

# Obtenção do Arcabouço Legal por meio da Adição das Remissões Externas aos Documentos Legais

Francisco de Oliveira<sup>1</sup>, José M. Parente de Oliveira<sup>1</sup>

<sup>1</sup>Instituto Tecnológico de Aeronáutica (ITA)  
São José dos Campos – SP – Brazil

{francisco.sismicro@gmail.com, parente@ita.br}

**Abstract.** *Legal documents have internal and external relationships. The hypothesis of this research is that the addition of external references between legal documents allows recovering the legal framework about a certain subject matter. To achieve this, legal documents were converted into RDF triples associated with a legal ontology. The resultant dataset was stored into a repository that allows SPARQL queries to be issued. The relevance of the data returned was classified by legal experts. The positive correlation of the measures assures the high similarity of the classification made by the specialists. F-Measure metric was applied and showed values, on average, twice the values obtained in the Google search engine, that has been adjusted to focus on the legal domain.*

**Keywords:** RDF, Ontology, Linked Open Data, Law

**Resumo.** *Documentos legais possuem relacionamentos internos e externos. A hipótese desta pesquisa é que a adição de referências externas entre documentos legais permite recuperar o arcabouço legal sobre um determinado assunto. Para conseguir isso, documentos legais foram convertidos em triplas RDF associadas a uma ontologia legal. O conjunto de dados resultante foi armazenado em um repositório que permite efetuar consultas SPARQL. A relevância dos dados retornados foi classificada por especialistas legais. A correlação positiva das medidas assegura a alta similaridade da classificação feita pelos especialistas. A aplicação da Medida-F mostrou valores, em média, o dobro dos valores obtidos no buscador do Google, que foi ajustado para focar no domínio legal.*

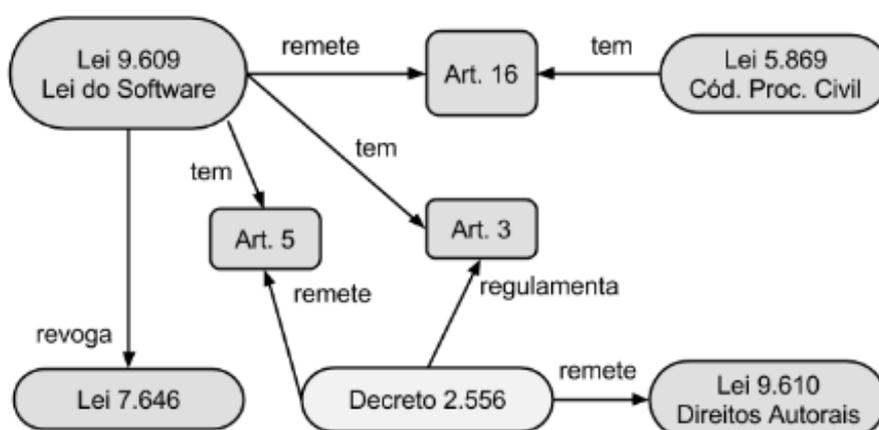
**Palavras-Chave:** RDF, Ontologia, Dados Ligados Abertos, Lei

## 1. Introdução

Desde o lançamento da web comercial em 1991 até os dias de hoje, o que se viu foi a explosão do volume de dados disponíveis [Baeza-Yates 2013], distribuídos em diversos domínios. Entre os vários campos de informação disponíveis na web está o domínio legal cujos principais provedores são os órgãos governamentais. A importância do domínio legal se faz presente no cotidiano de empresas, governos e, também, dos cidadãos. As normas, em sua maioria, são desconhecidas dos cidadãos comuns. Isso tem várias causas, entre elas a quantidade e a complexidade inerente ao domínio legal.

Além do grande número de textos legais publicados, há o alto grau de dependência entre as normas [Machado 2013], na forma de interligações, que constituem uma rica rede de informações [Lima 2008]. Por outro lado, em muitas situações do dia a dia é necessário recuperar o conjunto de dispositivos legais relacionados a um assunto de interesse com o propósito de encontrar soluções para problemas particulares.

No presente trabalho, um dispositivo legal é considerado como sendo uma unidade ou fragmento de informação legal. Dessa forma, um dispositivo legal poderá ser um artigo, um parágrafo, um inciso, uma alínea, um item e assim por diante. As interligações entre os vários dispositivos legais são efetuadas por meio de remissões internas ou externas, [Machado 2013], [Machado 2014]. As remissões internas interligam dispositivos contidos em um mesmo documento legal. As remissões externas referenciam dispositivos legais que estão em documentos distintos. No exemplo a seguir, o Decreto 2.556 que regulamenta o artigo 3o. da Lei 9.609, Lei do Software. A Lei 9.609, por sua vez, revoga a Lei 7.646. Esta lei, também, remete ao Artigo 16 da Lei 5.869. Na Figura 1 foram destacados apenas os elementos necessários para mostrar os relacionamentos do Decreto 2.556.



**Figura 1. Interligações entre dispositivos legais.**

O presente artigo tem como objetivo apresentar um método para recuperar o conjunto de dispositivos legais relevantes sobre um determinado assunto ou tema de interesse. Para tanto, adicionamos as remissões existentes aos documentos legais e assim acionamos a rede de informações relacionadas. O artigo está organizado da forma que segue. Na Seção 2 são apresentados alguns trabalhos relacionados. Na Seção 3, apresentamos o método proposto para recuperação do arcabouço legal de interesse. Na Seção 4, apresentamos os experimentos realizados e os resultados obtidos. Por fim, na Seção 5, apresentamos conclusões sobre o trabalho, as limitações encontradas e possíveis trabalhos futuros.

## **2. Trabalhos Relacionados**

Em [Machado 2013] foi realizada uma extensa pesquisa com o objetivo de se criar um Modelo Conceitual Formal de Relacionamentos do Ordenamento Jurídico Positivo (MROJ). O resultado do esforço de pesquisa é um conjunto de ontologias composto por

*Ontologia do Sistema Lógico Jurídico, Ontologia da Teoria da Norma Jurídica, Ontologia do Ordenamento Jurídico, Ontologia da Sistemática Legal Externa e Ontologia da Divisão Tricotômica das Pétreas.* Como benefícios do trabalho o autor cita a expressividade do modelo conceitual, a possibilidade de implementação do modelo em diversas linguagens ontológicas com diferentes níveis de expressividade de acordo com os requisitos computacionais, a interoperabilidade e a possibilidade de publicação do conhecimento legal e sua respectiva semântica. Tal trabalho serviu aos objetivos da pesquisa em curso por fornecer a *Ontologia da Sistemática Legal Externa*, base para a criação das triplas e a associação da ontologia ao repositório das triplas geradas.

Outro trabalho de conceitualização e modelagem ontológica vem de [Lima 2008] que desenvolveu na sua tese de doutorado um modelo conceitual que chamou de Modelo Genérico de Relacionamentos (MGR). O autor aplicou o seu modelo no projeto *Coletânea Brasileira de Normas e Julgados de Telecomunicações* por meio de uma pesquisa ação. O MGR é, segundo seu autor, formado por três entidades principais que são o Conceito, a Unidade de Informação e o Relacionamento. A partir delas são criadas subclasses para atender às necessidades do modelo. Como resultado final do modelo, são geradas uma ontologia para a organização da informação e uma tipologia.

No trabalho de mestrado de [Machado 2010], o autor aproveita o caráter hierárquico dos textos legais para melhorar o processo de busca de informação do ordenamento jurídico em relação aos métodos convencionais. O autor aplica técnicas amplamente utilizadas em recuperação de informação como a extração de radicais, indexação, lista de termos, entre outros. Diferentemente da pesquisa aqui relatada, no trabalho do autor a premissa é a busca do documento legal que contém certa informação, tal como fazem os mecanismos de buscas.

Na esfera governamental existe o projeto LexML-BR [PRODASEN 2016] que é baseado numa iniciativa multinacional. Pelo lado brasileiro, esse projeto visa organizar a informação jurídica disponibilizada pelos vários órgãos do governo brasileiro, conforme [Machado 2013]. Para a realização do objetivo, foi criado um portal de acesso público, no endereço [www.lexml.gov.br](http://www.lexml.gov.br), contendo mais de um milhão de documentos entre os quais aqueles referentes à legislação brasileira. Esse acervo pode ser pesquisado por meio de termos de busca que podem ser informados no portal. Por meio de algoritmos de recuperação de informação, é retornada uma lista dos documentos cujos termos de busca estão na respectiva ementa.

Na Finlândia, há o projeto FinLex [Frosterus 2013]. O projeto tem por finalidade publicar os dados legais finlandeses como dados ligados abertos mediante a aplicação das tecnologias de web semântica. A similaridade desse trabalho com o da atual pesquisa é a disponibilização da informação na forma de dados abertos.

Na Itália, além do projeto LexML-IT, similar ao LexML-BR, há vários esforços de pesquisa, como, por exemplo, [Soria 2007] que, por meio de anotação semântica em textos legais italianos, procurou melhorar a recuperação de informação jurídica.

Na área de conformidade legal encontra-se o trabalho alemão de [Thatmann 2014], cuja finalidade foi prover uma ontologia lightweight para representar o Ato Federal Alemão para Proteção de Dados (BDSG). Utiliza ontologia legal para tratar documentos jurídicos alemães relacionados à conformidade de SaS (Software

as service).

### 3. Método para Recuperação do Arcabouço Legal

O método consiste em explicitar as remissões internas e externas existentes nos documentos legais de forma a percorrer a rede de interligações dos documentos. A Figura 2 apresenta um esquema do método, contendo as etapas e produtos intermediários e final. Na sequência, o método é apresentado em detalhes.

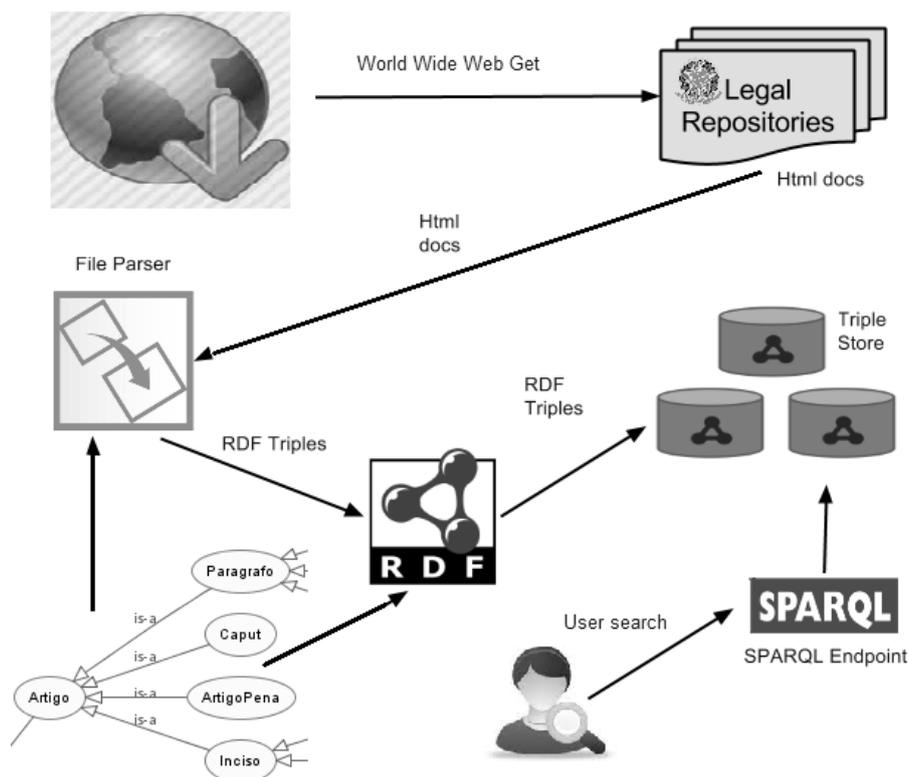


Figura 2. Abordagem utilizada para solucionar o problema

#### 3.1. Etapa 1 - Capturar os arquivos contendo os textos legais

Os documentos legais estão contidos em arquivos de computador de variados formatos, sendo os mais comuns o HTML, o DOC, o RTF e o PDF. Neste trabalho optou-se por trabalhar apenas com os documentos no formato HTML, que é o formato padrão atual para a publicação da legislação. A Tabela 1 apresenta um resumo dos formatos de documentos coletados em 06 de Agosto de 2017.

Formato do documento	Número de arquivos
HTML	70.016
PDF	8.302
DOC/RTF	1.737
Total	80.055

Tabela 1. Formatos de documentos legais

Após o download dos arquivos HTMLs, foi necessária uma reorganização dos arquivos legais, eliminação de redundâncias e separação dos arquivos que serão convertidos para RDF (leis, decretos e decretos-lei). Depois desta reorganização, o total de arquivos HTML a serem convertidos reduziu de 70.016 para 35.314 arquivos.

### 3.2. Etapa 2 - Converter o documento legal para a forma de triplas RDF

*RDF - Resource Description Framework* é uma linguagem formal com especificação definida por The World Wide Web Consortium (W3C) (<https://www.w3.org/RDF/>). Fornece uma maneira de descrever os recursos através de metadados na forma de triplas. Uma tripla é um conjunto ordenado formado por três elementos: um sujeito, um predicado e um objeto, nesta ordem. Como as triplas referenciam sujeitos e objetos, elas podem ser interligadas com outras triplas que se refiram ao mesmo recurso ou objeto. Assim, RDF forma um grafo rotulado e dirigido para representação de informações na Web.

A conversão dos documentos do formato HTML para o formato RDF é feita por meio de um programa de computador, aqui chamado de File Parser, criado especificamente para essa finalidade, que escaneia o diretório contendo os arquivos HTML. Cada arquivo HTML é passado para o File Parser que faz a conversão por meio de rotinas de tratamento de texto. A conversão é mediada pela ontologia *Ontologia Sistemática Externa Legal* de [Machado 2013], que foi modificada e estendida, ver Figura 3. A conversão utilizou a parte da articulação da ontologia, em destaque no retângulo na parte superior da figura.

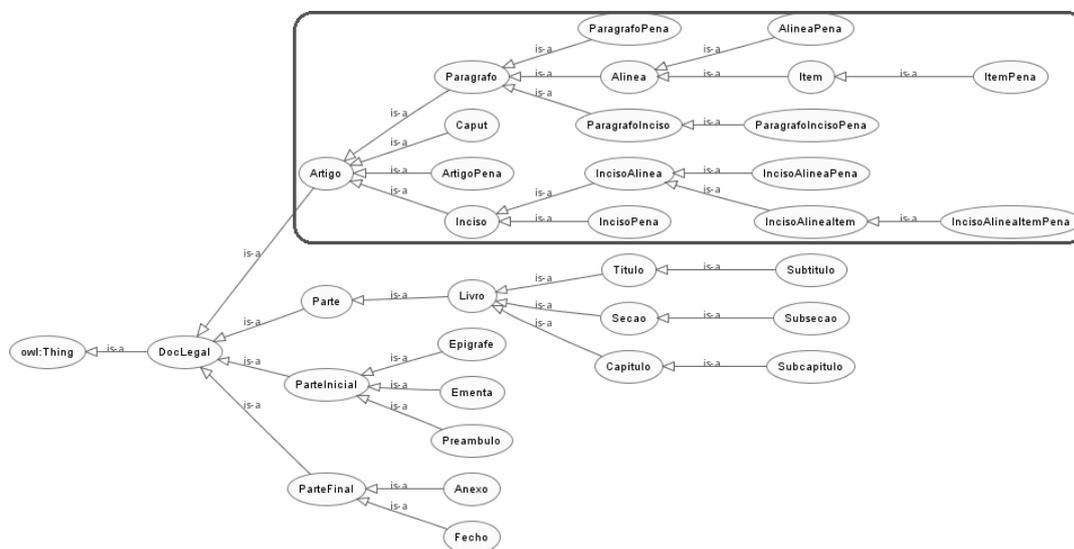


Figura 3. Ontologia Legal modificada e estendida

### 3.3. Etapa 3 - Armazenar as triplas em um repositório de triplas (triplestore)

Para efeito deste trabalho de pesquisa, o triplestore escolhido foi o TDB que faz parte do projeto de software Apache Jena, disponível em <https://jena.apache.org/>. É um repositório de triplas que pode ser acessado diretamente por meio da API Jena ou como um endpoint SPARQL. O File Parser gera um arquivo RDF serializado no formato Turtle. Depois envia-o para o triplestore por meio da API do Jena.

### 3.4. Etapa 4 - Disponibilizar o acesso às triplas por meio de um endpoint SPARQL

O acesso às triplas é disponibilizado por meio de um endpoint SPARQL. Neste trabalho optou-se pelo servidor Fuseki, que também faz parte do Apache Jena. Para os experimentos específicos deste trabalho, as queries (consultas) SPARQL foram enviadas diretamente de um programa em Java que utiliza a API do Jena.

### 3.5. Etapa 5 - Disponibilizar os arquivos legais como datasets RDF

Todas as triplas armazenadas no repositório também estão disponíveis em arquivos RDF serializados em Turtle. Os conjuntos de dados (datasets) RDF podem ser publicados na forma e sob os princípios de dados ligados abertos.

## 4. Experimentos e Resultados

Nesta seção são apresentados os experimentos realizados e os resultados obtidos nestes ensaios.

### 4.1. Experimentos

A conversão dos 35.314 documentos legais (em HTML) gerou 1.930.236 triplas no repositório. No quadro da Figura 4 tem-se os quantitativos de predicados pertencentes às triplas convertidas. Neste quadro, o nome do predicado está na coluna *Predicado* e a quantidade correspondente de ocorrências, geradas na conversão, está na coluna *Quantitativo*. Os predicados são oriundos da ontologia legal adotada. São responsáveis pela interligação das triplas de itens legais geradas. Dessa maneira, explicitam as remissões. Por exemplo, o predicado *lex:regulamentadoPor* liga um item legal *A*, sujeito da tripla, relacionado a um item legal *B*, objeto da tripla, de forma que *B* é a regulamentação legal de *A*. Nas colunas da esquerda na primeira linha do quadro da Figura 4, há o registro de 1.284 ocorrências para esse predicado.

QUANTITATIVO DE PREDICADOS					
Item	Predicado	Quantitativo	Item	Predicado	Quantitativo
1	<a href="#">lex:temConteudo</a>	"576098"^^xsd:integer	16	<a href="#">lex:regulamentadoPor</a>	"1284"^^xsd:integer
2	<a href="#">lex:ehUm</a>	"574420"^^xsd:integer	17	<a href="#">lex:temArtigoPena</a>	"1284"^^xsd:integer
3	<a href="#">lex:temArtigo</a>	"191453"^^xsd:integer	18	<a href="#">lex:temParagrafoPena</a>	"519"^^xsd:integer
4	<a href="#">lex:remetePara</a>	"176707"^^xsd:integer	19	<a href="#">lex:temIncisoPena</a>	"150"^^xsd:integer
5	<a href="#">lex:temParagrafo</a>	"128502"^^xsd:integer	20	<a href="#">rdf:type</a>	"74"^^xsd:integer
6	<a href="#">lex:temInciso</a>	"117816"^^xsd:integer	21	<a href="#">rdfs:domain</a>	"40"^^xsd:integer
7	<a href="#">lex:temParagrafoInciso</a>	"36802"^^xsd:integer	22	<a href="#">rdfs:range</a>	"40"^^xsd:integer
8	<a href="#">lex:temTextoIntegral</a>	"35314"^^xsd:integer	23	<a href="#">rdfs:subClassOf</a>	"32"^^xsd:integer
9	<a href="#">lex:temAlinea</a>	"28672"^^xsd:integer	24	<a href="#">lex:temAlineaPena</a>	"22"^^xsd:integer
10	<a href="#">lex:temIncisoAlinea</a>	"27687"^^xsd:integer	25	<a href="#">lex:temPena</a>	"5"^^xsd:integer
11	<a href="#">lex:revogadoPor</a>	"12915"^^xsd:integer	26	<a href="#">rdfs:subPropertyOf</a>	"4"^^xsd:integer
12	<a href="#">lex:revogou</a>	"12915"^^xsd:integer	27	<a href="#">owl:inverseOf</a>	"2"^^xsd:integer
13	<a href="#">lex:temIncisoAlineaItem</a>	"3652"^^xsd:integer	28	<a href="#">lex:temItemPena</a>	"1"^^xsd:integer
14	<a href="#">lex:temItem</a>	"2541"^^xsd:integer	29	<a href="#">owl:versionIRI</a>	"1"^^xsd:integer
15	<a href="#">lex:regulamenta</a>	"1284"^^xsd:integer			

Figura 4. Quantitativo de Predicados

Os experimentos consistiram de consultas por meio de termos relativos a certos temas. Os temas testados foram definidos aleatoriamente. São eles:

- (ACP) acumulação cargo público
- (DCO) direito consumidor
- (DHE) direitos humanos escola
- (FFI) fraude financeira
- (AIN) acesso informação
- (MIL) meios ilícitos
- (PRV) prevaricação
- (PCR) programa computador registrado
- (VDO) violência doméstica
- (TRD) tratamento desumano.

As consultas foram regidas por um conjunto de algoritmos. O principal deles é o Algoritmo *sparqlQuery*, que é mostrado a seguir em Algoritmo 1. Este recebe os termos de busca e ativa as subconsultas que fazem as buscas e retornam as triplas que casam com os termos informados. Para cada tripla retornada, o algoritmo faz a busca das triplas vizinhas até a distância de dois nós.

```
Data: Lista de termos
Result: Lista de triplas de dispositivos legais
triples = getQueryResults(Lista de termos);
while exists triple in triples do
    read current triple;
    add triple to Result;
    neighbours = getNeighbours(triple(subject));
    while exists neighbour in neighbours do
        read current neighbour;
        add neighbour to Result;
        nextNeighbours = getNextNeighbours(neighbour(object));
        while exists nextNeighbour in nextNeighbours do
            read current nextNeighbour;
            add nextNeighbour to Result;
        end
    end
end
```

**Algorithm 1:** sparqlQuery

A análise da relevância dos resultados das consultas se deu em duas etapas. Em primeiro lugar, os dispositivos legais retornados na forma de triplas RDF foram submetidos à apreciação de dois advogados. Cada um deles, pré-classificou cada tripla retornada como **relevante** ou **não relevante** em relação ao respectivo tema de busca.

Após a pré-classificação feita pelos advogados, a relevância dos resultados foi calculada com base nas métricas de Precisão, Cobertura e Medida-F, comumente utilizadas na área de Recuperação de Informação [Manning et al. 2008], que são definidas a seguir.

- **Precisão (Precision):** É uma medida que indica qual é a fração dos resultados retornados que são relevantes para as necessidades de informação.

- **Cobertura (Recall):** Indica qual fração de documentos relevantes na coleção foram retornados pelo sistema.
- **Medida-F (F-measure):** É utilizada para balancear eventuais distorções oriundas do cálculo da precisão ou cobertura. A Medida-F (F-measure) é a média harmônica entre as duas medidas anteriores.

## 4.2. Resultados

A classificação de relevância das triplas retornadas foi compilada, juntamente com os cálculos das métricas, e estão representados na Figura 5.

RESULTADOS DE BUSCA POR TEMAS VARIADOS NO REPOSITÓRIO DE TRIPLAS LEGAIS								
Tópicos pesquisados	Sigla	Triplas retornadas	Classificação Especialista 1			Classificação Especialista 2		
			Precisão	Cobertura	Medida-F	Precisão	Cobertura	Medida-F
Acumulação cargo público	ACP	20	0.50	1.00	0.67	0.70	1.00	0.82
Direito consumidor	DCO	213	0.11	1.00	0.19	0.10	1.00	0.19
Direitos humanos escola	DHE	13	0.46	0.67	0.55	0.00	0.00	0.00
Fraude financeira	FFI	14	0.57	1.00	0.73	0.50	1.00	0.67
Acesso informação	AIN	255	0.36	1.00	0.53	0.36	1.00	0.53
Meios ilícitos	MIL	8	0.75	0.83	0.79	0.88	1.00	0.93
Prevaricação	PRV	47	0.79	1.00	0.88	0.85	1.00	0.92
Programa computador registrado	PCR	10	0.40	1.00	0.57	0.30	1.00	0.46
Violência doméstica	VDO	94	0.60	1.00	0.75	0.52	1.00	0.69
Tratamento desumano	TRD	32	0.50	1.00	0.67	0.44	1.00	0.61
	<b>Média</b>	<b>71</b>	<b>0.50</b>	<b>0.95</b>	<b>0.63</b>	<b>0.46</b>	<b>0.90</b>	<b>0.58</b>

**Figura 5. Resultados: Precisão, Cobertura e Medida-F**

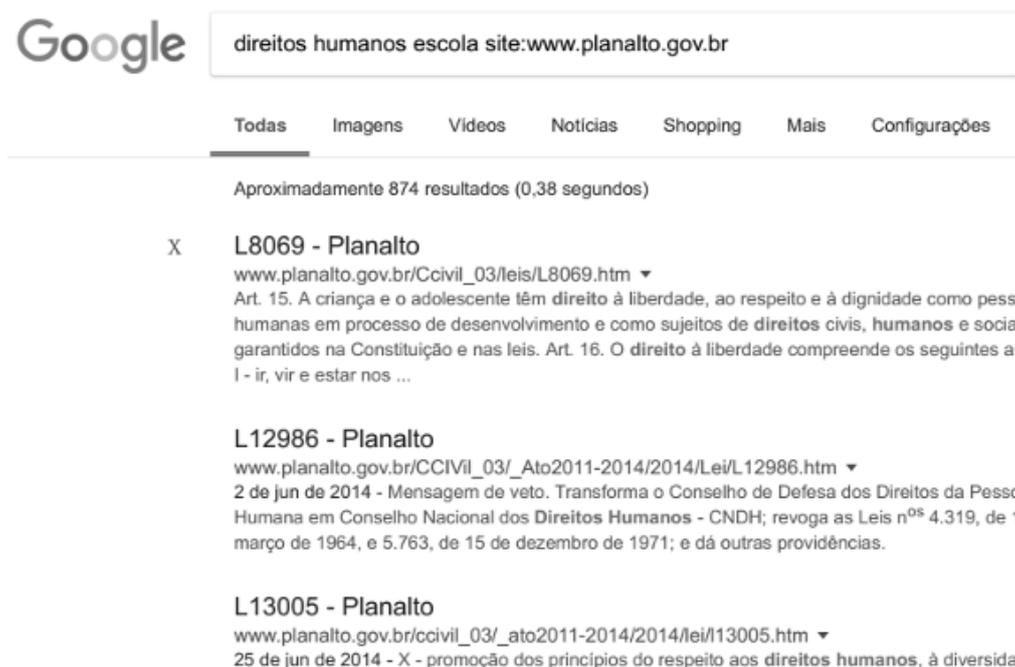
Os dados da figura foram utilizados para o cálculo da correlação da classificação feita pelos especialistas. Este cálculo objetiva verificar se os dados permitem uma classificação consistente. Eventuais divergências na classificação, observáveis na tabela, podem ter diversas causas, entre elas a própria subjetividade do classificador. O resultado da correlação é apresentado no final desta seção.

Com o propósito de obter mais evidências a respeito dos benefícios do método apresentado, as consultas anteriormente realizadas foram replicadas no buscador do Google ([www.google.com/search](http://www.google.com/search)). Para fazer as buscas no mecanismo do Google, os seguintes critérios foram adotados:

- **Indicação do site de busca** - para restringir a pesquisa às páginas contendo preferencialmente documentos legais, o site de busca foi restrito ao site oficial [www.planalto.gov.br](http://www.planalto.gov.br). Com isso, tentou-se eliminar o aspecto multidomínio do buscador. Isso foi feito por meio do uso de *site:www.planalto.gov.br* junto aos termos de busca no campo de pesquisa do buscador.
- **Critério de relevância** - a classificação feita pelo especialista 1 foi adotada, aleatoriamente, como critério de relevância.
- **Limitação da classificação de relevância** a um subconjunto das páginas retornadas em cada busca - foram consideradas as primeiras dez páginas retornadas. Para o cálculo da precisão, o quociente utilizado foi o menor valor entre 10 e o número de triplas retornadas conforme a Figura 5. Para o cálculo da cobertura, o quociente utilizado foi o menor valor entre 10 e o número de resultados relevantes classificados pelo especialista 1. Por exemplo, a busca por *Direitos humanos escola* retornou 13 triplas na consulta SPARQL, sendo que 6 delas foram

classificadas como relevantes pelo especialista 1. Então foi utilizado o valor 10 como denominador para o cálculo de precisão,  $\min(10, 13)$ . E o denominador 6 foi destacado para o cálculo de cobertura,  $\min(6, 10)$ .

A Figura 6 mostra um trecho do retorno de busca no Google. De acordo com especificação anterior, apenas os primeiros dez itens retornados foram considerados para a aplicação das métricas. Aqueles documentos resultantes da busca que, segundo o especialista no domínio legal, foram classificados como relevantes, receberam uma marcação ao lado esquerdo do item correspondente. Esta marcação foi feita por meio do símbolo **X** como se observa na figura.



**Figura 6. Retorno parcial de busca no Google.**

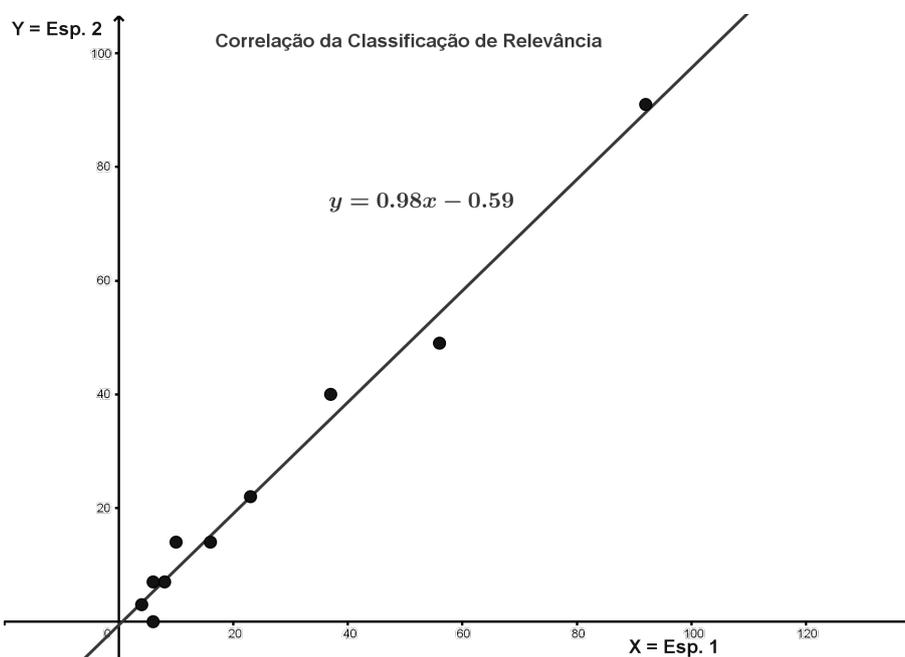
As métricas aplicadas aos resultados das buscas efetuadas no Google, considerando-se os critérios e restrições acima, são apresentadas na Figura 7. Além dos resultados específicos para o Google, a figura também mostra a Medida-F obtida na classificação do especialista 1 para efeito de comparação.

Os mecanismos de buscas em geral, devido às suas características multidomínio, não são bons parâmetros para comparações. Por esse motivo, no presente experimento, houve a necessidade de restringir as buscas ao domínio legal conforme destacado acima em *Indicação do site de busca*.

A Figura 8 apresenta graficamente o resultado do cálculo da correlação das classificações feitas pelos especialistas. O propósito é verificar se os dados obtidos nos retornos das consultas às triplas legais permitem uma classificação consistente. O resultado da medida estatística foi uma correlação positiva, no valor de 0,98. Esse resultado corrobora a intenção de que os itens legais retornados na consulta, pela aplicação do método proposto, fornecem subsídios para a classificação da relevância.

Sigla	Tópicos pesquisados	Google máx 10 docs	Classif. Espec. 1	Precisão Google	Cobertura Google	Medida-F Google	Medida-F Espec. 1
		# Relevantes	# Relevantes (máx=10 docs)				
ACP	Acumulação cargo público	3	10	0.30	0.30	0.30	0.67
DCO	Direito consumidor	2	10	0.20	0.20	0.20	0.19
DHE	Direitos humanos escola	3	6	0.30	0.50	0.38	0.55
FFI	Fraude financeira	1	8	0.10	0.13	0.11	0.73
AIN	Acesso informação	3	10	0.30	0.30	0.30	0.53
MIL	Meios ilícitos	3	6	0.38	0.50	0.43	0.79
PRV	Prevaricação	8	10	0.80	0.80	0.80	0.88
PCR	Programa computador registrado	3	4	0.30	0.75	0.43	0.57
VDO	Violência doméstica	2	10	0.20	0.20	0.20	0.75
TRD	Tratamento desumano	5	10	0.50	0.50	0.50	0.67
<b>Média</b>				<b>0.34</b>	<b>0.42</b>	<b>0.36</b>	<b>0.63</b>

**Figura 7. Resultados das buscas no Google.**

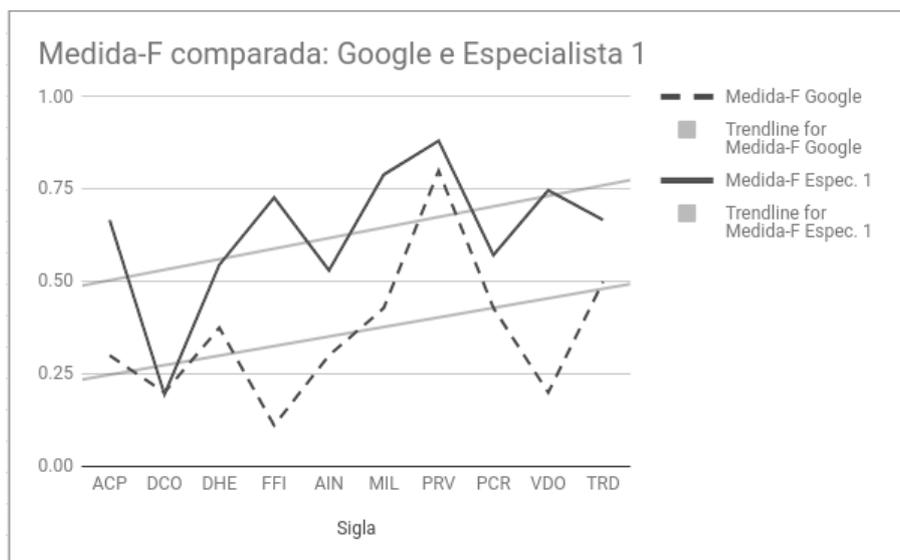


**Figura 8. Correlação das classificações dos especialistas**

Os resultados da métrica de Medida-F calculada comparativamente para o Especialista 1 e para o Google são apresentados na Figura 9. As linhas de tendência das duas medidas evidenciam a vantagem do método adotado nesta pesquisa sobre os resultados do buscador mais usado na atualidade.

## 5. Conclusão

Essa pesquisa abordou o problema de recuperar o arcabouço legal por meio da adição das remissões externas aos textos legais. O caminho para viabilizar este objetivo passou pela utilização de uma ontologia legal e conceitos e tecnologias da web semântica. O projeto e implementação de um método para coletar os textos legais, convertê-los em triplas RDF e armazená-las num triplestore, também fez parte da trilha. A complementação do caminho se deu pelas consultas ao repositório em linguagem SPARQL que possibilitam a recuperação do arcabouço legal. Os resultados foram avaliados por meio do uso de métricas do campo da Recuperação de Informação. Os resultados da pesquisa apresen-



**Figura 9. Gráfico de Resultados: Google**

taram Medida-F, em média, igual ao dobro daquelas obtidas no buscador do Google, que foi ajustado para focar no domínio legal.

A conversão dos documentos legais para o formato RDF precisa de aperfeiçoamentos. Os problemas concentram-se nos conteúdos dos arquivos em HTML. A língua portuguesa e sua acentuação característica trouxe dificuldades à execução do projeto. Outro problema encontrado foi a ocorrência de documentos legais mal formados, com marcação incompleta, com erros de digitação ou com nomenclatura não padronizada. O software construído para efetuar as consultas tem vários pontos a melhorar ou evoluir. Por isso tudo, vislumbra-se uma série de possibilidades de melhoria, complementação ou evolução do presente trabalho.

A automatização da coleta dos arquivos com os textos legais é importante para a manutenção e atualização da base de triplas RDF, assim como a melhoria do conversor ( File Parser) é outra necessidade. Tanto a interpretação sintática dos documentos legais originais quanto sua representação por meio de triplas devem ser melhoradas. Também, requer-se o aprimoramento da ontologia legal, mediante novas extensões. Outro aspecto a considerar em relação ao presente trabalho, seria complementá-lo com os resultados de pesquisas nos campos da recuperação de informação e do processamento de linguagem natural. O campo da semântica dos conteúdos revela-se promissor. Recuperar o arcabouço legal de maneira que os dispositivos de semântica similar aos argumentos de busca sejam retornados é tarefa em aberto.

## 6. Agradecimentos

É importante destacar e agradecer a contribuição dos advogados, Dr. Joaquim Carlos Paixão Júnior e Dr. Danilo Bueno Berber, que participaram ativamente efetuando a classificação da relevância dos resultados das buscas por dispositivos legais.

## References

- [Baeza-Yates 2013] Baeza-Yates, R.; Ribeiro-Neto, B. (2013). *Recuperação de Informação: conceitos e tecnologia das máquinas de busca*. Bookman, Porto Alegre, 2a edition.
- [Frosterus 2013] Frosterus, Matias; Tuominen, J. H. E. (2013). Facilitating re-use of legal data in applications - finnish law as a linked open data service. page 10. Extended Semantic Web Conference, Springer Berlin Heidelberg.
- [Lima 2008] Lima, J. A. O. (2008). *Modelo Genérico de Relacionamentos na Organização da Informação Legislativa e Jurídica*. Doutorado em ciência da informação, Universidade de Brasília - UnB, Brasília.
- [Machado 2013] Machado, A. L. (2013). *Modelo Conceitual Formal de Relacionamentos do Ordenamento Jurídico Positivo*. Doutorado em ciência da computação, Instituto Tecnológico de Aeronáutica - ITA, São José dos Campos.
- [Machado 2014] Machado, A. L.; Parente de Oliveira, J. M. (2014). A legal ontology of relationships for civil law system. In *Proceedings of 1st Joint Workshop ONTO.COM / ODISE on Ontologies in Conceptual Modeling and Information Systems Engineering-Co-located with 8th International Conference on Formal Ontology in Information Systems (FOIS 2014)*, Rio de Janeiro, Brasil.
- [Machado 2010] Machado, R. A. C. (2010). Seeklex - recuperação de informação em subdocumentos hierarquizados. Mestrado em engenharia de sistemas e computação, Universidade Federal do Rio de Janeiro - UFRJ, Rio de Janeiro.
- [Manning et al. 2008] Manning, C. D., Raghavan, P., and Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press, New York, 1st edition.
- [PRODASEN 2016] PRODASEN (2013 (accessado em 01 de Setembro de 2016)). Lexml - rede de informação legislativa e jurídica - portal especializado em informação jurídica e legislativa. <http://www.lexml.gov.br/>.
- [Soria 2007] Soria, C; Bartolini, R. L. A. M. S. P. V. (2007). Automatic extraction of semantics in law documents. In *Proceedings of the V Legislative XML Workshop*, pages 253–266, Florence, Italy.
- [Thatmann 2014] Thatmann, D.; Schuster, E. C. G. (2014). Mapping legal requirements to slas: an ontology based approach for cloud-based service consumption. In *SW4LAW 2014 and JURIX 2014-DC Semantic Web for the Law and Second Jurix Doctoral Consortium*, pages 5–16. Technische Universität, Technische Universität.