

Universidade Federal da Paraíba  
Centro de Informática  
Programa de Pós-Graduação em Informática

Consulta Semântica Baseada em *Linked Data* para Ambientes  
de Convergência Digital (TVDi e Web)

Manoel de Albuquerque Lira Amaro

Dissertação submetida à Coordenação do Curso de Pós-Graduação  
em Informática da Universidade Federal da Paraíba como parte dos  
requisitos necessários para obtenção do grau de Mestre em Informá-  
tica.

Área de Concentração: Ciência da Computação  
Linha de Pesquisa: Computação Distribuída

Natasha Correia Queiroz Lino  
(Orientadora)

João Pessoa, Paraíba, Brasil

©Manoel de Albuquerque Lira Amaro, 22 de Dezembro de 2015

## Resumo

O surgimento da TV Digital trouxe, além da alta definição, o potencial da interatividade e metadados relativos à programação da TV. Porém, algumas informações disponibilizadas pelas emissoras podem não ser suficientes para o usuário tomar uma decisão sobre o que vai assistir. Esse trabalho tem como objetivo propor uma abordagem de Consultas Semânticas no ambiente da TV Digital Interativa baseada nos conceitos e padrões da Web Semântica, especificando uma arquitetura de Serviço Web com o objetivo de prôver um meio de enriquecimento dos metadados vindos da emissora, expandindo as informações e adicionando relacionamentos semânticos em um grafo de dados provenientes da nuvem *Linked Data*. Um Serviço Web de Consulta Semântica integrado à plataforma *Knowledge-TV* foi desenvolvido para validar essa abordagem.

**Palavras-chave:** TV Digital, GINGA, Web Semântica, Consultas Semânticas, *Web Services*, Agentes de Software, *Linked Data*

## Abstract

The emergence of Digital TV brought, beyond high definition, the potential of the interactivity and metadata relative to the TV programmes, but some information provided by the broadcasts can be insufficient to the user make a decision on what he/she is going to watch. This work aims to propose a Semantic Query approach in the Interactive Digital TV environment, based on the concepts and designs of the Semantic Web, providing an advanced service of semantic query, and a way for doing metadata enrichment coming from the broadcasters, adding semantic relationships and expanding the information in a graph of data coming from the Linked Data cloud. A Web Service of Semantic Query integrated into the Knowledge-TV platform was developed to validate this proposal.

**Keywords:** Digital TV, GINGA, Semantic Web, Semantic Queries, Web Services, Software Agents.

# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>11</b>
1.1	Motivação . . . . .	12
1.2	Objetivo Geral . . . . .	14
1.3	Objetivos Específicos . . . . .	14
1.4	Organização da Dissertação . . . . .	15
<b>2</b>	<b>Fundamentação Teórica</b>	<b>16</b>
2.1	TV Digital Brasileira . . . . .	16
2.1.1	SBTVD e <i>Middleware</i> GINGA . . . . .	17
2.1.2	Padrão de metadados no SBTVD - MPEG-2 PSI/SI . . . . .	19
2.2	Representação do Conhecimento e Raciocínio na Web . . . . .	20
2.2.1	Ontologias . . . . .	21
2.2.1.1	Linguagens para Construção e Representação de Ontologias . . . . .	23
2.2.2	Web Semântica . . . . .	27
2.2.2.1	Agentes . . . . .	30
2.2.3	<i>Linked Data</i> . . . . .	31
2.2.4	<i>DBPedia</i> . . . . .	33
2.3	<i>Web Services</i> . . . . .	34
2.3.1	<i>Software as a Service</i> . . . . .	36
2.3.2	<i>Data as a Service</i> . . . . .	36
2.3.3	<i>Knowledge as a Service</i> . . . . .	36
2.4	Conclusões . . . . .	38

---

<b>3</b>	<b>Knowledge TV e Consultas Semânticas Baseadas em <i>Linked Data</i></b>	<b>39</b>
3.1	<i>Knowledge TV</i> . . . . .	39
3.2	Consulta Semântica Baseada em <i>Linked Data</i> em Ambientes de Convergência . . . . .	43
3.3	Arquitetura . . . . .	45
3.4	Módulos SQTV . . . . .	52
3.4.1	Request Receiver . . . . .	52
3.4.2	Local Searcher . . . . .	53
3.4.3	Linked Data Querier . . . . .	53
3.4.4	Semantic Reasoner . . . . .	55
3.4.5	CoreKTV Adapter . . . . .	58
3.4.6	Instance Saver . . . . .	58
3.5	Conclusão . . . . .	59
<b>4</b>	<b>Trabalhos Relacionados</b>	<b>60</b>
4.1	<i>NoTube TV</i> . . . . .	60
4.2	<i>LinkedTube</i> . . . . .	62
4.3	<i>Google Knowledge Graph</i> . . . . .	63
4.4	<i>Swoogle</i> , <i>Falcons</i> e o <i>GoPubmed</i> . . . . .	64
4.5	Conclusão . . . . .	65
<b>5</b>	<b>Resultados e Validação</b>	<b>66</b>
5.1	Resultados obtidos . . . . .	66
5.1.1	Resultados de consulta a partir do ambiente Web . . . . .	66
5.1.1.1	Consultas simuladas utilizando o <i>JUnit</i> . . . . .	67
5.1.1.2	Integração com o projeto <i>JCollab</i> . . . . .	67
5.1.2	Resultados com integração em outras plataformas . . . . .	69
5.2	Validação . . . . .	69
5.2.1	Métricas Utilizadas na Avaliação . . . . .	70
5.2.1.1	<i>Precision</i> . . . . .	70
5.2.1.2	<i>Recall</i> . . . . .	71
5.2.2	Comparação com ferramentas similares . . . . .	71

---

5.3	Resultados da validação . . . . .	72
5.3.1	Aplicação das métricas de Recuperação de Informações . . . . .	72
5.3.2	Comparação com ferramentas similares . . . . .	74
5.4	Conclusão . . . . .	76
<b>6</b>	<b>Considerações Finais</b>	<b>78</b>
6.1	Principais Contribuições . . . . .	78
6.2	Trabalhos Futuros . . . . .	80
	Referências Bibliográficas . . . . .	89
<b>A</b>	<b>Publicações</b>	<b>90</b>
<b>B</b>	<b>Lista de títulos utilizados na validação do trabalho</b>	<b>91</b>
<b>C</b>	<b>Interfaces presentes no projeto</b>	<b>96</b>

# Lista de Símbolos

**API** : *Application Programming Interface*

**BST** : *Band Segmented Transmission*

**DaaS** : *Data as a Service*

**EPG** : *Electronic Program Guide*

**FOAF** : *Friend of a Friend*

**HTML** : *Hypertext Markup Language*

**HTTP** : *Hyper-Text Transfer Protocol*

**JSON** : *JavaScript Object Notation*

**KaaS** : *Knowledge as a Service*

**KTV** : *Knowledge-TV*

**OFDM** : *Orthogonal Frequency Division Multiplexing*

**OLAP** : *Online Analytical Processing*

**OWL** : *Web Ontology Language*

**PC** : *Personal Computer*

**RDF** : *Resource Description Framework*

**RDFS** : *Resource Description Framework -Schema*

**SaaS** : *Software as a Service*

**SBTVD-T** : Sistema Brasileiro de TV Digital Terrestre

**SOAP** : *Simple Object Access Protocol*

**SPARQL** : *SPARQL Protocol and RDF Query Language*

**STB** : *Set-Top Box*

**TVD** : TV Digital

**TVDi** : TV Digital Interativa

**W3C** : *World Wide Web Consortium*

**WWW** : *World Wide Web*

**XML** : *Extensible Markup Language*



# Lista de Figuras

2.1	Camadas do SBTVD . . . . .	18
2.2	Estrutura básica de uma tripla RDF, com exemplos de utilização. . . . .	24
2.3	Exemplo do OWL da ontologia do domínio de programas de TV. . . . .	26
2.4	Estrutura da Web 3.0 . . . . .	28
2.5	<i>Datasets</i> conectados e presentes no <i>Linked Data</i> . . . . .	33
2.6	Exemplo da ontologia DBPedia . . . . .	34
2.7	Exemplo de uma arquitetura conceitual de Serviço Web KaaS . . . . .	37
3.1	Arquitetura geral do <i>Knowledge TV</i> . . . . .	40
3.2	Arquitetura conceitual do <i>Knowledge TV</i> . . . . .	41
3.3	Arquitetura cliente-servidor do <i>Knowledge TV</i> , apresentando os módulos e o Set Top Box. . . . .	46
3.4	Exemplos do serviço de Consultas Semânticas . . . . .	48
3.5	Arquitetura dos módulos da consulta semântica . . . . .	48
3.6	Diagrama de atividade do serviço de Consultas Semânticas . . . . .	50
3.7	Arquitetura conceitual do <i>Knowledge TV</i> , apresentando os módulos do serviço de consulta semântica. . . . .	51
3.8	Exemplo da aplicação da regra de inferência do Código Fonte 3.2 . . . . .	57
4.1	Fluxo de serviços do projeto NoTube TV. . . . .	61
4.2	Exemplo de funcionamento do LinkedTube. . . . .	63
4.3	Resultado do <i>Knowledge Graph</i> para o termo pesquisado. . . . .	64
5.1	Exemplo do serviço no ambiente Web integrado ao projeto <i>Jcollab</i> . . . . .	68
5.2	Exemplo do aplicativo Android . . . . .	70

---

5.3	Resultados SQTV . . . . .	73
5.4	Resultados Google Knowledge Graph . . . . .	74

# Lista de Tabelas

5.1	Tabela comparativa entre as ferramentas existentes . . . . .	75
-----	--	----

# Capítulo 1

## Introdução

A TV Digital Interativa (TVDi) é uma nova fase vivenciada pela TV brasileira. Ela tem como objetivo convergir as tecnologias digitais com a transmissão do sinal de forma digital ao invés do analógico. Isso produzirá grandes mudanças em todo o sistema de produção e consumo de mídias, como a alta definição de som e imagem e a interatividade. Nesse sentido, o telespectador atuará como um agente ativo capaz de receber e enviar informações para as emissoras de TV.

Uma nova versão da Internet, a 3.0, vem sendo gradativamente mais utilizada e apoiada por empresas e instituições acadêmicas. Essa nova versão objetiva disponibilizar não somente páginas Web e conteúdos multimídia, mas oferecer os dados de forma que aplicações e serviços também possam utilizá-los. Uma parte da Web 3.0 é a Web Semântica, que objetiva oferecer os dados de forma estruturada e semanticamente descritos, possibilitando uma ampla utilização das informações tanto por humanos quanto por máquinas. Além disso, ela permitirá a extração de conhecimento a partir dos relacionamentos e descrições sobre os dados.

Com o advento da Web Semântica, técnicas de raciocínio inteligente começaram a aparecer com objetivo de, a partir de um modelo conceitual descrito formalmente, recuperar informações e/ou relacionamentos que não estão explicitamente definidas nesse modelo, e assim novos conhecimentos em cima de dados descritos pelo modelo citado são gerados.

Objetivando criar uma grande rede de dados conectados por meio da Web, surgiu o movimento *Linked Data*, com definições e convenções acerca da forma como os dados

devem ser formatados, representados, relacionados e distribuídos. Essas convenções permitem a usuários e agentes computacionais acessar e navegar por dados e suas conexões mais facilmente, criando assim uma nuvem de informações e conhecimento interligados.

Com o aumento de informações descritas semanticamente na Web Semântica, o conceito de Consultas Semânticas surgiu para utilizar as vantagens dos dados descritos formalmente para aprimorar os resultados das consultas, retornando informações com mais precisão e relevância, assim como outros resultados relacionados.

O projeto *Knowledge-TV*, em cujo trabalho apresentado está inserido, tem como objetivo especificar uma camada semântica focada na plataforma da TV Digital Interativa Brasileira, provendo serviços avançados, como por exemplo, de recomendação de conteúdo e de consultas semânticas. Esses serviços são disponibilizados em formato de Serviços Web, provendo uma interoperabilidade entre as plataformas.

## 1.1 Motivação

Estudos e projetos sobre Consultas Semânticas estão cada vez mais evidentes seguindo a evolução da Web Semântica, que permite que mais dados sejam disponibilizados de forma estruturada conforme um formalismo semântico. Alguns desses projetos [Cheng, Ge e Qu 2008; Ding et al. 2004; Doms e Schroeder 2005] utilizam somente alguns conceitos da Web Semântica, como os documentos descritos formalmente. Porém, eles não utilizam dados interligados entre diferentes conjuntos de dados (*datasets*), ou seja, não são baseados nos princípios do *Linked Data*. O uso desses princípios permite que mais informações sejam recuperadas a partir de várias fontes interligadas, levando dados mais enriquecidos e com mais informações relacionadas para os usuários. O projeto [Tummarello, Delbru e Oren 2007] utiliza alguns conceitos do *Linked Data* (relacionamento entre os dados). Todavia, os documentos são recuperados e indexados, para depois serem consultados. Esse processo é oneroso e depende de um elevado poder computacional para realizar toda a recuperação e indexação dos dados. Um dos princípios do *Linked Data* é que cada *dataset* disponibilize um ponto de acesso (*endpoint*) para consultas. Utilizando esses *endpoints*, o poder computacional necessário será bem

menor, pois o processo de indexação deixará de ser necessário, e as informações serão obtidas diretamente da fonte, garantindo dados atualizados.

A convergência de tecnologias digitais, no ambiente da TV Digital, permite a descrição de programas a partir de metadados no formato de tabelas, seguindo os padrões e normas do Sistema Brasileiro de TV Digital (SBTVD)[NBR 2007]. Isso permite às emissoras enviar informações sobre os programas que vão, de dados técnicos como, por exemplo, a forma de decodificação do vídeo a informações textuais sobre o programa: título, sinopse, entre outros. O sub-conjunto dos metadados que descrevem as informações textuais do programa é pequena, limitando-se a título, descrição, gênero e classificação indicativa[NBR 2007]. Isso, muitas vezes, é insuficiente para o usuário decidir assistir ou não àquele programa, fato que frustra os usuários que buscam mais informações para essa decisão. Com a ajuda da Web Semântica e dos princípios do *Linked Data*, a busca e o enriquecimento de informações torna-se fácil, rápida e precisa, devido aos dados estruturados e interligados, sempre seguindo um modelo semântico formal. Por fim, a interatividade na TV Digital permite que aplicativos sejam desenvolvidos e, a partir do uso do canal de retorno, podem-se acessar serviços externos disponibilizados na Internet, a exemplo dos Serviços Web. Com isso, um ambiente de convergência (Web e TV Digital Interativa) pode ser criado a partir da integração dos dados vindos do ambiente da TV Digital com os dados estruturados da Web Semântica. Essa integração permite um enriquecimento semântico dos metadados relacionados aos programas, o que torna possível a expansão de conhecimento e a exibição de informações relacionadas e mais ricas em nível de dados.

O serviço de Consulta Semântica está disponível no contexto do projeto *Knowledge-TV* como um Serviço Web, possibilitando ser acessado por qualquer dispositivo, desde um *Set-Top Box* até dispositivos móveis. Porém, os dados disponibilizados têm foco nos domínios presentes: filmes, musicais, shows e programas de TV. Com a criação desse ambiente de convergência (Web e TV Digital Interativa), a seguinte hipótese deve ser verificada:

*A hipótese a ser testada e comprovada é que é possível, a partir dos conceitos da Web Semântica e das Consultas Semânticas aplicadas ao ambiente da TV Digital Interativa Brasileira, construir um Serviço de Consultas Semânticas convergente com uma arquitetura baseada em Linked Data, permitindo o enriquecimento semântico das informações dos programas para os usuários da TVDi.*

## 1.2 Objetivo Geral

O objetivo deste trabalho é propor um modelo de serviço Web avançado de Consultas Semânticas para ambientes de convergência TV e Web, por meio da utilização dos conceitos da Web Semântica, com foco nos domínios presentes na TV. Isso objetiva fornecer informações semanticamente enriquecidas a partir da Consulta Semântica, utilizando os metadados da TV Digital para os usuários da plataforma TV Digital Brasileira e outras plataformas, podendo ser utilizado também para dar suporte a outros serviços avançados no contexto do projeto *Knowledge-TV*.

## 1.3 Objetivos Específicos

Para atingir o objetivo geral, foram especificados os seguintes objetivos específicos:

1. Estudos teóricos sobre Ontologias, Web Semântica, movimento *Linked Data* e *Web Services*;
2. Estudos teóricos sobre técnicas de raciocínio semântico;

3. Levantamento de requisitos para Consultas Semânticas no contexto do projeto *Knowledge-TV*;
4. Pesquisa dos melhores métodos, *frameworks* e ferramentas para a extração de dados a partir da nuvem *Linked Data*;
5. Desenvolvimento de um modelo para consultas semânticas em cima da nuvem *Linked Data*;
6. Desenvolvimento de um modelo de enriquecimento semântico de metadados a partir da Consulta Semântica;
7. Desenvolvimento de uma arquitetura para o serviço de Consultas Semânticas.
8. Validação da arquitetura desenvolvida.

## 1.4 Organização da Dissertação

Este trabalho divide-se em 5 capítulos. Após a Introdução, segue o capítulo 2, que aborda a Fundamentação Teórica com os conceitos básicos sobre as principais áreas abordadas: Representação de Conhecimento e Raciocínio na Web, explanando os conceitos de Ontologias, Web Semântica e Linked Data e os Web Services com seu papel na Web Semântica. No capítulo 3, dá-se a apresentação, iniciando com a explanação do projeto Knowledge-TV; em seguida, a explanação em si detalhadamente de como ele foi desenvolvido. O capítulo 4 apresenta os resultados obtidos e as validações feitas no trabalho; o capítulo 5 apresenta as considerações finais, bem como trabalhos futuros.



# Capítulo 2

## Fundamentação Teórica

Neste capítulo, é apresentado o contexto teórico dos principais temas que envolvem este trabalho. Inicialmente serão apresentados os principais conceitos de TV Digital Interativa, em seguida sobre Representação de Conhecimento na Web, e por fim Web Services.

### 2.1 TV Digital Brasileira

Há mais de 50 anos, os sinais de TV aberta são transmitidos no padrão analógico. Seguindo a tendência mundial de digitalização de plataformas, a população brasileira começou a receber sinais no padrão digital em 2007[Fernandes, Filho e Silveira 2004]. Esse movimento teve início nos Estados Unidos na década de 90, espalhando-se em seguida para a Europa, Japão, Brasil e China. Mesmo sendo especificado e projetado em diferentes regiões, seguindo especificidades e objetivos regionais, os sistemas de TV Digital mais relevantes como o ATSC<sup>1</sup> americano, o DVB<sup>2</sup> europeu, o ISDB<sup>3</sup> japonês, o SBTVD<sup>4</sup> brasileiro e o DTMB chinês mantêm um núcleo de convergência definido pela União Internacional de Telecomunicação (ITU), através de um modelo de referência para transmissão de sinais de TV Digital.

A mudança do padrão analógico para o padrão digital de sinais trouxe, além da

---

<sup>1</sup><http://www.atsc.org/>

<sup>2</sup><http://www.dvb.org/>

<sup>3</sup><http://www.dibeg.org/>

<sup>4</sup><http://www.dtv.org.br/>

alta qualidade de som e imagem, a interatividade entre o usuário e a TV. A partir de um canal de retorno, a TV pode enviar e receber informações de um agente externo, tornando o telespectador um agente ativo capaz de interagir com a TV [Alves et al. 2006].

Com a convergência de tecnologias digitais, interatividade na TV e os serviços digitais oferecidos, a quantidade e complexidade de informações começam a aumentar [Lugmayr, Niiranen e Kalli 2004]. Com isso, foram adotadas normas e padrões de metadados de TV, com a finalidade de disponibilizar de forma mais estruturada e tratar os conteúdos e serviços de forma eficiente e com qualidade.

No Brasil, o sistema de TV Digital implantado é o Sistema de TV Digital Terrestre (SBTVD-T), regulamentado pela Associação Brasileira de Normas Técnicas – ABNT [NBR 2010], que define, entre outras especificações, a codificação, transmissão e o *middleware* utilizado.

No escopo desse trabalho, o SBTVD-T serve de plataforma para o serviço de Consulta Semântica. A partir da utilização da interatividade, o usuário da TV poderá obter mais informações relacionadas ao programa a que assiste. Nas próximas sub-seções, serão apresentados mais detalhes sobre o SBTVD-T.

### 2.1.1 SBTVD e *Middleware* GINGA

O *middleware* é “a camada de *software* posicionada entre o código das aplicações e a infra-estrutura de execução (plataforma de *hardware* e sistema operacional)” [Soares e Castro 2008]. Essa camada serve para abstrair os detalhes do sistema para as aplicações, como tipo de processador, interface de conexão com a TV, etc., bem como ocultar toda a complexidade de *hardware* e *software*, além de suportar a execução de aplicações interativas.

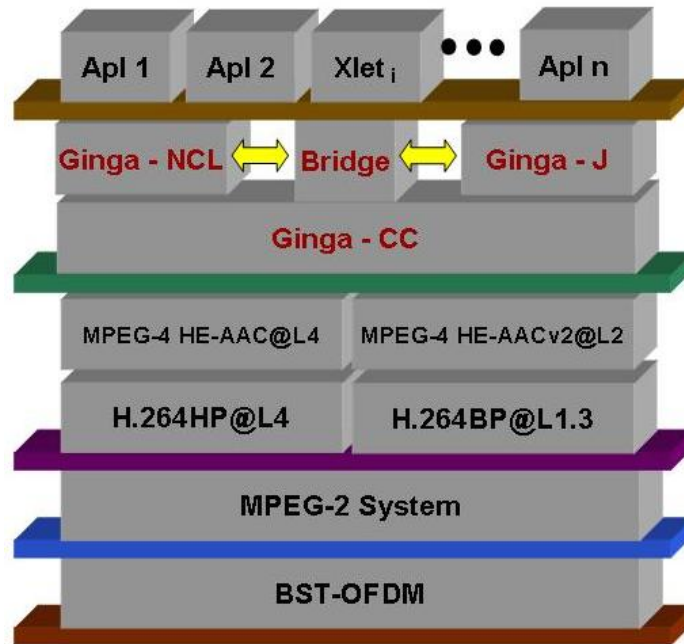
O Sistema Brasileiro de TV Digital (SBTVD) adota o *middleware* GINGA, cuja especificação de referência foi definida em conjunto entre o Laboratório de Aplicações de Vídeo Digital (LAVID)/UFPB e o Laboratório de Sistemas Multimídia TeleMídia/PUC-Rio [Soares e Filho 2007]. Assim como os demais padrões existentes, esse *middleware* é capaz de processar tanto aplicações declarativas<sup>5</sup> através do ambiente

---

<sup>5</sup>Aplicações declarativas são baseadas em uma linguagem declarativa, enfatizando uma descrição

Ginga-NCL[Soares, Rodrigues e Moreno 2007], como também aplicações imperativas<sup>6</sup> a partir do ambiente Ginga-J[Filho, Leite e Batista 2007]. Como pode ser observado na Figura 2.1, o SBTVD segue a tendência de outros padrões de TV Digital.

Figura 2.1: Camadas do SBTVD



Fonte: [Soares e Castro 2008]

De acordo com [Brackmann 2008], os módulos do SBTVD apresentados na Figura 2.1 podem ser detalhados como:

**BST-OFDM** : A modulação OFDM (*Orthogonal Frequency Division Multiplexing*) divide o canal em diversas sub-portadoras e as transmite paralelamente, cada uma carregando uma pequena parte da informação. Essa partição da informação permite que mesmo com interferências, poucas partes dos dados originais sejam perdidas, fazendo a tecnologia OFDM mais imune à interferências do ambiente. O BST (*Band Segmented Transmission*) é responsável pela divisão do seguimento de banda, ou seja, particionar o espectro do sinal (6 MHz) em vídeo, áudio e dados.

---

declarativa do problema.

<sup>6</sup>Aplicações declarativas são baseadas em uma linguagem procedural, decompondo o problema em uma implementação algorítmica

**MPEG-2 System** : O MPEG-2 (*Moving Pictures Expert Group*)[Haskell 1997] é o padrão adotado pelo SBTVD para a camada de transporte, responsável por conduzir o áudio e vídeo transmitidos.

**H.264** : O Padrão H.264[Wiegand et al. 2003] é responsável pela codificação/decodificação do vídeo. O SBTVD utiliza dois perfis do padrão H.264, o HP@L4 para vídeos de definição padrão à alta definição e o BP@L1.3 para vídeos em baixa definição. A escolha do H.264 é um dos diferenciais do SBTVD para o padrão de TV Digital Japonês (ISDB), que utiliza o padrão MPEG-2.

**MPEG-4** : Para a codificação/decodificação do áudio, o SBTVD utiliza o padrão MPEG-4[Koenen 2002], e assim como o H.264, o MPEG-4 apresenta dois perfis, o HE-AAC@L4 para áudio em alta definição e o HE-AACv2@L2 para áudio em baixa definição.

**Ginga** : A camada onde se encontra o *middleware* do SBTVD, o Ginga. Como explicado, os detalhes de *hardware* e outras camadas são abstraídas, de forma que as aplicações não precisam ter conhecimento de todos os detalhes técnicos para serem executadas.

**Aplicações** : Onde as aplicações enviadas pelo canal de dados são executadas.

Com isso, a partir da abstração do *middleware* Ginga, as aplicações disponibilizadas na camada de Aplicação poderão acessar os serviços disponibilizados por esse trabalho. Assim, com a utilização do canal de retorno, as aplicações podem-se conectar aos Serviços Web e obter informações. A próxima seção descreve como as informações sobre os programas são enviadas pelas emissoras e tratadas pelo SBTVD.

### 2.1.2 Padrão de metadados no SBTVD - MPEG-2 PSI/SI

Metadados podem ser considerados informações que descrevem, por exemplo, um documento, como data de criação, autor, entre outros, ou seja, são dados que descrevem outros dados. A International Federation of Library Associations (IFLA) define metadados como:

“Metadados são dados sobre dados. O termo se refere a qualquer informação utilizada para a identificação, descrição e localização de recursos.”

Na TV Digital, metadados são as informações que descrevem os dados que são enviados pelas emissoras e recebidos pelos *Set-Top Boxes* (STBs). Os principais sistemas de TV Digital (TVD), incluindo o SBTVD, adotam padrões que utilizam Tabelas de Informação de Serviço (SI)[NBR 2007] como metadados. O padrão adotado pelo SBTVD é o MPEG-2, sendo esse um padrão amplamente adotado por sistemas de TVD. Esse padrão utiliza descritores de baixo nível (tabelas PSI/SI), porém esses descritores, mesmo que sejam de fácil manipulação, são rígidos, ou seja, não podem ser estendidos, limitando a utilização e oferta de serviços mais avançados.

O conjunto de tabelas chamado PSI (*Program Specific Information*) é responsável pela descrição de informações essenciais para que o receptor possa tratar o fluxo de transporte e decodifique seus fluxos elementares. No entanto, outros dados são necessários para que o receptor identifique os serviços existentes no fluxo de transporte, sendo essas informações presentes nas tabelas SI (*Service Information*). Além dessas informações, as tabelas SI transportam também informações sobre o conteúdo de cada serviço, como título, descrição, gênero, classificação, etc.

A partir dos dados sobre o conteúdo dos serviços disponibilizados pelas tabelas SI, o trabalho apresentado nessa dissertação fornecerá informações enriquecidas, a partir de dados recuperados através da Internet. Esses dados disponibilizados na Internet também são estruturados, de forma a representar conhecimento e permitir raciocínio automático. A próxima seção detalha essa representação de conhecimento na Web.

## 2.2 Representação do Conhecimento e Raciocínio na Web

Nas últimas décadas, várias linguagens para representação de conhecimento foram propostas. Essas linguagens, mesmo não criadas para tal propósito, vêm sendo utilizadas na Web, para uma melhor organização e estruturação dos dados. [Breitman 2005] O *boom* da Internet fez com que surgissem linguagens que davam suporte e explora-

vam características da rede. Essas linguagens são conhecidas como “linguagens leves” (*lightweight*), ou “linguagens de *mark up*”, e ficaram conhecidas como linguagens de codificação genérica, diferenciando-se das linguagens de codificação específicas. Nesse sentido, a primeira utiliza marcações (*tags*) que fornecem uma descrição de como o *software* deve formatar o documento na tela. Um exemplo de linguagem genérica é o HTML (*Hypertext Markup Language*) [Raggett et al. 1999]. No entanto, a HTML não apresenta uma descrição formal para estruturação dos dados, impossibilitando a validação da informação exibida. Ou seja, essa linguagem é voltada para o consumo de informações por humanos, focada na exibição da informação em detrimento do processamento por máquinas e agentes computacionais. De modo a suprir essas limitações, o XML (*Extensible Markup Language*) [Bray et al. 1998] começou a ser adotado, oferecendo suporte a um grande número de aplicações Web e também ao HTML. O padrão XML separa o conteúdo da estrutura do documento, permitindo modificar a forma em que a informação é apresentada sem que seja necessário modificar o conteúdo dos documentos. Ou seja, as descrições e estrutura da informação não são alteradas, facilitando a interpretação dos dados por um agente computacional, por exemplo. Esse padrão também é extensível, sendo utilizado como uma forma de representação a partir de outros padrões que serão apresentados nas seções seguintes.

Em um ambiente como a Web, onde existe uma grande quantidade de informações, métodos de classificação e organização são necessários, assim como modelos para representar os dados dentro de um conjunto de conceitos. Um desses modelos é a Ontologia, detalhada na próxima seção.

### 2.2.1 Ontologias

As ontologias surgem como um modelo para organizar e gerenciar o conhecimento, seja ele genérico ou de domínios específicos dentro de aplicações, provendo, dentre outras vantagens, o acesso eficiente a grandes volumes de dados.

A definição de ontologia mais conhecida na literatura da Web Semântica é a proposta por Gruber:

“Ontologia é uma especificação formal e explícita de uma conceitualização compartilhada.”[Gruber et al. 1993]

Para exemplificar e descrever o conceito de Gruber, obtendo um significado mais amplo dos termos da sua definição, os autores [Ding e Foo 2002] explicam o conceito de Gruber como:

- Conceituação - “Modelo abstrato de um fenômeno no mundo”;
- Explícita - “Os tipos de conceitos usados e suas restrições devem estar explicitamente definido”;
- Formal - “a ontologia deve ser processada por máquina”;
- Compartilhada –“a ontologia deve capturar o conhecimento aceito por consenso pelas comunidades que dela faz uso”.

Outro conceito apresentado por [Breitman 2005] define uma ontologia sendo a expressão  $O := \{C, R, H^c, rel, A^o\}$ , cuja uma ontologia  $O$  é um conjunto de conceitos  $C$ , relações  $R$ , hierarquias  $H$  sobre os conceitos, uma função de relacionamento  $rel$  e axiomas<sup>7</sup>  $A$ .

Para representar os conceitos, geralmente são utilizados classes, atributos (propriedades), relacionamentos e restrições. As ontologias podem ser comparadas ao modelo de entidade-relacionamento [Chen 1976] presentes nos sistemas de banco de dados convencionais, porém possuindo mais expressividade. As classes representam um conceito do mundo real, como por exemplo, o conceito de Pessoa; os atributos podem ser considerados características de um conceito, como por exemplo, o nome e idade da Pessoa; os relacionamentos conectam os conceitos de forma semântica, como por exemplo, o relacionamento hierárquico pai-filho entre duas Pessoas.

---

<sup>7</sup>Na lógica, um axioma é uma sentença que não é provada ou demonstrada, porém considerada óbvia ou um consenso inicial para a construção ou aceitação de uma teoria.

As ontologias vêm sendo utilizadas em diversas áreas, com vários estudos dirigidos para a Web, onde a recuperação de informações ainda é imprecisa. A utilização de ontologias nessa área proporciona uma melhor organização e estruturação dos dados, permitindo a busca mais precisa.

Elas geralmente são especificadas com linguagens que permitem abstração das estruturas de dados e da implementação, ou seja, linguagens são mais expressivas. Devido a essa abstração, as ontologias podem ser utilizadas, por exemplo, para integração entre diferentes sistemas, possibilitando a interoperabilidade entre eles.

### 2.2.1.1 Linguagens para Construção e Representação de Ontologias

Para a construção e representação de ontologias, várias linguagens estão disponíveis como, por exemplo, OIL, DAML+OIL, RDF e OWL.[Gómez-pérez e Corcho 2002] Para padronizar a representação de ontologias, a *World Wide Web Consortium (W3C)*<sup>8</sup> recomenda a utilização de duas dessas linguagens citadas: (i)RDF e (ii)OWL.

#### i. Resource Description Framework (RDF)

O *Resource Description Framework (RDF)* [Klyne e Carroll], idealizada pela W3C, é um modelo simples de dados para processamento e manipulação por aplicações e agentes de *software*, fornecendo um meio universal para expressar informações e relacionamentos sobre recursos, além de permitir um intercâmbio de dados entre aplicações sem perda do significado.

A linguagem apresenta um conceito básico de triplas para descrever os recursos, sendo as triplas formadas por: (i) Sujeito, sendo o objeto ou recurso em si que deseja atribuir uma característica ou informação, (ii) Predicado, sendo a propriedade (característica) que deseja descrever, e (iii) Objeto, o valor referente à propriedade. A Figura 2.2 mostra a estrutura da tripla RDF, com um exemplo de utilização.

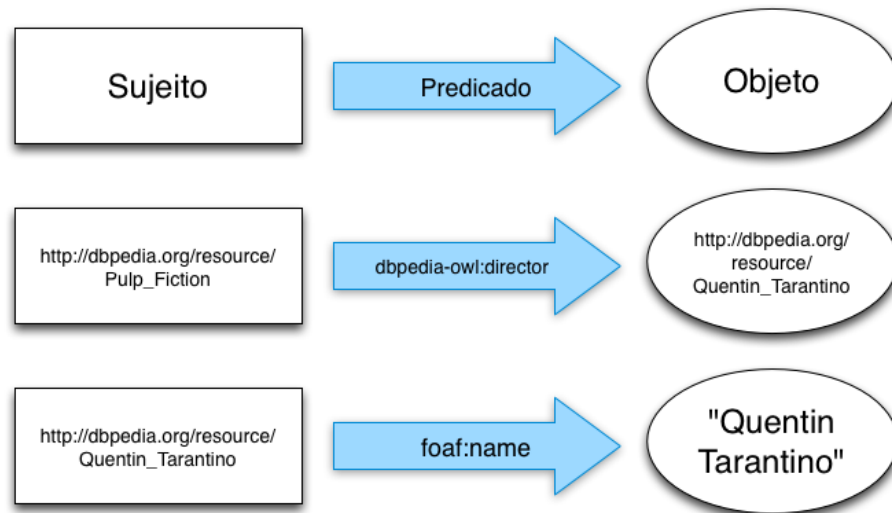
No exemplo apresentado na Figura 2.2, o recurso “Pulp Fiction”, identificado pela URI [http://dbpedia.org/resource/Pulp\\_Fiction](http://dbpedia.org/resource/Pulp_Fiction) é descrito com o pre-

---

<sup>8</sup>Consórcio de empresas, profissionais, cientistas e instituições acadêmicas responsável pela criação de padrões tecnológicos que regulam a *World Wide Web (WWW)*



Figura 2.2: Estrutura básica de uma tripla RDF, com exemplos de utilização.



dicado “dbpedia-owl:director” que refere-se à propriedade “diretor” do filme, relacionando a um outro recurso referente à “Quentin Tarantino”, identificado pela URI [http://dbpedia.org/resource/Quentin\\_Tarantino](http://dbpedia.org/resource/Quentin_Tarantino). O recurso referente ao diretor é descrito com o predicado “foaf:name”, conectando ao nome “Quentin Tarantino”. Os predicados são sempre identificados por um identificador único na Web (URI), como o “foaf:name” e o “dbpedia-owl:director”. Ou seja, existem conjuntos (vocabulários) de predicados pré-definidos para cada conceito ou domínio, como por exemplo o conjunto de descritores FOAF [Brickley e Miller 2010] para o domínio de pessoas (nome, sobrenome, relacionamentos, etc.).

Como o RDF é um modelo de dados, é necessária a utilização de outra linguagem para representar as informações no formato RDF. A principal linguagem para representação é a XML, introduzida anteriormente. O Código Fonte 2.1 apresentado a seguir representa o exemplo da Figura 2.2 escrito em XML.

#### Código Fonte 2.1: Exemplo de RDF em XML

```

1 <? xml version="1.0" ?>
2 <RDF xmlns = "http://w3.org/TR/1999/PR-rdf-syntax-19990105#"
3   xmlns:dc = "http://purl.org/DC#"
4   xmlns:foaf = "http://xmlns.com/foaf/0.1/"
5   xmlns:dbpedia-owl = "http://dbpedia.org/ontology/" >
6
7 <Description about = "http://dbpedia.org/resource/Pulp_Fiction" >

```

```

8     <foaf:name xml:lang="en">Pulp Fiction</foaf:name>
9     <dbpedia-owl:director rdf:resource="http://dbpedia.org/resource/
      Quentin_Tarantino"/>
10    </Description>
11
12    <Description rdf:about="http://dbpedia.org/resource/Quentin_Tarantino">
13      <type rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
14      <foaf:name xml:lang="en">Quentin Tarantino</foaf:name>
15      <foaf:givenName xml:lang="en">Quentin</foaf:givenName>
16      <foaf:surname xml:lang="en">Tarantino</foaf:surname>
17    </Description>
18  </RDF>

```

---

Nesse exemplo, os recursos são descritos a partir das *tags* XML pré-definidas para RDF, como **<Description>**. Essa *tag* inicia a descrição de um recurso, sendo o atributo **about** o identificador único (URI) para o recurso descrito. Dentro da *tag* **<Description>** são colocados outros valores referentes ao recurso, como por exemplo o nome do filme utilizando a *tag* **<foaf:name>**, e o diretor utilizando a *tag* **<dbpedia-owl:director>**, referenciando o recurso do diretor pela URI.

O RDF pode ser estendido, de forma a criar novos vocabulários para atender diversos conceitos e domínios. Um exemplo é o RDF Schema [Brickley e Guha 2004], onde são definidos conceitos abstratos como Recurso (**rdfs:Resource**), Classe (**rdfs:Class**) e relacionamentos, como Sub Classe (**rdfs:subClassOf**). A partir do RDF e RDF Schema, foi criada a *Ontology Web Language* (OWL), descrita na próxima seção.

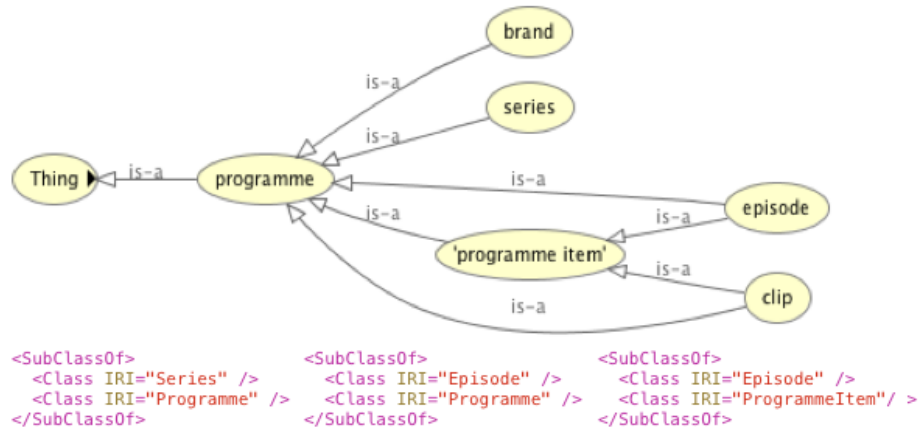
## ii. Ontology Web Language (OWL)

A *Ontology Web Language* (OWL) [McGuinness, Harmelen et al. 2004] é uma linguagem baseada em RDF e RDF Schema, que foi especificada pelo consórcio W3C como padrão para especificação de ontologias na Web. Nesse sentido, ela é uma revisão da linguagem para representação de ontologias existente, a DAML+OIL. [Breitman 2005].

O OWL pode descrever uma ontologia definindo classes, propriedades, indivíduos e restrições, assim como os valores e relacionamentos semânticos existentes entre as entidades. Os principais relacionamentos suportados pelo OWL são (i) genera-

lização (**superClassOf**), (ii) Especialização (**subClassOf**) e (iii) Equivalência (**equivalentClass**). A Figura 2.3 apresenta um exemplo da parte de uma ontologia do domínio relacionado à programação de TV, definindo que a classe Série (*serie*) é sub-classe de Programa (*programme*).

Figura 2.3: Exemplo do OWL da ontologia do domínio de programas de TV.



O OWL também define Indivíduos como objetos (instâncias) dos conceitos definidos pela ontologia, pertencendo a classes e podem ser relacionados a outros indivíduos, por meio de propriedades definidas pela classe a que pertence. A linguagem também define restrições, como por exemplo, a propriedade “gênero” só pode conter classes que sejam sub-classes de “genre” definida na ontologia de programas de TV.

A linguagem OWL possui 3 sublinguagens, listadas a seguir de acordo com sua capacidade de representatividade e computabilidade, com um detalhamento segundo [Breitman 2005]:

1. *OWL-Lite*: Suporta a criação de hierarquias simplificadas de classificação e suas restrições, não apresentando axiomas nem estruturas de relacionamentos sofisticadas.
2. *OWL-DL*: A *OWL-DL* (*Description Logic*) pode ser mapeada para linguagens de descrição lógica. É um meio termo entre o *Lite* e o *Full*, pois permite uma maior expressividade e garante a computabilidade.

3. *OWL-Full*: Segundo [McGuinness, Harmelen et al. 2004], “Suporta o máximo de expressividade enquanto mantém completude computacional (para todas as computações se garante tempo finito)”. Provê uma maior expressividade na representação de conhecimento.

Todas essas linguagens apresentadas são utilizadas para expressar conhecimento no ambiente da Web Semântica, sendo o RDF e o OWL as mais utilizadas para descrever os recursos e documentos nesse ambiente. A próxima seção traz mais detalhes sobre a Web Semântica e como essas linguagens são utilizadas para descrever os recursos.

### 2.2.2 Web Semântica

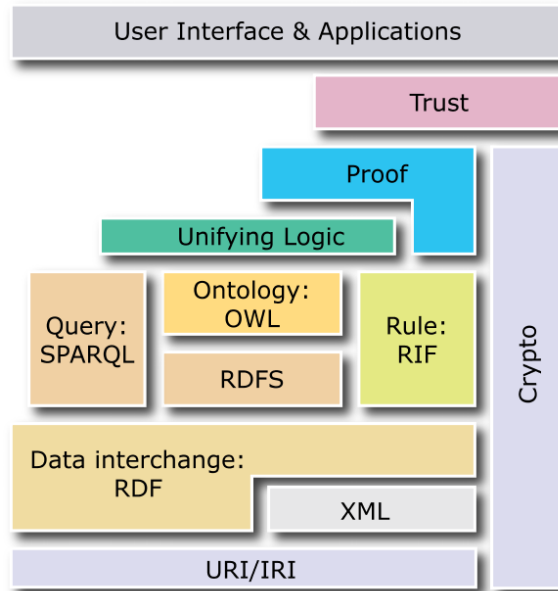
A Web atual, também chamada de Web 2.0 [O’Reilly e Battelle 2009], tem como principal foco a interatividade, com foco também no compartilhamento de dados entre *sites* e aplicativos, assim como o uso das redes sociais. Porém os dados que circulam na Web 2.0 são dados brutos, com pouca ou nenhuma estrutura que os descreve formalmente, não apresentando, muitas vezes, formas de disponibilizar as informações para consumo por outras aplicações, por exemplo.

“Não há como delimitar fronteiras para a web 2.0, pois trata-se de princípios e práticas para que diversos sites sigam. Um dos princípios fundamentais é a web como plataforma, ou seja, o usuário poder realizar atividades online que antes só eram possíveis com programas rodando em seu computador.” [O’reilly 2007]

A Web 3.0 [Hendler 2009] é um aprimoramento da Web 2.0, que objetiva oferecer as informações não somente aos usuários, mas também para outros aplicativos e serviços. Também possibilita a existência de serviços projetados somente para disponibilizar dados para outras aplicações, sendo essas as aplicações finais voltadas para o usuário. Esse aprimoramento representa um conteúdo fácil de ser processado, pois utiliza formatos padronizados e independentes da linguagem de programação para a troca das informações. A Web Semântica [Breitman 2005] faz parte da Web 3.0 com o objetivo principal de disponibilizar dados mais estruturados e com relacionamentos semânticos

entre si, facilitando assim o reuso das informações e também a descoberta de conhecimento a partir das conexões existentes entre as informações. A Figura 2.4 apresenta a estrutura da Web 3.0 segundo [Hendler 2009].

Figura 2.4: Estrutura da Web 3.0



Fonte: [Hendler 2009]

Conforme [Hendler 2009], as camadas apresentadas na Figura 2.4 são detalhadas como:

- **URI/IRI:** *Uniform Resource Identifier* (URI)[Masinter, Berners-Lee e Fielding 2005] e *Internationalized resource identifier* (IRI)[Dürst e Suignard 2005] são responsáveis por definir como um recurso na Web é identificado, utilizando um identificador único na Web como por exemplo `http://dbpedia.org/resource/Pulp_Fiction`.
- **Data interchange: RDF / XML:** O formato básico dos dados (RDF) e a linguagem para representar as informações em RDF (XML).
- **Query: SPARQL, Ontology: OWL, Rule: RIF, RDFS:** Define as linguagens/tecnologias recomendadas, como a SPARQL[Quilitz e Leser 2008] sendo a

linguagem para consultas em *datasets* e agentes semânticos, RDFS para descrição abstrata dos recursos, OWL como linguagem para representar ontologias e RIF[Kifer 2008] como *framework* para criar regras de inferência.

- **Unifying Logic:** Utilização das ontologias para aplicações baseadas em conhecimento.
- **Proof:** Prover meios de representação e validação das lógicas.
- **Trust:** Utiliza agentes de confiança ou certificados para prover conhecimento.

O conceito de Web Semântica surgiu em 2001, quando [Berners-Lee et al. 2001] junto ao W3C propôs uma nova forma de manipular o conteúdo na Web por meio de uma estruturação semântica dos dados, que conseqüentemente permite aos agentes inteligentes o processamento automático das informações. Além disso, essa estruturação permite uma interação entre eles com a troca de informações entre si por meio da infraestrutura de dados da Internet. Com isso, as informações são organizadas e disponibilizadas de maneira mais sofisticada, permitindo a descoberta e expansão de conhecimento em cima dos relacionamentos sobre os dados. O projeto da Web Semântica, em sua essência, é a criação e implantação de padrões tecnológicos para permitir esse cenário, que não somente facilitem as trocas de informações entre agentes pessoais, mas também estabeleçam uma linguagem padronizada para o compartilhamento de dados entre sistemas e agentes de modo geral.

“A Web Semântica não é uma Web separada, mas uma extensão da atual. Nela a informação é dada com um significado bem definido, permitindo melhor interação entre os computadores e as pessoas.” [Berners-Lee et al. 2001]

Para atingir tal propósito, é necessário criar uma padronização de tecnologias, linguagens e de vocabulários descritivos de forma que todos os usuários da Web sigam convenções como armazenar e disponibilizar os dados de forma que possam ser “consumidos” tanto por usuários humanos ou agentes computacionais, de maneira automática

e não ambígua<sup>9</sup>. Com a existência da infraestrutura da Internet, o foco do projeto da Web Semântica é convencionar o uso das tecnologias e linguagens apresentadas na Figura 2.4 e descritas na Seção 2.2.

O W3C possui grupos de interesses envolvidos na implementação de uma reorganização da Web pela perspectiva semântica. A Web Semântica é a principal tecnologia desenvolvida pelo W3C para esse objetivo. Essa tecnologia permite que máquinas compreendam o significado da informação na Web, possibilitando assim compartilhamento e reuso de dados por computadores e conexões semânticas entre os dados e objetos, algumas vezes modelado a partir de objetos do mundo real.

### 2.2.2.1 Agentes

A Web Semântica permite a criação de programas eficientes que coletem conteúdo da Web de diversas fontes, processem as informações e compartilhem os dados processados com outros programas, a partir das padronizações das tecnologias e linguagens para representar conhecimento e descrever os recursos. Esses programas são chamados de agentes. Embora não exista um conceito universal de “agente” na computação, pode-se considerar o conceito mais disseminado como sendo “um assistente de tarefas, ou seja, entidades de *software* que empregam técnicas de inteligência artificial com o objetivo de auxiliar o usuário na realização de uma determinada tarefa, agindo de forma autônoma”[Souza e Alvarenga 2004].

De acordo com [Wooldridge e Jennings 1995], um agente apresenta um conjunto de propriedades desejáveis, descritas à seguir:

- **autonomia**, onde o agente executa sem nenhum tipo de interação, possuindo controle sobre suas ações;
- **sociabilidade**, onde o agente interage com outros agentes, sejam artificiais ou humanos, por meio de uma linguagem padrão;
- **reatividade**, percebendo alterações no ambiente e reagindo a tempo;

---

<sup>9</sup>Não ambíguo significa que o resultado do processamento não apresente ambiguidades, ou seja, múltiplos sentidos ou interpretações.

- **proatividade**, apto a tomar iniciativas, em vez de somente responder ao ambiente;
- **continuidade temporal**, ou seja, executando continuamente, sem interrupção;
- **orientação a objetivos**, ser capaz de interagir e desempenhar ações isoladas para atingir um objetivo maior e complexo.

O projeto apresentado nessa dissertação segue a idéia de um agente computacional utilizando os padrões da Web Semântica, onde apresenta módulos para a captura de dados nos *datasets* semânticos, processa os dados e disponibiliza para o usuário informações mais ricas. A próxima seção detalha a fonte de dados utilizada pelo projeto, o *Linked Data*.

### 2.2.3 *Linked Data*

O *Linked Data*[Berners-Lee 2009] é um movimento liderado por Tim Berners-Lee para conectar os dados na Web de forma semântica e disponibilizá-lo de forma aberta para que qualquer indivíduo (humano ou agente computacional) possa utilizá-los. Para isso, foram criadas regras para padronizar a forma que os provedores de informações devem disponibilizar seus dados e como devem ser conectados a outros dados de outros provedores.

Existem quatro regras básicas para uma entidade colocar seus dados em um formato padronizado do *Linked Data*. Essas regras são:

1. Usar *Uniform Resource Identifiers* (URI)s para identificar os recursos na Internet. Por exemplo, um fabricante de carros pode utilizar para identificar os modelos dos seus carros URIs como *http://www.fabricante.com/carro/modeloA*. Com isso, cria-se um identificador único em toda a Internet, pois não pode existir dois *URIs* iguais.
2. Utilizar sempre o protocolo *Hypertext Transfer Protocol*(HTTP)<sup>10</sup> e forne-

---

<sup>10</sup>O *Hypertext Transfer Protocol* (HTTP)[Fielding et al. 1999] é um protocolo de aplicação base para a comunicação na *World Wide Web* (WWW). É o protocolo para a troca de *Hypertext* (textos estruturados com *links* entre eles).



cer dados caso o usuário acesse na URI de um recurso. Utilizando o mesmo exemplo do fabricante de carros, quando um usuário entrar na URI *http://www.fabricante.com/carro/modeloA*, serão retornadas informações sobre o carro *modeloA*.

3. Quando alguém acessar a URI, prover as informações em um padrão de representação de dados, como por exemplo o RDF/XML.
4. Incluir nessas informações *links* para outros dados, não só da mesma base, mas se possível para outras bases de dados também, fazendo assim um reuso de informações. Voltando ao exemplo, caso o carro *modeloA* utilize um motor fabricado por outra empresa, não colocar explicitamente as descrições completas do motor, e sim um *link* (URI) para o recurso referente ao motor disponibilizado por outro *dataset* em outro endereço.

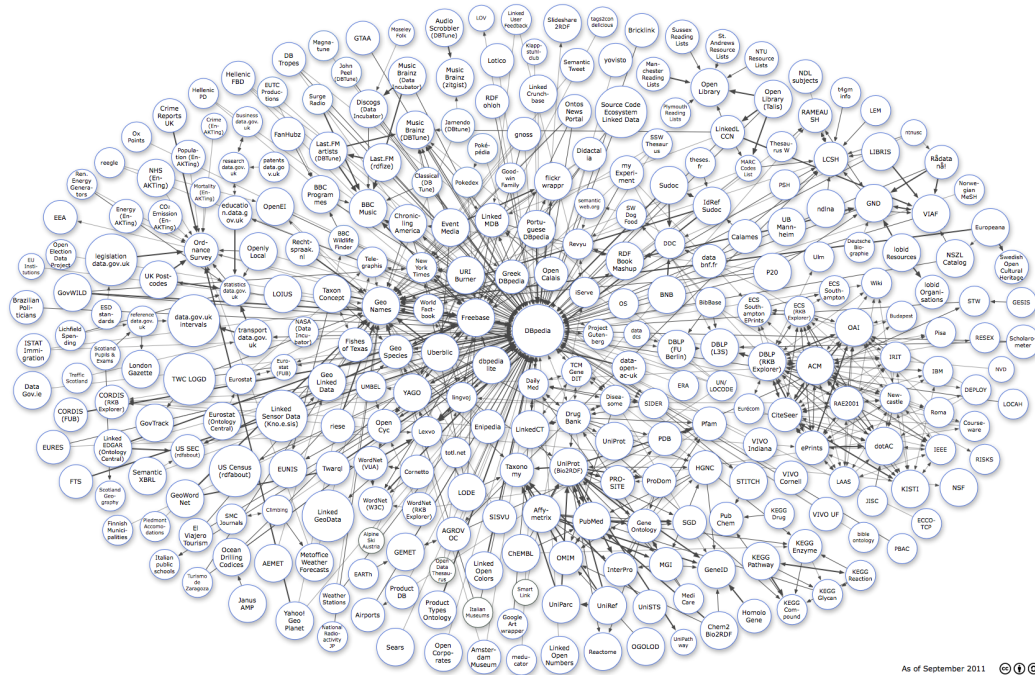
Uma regra importante para facilitar o acesso aos dados de um *dataset* é oferecer um ponto para consultas, utilizando a linguagem de consultas SPARQL[Prud'hommeaux e Seaborne][Hartig 2012]. Esses pontos são chamados de *endpoints* SPARQL. O *dataset* DBPedia, que será apresentado na próxima Seção, apresenta um *endpoint*<sup>11</sup>, onde podem ser feitas consultas, obtendo informações estruturadas semanticamente.

Com essas regras sendo seguidas, automaticamente é criada uma nuvem de *datasets* oferecendo dados conectados com outros dados de instituições distintas, fazendo com que o conhecimento em cima de todas essas informações se torne mais fácil de obter, assim como informações mais ricas. Com a utilização dos *links* entre os conjuntos de dados, mais informações relevantes e corretas podem ser obtidas, gerando assim um ambiente com informações estruturadas, fácil de serem lidas por uma máquina, e com informações relacionadas precisas. A Figura 2.5 mostra o estado da nuvem *Linked Data* no final de 2011, apresentando todos os conjuntos de dados presentes na nuvem, e os *links* presentes entre eles.

Nota-se na Figura 2.5 que existe um *dataset* maior e central, onde contém praticamente *links* com os outros *datasets*. Esse *dataset* central é o **DBPedia**, detalhado na Seção a seguir.

---

<sup>11</sup><http://dbpedia.org/sparql>

Figura 2.5: *Datasets* conectados e presentes no *Linked Data*

Fonte: Linked Data[Heath e Bizer 2011]

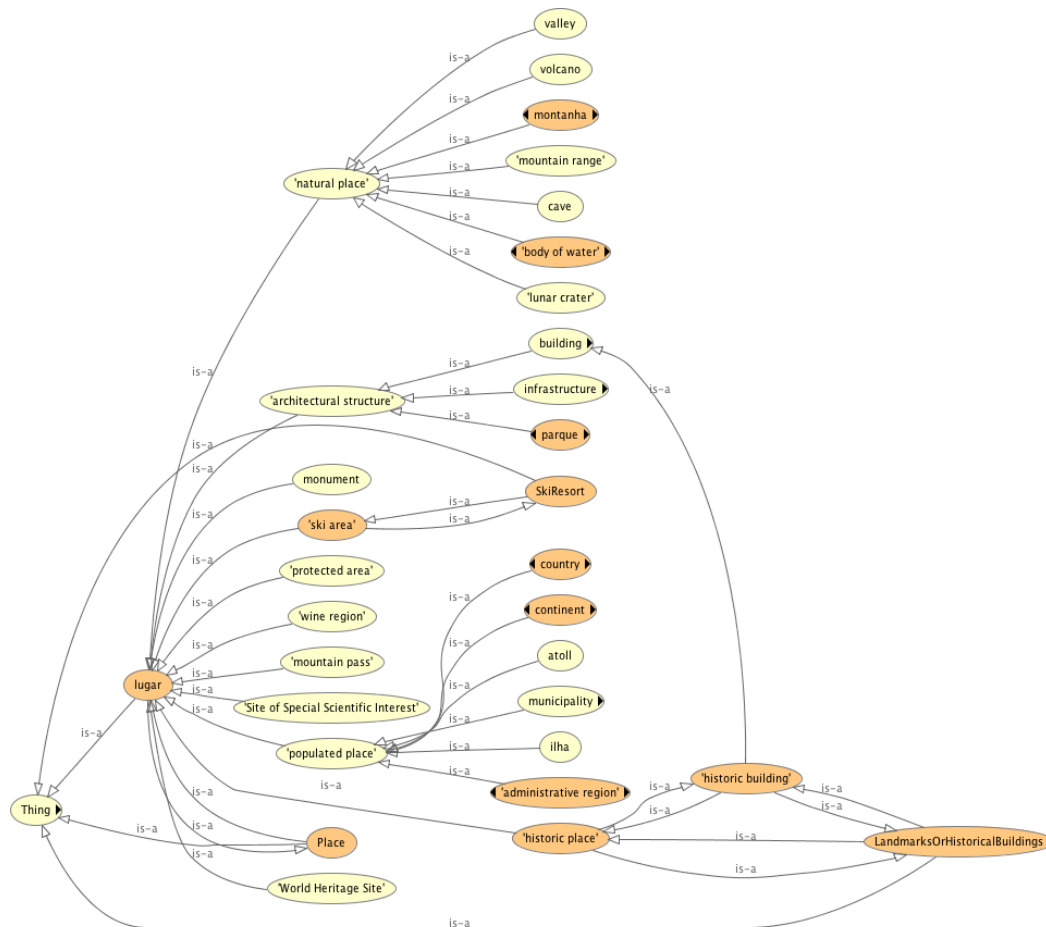
### 2.2.4 DBpedia

O DBpedia[Bizer et al. 2009] é um projeto comunitário com objetivo de extrair e transformar todos os dados da Wikipedia<sup>12</sup> para o modelo da Web Semântica, ou seja, com uma estrutura e descrições bem definidas. Com essa estruturação, os dados podem ser pesquisados e conectados com outros *datasets*, adicionando mais informações às já existentes da Wikipedia.

Para descrever todo o conteúdo da Wikipédia, uma ontologia foi criada, abrangendo todos os domínios presentes na Wikipédia. A Figura 2.6 apresenta uma parte da ontologia da DBpedia. Essa parte descreve o domínio de Lugar, como lugares naturais (cavernas, montanhas, etc.), monumentos, lugares históricos, etc.

<sup>12</sup><http://en.wikipedia.org/wiki/Wikipedia:About> A Wikipedia é uma enciclopédia livre aberta para edição, onde usuários são livres para contribuir com informações.

Figura 2.6: Exemplo da ontologia DBPedia



A DBPedia é um dos principais *datasets* do *Linked Data*. Devido a isso, o projeto apresentado nessa dissertação utiliza esse *dataset* como sendo a principal fonte de informações semânticas para enriquecer os metadados da TV Digital.

O trabalho dessa dissertação foi desenvolvido como um Serviço Web, disponibilizando o serviço de Consultas Semânticas para usuários (humanos ou agentes computacionais) por meio da infraestrutura de troca de dados da Internet. Os conceitos de Serviços Web (*Web Services*) são apresentados na seção a seguir.

## 2.3 Web Services

Um dos principais usos da Internet, além de interconectar computadores, é compartilhar dados como: informações meteorológicas, educacionais, governamentais, entre outros, além de textos e imagens sobre diversos assuntos. Todos esses dados são fa-

cilmente descobertos a partir de uma ferramenta de busca. Com a evolução, serviços começaram a ser oferecidos na Internet, como transações bancárias e compra de produtos, que são facilmente encontrados atualmente. A partir disso, instituições vêm tentando estabelecer padrões para a produção e utilização desses serviços, conhecidos como *Web Services*[Breitman 2005]. A seguir é apresentado algumas definições por diferentes autores:

*“Um Web Service é um aplicativo de software que pode ser acessado remotamente através de diferentes linguagens baseadas em XML. De modo geral, um Web Service é definido através de uma URL, da mesma forma que qualquer site na Internet. O que distingue um Web Service é o tipo de interação fornecida.”*

Ref.: [Kopack e Potts 2003]

*“Web Services são um novo tipo de aplicação para Internet. Eles são autocontidos, autodescritivos, modulares e podem ser publicados, localizados e chamados através da rede. Os Web Services realizam funções que vão das mais simples até processos de negócio complexos. Uma vez tornado público, outras aplicações (ou Web Services) podem “descobrir” e fazer uso do mesmo.”*

Ref.: [Leymann, Roller e Schmidt 2002]

*“Uma forma de expor a funcionalidade de um sistema de informação e disponibilizar através de padrões e tecnologias Web.”*

Ref.: [Alonso et al. 2004]

O ponto mais importante acerca da tecnologia que gera tanto entusiasmo é a questão da interoperabilidade dos ambientes computacionais. Como a arquitetura dos Web Services é baseada na troca de mensagens utilizando uma linguagem descritiva, como XML[Bray et al. 1998] ou JSON[Crockford 2006], faz com que essa tecnologia seja:

- Independente de plataforma;
- Independente de localidade, pois sendo disponibilizado na Internet, qualquer um pode acessar;

- Independente de linguagem no lado do cliente que utiliza o Web Service;
- Abstração das características técnicas de *hardware* onde o Web Service é executado.

Resumindo, por meio dos *Web Services*, as aplicações têm a potencialidade de se comunicar entre elas pela Internet, trocando mensagens conforme os protocolos de comunicação e os padrões estabelecidos. A seguir são apresentados modelos de *Web Services* utilizados atualmente.

### 2.3.1 *Software as a Service*

O *Software as a Service* (SaaS)[Gold et al. 2004] é um modelo consolidado na Internet, com objetivo de prover aplicações. Nesse modelo, as aplicações são disponibilizadas como um serviço Web, hospedado em um servidor remoto, e disponível para todos os usuários da Internet. A maior vantagem desse modelo é ser independente da plataforma que o usuário utiliza, removendo a necessidade do usuário ter sua própria infraestrutura de *hardware* para executar o *software*.

### 2.3.2 *Data as a Service*

O modelo *Data as a Service* (DaaS)[Truong e Dustdar 2009] provê uma nova forma de serviços Web, deixando de focar o processamento das informações em si, para disponibilizar dados e métodos para modificação desses, como: atualizar, remover e adicionar (CRUD - *Create, Read, Update, Delete*). Essa é a principal diferença entre o DaaS e o SaaS.

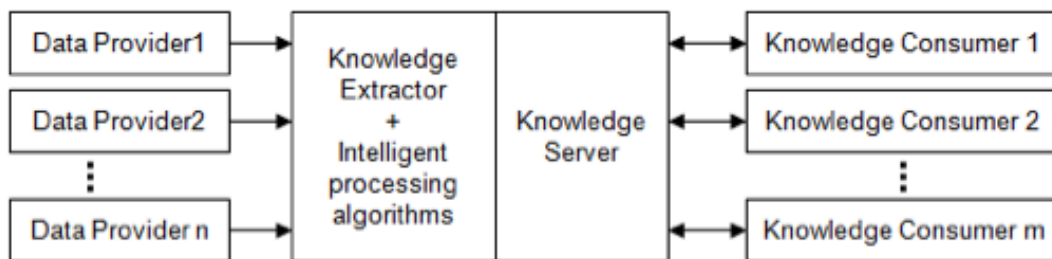
### 2.3.3 *Knowledge as a Service*

O *Knowledge as a Service* (KaaS)[Xu e Zhang 2005] é um novo modelo de Serviços Web, cujo foco passa a ser prover informações com descrições semânticas, aumentando a expressividade e provendo conhecimento. Essas informações são modeladas e descritas em cima de ontologias, permitindo uma padronização do formato e do significado.

Ademais e o conhecimento pode ser compartilhado com outros que tenham conhecimento do modelo da ontologia utilizada para descrever os dados. Os Serviços Web que seguem esse modelo têm a capacidade de acessar dados de diferentes fontes, instanciando suas representações e gerando conhecimento.

A Figura 2.7 apresenta um exemplo de arquitetura de um serviço KaaS. Os *Data Providers* são *datasets* existentes na Internet, onde o módulo *Knowledge Extractor + Intelligent Processing Algorithms* se conecta e extrai dados, processando-os com algoritmos, como por exemplo, de mineração de dados e de agregação dos resultados semelhantes, e então servindo-os a partir do módulo *Knowledge Server* para os consumidores (*Knowledge Consumer*).

Figura 2.7: Exemplo de uma arquitetura conceitual de Serviço Web KaaS



Fonte: [Xu e Zhang 2005]

É importante ressaltar três características essenciais do KaaS que fazem esse paradigma apropriado para criar ambientes de convergência:

1. KaaS é baseado em modelos semânticos, como ontologias, fazendo uma padronização no formato e no significado, e permitindo um melhor compartilhamento entre os participantes que adotam o mesmo modelo específico;
2. KaaS age como um componente de consolidação, desde que ele possa buscar informações para instanciar seus modelos, transformando essas informações em conhecimento;
3. Processos inteligentes usados nos servidores KaaS, como por exemplo algoritmos Mineração de Dados, podem enfatizar o conhecimento antes implícito ou impossível de ser entregue a partir das fontes primárias, devido ao conteúdo fraco e a

representação informal dessas informações.

O modelo de *Web Services* a ser seguido pelo trabalho dessa dissertação é o *Knowledge as a Service* (KaaS). Nele existe um módulo que captura informações semânticas de vários *datasets* presentes no *Linked Data*, incluindo o **DBPedia**, processa em cima de todas as informações recuperadas e retorna ao usuário as informações enriquecidas sobre os metadados da TV Digital.

## 2.4 Conclusões

Neste capítulo, foi apresentada uma visão geral da fundamentação teórica desse trabalho, discutindo os principais assuntos abordados. Foram apresentados conceitos sobre a TV Digital Interativa, pois essa dissertação propõe um ambiente de convergência (TVDi e Web).

Como esse trabalho utiliza os conceitos da Web Semântica e *Linked Data* para o desenvolvimento da Consulta Semântica, foram apresentados rapidamente os conceitos de ontologias e Web Semântica. Também é utilizado um formato de Web Service para disponibilizar o serviço de Consulta Semântica, o modelo *Knowledge as a Service* (KaaS), pois o o serviço proposto por esta dissertação tem como objetivo oferecer informações e conhecimento com a expressividade semântica.

No próximo capítulo, serão apresentados mais detalhes sobre o trabalho desenvolvido.

# Capítulo 3

## Knowledge TV e Consultas

### Semânticas Baseadas em *Linked Data*

Neste capítulo, é apresentado o trabalho de consultas semânticas baseadas em *Linked Data* em ambientes de convergência digital (TV Digital Interativa e Web). Inicialmente será apresentada a contextualização com o projeto Knowledge TV. Em seguida, será apresentada uma visão geral dos conceitos das consultas semânticas, seguindo da arquitetura detalhada do serviço desenvolvido, bem como da metodologia utilizada para seu desenvolvimento, a fim de afirmar a hipótese apresentada Seção 1.1.

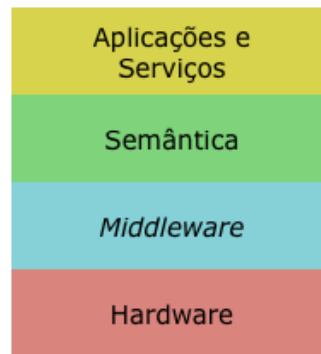
#### 3.1 *Knowledge TV*

O *Knowledge-TV* (KTV)[Lino et al. 2011] é um projeto financiado pela Rede Nacional de Pesquisa (RNP) e tem como objetivo melhorar a experiência dos usuários e produtores de conteúdo da TVDi, assim como investigar serviços avançados. Para isso, este trabalho apresenta a definição de uma camada semântica para a plataforma de TVDi, baseado em conceitos da Web Semântica, organizando o conteúdo e oferecendo serviços inteligentes, como consultas semânticas e recomendação.

A arquitetura simplificada do KTV no ambiente da TVDi é composta por quatro camadas: camada de *hardware*, camada de *middleware*, camada semântica e camada de aplicações e serviços, apresentadas na Figura 3.1.

De acordo com a Figura3.1, a camada de *hardware* concentra todos os componentes



Figura 3.1: Arquitetura geral do *Knowledge TV*

Fonte: [Lino et al. 2011]

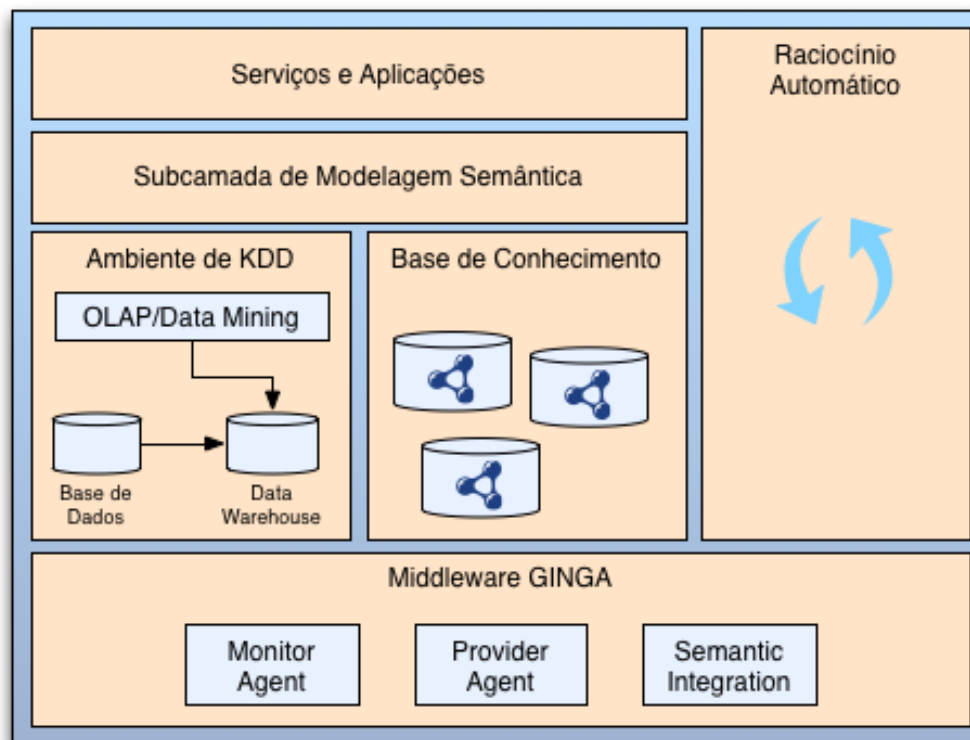
físicos que operam em um ambiente TVDi, como por exemplo a TV, entrada/saída de dados, conexão com a Internet, etc. A camada de *middleware*, no contexto do SBTVDi, representa o *middleware* GINGA, abstraindo toda a camada de *hardware* para as camadas superiores, fazendo com que não sejam necessários os detalhes do *hardware* ao desenvolver um aplicativo, por exemplo. A camada semântica é a camada proposta pelo projeto KTV, que fica entre a camada de aplicações e a do *middleware* GINGA, disponibilizando aos desenvolvedores de aplicações para a TVDi métodos para acesso aos serviços semânticos propostos pelo projeto. Essa camada também disponibiliza uma modelagem semântica rica e bases de conhecimento que descrevem os dados, recursos, serviços, aplicações e os relacionamentos entre eles, sendo toda a modelagem executada em uma linguagem formal padrão, como a OWL, permitindo o processamento automático por máquinas. A camada de Aplicação e Serviços representa os serviços e aplicações que utilizam os recursos e operações fornecidos pela camada semântica e pela camada de *middleware*.

O principal objetivo do Knowledge-TV é estender os conceitos da Web Semântica (como seus padrões, modelos e serviços) para o ambiente da TV Digital Interativa, fazendo as devidas adaptações à plataforma, buscando assim um ambiente de convergência digital. Com essa extensão, outras possibilidades serão disponibilizadas, como (i) o compartilhamento e reuso das informações e metadados presentes na TVDi, assim como os dados utilizados pelas aplicações; (ii) processamento automatizado das

informações compartilhadas, devido à descrição formal e utilização de linguagens como OWL ou RDF; (iii) conexões semânticas entre os dados, relacionando-os de semanticamente e oferecendo assim conhecimento sobre aquele domínio.

A Figura 3.2 mostra uma arquitetura conceitual do KTV, com ênfase na Camada Semântica.

Figura 3.2: Arquitetura conceitual do *Knowledge TV*



Fonte: próprio autor

A arquitetura conceitual apresentada na Figura 3.2 é composta dos seguintes módulos:

1. **Ambiente de KDD**, sendo responsável pelo ambiente de Descoberta de Conhecimento em Bases de Dados (KDD - *Knowledge Discovery in Databases*) [Han 2005]. Para tal, apresenta: (1) um banco de dados relacional; (2) um modelo analítico de dados multidimensional em um *Data Warehouse*, e (3) um módulo com ferramentas para análise de dados (OLAP - Processamento Analítico On-Line e

- Data Mining - Mineração de Dados)<sup>1</sup>;
2. **Base de Conhecimento**, onde as instâncias das entidades semânticas são armazenadas, a partir da modelagem especificada na subcamada de modelagem semântica;
  3. **Subcamada de modelagem semântica**, onde todo o conhecimento é descrito e especificado através de métodos e linguagens formais, permitindo assim processamento automático por agentes computacionais. Ou seja, onde são armazenadas as ontologias utilizadas pelo KTV;
  4. **Ambiente de Raciocínio Automático**, onde as operações de raciocínio automático (inferências, alinhamentos de ontologias e instâncias, etc.) são efetuadas com base nos modelos de conhecimento fornecidos, utilizando os mecanismos de raciocínio especificados;
  5. **Ambiente de Aplicações e Serviços**, tem o objetivo de fornecer os serviços e aplicações definidos pelo projeto KTV, como o serviço de Consultas Semânticas e de Recomendação de Conteúdo;
  6. **Middleware GINGA**, representa o *middleware* do SBTVD, fazendo parte da arquitetura do projeto, pois novos módulos necessários para o funcionamento do KTV serão inseridos no núcleo do GINGA.

Esse trabalho propõe o serviço de Consultas Semânticas, um dos serviços propostos pelo projeto *Knowledge TV*. Utilizando a arquitetura da Figura 3.2 como base e adotando os conceitos da Web Semântica, esse trabalho fornece um serviço de consultas semânticas em um ambiente de convergência TVDi e Web, com foco em enriquecer semanticamente as informações sobre os programas de TV a que os usuários assistem. Os detalhes do serviço desenvolvido serão apresentados nas próximas seções.

---

<sup>1</sup>*On-line Analytical Processing*(OLAP) [Chaudhuri e Dayal 1997] é um elemento essencial para o suporte a tomadas de decisão, provendo capacidade para manipular e analisar grandes volumes de dados sob várias perspectivas.

## 3.2 Consulta Semântica Baseada em *Linked Data* em Ambientes de Convergência

Para afirmar a hipótese apresentada na Seção 1.1, será apresentada à seguir a especificação de uma arquitetura baseada em *Linked Data* para criar um Serviço de Consultas Semânticas e oferecer um enriquecimento semântico das informações referentes aos programas de TV no ambiente da TV Digital Interativa Brasileira.

Uma das principais ferramentas existentes na Web atual são as de busca, disponibilizando aos usuários o poder de encontrar, a partir de palavras ou frases, páginas em toda a Web que contém as palavras pesquisadas. As ferramentas de busca, como o Google<sup>2</sup> e o Bing<sup>3</sup>, utilizam da técnica de indexação de páginas, que consiste em armazenar uma cópia dos textos de praticamente todas as páginas presentes na Web, e a partir de algoritmos específicos, indexa todos os dados das páginas, permitindo assim que consultas sejam feitas, retornando as páginas que apresenta os termos consultados. As pesquisas podem ser diferenciadas em dois tipos[Guha, Mccool e M 2003]:

1. **Pesquisa Navegacional**, onde o usuário pesquisa por um conjunto de palavras e espera encontrar documentos que contenham aquelas palavras. Essas palavras não têm nenhum significado explícito, nem uma interpretação, para explicitar em um conceito. O usuário somente navega em uma lista de resultados, com todas as páginas que apresentam os termos pesquisados, buscando uma página específica que contenha exatamente o que pesquisou;
2. **Pesquisa Inquisitiva**, onde o usuário pesquisa por um termo ou frase que denota um objeto, e esse objeto é do qual o usuário deseja obter informações. Não existe uma página em particular para o usuário buscar, e sim tenta agregar, a partir de várias páginas e fontes, as informações sobre aquele objeto.

Nos dois casos, o usuário vai obter uma lista de resultados e trabalhar em cima deles para obter as informações de que necessita. O conceito de Consulta Semântica[Guha,

---

<sup>2</sup><http://www.google.com>

<sup>3</sup><http://www.bing.com>

McCool e M 2003] surge para ajudar na obtenção desses resultados, retornando informações mais precisas e relevantes. A consulta semântica faz uso dos conceitos da Web Semântica para trazer resultados mais relevantes. Porém, a consulta semântica tem mais relevância quando utilizada em pesquisas do tipo Inquisitivas, pois a Web Semântica é focada nas entidades e conceitos sobre o mundo real. Por exemplo, se um usuário busca o termo “Paris”, invés de exibir todas as páginas da Web que contém esse termo, será identificado que “Paris” na verdade é uma entidade do tipo “Cidade”, e com isso retorna informações mais específicas sobre a cidade, como por exemplo, “capital da França”, ou total de habitantes, lugares históricos, etc. Outros tipos de pesquisa podem ser feitos, como por exemplo, “Qual a altura do Empire State Building?”, onde a partir de uma análise sintática da frase, e utilizando as entidades descritas formalmente na Web Semântica, retorna o valor “altura” da entidade “Empire State Building”.

Esse trabalho tem como objetivo levar os conceitos da consulta semântica para o ambiente da TV Digital Interativa e também para ambientes de convergentes (TVDi, Web e *mobile*). Com esses conceitos será possível, a partir das informações presentes nas bases de dados conectadas no *Linked Data*, enriquecer semanticamente os metadados vindos das emissoras por *broadcast*. Assim, informações mais ricas e também mais relevantes poderão ser exibidas ao telespectador, bem como outras informações relacionadas sobre o programa de TV assistido no momento, elevando o nível das informações acessadas pelo telespectador. Essas informações também possibilitam fornecer serviços avançados baseado em interatividade.

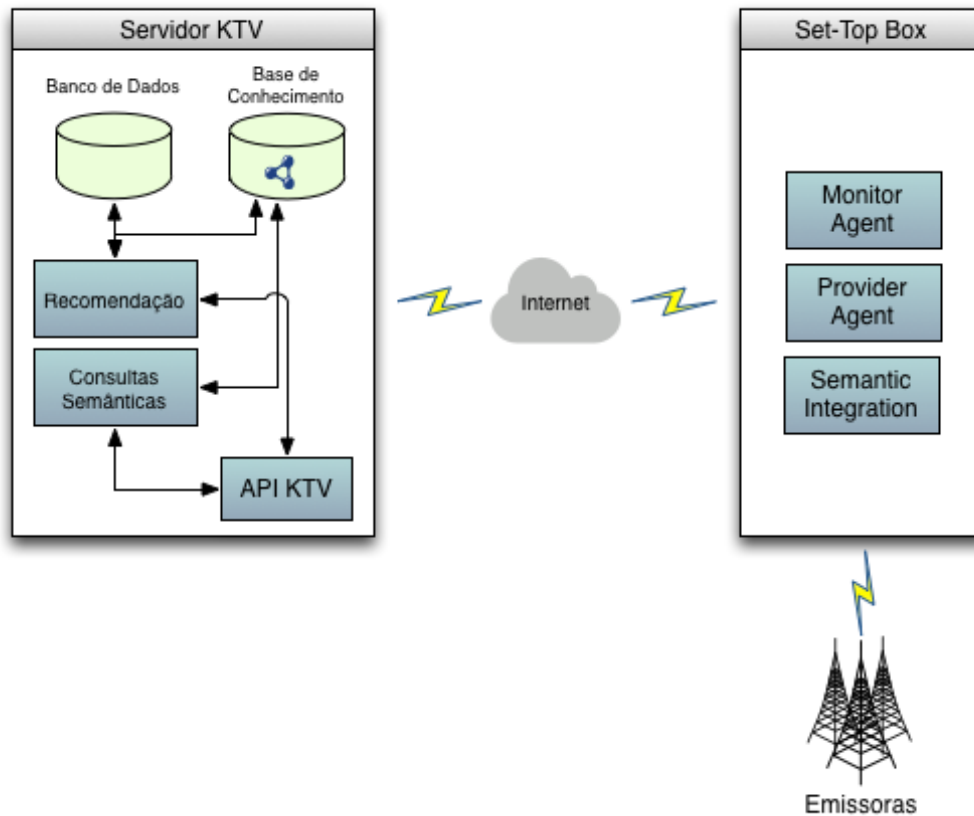
O presente trabalho é disponibilizado como um *Web Service*, deixando acessível não só para os clientes da TVDi, mas também para qualquer outro cliente que deseja obter informações enriquecidas. Como o foco desse trabalho é a TV Digital, então o serviço disponibiliza informações sobre os domínios específicos presentes na TVDi, como por exemplo, programas de TV, shows, musicais, filmes, etc. Na próxima seção será apresentada a arquitetura desse serviço.

### 3.3 Arquitetura

Sob a perspectiva da execução das operações previstas neste trabalho, tais como tarefas e técnicas de consultas semânticas a partir dos metadados provenientes das emissoras por *broadcast*, utilizando como base de dados os *datasets* presentes no *Linked Data*, foi definida e implementada uma arquitetura do tipo cliente-servidor, cujo serviço de consulta semântica é um *Web Service*, agindo como um agente de *software* e provendo uma API RESTful. Assim, não somente os usuários da TV poderão utilizar, mas qualquer outro cliente que possa conectar-se à API disponibilizada. Como o foco desse serviço é o ambiente de TV Digital e sendo um dos serviços propostos pelo projeto *Knowledge TV*, as consultas serão inicialmente restritas aos domínios de TV, tais como programas, shows, séries, atores, diretores, musicais, entre outros, porém com a possibilidade de expandir a abrangência de domínios.

Para o serviço de consultas semânticas no ambiente da TV Digital Interativa, foi especificada a arquitetura geral, apresentada na Figura 3.3. Nela existe um cliente (podendo ser o *Set Top Box* (STB), dispositivos móveis, etc.), sendo, nesse exemplo, o STB, com módulos inseridos no *middleware* e o servidor do *Knowledge TV*, sendo a consulta semântica um módulo inserido no servidor, onde reside todo o processamento das informações.

Figura 3.3: Arquitetura cliente-servidor do *Knowledge TV*, apresentando os módulos e o Set Top Box.



Os módulos inseridos no lado do STB fazem parte da especificação geral do projeto KTV[Lino et al. 2011], sendo eles:

1. *Monitor Agent*: Monitorar todas as atividades e comportamentos dos usuários do STB, registrando-os em um *log*. Os dados registrados servem para criar um perfil de recomendação para o usuário, no contexto do serviço de recomendação do KTV;
2. *Provider Agent*: Tem a função de capturar as informações provenientes das emissoras sobre o programa que o usuário está assistindo, ou seja, os metadados das tabelas SI referentes às informações da programação, como por exemplo título do programa, gênero, classificação etária, entre outros;
3. *Semantic Integration*: Tem a função de enviar as informações obtidas pelos outros

dois módulos para o servidor do *Knowledge-TV* através do canal de retorno<sup>4</sup> da TVDi;

De forma resumida, a partir dos dados enviados por esses três módulos inseridos no *middleware* do STB, o módulo de Consultas Semânticas enriquece semanticamente e disponibiliza os dados enriquecidos para que o próprio STB, ou outro dispositivo, como por exemplo, um dispositivo móvel, possa exibir as informações enriquecidas ao usuário. Exemplos de utilização são apresentados na Figura 3.4, onde exibe três usuários diferentes utilizando o serviço de Consultas Semânticas. O primeiro usuário utiliza uma TV Digital com o *middleware* GINGA, e a partir de um aplicativo executado no ambiente do *middleware*, uma requisição com os dados da programação atual são enviados para o serviço de Consultas Semânticas, e os resultados retornados são exibidos na tela da TV. Como o serviço de Consultas Semânticas segue o modelo de *Web Services*, tornando-o independente de plataforma, outros usuários podem acessá-lo através de um *smartphone* ou um aplicativo para PC como visto nos outros dois usuários da Figura 3.4.

A arquitetura mais detalhada do módulo de consultas semânticas no servidor é apresentada na Figura 3.5.

Os módulos na Figura 3.5 são representados pelas caixas cinzas, e as setas representam o fluxo de comunicação entre os módulos. Os módulos do serviço de consultas semânticas SQTV apresentados na Figura 3.5 são brevemente descritos à seguir. Mais detalhes de cada módulo serão apresentados na próxima seção:

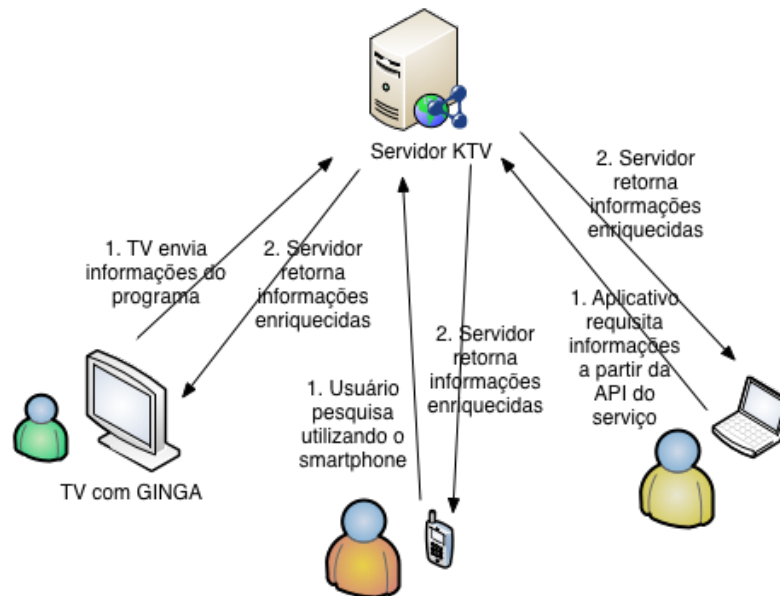
1. **Request Receiver:** o módulo que recebe todas as requisições dos clientes. Segue o modelo de *Web Services*, fazendo com que seja mais simples e fácil de utilizar a API. Ao receber os dados, faz validações antes de mandá-los ao próximo módulo, evitando assim erros;
2. **Local Searcher:** o primeiro módulo a receber os dados do módulo da API. Por ele, faz-se uma consulta no banco de dados local (*cache*) antes de transferir para

---

<sup>4</sup>Canal de retorno é um canal onde o Set-Top Box pode enviar dados, geralmente através da Internet, para a emissora ou algum Serviço Web, provendo interatividade ao usuário. [Montez e Becker 2005]

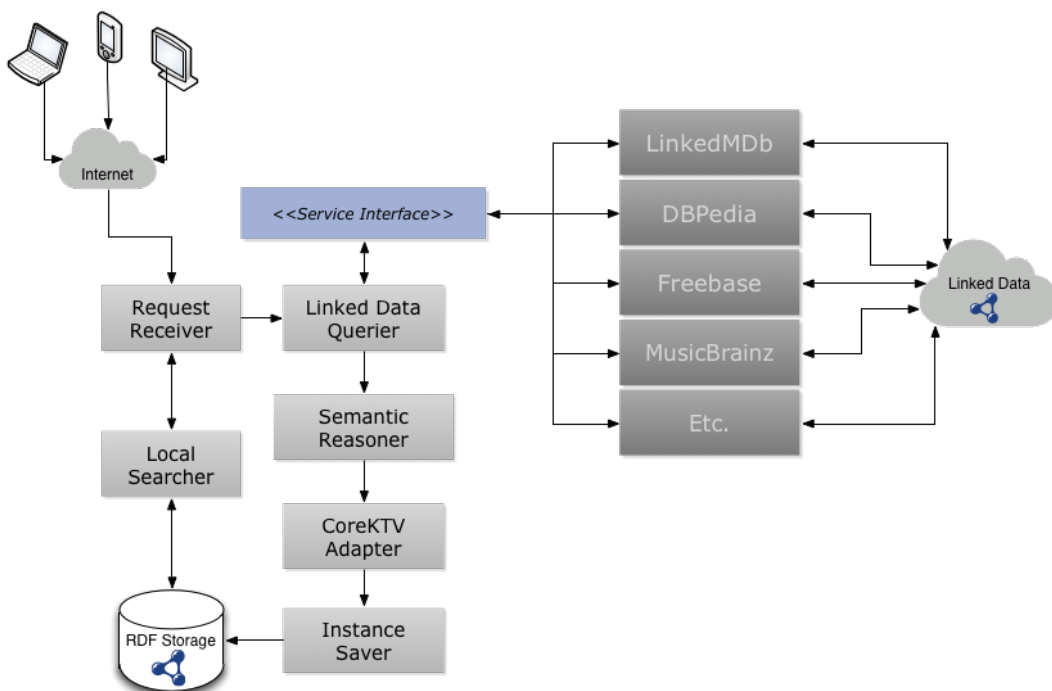


Figura 3.4: Exemplos do serviço de Consultas Semânticas



Fonte: próprio autor

Figura 3.5: Arquitetura dos módulos da consulta semântica



Fonte: próprio autor

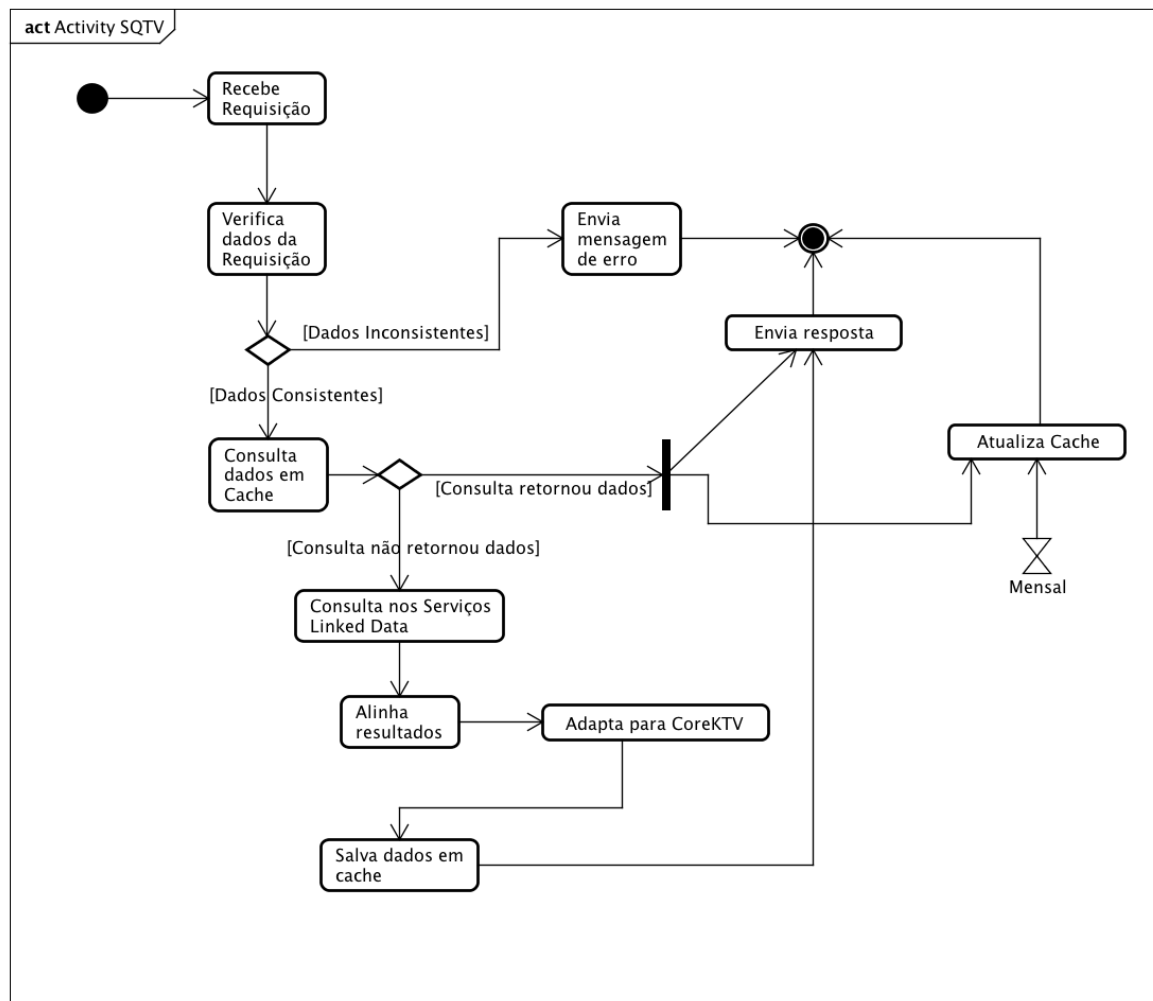
a consulta nos *datasets* do *Linked Data*. Caso encontre o resultado, retorna as informações e o cliente recebe os dados enriquecidos. Senão, retorna *vazio*. O módulo **Request Receiver** ao receber *vazio*, encaminha os dados para o próximo módulo;

3. **Linked Data Querier**: módulo responsável da consulta aos *datasets* do *Linked Data*. A partir de uma interface de programação *Service Interface*, cria uma instância dos serviços que implementam essa interface e chama os métodos implementados e armazena os resultados obtidos. Depois de obter os resultados, envia-os ao próximo módulo;
4. **Semantic Reasoner**: esse módulo é responsável por alinhar e combinar os resultados que são relacionados ou iguais. Como a consulta é realizada em vários *datasets* diferentes, o mesmo resultado pode ser encontrado em *datasets* distintos, porém descrevem o mesmo objeto pesquisado. A partir de regras de inferência (detalhadas na próxima seção) e o uso de um raciocinador, esses resultados duplicados são identificados e as propriedades são combinadas, formando uma instância única;
5. **CoreKTV Adapter**: Esse módulo é responsável por tratar todos os resultados vindos do módulo anterior, colocando-os em uma ontologia padrão utilizada pelo projeto, a CoreKTV[Araújo 2011]. Com os resultados tratados e instanciados na ontologia CoreKTV, essas instâncias são enviadas para o módulo seguinte;
6. **Instance Saver**: Esse módulo salva a instância no banco de dados de triplas local, fazendo com que as próximas consultas sobre a mesma entidade se tornem mais rápidas. Depois disso, as informações enriquecidas são retornadas ao cliente.

Para descrever melhor o fluxo seguido pelos módulos, a Figura 3.6 apresenta o diagrama de atividade do serviço de Consultas Semânticas.

O diagrama de atividades apresentado na Figura 3.6 pode ser entendido como o fluxo de uma requisição enviada por um cliente. O serviço de Consultas Semânticas inicia a execução recebendo uma requisição, que, após verificar os dados da requisição e validar a consistência dos dados, repassa os dados para realizar a consulta no *cache*.

Figura 3.6: Diagrama de atividade do serviço de Consultas Semânticas



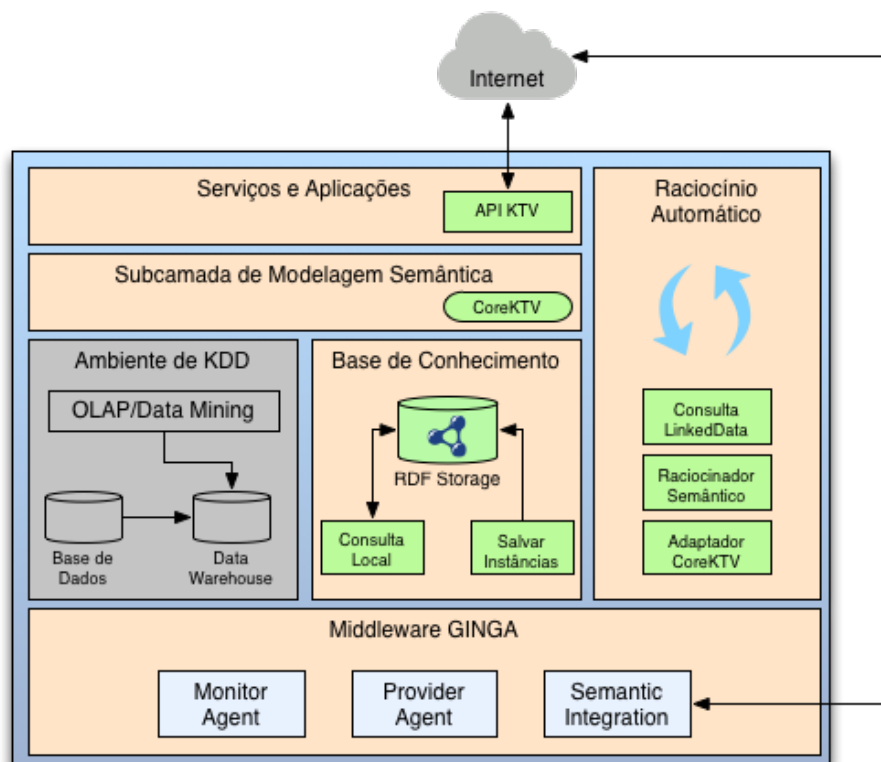
powered by Astah

Fonte: próprio autor

Caso essa consulta retorne dados, eles são enviados para o cliente e paralelamente inicia uma atualização do *cache* para os dados retornados. Caso nenhum dado seja retornado na consulta ao *cache*, inicia-se a consulta à nuvem *Linked Data* e, após obter os dados, eles são alinhados e adaptados à ontologia CoreKTV. Após isso, os dados são salvos em *cache* e então enviados ao cliente.

A Figura 3.7 apresenta os módulos definidos anteriormente instanciados na arquitetura conceitual do projeto *Knowledge TV*, mostrando o local onde cada módulo SQTV desenvolvido se encaixa.

Figura 3.7: Arquitetura conceitual do *Knowledge TV*, apresentando os módulos do serviço de consulta semântica.



Fonte: próprio autor

A próxima seção apresenta os detalhes dos módulos do serviço SQTV.

## 3.4 Módulos SQTV

Essa seção descreve com mais detalhes os módulos apresentados anteriormente.

### 3.4.1 Request Receiver

Esse módulo é responsável por receber todas as requisições, tratá-las e enviá-las para o módulo seguinte, o **Local Searcher**. Esse módulo utiliza a arquitetura para *Web Services* RESTful [Richardson e Ruby 2008]. Essa arquitetura de *Web Services* se caracteriza por ser simples e fácil de utilizar. Nela basicamente se definem convenções para a troca de informações entre o *Web Service* e o cliente, utilizando formato de dados simples, como por exemplo o XML ou JSON [Crockford 2006]<sup>5</sup>, e os métodos definidos na especificação do protocolo de aplicação *Hyper-Text Transfer Protocol* (HTTP) [Fielding et al. 1999], como o GET e POST<sup>6</sup>. A escolha por essa arquitetura de *Web Services* deu-se pela facilidade de utilização e integração do serviço em outras plataformas, frente a outra arquitetura conhecida, o *Simple Object Access Protocol* (SOAP) [Box et al. 2000]. O SOAP utiliza um padrão de XML para acessar os serviços disponibilizados por um *Web Service* a partir de chamada de métodos, utilizando uma anotação XML longa e complexa que aumenta o tamanho dos dados trafegados e a dificuldade de integração. [Mulligan e Gracanin 2009]

A partir da arquitetura RESTful, foi desenvolvido um *Web Service*, constituindo o módulo **Request Receiver**. Esse módulo, além de aceitar e receber as requisições vindas dos clientes, valida os dados antes de enviar os dados da requisição para o módulo seguinte, o **Local Searcher**. A validação feita é basicamente relacionada ao formato dos dados recebidos e enviados (se seguem o padrão e formatação). Depois de validados, os dados são enviados ao próximo módulo, descrito na próxima Seção.

---

<sup>5</sup> *JavaScript Object Notation* (JSON) é um formato de dados leve e estruturado para troca de informações baseado em pares nome/valor, independente de linguagem de programação.

<sup>6</sup> O HTTP apresenta vários tipos (métodos) especificados para a troca de mensagem, sendo os principais os GET, POST, PUT e DELETE. Esses métodos podem ser utilizados livremente, onde a semântica deles depende da forma que a aplicação os utiliza.

### 3.4.2 Local Searcher

Esse módulo, relativamente pequeno, executa uma consulta SPARQL no banco de dados de triplas local. Caso algum resultado seja encontrado, eles são retornados diretamente para o cliente que requisitou a consulta. Caso contrário, a requisição é enviada ao próximo módulo, o **Linked Data Querier**. A utilização desse módulo em conjunto com um *cache* de dados é necessário para reduzir a latência que ocorre na consulta nos *datasets* do *Linked Data*. Tal latência foi evidenciada a partir de testes realizados durante o desenvolvimento e validação dessa arquitetura. Para definir a política de atualização da base de instâncias, foi observado o período de atualização média dos *datasets* utilizados pelo projeto. A partir dessa análise, foi observado que o tempo entre as atualizações dos *datasets* são demasiadamente grandes, como por exemplo a DBPedia atualiza os dados a cada 6 meses, enquanto outros *datasets* apresentam atualizações mais frequentes. Com isso, foi definido uma política de atualização dos dados do *cache* em 1 mês, ou seja, a cada mês os dados são invalidados, forçando a consulta aos *datasets* novamente. A próxima Seção apresenta o módulo **Linked Data Querier**, sendo o sucessor do módulo **Local Searcher**, caso nenhum dado seja encontrado no *cache*, ou os dados se encontram inválidos.

### 3.4.3 Linked Data Querier

Esse é um dos principais módulos, responsável pelas consultas nos *datasets* do *Linked Data*. A arquitetura especificada permite que, a partir da implementação da *interface ServiceInterface*, mais *datasets* possam ser incluídos. Após acessar todos os serviços implementados, a lista de resultados obtida é enviada ao próximo módulo, para o processamento e refinamento dos mesmos.

Os *datasets* implementados são:

1. **Freebase**: Base de dados colaborativa[Bollacker et al. 2008], onde usuários podem adicionar e editar os dados presentes. Esse *dataset* implementa alguns conceitos da Web Semântica, como estruturação dos dados em domínios, tipos e propriedades, armazenados em um banco de dados de triplas. Porém o acesso a esses dados não segue os padrões do movimento *Linked Data*, ou seja, com um

SPARQL *endpoint*. Nesse caso foi necessário adaptar o acesso a esse *dataset*, de forma a recuperar os dados descritos por RDF.

2. **DbPedia**: Esse *dataset* é uma cópia da Wikipedia[Bizer et al. 2009], porém nesse processo os dados são adaptados e instanciados na ontologia própria da DbPedia. Esse *dataset* segue todas as regras do movimento *Linked Data*, provendo assim um *endpoint* SPARQL, onde consultas possam ser realizadas com mais facilidade.
3. **LinkedMDB**: Esse *dataset* é uma cópia do principal *site* de informações sobre filmes e séries[Hassanzadeh e Consens 2009], o IMDB<sup>7</sup>. Esse *dataset* também segue os padrões do *Linked Data*, possuindo informações estruturadas sobre filmes e séries de TV provenientes do IMDB.
4. **BBC Programmes**: Esse *dataset* é mantido pela emissora BBC, contendo informações sobre os programas e filmes transmitidos pela emissora. Também segue os conceitos e padrões do *Linked Data*, com um *endpoint* SPARQL.

Essa arquitetura de serviços “expansíveis”, a partir da implementação da *interface ServiceInterface*, foi definida devido à necessidade de ter mais de um *dataset* conectado ao serviço. Nos testes realizados no desenvolvimento e validação da arquitetura foi detectado que alguns *datasets* podem perder a conexão, e no cenário onde só um *dataset* é utilizado, isso acarretaria em retornar resultados vazios, mesmo que os resultados existam no *dataset*. Outro problema é que diferentes informações podem ser encontradas em diferentes *datasets*, e essas informações podem se complementar caso seja feita uma combinação dos mesmos, com a possibilidade de retornar informações mais ricas e precisas ao usuário.

O módulo realiza a consulta em cada serviço, armazenando os resultados e repassando-os para o próximo módulo, o **Semantic Reasoner**, detalhado na próxima seção.

---

<sup>7</sup><http://www.imdb.com>

### 3.4.4 Semantic Reasoner

Na consulta nos *datasets* no módulo anterior várias informações podem ser recuperadas, incluindo resultados relacionados à mesma entidade ou de entidades diferentes vindos de *datasets* diferentes. Por exemplo, se um usuário pesquisar por “*Friends*”, vários resultados podem ser obtidos vindos das diferentes fontes, como o seriado de TV *Friends* e também o filme *With Friends Like These*. Também é possível obter os mesmos resultados relacionados ao seriado de TV *Friends*, porém recuperados de *datasets* diferentes e provavelmente modelados com ontologias diferentes, apresentando informações que podem ser complementares entre si. O módulo **Semantic Reasoner** tem como objetivo alinhar os resultados, descobrindo se são iguais ou diferentes e tratando-os, combinando os resultados iguais e relacionando outros resultados similares.

A partir do motor de inferência baseado em regras<sup>8</sup> presente no *framework* Jena[Jena e Fuseki 2004] utilizado pelo projeto, algumas regras básicas foram definidas para inferir os resultados semelhantes. O uso do motor de inferência para o alinhamento dos resultados é uma abordagem simples, pois existe um trabalho de mestrado em andamento[Melo 2014] para fornecer um alinhamento mais sofisticado dos resultados. A partir dessa abordagem, foi possível desenvolver algumas regras de inferência simples, que utilizam os atributos e relacionamentos entre os dados obtidos para criar um resultado único, a partir de uma combinação de resultados referentes à mesma entidade.

Uma breve descrição da sintaxe das regras de inferência é apresentada no Código Fonte 3.1.

Código Fonte 3.1: Sintaxe das Regras de Inferência

---

```

1 regra      := termo, ... termo -> termo, ... termo
2
3 termo      := (node, node, node) ou function(node, ... node)
4
5 node       := uri-ref           // ex.: http://foo.com/eg
6             ou prefix:localname // ex.: rdf:type
7             ou ?varname         // variavel
8             ou 'um literal'     // valor em string (textual)
9             ou numero

```

---

<sup>8</sup><http://jena.apache.org/documentation/inference/index.html#RULEoverview>



A sintaxe apresentada no Código Fonte 3.1 define que uma **regra** é composta de dois conjuntos de **termos**, com o símbolo `->` separando-os. O conjunto do lado esquerdo indica a condição, enquanto o do lado direito indica a ação. Ou seja, caso os **termos** do lado esquerdo são verdadeiros, os **termos** do lado direito serão executados. Um **termo** pode ser um conjunto de três **nodes** (**node,node,node**) representando uma tripla RDF, ou uma **function**. Um **node** pode ser uma URI, predicado ou valor. Para exemplificar melhor, a seguir são apresentadas as regras de inferência definidas para o projeto. O Código Fonte 3.2 apresenta uma regra de inferência utilizada pelo projeto.

---

#### Código Fonte 3.2: Regra de Inferência

---

```

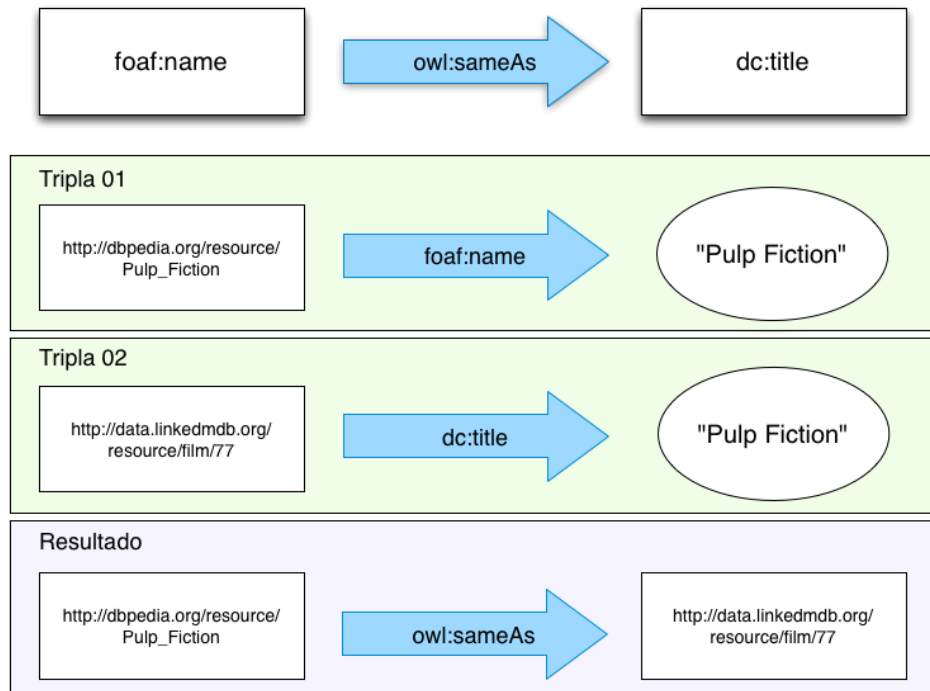
1 [rule1 :
2   (?r1 ?p1 ?v1)
3   (?r2 ?p2 ?v2)
4   (?p1 <http://www.w3.org/2002/07/owl#sameAs> ?p2)
5   equal(?v1,?v2)
6   ->
7   (?r1 <http://www.w3.org/2002/07/owl#sameAs> ?r2) ]

```

---

Nessa regra apresentada pelo exemplo 3.2, o raciocinador vai selecionar duas triplas “(?r1 ?p1 ?v1)” e “(?r2 ?p2 ?v2)”, sendo essas triplas apresentando propriedade igual ou equivalente “(?p1 <http://www.w3.org/2002/07/owl#sameAs> ?p2)” como também com os valores iguais “equal(?v1,?v2)”. Caso encontre duas triplas com essas condições, então os recursos são considerados iguais, associando-os com a propriedade “(?r1 <http://www.w3.org/2002/07/owl#sameAs> ?r2)”. Ou seja, para todos os recursos que apresentam a mesma característica com o mesmo valor, eles serão relacionados como iguais. Um exemplo prático é apresentado na Figura 3.8. Essa regra é genérica e é aplicada em todo e qualquer tripla no grafo dos resultados.

Figura 3.8: Exemplo da aplicação da regra de inferência do Código Fonte 3.2



Fonte: próprio autor

Outra regra utilizada é a apresentada no Código Fonte 3.3. Essa regra tem como objetivo relacionar todas as entidades para o tipo **MovieGender** (gênero de filme) que são do tipo `<http://dbpedia.org/ontology/Film>`. Ou seja, é feito um mapeamento entre todas as entidades do tipo *Film* da DbPedia para o tipo *MovieGender* da ontologia CoreKTV.

### Código Fonte 3.3: Regra de Inferência

```

1 [dbpedia_rule :
2   (?r1 rdf:type <http://dbpedia.org/ontology/Work>)
3   (?r1 ?p ?v)
4   (?r1 <http://xmlns.com/foaf/0.1/name> ?name)
5   (?r1 <http://dbpedia.org/ontology/abstract> ?abstract)
6   noValue(?r1, rdf:type, <http://www.semanticweb.org/ontologies/2011/8/CoreKTV.owl#
7     MovieGender>)
7   ->
8   (?r1 rdf:type <http://www.semanticweb.org/ontologies/2011/8/CoreKTV.owl#MovieGender
9     >)
9   (?r1 <http://purl.org/dc/terms/title> ?name)
10  (?r1 <http://purl.org/ontology/po/synopsis> ?abstract) ]

```

Com os resultados alinhados a partir das regras de inferência, os dados são enviados para o próximo módulo, o **CoreKTV Adapter**, para adaptar os resultados obtidos em instâncias da ontologia CoreKTV. Esse módulo é descrito na próxima Seção.

### 3.4.5 CoreKTV Adapter

A partir da lista de resultados alinhados vindos do módulo anterior (**Semantic Reasoner**), esse módulo cria uma instância da ontologia CoreKTV para cada resultado, padronizando assim as informações armazenadas na base local. Essa instanciação se dá pela extração dos valores dos resultados e colocados em uma instância nova da CoreKTV. A instância, no conceito de ontologia e OWL, é um indivíduo que utiliza os conceitos definidos na ontologia e os relaciona com valores. As instâncias geradas são então repassadas para o próximo módulo, o **Instance Saver**.

### 3.4.6 Instance Saver

Esse módulo recebe as instâncias do módulo anterior e salva todas no banco de dados de triplas Fuseki<sup>9</sup>. Esse banco de dados além de armazenar as instâncias, oferece também um *endpoint* SPARQL para consultas. O uso desse banco de dados de triplas foi motivado pela integração fácil com o *framework* Jena utilizado pelo projeto, além de ser robusto e oferecer um *endpoint* para consultas SPARQL através de um Serviço Web. Ou seja, com a utilização desse banco de dados de triplas, o projeto apresenta também um *endpoint* onde podem ser realizadas consultas em cima da base armazenada, facilitando e agilizando a recuperação das informações.

Depois que as instâncias são salvas, elas são enviadas para o cliente que requisitou a consulta, na forma de uma lista de dados em RDF, onde o processamento e tratamento se torna mais fácil no lado do cliente que requisitou.

---

<sup>9</sup>[http://jena.apache.org/documentation/serving\\_data/index.html](http://jena.apache.org/documentation/serving_data/index.html)

## 3.5 Conclusão

Neste capítulo, foram mostrados os detalhes do projeto desenvolvido neste trabalho, com a apresentação da contextualização no projeto *Knowledge-TV* e também a arquitetura desenvolvida para atender os requisitos do trabalho. Também foram detalhados os módulos desenvolvidos e definidos na arquitetura.

No próximo capítulo são discutidos alguns projetos relacionados ao proposto neste trabalho.

# Capítulo 4

## Trabalhos Relacionados

No âmbito das consultas semânticas, existe uma grande quantidade de trabalhos publicados e existentes, porém no domínio da TVDi são poucos os encontrados. Nesse Capítulo serão apresentados trabalhos relacionados que são relevantes e apresentam uma proximidade com a proposta apresentada, destacando as diferenças e melhorias em relação ao trabalho proposto.

### 4.1 *NoTube TV*

O projeto NoTube TV[Aroyo et al. 2009] é um projeto de financiado pela União Europeia que iniciou em fevereiro de 2009 e terminou em janeiro de 2012. Esse projeto visa conectar a TV e a Web a partir de modelos de dados compartilhados e redes sociais; tentar criar uma padronização de serviços para a convergência de TV e Web; e oferecer serviço de recomendação de conteúdo para os usuários. Esse projeto utiliza o enriquecimento dos metadados da TV com a finalidade de melhorar os resultados da recomendação de conteúdo, e também exibir informações enriquecidas semanticamente para o usuário da TV, tanto na tela da TV quanto em dispositivos móveis (*Second Screen*). Porém esse serviço de enriquecimento semântico dos metadados obtém os dados de várias fontes que disponibilizam a programação do canal em tempo real, como por exemplo, a BBC<sup>1</sup> e Skynet<sup>2</sup>. A partir desses dados da programação, é feito o

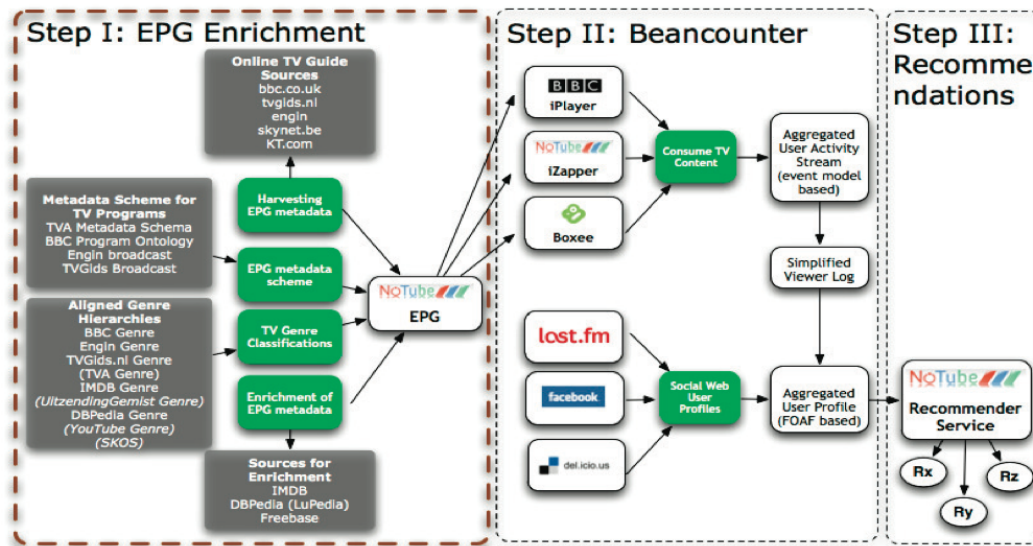
---

<sup>1</sup><http://bbc.co.uk>

<sup>2</sup><http://skynet.be>

enriquecimento semântico a partir das bases IMDB, DBPedia e Freebase. Com os resultados enriquecidos, um *Electronic Program Guide* (EPG)<sup>3</sup> é criado e disponibilizado, onde os usuários e outros serviços podem acessar e obter informações mais enriquecidas da grade de informação. A Figura 4.1 apresenta o fluxo de serviços do projeto NoTube.

Figura 4.1: Fluxo de serviços do projeto NoTube TV.



Fonte[Aroyo, Nixon e Miller 2011]

Na Figura, a parte referente ao enriquecimento de dados é a área destacada *Step 1*. A área *Step 2* apresenta a agregação das atividades do usuário, vindas de serviços como o BBC iPlayer<sup>4</sup>, onde usuários podem assistir a programação da BBC online e o Boxee<sup>5</sup>, outro serviço que disponibiliza sob-demanda programas e canais de TV e agrega outros serviços de vídeo sob-demanda, como o Netflix<sup>6</sup>. Também são coletados as atividades dos usuários nas redes sociais, como a Last.fm<sup>7</sup>, rede social especializada em compartilhar gostos musicais e o que está escutando agora. A partir desses dados, é criado um perfil do usuário, e então a recomendação é realizada em cima desse perfil.

A semelhança do projeto NoTube TV com este trabalho de mestrado é a parte do

<sup>3</sup>O *Electronic Program Guide* ou EPG é o guia de programação exibido ao usuário de uma TV, onde apresenta informações relacionadas aos atuais e futuros programas.

<sup>4</sup><http://www.bbc.co.uk/iplayer/radio>

<sup>5</sup><http://www.boxee.tv/>

<sup>6</sup><http://www.netflix.com.br>

<sup>7</sup><http://www.last.fm>

enriquecimento de metadados. Porém o projeto NoTube não é focado na TV Digital do usuário, e sim em vídeos *on-demand*, e também não disponibiliza uma API para consultas nem utiliza a abordagem KaaS, e sim um EPG enriquecido semanticamente. Esse trabalho de mestrado vai permitir, a partir da API a ser disponibilizada, que qualquer usuário, em especial os da TVDi, possam consultar qualquer coisa que esteja no domínio da TV e obter informações enriquecidas semanticamente. O atual trabalho também utilizará os agentes inseridos no *middleware* GINGA a fim de ter uma sincronização do que o usuário esteja assistindo, e então ele possa acessar essa informação por qualquer outro dispositivo.

## 4.2 *LinkedTube*

O LinkedTube[Batista e Schwabe 2009] é um projeto brasileiro, da Universidade de São Paulo, que tem como objetivo desenvolver uma plataforma para conectar os vídeos de serviços online, como o Youtube<sup>8</sup>, à entidades externas, como a Wikipedia<sup>9</sup>, IMDb, DBPedia, etc., a partir da análise dos textos encontrados nos vídeos (título, descrição, *tags*, comentários dos usuários, etc.), criando relacionamentos entre o vídeo e entidades relacionadas presentes no *Linked Data*. Esse projeto utilizou as ferramentas Zemanta<sup>10</sup> e OpenCalais<sup>11</sup>. Essas ferramentas disponibilizam uma API e são utilizadas para extrair, a partir de um texto, ligações para *sites* na Web, imagens, categorias, entidades nomeadas, entre outros. A partir do retorno dessas ferramentas, o projeto LinkedTube busca, a partir das entidades nomeadas, as entidades relacionadas no grafo *Linked Data*, e a partir daí conecta o vídeo aos resultados semânticos obtidos.

Esse projeto difere do trabalho proposto pelo fato de focar somente na questão dos vídeos online, como o serviço YouTube, enquanto essa proposta é focada no ambiente da TVDi, em especial o SBTVD. O SQTv também difere por fornecer não só *links*, mas uma infraestrutura de recuperação de informação e conhecimento.

---

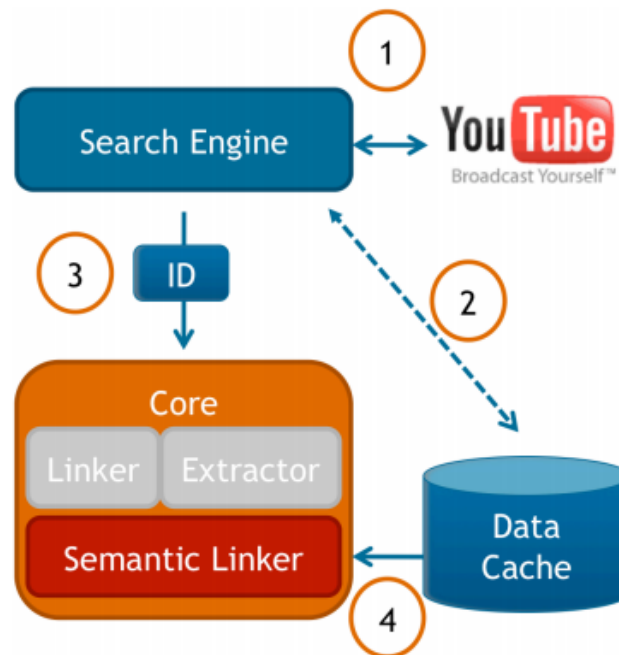
<sup>8</sup><http://www.youtube.com>

<sup>9</sup><http://www.wikipedia.org>

<sup>10</sup><http://www.zemanta.com/api/>

<sup>11</sup><http://www.opencalais.com>

Figura 4.2: Exemplo de funcionamento do LinkedTube.



Fonte: [Batista e Schwabe 2009]

### 4.3 Google Knowledge Graph

O Google Knowledge Graph [Singhal 2012] é um projeto lançado em 2012 da Google focado na Web Semântica. O objetivo desse projeto é deixar mais relevante as pesquisas feitas no Google, exibindo uma caixa de informações com dados sobre a entidade pesquisada. Por exemplo, se pesquisar no Google o termo “Paris”, além dos resultados normais de páginas indexadas, é exibido uma caixa de informações relacionadas à entidade pesquisada, que nesse caso seria a entidade Paris, capital da França, como mostra a Figura 4.3. As informações exibidas na caixa incluem além de descrições (população, clima, horário local) e imagens e/ou mapas, entidades navegáveis relacionadas ao termo pesquisado, como por exemplo os pontos de interesses em Paris.



Figura 4.3: Resultado do *Knowledge Graph* para o termo pesquisado.

Esse projeto não apresenta muitos detalhes liberados ao público. Funciona integrado aos produtos do Google e, até o momento, não apresenta uma API oficial para acesso de outros clientes, o que seria de importância fundamental para esse trabalho de mestrado. Outra questão importante é a origem dos dados, possivelmente da própria base de informações indexadas da Google e do *dataset* Freebase[Bollacker et al. 2008], uma base colaborativa mantida pelos usuários. O trabalho proposto propõe unir vários resultados de vários *datasets* disponíveis no grafo *Linked Data*, aumentando substancialmente a quantidade de resultados relevantes ao usuário.

#### 4.4 *Swoogle, Falcons e o GoPubmed*

O projeto Swoogle[Ding et al. 2004; Ding et al. 2005] e o projeto Falcon[Cheng, Ge e Qu 2008] são ferramentas de consulta sobre os documentos semânticos disponíveis na Web. O conceito de documentos semânticos para esse projeto são todos os arquivos

que apresentam a extensão RDF, ou seja, documentos apresentados na estrutura do *framework* RDF, descrito na Seção 2.2.1.1.

O projeto GoPubMed[Doms e Schroeder 2005; Doms e Schroeder 2009] também é uma ferramenta de consulta, porém os documentos obtidos são artigos na área médica, onde a indexação desses artigos é feita classificando-os à ontologias médicas, permitindo a busca dos documentos a partir da ontologia e sua classificação.

A partir do uso do Google<sup>12</sup> para consultar os documentos na Web com a extensão RDF, os resultados retornados são indexados, permitindo a consulta sobre eles. Isto é, esses projetos não utilizam o movimento *Linked Data*, onde os dados estão estruturados e conectados semânticamente, seguindo a convenção. O poder computacional necessário para realizar o *crawling*<sup>13</sup> é alto, o que pode gerar problemas como desempenho baixo e dados desatualizados, pois o serviço de *crawling*, por ser dispendioso, levará mais tempo para manter a base atualizada. O projeto apresentado nessa dissertação utiliza os *datasets* e padrões do *Linked Data*, onde são utilizadas base de dados atualizadas, realizando consultas diretamente à elas, obtendo o resultado mais rapidamente. Além disso, os projetos *Swoogle*, *Falcons* e o *GoPubMed* não são voltados nem apresentam uma *interface* API para o ambiente da TV Digital Interativa.

## 4.5 Conclusão

Neste capítulo, foram apresentados detalhes de alguns trabalhos que oferecem a funcionalidade de consulta semântica e/ou enriquecimento semântico, comparando-os com o trabalho desenvolvido nesta dissertação.

No próximo capítulo, são discutidos os experimentos realizados para validar a abordagem realizada e discutir os resultados obtidos.

---

<sup>12</sup><http://www.google.com>

<sup>13</sup>Web *crawler* é um agente Web que automaticamente navega em documentos disponíveis na Web, geralmente com a finalidade de indexar os dados

# Capítulo 5

## Resultados e Validação

Neste capítulo, são apresentados os resultados obtidos a partir do desenvolvimento do serviço, assim como uma descrição do método de validação utilizado, e o resultado da validação.

### 5.1 Resultados obtidos

Após a implementação da arquitetura da Consulta Semântica apresentada anteriormente, foram realizados testes a partir de uma lista de 500 títulos de filmes, todos em inglês, retirados de uma base de dados cedida pela empresa Netflix<sup>1</sup> para o *Netflix Prize*<sup>2</sup>, um evento da própria Netflix com objetivo de premiar quem aprimorar o algoritmo de recomendação utilizado pela empresa no serviço de *streaming* de filmes e programas de TV. A partir dos resultados obtidos da consulta semântica nessa lista, foram utilizadas métricas para medição da qualidade dos resultados, e também foram comparados com ferramentas similares. As próximas Seções descrevem os resultados obtidos e a validação desses resultados.

#### 5.1.1 Resultados de consulta a partir do ambiente Web

O primeiro teste foi a simulação de consultas a partir do ambiente Web, utilizando o JUnit[Cheon e Leavens 2006], uma ferramenta para testes unitários em Java, para

---

<sup>1</sup><http://www.netflix.com>

<sup>2</sup><http://www.netflixprize.com>

simular as requisições vindas de um cliente Web para o serviço de consultas semânticas. Também foi testada a integração do serviço com outra plataforma Web, o *JCollab*[Oliveira, Manguiera e Filho 2009], descrito na Seção 5.1.1.2.

#### 5.1.1.1 Consultas simuladas utilizando o *JUnit*

O *JUnit*[Cheon e Leavens 2006] é uma ferramenta de testes unitários para a linguagem Java[Gosling et al. 2005], com objetivo de facilitar a criação e automatização de testes, provendo suporte a captura de erros, execução e depuração, oferecendo um ambiente de testes controlado e seguro.

A partir dessa ferramenta, foram desenvolvidos testes sequenciais para o serviço de Consultas Semânticas, utilizando a base de 500 títulos de programas citados anteriormente. Nesse teste, para cada registro, foi realizada uma chamada ao serviço passando o título do programa, armazenando o resultado obtido. A partir dos resultados obtidos, algumas métricas de validação foram utilizadas. As métricas e os resultados obtidos serão descritos mais adiante.

#### 5.1.1.2 Integração com o projeto *JCollab*

O projeto *JCollab*[Oliveira, Manguiera e Filho 2009] tem como objetivo principal desenvolver uma ferramenta de criação e edição de jornais colaborativos. Para testar a interoperabilidade entre os ambientes, o *JCollab* integrou à sua ferramenta o serviço de Consultas Semânticas, oferecendo aos usuários uma forma de pesquisar e obter informações mais relevantes. A ferramenta *JCollab* é executada no ambiente Web, testando assim a integração com este ambiente. Alguns exemplos da integração podem ser vistos na Figura 5.1.

O serviço SQTV pode receber como entrada palavras-chaves que não necessariamente refletem o título ou nome de alguma entidade, como por exemplo, o nome de um filme ou de um país. Qualquer termo pode ser utilizado para realizar uma consulta, pois além do nome das entidades, outros campos, como descrição (*abstract*), são incluídos na consulta. Um exemplo de consulta que pode contribuir para um jornalista escrevendo uma matéria sobre cinema é oferecer, por exemplo, detalhes dos atores do filme “*A Beautiful Mind*”. Ao pesquisar o nome do filme, além das informações básicas

Figura 5.1: Exemplo do serviço no ambiente Web integrado ao projeto *Jcollab*

The image displays two overlapping screenshots of the JCollab Studio web application. The top screenshot shows the editing interface for a news item titled "Editando Pauta Novas tecnologias no tratamento do câncer (Telejornal da TV UFPB: 30/10/12)". The interface includes a sidebar on the left with sections for "Pauta", "Assunto", "Editoria", and "Abordagem". The main content area contains a "Busca Semântica" (Semantic Search) field with the placeholder text "palavras chaves separadas por vírgula" and a "Efetuar Busca" button. The bottom screenshot shows the same interface after a search has been performed. The "Busca Semântica" field now contains the word "cancer", and the "Resultados" (Results) section displays two search results: "The Fight: Science Against Cancer (Film)" and "Crazy Sexy Cancer (Film)". Each result includes a brief description of the film.

sobre o filme, também são obtidos os atores do filme, a partir dos relacionamentos semânticos existentes. A partir desses relacionamentos, podem-se obter mais informações sobre cada ator, por exemplo.

Para validar a integração, o campo de entrada das palavras-chave foi disponibilizado na interface Web do projeto *JCollab*, e algumas requisições ao Serviço de Consultas Semânticas foram feitas utilizando palavras-chaves sobre uma matéria jornalística digitadas por um usuário da ferramenta. A partir dessas palavras-chaves, o serviço de Consultas Semânticas consultou na nuvem *Linked Data* e retornou informações relacionadas à essas palavras-chaves. A interface Web da ferramenta *JCollab* recebeu esses dados estruturados e os apresentou de forma mais conveniente ao usuário. As requisições Web diferem das requisições do ambiente TVDi pelo fato de que no ambiente TVDi as palavras-chaves são criadas a partir dos metadados provenientes das emissoras, como os metadados obrigatórios apresentados na Seção 2.1.2.

Para uma melhor integração entre o projeto *JCollab* e o serviço de Consultas Semânticas, mais testes de usabilidade com jornalistas e colaboradores em matérias jornalísticas precisam ser realizados, utilizando a integração dos serviços e a Consulta Semântica para obter mais informações a partir de palavras-chave. Entretanto a usabilidade não foi o foco desse trabalho.

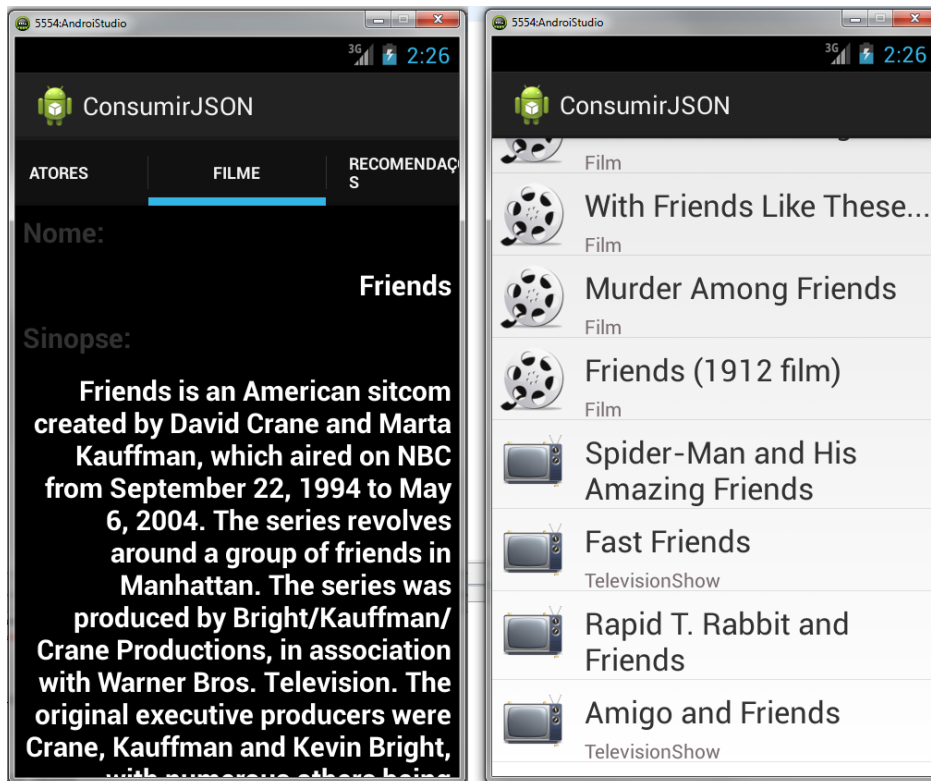
### 5.1.2 Resultados com integração em outras plataformas

A partir da implementação da API de acesso público ao serviço de consultas semânticas, foi desenvolvido um aplicativo para a plataforma de dispositivos móveis Android [Burnette 2009]. Esse aplicativo envia requisições para o servidor KTV, onde o módulo de Consultas Semânticas captura as requisições, processando os termos enviados e retornando os resultados da consulta, exibindo-os na tela. A Figura 5.2 a seguir mostra a interface do aplicativo, exibindo os resultados.

## 5.2 Validação

Para validar o trabalho desenvolvido, foram utilizados alguns métodos de validação. As métricas utilizadas na área de Recuperação de Informações (*Information Retrie-*

Figura 5.2: Exemplo do aplicativo Android



val)[Baeza-Yates, Ribeiro-Neto et al. 1999], *Precision* e *Recall*, são utilizadas, assim como um comparativo com outros serviços existentes similares ao SQTV. As próximas sub-seções detalham as métricas utilizadas.

## 5.2.1 Métricas Utilizadas na Avaliação

Para uma avaliação mais precisa, foram utilizadas métricas da área de Recuperação de Informações (*Information Retrieval*) para validar os resultados obtidos durante os testes de consulta, sendo essas (i) *Precision* e (ii) *Recall*. Essas métricas são apresentadas a seguir.

### 5.2.1.1 *Precision*

Essa métrica é bastante utilizada na área de Recuperação de Informações[Baeza-Yates, Ribeiro-Neto et al. 1999][Powers 2011], e tem como objetivo medir a relação entre os objetos relevantes obtidos com a quantidade de objetos recuperados, indiciando o quão

a consulta foi precisa. O cálculo é apresentado na Equação 5.1.

$$\text{Precision } P = \frac{\{\text{Documentos Relevantes}\} \cap \{\text{Documentos Recuperados}\}}{\{\text{Documentos Recuperados}\}} \quad (5.1)$$

### 5.2.1.2 *Recall*

Essa métrica também é bastante difundida na área de Recuperação de Informações[Baeza-Yates, Ribeiro-Neto et al. 1999][Powers 2011], que tem como objetivo medir a relação entre a quantidade de objetos recuperados com a quantidade total de objetos relevantes existentes. A medida de *recall* é expressa por um valor entre 1 e 0, sendo para o valor 1 todos os objetos esperados foram recuperados, e 0 nenhum objeto esperado foi recuperado. O cálculo é apresentado na Equação 5.2.

$$\text{Recall } R = \frac{\{\text{Documentos Relevantes}\} \cap \{\text{Documentos Recuperados}\}}{\{\text{Documentos Relevantes}\}} \quad (5.2)$$

Quando a recuperação de informações acontece no ambiente da Web Semântica, algumas considerações devem ser feitas para a utilização das métricas de Recuperação de Informação apresentadas. Levando em consideração que, na lista de termos utilizada para realizar os testes apresentam os títulos dos filmes e programas de TV, pode considerar que para cada item da lista pesquisado, é esperado que a consulta retorne pelo menos a entidade que descreve o item pesquisado. Por exemplo, se o termo pesquisado for “Pulp Fiction”, é esperado que seja retornado pelo menos 1 resultado, sendo esse resultado a entidade que descreve corretamente o filme “Pulp Fiction”. Ou seja, para esse teste, a quantidade de documentos relevantes esperados é no mínimo a quantidade de itens que a lista contém. Portanto, para todos os cálculos, a quantidade de documentos relevantes será no mínimo 500.

## 5.2.2 Comparação com ferramentas similares

Uma outra forma validar o trabalho foi comparar os resultados obtidos com ferramentas já existentes. Para isso, foram utilizados cinco serviços existentes de consultas semânticas apresentados no Capítulo 4: Google Knowledge Graph[Singhal 2012], Swoogle[Ding et al. 2004], GoPubMed[Doms e Schroeder 2005], e Falcons[Cheng, Ge e Qu 2008].



Para a comparação entre o SQTV e o Google Knowledge Graph, foi utilizada a lista de 500 títulos de programas citada na Seção 5.1. A partir dessa lista, as consultas foram feitas em ambos os serviços, passando o título do programa como a única palavra-chave, e comparando os resultados. Foi medido tanto a quantidade quanto a acurácia dos resultados.

Devido a alguns problemas, como falta de uma API para acesso, velocidade da consulta e domínios distintos, os outros serviços foram comparados a partir de características levantadas de artigos relacionados à eles.

Na próxima sessão os resultados da validação são apresentados.

## 5.3 Resultados da validação

A partir da consulta dos 500 títulos tanto no serviço SQTV quanto no Google Knowledge Graph, e utilizando as métricas descritas anteriormente, foram obtidos os seguintes resultados descritos a seguir.

### 5.3.1 Aplicação das métricas de Recuperação de Informações

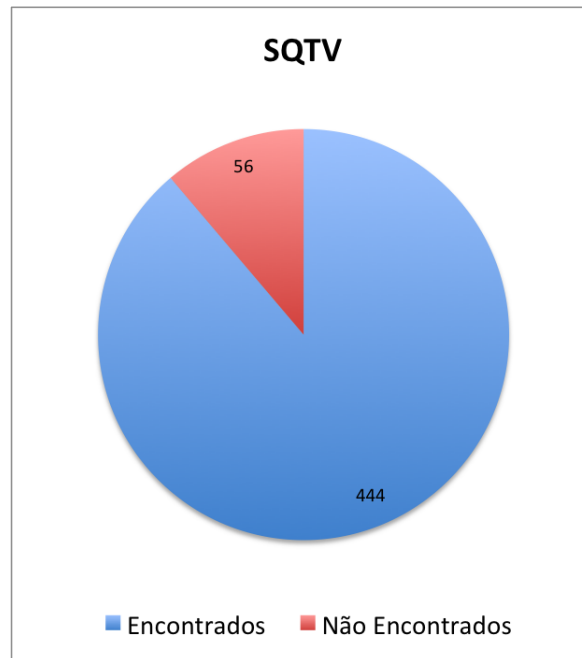
Na consulta SQTV, no total de 500 títulos pesquisados, 691 resultados foram obtidos. Porém, dentre os 691 resultados, 544 eram de fato a entidade ou objetos relacionados (por exemplo, versão mais antiga de um filme, ou a trilha sonora do filme) que o título pesquisado reflete, ou seja, documentos relevantes. A partir da Equação 5.2 para cálculo de *recall*, tem-se o resultado apresentado na Equação 5.3. Os resultados também são apresentados no gráfico da Figura 5.3. Para o cálculo de *precision*, utilizando a Equação 5.1, tem-se o resultado apresentado na Equação 5.4.

$$\text{Recall} = \frac{544 \cap 691}{544} = \frac{544}{544} = 1.000 \quad (5.3)$$

$$\text{Precision} = \frac{544 \cap 691}{691} = \frac{530}{691} = 0.767 \quad (5.4)$$

Na consulta com o Google Knowledge Graph, dos 500 itens consultados, foram recuperados 369 itens, sendo todos a entidade a que o título pesquisado se refere, ou seja,

Figura 5.3: Resultados SQTV



todas as 369 entidades encontradas são documentos relevantes. A partir da Equação 5.2, tem-se o resultado apresentado na Equação 5.5. Os resultados são apresentados no gráfico 5.4.

$$\text{Recall} = \frac{500 \cap 369}{500} = \frac{369}{500} = 0.696 \quad (5.5)$$

$$\text{Precision} = \frac{530 \cap 369}{530} = \frac{369}{530} = 1.0 \quad (5.6)$$

A partir desses resultados, foi observado que, no serviço SQTV, mais resultados relevantes foram obtidos a partir da métrica *recall*. Porém, retornou muito mais resultados do que o Google Knowledge Graph, diminuindo assim a precisão, a partir dos resultados da métrica *precision*. Devido ao uso de vários *datasets* diferentes, o SQTV retornou mais resultados para cada título pesquisado, porém após uma análise de todos os resultados, a maioria ainda pode ser considerada relevante, pois estão relacionadas à entidade principal que foi retornada. Por exemplo, ao pesquisar “*Lord of the Rings*”, o serviço SQTV retornou as entidades “*The Lord of the Rings trilogy*”<sup>3</sup> e “*The Lord of the*

<sup>3</sup>[http://en.wikipedia.org/wiki/The\\_Lord\\_of\\_the\\_Rings\\_\(film\\_series\)](http://en.wikipedia.org/wiki/The_Lord_of_the_Rings_(film_series))

Figura 5.4: Resultados Google Knowledge Graph



*Rings (1978)*<sup>4</sup>. As duas entidades são relacionadas, sendo a primeira a trilogia “*The Lord of the Rings*” lançada recentemente, e a segunda entidade refere-se ao primeiro filme da mesma ficção, lançado em 1978. Outro exemplo, quando consultado o título “*The Final Countdown*”, foi retornado 4 entidades diferentes, sendo a primeira relacionada à música<sup>5</sup>, a segunda relacionada ao álbum de músicas<sup>6</sup>, a terceira relacionada ao *tour*<sup>7</sup> de shows realizado pela banda, e a quarta entidade relacionada ao filme<sup>8</sup> de 1980. Ou seja, mesmo retornando mais resultados, e conseqüentemente reduzindo a métrica *precision*, o SQTV conseguiu obter mais resultados que o usuário possa achar relevante.

Na próxima seção outros serviços serão comparados com o SQTV.

### 5.3.2 Comparação com ferramentas similares

Outras ferramentas de pesquisa que utilizam conceitos da Web Semântica, citadas anteriormente, são comparadas nessa seção a partir das características levantadas em

<sup>4</sup>[http://en.wikipedia.org/wiki/The\\_Lord\\_of\\_the\\_Rings\\_\(1978\\_film\)](http://en.wikipedia.org/wiki/The_Lord_of_the_Rings_(1978_film))

<sup>5</sup>[http://en.wikipedia.org/wiki/The\\_Final\\_Countdown\\_\(song\)](http://en.wikipedia.org/wiki/The_Final_Countdown_(song))

<sup>6</sup>[http://en.wikipedia.org/wiki/The\\_Final\\_Countdown\\_\(album\)](http://en.wikipedia.org/wiki/The_Final_Countdown_(album))

<sup>7</sup>[http://en.wikipedia.org/wiki/The\\_Final\\_Countdown\\_Tour\\_1986](http://en.wikipedia.org/wiki/The_Final_Countdown_Tour_1986)

<sup>8</sup>[http://en.wikipedia.org/wiki/The\\_Final\\_Countdown\\_\(film\)](http://en.wikipedia.org/wiki/The_Final_Countdown_(film))

Tabela 5.1: Tabela comparativa entre as ferramentas existentes

	Base de Conhecimento	Estrutura da Base de Conhecimento	Estrutura da Base local	Ling. de Rep. de Conhecimento	Interface API	Multi-plataforma	Navegação dos resultados	Orientado a Linked Data	Domínio
<b>SQTV</b>	Datasets do Linked Data	Várias ontologias	CoreKTV	OWL	Sim	Sim	Sim	Sim	Geral, foco em TVDi
<b>GKG</b>	Freebase e datasets Google	Freebase e dataset Google	Formato próprio	Formato próprio	Não	Não	Sim	Não	Geral
<b>Swoogle</b>	RDFs indexados	Várias Ontologias	RDFs indexados	RDF	Não	Não	Não	Não	Geral
<b>GoPubmed</b>	RDFs indexados	Gene Ontology	RDFs indexados	RDF	Não	Não	Hierarquia de documentos	Não	Medicina
<b>Falcons</b>	RDFs indexados	Várias Ontologias	RDFs indexados	RDF	Sim	Sim	Sim	Não	Geral

análises e leituras de artigos relacionados às ferramentas. Primeiramente será apresentada a Tabela comparativa 5.1, e depois a explicação de cada ferramenta e sua característica.

O Google Knowledge Graph (GKG) utiliza como base de dados o Freebase [Bollacker et al. 2008], uma base de dados semântica e colaborativa e as páginas indexados pela própria Google. Por ser um produto de empresa privada, poucas informações sobre algoritmos e base de dados foram liberadas, portanto provavelmente utiliza um formato próprio para a representação de conhecimento. Não apresenta uma API aberta, porém o *dataset* Freebase disponibiliza, mas com restrições. Não apresenta opções de multi-plataforma e permite a navegação dos resultados através dos relacionamentos entre as entidades e propriedades. Um ponto importante é que o GKG, assim como as outras ferramentas pesquisadas, não seguem os conceitos do Linked Data, nem baseiam-se neles para oferecer e/ou recuperar as informações. E por último, o GKG trabalha no domínio geral, retornando resultados sobre qualquer coisa.

O Swoogle<sup>9</sup> [Ding et al. 2004] é um projeto que tem como objetivo indexar os RDFs

<sup>9</sup><http://swoogle.umbc.edu/>

encontrados a partir de buscas na Web. Essa indexação gera uma lista de documentos RDF estruturados, porém ao pesquisar é retornada uma lista de resultados de RDFs que apresentam o termo pesquisado. Também não segue as regras nem utiliza os *datasets* presentes no Linked Data. Por problemas técnicos no servidor do Swoogle, foi impossível a realização de mais testes.

O GoPubMed<sup>10</sup>[Doms e Schroeder 2005] tem como objetivo indexar todos os artigos na área de medicina, e a partir da análise dos títulos e resumos dos artigos, categorizá-los a partir da *Gene Ontology*[Ashburner et al. 2000]. Esse projeto disponibiliza somente a busca por palavras-chaves e navegação por categorias, retornando os resultados em listas, onde o usuário pode navegar a partir de outras palavras-chaves presentes nos artigos. Esse serviço também não utiliza os conceitos do Linked Data, nem provê uma API para outros serviços e aplicações usarem os dados.

O projeto Falcons<sup>11</sup>[Cheng, Ge e Qu 2008] também tem como objetivo a indexação de vários arquivos RDF na Web, identificando ontologias e entidades na Web Semântica. Apresenta os resultados na forma de lista, onde o usuário consegue identificar o tipo e a ontologia do RDF retornado. Assim como o SQTV, apresenta uma API onde outros serviços e aplicativos podem realizar consultas. A diferença entre esse projeto e o SQTV é a utilização dos *datasets* conectados no *Linked Data*. O Falcons faz a busca em cima de documentos RDFs indexados por um *Web crawler*, enquanto o SQTV, utilizando diretamente os *endpoints* dos *datasets* conectados no *Linked Data*. Essa diferença faz com que as informações obtidas estejam sempre atualizadas, pois os dados vêm diretamente da fonte. Outra vantagem é a utilização dos relacionamentos existentes entre os dados dos *datasets* presentes na nuvem *Linked Data*, aumentando a quantidade de informações relacionadas.

## 5.4 Conclusão

Neste capítulo, foi apresentado e detalhado o experimento feito para validar o trabalho desenvolvido. Também foi feito um comparativo entre esse trabalho e outras ferramen-

---

<sup>10</sup><http://gopubmed.com/>

<sup>11</sup><http://ws.nju.edu.cn/falcons/objectsearch/index.jsp>

tas existentes, apresentando uma tabela comparativa e descrevendo as semelhanças e diferenças entre cada ferramenta e o SQTV.

No próximo capítulo, serão apresentadas as conclusões desse trabalho, assim como os trabalhos futuros.

# Capítulo 6

## Considerações Finais

Este capítulo discute as conclusões acerca deste trabalho, bem como as principais contribuições do mesmo e trabalhos futuros.

### 6.1 Principais Contribuições

Inicialmente, o foco foi dado aos estudos teóricos relacionados a Ontologias, Web Semântica, *Linked Data* e raciocínio automático, bem como o emprego desses conceitos no ambiente da TV Digital Interativa Brasileira. A integração desses conceitos no ambiente da TVDi é relativamente novo, com um número reduzido de trabalhos relacionados. Esses estudos alcançaram os objetivos específicos 1 e 2.

Após os estudos aprofundados sobre Web Semântica e *Linked Data*, bem como o processo de convergência digital, foi possível investigar e levantar requisitos para a criação de um Serviço de Consultas Semânticas no contexto do projeto *Knowledge-TV*. Nesse levantamento, foi identificado que a utilização de *Web Services* e o modelo *Knowledge as a Service* foram boas escolhas devido à independência de plataforma e fácil integração com o ambiente da TVDi, com a utilização do canal de retorno para a comunicação com a Internet e acesso ao serviço. Com isso, o objetivo específico 3 foi alcançado.

Com os requisitos levantados, foi possível pesquisar melhores métodos, ferramentas e *frameworks* para a extração de dados na nuvem *Linked Data*. Foi identificado que ainda não existe um modelo consolidado para automatizar a descoberta dos *datasets*

conectados, portanto foi definido o método para implementar especificamente os datasets necessários, ou seja, os referentes aos domínios necessários para a TV. A escolha do *framework* Jena foi motivada pela robustez e facilidade de uso, com ferramentas que cobrem desde o acesso a dados no formato RDF até a persistência de dados em banco de dados de triplas. Com os métodos, ferramentas e frameworks escolhidos, o objetivo 4 foi alcançado.

A partir dos requisitos e tecnologias definidos, foi possível o desenvolvimento dos modelos para Consultas Semânticas e enriquecimento de dados. O modelo definido foi baseado na entrada de palavras-chave enviadas do ambiente da TVDi (metadados) e/ou enviadas por clientes no ambiente Web. A partir das palavras-chave, a consulta é realizada em cada *endpoint* implementado, e os dados são combinados para gerar informações mais ricas. Com isso, os objetivos específicos 5 e 6 foram alcançados.

Após o modelo definido, foi especificada a arquitetura com os módulos necessários para atendê-los. A arquitetura utilizou as tecnologias e *frameworks* definidos para tornar o serviço de Consultas Semânticas disponível para o ambiente da TVDi, e com isso foi alcançado objetivo específico 7.

Com a arquitetura especificada, o serviço de Consultas Semânticas foi implementado, a fim de validar os modelos e arquitetura propostas. A validação foi feita utilizando uma lista de títulos de filmes e programas de TV, aplicando as métricas de Recuperação de Informação e comparando com outro serviço de referência para consultas na Internet. Portanto, o objetivo específico 8 foi alcançado completamente, validando a arquitetura proposta.

A implementação validou a proposta da arquitetura apresentada, que comprova a criação do ambiente de convergência TVDi e Web, por meio da Web Semântica. Os dados apresentados nas métricas de validação, bem como nos testes de integração com outras plataformas, afirmam a hipótese apresentada na Seção 1.1, Nela, o ambiente da TV Digital as informações enriquecidas retornadas seriam equivalentes aos dados obtidos na validação, visto que os dados utilizados nos testes são aproximados a um ambiente de utilização real do serviço.



## 6.2 Trabalhos Futuros

Como trabalhos futuros, foi identificada inicialmente a necessidade de integração de mais *datasets* e serviços semânticos, aumentando assim a quantidade de fontes de dados, bem como ampliando o domínio do serviço de consultas semânticas. Também se identificou que outras abordagens para o alinhamento dos resultados obtidos podem ser utilizadas por meio de outras regras de inferências e raciocinadores semânticos, por exemplo. A seguir uma lista de possíveis trabalhos futuros:

- Validação do serviço de Consultas Semânticas com usuários da SBTVDi, onde serão validado os benefícios reais que as informações enriquecidas podem trazer para os usuários;
- Especificação de um módulo para definir previamente o domínio que o usuário deseja pesquisar, retornando assim menos resultados, porém mais precisos[Pires 2013];
- Especificação de novas regras/métodos para alinhar os resultados obtidos de diferentes *datasets*, normalizando-os para uma única ontologia padrão[Melo 2014];
- Desenvolvimento e disponibilização de serviços aplicativos para acessar o serviço SQTV;
- Testes de usabilidade na integração do serviço de Consultas Semânticas e o projeto *JCollab*, a partir de usuários (jornalistas e colaboradores) utilizando a ferramenta.

# Bibliografia

- [Alonso et al. 2004]ALONSO, G. et al. *Web services*. [S.l.]: Springer, 2004.
- [Alves et al. 2006]ALVES, L. G. P. et al. Análise comparativa de metadados em tv digital. In: *Simpósio brasileiro de redes de computadores*. [S.l.: s.n.], 2006. v. 50, p. 87–98.
- [Araújo 2011]ARAÚJO, J. P. C. Corektv - uma infraestrutura baseada em conhecimento para tv digital interativa: um estudo de caso para o middleware ginga. In: CENTRO DE INFORMÁTICA, UNIVERSIDADE FEDERAL DA PARAÍBA (UFPB). *Master of Science Dissertation at Programa de Pós-Graduação em Informática (PPGI)*. [S.l.], 2011.
- [Aroyo et al. 2009]AROYO, L. et al. NoTube – Making TV a Medium for Personalized Interaction. In: DONOSO, V. et al. (Ed.). *Networked Television : Adjunct proceedings of EuroITV 2009*. [S.l.]: Univ. Leuven, 2009. p. 22–25. ISBN: 9789081428002 EAN: 9789081428002.
- [Aroyo, Nixon e Miller 2011]AROYO, L.; NIXON, L.; MILLER, L. Notube: The television experience enhanced by online social and semantic data. In: *Consumer Electronics - Berlin (ICCE-Berlin), 2011 IEEE International Conference on*. [S.l.: s.n.], 2011. p. 269 –273.
- [Ashburner et al. 2000]ASHBURNER, M. et al. Gene ontology: tool for the unification of biology. *Nature genetics*, Nature Publishing Group, v. 25, n. 1, p. 25–29, 2000.
- [Baeza-Yates, Ribeiro-Neto et al. 1999]BAEZA-YATES, R.; RIBEIRO-NETO, B. et al. *Modern information retrieval*. [S.l.]: ACM press New York, 1999.

- [Batista e Schwabe 2009]BATISTA, C. E. C. F.; SCHWABE, D. Linkedtube: semantic information on web media objects. In: *Proceedings of the XV Brazilian Symposium on Multimedia and the Web*. ACM, 2009. (WebMedia '09), p. 22:1–22:7. ISBN 978-1-60558-880-3. Disponível em: <<http://doi.acm.org/10.1145/1858477.1858499>>.
- [Berners-Lee 2009]BERNERS-LEE, T. *Linked-data design issues*. June 2009. W3C design issue document. [Http://www.w3.org/DesignIssue/LinkedData.html](http://www.w3.org/DesignIssue/LinkedData.html). Disponível em: <<http://www.w3.org/DesignIssues/LinkedData.html>>.
- [Berners-Lee et al. 2001]BERNERS-LEE, T. et al. The semantic web. *Scientific american*, New York, NY, USA:, v. 284, n. 5, p. 28–37, 2001.
- [Bizer et al. 2009]BIZER, C. et al. Dbpedia-a crystallization point for the web of data. *Web Semantics: Science, Services and Agents on the World Wide Web*, Elsevier, v. 7, n. 3, p. 154–165, 2009.
- [Bollacker et al. 2008]BOLLACKER, K. et al. Freebase: a collaboratively created graph database for structuring human knowledge. In: ACM. *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*. [S.l.], 2008. p. 1247–1250.
- [Box et al. 2000]BOX, D. et al. *Simple object access protocol (SOAP) 1.1*. 2000.
- [Brackmann 2008]BRACKMANN, C. P. Sistema brasileiro de tv digital. *UNIVERSIDADE CATÓLICA DE PELOTAS, Pelotas*, 2008.
- [Bray et al. 1998]BRAY, T. et al. Extensible markup language (xml). *World Wide Web Consortium Recommendation REC-xml-19980210*. <http://www.w3.org/TR/1998/REC-xml-19980210>, 1998.
- [Breitman 2005]BREITMAN, K. K. *Web semântica: a internet do Futuro*. Rio de Janeiro: LTC, 2005. [S.l.]: LTC, 2005.
- [Brickley e Guha 2004]BRICKLEY, D.; GUHA, R. V. Rdf vocabulary description language 1.0: Rdf schema. 2004.

- [Brickley e Miller 2010]BRICKLEY, D.; MILLER, L. Foaf vocabulary specification 0.98. 2010.
- [Burnette 2009]BURNETTE, E. *Hello, Android: introducing Google's mobile development platform*. [S.l.]: Pragmatic Bookshelf, 2009.
- [Chaudhuri e Dayal 1997]CHAUDHURI, S.; DAYAL, U. An overview of data warehousing and olap technology. *ACM Sigmod record*, ACM, v. 26, n. 1, p. 65–74, 1997.
- [Chen 1976]CHEN, P. P.-S. The entity-relationship model - toward a unified view of data. *ACM Trans. Database Syst.*, ACM, New York, NY, USA, v. 1, n. 1, p. 9–36, mar. 1976. ISSN 0362-5915. Disponível em: <<http://doi.acm.org/10.1145/320434.320440>>.
- [Cheng, Ge e Qu 2008]CHENG, G.; GE, W.; QU, Y. Falcons: searching and browsing entities on the semantic web. In: ACM. *Proceedings of the 17th international conference on World Wide Web*. [S.l.], 2008. p. 1101–1102.
- [Cheon e Leavens 2006]CHEON, Y.; LEAVENS, G. T. A simple and practical approach to unit testing: The jml and junit way. In: *ECOOP 2002—Object-Oriented Programming*. [S.l.]: Springer, 2006. p. 231–255.
- [Crockford 2006]CROCKFORD, D. Json: The fat-free alternative to xml. In: *Proc. of XML*. [S.l.: s.n.], 2006. v. 2006.
- [Ding et al. 2004]DING, L. et al. Swoogle: a search and metadata engine for the semantic web. In: ACM. *Proceedings of the thirteenth ACM international conference on Information and knowledge management*. [S.l.], 2004. p. 652–659.
- [Ding et al. 2005]DING, L. et al. Search on the semantic web. *Computer*, IEEE, v. 38, n. 10, p. 62–69, 2005.
- [Ding e Foo 2002]DING, Y.; FOO, S. Ontology research and development. part 1-a review of ontology generation. *Journal of information science*, Sage Publications, v. 28, n. 2, p. 123–136, 2002.

- [Doms e Schroeder 2005]DOMS, A.; SCHROEDER, M. Gopubmed: exploring pubmed with the gene ontology. *Nucleic acids research*, Oxford Univ Press, v. 33, n. suppl 2, p. W783–W786, 2005.
- [Doms e Schroeder 2009]DOMS, A.; SCHROEDER, M. Semantic search with gopubmed. In: *Semantic techniques for the web*. [S.l.]: Springer, 2009. p. 309–342.
- [Dürst e Suignard 2005]DÜRST, M.; SUIGNARD, M. *Internationalized resource identifiers (IRIs)*. [S.l.], 2005.
- [Fernandes, Filho e Silveira 2004]FERNANDES, J. H. C.; FILHO, G. L. S.; SILVEIRA, G. E. da. Introdução à televisão digital interativa: Arquitetura, protocolos, padrões e práticas. In: *XXIII JAI - Livro Texto, Anais do XXIV Congresso da Sociedade Brasileira de Computação*. [S.l.]: Sociedade Brasileira de Computação, 2004. v. 2.
- [Fielding et al. 1999]FIELDING, R. et al. *Hypertext transfer protocol–HTTP/1.1*. [S.l.]: RFC 2616, June, 1999.
- [Filho, Leite e Batista 2007]FILHO, G. L. d. S.; LEITE, L. E. C.; BATISTA, C. E. C. F. Ginga-j: The procedural middleware for the brazilian digital tv system. *Journal of the Brazilian Computer Society*, SciELO Brasil, v. 12, n. 4, p. 47–56, 2007.
- [Gold et al. 2004]GOLD, N. et al. Understanding service-oriented software. *Software, IEEE*, v. 21, n. 2, p. 71 – 77, march-april 2004. ISSN 0740-7459.
- [Gómez-pérez e Corcho 2002]GÓMEZ-PÉREZ, A.; CORCHO, O. Ontology languages for the semantic web. *IEEE INTELLIGENT SYSTEMS*, v. 17, p. 54–60, 2002.
- [Gosling et al. 2005]GOSLING, J. et al. *Java (TM) Language Specification, The (Java (Addison-Wesley))*. [S.l.]: Addison-Wesley Professional, 2005.
- [Gruber et al. 1993]GRUBER, T. R. et al. A translation approach to portable ontology specifications. *Knowledge acquisition*, Academic Press, v. 5, n. 2, p. 199–220, 1993.

- [Guha, Mccool e M 2003]GUHA, R.; MCCOOL, R.; M, E. Semantic search. In: *INTERNATIONAL WORLD WIDE WEB CONFERENCE*. [S.l.]: ACM, 2003. p. 700–709.
- [Han 2005]HAN, J. *Data Mining: Concepts and Techniques*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2005. ISBN 1558609016.
- [Hartig 2012]HARTIG, O. Sparql for a web of linked data: Semantics and computability. In: *9th Extended Semantic Web Conference (ESWC2012)*. [S.l.: s.n.], 2012.
- [Haskell 1997]HASKELL, B. G. *Digital video: an introduction to MPEG-2*. [S.l.]: Springer, 1997.
- [Hassanzadeh e Consens 2009]HASSANZADEH, O.; CONSENS, M. P. Linked movie data base. In: *LDOW*. [S.l.: s.n.], 2009.
- [Heath e Bizer 2011]HEATH, T.; BIZER, C. *Linked Data: Evolving the Web Into a Global Data Space*. Morgan & Claypool, 2011. (Synthesis Lectures on the Semantic Web: Theory and Technology). ISBN 9781608454303. Disponível em: <<http://books.google.com.br/books?id=OFv59Wcfx8C>>.
- [Hendler 2009]HENDLER, J. Web 3.0 emerging. *Computer*, v. 42, n. 1, p. 111–113, jan. 2009. ISSN 0018-9162.
- [Jena e Fuseki 2004]JENA, J.; FUSEKI, P. *Semantic web frameworks*. 2004.
- [Kifer 2008]KIFER, M. Rule interchange format: The framework. In: *Web reasoning and rule systems*. [S.l.]: Springer, 2008. p. 1–11.
- [Klyne e Carroll]KLYNE, G.; CARROLL, J. J. *Resource Description Framework (RDF)*. Disponível em: <<http://www.w3.org/RDF/>>.
- [Koenen 2002]KOENEN, R. Overview of the mpeg-4 standard. *ISO/IEC JTC1/SC29/WG11 N*, v. 1730, p. 11–13, 2002.
- [Kopack e Potts 2003]KOPACK, M.; POTTS, S. *Sams Teach Yourself Web Services in 24 Hours*. [S.l.]: Pearson Education, 2003.

- [Leymann, Roller e Schmidt 2002]LEYMANN, F.; ROLLER, D.; SCHMIDT, M.-T. Web services and business process management. *IBM systems Journal*, IBM, v. 41, n. 2, p. 198–211, 2002.
- [Lino et al. 2011]LINO, N. et al. Knowledge tv. In: *Proceedings of the 9th international interactive conference on Interactive television*. ACM, 2011. (EuroITV '11), p. 29–38. ISBN 978-1-4503-0602-7. Disponível em: <<http://doi.acm.org/10.1145/2000119.2000126>>.
- [Lugmayr, Niiranen e Kalli 2004]LUGMAYR, A.; NIIRANEN, S.; KALLI, S. *Digital interactive TV and metadata: future broadcast multimedia*. [S.l.]: Springer, 2004.
- [Masinter, Berners-Lee e Fielding 2005]MASINTER, L.; BERNERS-LEE, T.; FIELDING, R. T. Uniform resource identifier (uri): Generic syntax. 2005.
- [McGuinness, Harmelen et al. 2004]MCGUINNESS, D. L.; HARMELEN, F. V. et al. Owl web ontology language overview. *W3C recommendation*, v. 10, n. 2004-03, p. 10, 2004.
- [Melo 2014]MELO, D. M. N. de. Interoperabilidade de ontologias no contexto web tvdi. Dissertação de Mestrado em Andamento. 2014.
- [Montez e Becker 2005]MONTEZ, C.; BECKER, V. *TV digital interativa: conceitos, desafios e perspectivas para o Brasil*. [S.l.: s.n.], 2005.
- [Mulligan e Gracanin 2009]MULLIGAN, G.; GRACANIN, D. A comparison of soap and rest implementations of a service based interaction independence middleware framework. In: IEEE. *Simulation Conference (WSC), Proceedings of the 2009 Winter*. [S.l.], 2009. p. 1423–1432.
- [NBR 2007]NBR, A. A. Abnt nbr 15603-1, multiplexação e serviços de informação (si) parte 1: Si do sistema de radiodifusão. *ABNT, primeira edição*, 2007.
- [NBR 2010]NBR, A. A. Abnt nbr 15606-1, codificação de dados e especificações de transmissão para radiodifusão digital, parte 1: Codificação de dados. *ABNT, segunda edição*, 2010.

- [Oliveira, Mangueira e Filho 2009]OLIVEIRA, F. Soares de; MANGUEIRA, J. A.; FILHO, G. L. de S. Architecture for production and collaboration of tv content on the internet. In: *Proceedings of the XV Brazilian Symposium on Multimedia and the Web*. ACM, 2009. (WebMedia '09), p. 51:1–51:4. ISBN 978-1-60558-880-3. Disponível em: <<http://doi.acm.org/10.1145/1858477.1858528>>.
- [O'reilly 2007]O'REILLY, T. What is web 2.0: Design patterns and business models for the next generation of software. *Communications & strategies*, n. 1, p. 17, 2007.
- [O'Reilly e Battelle 2009]O'REILLY, T.; BATTELLE, J. Web squared: Web 2.0 five years on. June 2009. Disponível em: <<http://www.web2summit.com/web2009/public/schedule/detail/10194>>.
- [Pires 2013]PIRES, A. D. de S. *Uma Abordagem para Identificação de Domínios de Aplicação em Ambiente de Convergência Digital*. 2013.
- [Powers 2011]POWERS, D. Evaluation: From precision, recall and f-measure to roc., informedness, markedness & correlation. *Journal of Machine Learning Technologies*, v. 2, n. 1, p. 37–63, 2011.
- [Prud'hommeaux e Seaborne]PRUD'HOMMEAUX, E.; SEABORNE, A. *SPARQL Query Language for RDF*. Disponível em: <<http://www.w3.org/TR/rdf-sparql-query/>>.
- [Quilitz e Leser 2008]QUILITZ, B.; LESER, U. Querying distributed rdf data sources with sparql. In: *The Semantic Web: Research and Applications*. [S.l.]: Springer, 2008. p. 524–538.
- [Raggett et al. 1999]RAGGETT, D. et al. Html 4.01 specification. *W3C recommendation*, v. 24, 1999.
- [Richardson e Ruby 2008]RICHARDSON, L.; RUBY, S. *RESTful web services*. [S.l.]: O'Reilly, 2008.
- [Singhal 2012]SINGHAL, A. Introducing the knowledge graph: things, not strings. *Official Google Blog*, May, 2012.



- [Soares e Castro 2008]SOARES, L. F.; CASTRO, P. As múltiplas possibilidades do middleware ginga. *Produção Profissional: Revista de Comunicação e Técnica Audio-visual, São Paulo*, p. 76–83, 2008.
- [Soares e Filho 2007]SOARES, L. F. G.; FILHO, G. L. d. S. Interactive television in brazil: System software and the digital divide. In: *European Conference on Interactive TV (EuroITV), Áustria*. [S.l.: s.n.], 2007.
- [Soares, Rodrigues e Moreno 2007]SOARES, L. F. G.; RODRIGUES, R. F.; MORENO, M. F. Ginga-ncl: the declarative environment of the brazilian digital tv system. *Journal of the Brazilian Computer Society, SciELO Brasil*, v. 12, n. 4, p. 37–46, 2007.
- [Souza e Alvarenga 2004]SOUZA, R. R.; ALVARENGA, L. A web semântica e suas contribuições para a ciência da informação. *Ciência da Informação, Brasília, SciELO Brasil*, v. 33, n. 1, p. 132–141, 2004.
- [Truong e Dustdar 2009]TRUONG, H.-L.; DUSTDAR, S. On analyzing and specifying concerns for data as a service. In: *Services Computing Conference, 2009. APSCC 2009. IEEE Asia-Pacific*. [S.l.: s.n.], 2009. p. 87 –94.
- [Tummarello, Delbru e Oren 2007]TUMMARELLO, G.; DELBRU, R.; OREN, E. Sindice. com: Weaving the open linked data. In: *The Semantic Web*. [S.l.]: Springer, 2007. p. 552–565.
- [Wiegand et al. 2003]WIEGAND, T. et al. Overview of the h. 264/avc video coding standard. *Circuits and Systems for Video Technology, IEEE Transactions on, IEEE*, v. 13, n. 7, p. 560–576, 2003.
- [Wooldridge e Jennings 1995]WOOLDRIDGE, M.; JENNINGS, N. R. Intelligent agents: Theory and practice. *The knowledge engineering review, Cambridge Univ Press*, v. 10, n. 02, p. 115–152, 1995.
- [Xu e Zhang 2005]XU, S.; ZHANG, W. Knowledge as a service and knowledge breaching. In: *Proceedings of the 2005 IEEE International Conference on Services Computing - Volume 01*. Washington, DC, USA: IEEE Computer So-

---

ciety, 2005. (SCC '05), p. 87–94. ISBN 0-7695-2408-7-01. Disponível em:  
<<http://dx.doi.org/10.1109/SCC.2005.60>>.

# Apêndice A

## Publicações

A seguir uma lista de trabalhos publicados até o momento relativos a essa proposta.

Amaro, M. A. L.; Lino, N. C. Q.; Siebra, C. A. and Guedes, A. L. V. (2012). The Knowledge as a Service Metaphor and its Use for Building Convergence Environments. *WebMedia 2012 - 18th Brazilian Symposium on Multimedia and the Web*, São Paulo, Brazil, 2012.

Santos, R. N.; Martins, M. B.; Amaro, M. A. L.; Patricio Jr., J. C. A. and Lino, N. C. Q. (2011). Semantic Queries in Knowledge TV Supported by Knowledge Discovery. *Proceedings of IADIS International Conference - WWW/INTERNET 2011*, Rio de Janeiro, Brazil, 2011.

Santos, R. N., Amaro, M. A. L. e Lino, N. C. Q. (2011). Consulta Semântica na TV: Uma Abordagem de Consulta Semântica em Ambientes de Convergência (Web e TV). *Workshop de Pós-Graduação (WPG) - I Escola Paraibana de Informática (EPI 2011)*, João Pessoa, PB, Brasil, 2011.

Lino, N. C. Q.; Araújo, J. P.; Anabuki, D.; Patrício Junior, J. C.; Batista, M.; Nóbrega, R.; Amaro, M.; Siebra, C. A. and Souza Filho, G. L. de. (2011). Knowledge TV. *Proceedings of the 9th EuroITV Conference (EuroITV 2011)*, Lisbon, Portugal, 2011.

## Apêndice B

# Lista de títulos utilizados na validação do trabalho

Dinosaur Planet; The Rise and Fall of ECW; Sick; Fighter; My Favorite Brunette; Lord of the Rings; Screamers; 7 Seconds; Immortal Beloved; By Dawn's Early Light; Seeta Aur Geeta; Strange Relations; Chump Change; My Bloody Valentine; Never Die Alone; Boycott; Something's Gotta Give; Ferngully 2: The Magical Rescue; Lady Chatterley; Searching for Paradise; Silent Service; Spitfire Grill; Rudolph the Red-Nosed Reindeer; The Bad and the Beautiful; Justice League; A Yank in the R.A.F.; The Weather Underground; We're Not Married; Carandiru; Richard III; Dragonheart; The Libertine; Crash Dive; Outside the Law; Invader Zim; Maya Lin: A Strong Clear Vision; Congo; Jingle All the Way; The Killing; Winter Kills; The Frogmen; Silkwood; The Powerpuff Girls Movie; Louder Than Bombs; Chain of Command; The Lemon Drop Kid; G3: Live in Concert; Mostly Martha; Notre Musique; The Great Race; Obsessed; Spartan; Silk Stockings; Rambo: First Blood Part II; Cube 2: Hypercube; Chain of Fools; Cold Blooded; Nighthawks; Fatal Beauty; Drowning on Dry Land; Arachnid; Spirit Lost; Cat and the Canary; Goddess of Mercy; Get Out Your Handkerchiefs; Cannibal Women in the Avocado Jungle of Death; Where Sleeping Dogs Lie; Sweet November; A Little Princess; Husbands and Wives; A Killer Within; Gentlemen of Fortune; Ninotchka; One Last Dance; The Chorus; Smokey and the Bandit Part 3; Tomorrow; Funny Face; Jack; The Devil's Brigade; Reservoir Dogs; Pressure; Scratch; The Last Shot; Panic in the Streets; Death to Smoochy; Dead Birds; Airplane II: The

---

Sequel; Blood and Sand; Taking Lives; Onmyoji; The Agronomist; Querelle; Dinner Rush; Cherry 2000; Neon Bible; The Cookout; Butch and Sundance: The Early Days; Marooned in Iraq; Gross Anatomy; Chato's Land; Unhook the Stars; Broken Blossoms; Woman of the Year; North by Northwest; Pressure Point; The Boston Strangler; Oleanna; Cast a Giant Shadow; Stuart Little 2; A Night at the Opera; The Hunchback of Notre Dame II; Ghost Dog: The Way of the Samurai; Charlotte's Web; Animal; The Big Clock; Herbie Rides Again; The Final Countdown; Parenthood; Saludos Amigos; Taxi; Crazy as Hell; 100 Days Before the Command; Sarah Brightman: In Concert; If These Walls Could Talk; The Devil's Own; Thunderbirds Are Go; The Avengers; Lennon Legend: The Very Best of John Lennon; In His Life: The John Lennon Story; The Hebrew Hammer; Bridget Jones's Diary; Les Liaisons Dangereuses; Jack; The Day of the Locust; Leila; Ed Wood; High Fidelity; Pay It Forward; Mandela and de Klerk; Straight to Hell; Flesh Gordon; Modern Vampires; Mother's Boys; The Plainsman; Storefront Hitchcock; Dogma; Wild Things; Chasing Amy; The Pacifier; Fear No Evil; Star 80; Tremors 4: The Legend Begins; Noir; The Taming of the Shrew; Iceman Cometh; The Last Temptation of Christ; The Swarm; Guns of the Magnificent Seven; Life or Something Like It; Sadomania; House of Sand and Fog; The Onion Field; Lies; The Flintstones in Viva Rock Vegas; Stark Raving Mad; Spun; Playing Mona Lisa; Chaplin; The Trouble with Angels; Jeans; Third World Cop; The Little Girl Who Lives Down the Lane; Party Monster; Crash Dive; Live Wire; Out for Justice; Crossroads; The Santa Clause 2; Princess Caraboo; The Replacement Killers; 20,000 Leagues Under the Sea; Captain Blood; A Night in Casablanca; Rio Lobo; Pan Tadeusz; Substitute 4: Failure is Not an Option; Wings of Desire; Hostage; The Abductors; Nightbreed; Frank Lloyd Wright; Vendetta; Igby Goes Down; Girl; Elephant; The Unsinkable Molly Brown; Happiness; The 10th Kingdom; Recess: School's Out; Graveyard Shift; Zeus and Roxanne; Millennium Actress; Untamed Heart; Girls Just Want to Have Fun; Town Without Pity; Mississippi Burning; Rabbit-Proof Fence; Michael Collins; Scorched; Simple Men; Blast; Basquiat; Pink Narcissus; Topper Returns; The Matrix: Revolutions; First Night of My Life: 2000 Seen By...; Beneath Loch Ness; City Lights; Mr. Jealousy; Princess Mononoke; Legend of 1900; Surf Ninjas; Veggie-Tales Classics: Where's God When I'm Scared?; George Carlin: Personal Favorites;

---

The Beverly Hillbillies; George Carlin: What Am I Doing in New Jersey?; Flypaper; Building the Great Pyramid; Frida; Rush Hour 2; Danielle Steel's Jewels; The Avengers '67; Journey to the Center of the Earth; Doctor Who: Lost in Time: The Patrick Troughton Years; His Secret Life; Return to Paradise; The Passion of Ayn Rand; Torn Curtain; Mrs. Soffel; The Giallo Collection: Who Saw Her Die?; Monty Python's The Meaning of Life: Special Edition; Mission Kashmir; Primus: Hallucino-Genetics Live 2004; Broadway's Lost Treasures; Glory: Bonus Material; Under Suspicion; Mail Call: The Best of Season 1; Three Days of the Condor; Mitch Hedberg: Mitch All Together; Max and Ruby: Springtime for Max and Ruby; The Twilight Zone: Vol. 27; A Thousand Acres; Blank Generation / Dancin' Barefoot; Nevada Smith; Saint Jack; The Volcano Disaster; Blue's Clues: Blue's Room: It's Hug Day; Avenging Angelo; Bubblegum Crisis: Hurricane Live!; Butterfly Kiss; Santana: Supernatural Live; Avia Vampire Hunter; Monsoon Wedding; Two Can Play That Game; The Prime of Miss Jean Brodie; From Mao to Mozart: Isaac Stern in China; E's Otherwise; Presenting Felix the Cat; Love Songs; My Side of the Mountain; Mumford; The Last Seduction II; Angie; Barbarian Queen; The Hitchhiker's Guide to the Galaxy; Summer of the Monkeys; Return to Horror High; You're Invited to Mary-Kate and Ashley's Vacation Parties; Young Einstein; Drop Dead Fred; With a Friend Like Harry; The Alamo; Sol Goode; Here is Greenwood; A Crime of Passion; Rumpole of the Bailey: Series 4; Ghosts of the Abyss: Bonus Material; King Cobra; The Blackout; Roughnecks: The Starship Troopers Chronicles: The Pluto Campaign; Maurice: Bonus Material; Bob the Builder: The Big Game; Kuffs; Hudson Hawk; Rage at Dawn; First Knight; The Santa Clause; Radioland Murders; Rangeela; Solaris; Cadfael: A Morbid Taste for Bones; New Waterford Girl; Liebestraum; Fantastic Four; Rebecca: Bonus Material; Star Trek: Enterprise: Season 3; High and Low; Bloody Mallory; Jimmy Neutron: Boy Genius; Predator 2; Tom Petty and the Heartbreakers: Live at the Olympic: The Last DJ; Bullet; Dunsmore; Knights of Ramune; The In-Laws; The Visit; American Beauty; The Masque of the Red Death / The Premature Burial; Winners & Sinners; Strangers When We Meet; Highlander: Season 4; Popular: Season 2; The Twilight Zone: Vol. 36; The Terror; The Autobiography of Miss Jane Pittman; THX 1138: Special Edition; Black Jack: The Movie; Jo Jo Dancer, Your Life is Calling; Third

---

Man on the Mountain; The Curve; My Family; Shanghai Triad; The Scarlet and the Black; Blue Planet: IMAX; Northanger Abbey; Michael McDonald: A Gathering of Friends; Particles of Truth; House of Whipcord; Baby Genius: Mozart and Friends; By Hook or By Crook; Monarch of the Glen: Series 2; Alice in Wonderland (TV Version); The Brady Bunch Variety Hour; Bobby Darin: Mack is Back; Sea of Love; What's Up, Tiger Lily?; The Smiths: The Complete Picture; Tokyo Raiders; Burn Up Excess: Vol. 1: To Serve and Protect; Walt Disney's Timeless Tales: Vol. 2; The Secrets of the Dead Sea Scrolls; Quatermass and the Pit; Speed; The Lickerish Quartet; Jimmy Neutron: Attack of the Twonkies; Offspring: Americana; Spooks & Creeps; Cloak and Dagger; The Dr. Who Collection; Bad Boy Bubby; An Ideal Husband; SCTV Network 90: Vol. 4; UFC: Ultimate Fighting Championship: Ultimate Knockouts 3; Doctor Who: The Leisure Hive; Hot Wheels: World Race; Armageddon; Talkin' Dirty After Dark; Dario Argento Collection: Vol. 2: Demons 2; My Mother Likes Women; Underground Comedy Movie; They Died with Their Boots On; Beware of a Holy Whore; Rescue Heroes: The Movie; Firestarter; Empire of the Ants; Unleashed; Shaft's Big Score!; Baazigar; Christmas with The Simpsons; The Fassbinder Collection: Pioneer in Ingolstadt; Stir Crazy; National Geographic: Inside American Power: Air Force One; The Last Dragon; Jim Brown: All-American; To Catch a Killer; Charly; Mystery Science Theater 3000: The Hellcats; Buying the Cow; Cheyenne Warrior; Dear Frankie; Matewan; Cinderella; John Mayall & The Bluesbreakers and Friends: 70th Birthday Concert; Ricky Martin: The Ricky Martin Video Collection; Elvira's Haunted Hills; The Rainbow Fish / Dazzle the Dinosaur; Marvin's Room; American Tragedy; Return of the Chinese Boxer; Green Acres: Season 1; Veronica 2030; Highlander: Season 5; Robin Hood: Prince of Thieves; The Last House on the Left; Saving Grace; Who is Cletis Tout?; Rob Roy; La Femme Nikita: Season 3; Office Killer; Lipstick; The Man Who Came to Dinner; Gilligan's Island: Season 2; Saturday Night Live: The Best of Will Ferrell 2; Real Kung Fu of Shaolin; He Loves Me, He Loves Me Not; The Winter People; He Said, She Said; Fight Club: Bonus Material; Hellbound: Hellraiser II; Rugrats All Grown Up: O' Brother!; Mafia!; In the Mood for Love; Loves of a Blonde; R.E.M.: Pop Screen; Better Than Sex; Police Academy 3: Back in Training; The Mighty; Cavalry Charge; Visitors of the Night; Running Out of Time; Bedazz-

---

led; Peter Jennings Reports: The Kennedy Assassination: Beyond Conspiracy; The History Channel Presents: The Alamo; One True Thing; Baby Einstein: Baby Da Vinci; Angel's Dance; The Hand that Rocks the Cradle; The Quest; When Dinosaurs Roamed America; Koma; Last Man Standing; Caillou: Caillou's Treasure Hunt and Other Adventures; Daud; Mahoromatic: Automatic Maiden: Summer Special; Todd McFarlane's Spawn; Nine Lives; For the Moment; Japan: Memoirs of a Secret Empire; A Change of Heart; Major League; Mobsters; The Return; The Perfect Storm; Africans in America: America's Journey Through Slavery; Inspector Morse 9: The Last Enemy; Dolores Claiborne; Homicide: Life on the Street: Season 7; Cause Celebre; Whisper Kill; Film Noir Collection: D.O.A.; The Bravados; Testosterone; 44 Minutes; City of Industry; Roger & Me; Royal Deceit; The Wire: Season 1; Curly Sue; Yu Yu Hakusho; Woman Thou Art Loosed; The Haunting of Morella; Carlton-Browne of the F.O.; Abigail's Party; Fast Food, Fast Women; Bagdad Cafe; Mr. Magoo; 11:14; Dead Ringers; Indochine; Hillary & Tenzing: Climbing to the Roof of World; The Last Man; Great Pas de Deux; A Slipping Down Life; Treading Water; The House on Carroll Street; Ken Burns' America: The Congress; Sherlock: Case of Evil; Widows' Peak; Hot War; Fantasy Mission Force; Domestic Disturbance; Saber Marionette J; I Can Do Bad All By Myself; Ernest Goes to Jail; The Librarian: Quest for the Spear; Halloween: H2O; Star Trek: The Next Generation: Season 7; Mary Poppins: Bonus Material; The Lonely Guy



# Apêndice C

## Interfaces presentes no projeto

---

### Código Fonte C.1: ServiceInterface.java

---

```
1 package br.com.ktv.linkedquery.services;
2
3 public interface ServiceInterface {
4
5     public String readFileSparql(String file, String... replaces);
6
7     public ResultSet executeQuery(String endpoint, String sparql);
8
9     public Model executeConstruct(String endpoint, String sparql);
10
11    public Model describeObject(String endpoint, String sparql);
12
13    public void saveCache(Model modelToSave);
14
15    public Model loadCache(String sparql);
16
17    public JSONArray searchByKeyword(String keyword);
18
19    public JSONObject accessObject(String uri);
20 }
```

---

### Código Fonte C.2: ApiController.java

---

```
1 package br.com.ktv.api;
2
3 @Resource
4 public class ApiController {
5
6     @Path("/")
7     public void index() {
```

---

```
8     }
9
10    @Post
11    @Path("/api/v1/receiver")
12    @Consumes
13    public void receiver(String timestamp, String id, String nome_programa, String
        nome_canal, String data_inicio, String data_fim) {
14    }
15
16    @Get
17    @Path("/api/v1/query/{keyword}")
18    public void query(String keyword) {
19    }
20
21
22    @Get
23    @Path("/api/v1/describe")
24    public void describe(String id) {
25    }
26
27 }
```

---