

Audiovisuais e Linked data: um estudo das bases DBpedia e LMDB

Ana Carolina Simionato

Doutora; Universidade Federal de São Carlos, São Carlos, SP, Brasil;
acsimionato@ufscar.br

Caio Saraiva Coneglian

Doutorando; Universidade Estadual Paulista, Marília, SP, Brasil;
caio.coneglian@gmail.com

Paula Regina Ventura Amorim Gonçalves

Doutora; Universidade Federal do Espírito Santo, Vitória, ES, Brasil;
paulaventuramorim@gmail.com.

José Eduardo Santarem Segundo

Doutor; Universidade Estadual Paulista, Marília, SP; Universidade de São Paulo, Ribeirão Preto, SP,
Brasil;
santarem@usp.br

Resumo: Como proponente da *Web Semântica* e dos princípios *Linked Data*, a iniciativa *Linking Open Data* oferece uma enorme proporção de dados de audiovisuais que podem auxiliar nas buscas e na recuperação de informações mais precisas. Nesse cenário, o objetivo é explorar os possíveis relacionamentos de bases de dados de audiovisuais no *Linking Open Data*, no intuito de apresentar o potencial dessa iniciativa para usuários que buscam fontes de informação detalhadas sobre os audiovisuais. Utilizou-se uma pesquisa de natureza qualitativa, com caráter exploratório e aplicado, baseada na literatura científica dos temas *Linked Data*, *Web Semântica* e audiovisual. Posteriormente, foram consultadas as bases *DBpedia* e *LMDB* com o uso do protocolo *SPARQL*. Considera-se que os conjuntos de dados disponíveis no *Linking Open Data* não só podem auxiliar na ligação entre informações sobre recursos audiovisuais, como também, podem ser fonte para a construção de catálogos mais dinâmicos, reduzindo o retrabalho durante o processo de descrição de recursos informacionais.

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

1 Introdução

O exponencial crescimento de dados pode ser atribuído, principalmente, à

popularização da *Web*. Uma consequência deste contexto está no valor que os dados adquiriram, tornando-se capital para todas as áreas do conhecimento. Davenport, Barth e Bean (2012) afirmam que uma das habilidades mais desejadas na última década é saber manusear grandes quantidades de dados originários de diversas localidades e em diferentes formatos.

Dentre as novas perspectivas para semântica dos recursos disponibilizados na *Web*, destacam-se os princípios *Linked Data* propostos por Berners-Lee (2006) que não só oferecem orientações à criação de relacionamentos entre coleções de dados conectados na *Web*, como também, possibilitam novas e mais específicas formas aos usuários, com o intuito de multidimensionar suas escolhas e agilizar suas opções, atreladas às tecnologias da *Web Semântica*.

Um dos principais difusores do *Linked Data* é o projeto *Linking Open Data*, criado em 2007, gerenciado por Cyganiak e Jentzsch e mantido pelo *World Wide Web Consortium (W3C)*. O objetivo do *Linking Open Data* é evidenciar a prática dos princípios *Linked Data* e do formato aberto (*Open Data*) para os conjuntos de dados (*datasets*), utilizando o modelo de dados da *Web Semântica*, *Resource Description Framework (RDF)*. A iniciativa *Linking Open Data* está disponível no portal lod-cloud.net.

O *Linking Open Data* é apresentado em forma de diagrama e entre suas ligações está a divisão de suas temáticas, categorizado por dados de publicação, ciências da vida, domínio geral (*cross-domain*), geográficos, governamentais, mídia, uso geral, redes sociais e linguística. O principal requisito para pertencer a nuvem do *Linking Open Data* é que os dados devem estar abertos e disponibilizados para download (ABELE; MCCRAE, 2017).

Neste cenário, destaca-se a importância do *Linking Open Data* ao disponibilizar fontes de informações completas e com alta qualidade, visto que, em determinados campos de mercado e da pesquisa, a atividade de realizar buscas em específicas temáticas pode ser considerada como uma tarefa desafiadora.

Especificamente na temática mídia, há uma enorme proporção de dados de audiovisuais no *Linking Open Data* que potencialmente pode corroborar nas

buscas e na compreensão dos usuários em diferentes contextos sociais. Grisoto (2016) ressalta que as possibilidades de processar e interpretar o conteúdo para uma recuperação de informações torna-se mais eficiente devido a utilização dos dados ligados relativos aos audiovisuais.

Como base de dados de audiovisuais, a *Linked Movie Data Base* (LMDB) destaca-se pela especificidade de dados relacionados a filmes, com informações sobre os filmes, atores, diretores, personagens, trilhas sonoras, compositores, entre outras; capaz de fornecer informações estruturadas em RDF, que auxilia quaisquer aplicações a construir e relacionar dados sobre essa temática. Outra base de dados é a *DBpedia* que, diferentemente da LMDB, não contém somente informações de audiovisuais, sendo uma base de domínio geral, com dados de autoridade, filmes, locais, datas, programas de televisão, entre diversas outras temáticas.

Entretanto, como percebido no uso do *Linking Open Data*, muitas bases de dados ainda necessitam de um suporte para sua apresentação com caráter mais dinâmico. Assim, questiona-se como os princípios *Linked Data* e as tecnologias da *Web Semântica* podem propiciar informações precisas aos usuários que não conseguem obter essas informações em buscadores de uso geral? Objetiva-se explorar os possíveis relacionamentos de conteúdos audiovisuais nas bases de dados LMDB e *DBpedia*, com o intuito de apresentar consultas e cruzamentos entre as bases e apresentar resoluções para a problemática incitada.

2 Recursos audiovisuais, *Linked Data* e *Web Semântica*

Desde a sua origem em 1896 e a sua institucionalização em 1908, o cinema foi desenvolvido, a partir dos seus suportes físicos, conjunto à formação de seu espaço cultural (AUMONT, 2012). Adicionalmente, à categoria de filmes, encontramos os gêneros referentes à produção cinematográfica. Sendo que os principais gêneros cinematográficos originaram-se:

[...] meses depois das projeções dos irmãos Lumière, o ex-ilusionista Georges Méliès (1861-1938) criava filmes de fantasia,

terror e ficção científica. O documentário, naturalmente, existiu desde o começo, pois muitos dos primeiros cineastas simplesmente apontavam as câmeras para o mundo que os cercava. A comédia veio logo em seguida, junto com os dramas de época, romances, filmes de ação, drama psicológico, filmes de guerra, farsa, épicos da antiguidade e até mesmo pornografia. (FRAYLING; KEMP, 2011, p. 9).

Os gêneros cinematográficos também podem ser classificados conforme o seu tema principal ou mesmo pela combinação de vários temas, que originam os subgêneros. Essa categorização não é o bastante no processo de representação nos centros informacionais. O audiovisual é parte de um contexto, sendo um recurso com propriedades específicas e, ao mesmo tempo, plurais em relação aos outros recursos tradicionais, como por exemplo, o textual. Além disso, suas propriedades são próximas dos recursos sonoros e imagéticos. (BUARQUE, 2008; SIMIONATO, 2017).

Por essa razão, os dados disponibilizados nos catálogos ou *sites* temáticos são uma pequena parte do que a produção cinematográfica gera. Leone e Mourão (1987) destacam que durante a criação do audiovisual, a confecção envolve a articulação entre as etapas de encenação da peça à organização da montagem e, de acordo com a narrativa, há o roteiro conjunto à temporalidade de cada cena (*shot*). Da mesma forma, no âmbito da Web Semântica e dos princípios *Linked data*, os dados dos recursos audiovisuais podem aprimorar e oferecer formas mais avançadas de uso e reuso dos recursos audiovisuais.

Nesse sentido, as organizações, instituições e comunidades estão identificando a necessidade de compreender o ambiente em que tanto os dados, quanto os usuários se encontram. Rasmussen-Pennington (2016) afirma que a aproximação do *Linked Data* e da *Web Semântica* aos centros informacionais configura-se como uma maior compressão das máquinas para o conteúdo dos recursos da *Web*. Tal processo está ligado com a chamada ‘materialização’ da *Web Semântica* (SANTAREM SEGUNDO; CONEGLIAN, 2016), momento em que iniciativas passam a aplicar os conceitos e as tecnologias dessa proposta na prática, que tem como principal expoente o *Linked Data*.

O cenário apresentado aponta para um futuro em que as aplicações deverão estar integradas aos principais serviços da *Web*, obrigando os bibliotecários, os desenvolvedores e as organizações a pensarem na aplicação das tecnologias semânticas para disponibilizar seu conteúdo.

Neste sentido, Matthews insere as tecnologias e os padrões da *Web Semântica* como um meio para:

[...] proporcionar soluções neutras para fornecedores [...] (permitindo dados estruturados e semi-estruturados, descrições formais e informais e uma arquitetura aberta e extensível) e ajuda a suportar soluções descentralizadas [...] (MATTHEWS, 2005, p. 8, tradução nossa).

Desta forma, a *Web Semântica*, por meio de suas tecnologias, cria uma estrutura capaz de auxiliar na publicação e na disponibilização de dados, seguindo princípios que favoreçam o reuso, o compartilhamento, a organização e a descrição das informações.

Para a disponibilização do conteúdo, diferentes domínios são estruturados em conjuntos de dados, enfatiza-se ainda a importância de tornar a *Web Semântica* real e materializada, como uma estrutura capaz de ser utilizada pelas mais diversas aplicações da *Web*. Segundo Berners-Lee, “A *Web Semântica* não é apenas colocar dados na web. Trata-se de fazer *links*, para que uma pessoa ou máquina possa explorar a *Web* de dados.” (BERNERS-LEE, 2006, tradução nossa, doc. não paginado).

Por esse viés, os princípios *Linked Data* foram criados e são definidos em: (1) usar URIs (*Uniform Resource Identifier*) como nomes para os itens; (2) usar URIs HTTP para que as pessoas possam consultar esses nomes; (3) quando alguém consulta uma URI, deve provir de uma informação RDF útil; (4) incluir sentenças em RDF com links para outras URIs, a fim de permitir que itens relacionados possam ser descobertos.

Após uma década da concepção dos princípios, é possível verificar o crescimento e a importância dos princípios em permitir a criação e a ligação de

dados, no intuito de facilitar o acesso, o uso e o reuso à humanos e não-humanos, Arakaki destaca que os princípios são “[...] melhores práticas para estruturar e ligar dados. Esse processo facilita a busca de agentes humanos e não humanos e os direcionam em diferentes bases a partir desses dados ligados.” (ARAKAKI, 2016, p. 27).

Os princípios *Linked data* oferecem às instituições culturais formas mais semânticas e capazes de se interoperar juntamente com as melhores ferramentas de gerenciamento de dados, aumentando assim, a quantidade e a qualidade da informação (GODDARD; BYRNE, 2010). Bem como, a “[...] possibilidade de otimizar a recuperação e navegação dos dados de acordo com as necessidades informacionais de seus usuários.” (SIMIONATO; ARAKAKI; SANTOS, 2017, p. 458).

Dentre as características das instituições culturais, no âmbito dos recursos audiovisuais, destaca-se a custosa tarefa dos usuários de reunir e relacionar dados deste domínio no âmbito da *Web* tradicional e, com grande dificuldade, localizar determinadas informações e compilar os dados necessários para uma compreensão clara do cenário que se deseja investigar.

Nesse sentido, algumas tecnologias da *Web Semântica* foram construídas no intuito de representar e de recuperar os dados estruturados. As URIs, no contexto *Linked data*, têm uma função de desambiguação, uma vez que são usadas para identificar recursos de forma exclusiva. O risco relacionado a este elemento é a falta de estabilidade dos URLs, o que pode determinar perda de validade, portanto, uma conexão perdida. (VAN HOOLAND; VERBORGH, 2014). Além das URIs, duas tecnologias apresentam maior destaque para essa pesquisa, o RDF e o SPARQL.

O RDF é o modelo de descrição de recursos dentro da *Web*, por uma estrutura de triplas de associação em que um sujeito está ligado a um objeto por meio de um predicado. Ferreira e Santos (2013, p. 21) complementam tal questão afirmando que:

[...] o modelo RDF oferece a possibilidade para as comunidades de descrição de recursos definirem a semântica de seus metadados de maneira formal, isto é, definindo o significado dos elementos de metadados, conforme as suas necessidades específicas de descrição, em um modelo processável por máquinas.

Enquanto o RDF está focado na representação e na descrição dos dados, o SPARQL tem suas funções relativas à recuperação destes dados. Para realizar tal tarefa, o SPARQL foi construído utilizando o princípio do RDF, estruturando as consultas nas chamadas triplas RDF (DUCHARME, 2013). Santarém Segundo (2014, p. 3870) afirma que: “[...] o SPARQL é um conjunto de especificações que fornecem linguagens e protocolos para consultar e manipular o conteúdo publicado em RDF na Web.”.

As duas tecnologias apresentadas estão fortemente atreladas ao *Linked Data*, uma vez que o RDF é a tecnologia utilizada nesta proposta para a estruturação dos dados, enquanto o SPARQL é a principal forma de se acessar os dados das bases *Linking Open Data*. Desta maneira, para se explorar bases construídas seguindo tais princípios, é necessário o conhecimento acerca de ambas as tecnologias, o que permite que sejam identificadas e traçadas as relações entre os dados estruturados.

3 Procedimentos metodológicos

A identificação das aproximações interdisciplinares entre a *Web Semântica*, mais especificamente, entre *Linked Data* e audiovisual pode ser realizada a partir de diversas perspectivas. Uma forma é apresentar como os princípios e o audiovisual podem dialogar, demonstrando e explorando algumas possibilidades dos conjuntos desses tipos de dados.

Para atingir os objetivos desta pesquisa, utilizou-se uma pesquisa de natureza qualitativa com caráter exploratório, uma vez que se buscou a literatura de *Web Semântica* e audiovisual como embasamento teórico. Para construção de uma pesquisa aplicada, analisaram-se as possibilidades de extração de

conhecimento em bases de dados *Linking Open Data* focados nos materiais audiovisuais.

Primeiramente, foram definidas as bases de dados para centrar a pesquisa. Dentro do domínio do *Linking Open Data*, há diversos *datasets* que contém dados de conteúdos audiovisuais, dentre eles destaca-se, a *Linked Movie Database* (LMDB) e a *DBpedia*. Essas duas bases apresentam diferenças quanto ao domínio que cada um atinge, sendo que a LMDB apresenta, em suma, os dados de filmes, enquanto a *DBpedia* contém dados de domínio geral, que abrange as mais diversas áreas existentes.

Ao analisar a *DBpedia*, verifica-se que o conteúdo audiovisual desta base mostra-se bastante amplo, devido aos inúmeros relacionamentos existentes dentro desse conjunto de dados. Devido a essa característica, há relações entre diferentes elementos, como filmes, atores e obras literárias. Por essa razão, a *DBpedia* torna-se uma importante fonte informacional de audiovisual, que possibilita uma exploração pertinente sobre as relações existentes no contexto do audiovisual a partir do uso dos recursos *Linked Data*.

Diante do exposto, o processo de exploração ocorreu nas duas bases mencionadas, LMDB e *DBpedia*. Tal processo não objetivou realizar uma comparação, mas sim, traçar um panorama dos relacionamentos existentes. Essas bases foram escolhidas para demonstrar que as informações utilizadas na pesquisa estão disponíveis na rede, tanto em formatos preferenciais para a leitura humana, como a *Wikipédia* e a *IMDb*, quanto em bases de dados estruturados, *DBpedia* e LMDB, que contém dados estruturados em RDF de conteúdo similar aos dois serviços citados anteriormente.

Para ser possível explorar as bases de dados citadas, identificando as possibilidades que *Linking Open Data* fornece no tratamento dos materiais audiovisuais, utilizou-se o SPARQL, para realizar consultas para apontar as relações que o conjunto de dados permitia. A partir dos resultados encontrados, utilizou-se da literatura e das relações realizadas para identificar e traçar como *Linked Data* pode contribuir para pesquisas e para o usuário interessado em explorar informações sobre audiovisuais.

4 Resultados e discussões

A partir dos métodos selecionados, foram realizadas as análises sobre as bases de dados da LMDB e da *DBpedia*, no intuito de estabelecer conexões entre *Linked Data* e recursos audiovisuais.

Explicita-se que as duas bases permitem o acesso aos seus dados pelo SPARQL *Endpoint*, que consiste em um *web service* em que os usuários podem interagir e, utilizando-se do protocolo SPARQL, podem recuperar os dados oriundos das bases mencionadas. Este ambiente fornece uma interface em que o usuário insere uma consulta SPARQL, retornando na sequência os dados que atenderam às regras definidas.

Dessa forma, buscou-se apresentar combinações de dados que não estão disponíveis facilmente na *Web*. Isto é, foram realizadas associações entre diferentes informações que, caso fossem realizadas por um indivíduo em páginas *Web*, teria um alto nível de complexidade e demandaria pesquisas em diferentes *websites* e ambientes para serem construídas.

A questão levantada vai ao encontro da proposta da *Web Semântica* de Berners-Lee, Hendler e Lassila (2001), quando os autores indicam que os agentes computacionais poderiam auxiliar os usuários a realizarem cruzamentos e relações entre as informações, sem a necessidade de o usuário navegar por diversas páginas fazendo a coleta de cada um dos dados. As consultas construídas permitem que os dados sejam coletados em um único ambiente, sem que um usuário navegue por inúmeras plataformas coletando e cruzando cada uma das informações.

Para demonstrar a riqueza de informações audiovisuais no *Linking Open Data*, o primeiro *dataset* analisado foi a LMDB. Nessa base, há uma predominância de informações sobre as películas, como informações acerca do elenco, do tempo de duração, dos produtores e dos diretores.

Um exemplo da granularidade e do tipo de dados que são apresentados nessa base de dados pode ser visto quando se busca consultar e quantificar os gêneros dos filmes que um determinado ator fez em sua carreira. Nesse caso,

realizou-se uma consulta com a contagem dos gêneros dos filmes que o ator *Ian McKellen* participou. O resultado obtido pode ser observado na figura 1.

Figura 1 - Contagem do gênero dos filmes estrelados por um ator

	Nome Gênero	Contagem
<pre> SELECT ?nome_genero (count(*) as ?contagem) WHERE { ?ator movie:actor_name "Ian McKellen". ?filmes movie:actor ?ator; movie:genre ?genero; rdfs:label ?nome_filme. ?genero movie:film_genre_name ?nome_genero } GROUP BY (?nome_genero) ORDER BY DESC (?contagem) </pre>	"Action"	1
	"Adventure"	6
	"Biographical"	1
	"Drama"	6
	"Mystery"	1
	"Superhero"	1

Fonte: Elaborado pelos autores.

A partir da consulta demonstrada na figura 1, percebe-se a existência de diversas relações que são traçadas, buscando agrupar informações. Nesse sentido, a consulta é construída partindo da identificação de uma entidade do *dataset* que possui o nome *Ian McKellen*, realizando uma vinculação a todos os filmes que este ator estrelou; além da recuperação dos gêneros destes filmes. Com tais informações, realiza-se o agrupamento dos dados pelo nome do gênero, apresentando uma contagem do total de filmes de cada um dos gêneros que o referido ator atuou.

Outro exemplo que expande as formas tradicionais de buscas e de recuperação da informação seria a realização de consultas dentro de um universo de personagens criados por um estúdio. A obtenção de informações desse tipo demonstra o nível de relações que podem ser extraídas do *dataset*, além de apresentar como as tecnologias da *Web Semântica* podem auxiliar na obtenção de dados do domínio audiovisual e permitir a construção de relacionamentos de forma automatizada para os pesquisadores e interessados nesta área.

Uma busca com o objetivo discutido permite, por exemplo, cruzar informações entre os atores e os seus respectivos personagens, buscando somente as atuações em filmes produzidos pelo estúdio escolhido. A figura 2 demonstra um exemplo que contempla não só a consulta realizada, bem como os resultados obtidos a partir dela.

Figura 2 - Consulta dos atores e dos personagens de um estúdio

	Nome Ator	Nome Personagem
SELECT ?nome_ator ?nome_personagem	"Bruce Campbell"	"Wrestling Announcer"
WHERE {	"Paul Nakauchi"	"Wong"
?estudio	"Steven Blum"	"Wolverine"
movie:production_company_name "Marvel Studios".	"Hugh Jackman"	"Wolverine"
?s movie:production_company ?estudio;	"Hugh Jackman"	"Wolverine"
dc:title ?nome_filme;	"Mickey Rourke"	"Whiplash"
movie:performance ?perf.	"Don Cheadle"	"War Machine"
?perf movie:performance_actor ?nome_ator;	"Don Cheadle"	"War Machine"
movie:performance_character	"Rodney Saulsberry"	"War Machine"
?nome_personagem.	"Shawn MacDonald"	"Vision"
}
ORDER BY DESC (?nome_personagem)		

Fonte: Elaborado pelos autores.

A consulta da figura 2 apresenta os dados recuperados sobre todos os filmes produzidos pelo estúdio *Marvel Studios*, sendo identificados as personagens e os atores desses filmes. Cabe como ressalva que parte das performances que foram recuperadas não continham a informação acerca do nome das personagens, tornando os dados incompletos.

As duas consultas realizadas na LMDB exploram alguns dos principais relacionamentos que podem ser realizados dentro desse conjunto de dados. As consultas podem ser expansivas com a localização dos produtores, os escritores e os diretores de um filme, além de permitir recuperar os filmes de uma determinada série, os locais de gravação e o país que produziu um filme.

Ao aprofundar a exploração realizada, efetuou-se consultas com o SPARQL dentro do conjunto de dados da *DBpedia*. Como relatado, a *DBpedia* apresenta dados de domínio geral o que permite explorar outros tipos de informações relacionadas aos filmes. Além disso, a *DBpedia* contém um grande número de classes e de entidades que permitem recuperar os dados, explorando uma quantidade maior de relações.

A primeira consulta realizada com essa base de dados buscou investigar algumas relações entre os conteúdos audiovisuais relacionados a filmes. Tal consulta utilizou de uma categoria da *DBpedia*, com o nome *Marvel Cinematic Universe films*, que contém relações com todos os filmes pertencentes a série de títulos produzidas pelo *Marvel Studios*, chamado de Universo Cinematográfico *Marvel*.

Neste sentido, a partir da identificação dessa classe, é possível obter todos os filmes que pertencem a citada série de títulos cinematográficos. Com a identificação de todos os filmes, foi possível constatar o músico responsável pela trilha sonora de cada película, sendo que, ao identificar esta informação, coletaram-se dados do nome e do nascimento deste músico. A consulta construída e os resultados obtidos estão apresentados na figura 3.

Figura 3 - Músicos responsáveis pela trilha sonora de uma série de filmes

Nome Filme	Nome Compositor	Sobrenome Compositor	Nascimento Compositor
"Avengers: Age of Ultron"@de	"Brian"@en	"Tyler"@en	"1972-05-08"^^xsd:date
"Iron Man (2008 film)"@en	"Ramin"@en	"Djawadi"@en	"1974-07-19"^^xsd:date
"Thor: Ragnarok"@en	"Mark"@en	"Mothersbaugh"@en	"1950-5-18"^^xsd:date
"Captain America: The First Avenger"@en	"Alan"@en	"Silvestri"@en	"1950-03-26"^^xsd:date
"Guardians of the Galaxy (film)"@en	"Tyler"@en	"Bates"@en	"1965-6-5"^^xsd:date
...

Fonte: Elaborado pelos autores.

A partir dos resultados, verifica-se a existência de diversas ligações entre os dados, possibilitando localizar informações específicas que não são encontradas em uma única página de forma organizada. O SPARQL permite que as relações existentes dentro do RDF sejam exploradas e explicitadas, possibilitado que o usuário possa obter dados acerca de informações audiovisuais, com os cruzamentos existentes entre estes.

No exemplo da figura 3, os usuários podem explorar informações de diversas mídias, e tal exemplo aponta as possibilidades que o projeto *Linking Open Data* permite. Nessa circunstância, há uma série de elementos que auxiliam aos pesquisadores sobre as possibilidades de conexão entre os processos de criação, de filmagem, consultas de trilhas sonoras, cenários, localizações, figurinos de época e personagens.

Vale destacar que os níveis de conexões existentes são inúmeros, havendo possibilidades de aprofundar consideravelmente as ligações realizadas, com a finalidade de tornar explícito um conjunto grande de relações. Ressalta-se que a complexidade do âmbito audiovisual demonstra, claramente, como o

SPARQL e o RDF podem trazer contribuições para este cenário, em que diversas particularidades podem se relacionar.

O processo de relacionar e de obter conhecimentos sobre os recursos audiovisuais é árduo, devido, sobretudo a tais características expostas. Assim, a natureza interoperável das bases de dados apresentadas, *DBpedia* e *LMDB*, devido, essencialmente, aos princípios da *Web Semântica*, *Linked Data* e *Linking Open Data*, possibilita a quaisquer projetos que visem reunir recursos audiovisuais sob uma perspectiva semântica utilizar os dados desses projetos.

Em um segundo momento, foi realizada uma integração entre a *DBpedia* e a *LMDB*, por meio das consultas SPARQL. Esse processo é chamado de consulta federada e busca relacionar os dados de diversos *datasets*. No âmbito dessa pesquisa, é apresentada como essa relação contribui para tornar a recuperação e o acesso aos recursos audiovisuais mais eficiente.

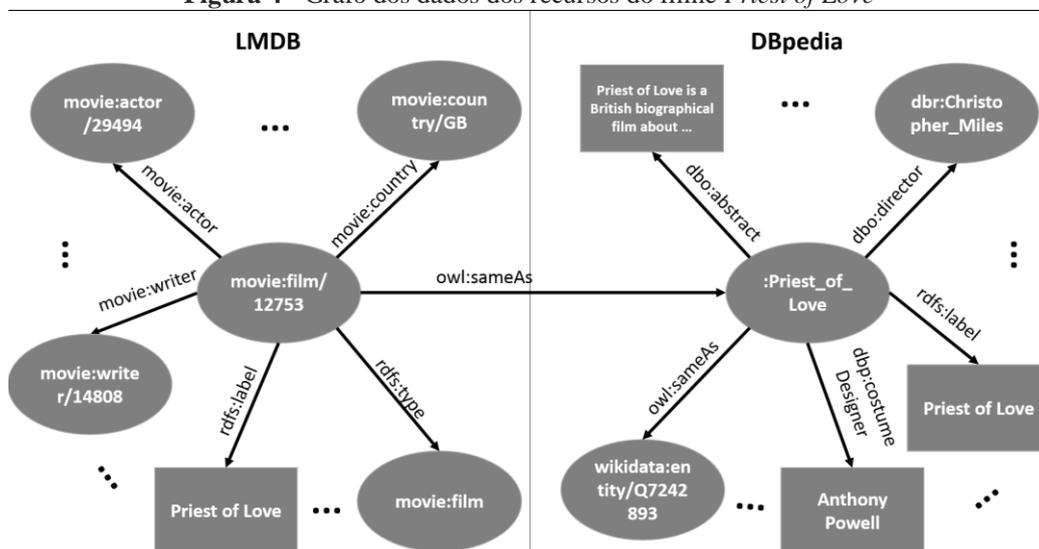
Assim, utilizando como base a *DBpedia* e a *LMDB*, construiu-se uma consulta federada que relacionava os dados entre esses dois *datasets*. Tal consulta demonstra como uma aplicação que deseja relacionar dados de recursos audiovisuais poderia conectar as informações destes *datasets*, fornecendo um meio de estabelecer vínculos e possibilidades para quaisquer ferramentas que busquem tornar a procura por dados de recursos audiovisuais mais semântica e inserida na seara do *Linked Data*.

A consulta federada foi feita a partir da análise dos dados de ambos os *datasets* e, também, reduzida a um recurso informacional específico. No caso, o recurso escolhido foi *Priest of Love*, em razão de um dos atuantes ser *Ian McKellen*, que foi utilizado como exemplo anteriormente. Desta forma, tanto a *LMDB* quanto a *DBpedia* contemplam as informações acerca deste recurso, existindo inclusive uma ligação entre o recurso da *LMDB* para com o recurso da *DBpedia*.

No intuito de demonstrar esta relação, bem como as próprias ligações e informações contidas nos recursos, a figura 4 demonstra uma parte dos dados apresentada em esquema de grafos. As reticências inseridas nos grafos

representam que há outras relações e recursos diversos, e que não foram inseridos nesta representação.

Figura 4 - Grafo dos dados dos recursos do filme *Priest of Love*



Fonte: Elaborado pelos autores.

Ao observar a figura 4, é possível visualizar que há uma relação entre os dois recursos que ocorre por meio da relação *owl:sameAs*, que tem a função de realizar ligações entre recursos que representam uma mesma informação. Outro ponto apresentado são as relações existentes entre os recursos e os dados, por exemplo, os atores, o diretor, o roteirista, o país de origem e o resumo do filme. Vale ainda destacar que há informações distintas em cada um dos recursos, o que justifica a realização de relacionamentos e ligações entre as bases por meio das consultas federadas.

Partindo dessa relação entre os dois *datasets*, foi realizada a consulta federada do filme de título *Priest of Love*, em que a ligação entre os *datasets* se dava pela relação *owl:sameAs*, conforme demonstrado na figura 4. Desse modo, a busca localizou algumas informações disponíveis em ambas as bases de dados, por meio de uma única consulta, como demonstrado pela figura 5.

Figura 5 - Consulta Federada LMDB e DBpedia

<pre> SELECT ?nome_filme ?resumo ?nome_diretor ?nascimento_diretor WHERE { ?filme rdfs:label "Priest of Love". ?filme owl:sameAs ?igual. SERVICE <http://dbpedia.org/sparql>{ ?igual <http://www.w3.org/2000/01/rdf-schema#label> ?nome_filme; <http://dbpedia.org/ontology/director> ?diretor; <http://dbpedia.org/ontology/abstract> ?resumo. ?diretor <http://dbpedia.org/ontology/birthDate> ?nascimento_diretor; <http://www.w3.org/2000/01/rdf-schema#label> ?nome_diretor. } } </pre>	Nome Filme	Resumo	Nome Diretor	Nascimento Diretor
	"Priest of Love"@en	"Sacerdote del amor (titulo original en inglés, Priest of Love) es una película británica basada en la vida del escritor ...	"Christopher Miles"@en	"1939-04-19"^^xsd:date
	"Priest of Love"@en	"Priest of Love est un film britannique réalisé par Christopher Miles et sorti en 1981. ...	"Christopher Miles"@en	"1939-04-19"^^xsd:date
	"Priest of Love"@en	"Priest of Love is a British biographical film about D. H. Lawrence and his wife Frieda (née Von Richthofen....	"Christopher Miles"@en	"1939-04-19"^^xsd:date
	"Priest of Love"@en	"Sacerdote del amor (titulo original en inglés, Priest of Love) es una película británica basada en la vida del escritor D. H. Lawrence. Estrenada el 11 ...	"Miles, Christopher"@en	"1939-04-19"^^xsd:date

Fonte: Elaborado pelos autores.

Os resultados obtidos com a consulta federada permitem identificar que as duas bases complementam-se ao fornecer informações acerca de recursos audiovisuais. Neste sentido, os *links* entre as bases, na maioria dos casos, já estão inseridos, permitindo que a realização de consultas federadas possam ligar as duas bases de dados.

Destaca-se que a base da LMDB contempla informações mais técnicas e com uma riqueza elevada no que se refere a dados sobre atores, diretores e produtores de uma película. Enquanto o *dataset* da DBpedia fornece dados que estão vinculados a estes indivíduos e a própria relação de cada filme com outros recursos, como livros, histórias em quadrinho e músicas. Essas bases ao se relacionarem por meio de consultas federadas são capazes de fornecer informações ricas sobre a conjuntura de um recurso audiovisual, que são dificilmente localizadas em outros ambientes da *Web*.

Outro ponto que merece destaque são as relações e as ligações que tanto a LMDB quanto a DBpedia fornecem a outras bases e plataformas da *Web*, como ao IMDb, *site* popular com informações sobre programas de televisão e filmes. Tais relacionamentos podem fornecer diversas possibilidades para que o usuário navegue por outros ambientes, encontrando informações mais específicas, quando se desenvolve um sistema baseado nas bases de dados supracitadas, explicitando as relações encontradas nesse trabalho.

Na perspectiva das duas bases exploradas, LMDB e DBpedia, identifica-se que os recursos audiovisuais estão devidamente inseridos no âmbito do

Linking Open Data, permitindo que aplicações embasadas nos princípios da *Web Semântica* sejam capazes de fornecer aos usuários dados e informações mais estruturadas para esse domínio. Sendo que em um segundo momento, os usuários, a partir dos dados recuperados, possam reutilizá-los com inferências sobre tais dados, contribuindo para pesquisadores e usuários que buscam uma fonte de informações mais semântica desse cenário, e, portanto, tenham suas necessidades informacionais supridas.

5 Considerações finais

Os conceitos e as tecnologias da *Web Semântica* estão evoluindo e têm sido desenvolvidas um conjunto de ferramentas estáveis, capazes de contribuir para os estudos em diversos contextos e áreas do conhecimento.

Diante da relação entre os estudos teóricos e pragmáticos sobre *Linked Data* e de audiovisuais, esse trabalho aponta para uma eminente relação entre estes campos de estudos. A realização de consultas dentro de conjuntos de dados DBpedia e LMDB, reúnem quantidades significativas de informações sobre os audiovisuais, esses dados como demonstrados podem atuar como principais contribuintes ao traçar as relações entre os mesmos, promovendo maiores informações sobre esse cenário a todos interessados.

As relações ampliam as dimensões de recuperação das informações para elementos característicos aos audiovisuais e significativos, como a conexão da trilha sonora com os filmes. No mesmo sentido, a persuasão encontrada pela exploração dos relacionamentos LMDB de dados audiovisuais e a utilização da *DBpedia* e da LMDB, como fontes para a satisfação das necessidades informacionais dos usuários, permite reunir uma gama de informações de difícil cruzamento para um usuário que não utiliza as tecnologias da *Web Semântica*.

Foi possível verificar que os relacionamentos entre audiovisuais são eficientes, e os resultados obtidos com as consultas SPARQL apontaram a quantidade de relações existentes se tratando de dados sobre recursos audiovisuais. Ainda que as consultas realizadas tenham demonstrado que há um

alto número de dados incompletos dentro dos conjuntos de dados, o que torna os resultados obtidos menos íntegros. Essa questão foi averiguada, em diversas consultas, o que pode estar vinculado a uma não atualização dos dados, ou ainda, de uma não manutenção das URIs utilizadas nos recursos.

Apesar de tal fato, os resultados obtidos apontam que os princípios *Linked Data*, juntamente com as tecnologias da *Web Semântica*, podem auxiliar na ligação entre informações sobre recursos visuais, sendo uma fonte de extrema relevância para a representação destes objetos. A ligação dos dados não só poderá ser fonte para a construção de catálogos mais dinâmicos, mas também poupará o retrabalho durante o processo de descrição de recursos informacionais.

Como trabalhos futuros, busca-se construir um ambiente informacional digital que permita aos usuários consultarem as ligações existentes entre recursos audiovisuais, em uma linguagem mais clara para o usuário, sendo possível a construção desse ambiente utilizando-se dos princípios *Linked Open Data* e SPARQL como ferramentas para localização dos dados.

Financiamento

O trabalho tem o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Referências

ABELE, A.; MCCRAE, J. **The linking open data cloud diagram**. 2017.

ARAKAKI, F. A. **Linked data: ligação de dados bibliográficos**. 2016. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2016.

AUMONT, J. **A estética do filme**. Campinas: Papyrus, 2012.

BERNERS-LEE, T. **Linked data**. 2006.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web. **Scientific American**, London, v. 284, n. 5, p. 28-37, 2001.

- BUARQUE, M. D. Estratégias de preservação de longo prazo em acervos sonoros e audiovisuais. 2008. In: ENCONTRO NACIONAL DE HISTÓRIA ORAL, 9., 2008, São Leopoldo, RS. **Anais...** Rio de Janeiro: Associação Brasileira de História Oral; São Leopoldo: UNISINOS, 2008.
- DAVENPORT, T. H.; BARTH, P.; BEAN, R. How big data is different. **MIT Sloan Management Review**, Boston, v. 54, n. 1, p. 43, 2012.
- DUCHARME, B. **Learning SPARQL: querying and updating with SPARQL 1.1**. Massachusetts: O'Reilly Media, 2013.
- 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**, João Pessoa, v. 23, n. 2, p. 13-23, maio/ago. 2013.
- FRAYLING, C.; KEMP, P. **Tudo sobre o cinema**. Rio de Janeiro: Sextante, 2011.
- GODDARD, L.; BYRNE, G. The strongest link: libraries and linked data. **D-Lib magazine**, Reston, v. 16, n. 11/12, 2010.
- GRISOTO, A. P. **Um estudo acerca dos recursos audiovisuais no contexto do Linked data**. 2016. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2016.
- LEONE, E.; MOURÃO, M. D. **Cinema e montagem**. São Paulo: Ática, 1987.
- MATTHEWS, B. Semantic web technologies. **JISC Technology and Standards Watch**, Bristol, v. 6, n. 6, p. 8, 2005
- RASMUSSEN PENNINGTON, D. Demystifying Linked Data: are you ready for what's next?. **CILIP Update**, London, n. Jul./Aug., p. 34-36, 2016.
- SANTAREM SEGUNDO, J. E. Web Semântica: introdução a recuperação de dados usando SPARQL. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 15., 2014, Belo Horizonte, MG. **Anais eletrônicos...** Belo Horizonte: ANCIB, 2014.
- SANTAREM SEGUNDO, J. E.; CONEGLIAN, C. S. Web semântica e ontologias: um estudo sobre construção de axiomas e uso de inferências. **Informação & Informação**, Londrina, v. 21, n. 2, p. 217-244, dez. 2016.
- SIMIONATO, A. C. O tratamento descritivo para recursos audiovisuais: a representação de filmes. In: ALBUQUERQUE, A. C.; SIMIONATO, A. C. **Recursos audiovisuais: sua contemporaneidade na organização e representação da informação e do conhecimento**. Rio de Janeiro: Interciência, 2017.

SIMIONATO, A. C.; ARAKAKI, F. A.; SANTOS, P. L. V. A. C. Descrição em bibliotecas, arquivos, museus e galerias de arte: linkando recursos e comunidades. **Informação & Informação**, Londrina, v. 22, n. 2, p. 449-466, out. 2017.

VAN HOOLAND, S.; VERBORGH, R. **Linked data for libraries, archives and museums**: how to clean, link and publish your metadata. London: Facet publishing, 2014.

Audiovisual and Linked Data: a study of the DBpedia and LMDB databases

Abstract: As proponent of the Semantic Web and Linked data principles, the Linking Open Data initiative offers an enormous proportion of audiovisual data, which can assist in the search and retrieval of more accurate information. In this sense, the objective of this work is to explore the possible relations between audiovisual databases and Linking Open Data, to present the potential of this initiative for users who seek detailed sources of information about the audiovisual resources. A qualitative research was used, with an exploratory and applied nature, based on the scientific literature of the Linked Data, Semantic Web and audiovisual, and subsequently the DBpedia and LMDB with the use of the SPARQL protocol. The study considers that datasets available in the Linking Open Data can assist the link between information on audiovisual resources, as well as, it may be a source for the construction of more dynamic catalogues, reducing rework during the process of description of informational resources.

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

Recebido: 24/11/2017
Aceito: 23/03/2018