) buscar por ontologias já desenvolvidas por outras aplicações no sentido de sua reutilização. As ontologias reutilizadas demandariam refinamento e extensão para serem usadas na nova aplicação.

Método CYC

No dealbar da década de 80 do século passado, a *Microelectronics and Computer Technology* (MCC) deu início a criação da Cyc, uma ampla base de conhecimento que considera o conhecimento consensual sobre o mundo, incluindo regras e heurísticas para dedução sobre objetos e eventos do quotidiano. A linguagem de representação da Cyc é a CycL, considerada híbrida por combinar *frames* com cálculos de predicado. Tal linguagem possui uma máquina de inferência que permite herança múltipla, classificação automática, manutenção de *links* inversos, verificação de restrições, busca ordenada, detecção de contradição e módulo de resolução (Silva et al., 2008).

Ainda segundo os autores Silva et al. (2008), a base de conhecimento Cyc foi desenvolvida em 1990 por Douglas Lenat e Ramanathan Guha, em que três processos foram considerados em tal desenvolvimento, a saber:

- 1) Extração do conhecimento de senso comum;
- 2) Extração auxiliada por computador;
- 3) Extração gerida por computador.

No primeiro processo, o conhecimento requerido para a ontologia foi obtido de forma manual em diferentes fontes como artigos, livros e jornais. O segundo processo foi conduzido de maneira automática, isto é, com uso de ferramentas computacionais de processamento de linguagem natural e aprendizado de máquina capazes de usar conhecimento de senso comum suficiente para investigar e descobrir novos conhecimentos. E, finalmente, o terceiro processo foi conduzido por um número maior de ferramentas no sentido de gerir a extração de conhecimento de senso comum (partes consideradas difíceis de serem interpretadas nas fontes de conhecimento envolvidas) na base Cyc.

Método Sensus

A ontologia Sensus foi desenvolvida pelo grupo de linguagem natural *Information Sciences Institute* (ISI), com o propósito de ser usada para fins de processamento de linguagem natural. A ontologia Sensus possui aproximadamente 70 mil conceitos organizados numa hierarquia, de acordo com seu nível de abstração que vai de médio a alto.

O método Sensus, baseado na ontologia Sensus, propõe alguns processos para estabelecer as ligações entre os termos específicos e os termos da ontologia de alto nível. O resultado de tal processo é uma estrutura de uma nova ontologia, que é generalizada automaticamente através de uma ferramenta denominada *OntoSaurus* (Swartout *et al.* 1996; Fernandez; Gomez-Perez; Corcho, 2004).

De acordo com o método, os processos envolvidos na construção da ontologia de um domínio específico seriam: **i)** identificar termos-chave do domínio; **ii)** ligar manualmente os termos-chave à ontologia Sensus; **iii)** adicionar caminhos até o conceito de hierarquia superior da Sensus; **iv)** adicionar novos termos para o domínio; **v)** adicionar subárvores completas (Silva *et al.*, 2008).

Os três últimos métodos apresentados estão mais direcionados para ontologistas já com experiência de construção de ontologias, uma vez que descuram um pouco na elucidação dos passos para construção de ontologias. Por seu turno, ontologias como *Methontology* e *Ontology 101*, fornecem com detalhe a forma de conduzir cada atividade.

4. RESULTADOS

O objetivo deste capítulo é de apresentar o vocabulário controlado resultante do trabalho de organização dos termos do *website* da ANACOM.

Antes de passarmos à apresentação do vocabulário controlado propriamente dito, dispensamos algumas páginas do presente capítulo à caracterização da ANACOM.

4.1. Caracterização da ANACOM

Segundo pudemos apurar no sítio¹⁸ da instituição, a Autoridade Nacional de Comunicações (ANACOM) «regula e supervisiona o sector das comunicações eletrónicas e postais em Portugal (conforme resulta da própria lei de bases dos serviços postais (artigo 18º da Lei n.º 102/99, de 26 de julho) e da lei das comunicações eletrónicas (artigos 4º e 5º da Lei n.º 5/2004, de 10 de fevereiro), assegurando a representação nacional nos diversos fora internacionais relevantes. Promove a concorrência e defende os interesses dos cidadãos, garantindo a prestação de informações claras e a transparência nas tarifas e nas condições de utilização dos serviços. Visa também o desenvolvimento dos mercados e das redes de comunicações».

A ANACOM tem por objeto a regulação, supervisão e representação do setor das comunicações. Para o efeito, são atribuições da ANACOM:

1 - No âmbito da regulação do mercado:

- a) Garantir o acesso dos operadores de comunicações às redes, em condições de transparência e igualdade;
- b) Promover a competitividade e o desenvolvimento nos mercados das comunicações, nomeadamente no contexto da convergência das telecomunicações, dos meios de comunicação social e das tecnologias da informação;

¹⁸ <http://www.anacom.pt>

- c) Atribuir os títulos de exercício da atividade postal e de telecomunicações; assegurar a gestão do espectro radioelétrico, garantindo a coordenação entre as comunicações civis, militares e paramilitares, e a gestão da numeração no sector das comunicações.

2 - Em matéria de supervisão do mercado:

- a) Velar pela aplicação e fiscalização das leis, regulamentos e requisitos técnicos aplicáveis no âmbito das suas atribuições, bem como o cumprimento, por parte dos operadores de comunicações, das disposições dos respetivos títulos de exercício de atividade ou contratos de concessão;
- b) Garantir a existência e disponibilidade de um serviço universal de comunicações, assegurando o cumprimento das obrigações correspondentes;
- c) Velar pela correta utilização dos recursos espectrais e de numeração atribuídos; proteger os interesses dos consumidores, especialmente os utentes do serviço universal, em coordenação com as entidades competentes, promovendo designadamente o esclarecimento dos consumidores.

3 - Em relação à representação do setor das comunicações:

- a) Assegurar a representação técnica do Estado Português nos organismos internacionais congéneres, acompanhar a atividade das entidades reguladoras afins e as experiências estrangeiras de regulação das comunicações e estabelecer relações com outras entidades reguladoras;
- b) Colaborar com outras entidades públicas e privadas na promoção da investigação científica aplicada às telecomunicações, bem como na divulgação nacional e internacional do setor; promover a normalização técnica, em colaboração com outras organizações, no setor das comunicações e áreas relacionadas;
- c) Colaborar na definição das políticas de planeamento civil de emergência do sector das comunicações, apoiando tecnicamente os organismos e serviços responsáveis pelo estabelecimento e gestão da rede integrada de comunicações de emergência;

- d) Assegurar a realização de estudos nas áreas das comunicações postais e de telecomunicações, bem como a execução de projetos no âmbito da promoção do desenvolvimento do acesso à sociedade de informação e do conhecimento.

Para prosseguir as suas atribuições, compete nomeadamente à ANACOM:

- Assessorar o Governo, a pedido deste ou por iniciativa própria, na definição das linhas estratégicas e das políticas gerais das comunicações e da atividade dos operadores de comunicações, sugerindo ou propondo medidas de natureza política ou legislativa nas matérias relacionadas com as suas atribuições, e participar na definição estratégica global de desenvolvimento das comunicações, nomeadamente no contexto da convergência, realizando os estudos para o efeito necessários;
- Elaborar regulamentos, nos casos previstos na lei e quando se mostrem necessários ao exercício das suas atribuições, e promover processos de consulta pública e de manifestação de interesse, nomeadamente no âmbito da introdução de novos serviços ou tecnologias;
- Atribuir recursos espectrais e de numeração;
- Coordenar com a entidade competente a aplicação da lei da concorrência no setor das comunicações;
- Proceder à avaliação da conformidade de equipamentos e materiais e definir os requisitos necessários para a sua comercialização;
- Arbitrar e resolver litígios que surjam no âmbito das comunicações.

A eficiente concretização das atribuições que lhe estão cometidas e a especificidade do sector das comunicações, associada às constantes inovações que sofre, impõem a atribuição à ANACOM de um amplo poder normativo que faz desta uma verdadeira autoridade de regulação e supervisão das comunicações. Assim, além de emitir atos vinculativos individuais e concretos e de formular recomendações concretas, de instaurar e instruir processos e de punir as infrações que sejam da sua competência, de fiscalizar o cumprimento das leis e regulamentos aplicáveis ao sector das comunicações, de vigiar a atividade das entidades sujeitas à sua supervisão e o funcionamento do mercado das

comunicações, a ANACOM pode emitir os regulamentos que se revelem necessários ao exercício das suas funções.

Para além deste quadro de atribuições, com a publicação do Decreto-Lei n.º 7/2004, de 7 de janeiro, que disciplina certos aspectos legais dos serviços da sociedade da informação, em especial do comércio eletrónico, em transposição da Diretiva 2000/31/CE, do Parlamento Europeu e do Conselho, de 8 de junho de 2000, a ANACOM passou a desempenhar funções de entidade de supervisão central, com atribuições em todos os domínios regulamentados no referido diploma, salvo nas matérias em que lei especial atribua competência sectorial a outra entidade.

4.2. Proposta de organização dos termos em SKOS

Para sustentar a atividade de construção de ontologias, segundo Breitman (2005), citado por Gonçalves (2007), existem diversas ferramentas, normalmente associadas a cada um dos tipos de linguagens ontológicas, tais como: Protégé, SemanticWorks, SWOOP, SMORE, OntoEdit, etc.

No nosso caso específico, toda a organização foi realizada recorrendo à ferramenta Protégé¹⁹, versão 4.3.

Antes de apresentarmos o nosso vocabulário propriamente dito, e parafraseando Gonçalves (2007), podemos referir que «o Protégé é um ambiente *open source* interativo para o projeto de ontologias que oferece uma interface gráfica e intuitiva para a edição de ontologias, para além de disponibilizar uma arquitetura para a criação de estruturas baseadas em conhecimento. Esta é uma das ferramentas gráficas de edição de ontologias e aquisição de conhecimento mais utilizadas na modelação de domínios do conhecimento, uma vez que permite a edição de ontologias em diversas linguagens para a Web Semântica de forma personalizada. Com Protégé, a sintaxe deixa de ser uma preocupação para o

¹⁹ <http://protege.stanford.edu>

utilizador, podendo este canalizar a sua atenção para a representação dos conceitos e suas relações. De referir ainda que é uma ferramenta aberta baseada em Java que permite a adição de novas funcionalidades recorrendo aos *plug-ins* apropriados».

Vocabulário Controlado (VC) Proposto

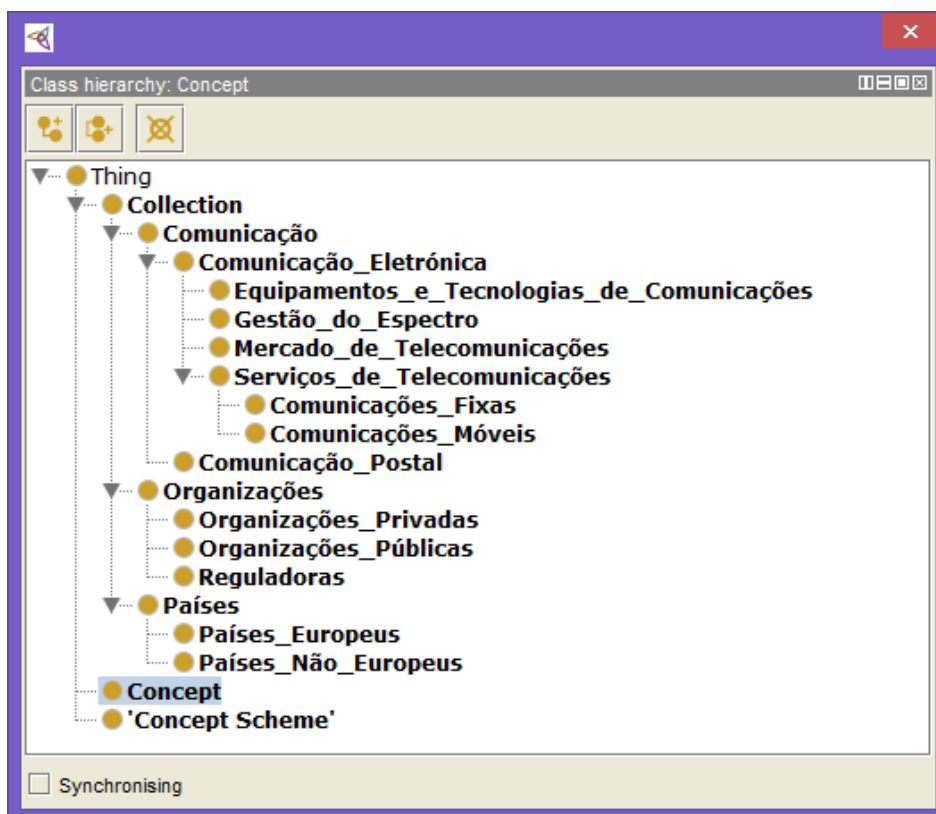


Figura 12 - Hierarquia de Classes do VC Proposto

Esta proposta de organização resultou de uma interação com responsáveis da ANACOM, em que se apresentou uma versão que foi alvo de sugestões. A solução que apresentamos na Figura 12, contém essas alterações nomeadamente com a introdução da classe Reguladoras, subclasse de Organizações, da classe Gestão do Espectro, subclasse de Comunicação Eletrónica, só para citar alguns. Salientamos que para a organização proposta contribuiu igualmente o CCIO (Classificação Portuguesa do Consumo Individual por Objectivo), onde baseamos na divisa 8, referente à Classificação de Correios e Telecomunicações, editado em 2009 pelo INE (Insituto Nacional de Estatística).

A organização proposta é composta pelas seguintes classes:

- **Comunicação** - que possui duas subclasses, Comunicação Eletrónica e Comunicação Postal; A primeira subclasse, está ainda dividida em 4 subclasses: Equipamentos e Tecnologias de Comunicação, Gestão do Espectro, Mercado de Telecomunicações e Serviços de Telecomunicações; A subclasse Serviços de Telecomunicações é subdividida em Comunicações Fixas e Comunicações Móveis;
- **Organizações** - tem três subclasses, Organização Pública, Organização Privada e Reguladoras;
- **Países** - dividimos em duas subclasses, Países Europeus e Países Não Europeus;

Tendo essa hierarquia definida, o próximo passo seria a afetação dos diversos termos que constavam do website da ANACOM. Toda essa informação encontrava-se em linguagem natural, pelo que a sua importação não foi linear. Houve a necessidade de carregar individualmente cada termo para o software, Protégé. Na Figura 13, podemos observar uma vista parcial dos termos importados.

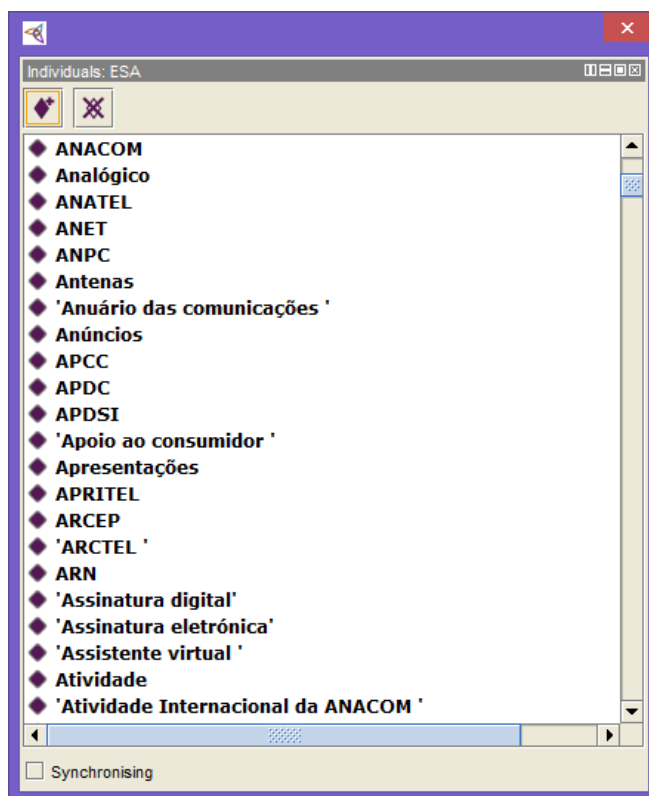


Figura 13 - Lista de Termos

O próximo passo seria pegar em cada termo e efetuar a sua etiquetagem segundo o modelo SKOS. Para o efeito utilizamos as etiquetas preferenciais (`skos:prefLabel`), alternativas (`skos:altLabel`) e ocultas (`skos:hiddenLabel`) em dois idiomas, português e inglês.

Seguidamente ilustramos com recurso a figuras, os termos que evidenciam os rótulos SKOS acima mencionados.

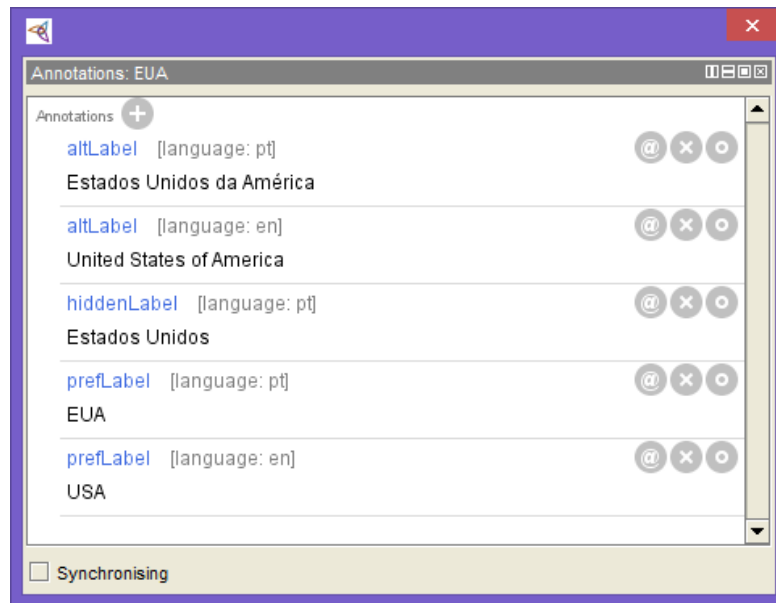


Figura 14 - Etiquetagem do termo EUA

Como se pode verificar na Figura 14, o termo EUA, foi etiquetado de acordo com o modelo SKOS, recorrendo aos três tipos de etiquetas e em dois idiomas diferentes.

OntoGraf

Com recurso ao *plug-in* OntoGraf, poderemos visualizar graficamente qualquer conceito (termo). Estando na vista do OntoGraf, ao passarmos o rato em cima de qualquer conceito, obtemos informações que estão associados a este mesmo conceito. É-nos dado a conhecer o URI desse termo, e as suas anotações.

Tomemos ainda como exemplo o termo EUA, podemos observar na Figura 15 que o termo em questão possui 5 anotações.

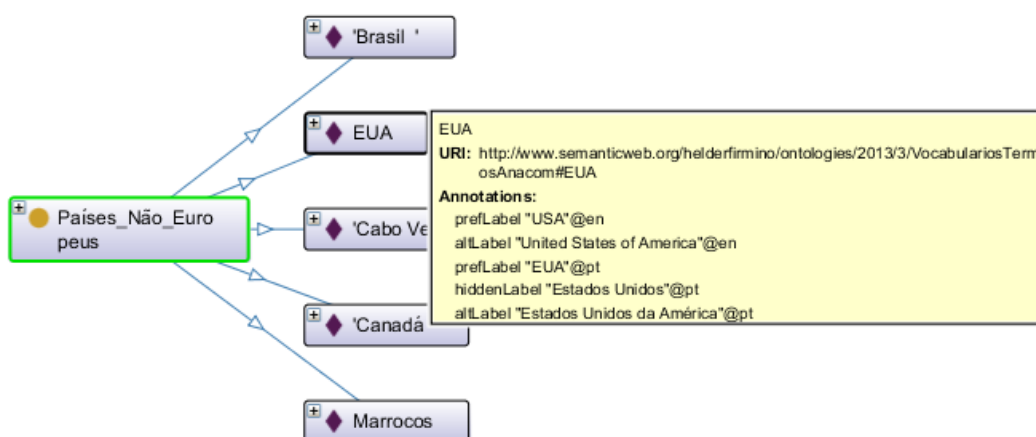


Figura 15 - Vista OntoGraf do termo EUA

Seguidamente ilustramos relações hierárquicas entre termos constantes do nosso vocabulário controlado. Como foi anteriormente referido, as relações hierárquicas permitem a representação de um termo mais específico para um termo mais geral. Para isso utilizamos as propriedades padrão do SKOS, o `skos:broader` e o `skos:narrower`.

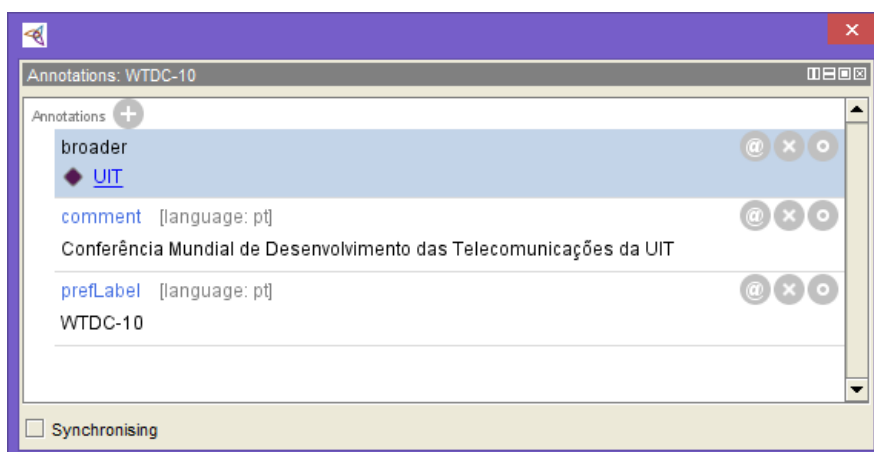


Figura 16 - Relação semântica hierárquica (propriedade `skos:broader`)

No caso da Figura 16 ilustramos o termo “WTDC-10”, que é um termo cujo `broader` é outro termo, no caso a “UIT”. Como gozam de propriedade transitiva pode-se inferir o caso inverso, que poderá ser confirmado pela Figura 17.



Figura 17 - Relação semântica hierárquica (propriedade skos:narrower)

Como é visível na Figura 17, o termo UIT é um termo que tem um vínculo com o termo WTDC-10, que é consubstanciada pela propriedade skos:narrower.

Protégé SPARQL *Plug-in*

O Protégé possui um SPARQL *plug-in* que nos permite manipular a informação armazenada no sistema. Para se obter qualquer resultado, basta escrever a cláusula respectiva, selecionando o triplo que se pretende obter qualquer informação.

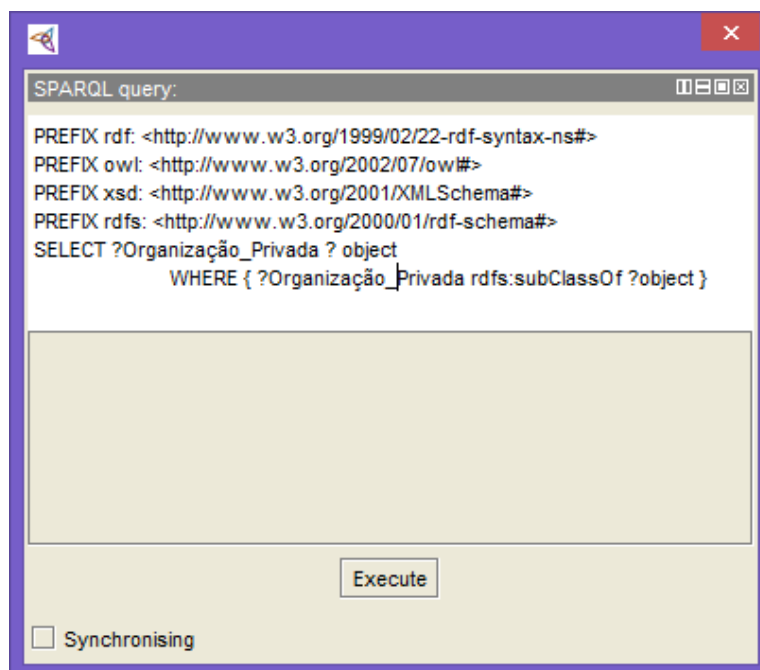


Figura 18 - Protégé SPARQL *Plug-in*

Como se pode observar na Figura 18, é possível obter a superclasse da classe Organização_Privada, através da propriedade RDF Schema, rdfs:subClassOf, onde se pode ver igualmente o PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>.

Da mesma forma que se obteve um resultado aplicando uma propriedade RDFS, pode-se igualmente obter resultados para cláusulas que contenham propriedades SKOS, basta para isso adicionar o PREFIX skos:<http://www.w3.org/2004/02/skos/core#>.

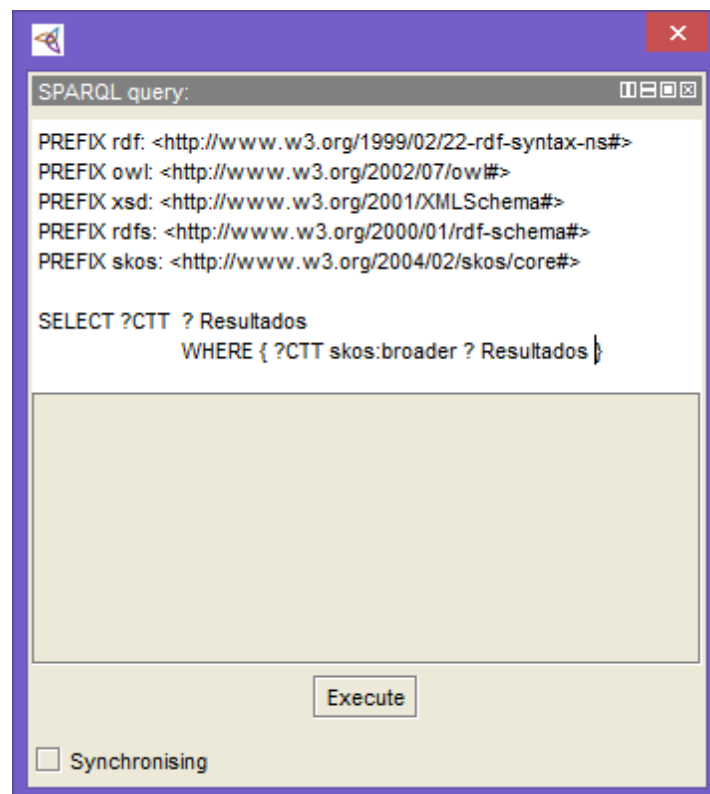


Figura 19 – Interrogação SPARQL com a propriedade skos:broader

Com esta *querie*, ilustrado na Figura 19, iremos obter como resultado todos os termos que sejam *broader* do termo CTT.

4.3 Discussão

Terminado todo o processo de organização dos termos com recurso a propriedades SKOS e também de outros vocabulários (*Dublin Core*, *Geonames*, etc), estamos em condições de afirmar que seguindo todas as recomendações, quer das metodologias (Metodologia da ISA e Metodologia de Noy e McGuinness), quer dos Esquemas de Interoperabilidade (Níveis de Interoperabilidade *Dublin Core* e *Five Star Deployment Scheme*) é possível conseguir uma web de dados, totalmente ligada e em acesso livre, como fora idealizado pelos mentores da web semântica.

Para que isso seja uma realidade, é preciso alguns pressupostos, que estão descritos nas metodologias e tecnologias, como por exemplo, todo o recurso possuir um identificador único (URI), para que esse recurso possa referenciar a outro recurso e ser referenciado.

Todos os nossos termos possuem um URI, que os identifica, tornando possível por exemplo, utilizando predicados SKOS, tais como `skos:exactMatch`, ou eventualmente predicados OWL, como o `owl:sameAs`, realizar *matching* para outros vocabulários.

Tomemos como exemplo o vocabulário *Geonames*. Podemos ter na nossa organização um termo (país) de nome Portugal e pretendemos utilizar o mesmo termo (Portugal) definido naquele vocabulário (*geonames*), basta para isso apontar para esse URI, utilizando predicados adequados.

`<http://sws.geonames.org/2264397/>` é o URI que identifica Portugal no vocabulário *Geonames*. Para efectuar um *matching* com o nosso vocabulário podemos ter o seguinte triplo:

```
base:Portugal skos:closeMatch <http://sws.geonames.org/2264397/>
```

Em que *base* é o *namespace* do nosso vocabulário, `skos:closeMatch` é o predicado e o objeto pretendido é `<http://sws.geonames.org/2264397/>`. Ou podemos utilizar um predicado OWL, em que teríamos:

```
base:Portugal owl:sameAs <http://sws.geonames.org/2264397/>
```

Onde base é o *namespace* do nosso vocabulário, owl:sameAs é o predicado e o objeto pretendido é <http://sws.geonames.org/2264397/>.

Acabamos de mostrar dois exemplos práticos de como relacionar termos de vocabulários diferentes. Um com uma propriedade do SKOS e outro com uma propriedade do OWL.

Todos os termos do vocabulário controlado proposto foram etiquetados em pelo menos uma propriedade, perfazendo um total de aproximadamente cinco centenas de termos.

Utilizamos essencialmente propriedades SKOS, mescladas com propriedades RDFS, OWL e Dublin Core. skos:prefLabel, skos:altLabel, skos:hiddenLabel, skos:broader, skos:narrower, skos:related; rdfs:comment, rdfs:subClassOf, dc:title, dc:author, dc:contributor; dc:date. Disponibilizamos no Apêndice A, trechos de código no formato RDF/Turtle, onde poderemos encontrar todas as propriedades acima referidas.

Um outro objetivo proposto era o de rotular os termos em dois idiomas diferentes, em português e em inglês. Este desiderato foi conseguido, uma vez que se conseguiu etiquetar a maioria dos quinhentos termos identificados no *website* da ANACOM (ver Figura 20).

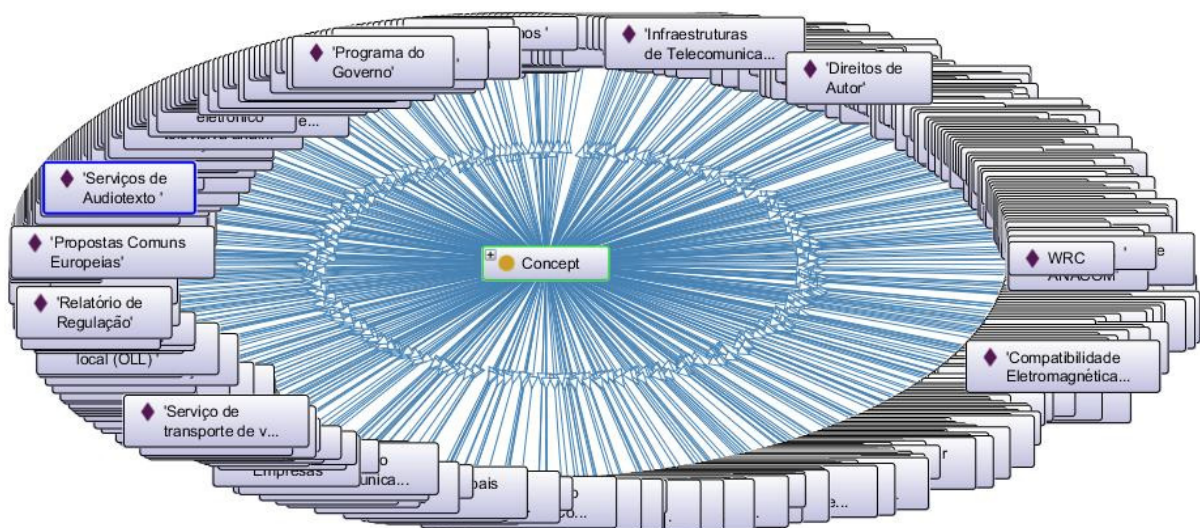


Figura 20 – Vista Ontograf de todos os termos

Reconhecemos igualmente limitações ao nosso trabalho, que se prendem ao facto de não ser possível, no momento imediato, a pesquisa dos nossos termos diretamente da Web, mas que poderá ser uma atividade a ser tratada no futuro, nomeadamente a migração do nosso vocabulário para o Web Protégé.

Web Protégé é um editor de ontologias e uma ferramenta de aquisição de conhecimento para a Web, desenvolvido por Tudorache, Nyulas, Noy, & Musen (2011). É de acesso livre e *open source*.

Uma outra limitação que podemos apontar ao nosso trabalho, reside no facto de que o SKOS, por ser uma tecnologia vocacionada para modelação de ambientes simples não permite modelar certas situações. Nessas situações devemos combinar com outras tecnologias, nomeadamente com a utilização de construtores RDFS e OWL, só para citar alguns. Embora seja uma limitação, o SKOS tenta mitigar esta faceta mais simplista, aproveitando das vantagens de estar assente em RDF, permitindo extensões a outras redes de informação ao longo da Web.

5. CONCLUSÃO E TRABALHO FUTURO

Chegado à esta fase da dissertação podemos afirmar que todo o conhecimento adquirido foi fruto de árduo trabalho, onde foi-nos transmitido muito conhecimento, conseguimos novas conquistas, houve lugar a encontros, mas também há que assinalar alguns desencontros.

Com este trabalho introduzimos alguns princípios da Web Semântica e de *Linked Open Data*, que definem como devemos publicar e interligar dados na Web, que sejam entendidos quer por nós humanos, quer pelas máquinas. O nosso objetivo principal era o de propor uma organização dos termos do website da ANACOM, alicerçada em tecnologias da Web Semântica. Para isso, partimos de uma estrutura não organizada e com recurso à ferramenta Protégé, conseguimos modelar todo o nosso vocabulário.

No capítulo introdutório foi feito o enquadramento do trabalho, onde explicamos a relevância do problema que tencionávamos resolver. Partindo do problema (termos não organizados), enunciamos os objetivos do trabalho, relacionando-os com o contexto apresentado. Apresentamos igualmente a motivação e a estruturação (organização) da dissertação. Ainda dentro do capítulo inicial demos conta da publicação de artigos, que foram submetidos a duas conferências internacionais.

O capítulo 2, Fundamentos, serviu para discorrermos sobre assuntos, temas e tecnologias relacionadas com a Web Semântica, onde fizemos uma revisão da literatura, sobre temáticas relacionadas com o nosso campo de estudo. Tivemos acesso a diversos documentos de autores, que antes de nós estudaram e abordaram as mesmas questões. Focamos igualmente no Sistema de Organização de Conhecimento, o SKOS, que era a base tecnológica principal do nosso trabalho, onde destacamos as principais propriedades, àquelas que iriam sustentar todo o processo de organização de termos.

No capítulo 3, introduzimos o método de investigação (muitas vezes referido em Sistemas de Informação como método em vez de metodologia) que iria nos auxiliar em todo o processo de dissertação. A nossa escolha recaiu sobre o método de investigação

qualitativo, que a nosso ver era o método que melhor se adequava ao problema que tínhamos que dar resposta. Identificamos um conjunto de metodologias de construção de Ontologias e de Vocabulários Controlados, de onde destacaríamos as metodologias *Ontology Development 101*, de Noy & McGuinness e a Metodologia *Core Vocabularies* do Programa Europeu, ISA, que serviram de base à construção do nosso vocabulário.

No capítulo 4, mostramos o resultado do trabalho que foi conseguido com o auxílio da ferramenta Protégé. Trata-se de uma das ferramentas mais completas e amplamente recomendada. Aliás, uma das metodologias que orientaram o nosso trabalho, sugere a adoção desta ferramenta. Ainda no capítulo 4, efetuamos uma discussão dos resultados obtidos, dando especial ênfase à característica diferenciadora do SKOS, que é a capacidade de relacionar vocabulários, que estejam publicados na Web, segundo os princípios da Web de Dados.

Achamos importante a inclusão de apêndices e anexos, com o objetivo de complementar a informação tratada ao longo dos capítulos. No caso dos apêndices, disponibilizamos: algumas partes do código gerado, em formato RDF/Turtle, do vocabulário proposto; exemplos de vocabulários SKOS; tabela com indicação dos prefixos de *namespaces* mais comuns e exemplos de fontes de *Linked Data*. Adicionamos ainda dois anexos relativos à listagem inicial (em linguagem natural) dos termos do *website* da ANACOM.

A consciência que um trabalho nunca é um produto final e acabado deverá nortear qualquer atividade humana. Deve-se reconhecer que há sempre lugar a melhoramentos e outrossim devemos sinalizar os limites do trabalho proposto. Derivado à própria natureza do ramo de ciência da computação em constante mudança, onde novas tecnologias surgem a uma velocidade alucinante é quase impossível não apontar pistas para trabalhos futuros.

No caso específico do nosso vocabulário controlado identificamos uma série de questões que poderão ser exploradas no futuro, que listamos de seguida:

- Incluir outros idiomas, nomeadamente a língua castelhana (trata-se de uma língua que é amplamente utilizada como língua nativa, especialmente na América latina) e o alemão (pela sua importância crescente na Europa, como economia pujante, com especial influência na Europa do Leste e em toda a Europa de uma forma geral);

- O desenvolvimento de um ambiente específico de edição e manutenção dos vocabulários produzidos e sua integração com outras ferramentas e tecnologias existentes e acreditadas pela comunidade web semântica;
- O desenho e implementação de uma estratégia integrada de web semântica, integrando todos os *stakeholders*, especialmente outras autoridades de comunicações, instituições governamentais, fornecedores de serviços e consumidores finais.

6. REFERÊNCIAS

- Allemang, D., & Hendler, J. (2011). *Semantic Web for the Working Ontologist Effective Modeling in RDFS and OWL* (Second Edition.). Morgan Kaufmann Publishers.
- Assem, M. van. (2010). *Converting and Integrating Vocabularies for the Semantic Web* (Tese de Doutorado). Vrije Universiteit.
- Baker, T., Allinson, J., Johnston, P., Miyazawa, A., Rühle, S., & Tennis, J. (2012, June 14). DC Terms. Retrieved from <http://dublincore.org/documents/dcmi-terms/>
- Baonza, M. del C. (2010). *NeOn Methodology for Building Ontology Networks: Specification, Scheduling and Reuse* (Tese de Doutorado). Universidad Politécnica de Madrid, Madrid. Retrieved from http://oa.upm.es/3879/2/MARIA_DEL_CARMEN_SUAREZ_DE_FIGUEROA_BAONZA.pdf
- Baptista, A. A. R. P. (2002). *Informattica Online Um Enquadramento para a Publicação em Linha de Revistas Científicas Electrónicas* (Tese de Doutorado). Universidade do Minho, Guimarães.
- Berndtsson, M., Hansson, J., Olsson, B., & Lundell, B. (2008). *Thesis Projects A Guide for Students in Computer Science and Information Systems* (2nd ed.). Springer.
- Berners-Lee, T. (2010). Linked Open Data. Retrieved January 20, 2013, from <http://www.w3.org/DesignIssues/LinkedData.html>
- Bôaventura, R. S. (2009). *Web Semântica e Ontologias: Teoria e Aplicação*. Faculdade Politécnica de Uberlândia.

- Catarino, M.-E. (2009). *Integração das folksonomias nos metadados: identificação de novos elementos como contributo para a descrição de recursos em repositórios*. Universidade do Minho, Guimarães. Retrieved from http://repositorium.sdum.uminho.pt/bitstream/1822/9564/1/Tese_CatarinoMElisabete.pdf
- Catarino, M.-E., & Baptista, A.-A. (2010). Ontologia STAP: um vocabulário de termos de metadados. In *Inovação e inclusão social: questões contemporâneas da informação*. Rio de Janeiro. Retrieved from <http://congresso.ibict.br/index.php/enancib/xienancib/paper/view/213>
- Codina, L., & Pedraza-Jiménez, R. (2011). Tesoros y ontologías en sistemas de información documental. *Profesional De La Informacion*, 20(5), 555–563.
- Davies, J., Studer, R., & Warren, P. (2006). *Semantic Web Technologies Trends and Research in Ontology-based Systems*. England: John Wiley & Sons Ltd.
- Eckert, K. (2013). Provenance and Annotations for Linked Data. In *International Conference on Dublin Core and Metadata Applications*. Lisboa, Portugal.
- Fahmi, I., Ellermann, H., Scholing, P., & Zhang, J. (2008). Improving Berrypicking Techniques Using Semantically Similar Information in a Historical Digital Library System. In *Proceedings of the First International Workshop, SIEDL 2008*. Tenerife, Spain.
- Gonçalves, V. (2007). *A web semântica no contexto educativo: um sistema para a recuperação de objectos de aprendizagem baseado nas tecnologias para a web semântica, para o e-learning e para*

os agentes. Universidade do Porto - Faculdade de Engenharia, Porto. Retrieved from https://bibliotecadigital.ipb.pt/bitstream/10198/620/1/tese_phd_vg.pdf

Gray, A. J. G., Gray, N., & Ounis, I. (2009). Searching and exploring controlled vocabularies. In *Proceedings of the WSDM '09 Workshop on Exploiting Semantic Annotations in Information Retrieval* (pp. 1–5). New York, NY, USA: ACM. doi:10.1145/1506250.1506252

Grüniger, M., & Fox, M. S. (1995). Methodology for the Design and Evaluation of Ontologies.

Guarino, N., Oberle, D., & Staab, S. (2009). What is an ontology? *Handbook on Ontologies*. doi:10.1007/978-3-540-92673-3

Harris, S., & Seaborne, A. (2013, March). SPARQL 1.1 Query Language. Retrieved from <http://www.w3.org/TR/sparql11-query/>

Hassanzadeh, O. (2011). Introduction to Semantic Web Technologies & Linked Data.

Heath, T., & Bizer, C. (2011). *Linked Data: Evolving the Web into a Global Data Space. Synthesis lectures on the Semantic Web: Theory and Technology* (1st Edition., Vol. 1). Morgan & Claypool. Retrieved from <http://linkeddatabook.com/editions/1.0/>

Hedden, H. (2010). Taxonomies and controlled vocabularies best practices for metadata. *Journal of Digital Asset Management*, 6(5), 279–284. doi:10.1057/dam.2010.29

Hendler, J. (2009). Web 3.0 Emerging. IEEE Computer Society.

Hickson, I. (2011). HTML Microdata. W3C. Retrieved from <http://dev.w3.org/html5/md-LC/>

- Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P. F., & Rudolph, S. (2012). OWL 2 Web Ontology Language Primer (Second Edition). Retrieved from <http://www.w3.org/TR/2012/REC-owl2-primer-20121211/>
- Isaac, A., & Summers, E. (2009, August). SKOS Simple Knowledge Organization System Primer. Retrieved November 5, 2012, from <http://www.w3.org/TR/2009/NOTE-skos-primer-20090818/>
- Kaplan, B., & Maxwell, J. A. (1994). Qualitative Research Methods for Evaluating Computer Information Systems. *Evaluating Health Care Information Systems: Methods and Applications*, pp. 45–68. Thousand Oaks, CA.
- Lévy, P. (1999). *Cibercultura*. São Paulo: Editora 34.
- Librelotto, G., Ramalho, J., & Henriques, P. (2005). Representação de Conhecimento na Semantic Web. In *A Universalidade da Computação: Um Agente de Inovação e Conhecimento*. UNISINOS - São Leopoldo/RS. Retrieved from <http://www.lbd.dcc.ufmg.br/colecoes/jai/2005/001.pdf>
- Manola, F., & Miller, E. (2004, October 1). RDF Primer. Retrieved November 5, 2012, from <http://www.w3.org/TR/2004/REC-rdf-primer-20040210/>
- McGuinness, D. (2003). *Ontologies Come of Age*. MIT Press. Retrieved from <http://www.ksl.stanford.edu/people/dlm/papers/ontologies-come-of-age-mit-press-%28with-citation%29.htm>
- Méndez, E., & Greenberg, J. (2012). Linked Data for Open Vocabularies and HIVE's global framework. *El Profesional De La Información*, 21(3). Retrieved from

<http://www.elprofesionaldelainformacion.com/contenidos/2012/mayo/mendez-greenberg.html>

Moreira, A. (2003). *Tesauros e ontologias: Estudo de definições presentes na literatura das áreas das Ciências da Computação e da Informação, utilizando-se o método analítico-sintético*. Universidade Federal de Minas Gerais, Belo Horizonte.

Myers, M. (1997). Qualitative Research in Information Systems. *MIS Quarterly*, 2(21), 241 – 242.

Nilsson, M., Baker, T., & Johnston, P. (2009, May). Interoperability Levels for Dublin Core Metadata. Recomendação da DCMI. Retrieved June 9, 2010, from <http://dublincore.org/documents/interoperability-levels/>

Noy, N., & McGuinness, D. (2001). Ontology development 101:A Guide to Creating Your First Ontology. Stanford University. Retrieved from <http://www.ksl.stanford.edu/people/dlm/papers/ontology101/ontology101-noy-mcguinness.html>

Ontologia e a Informação. (2011). Retrieved from <http://vimeo.com/28773336>

Pastor-Sanchez, J.-A., Martinez-Mendez, F.-J., & Rodriguez-Munoz, J.-V. (2012). SKOS application for interoperability of controlled vocabularies in the field of linked open data. *PROFESIONAL DE LA INFORMACION*, 21(3), 245–253.

Pereira, P. H. R. (2009). *Extensible metadata repository for information systems*. Universidade Nova de Lisboa - Faculdade de Ciências e Tecnologia, Lisboa. Retrieved from http://run.unl.pt/bitstream/10362/2290/1/Pereira_2009.pdf

- Prazeres, C. (2009). *Serviços Web Semânticos: da modelagem à composição* (Tese de Doutorado). Universidade de São Paulo, São Carlos.
- PwC EU Services EESV. (2012). *D3.1 – Process and Methodology for Core Vocabularies* (p. 59). Brussels: European Commission. Retrieved from http://joinup.ec.europa.eu/sites/default/files/D3.1-Process%20and%20Methodology%20for%20Core%20Vocabularies_v1.01.pdf
- Sewald, E., Rotta, M. J. R., Silva, E. R. G. da, Gauthier, F. O., Carmo, V. M., & Rover, A. J. (2013). Processo de Desenvolvimento de Ontologia de Domínio da Área Jurídica, Caso Acidentes de Trânsito. In *8ª Conferência Ibérica de Sistemas e Tecnologias de Informação* (Vol. II, p. 366). Lisboa, Portugal: AISTI - Associação Ibérica de Sistemas de Informação e Tecnologias de Informação.
- Silva, D., Souza, R., & Almeida, M. (2008). Ontologias e vocabulários controlados: comparação de metodologias para construção. *Revista Ciência da Informação*, 37(3), 60–75.
- Silva, D. L. da. (2008). *Uma Proposta Metodológica para Construção de Ontologias: Uma perspectiva interdisciplinar entre as Ciências da Informação e da Computação*. Universidade Federal de Minas Gerais, Belo Horizonte. Retrieved from <http://www.bibliotecadigital.ufmg.br/dspace/handle/1843/ECID-7NRQZ2>
- Stuckenschmidt, H. (2003). *Information Sharing on the Semantic Web*. Springer.
- Tarasova, T. (2012). *Publicação de linked data: o caso de uso Pordata*. Universidade Nova de Lisboa - Faculdade de Ciências e Tecnologia, Lisboa. Retrieved from http://run.unl.pt/bitstream/10362/7859/1/Tarasova_2012.pdf

- Tennis, J. T. (2007). Scheme Versioning in the Semantic Web. *Cataloging & Classification Quarterly*, 43(3-4), 85–104. doi:10.1300/J104v43n03_05
- Thuraisingham, B. (2002). *XML databases and the semantic web*. Boca Raton, USA: CRC Press.
- Tudorache, T., Nyulas, C., Noy, N. F., & Musen, M. A. (2011). WebProtégé: A Collaborative Ontology Editor and Knowledge Acquisition Tool for the Web. Harith Alani, Knowledge Media institute, The Open University, UK.
- Uschold, M., & King, M. (1995). Towards a Methodology for Building Ontologies. Artificial Intelligence Applications Institute, University of Edinburgh. Retrieved from http://staff.um.edu.mt/cabe2/lectures/webscience/docs/uschold_king_95.pdf
- W3C SPARQL Working Group. (2013, March). SPARQL 1.1 Overview. Retrieved from <http://www.w3.org/TR/sparql11-overview/>

7. APÊNDICES

Apêndice A – Trechos de Código do Vocabulário Controlado no formato RDF/Turtle

```
@prefix owl: <http://www.w3.org/2002/07/owl#>
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
@prefix xml: <http://www.w3.org/XML/1998/namespace>
@prefix xsd: <http://www.w3.org/2001/XMLSchema#>
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
@base <http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM> rdfs:type :Ontology

<http://purl.org/dc/terms/title> "Controlled vocabulary of terms used in the ANACOM website
(Communications Authority of Portugal)"@en ;
<http://purl.org/dc/terms/date> "2013-05-27"@pt ;
<http://purl.org/dc/terms/contributor> "Ana Alice Baptista"@pt ;
<http://purl.org/dc/terms/creator> "Helder Firmino"@pt ;
<http://purl.org/dc/terms/title> "Vocabulário controlado da lista de termos do website da ANACOM
(Autoridade de Comunicações de Portugal)"@pt ;

:imports <http://www.w3.org/TR/skos-reference/skos-owl1-dl.rdf> .

#####
#
# Annotation properties
#
#####

### http://purl.org/dc/terms/contributor
<http://purl.org/dc/terms/contributor> rdfs:type :AnnotationProperty .

### http://purl.org/dc/terms/creator
<http://purl.org/dc/terms/creator> rdfs:type :AnnotationProperty .

### http://purl.org/dc/terms/date
<http://purl.org/dc/terms/date> rdfs:type :AnnotationProperty .

### http://purl.org/dc/terms/title
<http://purl.org/dc/terms/title> rdfs:type :AnnotationProperty .

### http://www.w3.org/2004/02/skos/core#altLabel
<http://www.w3.org/2004/02/skos/core#altLabel> rdfs:type :AnnotationProperty .

### http://www.w3.org/2004/02/skos/core#broadMatch
<http://www.w3.org/2004/02/skos/core#broadMatch> rdfs:type :AnnotationProperty .

### http://www.w3.org/2004/02/skos/core#broader
<http://www.w3.org/2004/02/skos/core#broader> rdfs:type :AnnotationProperty .
```

<http://www.w3.org/2004/02/skos/core#broaderTransitive>
<<http://www.w3.org/2004/02/skos/core#broaderTransitive>> **rdf:type** :AnnotationProperty .

<http://www.w3.org/2004/02/skos/core#changeNote>
<<http://www.w3.org/2004/02/skos/core#changeNote>> **rdf:type** :AnnotationProperty .

<http://www.w3.org/2004/02/skos/core#closeMatch>
<<http://www.w3.org/2004/02/skos/core#closeMatch>> **rdf:type** :AnnotationProperty .

<http://www.w3.org/2004/02/skos/core#editorialNote>
<<http://www.w3.org/2004/02/skos/core#editorialNote>> **rdf:type** :AnnotationProperty .

<http://www.w3.org/2004/02/skos/core#exactMatch>
<<http://www.w3.org/2004/02/skos/core#exactMatch>> **rdf:type** :AnnotationProperty .

<http://www.w3.org/2004/02/skos/core#hasTopConcept>
<<http://www.w3.org/2004/02/skos/core#hasTopConcept>> **rdf:type** :AnnotationProperty .

<http://www.w3.org/2004/02/skos/core#hiddenLabel>
<<http://www.w3.org/2004/02/skos/core#hiddenLabel>> **rdf:type** :AnnotationProperty .

<http://www.w3.org/2004/02/skos/core#historyNote>
<<http://www.w3.org/2004/02/skos/core#historyNote>> **rdf:type** :AnnotationProperty .

<http://www.w3.org/2004/02/skos/core#inScheme>
<<http://www.w3.org/2004/02/skos/core#inScheme>> **rdf:type** :AnnotationProperty .

<http://www.w3.org/2004/02/skos/core#mappingRelation>
<<http://www.w3.org/2004/02/skos/core#mappingRelation>> **rdf:type** :AnnotationProperty .

<http://www.w3.org/2004/02/skos/core#member>
<<http://www.w3.org/2004/02/skos/core#member>> **rdf:type** :AnnotationProperty .

<http://www.w3.org/2004/02/skos/core#narrower>
<<http://www.w3.org/2004/02/skos/core#narrower>> **rdf:type** :AnnotationProperty .

<http://www.w3.org/2004/02/skos/core#narrowerTransitive>
<<http://www.w3.org/2004/02/skos/core#narrowerTransitive>> **rdf:type** :AnnotationProperty .

<http://www.w3.org/2004/02/skos/core#notation>
<<http://www.w3.org/2004/02/skos/core#notation>> **rdf:type** :AnnotationProperty .

<http://www.w3.org/2004/02/skos/core#note>
<<http://www.w3.org/2004/02/skos/core#note>> **rdf:type** :AnnotationProperty .

<http://www.w3.org/2004/02/skos/core#prefLabel>
<<http://www.w3.org/2004/02/skos/core#prefLabel>> **rdf:type** :AnnotationProperty .

<http://www.w3.org/2004/02/skos/core#related>
<<http://www.w3.org/2004/02/skos/core#related>> **rdf:type** :AnnotationProperty .

<http://www.w3.org/2004/02/skos/core#relatedMatch>
<<http://www.w3.org/2004/02/skos/core#relatedMatch>> **rdf:type** :AnnotationProperty .

<http://www.w3.org/2004/02/skos/core#semanticRelation>

<http://www.w3.org/2004/02/skos/core#semanticRelation> **rdf:type** :AnnotationProperty .

http://www.w3.org/2004/02/skos/core#topConceptOf

<http://www.w3.org/2004/02/skos/core#topConceptOf> **rdf:type** :AnnotationProperty .

#####

#

Classes

#

#####

<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/

VocabulariosTermosANACOM#ConceptScheme> .

<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/

VocabulariosTermosANACOM#Comunicação> **rdf:type** :Class ;

rdfs:subClassOf <http://www.semanticweb.org/helderfirmino/ontologies/2013/3/

VocabulariosTermosANACOM#Collection> .

http://www.semanticweb.org/helderfirmino/ontologies/2013/3/

VocabulariosTermosANACOM#Comunicação_Eletrónica

<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/

VocabulariosTermosANACOM#Comunicação_Eletrónica> **rdf:type** :Class ;

rdfs:subClassOf <http://www.semanticweb.org/helderfirmino/ontologies/2013/3/

VocabulariosTermosANACOM#Comunicação> .

http://www.semanticweb.org/helderfirmino/ontologies/2013/3/

VocabulariosTermosANACOM#Comunicação_Postal

<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/

VocabulariosTermosANACOM#Comunicação_Postal> **rdf:type** :Class ;

rdfs:subClassOf <http://www.semanticweb.org/helderfirmino/ontologies/2013/3/

VocabulariosTermosANACOM#Comunicação> .

http://www.semanticweb.org/helderfirmino/ontologies/2013/3/

VocabulariosTermosANACOM#Equipamentos_e_Tecnologias_de_Comunicações

<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/

VocabulariosTermosANACOM#Equipamentos_e_Tecnologias_de_Comunicações> **rdf:type** :Class ;

rdfs:subClassOf <http://www.semanticweb.org/helderfirmino/ontologies/2013/3/

VocabulariosTermosANACOM#Comunicação_Eletrónica> .

http://www.semanticweb.org/helderfirmino/ontologies/2013/3/

VocabulariosTermosANACOM#Gestão_do_Espectro

<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/

VocabulariosTermosANACOM#Gestão_do_Espectro> **rdf:type** :Class ;

rdfs:subClassOf <http://www.semanticweb.org/helderfirmino/ontologies/2013/3/

VocabulariosTermosANACOM#Comunicação_Eletrónica> ;

<http://www.w3.org/2004/02/skos/core#prefLabel> "Spectru"@en , "Espectro"@pt ;

<http://www.w3.org/2004/02/skos/core#broader>

<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/

VocabulariosTermosANACOM#Gestão_do_Espectro> .

http://www.semanticweb.org/helderfirmino/ontologies/2013/3/

VocabulariosTermosANACOM#Mercado_de_Telecomunicações

<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/

```

VocabulariosTermosANACOM#Mercado_de_Telecomunicações> rdf:type :Class ;
rdfs:subClassOf <http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#Comunicação_Eletrónica> ;
rdfs:comment "Esta classe (subclasse de Comunicação_Eletrónica) comporta todos os elementos
relativos ao mercado das telecomunicações."@pt .

### http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#Organizações
<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#Organizações> rdf:type :Class ;
rdfs:subClassOf <http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#Collection> .

### http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#Organizações_Privadas
<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#Organizações_Privadas> rdf:type :Class ;
rdfs:subClassOf http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#Organizações> .

### http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM #Organizações_Públicas
<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#Organizações_Públicas> rdf:type :Class ;
rdfs:subClassOf <http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#Organizações> .

### http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#Países
<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#Pais> rdf:type :Class ;
rdfs:subClassOf <http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#Collection> .

### http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#Países_Europeus
<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#Países_Europeus> rdf:type :Class ;
rdfs:subClassOf <http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#Países> .

### http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#Países_Não_Europeus
<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#Países_Não_Europeus> rdf:type :Class ;
rdfs:subClassOf <http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#Países> ,
[
rdf:type :Class ;
complementOf <http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#Países_Europeus>
] .

### http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#Serviços_de_Telecomunicações

```

<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#Serviços_de_telecomunicações> **rdf:type** :Class ;
rdfs:subClassOf <http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#Comunicação_Eletrónica> .

http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#Comunicações_Fixas
<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#Comunicações_Fixas> **rdf:type** :Class ;
rdfs:subClassOf <http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#Serviços_de_Telecomunicações> .

http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#Comunicações_Móveis
<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#Comunicações_Móveis> **rdf:type** :Class ;
rdfs:subClassOf <http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#Serviços_de_Telecomunicações> .

Individuals

#####

http://www.semanticweb.org/helderfirmino/ontologies/2013/3/VocabulariosTermosANACOM#ACEP

<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#ACEP> **rdf:type**
<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#Concept> ,
<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#Organizações_Privadas> ,
:NamedIndividual ;
<http://www.w3.org/2004/02/skos/core#prefLabel> "ACEP"@pt ;
rdfs:comment "Associação de Comércio Eletrónico de Portugal"@pt ;
<http://www.w3.org/2004/02/skos/core#broader>
<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#Organizações_Privadas> .

http://www.semanticweb.org/helderfirmino/ontologies/2013/3/VocabulariosTermosANACOM#ADSL

<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#ADSL> **rdf:type**
<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#Concept> ,
<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#Equipamentos_e_Tecnologias_de_Comunicações> ,
:NamedIndividual ;
<http://www.w3.org/2004/02/skos/core#prefLabel> "ADSL"@en ;
rdfs:comment "Asymmetric Digital Subscriber Line"@en ;
<http://www.w3.org/2004/02/skos/core#broader>
<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
VocabulariosTermosANACOM#Equipamentos_e_Tecnologias_de_Comunicações> .

(...)

<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/VocabulariosTermosANACOM#AICEP>

```
<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
  VocabulariosTermosANACOM#AICEP> rdf:type
<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
  VocabulariosTermosANACOM#Concept> ,

<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
  VocabulariosTermosANACOM#Organizações_Privadas> ,
  :NamedIndividual ;

  <http://www.w3.org/2004/02/skos/core#prefLabel> "AICEP"@en , "AICEP"@pt ;
  rdfs:comment "Associação Internacional das Comunicações de Expressão Portuguesa"@pt ;

  <http://www.w3.org/2004/02/skos/core#broader>
<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
  VocabulariosTermosANACOM#Organizações_Privadas> .
```

[http://www.semanticweb.org/helderfirmino/ontologies/2013/3/VocabulariosTermosANACOM#AM\(R\)S](http://www.semanticweb.org/helderfirmino/ontologies/2013/3/VocabulariosTermosANACOM#AM(R)S)

```
<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
  VocabulariosTermosANACOM#AM(R)S> rdf:type
<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
  VocabulariosTermosANACOM#Concept> ,
  <http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
  VocabulariosTermosANACOM#Comunicações_Móveis> ,
  :NamedIndividual ;

  rdfs:comment " Aeronautical Mobile (R) Service "@en ;
  <http://www.w3.org/2004/02/skos/core#prefLabel> "AM(R)S"@en , "AM(R)S"@pt ;
  <http://www.w3.org/2004/02/skos/core#broader>
  <http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
  VocabulariosTermosANACOM#Comunicações_Móveis> .
```

<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/VocabulariosTermosANACOM#ANACOM>

```
<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
  VocabulariosTermosANACOM#ANACOM> rdf:type
<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
  VocabulariosTermosANACOM#Concept> ,
  <http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
  VocabulariosTermosANACOM#Organizações_Públicas> ,
  :NamedIndividual ;

  <http://www.w3.org/2004/02/skos/core#prefLabel> "ANACOM"@en , "ANACOM"@pt ;
  rdfs:comment "Autoridade de Comunicações de Portugal"@pt ;
  <http://www.w3.org/2004/02/skos/core#related>
  <http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
  VocabulariosTermosANACOM#Reguladoras> ;

  <http://www.w3.org/2004/02/skos/core#broader>
  <http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
```

VocabulariosTermosANACOM#Organizações_Públicas> .

```
### http://www.semanticweb.org/helderfirmino/ontologies/2013/3/VocabulariosTermosANACOM#ANATEL
<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
  VocabulariosTermosANACOM#ANATEL> rdf:type
<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
  VocabulariosTermosANACOM#Concept> ,
<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
  VocabulariosTermosANACOM#Organizações_Públicas> ,
  :NamedIndividual ;

  <http://www.w3.org/2004/02/skos/core#prefLabel> "ANATEL"@en , "ANATEL"@pt ;
  rdfs:comment "Agência Nacional de Telecomunicações do Brasil"@pt ;
  <http://www.w3.org/2004/02/skos/core#related>
<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
  VocabulariosTermosANACOM#Reguladoras> ,
<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
  VocabulariosTermosANACOM#Brasil> ;
  <http://www.w3.org/2004/02/skos/core#broader>
<http://www.semanticweb.org/helderfirmino/ontologies/2013/3/
  VocabulariosTermosANACOM#Organizações_Públicas> .
```

(...)

Generated by the OWL API (version 3.4.2) <http://owlapi.sourceforge.net>

Apêndice B – Prefixos de *Namespaces* mais comuns

Tabela 5 - Prefixos de *Namespaces* mais comuns

| Prefixo | Namespace do URI | Vocabulário |
|---------|---|---|
| rdf | http://www.w3.org/1999/02/22-rdf-syntax-ns# | RDF core |
| rdfs | http://www.w3.org/2000/01/rdf-schema# | RDF Schema |
| owl | http://www.w3.org/2002/07/owl# | Web Ontology Language |
| skos | http://www.w3.org/2004/02/skos/core# | Simple Knowledge Organization System |
| foaf | http://xmlns.com/foaf/0.1/ | Friend Of A Friend |
| void | http://rdfs.org/ns/void# | Vocabulary of Interlinked Datasets |
| dc | http://purl.org/dc/elements/1.1/ | Dublin Core |
| dcterms | http://purl.org/dc/terms/ | Dublin Core terms |
| scovo | http://purl.org/NET/scovo# | Statistical Core Vocabulary |
| qb | http://purl.org/linked-data/cube# | The Data Cube vocabulary |
| tl | http://purl.org/NET/c4dm/timeline.owl# | The Timeline ontology |
| xsd | http://www.w3.org/2001/XMLSchema# | XML Schema |
| dbpedia | http://dbpedia.org/resource/ | Linked Data version of Wikipedia |
| geo | http://www.w3.org/2003/01/geo/wgs84_pos# | Basic Geo Vocabulary |
| xml | http://www.w3.org/XML/1998/namespace | XML |
| adms | http://www.w3.org/ns/adms# | Asset Description Metadata Schema |
| cc | http://creativecommons.org/ns# | Creative Commons |
| doap | http://usefulinc.com/ns/doap# | Description of a Project |
| rov | http://www.w3.org/ns/regorg# | Registered Organization Vocabulary |
| org | http://www.w3.org/ns/org# | Organization Vocabulary |
| sioc | http://rolfs.org/sioc/ns# | Semantically-Interlinked Online Communities Project |

Apêndice C – Exemplos de Vocabulários SKOS

Desde que o W3C tornou o SKOS uma recomendação *de jure*, muitos vocabulários e tesouros de larga escala foram publicados em SKOS.

Na Tabela 6, listamos alguns vocabulários de larga escala, que utilizam o standard SKOS.

Tabela 6 - Vocabulários publicados em SKOS

| Vocabulário | Entidade promotora |
|--|---|
| Dewey Decimal Classification | Online Computer Library Center (OCLC) |
| Art and Architecture Thesaurus | Getty Research Institute |
| NAL - National Agriculture Library ²⁰ | USDA (Departamento Agricultura dos EUA) |
| Computing Classification System | ACM |
| Agrovoc ²¹ | FAO |
| LCSH - Library of Congress Subject Headings | Library of Congress |
| Eurovoc ²² | União Europeia |

²⁰ NAL – National Agriculture Library (URL: <http://agclass.nal.usda.gov/>)

²¹ AGROVOC – United Nations Agriculture Vocabulary (URL: <http://aims.fao.org/website/AGROVOC-Thesaurus>)

²² Eurovoc (URL: <http://eurovoc.europa.eu>)

DBpedia

Fonte de *linked data* criado para:

- Extração de informação estruturada do Wikipedia, utilizando “*infobox*” dos artigos;
- Estabelecer ligações para outras fontes externas.



Figura 22 - Artigo no Wikipedia

Para evitar que a informação seja ambígua, a extração da informação é realizada a partir dos *infoboxes*, conforme se pode verificar na Figura 23.



Figura 23 - Extração de informação a partir da *infobox* de um artigo do Wikipédia

(fonte: Baseado na apresentação de Anja Jentzsch, disponível em http://www.swib09.de/vortraege/20091124_jentzsch.pdf)

A informação é extraída das caixas de informação, para evitar possíveis situações de ambiguidade. Se a extração de informação for no texto podemos deparar com situações ilustradas na Figura 24.

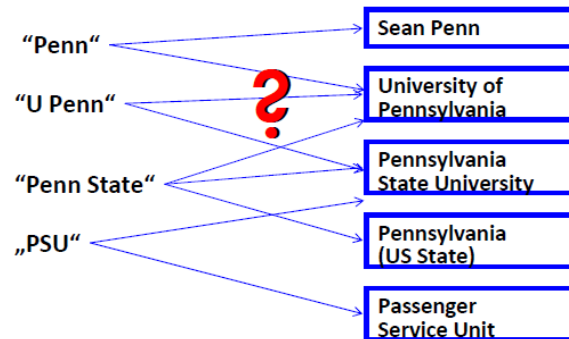


Figura 24 - Desambiguação de entidades

(fonte: Baseado na apresentação de Lauw, Schenkel, Suchanek, Theobald and Weikum, disponível em <http://www.mpi-inf.mpg.de/yago-naga/CIKM10-tutorial/>)

Freebase

Trata-se de um repositório Web livre (aberto, com licença *Creative Commons*), com mais de 12 milhões de objetos (*things*), de quase todo o tipo (filmes, livros, celebridades, lugares, companhias e muito mais). De acordo com Hassanzadeh (2011), possui as seguintes características:

- Centralizada, com uma abordagem focada na publicação de *linked data*;
- Cada objeto tem um identificador global único (GUID);
- Os dados podem ser obtidos em RDF ou JSON;
- APIs apelativos, uma linguagem de manipulação de dados (MQL) e um conjunto de ferramentas que simplificam a edição, publicação ou recuperação de dados.

Seguidamente damos alguns exemplos de aplicações Freebase.

- **Google Refine** – uma poderosa ferramenta para limpeza de dados e descoberta de dados.

- **Powerset** – um motor de busca semântica, que o Freebase utiliza nas suas pesquisas em linguagem natural (foi adquirido pela Microsoft e utilizada no seu motor de busca Bing).
- **Freebase genealogy** – visualizador de árvores genealógicas.
- **FMDb** – Um IMDB Freebase.
- **Freebase sets** – um clone do Google sets que utiliza dados Freebase.
- **Parallax data viewer** – um interface-utilizador alternativo.
- **Freebase Schema Explorer** – um visualizador de ontologias Freebase.
- **2D Visualiser** – um browser para aplicações Java.
- **Thinkbase** – uma ferramenta de exploração visual de grafos.

Linked MDB (Linked Movie DataBase)

De acordo com Hassanzadeh (2011), trata-se da primeira fonte de *Linked Data* dedicada a filmes e assuntos relacionados com filmes.

É publicado pelo Servidor D2R (*Linked Data interface to relational data*) e fornece ligações a outras fontes de *linked data* e *websites* de filmes. É composto por aproximadamente 6 milhões de triplos e cerca de meio milhão de ligações externas.

Em cima do Linked MDB foi construído uma aplicação (aplicação web com menos de 10KB) que permite a geração, por exemplo, de interrogações sobre atores ou atrizes de cinema.

Pode-se utilizar o seu SPARQL *Endpoint*, para efetuar *queries*.

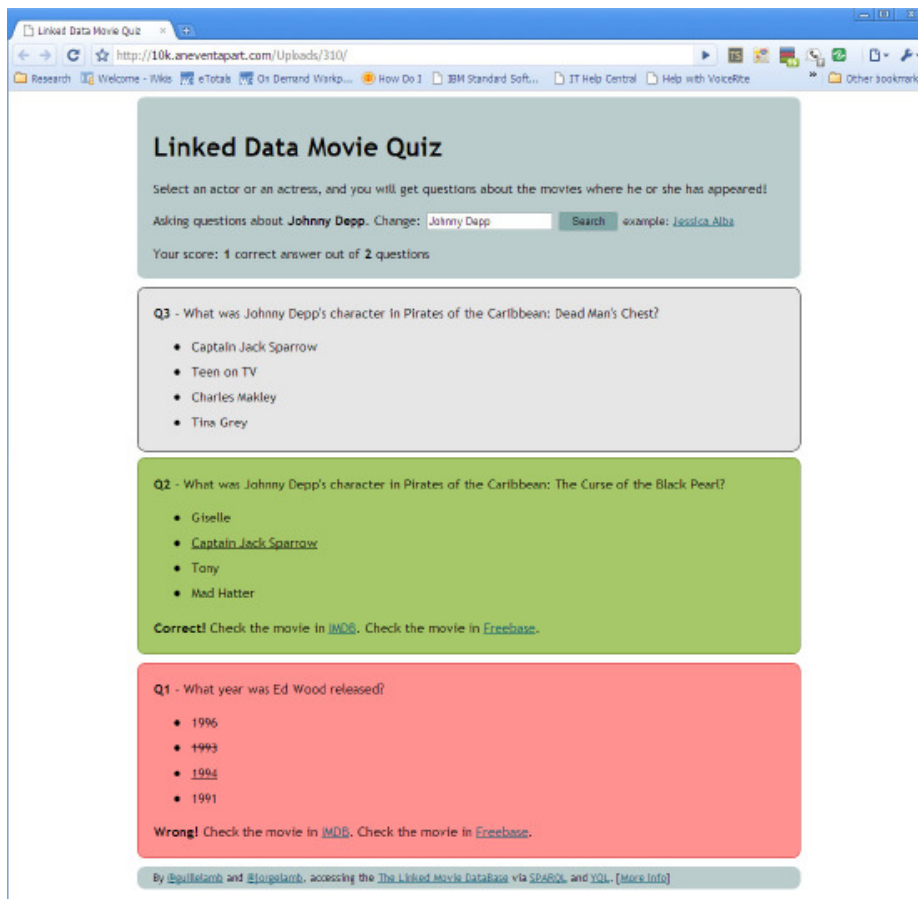


Figura 25 - Pesquisa no Linked MDB, via SPARQL Endpoint (Hassanzadeh, 2011)

Linked Clinical Trials

Segundo Hassanzadeh (2011), faz parte do *Linking Open Drug Data Project* pertencente ao *task force* da W3C, *Health Care and Life Sciences* (HCLS).

Pode-se aceder ao *link* através de <http://linkedct.org>, publicado pelo Servidor D2R, com dados recolhidos de ClinicalTrials.gov. Trata-se de um registo online de experiências clínicas conduzidas nos Estados Unidos e por todo o mundo e publicado em XML.

Fornecer igualmente ligações a outras fontes externas que abordam temas sobre, droga, doenças, etc.

BibBase

Ainda de acordo com o autor, Hassanzadeh (2011), os objetivos da BibBase é tornar mais facilitada a tarefa de manutenção de páginas de publicação para cientistas e grupos de investigação.

- Os utilizadores só precisam manter um ficheiro BibTex, que a BibBase trata do resto;
- A BibBase, publica uma apelativa página HTML parametrizável;
- Publica os dados em RDF, com um *endpoint* SPARQL;
- Interliga suas entradas com a *cloud* LOD.

Trata-se de uma base de dados bibliográfica de alta qualidade, que é mantida graças ao apoio de cientistas.

É um conjunto de dados de incalculável valia para avaliação por deteção dupla e descoberta de conhecimento em ligações semânticas.

Mapa de Sismos Interativo

Com *linked data* de Data.gov.

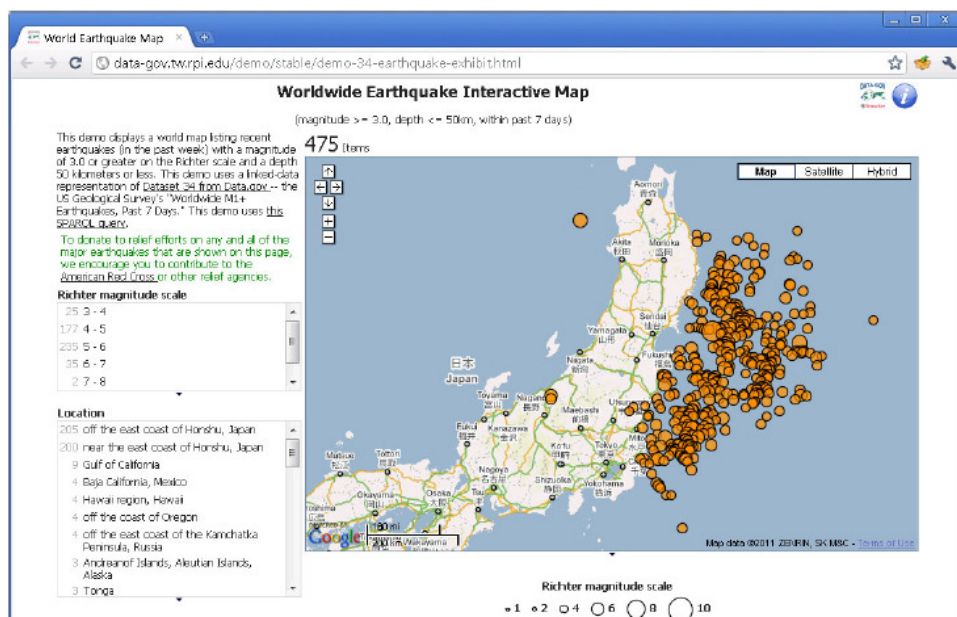


Figura 26 - Mapa interativo de sismos mundiais (Hassanzadeh, 2011)

Outras aplicações *Linked Data*

Browsing

- Marbles
- DBpedia Mobile

Pesquisa

- Falcons

Outros

- Revyu, BBC Music, Pipe
- New York Times Data
- Revyu.com: Review Anything
- Enterprise analytics

Browsers Linked Data

- Tabulator Browser (MIT, USA)
- Marbles (FU Berlin, DE)
- OpenLink RDF Browser (OpenLink, UK)
- Zitgist RDF Browser (Zitgist, USA)
- Humboldt (HP Labs, UK)
- Disco Hyperdata Browser (FU Berlin, DE)
- Fenfire (DERI, Irland)

8. ANEXOS

Anexo A – Lista inicial dos termos do *website* da ANACOM em inglês

| | | | |
|-----------------------------------|------------------------------------|--|---------------------------------------|
| Access (188) | Accessibility (70) | ACEP (2) | Activities Report (68) |
| Activity (4) | Addressing (3) | ADSL (29) | Advisory Committee (2) |
| Advisory Council (17) | Aerials (15) | Aeronautic Mobile Service (4) | AICEP (6) |
| Amateur Satellite Service (12) | Amateur Service (16) | ANACOM (311) | ANACOM events (106) |
| ANACOM images (3) | ANACOM Laboratories (16) | ANACOM publications (486) | Analogue (3) |
| Analytical accounting (34) | ANATEL (4) | ANET (1) | Annual Report and Accounts (76) |
| ANPC (1) | APRITEL (3) | Arbitration Centres for Consumer support (2) | ARCTEL-CPLP (7) |
| Atualidades (1) | Auctions (27) | Audiotext services (51) | Audit (34) |
| Authorisations (4) | Autoridade da Concorrência (2) | Benchmarking (4) | BEREC (91) |
| Broadband (198) | Broadcasting (15) | Budget (1) | BWA (52) |
| Cable (6) | Cable distribution (3) | Call Signs (1) | Calling Line Identification (CLI) (1) |
| Candidatures (1) | CEPT (202) | CERP (27) | Certificates (2) |
| Charging (6) | Chronology (1) | Clarification (49) | COCOM (18) |
| Co-mingled operators (2) | Com-ITU (14) | Committee (6) | Common European Proposals (3) |
| Communications (47) | Communications equipment (37) | Communications Yearbook (6) | Competition (32) |
| Complaints (37) | Conclusions (1) | Conditional access (1) | Conduit access (1) |
| Conferences (28) | Confidentiality in information (2) | Congress (1) | Consulta Pública (1) |
| Consumer (24) | Consumer protection (77) | Consumer support (16) | Contact Network (13) |
| Contracts (2) | Conventions (11) | Convergence (10) | Cooperation (71) |
| Cost models (22) | Costing (1) | Costs (66) | Council (1) |
| Coverage (network) (2) | CPG (1) | CPG PTC (3) | CPG PTD (2) |
| CPLP (1) | CTT (31) | Cybercentres (5) | Cybercrime (3) |
| Cybersecurity (4) | Data (4) | Data transmission services (10) | Debriefings (1) |
| Decisions (1696) | Declaration (61) | DECO (11) | Delegation of powers (2) |
| Determination (25) | Digital (12) | Digital Agenda for Europe (6) | Digital Dividend (27) |
| Digital Library (4) | Digital signature (2) | Digital Television (223) | Digital Terrestrial Television (80) |
| Directive (6) Draft decision (15) | DSL (7) Ducts (12) | ECC decisions (34) | ECC PT SE 42 (1) |
| ECC PT SE 43 (9) | eCall (1) ECC (98) | ECC WG FM PT46 (3) | ECO (12) |
| | ECC TG4 (1) | | |

| | | | |
|---|--|---|---|
| ECTA (3) | e-Government (7) | e-initiatives (4) | E/IRG (20) |
| Elections (2) | Electricity supplies (2) | Electromagnetic Compatibility (13) | Electromagnetic fields (5) |
| Electronic billing (2) | Electronic Certification (2) | Electronic Commerce (76) | Electronic Communications (663) |
| Electronic Money (2) | Electronic payment (3) | Electronic services (8) | Electronic signatures (11) |
| Electronics (7) | EMERG (14) | Emergency number (112) (8) | ENISA (14) |
| ENUM (16) | ERG (25) | ERGP (16) | ERO (7) |
| ESA (4) | Establishments (2) | ETO (3) ETSI (23) | Europe 2020 (1) |
| European Commission (232) | European Council (3) | European Digital Agenda (3) | European Parliament (4) |
| European Union (211) | European Union Law (215) | Eurostat (2) | Eurotariff (2) |
| EUTELSAT (5) | Eutelsat (5) | Events (82) | Expert Group on Consumer Complaints (3) |
| E112 (5) | FAQ (386) | FCCN (4) | Fees (54) |
| Fibre Optic (11) | Fibre-to-the-Home (5) | Fiscalization (4) | Fixed network (3) |
| Fixed wireless systems and others (4) | Forms (5) | Frequencies (57) | Frequencies Use (14) |
| FWA (35) | General Assembly (1) | Government (4) | Green Paper (1) |
| Group of Experts on Electronic Commerce (4) | Groups of Experts on Consumer Complaints (1) | GSM (36) | GSM-R (7) |
| GSS (2) HDTV (1) | High-speed services (5) | Human Resources (2) | IMSO (4) |
| INE (1) | Information (4) | Information and Communication Technologies (34) | Information Society (80) |
| Information Technologies (7) | Infrastructures (3) | Infrastructures for telecommunications (87) | Intelligent Transport (1) |
| Interconnection (203) | International Activity of ANACOM (430) | International organizations (1) | Internet (116) |
| Internet access services (37) | Invoicing (2) | International roaming (117) | IRG (65) |
| ISP (1) | ITED (96) | IPTV (3) | ITU (149) |
| ITU-D (11) | ITU-R (17) | ITSO (14) | ITU-T (8) |
| i2010 strategy (1) | Land Mobile Service (5) | ITUR (41) | Leased vehicles (9) |
| Legislation (599) | Letter Writing (6) | Leased lines (131) | Library (10) |
| Licenses (11) | List of participants (1) | Liberalisation (4) | LLRO (13) |
| LLU (11) | Local Loop Unbundling (47) | Lists (3) | Loop (1) |
| Management Plan (34) | Maritime Mobile Service (6) | Long Term Evolution technology (1) | MCA (7) |
| Meetings (584) | MFRW (3) | Markets (11) | Mobile communication services on board vessels (13) |
| Mobile Networks and Services (252) | Mobile satellite services (27) | MMS (22) | Mobile trunking service (5) |
| Mobile TV (1) | Mobility (2) | Mobile Telephone Service (112) | MOPTC (1) |
| Mobiles (1) | | Monitor (1) | |
| | | Monitoring (11) | |

| | | | |
|------------------------------------|--|--|---|
| Multiplexers (27) | Municipalities set percentages (23) | NATO (2) | Net neutrality (14) |
| Netsonda (2) | Network infrastructure (21) | Network security (8) | Networks (66) |
| News (3415) | Newsletter (14) | Next Generation Access Networks (16) | Next Generation Networks (129) |
| NFAP (11) | NNA (4) | NNP (158) | Notice (3) |
| Notifications (8) | NRA (3) | NTFA (42) | Number (5) |
| Numbering (236) | Numbering ranges (15) | OberCom (6) | Offers (3) |
| Online Services (8) | Operators (1) | OECD (10) | Opinion (12) |
| Order (12) | Origination (1) | Operators/providers (83) | Personal data (23) |
| Personal data protection (6) | Plan for the Prevention of Risks of Corruption (5) | Paid Television (2) | Personal data (23) |
| Portability (145) | Portugal (19) | Plenary (18) | Plenipotentiary (1) |
| Postal services (208) | Post-paid (6) | Postal Directive Committee (7) | Postal service (26) |
| Pre-paid (6) | Presentations (64) | Posts (148) | PP10 (6) |
| Privacy of each user's data (6) | Private (27) | Press releases (140) | Prices (56) |
| Protocols (17) | Providers (7) | Prizes (7) | Programme of Government (1) |
| PT NGN (1) | PT PMS (1) | Provision of voice telephone traffic resale services (2) | PT ADSL Network (2) |
| Public consultations (1027) | Public contracts (8) | PT Revision 2006 (1) | PT SMS (2) |
| Public Sector (1) | Public telecommunications service (18) | Public payphones (30) | Public Procurement (37) |
| Purchases (3) | QSI (4) | Public tender (93) | Publics (7) |
| Questionnaires (32) | Radiations (8) | Publications (516) | Quality of service (123) |
| Radio Licensing (44) | Radio spectrum (20) | Quality (14) | Radio Broadcasting (88) |
| Radio-determination Service (2) | Radiofrequencies (5) | Radio Amateurs (1) | Radiocommunications Satellite Service (3) |
| Recommendation (3) | Recruitment (100) | Radiocommunications (180) | RDAO (81) |
| Regulatel (15) | Regulation (442) | Rates (3) | Reform (1) |
| Relevant markets (193) | RELO (5) Repeal (3) | Reference offer (11) | Regulatory Framework (19) |
| Resale voice telephone traffic (1) | Residential (1) | Regulators (16) | Resale Fixed Telephone Service (1) |
| Revoked legislation (74) | RFID (2) | Report (239) | Resolution of disputes (1) |
| Rights of use of numbers (5) | RIO (74) | Resolution (2) | Rights (8) Rights of use (3) |
| Rules Governing Companies (52) | RUO (61) | Right of use for frequencies (226) | R&TTE (30) |
| Satisfaction Questionnaire (2) | SDSTv (1) | RSPG (17) | Satellite (25) |
| Security/privacy (1) | Selection (8) | Sanctions (6) | Security and Emergency Communications (2) |
| | | Security and Emergency Communications (63) | Services (71) |
| | | Seminars (31) | |

| | | | |
|----------------------------------|---|--|--|
| Serviço de Postos Públicos (1) | Serviço Universal (1) | SIC (77) | Signalling (5) |
| Significant Market Power (20) | Site (1) | SLRO (25) | SMP (1) |
| SMRP-CDMA (3) | SMS (21) | Software (1) | Spam (13) |
| Spectru (160) | Spectrum management (107) | Speech (5) | Standard contracts (16) |
| Standardization (12) | State (3) | State of Communications (7) | Statement (4) |
| Station (17) | Statistics (269) | Studies/Reports (229) | Subscription (4) |
| Subscription television (34) | Summit (2) | Switch-off (77) | Subscription contract (2) |
| Tariff (92) | Tariff Monitor (5) | Technologies (2) | Systems and network information security (7) |
| Telecommunications operators (3) | Telephone directories / information services (25) | Telephone Service at a Fixed Location (375) | Telecommunications (99) |
| Television Broadcasting (4) | Television signal (2) | Telex (2) | Television (195) |
| Termination Charges (5) | Terrestrial (8) | Terrestrial Digital Audio Broadcasting (13) | Tenders (14) |
| Thematic leaflets (8) | Training (5) | Treaty of Lisbon (1) | TETRA (2) |
| TSAG (1) | Twitter (4) | UMIC (2) | Trunking Mobile Service (15) |
| Unbundled (2) | Unbundling (1) | UNESCO (2) | UMTS (83) |
| UPAEP (16) | UPU (28) | URSI (37) | Universal Service (396) |
| Use (6) Users (9) | UWA (2) | UWB (6) | Usability (50) |
| Value added services (31) | Vertical Functional Separation (1) | Videoconference (1) | Vacancy (3) |
| Voice (1) | Voice Mail (4) | Voice Transport Services in Closed User Groups (1) VoIP (40) | Virtual calling card service (14) |
| Web 3.0 (1) | WG FM PT22 (3) | WG NNA (1) | VPN (4) |
| Wholesale offers (12) | Wi-Fi (1) | WIMAX (2) | Wholesale (13) |
| Wireless (6) | Withdrawal (2) | Working Group (5) | Win-back (7) |
| WP 4B (1) | WRC (9) | WSIS (4) | Workshops (17) |
| WTPF-09 (2) | WTSA (2) | WTSA08 (1) | WTDC-10 (2) |
| 4G/4th Generation (51) | | | 3G/3rd Generation (1) |

Anexo B – Lista inicial dos termos do *website* da ANACOM em português

| | | | |
|--------------------------------------|---|--|---|
| Acessibilidade (81) | Acesso (206) | Acesso Universal (2) | ACIST-AET (2) |
| ADSL (32) | Aeronaves (1) | AETTUA (1) | Agenda Digital Europeia (10) |
| AICEP (9) | Alemanha (3) | AM(R)S (2) | ANACOM (513) |
| Analógico (5) | ANATEL (6) | ANET (2) | ANPC (2) |
| Antenas (22) | Anuário das comunicações (6) | Anúncios (10) | APCC (1) |
| APDC (6) | APDSI (6) | Apoio ao consumidor (102) | Apresentações (91) |
| APRITEL (5) | ARCEP (1) | ARCTEL (3) | ARCTEL-CPLP (7) |
| ARN (4) | Assinatura digital (2) | Assinatura eletrónica (18) | Assistente virtual (1) |
| Atividade (16) | Atividade Internacional da ANACOM (482) | Atualidades (4688) | Auditoria (38) |
| Autoridade da Concorrência (12) | Autorizações (75) | Avarias (1) | Balcão Virtual (20) |
| Banda estreita (8) | Banda larga (266) | Bélgica (2) | Benchmarking (4) |
| BEREC (92) | Bibliotecas (16) | Bibliotecas Digitais (4) | BNetZA (1) |
| Boas práticas (1) | Brasil (3) | BWA (55) | Cabo Verde (1) |
| Campo Eletromagnético (8) | Canadá (3) | Candidaturas (8) | Carregamentos (6) |
| CDMA (1) | CENELEC (7) | Centrais (3) | Centros de Arbitragem de Conflitos de Consumo (2) |
| CEPT (348) | CERP (28) | Certificação eletrónica (1) | Chamadas de voz (2) |
| Cibercrime (6) | Cibersegurança (6) | Cimeira (3) | Circuitos alugados (157) |
| CLI (1) | Cloud computing (1) | CMT (2) | Cobertura (de rede) (2) |
| COCOM (20) | Código (1) | Colóquios (1) | Comércio eletrónico (110) |
| Comissão Europeia (247) | Comité Consultivo (4) | Comité da Diretiva Postal (7) | Comités (16) |
| Com-ITU (16) | Compatibilidade Eletromagnética (20) | Competitividade (10) | Compras Públicas (126) |
| ComReg (1) | Comunicações (61) | Comunicações de emergência e segurança (101) | Comunicações Eletrónicas (1078) |
| Comunicações móveis via satélite (3) | Comunicado (7) | Conclusões (4) | Concorrência (25) |
| Concurso Público (407) | Concursos (32) | Condições de oferta (2) | Condutas (16) |
| Conferências (42) | Confidencialidade da informação (2) | Congressos (15) | Conselho (5) |
| Conselho Consultivo (23) | Conselho Europeu (2) | Consulta Pública (1451) | Consumidor (70) |
| Contabilidade | Contratação | Contrato (19) | Contratos de adesão |

| | | | |
|---|---|-----------------------------|-----------------------------------|
| analítica (24) | pública (24) | | (74) |
| Convênios (11) | Convergência (14) | Cooperação (82) | Correios (174) |
| Corrupção (1) | CPG (1) | CPG PTC (4) | CPG PTD (3) |
| CPLP (5) | Cronologia (1) | CRTC (1) | CTT (71) |
| Cursos (8) | Custeio (9) | Custos (61) | Cybercentros (6) |
| Dados (6) | Dados Pessoais (22) | Decisões (2468) | Decisões ECC (146) |
| Declarações (87) | DECO (15) | DECT (2) | Defesa do Consumidor (89) |
| Delegação de Competências (14) | Desagregado (3) | Despachos (19) | Difusão (4) |
| Digital (21) | Diplomas (1) | Direitos Conexos (1) | Direitos de Autor (1) |
| Direitos de utilização de frequências (335) | Direitos de utilização de numeração (8) | Diretiva (9) | Discursos (5) |
| Distribuição (5) | Distribuição por Cabo (8) | Dividendo Digital (30) | DSL (7) |
| eCall (1) | ECC (88) | ECC PT FM 46 (5) | ECC PT SE 42 (13) |
| ECC PT SE 43 (9) | ECC TG3 (1) | ECC TG4 (1) | ECO (17) |
| Economia (1) | Económico (1) | eContent (20) | ECSI (3) |
| ECTA (7) | eGovernment (10) | e.iniciativas (13) | E/IRG (6) |
| Eleições (5) | Eletrónica (1) | Eletrotecnia (2) | Emendas (2) |
| EMERG (14) | Emergência (20) | Emissor (1) | Encontro (2) |
| Endereçamento (3) | ENISA (19) | E-Nova (1) | Entidade Certificadora (141) |
| Entidades formadoras (25) | ENUM (21) | Epistolar (9) | Equipamentos de comunicações (44) |
| ERC (1) | ERG (40) | ERGP (16) | Ericsson (1) |
| ERO (9) | ESA (4) | Esclarecimentos (69) | Espaciais (1) |
| Espanha (3) | Espectro (60) | Espectro Radioelétrico (27) | Estabelecimentos (3) |
| Estação (15) | Estações móveis (1) | Estações terrenas (2) | Estatísticas (312) |
| Estratégia i-2010 (1) | Estudos/Relatórios (339) | eTEN (6) | ETO (4) |
| ETSI (27) | EUA (1) | Europa 2020 (1) | Eurostat (2) |
| Eurotarifa (2) | EUTELSAT (5) | Eutelsat (5) | Eventos (496) |
| Eventos ANACOM (148) | FAQ (524) | Fatura eletrónica (4) | Faturação (6) |
| FCCN (6) | Fibra Ótica (18) | Fiscalização (132) | FM - Frequência Modelada (7) |
| Folhetos/Brochuras (22) | Formação (9) | Formulários (5) | Fórum (4) |
| França (4) | Frequências (70) | FTTH (8) | FWA (46) |
| Gamas de numeração (16) | Gestão do espectro (183) | Governo (5) | Grossistas (53) |
| Grupo de Peritos em Comércio Eletrónico (3) | Grupo de Peritos em Reclamações de Consumidores (5) | Grupos de Trabalho (5) | GSM (52) |
| GSM-R (8) | GSS (2) | HDTV (1) | Holanda (1) |
| IDC (1) | iGOV (1) | Imagens ANACOM (3) | IMSO (4) |
| Indicativos de Chamada (2) | Indústria (1) | INE (2) | Infraestruturas de rede (26) |
| Infraestruturas de Telecomunicações (123) | Inquéritos satisfação (2) | Interligação (234) | Internacional (6) |
| Internet (160) | Investigação e | IPTV (6) | IQS (4) |

| | | | |
|--|---------------------------------------|--------------------------------------|---|
| | Desenvolvimento (2) | | |
| IRG (77) | Irlanda (3) | ISP (1) | ITED (479) |
| ITSO (18) | ITUR (119) | Jornal Oficial da União Europeia (3) | Laboratórios ANACOM (15) |
| Legislação (979) | Legislação Comunitária (243) | Legislação revogada (84) | Leilões (30) |
| Liberalização (6) | Licenciamento radioelétrico (52) | Listas (3) | Listas telefónicas e serviços informativos (23) |
| Livro de reclamações (3) | Livro Verde (2) | Locação de viaturas (26) | Marrocos (1) |
| MCA (7) | Media (6) | Mercados relevantes (232) | Metrologia (1) |
| MMS (32) | Mobilidade (2) | Modelos de custeio (30) | Moeda eletrónica (3) |
| Monitorização (14) | MOPTC (2) | Multimedia (3) | Multiplexers (74) |
| MVNO (2) | Nacional (27) | NATO (2) | Neutralidade da rede (18) |
| NGN (3) | NNA (3) | Normalização (40) | Normas (2) |
| Nota de Imprensa (170) | Notificações (9) | Novas Redes de Acesso (10) | Numeração (340) |
| Número de emergência (112) (10) | OberCom (7) | Observatório de tarifários (9) | OCDE (13) |
| Ofcom (1) | Oferta de referência (12) | Oferta do lacete local (OLL) (59) | Ofertas grossistas (26) |
| OHIM (1) | Operadores coinstalados (4) | Operadores/Prestadores (100) | OPTA (1) |
| ORAC (117) | ORALL (62) | ORCA (26) | Orçamento (2) |
| ORCE (16) | Organizações internacionais (2) | ORLA (28) | Pacotes de Serviços (1) |
| Pagamentos eletrónicos (5) | Paging (1) | Parecer (18) | Parlamento Europeu (6) |
| Plano de Gestão de Riscos de Corrupção (7) | Plano estratégico (45) | Plano Nacional de Numeração (184) | Plenárias (20) |
| Poder de Mercado Significativo (2) | Poder de mercado significativo (24) | Portabilidade (163) | Pós-pago (6) |
| PP10 (6) | Preços (128) | Prémios (12) | Pré-pago (6) |
| PRI (87) | Privacidade (dados) (9) | Privativas (27) | Privatização (1) |
| Programa do Governo (1) | Programas Comunitários (26) | Propostas Comuns Europeias (3) | Proteção (1) |
| Proteção de dados pessoais (7) | Protocolos (23) | PT NGN (1) | PT Revisão 2006 (1) |
| Publicações (1788) | Publicações ANACOM (1714) | Publicidade institucional (4) | QNAF (44) |
| Quadro regulamentar (22) | Qualidade de serviço (288) | Questionários (46) | Radiações (12) |
| Radioamadores (6) | Radiocomunicações (7) | Radiodifusão (7) | Radiodifusão sonora (131) |
| Radiodifusão televisiva (10) | Radiodifusão televisiva analógico (1) | Radiofrequências (7) | Radiomóvel (3) |
| Reclamações (45) | Recomendações (4) | Recrutamento (124) | Recursos humanos (6) |
| Rede ADSL PT (13) | Rede de Contactos (14) | Rede fixa (4) Redes (77) | Redes de nova geração (269) |
| Redes e serviços móveis (291) | Redes públicas (4) | Regime de Empresas (67) | Regional (1) |

| | | | |
|---|---|--|---|
| Regulação (188) | Reguladores (24) | Regulamentos (97) | Regulatel (15) |
| Reino Unido (3) | Relatório de Atividades (67) | Relatório de Regulação (213) | Relatório e Contas (149) |
| Resolução de conflitos (1) | Reuniões (630) | Revenda Serviço Telefónico Fixo (1) | Revogação (25) |
| RFID (2) | Roaming internacional (139) | RSPG (18) R&TTE (39) | Sanções (134) |
| Satélite (36) | SDSTv (1) | Segurança (23) | Segurança das redes (10) |
| Segurança de comunicações (23) | Segurança dos Sistemas e Redes de Informação (29) | Segurança/privacidade (6) | Seguros (6) |
| Seleção (10) | Seminários (48) | Sentidos prováveis de decisão (19) | Separação Vertical Funcional (10) |
| Serviço de Acesso à Internet (83) | Serviço de Amador (20) | Serviço de Amador por Satélite (14) | Serviço de cartões virtuais de chamadas (16) |
| Serviço de cartões virtuais de chamadas (2) | Serviço de Postos Públicos (52) | Serviço de Radiocomunicações por Satélite (15) | Serviço de Radiodeterminação (2) |
| Serviço de Radiodifusão Televisiva (15) | Serviço de revenda de tráfego telefónico de voz (3) | Serviço de transmissão de dados (12) | Serviço de transporte de voz em Grupos Fechados de Utilizadores (1) |
| Serviço Móvel Aeronáutico (4) | Serviço Móvel Marítimo (5) | Serviço móvel por satélite (24) | Serviço Público de Telecomunicações (24) |
| Serviço Público Essencial (4) | Serviço radiodifusão sonora digital terrestre (17) | Serviço Telefónico em Local Fixo (645) | Serviço Telefónico Móvel (118) |
| Serviço Universal (608) | Serviços de Acesso Universal (3) | Serviços de alta velocidade (5) | Serviços de Audiotexto (60) |
| Serviços de comunicações móveis a bordo de embarcações (21) | Serviços de energia elétrica (8) | Serviços de Radiocomunicações (284) | Serviços de retalho (1) |
| Serviços de valor acrescentado (51) | Serviços eletrónicos (8) | Serviços informativos (3) | Serviços Móveis (104) |
| Serviços postais (381) | Sector (4) | SIC (78) | Sinal de televisão (7) |
| Sinalização (5) | Sistema de Contabilidade Analítica (46) | Sistemas fixos sem fios e outros (5) | Sítio da ANACOM (2) |
| Situação das Comunicações (22) | SMRP (68) | SMS (34) | SMT (5) |
| Sociedade de Informação (148) | Software (2) | Spam (12) | Spectru (1905) |
| STM (13) | Subscrição (4) | Suspensão (3) | Suspensão do Serviço (1) |
| Switch-off (86) | Tarifários (105) | Taxas (92) | Tecnologia Long Term Evolution (2) |
| Tecnologias da informação e comunicação (89) | Telecomunicações (147) | Televisão (213) | Televisão Digital (475) |
| Televisão móvel (3) | Televisão por Assinatura (2) | Televisão por subscrição (41) | Telex (3) |
| Terminação (11) | TETRA (2) | TI (1) | TIC (8) |
| Titulos (82) | TMDP (224) | Tratado de Lisboa (1) | TSAG (3) |

| | | | |
|-------------------------|------------------|--------------|----------------------|
| Twitter (4) | UIT (164) | UIT-D (11) | UIT-R (18) |
| UIT-T (14) | UMIC (3) | UMTS (116) | UNESCO (2) |
| União Europeia (247) | UPAEP (16) | UPU (37) | URSI (45) |
| Usabilidade (56) | Utilizadores (7) | UWB (9) | Videoconferência (1) |
| Voice Mail (4) | VoIP (58) | VPN (4) | WAP (1) |
| Web 3.0 (1) | WG FM PT22 (3) | WG NNA (1) | Wi-Fi (2) |
| Wik-Consult (1) | WIMAX (2) | Win-back (7) | Wireless (10) |
| Workshops (16) | WRC (13) | WSIS (4) | WTDC-10 (2) |
| WTPF-09 (2) | WTSA-08 (3) | 3G (1) | 4G (58) |
| WTSA (4) | | | |