

5

Estudo de caso

De modo a ilustrar o processo de recomendação proposto no capítulo anterior, apresentamos importantes informações que são extraídas da fonte de dados `data.open.ac.uk` da universidade *The Open University*, de acordo com opções escolhidas no processo de recomendação de fontes RDF. As informações extraídas definem estatísticas que o usuário utiliza durante o processo de recomendação, que são analisadas segundo as etapas correspondentes do processo (ilustradas na figura 4.1).

Este capítulo está organizado da seguinte forma, a seção 5.1 apresenta uma análise dos resultados da etapa de geração de estatísticas, que foi definida na seção 4.2, onde um conjunto de estatísticas sobre a fonte de dados é extraído e analisado. A seção 5.2 analisa os resultados da etapa de escolha de classes, definida na seção 4.5; as estatísticas apresentadas permitem analisar como as propriedades e relações entre indivíduos influenciam na escolha de informação relevante. Por último, a seção 5.3 apresenta os resultados da etapa de escolha de indivíduos, definida na seção 4.4; descreve ainda os resultados finais do processo de recomendação junto com o ranking de fontes de dados RDF. Além disso, a seção 5.3 contém uma comparação dos tipos de busca (apresentados na seção 4.4.2), assim como uma avaliação de como as diferentes formas de envio do conjunto das palavras-chave afeta o ranking de fontes RDF.

5.1

Etapa de geração de estatísticas

Depois que o processamento de dados da etapa de geração de estatística finaliza, os dados são consumidos pelo serviço de indexação e eles são disponibilizados para o usuário.

Na tabela 5.1 apresenta-se uma amostragem das classes da fonte de dados (das 68 classes recuperadas), ordenadas de acordo com o número de indivíduos. Este tipo de informação pode melhorar os planos de execução de um mediador, no âmbito de consultas federadas [2, 24]. O envio de uma consulta pode-se priorizar, só se podemos antecipar os tipos de dados que podem ser encontrados em uma fonte de dados e assim otimizar o plano de execução da consulta [2, 24].

Bancos de dados de triplas, tais como *Virtuoso Quad Store*, geram estas estatísticas automaticamente usando o vocabulário VoiD [24, 2]. No entanto, gerar estatísticas sobre dados remotos e manter esta informação atualizada não é uma tarefa fácil [2]. O processo proposto reduz o custo desta tarefa.

Tabela 5.1: Lista de classes da fonte data.open.ac.uk

	Nome	Núm. de Indiví- duos
1	http://purl.org/goodrelations/v1#Offering	61355
2	http://xmlns.com/foaf/0.1/Document	35577
3	http://purl.org/ontology/bibo/Article	24061
4	http://data.open.ac.uk/ontology/red/Experience	19574
5	http://purl.org/ontology/bibo/AcademicArticle	16013
6	http://purl.org/media#Collection	15505
7	http://xmlns.com/foaf/0.1/Person	13610
8	http://purl.org/ontology/bibo/Book	13176
9	http://purl.org/ontology/bibo/Document	8767
10	http://purl.org/ontology/bibo/Collection	8303

Existem algumas observações sobre as classes listadas na tabela 5.1. Na ontologia <http://purl.org/ontology/bibo>, a classe **Document** é equivalente a <http://xmlns.com/foaf/0.1/Document> através da propriedade *owl:equivalentClass*. Um indivíduo então pode pertencer à classe **Document** de ambas ontologias. Se no processo da escolha de classes (descrito na seção 4.5) são escolhidas classes com sobreposição de indivíduos, apenas um dos indivíduo é considerado no processo de indexação, para evitar sobrecarga de processamento de dados.

Tabela 5.2: Lista de prefixos usados nas classes

	Nome
1	http://purl.org
2	http://data.open.ac.uk
3	http://www.w3.org
4	http://xmlns.com
5	http://xcri.org
6	http://vocab.deri.ie
7	http://courseware.rkbexplorer.com
8	http://dbpedia.org
9	http://linkedevents.org
10	http://www.daml.org
11	http://rdfs.org

A tabela 5.2 apresenta os diferentes prefixos usados nas URIs das classes definidas na fonte de dados. Pode-se observar que os recursos são descritos com o vocabulário da fonte `data.open.ac.uk`, que representa o mesmo domínio da fonte. Na descrição da fonte também são usados vocabulários externos.

5.2

Etapas de escolha classes

5.2.1

Resumo das estatísticas

Depois do processamento dos dados na etapa de escolha de classes, os documentos gerados, para cada um dos indivíduos coletados, contêm informações que são enviadas para o serviço de indexação para serem analisadas. O serviço de indexação caracteriza a informação e retorna as seguintes estatísticas:

- **R1**: Número de triplas por documento e por coleção de documentos.
- **R2**: Número de propriedades de tipo de dados (do inglês *Data properties*), por documento e por coleção de documentos. Por exemplo, a tripla a seguir é considerada na contagem:

$\langle \textit{Sujeito} \rangle \langle \textit{Predicado} \rangle \langle \textbf{Literal} \rangle$.

- **R3**: Número de propriedades de objeto (do inglês *Object property*) por documento e por coleção de documentos. Por exemplo, a tripla a seguir é considerada na contagem:

$\langle \textit{Sujeito} \rangle \langle \textit{Predicado} \rangle \langle \textbf{Objeto} \rangle$.

- **R4**: Número de vezes que o indivíduo aparece como sujeito da tripla. Por exemplo, a tripla a seguir é considerada na contagem:

$\langle \textbf{Indivíduo} \rangle \langle \textit{Predicado} \rangle \langle U \rangle$.

Considerando que $U = \{\textit{Objeto} \cup \textit{Literal}\}$ e $R4 = \{R2 \cup R3\}$.

- **R5**: Número de relações onde o indivíduo participa, considerando a posição dele como sujeito ou objeto. Por exemplo, as triplas a seguir são consideradas na contagem:

Tipo a. $\langle \textbf{Indivíduo} \rangle \langle \textit{Predicado} \rangle \langle \textit{Objeto} \rangle$.

Tipo b. $\langle \textit{Sujeito} \rangle \langle \textit{Predicado} \rangle \langle \textbf{Indivíduo} \rangle$.

Dessa forma, Pode-se identificar qual é o indivíduo que participa de mais relações.

- **R6**: Predicado mais usado por documento e por coleção de documentos.

- **R7**: Distribuição de palavras por documento e por coleção de documentos. Fornece informação aos componentes para calcular a frequência de palavras. Assim, valores como frequência do termo no documento (**TF** do inglês *Document Term Frequency*), frequência do termo entre os documentos (**DF** do inglês *Document Frequency*) e inverso da frequência do termo entre os documentos (**IDF** do inglês *Inverse Document Frequency*) [61] são calculados no processo de indexação.
- **R8**: Informação geral sobre o vocabulário usado na descrição dos indivíduos.

A seguir, para as estatísticas listadas anteriormente apresenta-se uma descrição dos resultados obtidos, considerando as classes definidas na tabela 4.4, no capítulo 4.

5.2.2

Artigos acadêmicos

As estatísticas sobre a coleção de documentos são apresentadas na tabela 5.3. Estas estatísticas são baseadas na quantidade de documentos coletados.

Tabela 5.3: Lista de estatísticas na coleção de documentos

ID	Nome	Valor
R1	Triplas	293976
R2	Propriedades de tipo de dado	110527
R3	Propriedades de objeto	181160
R5	Relação da coleção (tipo <i>b</i>)	2289

Na tabela 5.3 observa-se uma considerável quantidade de relacionamentos (de **R3** e **R5**). A presença do predicado *owl:sameAs*, mostra a uma possível existência de relacionamentos a outras fonte de dados, mas, do total de 4578 predicados *owl:sameAs* encontrados, 2265 apontam para URIs do tipo `<info:xxxxx>`, que não representam um *link* válido. Além disso, o restante das URIs apontam para recursos dentro da mesma fonte.

A tabela 5.4 apresenta uma distribuição do número de propriedades e de relações por documento. Esta informação ajuda a decidir quais indivíduos podem ser utilizados na busca por fontes de dados relevantes, o que pode ser bastante útil quando o usuário não tem conhecimento da fonte de dados em análise.

Ressalta-se a presença do predicado `http://purl.org/dc/terms/creator`, usado para conectar recursos do domínio `kmi.open.ac.uk/` no *site*

Tabela 5.4: Estatísticas por documento na classe de Artigo Acadêmico

ID	Estatística	Máximo	Mínimo	Media
R1	Triplas	287	9	18.37
R2	Propriedades de tipo de dados	138	1	6.9
R3	Propriedades de objeto	277	6	11.3
R5	Relações (tipo <i>b</i>)	3	0	0.14

*The Knowledge Media Institute*¹. A organização *The Knowledge Media Institute* trabalha em parceria com a universidade *The Open University* na geração de material didático de tipo multimídia, o que explica-se a relação entre as duas organizações.

Existem alguns predicados como `http://purl.org/dc/terms/isPartOf` que a apontam para URIs do tipo `<urn:xxxxx>` que não são dereferenciáveis. Além dos domínios usados no vocabulário na fonte `data.open.ac.uk`, como foi descrito na etapa de geração de estatísticas, na seção 4.2, não se identifica outros domínios que permitam a descoberta de mais conexão.

Assim fica evidente que não há relação entre a fonte de dados `data.open.ac.uk` com fontes externas, sendo o objetivo deste trabalho atingir este tipo de necessidade.

É necessário ressaltar que na coleta dos dados não foram considerados os nós em branco encontrados nas triplas. Os identificadores dos nós em branco são locais para o documento onde são usados, mas não podem ser referenciados externamente [49]. As boas práticas de *Linked Data* recomendam minimamente o uso de nós em branco na publicação de dados [49].

5.2.3

Material multimídia

A tabela 5.5 apresenta estatísticas sobre a coleção de documentos de material multimídia. A tabela 5.6 apresenta estatísticas do número de propriedades e de relações por documento. Estas estatísticas são consideradas importantes para uma adequada escolha de indivíduos relevantes.

Tabela 5.5: Lista de estatísticas na coleção de documentos

ID	Nome	Valor
R1	Triplas	100370
R2	Propriedades de tipo de dado	48834
R3	Propriedades de objeto	51536
R5	Relação da coleção (tipo <i>b</i>)	0

¹<http://kmi.open.ac.uk/>

Pode-se afirmar que existe uma relação direta entre o número de computadores e a quantidade de dados que se pretende analisar para alcançar um resultado relevante. Quanto maior é o conteúdo do documento que o serviço de indexação processa, menos disponível está o serviço. Tudo isto comparando as tabelas 5.3 e 5.5, que apresentam as estatísticas do número de propriedades.

Tabela 5.6: Estatísticas por documento nas classes de material multimídia

ID	Estatística	Máximo	Mínimo	Media
R1	Triplas	33	22	26.50
R2	Propriedades de tipo de dados	19	12	12.89
R3	Propriedades de objeto	15	10	13.60
R5	Relações (tipo <i>b</i>)	0	0	0

Se o serviço de indexação está indisponível, a operação de *merge* (definida na seção 4.3.2) reconhece o estado e ativa um tempo de espera. Vale ressaltar que o tempo de espera não é sincronizado entre os diferentes processos MapReduce. A não uniformidade dos dados processados agrega uma probabilidade favorável na requisição da escrita no serviço de indexação, assim as requisições tem baixa probabilidade de concorrer. No entanto, este tempo de espera em uma requisição deve ser considerado nas opções de otimização de um fluxo de tarefas.

Na tabela 5.5, apresenta resultados sobre indivíduos (correspondentes os documentos coletados) que não participam como objeto nas triplas. Nota-se a ausência do predicado *owl:sameAs* na coleção de documentos .

Os valores do predicado <http://www.w3.org/TR/2010/WD-mediaont-10-20100608/publisher> contém conexões ao *site HM Government*², *site* que publica informação da educação do Governo do Reino Unido no domínio <http://education.data.gov.uk/>, um total de 3788 conexões. O valor único deste predicado corresponde ao identificador da universidade *The Open University*, que é atribuído pelo governo e registra a instituição como organização de educação.

A maior quantidade de conexões se refere ao repositório de *podcast* com domínio <http://podcast.open.ac.uk> da universidade, com 14650 conexões. As conexões usam URIs da localização física dos recursos, por exemplo <http://podcast.open.ac.uk/feeds/a200-history/a200-palais-jacques-coeur-decoration.m4v>. Destes objetos, poderiam-se usar o nome do recurso no processo de busca, mas no trabalho de Nikolov [15] recomenda-se usar propriedades de tipo texto.

²*His Majesty's Government Education*, <http://.data.gov.uk/>

Existem conexões para o domínio <http://www.lingvoj.org/>. Esta fonte de dados disponibiliza informações sobre entidades da linguagem humano, por exemplo: línguas, caracteres, e palavras. A fonte analisada usa recursos que definem o idioma do material acadêmico publicado, que na grande maioria pertence ao idioma Inglês.

Por último, detectam-se predicados com valores no objeto de tipo texto que referenciam recursos do site *iTunes Preview*³, um total de 3230 conexões. De mesma forma que nos *podcast*, estes últimos predicados contêm valores com a localização do recurso, por exemplo, a URI <http://itunes.apple.com/gb/itunes-u/rio+20-united-nations-conference/id538502636?mt=2>. Estas URIs contêm informações relevantes para a busca. Como nem sempre a descrição de uma URI contém uma informação útil, estes valores não são considerados na busca.

5.3

Etapa de escolha de indivíduos

Finalizado o processamento na etapa de escolha de indivíduos, os resultados são enviados até o serviço de indexação, onde o usuário pode gerar um conjunto de estatísticas e identificar as fontes relevantes segundo o tópico de interesse.

A seguir apresenta-se o ranking das fontes encontradas, ordenadas pelo número de indivíduos e o ranking das classes. Os rankings se relacionam com as classes escolhidas pelo usuário na etapa de escolha de classes (definida na seção 4.5). Realiza-se uma comparação dos tipos de busca, apresentados na seção 4.4.2, juntamente com o ranking de fontes e o de classes, avalia-se também o comportamento das buscas usando diferentes amostras de indivíduos.

5.3.1

Ranking de fontes de dados

O resultado da busca por palavras-chave é apresentado nas figuras 5.1 e 5.2. As figuras apresentam o resultado das buscas definidas na seção 4.4.2, para as classes “Artigo Acadêmico” e “Material Multimídia”, respectivamente.

Artigos acadêmicos

O resultado da busca por palavras-chave apresentado na figura 5.1 foi obtido com uma amostra de 1280 indivíduos, usando um conjunto de 10 palavras chaves por busca. O processo de busca foi executado usando a

³<http://itunes.apple.com/>

configuração definida na tabela 4.7, no capítulo 4, onde define-se um *cluster* de 5 computadores.

Um *cluster* é definido segundo a capacidade de atendimento de requisições por parte do Síndice. O Síndice não consegue um atendimento aceitável com uma quantidade de computadores maior à definida na tabela 4.7. Onde é definido um *cluster* limitado a 5 computadores, com 8 buscas paralelas por computador. As limitações foram encontradas durante o desenvolvimento deste trabalho, e resolvidas com ajuda dos mesmos desenvolvedores do Síndice [12].

Porém, a arquitetura, definida no capítulo 6, suporta um número maior de computadores no processo de busca, é necessário considerar as limitações das tecnologias usadas.

O resultado obtido no trabalho de Nikolov [15] é muito similar ao apresentado na tabela 5.1. A abordagem de Nikolov usa um subconjunto dos dados definidos na fonte `data.open.ac.uk`, que foi usada para ilustrar o processo proposto no capítulo 4. Os resultados de Nikolov foram baseados na classe **bibo:Journal** da fonte `data.open.ac.uk`, mas esta classe não faz parte do vocabulário usado na fonte. Portanto, na ausência da classe **bibo:Journal**, a classe **bibo:AcademicArticle** correspondente à artigo acadêmico foi escolhida no processo de busca.

Na figura 5.1, o ranking das fontes encontradas apresenta a DBPedia com a maior quantidade de indivíduos coletados, junto com outras fontes de dados como RKBExplorer e *open EAN* correspondente a `openean.kaufkauf.net`. O RKBExplorer é uma fonte que disponibiliza informação de material acadêmico, *courseware*, pessoas e organizações. O *open EAN* é uma fonte relacionada com comércio eletrônico. Na figura 5.1 juntamente com a lista de nomes das fontes é exibida uma anotação de relevância de tipo: “+/-” (pode ser considerada relevante), “+” (relevante), “-” (não relevante). Estas anotações denotam a relevância da fonte em um determinado contexto, segundo os interesses do usuário.

Material Multimídia

O resultado do processo de busca por palavras-chave do material multimídia é apresentado na figura 5.2. Neste resultado foi obtida uma amostra de 307 indivíduos, usando um conjunto de 10 palavras chaves por busca. O processo de busca também foi executado usando a lista de parâmetros definida na tabela 4.7, usando um *cluster* de 5 computadores. Lembrando que os indivíduos foram filtrados pelo tópico `http://data.open.ac.uk/topic/engineering_and_technology`, melhorando assim o processo de busca.

Na figura 5.2, no ranking de fontes do material multimídia a fonte

Legenda	
M'	Amostragem de indivíduos
F'	Número de fontes recuperadas
R'	Número de indivíduos recuperados
T'	Tempo de execução da busca
+/-	Relevância das fontes

Lista de fontes de dados		
A	dbpedia.org	+
B	oai.rkbexplorer.com	+
C	openean.kaufkauf.net	+/-
D	dotac.rkbexplorer.com	+
F	api.hi5.com	-
G	eprints.rkbexplorer.com	+
H	www.bibsonomy.org	+/-
I	www.articlestoreprint.net	+
J	rdf.freebase.com	+
K	bio2rdf.org	+
L	products.semweb.bestbuy.com	+/-
M	www.rdfabout.com	+/-
N	sw.opencyc.org	+/-
O	data-gov.tw.rpi.edu	+
P	identi.ca	-
Q	data.dcs.shef.ac.uk	+
R	www.sudoc.fr	+
S	www.ecademy.com	+
T	dblp.rkbexplorer.com	+
U	www.talkdigger.com	+/-
V	chem2bio2rdf.org	+
W	foaf.qdos.com	+/-

M':	1280
F':	11549
R':	59083
T':	2h13min

A	1
C	2
B	3
F	4
D	5
G	6
Q	7
S	8
M	9
H	10
U	11
I	12
R	13
O	14
W	15
V	16
L	17
J	18
N	19
P	20

Ranking pelo número de indivíduos

M':	1280
F':	9770
R':	52142
T':	1h55min

A	1
C	2
B	3
F	4
D	5
G	6
S	7
Q	8
M	9
U	10
R	11
H	12
I	13
O	14
V	15
W	16
L	17
N	18
P	19
J	20

Ranking pelo número de indivíduos

Figura 5.1: Resultados sobre Artigos Acadêmicos: (a) Busca por combinação independente de palavras-chave, (b) Busca por acúmulo de palavras-chave

`openean.kaufkauf.net` aparece como a primeira no ranking nas duas buscas efetuadas. Os rankings gerados são semelhantes com algumas variações nas localizações das fontes. Este ranking trouxe fontes como: a `disi.unitn.it`, que representa o departamento de Engenharia de Software e Computação da Universidade de Trento na Itália, a `cordis.rkbexplorer.com`, que representa à *Community Research and Development Information Services*, e algumas outras sobre comércio eletrônico.

Na figura 5.2, os resultados do material multimídia são mais especializados pelo uso de um tópico específico. O Material Multimídia está ligado ao material acadêmico na fonte de dados de `data.open.ac.uk`, por isso que a maioria das fontes tem alguma semelhança com os resultados da figura 5.1, correspondente à classe de artigos acadêmicos.

Em relação aos tipos de busca usados, a busca de combinação de palavras mantém uma quantidade maior de informação recuperada em relação à busca por acúmulo de palavras-chave. Uma outra observação comparando os resultados das figuras 5.1 e 5.2 é a quantidade de resultados obtidos. A amostra do material multimídia (307 indivíduos) é menor que da classe de

Legenda		
M'	Amostragem de indivíduos	
F'	Número de fontes recuperadas	
R'	Número de indivíduos recuperados	
T'	Tempo de execução da busca	
+/-	Relevância das fontes	

Lista de fontes de dados		
A	openean.kaufkauf.net	+
B	oai.rkbexplorer.com	+
C	dbpedia.org	+
D	sw.opencyc.org	+/-
F	data.dcs.shef.ac.uk	+
G	dotac.rkbexplorer.com	+
H	www.bibsonomy.org	+/-
I	data-gov.tw.rpi.edu	+
J	www.articlestoreprint.net	+
K	sw.cyc.com	+/-
L	www.rdfabout.com	+/-
M	eprints.rkbexplorer.com	+
N	www.talkdigger.com	+
O	products.semweb.bestbuy.com	+
P	cordis.rkbexplorer.com	+
Q	purl.org	+/-
R	rdf.freebase.com	+
S	bio2rdf.org	+
T	identi.ca	+
U	disi.unitn.it	+
V	rs.tdwg.org	+/-

M':	307
F':	14383
R':	79459
T':	1h10min

A	1
B	2
C	3
D	4
F	5
G	6
H	7
I	8
J	9
K	10
L	11
M	12
N	13
O	14
P	15
Q	16
R	17
S	18
T	19
U	20

M':	307
F':	11489
R':	59021
T':	47min

A	1
B	2
C	3
F	4
G	5
L	6
D	7
I	8
J	9
H	10
N	11
M	12
P	13
O	14
S	15
K	16
U	17
V	18
Q	19
R	20

Figura 5.2: Resultados sobre o material multimídia: (a) Busca por combinação independente das palavras-chave, (b) Busca por acúmulo de palavras-chave

artigos acadêmicos (1280 indivíduos), no entanto, o número de indivíduos retornados pelo índice semântico é maior nas buscas de material multimídia. Uma possível explicação para esta discrepância está baseada na quantidade de propriedades de tipo de dados usadas na descrição dos indivíduos, e como os valores das propriedades forem escritos. O que afeta diretamente o processo de extração de palavras chaves. A tabela 4.6, do capítulo 4 e a tabela A.1 do apêndice A apresentam as diferentes propriedades das classes artigo acadêmico e de material multimídia, respectivamente.

5.3.2

Ranking de classes

Artigos acadêmicos

Na figura 5.3 apresenta-se o ranking de classes para artigos acadêmicos. As classes estão ordenadas em forma decrescente pela quantidade de indivíduos, em um total de 19480 classes. Uma amostra de 30 classes é apresentada na figura 5.3 e algumas outras classes estão relacionadas com publicações

acadêmicas. Estas classes são identificadas considerando o contexto da busca. Apresenta-se também a definição da classe **bibo:AcademicArticle**, para explicar a relação com as classes encontradas.

O resultado mostra a classe **dc:agent** do vocabulário Dublin Core (DC)⁴ como a primeira do ranking, junto com ela várias classes do vocabulário FOAF⁵ e GoodRelations⁶ dominam o ranking. Como se indica no trabalho de Hogan [49] estas classes são bastante usadas na descrição de conjuntos de dados RDF. Aparece também o vocabulário *Advanced Knowledge Technologies* (AKT)⁷ especializado na modelagem de material acadêmico.

Além disso, o ranking é dominado por superclasses que representam conceitos de alto nível, como por exemplo: **owl:ontology**, **akt:thing**, **owl:class**, **foaf:document**, **dc:bibliographicresource**, **gr:businessentity**, etc. Isto é observado também no trabalho de Nikolov [15].

O trabalho de Nikolov apresenta uma abordagem de *ontology matching*, que pode-se aplicado no filtro das superclasses e melhorar o ranking. Para fazer esta filtragem, a abordagem de Nikolov compara cada uma das classes encontradas com a classe que indica o usuário (que em nosso caso seria **bibo:AcademicArticle**), tudo isto no nível de metadados das classes para assim calcular o grau de similaridade das classes. Este processo não é abrangido pelo estudo feito neste documento, mas percebe-se uma necessidade deste processo para melhorar o ranking apresentado.

Material Multimídia

Na figura 5.4 apresenta-se o ranking de classes para o material acadêmico, especificamente das classes **http://data.open.ac.uk/podcast/ontology/VideoPodcast** e **http://data.open.ac.uk/podcast/ontology/AudioPodcast**, definidas no vocabulário da fonte **http://data.open.ac.uk/**. As classes estão ordenadas em forma decrescente pela quantidade de indivíduos, e um total de 57277 classes são recuperadas na busca. Uma amostra de 30 classes é apresentada na figura 5.4 e algumas classes relacionadas com recursos multimídia.

Como nos resultados da classe artigos acadêmicos, o ranking da figura 5.4 mostra a classe **dc:agent** do vocabulário Dublin Core como a primeira do ranking. Junto com ela várias classes de GoodRelations dominam o ranking. Várias delas estão relacionadas com comércio eletrônico. O vocabulário FOAF também aparece nas primeiras posições.

⁴<http://dublincore.org/>

⁵<http://xmlns.com/foaf/spec/>

⁶<http://www.heppnetz.de/projects/goodrelations/>

⁷<http://www.aktors.org/akt/>

Pos	Nome da classe
1	http://purl.org/dc/terms/agent
2	http://xmlns.com/foaf/0.1/person
3	http://xmlns.com/foaf/0.1/document
4	http://xmlns.com/foaf/0.1/agent
5	http://www.w3.org/2003/01/geo/wgs84_pos#spatialthing
6	http://www.w3.org/2000/10/swap/pim/contact#person
7	http://www.w3.org/2000/10/swap/pim/contact#socialentity
8	http://xmlns.com/foaf/0.1/image
9	http://blogs.yandex.ru/schema/foaf/posts
10	http://www.w3.org/2002/07/owl#ontology
11	http://www.w3.org/2002/07/owl#annotationproperty
12	http://purl.org/goodrelations/v1#productorservicemodel
13	http://purl.org/dc/terms/licensedocument
14	http://purl.org/goodrelations/v1#productorservice
15	http://purl.org/goodrelations/v1#businessentity
16	http://purl.org/dc/terms/bibliographicresource
17	http://www.w3.org/2002/07/owl#class
18	http://www.aktors.org/ontology/support#thing
19	http://www.aktors.org/ontology/support#intangible-thing
20	http://www.aktors.org/ontology/portal#generic-agent
21	http://www.aktors.org/ontology/portal#abstract-information
22	http://www.aktors.org/ontology/portal#publication-reference
23	http://www.aktors.org/ontology/support#calendar-date
24	http://www.w3.org/1999/02/22-rdf-syntax-ns#seq
25	http://www.aktors.org/ontology/portal#article-reference
26	http://xmlns.com/foaf/0.1/personalprofiledocument
27	http://www.w3.org/2000/01/rdf-schema#container
28	http://zeitkunst.org/bibtex/0.1/bibtex.owl#entry
29	http://purl.org/rss/1.0/item
30	http://purl.org/rss/1.0/channel
45	http://purl.org/ontology/bibo/document
52	http://swrc.ontoware.org/ontology#publication
57	http://swrc.ontoware.org/ontology#article
60	http://purl.org/ontology/bibo/book
88	http://www.aktors.org/ontology/portal#journal
111	http://www.articlestoreprint.net:8080/vocab/resource/article
121	http://www.aktors.org/ontology/portal#thesis-reference
131	http://purl.org/ontology/bibo/thesis
999	http://dbpedia.org/resource/literature

(a)

```

<!-- http://purl.org/ontology/bibo/AcademicArticle -->
<owl:Class rdf:about="AcademicArticle">
  <rdfs:subClassOf rdf:resource="Article"/>
  <rdfs:comment xml:lang="en">A scholarly academic article,
  typically published in a journal.</rdfs:comment>
</owl:Class>
<!-- http://purl.org/ontology/bibo/Article -->
<owl:Class rdf:about="Article">
  <rdfs:subClassOf rdf:resource="Document"/>
  <rdfs:comment xml:lang="en">A written composition in prose,
  usually nonfiction, on a specific topic, forming an independent
  part of a book or other publication, as a newspaper or magazine.
  </rdfs:comment>
</owl:Class>

```

(b)

Figura 5.3: Ranking de artigo acadêmico: (a) Ranking de classes, (b) Definição de classe `bibo:AcademicArticle`

Este ranking também é dominado por superclasses, como por exemplo: **owl:ontology**, **akt:thing**, **owl:class**, **foaf:document**, **gr:productorservice**, **dbpedia:work**, etc.

Existem classes e fontes recuperadas neste processo que não foram apresentadas nos rankings descritos nesta seção e que provavelmente tenham uma proximidade com as classes escolhidas pelo usuário. Para melhorar o ranking é necessária uma análise especializada no nível de metadados, assim eliminar classes e fontes irrelevantes nos rankings apresentados, como é recomendado por Nikolov [15].

Pos	Nome da classe
1	http://purl.org/dc/terms/agent
2	http://www.w3.org/2002/07/owl#ontology
3	http://www.w3.org/2002/07/owl#annotationproperty
4	http://xmlns.com/foaf/0.1/document
5	http://purl.org/goodrelations/v1#productorservicemodel
6	http://purl.org/dc/terms/licensedocument
7	http://purl.org/goodrelations/v1#businessentity
8	http://purl.org/goodrelations/v1#productorservice
9	http://xmlns.com/foaf/0.1/agent
10	http://xmlns.com/foaf/0.1/person
11	http://xmlns.com/foaf/0.1/image
12	http://www.w3.org/2003/01/geo/wgs84_pos#spatialthing
13	http://www.w3.org/2002/07/owl#class
14	http://www.w3.org/2000/10/swap/pim/contact#person
15	http://www.w3.org/2000/10/swap/pim/contact#socialentity
16	http://purl.org/dc/terms/bibliographicresource
17	http://blogs.yandex.ru/schema/foaf/posts
18	http://www.w3.org/2002/07/owl#objectproperty
19	http://www.w3.org/2002/07/owl#functionalproperty
20	http://www.w3.org/1999/02/22-rdf-syntax-ns#seq
21	http://zeitkunst.org/bibtex/0.1/bibtex.owl#entry
22	http://www.aktors.org/ontology/support#thing
23	http://www.w3.org/2000/01/rdf-schema#container
24	http://www.aktors.org/ontology/support#intangible-thing
25	http://purl.org/rss/1.0/item
26	http://rdfs.org/sioc/ns#post
27	http://purl.org/rss/1.0/channel
28	http://rdfs.org/sioc/ns#container
29	http://www.aktors.org/ontology/porta#generic-agent
30	http://www.aktors.org/ontology/porta#abstract-information
31	http://www.aktors.org/ontology/support#temporal-thing
82	http://www.w3.org/2004/02/skos/core#concept
83	http://rdf.freebase.com/ns/common.topic
84	http://dbpedia.org/ontology/work
90	http://dbpedia.org/ontology/musicalartist
106	http://www.aktors.org/ontology/porta#conference-proceedings-reference
116	http://www.aktors.org/ontology/porta#publishing-house
131	http://dbpedia.org/ontology/film
153	http://purl.org/dc/terms/mediatypeorextent
168	http://purl.org/omapi/0.2/#track
191	http://dbpedia.org/ontology/Software

(a)

```

<!-- http://data.open.ac.uk/podcast/ontology/VideoPodcast -->
<owl:Class rdf:about="VideoPodcast">
  <rdfs:subClassOf rdf:resource="PodCast"/>
  <rdfs:comment xml:lang="en">
    OU podcasts, part of which is being made available through iTunes U,
    is a collection of Video material related to education
    and research at the Open University. This dataset is defined using
    a variety of ontologies, including the W3C Media Ontology,
    as well as our own SKOS representation of the iTunes U topic categories.
  </rdfs:comment>
</owl:Class>

<!-- http://data.open.ac.uk/podcast/ontology/AudioPodcast -->
<owl:Class rdf:about="AudioPodcast">
  <rdfs:subClassOf rdf:resource="PodCast"/>
  <rdfs:comment xml:lang="en">
    OU podcasts, part of which is being made available through iTunes U,
    is a collection of Audio material related to education
    and research at the Open University. This dataset is defined using
    a variety of ontologies, including the W3C Media Ontology,
    as well as our own SKOS representation of the iTunes U topic categories.
  </rdfs:comment>
</owl:Class>

```

(b)

Figura 5.4: Ranking do material multimídia: (a) Ranking de classes, (b) Definição das classes PodCast