

# Recursos audiovisuais na *Web* de Dados: a construção do Portal Semântico Audiovisual

## *Audiovisual resources in the Data Web: The construction of the Semantic Audiovisual Portal*

Caio Saraiva CONEGLIAN<sup>1</sup>  0000-0002-6126-9113

Ana Carolina Simionato ARAKAKI<sup>2</sup>  0000-0002-0140-9110

Paula Regina Ventura Amorim GONÇALEZ<sup>3</sup>  0000-0002-5480-4106

José Eduardo SANTAREM SEGUNDO<sup>1</sup>  0000-0003-3360-7872

### Resumo

Na produção cinematográfica, uma variedade de dados filmicos é criada em todas as etapas de produção do recurso audiovisual, sendo tais dados valiosos para o aprimoramento das bases de dados e posterior recuperação de informações. As informações gerais sobre filmes podem ser encontradas, por exemplo, na Wikipedia e, de forma mais especializada, na base *Internet Movie Database*. Contudo, as buscas podem ser mais semânticas, utilizando o *Linked Movie Database* e a *DBpedia*, que demandam conhecimento acerca das tecnologias da *Web Semântica*. Por essa razão, o presente artigo objetiva apresentar o desenvolvimento do Portal Semântico Audiovisual, que reúne, de forma dinâmica, dados de diversas bases do *Linked Data*, permitindo que os usuários possam ter acesso a informações que expressam as relações e o significado que aqueles dados possuem. Para que os objetivos fossem atendidos, a pesquisa teve como procedimentos metodológicos duas etapas: na primeira etapa, foi feito um levantamento bibliográfico sobre a temática, e, na segunda etapa, foi desenvolvido o Portal Semântico Audiovisual, utilizando as bases *Linked Movie Database*, *DBpedia*, *Internet Movie Database* e *Wikipedia*. Os resultados evidenciaram o potencial semântico das bases elencadas, bem como o aprimoramento das consultas realizadas em *SPARQL Protocol and RDF Query Language*. Por fim, o estudo considera que a criação da arquitetura e o desenvolvimento do Portal Semântico Audiovisual são favorecidos com o uso das tecnologias da *Web Semântica* e das bases de dados do *Linked Data*, promovendo a materialização da *Web Semântica*, ao mesmo tempo que contribui para o domínio do audiovisual.

**Palavras-chave:** Filmes. *Linked data*. Recursos audiovisuais. SPARQL. *Web Semântica*.

### Abstract

*In film production, a variety of filmic data is created in all stages of production of the audiovisual resource. These data are valuable for the improvement of databases and later retrieval of information. General information about movies can be found, for example, on*

<sup>1</sup> Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Programa de Pós-Graduação em Ciência da Informação. Marília, SP, Brasil.

<sup>2</sup> Universidade Federal de São Carlos, Centro de Educação e Ciências Humanas, Programa de Pós-Graduação em Ciência da Informação. Rod. Washington Luís, Km 235 SP-310, 13565-905, São Carlos, SP, Brasil. Correspondência para/Correspondence to: A.C.S. ARAKAKI. E-mail: <acsimionato@ufscar.br>.

<sup>3</sup> Universidade Federal do Espírito Santo, Centro de Ciências Jurídicas e Econômicas, Departamento de Ciência da Informação. Vitória, ES, Brasil.

Apoio: Conselho Nacional de Desenvolvimento Científico e Tecnológico, Processos (141207/2018-3 Doutorado) e (431612/2016-1 Universal).

Recebido em 11 de junho de 2018, versão final rerepresentada em 13 de dezembro de 2018 e aprovada em 8 de março de 2019.

Como citar este artigo/How to cite this article

Coneglian, C.S et al. Recursos audiovisuais na *Web* de Dados: a construção do Portal Semântico Audiovisual. *Transinformação*, 31, e180046, 2019. <http://dx.doi.org/10.1590/1678-9865201931e180046>



Wikipedia and more specifically on the Internet Movie Database. Nonetheless, searches can be more semantic using the Linked Movie Database and DBpedia, but they still need to have understood of the Semantic Web technologies. For this reason, it aims to present the development of the Semantic Audiovisual Portal, which dynamically gathers information from several databases of Linked Data, allowing users to have access to the information that expresses the data relationships and the data meaning. In order to achieve the objectives, the research had as methodological procedures two stages, the first step was a bibliographic survey on the subject, and the second stage, the Semantic Audiovisual Portal was developed, using the o Linked Movie Database, DBpedia, Internet Movie Database and Wikipedia. The results evidenced the semantic potential of the bases listed, as well as the improvement of the SPARQL Protocol and RDF Query Language. It considers that the creation of the architecture and the development of the Semantic Audiovisual Portal is favored by the use of Semantic Web technologies and Linked Data databases, promoting the materialization of the Semantic Web, at the same time contributing to the audiovisual domain.

**Keywords:** Movies. Linked data. Audiovisual resources. SPARQL. Semantic Web.

## Introdução

O crescimento e a popularização da *Web* alteraram o modo com as informações são disponibilizadas, bem como o comportamento do usuário frente ao exponencial crescimento da quantidade de dados, evidenciado pela diversidade de imagens e vídeos que são produzidos atualmente.

Nesse cenário, os recursos audiovisuais são aqueles compreendidos como artefatos que comunicam a informação por meio de signos visuais e sonoros, os quais passaram a ser amplamente utilizados como complemento da informação textual, fato que tem ocorrido em todos os segmentos da sociedade. Machado (2001, p.48), ao discorrer sobre os avanços das tecnologias, faz a seguinte observação sobre os recursos audiovisuais: “[...] à medida que avançam os progressos na área da tecnologia, percebemos que a imagem eletrônica (e digital) invade todos os setores da produção audiovisual, apontando para o horizonte da mídia única de mil faces diferentes”. As telas de vídeos são os locais para os quais convergem a visualidade e a comunicação, o que indica a necessidade de estudos e propostas de ambientes que facilitem o acesso às informações referentes aos recursos audiovisuais.

O acesso a conteúdo relevante, de maneira rápida e precisa, é o desafio da *Web*, tanto que Berners-Lee, Hendler e Lassila (2001) propuseram no ano de 2001 a *Web Semântica* como um meio de tornar a *Web* mais organizada e com um nível de semântica formal mais elevado. Consequentemente à proposta da *Web Semântica*, surgiu o conceito de *Web de Dados*, estruturada pelo *Linked Data* (Berners-Lee, 2006).

O *Linked Data* diz respeito à publicação e à ligação de dados estruturados na *Web*, isto é, oferece orientações para a criação de relacionamentos entre coleções de dados conectados na *Web*, de maneira escalável e generalizada, aproximando informações antes não ligadas em um espaço de dados global. Heath e Bizer (2011, p.10, tradução nossa) expõem que “A *Web* de dados apresenta uma oportunidade revolucionária para se obter insights e valor dos dados, ao possibilitar conexões diretas entre conjuntos de dados”. Ainda segundo os autores, a utilidade e o valor desses dados aumentam conforme aumentam o número de acesso e as combinações entre eles, sendo esse o propósito da *Web de Dados*. A partir da criação de relacionamentos entre coleções de dados conectados na *Web*, propiciados pelo *Linked Data*, os usuários podem fazer consultas vinculadas às tecnologias da *Web Semântica*.

Nesse sentido, esta pesquisa objetiva apresentar o desenvolvimento de um Portal Semântico de Dados Audiovisual. O Portal Semântico Audiovisual consulta diversas bases dinamicamente por meio do *SPARQL Protocol and RDF Query Language* (SPARQL), possibilitando uma melhor busca e recuperação de informações e dados de um filme elencado, como também oferece opções relacionadas à busca.

### Web Semântica

Os ambientes informacionais digitais estão evoluindo na busca de aprimorar o nível de semântica formal que os mecanismos computacionais conseguem extrair dos dados e dos documentos. Assim, uma série de projetos

e de tecnologias foram e estão sendo desenvolvidos visando permitir que essa compreensão da semântica e do contexto dos conteúdos se torne real e implementável.

A principal iniciativa com tal intuito é a *Web Semântica*, proposta em 2001 por Berners-Lee, Hendler e Lassila, visando permitir que os agentes computacionais fossem capazes de compreender o sentido dos conteúdos disponíveis na *Web*. Os autores complementam ainda que: “A *Web Semântica* não é uma *Web* separada, mas uma extensão desta, em que as informações possuem um significado bem definido, permitindo que os computadores e as pessoas trabalhem melhor em cooperação” (Berners-Lee; Hendler; Lassila, 2001, p.30).

Inicialmente, a *Web Semântica* foi concebida conforme conceitos e princípios em que a proposta deveria se basear. Com o passar dos anos, uma série de tecnologias foram desenvolvidas visando tornar implementável e real a *Web Semântica*, possibilitando que os mecanismos computacionais fossem capazes de interagir e compreender o significado os conteúdos dispostos nos ambientes informacionais digitais.

Atualmente, há uma gama de tecnologias da *Web Semântica* que são utilizadas não apenas com foco em aplicações da *Web*, mas buscando inserir um nível mais elevado de semântica formal em diversos tipos de aplicações e processos, como a Recuperação e a Representação da Informação. Dessa forma, as tecnologias da *Web Semântica* auxiliam o desenvolvimento de diversas aplicações e ferramentas, aprimorando a eficiência desses sistemas ao atender às necessidades de informação dos usuários, permitindo uma melhor compreensão do contexto tanto do conteúdo quanto do usuário.

No âmbito deste trabalho e de grande parte das aplicações de *Web Semântica* atual, as tecnologias que mais têm tido destaque são: *Uniform Resource Identifier* (URI), *Resource Description Framework* (RDF), *Web Ontology Language* (OWL) e SPARQL.

Os identificadores únicos são essenciais, pois eles permitem que os recursos sejam tratados unicamente e que possam ser localizados com precisão. No caso, o URI é utilizado para identificar unicamente os recursos. “Os URIs identificam recursos e, portanto, são centrais para a empresa da *Web Semântica*. O uso de uma convenção global de nomenclatura (por mais arbitrária que seja a sintaxe) fornece os efeitos de rede global que impulsionam os benefícios da *Web*” (Shadbolt; Hall; Berners-Lee, 2006, p.98, tradução nossa).

Outra tecnologia é o RDF, a estrutura básica de organização dos dados, em que os recursos são interligados por meio de triplas. As triplas RDF são baseadas no princípio de que um recurso (sujeito) está interligado por meio de uma propriedade (predicado) a outro recurso (objeto). Ferreira e Santos (2013, p.21) complementam dizendo que o RDF “[...] pode ser utilizado para representar recursos de maneira simples, flexível e interpretável, podendo conter representações abrangentes, do ponto de vista cognitivo, e sintéticas para a tomada de decisão dos usuários sobre os recursos buscados”.

O RDF é uma estrutura fundamental para a interligação de recursos. No entanto, essa tecnologia não é capaz de inserir um nível semântico formal e necessário para que os dados sejam passíveis da realização de inferências e com ligações que permitam uma interpretação maior do contexto dos dados. Assim, as ontologias tornaram-se um importante instrumento da *Web Semântica*, por ser um poderoso ferramental que permite uma série de relações e contextualização dos dados. Nesse contexto, “Para o uso como tecnologia da *Web Semântica*, entende-se as ontologias como: artefatos computacionais que descrevem um domínio do conhecimento de forma estruturada, através de: classes, propriedades, relações, restrições, axiomas e instâncias” (Santarem Segundo; Coneglian, 2015, p.227).

Além disso, embora exista uma série de linguagens para a construção de ontologias, a *World Wide Web Consortium* (W3C) recomenda o uso da OWL. A OWL consegue formalizar computacionalmente as especificações das ontologias, possibilitando que os computadores tenham uma compreensão melhor do significado dos dados, além de ter como base o RDF na realização das associações. Nesse âmbito, o SPARQL é a tecnologia que permite a realização de consultas, tendo como base de sua sintaxe o RDF.

Ducharme (2013, p.19, tradução nossa) afirma que: “SPARQL é uma linguagem de consulta para dados que segue um modelo específico, mas a *Web Semântica* não é sobre a linguagem de consulta ou sobre o modelo – é sobre os dados. [...] SPARQL é a melhor maneira de obter esses dados e colocá-lo para trabalhar em seus aplicativos”.

O autor considera que, como a *Web Semântica* trata de dados, o SPARQL deve ser baseado nesse princípio, permitindo todos os benefícios que a proposta oferece.

Com a maturação das tecnologias da *Web Semântica*, esta passou por um processo denominado de materialização, em que foram desenvolvidas aplicações utilizando os seus princípios. O principal expoente desse processo é o *Linked Data*, que busca interligar os recursos utilizando as tecnologias supracitadas. Berners-Lee (2006), ao apresentar o *Linked Data*, relata quatro princípios que são a base dessa proposta:

1. Usar URIs como nomes para as coisas;
2. Usar HTTP URIs para que as pessoas possam procurar esses nomes;
3. Quando alguém procura uma URI, fornecer informações úteis, usando padrões (RDF e SPARQL);
4. Incluir *links* para outras URIs para que se possa descobrir mais coisas (Berners-Lee, 2006, s/p., tradução nossa).

Nesse sentido, Santarem Segundo e Coneglian (2016, p.240) entendem que: “A *Web Semântica*, suas tecnologias e a materialização de suas propostas na constituição do *Linked Data* são uma realidade. Isso tem mudado a maneira de realizar pesquisa na Ciência da Informação moderna”.

Diversas aplicações estão seguindo os princípios do *Linked Data* e da *Web Semântica* para interligar os seus dados e disponibilizá-los. Um projeto denominado *Linking Open Data* (LOD) reúne algumas dessas iniciativas, em que os conjuntos de dados, chamados de *datasets*, apresentam dados ligados em RDF, os quais estão necessariamente vinculados a outros *datasets* pertencentes ao LOD.

Nesse projeto, há *datasets* de diversos domínios, como saúde, publicação e audiovisual, dentre outros. Assim, na próxima seção apresenta-se o conceito de recurso audiovisual, bem como alguns *datasets* de *Linked Data* que reúnem informações acerca desse domínio.

## Recursos audiovisuais: o recurso fílmico

A presença dos recursos audiovisuais a partir do século XX tem sido a marca da sociedade em todos os setores. O termo audiovisual, composto pelas palavras áudio e visual, tem sua origem no latim. Conforme Araújo (1992) aponta, a conjunção dos termos ‘áudio’ e ‘visual’ iniciou-se na década de 30, quando os Estados Unidos desenvolveram as técnicas de som e imagem. Na França, o termo audiovisual foi adotado anos mais tarde, referindo-se ao uso simultâneo de imagens e som. No Brasil, começou ser empregado na década de 50, designando a impressão de movimento independente do suporte e captura de seu registro.

Os recursos audiovisuais apresentam-se em diferentes manifestações. A Federação Internacional de Associações de Bibliotecários (FIAB) considera que o termo agrupa os “[...] discos, fitas magnéticas, filmes, diapositivos, diafilmes, videoteipes, transparências e microfílmicos” (McCarthy; Targino, 1984, p.304). Contudo, o foco deste trabalho são recursos audiovisuais como expressões relativas aos processos fílmicos, como Simionato (2017, p.78) sintetiza: “[...] o recurso audiovisual como registro final do registro imagético em movimento, independentemente do lugar em que será armazenado ou da sua forma de produção [...]”. Isto é, os filmes compõem as produções que capturam imagem e som para serem armazenados e disponibilizados em ambientes informacionais digitais.

Moran, Masetto e Behrens (2000) pontuam que a informação audiovisual, especificamente quando em formato de filme, é instrumento valioso para o aprendizado, não só por seu caráter dinâmico, mas também pelo fato de empregar diversas linguagens.

Parente e Carvalho (2009, p.20) observam que, no cinema, a base para a composição de todas as categorias excede a origem do dispositivo, cujo conceito são as estruturas a partir das quais o cinema se expandiu, seja na sua forma tradicional ou nas formas digitais. O dispositivo pode contribuir “[...] para uma renovação da teoria do cinema, sobretudo no que diz respeito a um cinema expandido sob todas suas modalidades [...]”. No caso, os dispositivos “[...] acionam variações, transformações e posicionamentos que determinam o horizonte e uma prática em ocorrência [...]” (Parente; Carvalho, 2009, p.16). Os autores complementam que “[...] o dispositivo designa a forma como a apresentação material da obra se inscreve em uma visada sistêmica, estrutural” (Parente; Carvalho, 2009, p.23).

Segundo Martin (2003, p.18) o cinema difere “de todos os meios de expressão” e tem um “poder excepcional que vem do fato de sua linguagem funcionar a partir da reprodução fotográfica da realidade”. Por essa razão, a informação fílmica é almejada não só por produtores e profissionais da área, mas também por amadores e até mesmo usuários que procuram por informações indicativas de filmes.

O recurso fílmico apresenta várias etapas no processo de criação e fabricação da película. Conforme apontam Goliot-Lété e Vanoye (1994, p.12), a “[...] escrita do roteiro, a decupagem técnica, a filmagem, a montagem e a mixagem constituem as etapas de um processo de criação de fabricação [...]” de um filme. Nessas etapas, a produção de dados é inerente, sendo que eles já apresentam uma estrutura que poderá antecipar as condicionantes para a publicação e curadoria dos dados e informações relacionadas à fabricação de filmes.

Algumas bases já utilizam dados extraídos da fabricação de filmes para a publicação e relacionamento de dados para melhor busca e recuperação de informações sobre os filmes. A fonte de dados IMDb (disponível pelo *link* <https://www.imdb.com/>) é um bom exemplo, sendo uma das principais bases de informações sobre filmes e programas de televisão, com a apresentação de pontuações e *rankings* dos conteúdos. Um desses *rankings* apresenta a avaliação dos filmes pelos usuários, com nota até dez (10). Segundo informação disponível na base Internet Movie Database (2017), esta contém 250 milhões de itens de dados, incluindo mais de quatro milhões de filmes, programas de TV e entretenimento, e oito milhões de membros do elenco e da equipe. O IMDb foi lançado *online* em 1990 e é uma subsidiária da Amazon.com desde 1998.

As buscas podem ser mais semânticas do que aquelas realizadas no IMDb, com o uso das bases *LinkedMDB* (disponível pelo *link*: <http://www.linkedmdb.org/>) e *DBpedia* (disponível pelo *link*: <http://wiki.dbpedia.org/>). A *LinkedMDB* é uma base que disponibiliza informações sobre filmes, atores, diretores, personagens, trilhas sonoras e compositores, dentre outras, sendo capaz de fornecer informações estruturadas em RDF, auxiliar a construção de aplicações e conectar os dados sobre a temática.

A *DBpedia*, diferentemente do *LinkedMDB*, não contém somente informações de audiovisuais, sendo uma base de domínio geral, com dados sobre pessoas, filmes, locais, datas e programas de televisão, dentre diversas outras temáticas. A *DBpedia* é uma base de domínio geral bastante completa, que tem a maioria da sua informação extraída da *Wikipedia*. Por essa característica de informações gerais e por ser uma das bases mais amplas e utilizadas do *Linked Data*, há a extensão de informações sobre personagens e locais, dentre outros dados que são fornecidos pelos próprios usuários.

A partir dos pressupostos teóricos discutidos, na sequência apresentam-se os resultados e as discussões dessas pesquisas, contemplando a construção do Portal Semântico Audiovisual.

## Procedimentos Metodológicos

Para atingir os objetivos desta pesquisa – de natureza qualitativa, com caráter exploratório –, optou-se pelos procedimentos metodológicos em duas etapas: teórica e aplicada.

A etapa teórica buscou na literatura os temas de *Web Semântica* e *Audiovisual*, especificamente os recursos fílmicos, complementando o contexto teórico na área de Ciência da Informação. Os resultados dessa etapa

refletem-se nos Tópicos 2 e 3 deste trabalho, e também nos estudos já realizados sobre as bases *DBpedia* e *Lightning Memory-Mapped Database* (LMDB) (Simionato *et al.*, 2018). Na segunda etapa, de nível aplicado, foi realizado o desenvolvimento do Portal Semântico Audiovisual apresentado no decorrer deste trabalho, destacando-se o Tópico 4.

O projeto utilizou a linguagem de programação *Python* e o *framework Django*, tendo sido a coleta de dados realizada a partir dos conjuntos de dados contidos no *Linking Open Data* (*Linked Movie Database-LinkedMDB* e *DBpedia*), e na *Internet Movie Database* (IMDb) e *Wikipedia*, apresentando resultados dinâmicos da consulta realizada no Portal Semântico Audiovisual.

## Resultados e Discussão

### Portal Semântico Audiovisual

A criação de aplicações que se utilizam da base de dados do *Linked Data* é um passo subsequente para a maturação desta proposta, uma vez que há uma série de iniciativas que estão publicando e disponibilizando os seus dados abertamente. Consumir esses dados, gerando valor e permitindo inferências, é fundamental para aproveitar o potencial que eles possuem.

Nesse sentido, os diversos domínios em que os dados estão sendo publicados devem avaliar e criar ambientes que permitam tirar o máximo proveito possível, visando o aproveitamento dos dados existentes, de forma orientada aos princípios do *Linked Data*. Para ilustrar tal situação, por exemplo, no domínio das Ciências Biológicas e da Saúde, essas aplicações podem auxiliar os especialistas no momento do diagnóstico, utilizando da fonte de informação *Linked Data* como um elemento auxiliar.

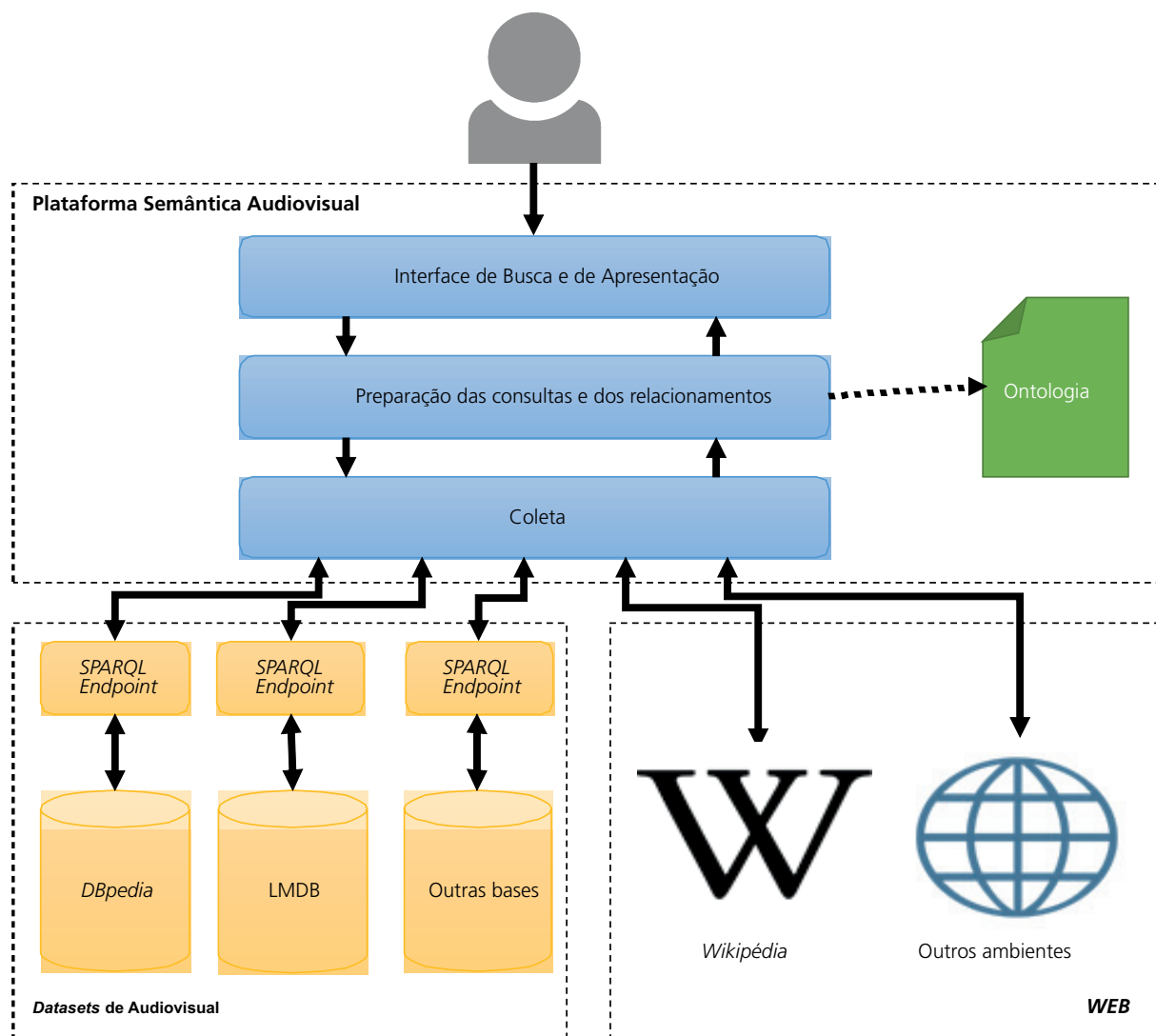
No campo dos audiovisuais, há algumas bases de domínio geral e outras bases específicas que tratam de dados da temática. A *DBpedia* e o *LinkedMDB* são as duas bases mais destacadas e completas que contemplam dados sobre audiovisuais e, mais especificamente, sobre filmes.

Com o uso principalmente dessas duas bases, é possível extrair uma quantidade grande de informação sobre filmes, especialmente quando elas estão vinculadas a outras fontes mais específicas. A partir desse cenário, refletiu-se acerca da construção de uma plataforma que reunisse dados de *Linked Data* sobre filmes, de forma com que os usuários fossem capazes de explorar as estruturas desses dados, em uma linguagem natural. Vale destacar que o *Linked Data* tem como estrutura fornecer os dados para a leitura das máquinas, de modo que essa plataforma tem o intuito de ser uma interface entre os dados estruturados para a leitura dos agentes computacionais e os usuários.

Diante do exposto, construiu-se a arquitetura do Portal Semântico Audiovisual, que utiliza como base de dados os *datasets* do *Linked Data*, além de outras fontes de informações da *Web*, como a *Wikipedia* e a *IMDb*, tendo como princípio o uso das tecnologias da *Web Semântica*, como o SPARQL, para a realização das consultas e ontologias em OWL, a fim de enriquecer os processos de consulta e a realização de relacionamentos. A Figura 1 apresenta a arquitetura da Portal Semântico Audiovisual.

Por meio da Figura 1, é possível identificar que a arquitetura é composta essencialmente de duas partes: a primeira, correspondente à composição estrutural da plataforma; e a segunda, das fontes que serão consultadas para apresentar as informações aos usuários.

Vale destacar que um protótipo da arquitetura já foi desenvolvido, faltando apenas aprimorar a relação entre a preparação da consulta e os relacionamentos com as ontologias. O protótipo visa demonstrar a viabilidade da arquitetura, apresentando o funcionamento bem como o relacionamento com as diversas fontes informacionais.



**Figura 1.** Arquitetura do Portal Semântico Audiovisual.

Fonte: Elaborado pelos autores (2018).

No que tange à plataforma, há uma interface em que o usuário irá realizar a sua busca e receber os resultados. Há uma segunda camada em que serão preparadas as consultas que serão feitas nas fontes e os consequentes relacionamentos existentes. Nessa camada poderá haver ontologias que fornecerão maior subsídio quanto à semântica formal dos dados, para preparar as consultas com um nível de relacionamento mais elevado. Por fim, a terceira camada compõe a coleta dos dados, ou seja, será nessa camada que haverá a busca e a consulta das fontes de informação.

A camada das fontes de informação é composta por dois elementos: o primeiro são os *datasets DBpedia* e o *LinkedMDB* do *Linked Data*, além de outras bases eventuais que podem ser utilizadas para coletar dados, enquanto o segundo são informações espalhadas na *Web*, em outros formatos, como, por exemplo, informações da *Wikipedia*.

Posteriormente à consulta, o usuário tem acesso a uma interface que reúne as informações obtidas, de forma que estas estarão relacionadas, buscando tirar proveito das ligações existentes, bem como da semântica formal que os dados possuem.



A princípio, o projeto está focado em permitir uma forma de recuperar as informações e apresentá-las de forma semântica, utilizando ontologias e demais instrumentos para oferecer os dados e relacioná-los. Nesse contexto, o uso do protocolo SPARQL é necessário para criar as consultas que relacionam os dados e buscá-los nas bases de dados existentes.

Vale destacar que embora o SPARQL seja o principal instrumento para recuperar dados na *Web* de Dados, o seu uso é restrito aos profissionais que conhecem a linguagem ou aos agentes computacionais construídos que geram tais consultas automaticamente. Nesse sentido, a plataforma busca permitir aos usuários tirar proveito dos benefícios e das relações que as tecnologias da *Web* Semântica propiciam, como as ontologias e o SPARQL, de forma natural, fazendo uma busca e navegando pela interface gráfica.

Outro ponto do projeto está na forma dinâmica como as informações são apresentadas, uma vez que as consultas ocorrem em tempo de execução, para que os dados estejam o mais atualizados possível. Não se exclui a possibilidade de haver um armazenamento em cache, para melhorar o desempenho computacional, mas, enquanto proposta, a ideia é utilizar os dados mais atualizados possível e os principais *datasets*.

A partir do modelo discutido (Figura 1), iniciou-se o processo de implementação do projeto, que visa ser uma forma de apresentar abertamente informações acerca de audiovisuais, especialmente sobre filmes. A ideia de implementar esse projeto é torná-lo uma referência para a construção de aplicações que utilizam informações publicadas abertamente no *Linked Data*, focadas num público especializado na temática audiovisual, mas que não necessita ter conhecimentos específicos sobre as tecnologias da *Web* Semântica, como SPARQL.

Dessa forma, o desafio consiste em facilitar os processos relativos às consultas e ao uso das tecnologias da *Web* Semântica, em que de uma forma simples o usuário somente insira sua expressão de busca, no caso um filme, e o sistema irá realizar uma série de buscas em bases conectadas, além das inferências relacionadas às ontologias, e apresentar informações refinadas sobre o assunto.

Para isso, o projeto foi desenvolvido utilizando a linguagem de programação *Python*, com o auxílio do *framework Django*.

O primeiro passo foi definir as fontes utilizadas para a obtenção dos dados, tanto as fontes estruturadas do *Linked Data*, quanto as da *Web* como um todo. Dessa forma, definiu-se que as fontes de *Linked Data* seriam a *DBpedia* e o *LinkedMDB*, que apresentam dados gerais e específicos sobre filmes, como exposto anteriormente, enquanto a *Wikipedia* e o *IMDb* seriam as fontes da *Web* a serem consultadas para obter dados não disponíveis nas bases de *Linked Data*.

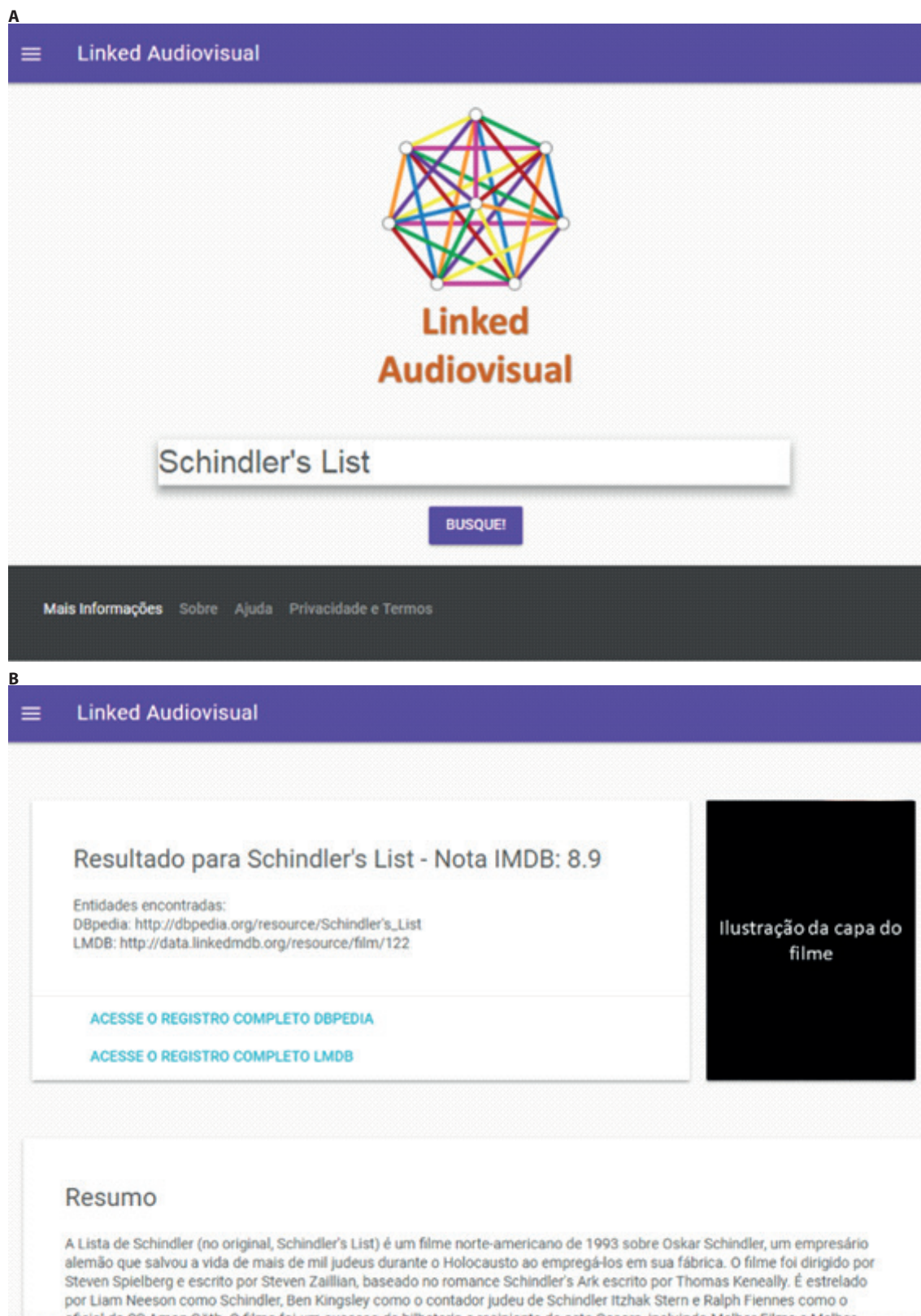
A *DBpedia* deveria fornecer informações mais gerais, relacionadas possivelmente com outras entidades, como o resumo do filme, os atores, as informações sobre eles e o *link* com a *Wikipedia*. Já o *LinkedMDB* forneceria informações mais técnicas dos filmes, como diretores, editores, tempo do filme, datas e o *link* com a página do *IMDb*.

A pesquisa nessas bases ocorre com o uso do SPARQL, em que as consultas são geradas dinamicamente de acordo com as estruturas de cada entidade, dado que em algumas entidades os dados estão em propriedades distintas e que algumas informações são apresentadas somente em algumas entidades, o que necessita de um tratamento na realização da consulta.

As ontologias podem ser utilizadas nesse momento, fornecendo elementos que contribuam para definir inferências com os dados coletados, fornecendo o contexto e significado das propriedades e dos dados. Além disso, com a evolução do projeto, procurar-se-á construir uma ontologia que reflita as buscas dos usuários e os dados que foram coletados nas diversas bases utilizadas.

Na sequência, após serem coletados os dados, um *script* de extração deverá ir à *Wikipedia* coletar dados não encontrados nas bases supracitadas. Por exemplo: na *Wikipedia* busca-se a imagem do cartaz do filme, e na





**Figura 2.** Telas da Implementação (A- Busca; B- Resultados).

Fonte: Elaborado pelos autores (2018).

IMDb busca-se a nota que o filme possui. Vale destacar que outras informações podem ser buscadas, tanto em outras bases de dados ligadas e nas próprias bases *DBpedia* e *LinkedMDB*, quanto na *Wikipedia*, no IMDb e em outros portais sobre filmes. No entanto, no escopo atual, o projeto limitou-se a essas informações, pois ele será incrementado de acordo com o *feedback* recebido dos usuários.

As duas telas construídas para o usuário interagir com o portal semântico estão apresentadas na Figura 2. A tela de busca é apresentada na Figura 2A, em que o usuário tem um campo para digitar o nome do filme cuja informação ele procura. Optou-se por construir uma tela simples, em que o usuário tem apenas a opção de realizar uma busca, pois o objetivo do portal é que ele busque os filmes e posteriormente navegue e explore as relações existentes. Vale destacar que, além de ser a página em que usuário realiza a busca, essa tela também é a página inicial do portal, onde são apresentados *links* para outras páginas com informações gerais do projeto, como os objetivos, os pesquisadores e as publicações oriundas dessa proposta.

Após o usuário realizar a busca, o portal o direciona a uma tela semelhante à representada na Figura 2B, em que são apresentadas as informações recuperadas dos *datasets* do *Linked Data* e de outras fontes de informação. Nessa tela, são apresentadas aos usuários as principais informações localizadas, além das relações que foram traçadas a partir das bases de dados. Os usuários poderão, assim, ter uma visão diferenciada sobre uma película, com informações e com relações que em geral não são encontradas em outros ambientes informacionais digitais.

Outro destaque é a possibilidade de se acessarem as fontes dos dados que foram a base das informações apresentadas. Nesse sentido, o usuário poderá acessar os recursos da *DBpedia*, do *LinkedMDB* e da *Wikipedia*, por exemplo, para explorar outras informações de seu interesse.

Descrita a criação do modelo e o desenvolvimento do protótipo do Portal Semântico, apresenta-se a seguir um caso em que são representadas as informações audiovisuais de um filme, apontando-se como estas, oriundas de diversas fontes, são relacionadas e apresentadas aos usuários.

Para demonstrar a viabilidade do Portal Semântico Audiovisual foi representado um exemplo de como as informações e as ligações são estruturadas pelo sistema quando um usuário o utiliza. Neste exemplo, as informações são referentes ao filme *Schindler's List* (em português "A Lista de Schindler"), escolhido por ser uma obra premiada e que contém um conjunto bem completo de informações nos *datasets* do *Linked Data*. Assim, o exemplo apresentado contempla dados do *LinkedMDB*, da *DBpedia*, da *Wikipedia* e do IMDb, sendo possível demonstrar as relações que podem ser construídas com esses dados.

A Figura 3 apresenta a representação desse exemplo em formato de grafos, em que os dados estão vinculados, a princípio, com o Portal Semântico Audiovisual, por ser a base que utiliza esses dados no princípio.

Na Figura 3 são apresentadas diversas informações relacionadas ao filme em questão, obtidas em distintas fontes. No centro, estão interligadas as duas entidades principais das quais foram extraídas informações, o recurso que representa o filme da *LinkedMDB* e da *DBpedia*. Além disso, estão vinculadas informações da *Wikipedia*, de onde foi extraída uma foto do filme, e do IMDb, em que se obteve a nota que o filme possui nesse portal.

No que tange à entidade da *DBpedia*, foram extraídas informações mais gerais do filme, como os atores que o estrelaram e um resumo. Já sobre a entidade da *LinkedMDB*, foram obtidas informações mais técnicas, como gênero, data e tempo de duração.

A partir da informação sobre o gênero do filme, foi possível recuperar outros filmes do mesmo gênero, o que permite apresentar ao usuário, enquanto navega pelo Portal Semântico Audiovisual, outros filmes relacionados àquele buscado.

O exemplo acima apresentado demonstra como o sistema organiza as informações das diversas fontes e as apresenta ao usuário. No exemplo, o Portal apresenta algumas das possibilidades, podendo ser expandido com a inserção de novas bases de dados e com o uso de ontologias que favorecem a realização de inferências e a

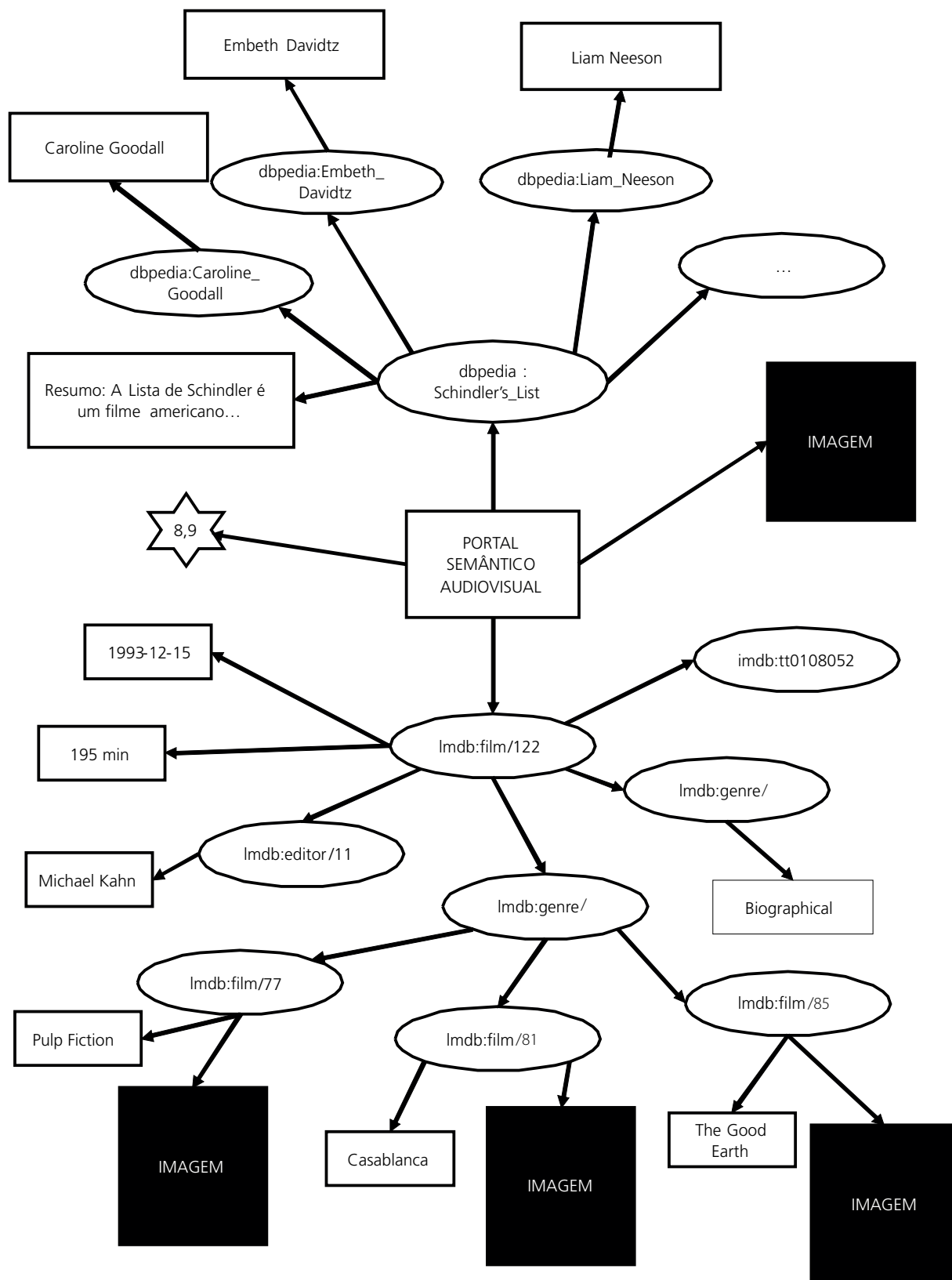


Figura 3. Exemplo das informações apresentadas pelo Portal Semântico Audiovisual.

Fonte: Elaborado pelos autores (2018).

construção de novos conhecimentos. Nesse momento, é importante ressaltar que o Portal apresenta uma estrutura semântica que poderá ser revisitada pelos catálogos *online* disponíveis atualmente nas bibliotecas, o que torna mais viável e real a possibilidade de sustentabilidade e modelagem de dados para todos os profissionais de informação.

## Considerações Finais

As tecnologias da *Web Semântica* estão se tornando cada vez mais presentes nas novas aplicações e em sistemas que buscam realizar inferências e aumentar o nível de semântica formal desses ambientes informacionais. Nessa seara, tais aplicações conseguem aprimorar os processos de recuperação da informação e, conseqüentemente, atender mais eficientemente às necessidades informacionais dos usuários.

Há exemplos do uso das tecnologias da *Web Semântica* em diversos domínios no âmbito dos repositórios digitais, como em mecanismos de buscas, em aplicações de natureza acadêmica e em redes sociais. O presente trabalho explorou um outro domínio que abarca uma grande quantidade de informações e especificidades: o dos recursos audiovisuais.

Assim, esta pesquisa propôs uma arquitetura de um portal audiovisual, em que as tecnologias da *Web Semântica*, em adição ao uso de bases de dados de *Linked Data*, possibilitam um ambiente digital de busca por filmes, no qual é realizada uma série de consultas e relações, a fim de apresentar ao usuário informações interligadas. O portal é capaz também de realizar inferências para apresentar conteúdos que possam ser de interesse aos usuários, o que é possível a partir do uso de ontologias e das bases de *Linked Data* que apresentam um alto nível de semântica formal.

Para demonstrar a validade dessa arquitetura, realizou-se a implementação do Portal Semântico Audiovisual utilizando linguagem de programação, bem como tecnologias da *Web Semântica*, especialmente SPARQL e RDF. Por meio dessa implementação, foi possível verificar que as bases de dados de *Linked Data* utilizadas (*DBpedia* e *LinkedMDB*) e outras fontes (*Wikipedia* e *IMDb*), quando relacionadas, são capazes de fornecer informações interessantes, que não são encontradas com facilidade na *Web*. Além disso, as tecnologias da *Web Semântica* permitem a realização dos relacionamentos e das inferências, o que facilita não só a recuperação da informação, como também a sua apresentação aos usuários.

Por fim, demonstrou-se um caso que simula como o Portal Semântico Audiovisual relaciona as informações e as apresenta aos usuários. No exemplo, é possível visualizar como as diversas bases de dados são relacionadas e apresentadas aos usuários. Além disso, é possível visualizar alguns pontos em que o sistema pode favorecer a realização de inferências, como a apresentação de filmes relacionados àquele buscado originalmente.

Portanto, a criação da arquitetura e o desenvolvimento do Portal Semântico Audiovisual demonstram como um contexto específico, no caso o recurso fílmico, é favorecido com o uso das tecnologias da *Web Semântica* e das bases de dados do *Linked Data*. Destaca-se que este trabalho se mostra como mais uma iniciativa que promove a materialização da *Web Semântica*, ao mesmo tempo que contribui para o domínio do audiovisual.

Como trabalho futuro, busca-se aprimorar o Portal Semântico Audiovisual inserindo outras funcionalidades e tornando este projeto acessível na *Web*, como também implementar uma ontologia que reflita as buscas dos usuários e os dados que foram coletados nas bases elencadas durante a pesquisa.

## Colaboradores

Todos os autores contribuíram igualmente para a concepção e delineamento deste estudo e para a análise dos dados, redação do manuscrito e revisão da versão final.

## Referências

- Araújo, W.T. Uso da informação audiovisual em bibliotecas: dados de pesquisas. *Informação & Sociedade*, v.2, n.1, 1992. Disponível em: <http://www.brapci.inf.br/index.php/article/download/13963>. Acesso em: 23 mar. 2018.
- Berners-Lee, T. *Linked data principles*. London: W3C, 2006. Available from: <https://www.w3.org/DesignIssues/LinkedData.html>. Cited: Mar. 23, 2018.
- Berners-Lee, T.; Hendler, J.; Lassila, O. The Semantic Web. *Scientific American*, v.284, n.5, p.28-37, 2001.
- Ducharme, B. *Learning SPARQL: Querying and updating with SPARQL 1.1*. Washington: O'Reilly Media, 2013. p.19.
- Ferreira, J.A.; Santos, P.L.V.A.C. O modelo de dados Resource Description Framework (RDF) e o seu papel na descrição de recursos. *Informação & Sociedade*, v.23, n.2, p.13-23, 2013. Disponível em: <http://www.ies.ufpb.br/ojs/index.php/ies/article/view/15436/9681>. Acesso em: 23 mar. 2018.
- Goliot-Lété, A.; Vanoye, F. *Ensaio sobre a análise filmica*. Campinas: Papyrus, 1994. p.12.
- Heath, T.; Bizer, C. Linked data: Evolving the Web into a global data space. *Synthesis Lectures on the Semantic Web: Theory and Technology*, v.1, n.1, p.1-136, 2011.
- Internet Movie Database. *General information of IMDb*. Grand Junction: IMDb, 2017. Available from: [https://help.imdb.com/article/imdb/general-information/what-is-imdb/G836CY29Z4SGN\\_MK5?ref\\_=helpart\\_nav\\_1#](https://help.imdb.com/article/imdb/general-information/what-is-imdb/G836CY29Z4SGN_MK5?ref_=helpart_nav_1#). Cited: Mar. 23, 2018.
- Machado, A. *Máquina e imaginário: o desafio das poéticas tecnológicas*. São Paulo: Edusp, 2001. p.48.
- Martin, M. *A linguagem cinematográfica*. São Paulo: Brasiliense, 2003.
- McCarthy, C.M.; Targino, M.G. Materiais audiovisuais na sociedade e nas bibliotecas brasileiras. *Revista da Escola de Biblioteconomia da UFMG*, v.13, n.2, p.302-321, 1984.
- Moran, J.M.; Masetto, M.T.; Behrens, M.A. *Novas tecnologias e mediação pedagógica*. Campinas: Papyrus, 2000.
- Parente, A.; Carvalho, V. Entre cinema e arte contemporânea. *Galáxia: Revista do Programa de Pós-Graduação em Comunicação e Semiótica*, n.17, 2009. Disponível em: <https://revistas.pucsp.br/index.php/galaxia/article/view/2093>. Acesso em: 23 mar. 2018.
- Santarem Segundo, J.E.; Coneglian, C.S. Tecnologias da Web Semântica aplicadas a organização do conhecimento: padrão SKOS para construção e uso de vocabulários controlados descentralizados. *Organização do Conhecimento e Diversidade Cultural*, v.3, p.224-233, 2015. Disponível em: <http://isko-brasil.org.br/wp-content/uploads/2015/09/Organiza%C3%A7%C3%A3o-do-Conhecimento-e-Diversidade-Cultural-ISKO-BRASIL-2015.pdf>. Acesso em: 23 mar. 2018.
- Santarem Segundo, J.E.; Coneglian, C.S. Web semântica e ontologias: um estudo sobre construção de axiomas e uso de inferências. *Informação & Informação*, v.21, n.2, p.217-244, 2016. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/viewFile/26417/20131>. Acesso em: 23 mar. 2018.
- Shadbolt, N.; Berners-Lee, T.; Hall, W. The Semantic Web revisited. *IEEE Intelligent Systems*, v.21, n.3, p.96-101, 2006. Available from: [https://eprints.soton.ac.uk/262614/2/OLD\\_Semantic\\_Web\\_Revisited.pdf](https://eprints.soton.ac.uk/262614/2/OLD_Semantic_Web_Revisited.pdf). Cited: Mar. 23, 2018.
- Simionato, A.C. O tratamento descritivo para recursos audiovisuais: a representação de filmes. In: Albuquerque, A.C.; Simionato, A.C. (Org.). *Recursos audiovisuais: sua contemporaneidade na organização e representação da informação e do conhecimento*. Rio de Janeiro: Interciência, 2017. v.1, p.77-94.
- Simionato, A.C. et al. Audiovisuais e Linked data: um estudo das bases DBpedia e LMDb. *Em Questão*, v.24, n.3, 2018. Disponível em: <https://seer.ufrgs.br/EmQuestao/article/view/78206>. Acesso em: 6 out. 2018.