

# PERCURSO METODOLÓGICO PARA CONSTRUÇÃO DE ONTOLOGIAS JURÍDICAS<sup>1</sup>

## METHODOLOGICAL PATHWAY FOR THE CONSTRUCTION OF LEGAL ONTOLOGIES

*Felipe Mozart Santana Nascimento<sup>2</sup>*

*Renato Fernandes Corrêa<sup>3</sup>*

*Fábio Assis Pinho<sup>4</sup>*

### RESUMO

Propõe e discute um percurso metodológico para construção de ontologias do domínio jurídico, tendo como foco o Processo Judicial Eletrônico e adotando técnicas de mineração de textos, mais especificamente de aprendizado de ontologia, além de reuso de ontologias e linguagens documentárias já existentes para o domínio. O método da pesquisa consistiu das seguintes etapas: descrição do domínio jurídico bem como o processo judicial eletrônico, a identificação e descrição de ontologias com o propósito de compreender sua arquitetura, a revisão de trabalhos aplicados de ontologias jurídicas e, por fim, a análise da aplicação de técnicas de mineração de texto. O percurso metodológico se apoiou nas camadas para construção de ontologias a partir de textos, as quais procuram definir: termos, sinônimos, conceitos, hierarquização de conceitos, relações e regras. Como resultado, o percurso metodológico se mostra exequível e coerente para construção ontologias do domínio jurídico e em particular para representação do conhecimento em torno do Processo Judicial Eletrônico.

**Palavras-Chave:** Ontologia. Sistema de Organização do Conhecimento. Mineração de Textos. Informação jurídica. Ontologia jurídica.

### ABSTRACT

Proposes and discusses a methodological path for the construction of ontologies of the legal domain, focusing on the Electronic Judicial Process and adopting techniques of text mining, specifically ontology learning, as well as the reuse of ontologies and documentary languages for the domain. The research method consists of the following steps: description of the legal domain as well as the electronic judicial process, identification and description of ontologies with the purpose of understanding its architecture, the review of applied works of ontologies and, finally, analysis of the application of text mining techniques. The methodological path was supported by the layers for the construction of ontologies based on texts, which seek to define: terms, synonyms, concepts, hierarchization of concepts, relationships and rules. As a result, the methodological path is feasible and coherent to construct ontologies of the legal domain and in particular for the representation of knowledge around the Electronic Judicial Process.

**Keywords:** Ontology. System of knowledge organization. Text mining. Legal information. Legal ontology.

Artigo submetido em 12/08/2019 e aceito em 19/09/2019

1 O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001. E pela Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE).

2 Mestre em Ciência da Informação pela Universidade Federal de Pernambuco, Brasil. Discente Pesquisador da Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco, Brasil. E-mail: mozart.felipe@gmail.com.

3 Doutor em Ciências da Computação pela Universidade Federal de Pernambuco, Brasil. Professor Adjunto da Universidade Federal de Pernambuco, Brasil. E-mail: fc\_renato@yahoo.com.br.

4 Doutor em Ciência da Informação pela Universidade Estadual Paulista Júlio de Mesquita Filho, Brasil. Professor Associado da Universidade Federal de Pernambuco, Brasil -mail: fabiopinho@ufpe.br.

## 1 INTRODUÇÃO

Percebida como uma das atividades mais influentes no campo da Ciência da Informação (CI), o exercício da Organização e Representação da Informação e do Conhecimento é justificado como um dos pilares fundamentais para desenvolvimento desta ciência e conseqüentemente valida sua legitimidade frente às preocupações dos processos de gestão de informação a ela empregada. A esses processos somam-se: seleção, extração, mineração, indexação, busca e recuperação da informação.

Em conseqüência dos processos citados acima, organizações que exploram a informação como commodities enfrentam um cenário competitivo informacional a qual aplicam parte considerável de seus investimentos em pesquisas que contemplem práticas assertivas para gestão das informações estratégicas, além da adesão de ferramentas que auxiliem nas resoluções de suas necessidades informacionais e que apoie o processo de tomada de decisão de seus usuários finais (NASCIMENTO; PINHO, 2018a). Essas ferramentas são dotadas de expertises, principalmente quando exploradas à luz da inteligência artificial ao qual contribuiu significativamente na ascensão de pesquisas relacionadas aos processos informacionais.

Diante do contexto acima, essa pesquisa justifica-se pela caracterização de uma das principais ferramentas estudadas no contexto da Ciência da Computação (CC) e CI, o sistema ontologia. Dotada de funções e propriedades que exploram os interesses organizacionais que dizem respeito à gestão da informação nas empresas que utilizam como negócio a informação jurídica, o sistema de ontologia pode explorar contribuições significativas, como: representação da informação, reutilização de conhecimento, geração de conhecimento para tomada de decisão e compreensão da estrutura das informações estratégicas de domínio jurídico (NASCIMENTO; PINHO, 2018b).

De acordo com o contexto acima, deparamo-nos com a seguinte questão de pesquisa: qual o percurso para construção de uma ontologia de domínio jurídico?

Para Batres *et al.* (2005) a ontologia permite, além das contribuições sinalizadas anteriormente, uma ampliação de um outro processo de gestão: a extração de informação dos documentos, de forma a fornecer um sistema conceitual que integra um conjunto de termos, identificando os termos mais relevantes para extração, a localização desses termos nos documentos e suas devidas relações. Valente (1995) citado por Batres *et al.* (2005, p. 77) ressalta outras contribuições importantes que justificam o uso de uma ontologia de domínio jurídico pelos seguintes pontos: (1) é inevitável o uso de ontologias pelo fato de qualquer sistema baseado no domínio jurídico ter comprometimento ontológico que determine o que pode ou não ser executado, sendo assim, para que esses comprometimentos estejam representados

de forma clara e concisa é fundamental o uso de ontologias; (2) ontologias podem ser utilizadas por dois objetivos que se concatenam, para produzir linguagens especializadas e para a representação dos formalismos do domínio jurídico; (3) a ontologia define o domínio em termos de categorias primitivas do conhecimento, e neste ponto compreende os elementos relacionais do domínio que representa; (4) há uma forte tendência para que ontologias jurídicas comecem a se interoperar e, desta forma, apresentar uma visão global das potencialidades de seu uso para explicar um fenômeno jurídico e, por fim (5) o uso de ontologia ser significativo por definir que tipos de conclusões, garantias e cadeias de argumentos são usuais ou válidos no domínio jurídico, visto que este tipo de análise é demasiadamente importante, considerando o raciocínio jurídico como produtor e avaliador de argumentos legais.

Salienta-se que a construção deste tipo de sistema exige esforços e recursos significativos para seu desenvolvimento. Esses são os motivos pelos quais a aplicação deste sistema ainda é considerada como um assunto complexo e, por isso, demanda de estudos como este, o qual reflete a necessidade de apontar o real significado e potencialidade da construção e uso de ontologias. Ainda neste sentido, um dos fatores que mais requer tempo e recursos é adquirir conhecimento do domínio, por ser visto como um trabalho que exige atenção do profissional às pesquisas na literatura ou consulta de especialistas com a intenção de identificar os termos relevantes, organizá-los hierarquicamente e identificar as relações existentes.

Na tentativa de minimizar parte desses esforços surge o processo de aprendizagem de ontologias a partir de textos, a qual compreende as fases de extração dos termos relevantes, a descoberta dos sinônimos entre esses termos, a ordenação hierárquica dos termos, ou seja, a geração de taxonomias, e por fim, estabelecer os relacionamentos existentes entre eles. Mesmo que ainda em caráter tímido, pode-se encontrar estudos já desenvolvidos que identificam este tipo de método como Zahra *et al.* (2014) e Farias e Pinho (2016), esses dois trabalhando na perspectiva dos estudos médicos e Batres *et al.* (2005), Nascimento e Siebra (2015) e Nascimento e Pinho (2018a, b) na perspectiva dos estudos jurídicos.

Portanto, considera-se como objetivo desta pesquisa propor e discutir um percurso metodológico para construção de ontologias do domínio jurídico, tendo como foco o Processo Judicial Eletrônico, adotando técnicas de mineração de textos, mais especificamente de aprendizado de ontologia.

## 2 METODOLOGIA

De acordo com as buscas por produções científicas nas bases: Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (BRAPCI), Google Acadêmico e a Biblioteca Digital

Brasileira de Teses e Dissertações (BDTD), usando como termos de busca “ontologia de domínio jurídico”, “ontologia jurídica” e “ontologia legal” entre os anos 2010 e 2019, evidenciaram-se os estudos mais recentes, nos quais apontam o uso deste sistema e como ele contribui para gestão e organização da informação em domínio jurídico brasileiro. Ainda como estratégias de buscas, foram utilizados os termos “ciência da informação” e “ciência da computação”, aos quais acompanharam os termos citados anteriormente. Justificam-se as bases consultadas pelo domínio jurídico brasileiro possuir caráter e especificações próprias que regem as ordens executiva, legislativa e judiciária.

O resultado das buscas totalizou cerca de 20 pesquisas, entre elas as que abordam a ontologia no contexto epistemológico e filosófico do Direito. Entretanto, as pesquisas relacionadas ao objeto investigativo deste trabalho estão representadas no quadro 1, as quais constam como alicerce para desenvolvimento da análise e discussão da proposição do percurso metodológico para construção de ontologia jurídica. Justifica-se como seleção das pesquisas em destaque o potencial estudo técnico orientado às ontologias na visão da Ciência da Informação, ou seja, como um sistema de organização do conhecimento e, também, por explorarem as características oriundas do domínio jurídico brasileiro, conforme estão caracterizadas no quadro abaixo:

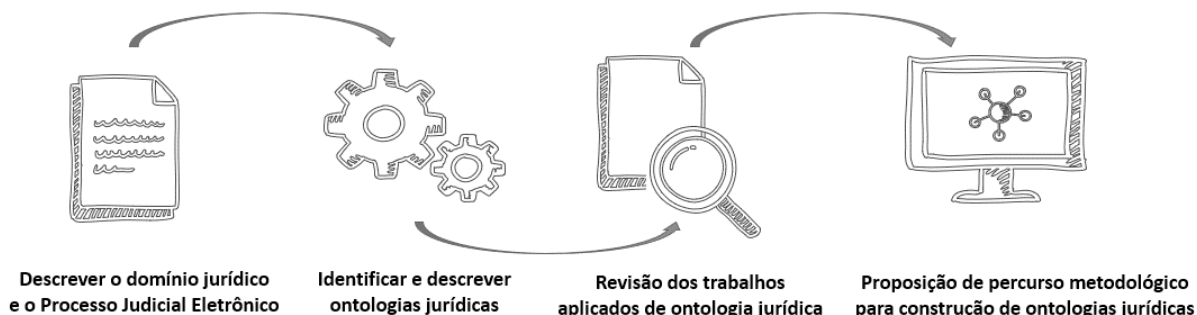
**Quadro 1** - Pesquisas relevantes recuperadas.

<b>Autores</b>	<b>Título</b>	<b>Objetivo</b>
Beccalli (2018)	UFO-L: uma ontologia núcleo de aspectos jurídicos construídas sob a perspectiva das relações jurídica	Definir a fundamentação teórica para relações jurídicas e representá-la por meio de uma ontologia núcleo baseada na Teoria dos Direitos Fundamentais de Robert Alexy e fundamentada na Ontologia de Fundamentação Unificada (UFO) proposta em Guizzardi.
Minghelli (2011)	A relação de meronímia em uma ontologia jurídica	Investigar o papel da relação de meronímia como uma relação organizadora do conteúdo da área do Direito Processual Penal, especialmente do procedimento do tribunal do júri, a fim de inseri-los em uma ontologia jurídica linguística.
Nascimento e Siebra (2015)	Organização da Informação Jurídica: A Ontologia no Contexto do Business Intelligence	Contribuir com o desenvolvimento de uma solução para aprimorar o processo de recuperação, armazenamento e compartilhamento de informações jurídicas, no contexto da inteligência de negócios, por meio do planejamento, construção e validação de uma ontologia para a área jurídica.
Constâncio (2017)	Ontologia para um motor de busca semântica para recuperação jurisprudencial no Brasil	Elaborar uma ontologia que suporte um Motor de Busca Semântica que possa reduzir o Problema da Recuperação da Informação no domínio da consulta jurisprudencial no Brasil.
Nascimento e Pinho (2018a)	Ontologia na gestão de conhecimento jurídico	Identificar a contribuição do uso de ontologias na gestão do conhecimento jurídico.
Nascimento e Pinho (2018b)	Uso estratégico de ontologia para organização e gestão da informação da informação jurídica	Identificar as contribuições do uso estratégico de ontologias para organização e gestão da informação jurídica.

**Fonte:** elaborada pelos autores.

Adiante, foram adotados como percurso investigativo os processos representados na figura 1, ao qual contribuiriam diretamente para o alcance do objetivo que se propôs nesta pesquisa:

**Figura 1** - Percurso investigativo da pesquisa.

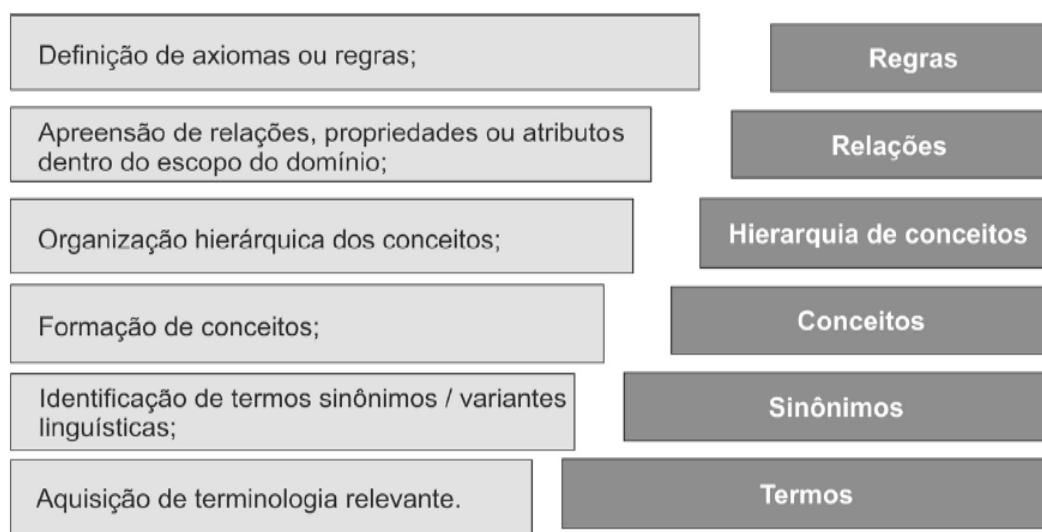


Fonte: elaborado pelos autores.

O primeiro passo, de *descrever o domínio jurídico e o Processo Judicial Eletrônico*, nos leva a compreensão da natureza da informação e documentação jurídica. Em seguida, a *identificação e descrição de ontologias jurídicas* teve a finalidade de caracterizar o sistema no domínio em estudo. A *revisão dos trabalhos aplicados de ontologias* buscou a categorização das pesquisas de acordo com a natureza do domínio e a tarefas de mineração de textos realizadas nelas.

O delineamento do aprendizado de ontologias por camadas baseou-se em Schiessl e Brascher (2011, p. 308). Conforme apresentado na figura 2, a seguir.

**Figura 2** - Aprendizado de ontologias em camadas.



Fonte: Schiessl e Brascher (2011, p. 308).

Ao que concerne às camadas apresentadas na figura 2, Schiessl e Brascher (2011, p. 308 – 309) aportados das reflexões de Buitelaar *et al.* (2005), descrevem as camadas para construção de ontologias e como ela está aliada às técnicas de mineração de texto:

- **Termos:** a extração de termos é pré-requisito para o aprendizado de ontologias a partir de textos. Os termos são as expressões linguísticas que possibilitam a identificação de conceitos e relações e, portanto, são os pilares para construção base em que se apoia todo o processo de desenvolvimento.
- **Sinônimos:** a identificação de sinônimos consiste em descobrir termos que denotam o mesmo conceito, ou seja, são considerados sinônimos os termos que compartilham do mesmo significado.
- **Conceitos:** considerando o aprendizado de ontologias a partir de textos, os autores aportam-se de Cimiano (2006) para compreender as propriedades pertencentes a conceitos, tais como a intensão do conceito, a extensão do conceito e o conjunto de realizações linguísticas, isto é, de termos para conceitos.
- **Hierarquia de conceitos:** como já citado anteriormente, o processo de hierarquização é definido por meio de taxonomias, em que são utilizadas para organizar o conhecimento ontológico empregando relações de generalização e especialização por meio das quais o conceito de herança é aplicado.
- **Relações:** As relações não taxonômicas referem-se a qualquer relação entre conceitos, exceto a relação “é-um”, quais sejam partitivas, de oposição, funcional e de identidade. Ainda na camada de relações pode-se apontar uma subcamada conceituada hiponímia e hiperonímia: é a relação existente entre uma palavra de sentido mais específico e outra de sentido mais genérico, que tem com a primeira, traços semânticos comuns (BUIELAAR, *et al.*, 2005).
- **Regras:** regras ou axiomas são utilizados para modelagem de sentenças que são sempre verdadeiras. Elas podem ser incluídas em uma ontologia com o propósito de restringir, de verificar a informação do domínio ou de deduzir uma nova.

Ainda como delineamento metodológico, além do apoio das análises bibliográficas, foi abordado o uso do *software Protégé*<sup>5</sup>, na tentativa de evidenciar a estrutura das propriedades e funções manifestadas nos processos judiciais. Por fim, é desenvolvida uma proposta de percurso metodológico para construção de ontologia jurídica frente às técnicas de mineração de textos.

5 <https://protege.stanford.edu/>

### 3 INFORMAÇÃO E DOCUMENTAÇÃO JURÍDICA BRASILEIRA

A natureza da informação é classificada por meio de sua área de domínio, nesta seção é discutido sobre a informação em domínio jurídico. Definida por Guimarães (1999, p. 11), a informação jurídica é “o conjunto de conteúdos pertencentes ao universo conceitual da Ciência Jurídica, que se expressam por meio das formas documentárias reveladoras da atuação jurídica”, é por meio desta consideração que se justifica o tratamento complexo entre informação e documento baseados sob o domínio jurídico.

A ideia de informação jurídica é compreendida por Silva (2010, p. 3) como “todo e qualquer conhecimento produzido pelo homem com o intuito de fundamentar as atividades profissionais desenvolvidas pelos operadores do Direito”. A documentação jurídica é determinada por meio do corpus teórico que fundamenta a construção e constituição documental, devendo ser considerado meio, conteúdo e forma. É necessária, também, a observação quanto à percepção destes documentos representadas em formas conceituais, estruturais e funcionais, sejam essas informações em suporte tradicional ou em formato digital.

De acordo com o quadro conceitual acima, em que trabalham as compreensões de informação e documentação jurídica, Nascimento (2018) ordenou, com base em outros autores (MIRANDA, 2004; PASSOS; BARROS, 2009), em um tripé constitucional, os tipos de informação e documento deste domínio considerando as categorias jurídicas nos quais se manifestam. O autor buscou, além da conceituação, a classificação tipológica para cada categoria, sendo essas: *doutrina*, *legislação* e *jurisprudência*, conforme a figura 3:

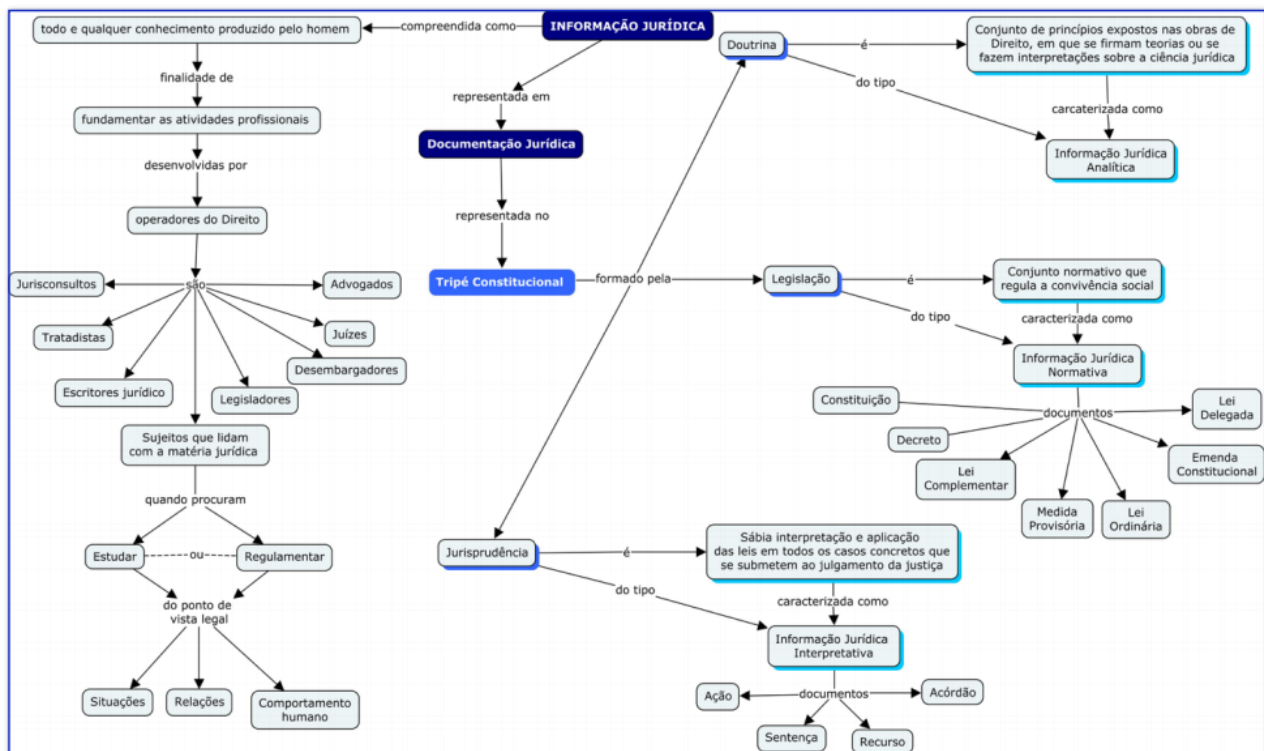
**Figura 3** - Tipos de informação e documentação jurídica.



Fonte: Nascimento (2018, p. 25).

No sentido de melhor compreender esta dinâmica, Nascimento, Braz e Pinho (2018) desenvolveram mapas conceituais com o objetivo de representar o domínio jurídico. Assim, a figura 4 trata sobre informação e documentação jurídica brasileira.

**Figura 4 -** Mapa conceitual da representação da informação jurídica.



Fonte: Nascimento, Braz e Pinho (2018, p. 18)

Como exemplo dos documentos jurídicos que circulam em meios institucionais como, por exemplo, os ramos do Poder Judiciário – Federal, Estadual e Trabalhista, encontra-se o Processo Judicial Eletrônico (PJE)<sup>6</sup>, documento alvo de estudo desta pesquisa. Fundamental destacar, conforme Zamur Filho (2011), que o PJE foi autorizado pela Lei nº 11.419/06 e deve ser entendido em um cenário mais amplo e sistêmico, ou seja, considerar que a aplicação da tecnologia para o desenvolvimento do processo e a jurisdição que lhe são prestadas, o arcabouço legal e regulamentar que sustenta seu desenvolvimento estão sendo estudado no sentido de ampliar e reorganizar em suas concepções sistemáticas e o limite de seu alcance.

6 [www.pje.jus.br/](http://www.pje.jus.br/)



O desenvolvimento e aplicação do PJE foi antecedido por um outro sistema eletrônico – o PROJUDI<sup>7</sup>, que foi o primeiro meio de informatizado dos processos judiciais desenvolvido pelo Conselho Nacional de Justiça (CNJ)<sup>8</sup>. Contudo, com o alto volume de modificações feitas por particularidades de cada tribunal, o CNJ optou por uma padronização controlada e coordenada pelo próprio conselho. Nesse sentido, partiu-se para o desenvolvimento do PJE, tornando como foco do CNJ seu amadurecimento no objetivo de torná-lo um meio unificado e que também considera as particularidades de cada tribunal (SOUZA NETO, 2014).

O alto volume de expedição de informações e documentações jurídicas, tal como os volumes processuais demandados atualmente, implica em métodos orientados à sua organização, uma vez considerando os processos judiciais como negócios para grandes departamentos jurídicos e escritórios de advocacia, o que coopera diretamente para as atividades dos operadores do Direito, tais como advogados, juízes, juristas, tratadistas, desembargadores e todos os indivíduos que estudam a informação e documentação jurídica. Esses métodos são alvo de estudos dessas organizações ao qual aplicam parte considerável de seus investimentos em pesquisas com o objetivo de desenvolver soluções que atendam suas necessidades informacionais. Consideravelmente, a ontologia é um dos sistemas que pode ser usado para as atividades de gestão e organização das informações estratégicas, entre essas atividades considera-se as de mineração de texto. Ontologia e mineração de textos são temas discutidos a seguir.

#### **4 ONTOLOGIA E MINERAÇÃO DE TEXTO: CONCEITUAÇÕES INTRODUTÓRIAS**

O termo ontologia, em seus aspectos conceituais, tem origem nas ciências filosóficas, um campo que busca compreender o conhecimento dos princípios e fundamentos dos seres e suas realidades. Para Nascimento (2018, p. 39) o termo é utilizado demasiadamente em algumas ciências, como na Filosofia, na Ciência da Computação e na Ciência da Informação, entre outras. Almeida (2014) apresentou um quadro, ilustrado abaixo, em que trata do sentido do uso do termo nas ciências mencionadas, trabalhando o conceito por meio de suas definições, objetivos e exemplos:

---

7 [www.cnj.jus.br/sistemas/projudi](http://www.cnj.jus.br/sistemas/projudi)

8 <https://www.cnj.jus.br/>

**Quadro 2:** Aplicação do conceito de ontologia na Filosofia, Ciência da Computação e Ciência da Informação

Distinção	Campo	O que é?	Propósito	Exemplo
Ontologia como uma disciplina	Filosofia	Ontologia como um sistema de categorias	Entender a realidade, as coisas que existem e suas características	Sistemas de Aristóteles, Kant, Husserl
Ontologia como um artefato	Ciência da Computação	ontologia como uma teoria (baseada em lógica)	Entender um domínio e reduzi-lo à modelos	BFO, DOLCE (genéricas)
		ontologia como um artefato de software	Criar um vocabulário para representação em sistemas e para gerar inferências	OWL (linguagem de RC)
	Ciência da Informação	ontologia como uma teoria (informal)	Entender um domínio e classificar termos	Sistema de classificação de Ranganathan
		ontologia como um sistema conceitual informal	Criar vocabulários controlados para recuperação da informação a partir de documentos	um catálogo, um glossário, um tesouros

Fonte: Almeida (2014, p. 52).

Na tentativa de reunir os conceitos trabalhados em cada uma das ciências vistas no quadro acima, Nascimento (2018, p. 44) define ontologia como a representação do conhecimento de um domínio, realizada por indivíduos em computadores, em que o conhecimento representado é um entendimento consensual para a comunidade a qual é orientada.

A ontologia na visão da Ciência da Informação é tratada como um dos Sistemas de Organização do Conhecimento, tal como tesouro, taxonomia, folksonomia, mapas conceituais e até mesmo os sistemas de classificação bibliográfica. Como um sistema a ontologia comporta-se de acordo com os seus elementos e a forma na qual estão configurados. Os elementos básicos, digamos, são: as classes, os relacionamentos entre as classes, os axiomas definidos no momento do planejamento da ontologia, as instâncias e as funções definidas para o domínio que se quer alcançar. Ainda sobre esses elementos, para Nascimento (2018, p. 58) “os axiomas são usados para modelar as sentenças, as instâncias usadas para representações dos objetos e as funções que são tidas como os relacionamentos especiais onde um objeto é único para outros objetos precedentes”.

A construção de um sistema como ontologia prevê uma estrutura baseadas em outros sistemas como tesouro e taxonomia. Quanto ao tesouro considera-se sua aplicação na ontologia devido aos estudos terminológicos que cercam o domínio, quanto à taxonomia considera-se sua aplicação na ontologia pela estrutura (esqueleto) que se configura o ordenamento das classes e subclasses no sistema, tal como seus ordenamentos por meio da hierarquização.

Braga (2016, p. 177) aponta que o “processo de construção e manutenção de uma taxonomia, quando envolve grandes coleções de textos, demanda tempo e é custoso, tornando-se extremamente complexo.”

Sendo assim, a autora explorou o processo de geração semiautomática de taxonomia automática, estrutura comum a todos os sistemas de ontologias, utilizando técnicas de mineração de textos.

Sanga (2017, p. 38 -39) se apóia de autores como Hotho, Nürnberger e Paaß, (2005) e Kaur e Aggarwal (2013) para caracterizar a técnica de mineração de texto, o qual a considera como descoberta de conhecimento a partir de textos. Para Sanga, a técnica de mineração de textos “utiliza técnicas de recuperação de informações e de processamento de linguagem natural em conjunto com algoritmos e métodos de descobrimento de conhecimento em banco de dados, mineração de dados, aprendizado de máquinas e estatísticas”. Destarte, a mineração de texto é tida pelo autor como uma “técnica emergente no campo de mineração de dados”.

A técnica em questão surgiu “a partir da necessidade de se descobrir, de forma automática, informações em textos” (SANGA, 2017, p. 25). Nesse sentido, as funcionalidades desta técnica fornecem subsídios para solucionar as necessidades de informações para os sistemas de recuperação da informação baseados em ontologia, tidos nessa pesquisa com um olhar para as soluções de domínio jurídicos.

A dinâmica da relação de ontologia e mineração de textos será discutido na seção a seguir, configurando perspectivas da caracterização deste sistema na mineração de textos jurídicos, tendo como elemento documental o PJE.

## 5 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Para Sanga (2017, p. 25-26) apenas o esforço de equipes de pessoas, ou seja, os operadores de Direito, não é suficiente para as atividades de preparação, análise e classificação de informações, promovendo como motivação, para diminuir tais esforços, a técnica de mineração. O autor considera que “a utilização de algoritmos e técnicas automatizam o processo de preparação do ambiente, deixando apenas o trabalho de análise das informações por parte dos especialistas para descobrir novos fatos que possam ser utilizados em tomadas de decisões”.

Sendo assim, a caracterização de ontologia às técnicas de mineração de textos jurídicos possui potencialidades inteligentes no que diz respeito às atividades de gestão estratégica de informações jurídicas, principalmente ao que contribui para o descobrimento de novos conhecimentos. Tão logo, apresentam-se a reunião das propriedades do sistema ontologia por mineração de texto de processos jurídicos eletrônicos.

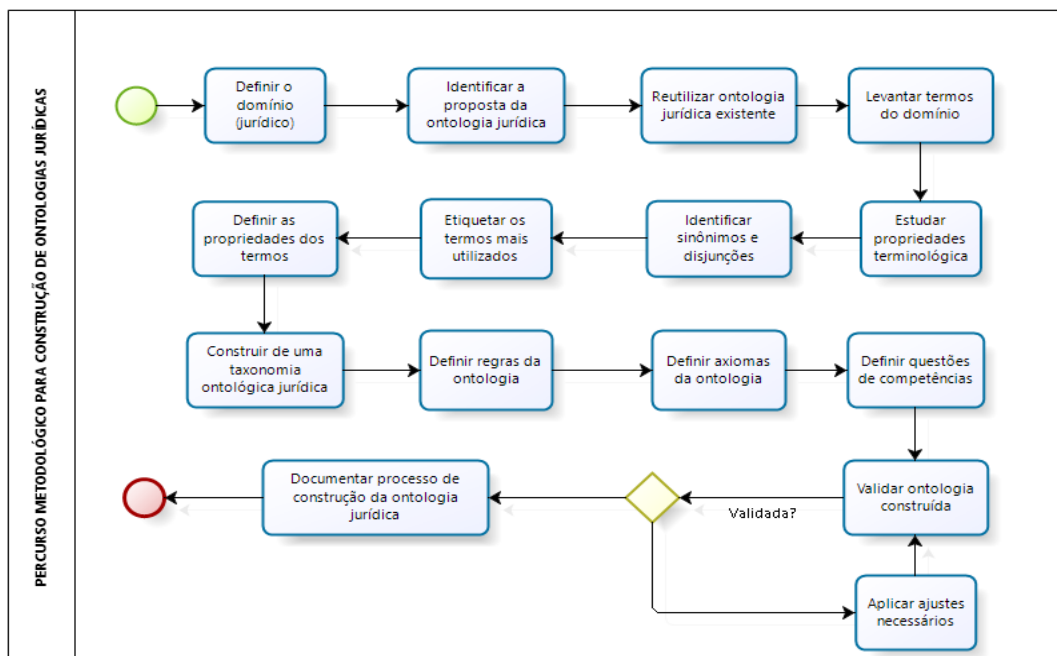
Na finalidade de compreender os processos para mineração de texto, Gajze (2010, p. 223-224), citado por Bezerra e Guimarães (2014, p. 135), apresenta as seguintes etapas: *transformação do texto*

(a), a qual consiste na transformação do documento para um formato de texto; *separação de palavras* (b), tratamento para identificar as características importantes do texto; *stemming* (c), tratamento dos termos por meio da remoção de sufixos e prefixos e a identificando nos formatos mais puros; *matriz de frequência* (d), categorização dos termos associando-os às frequências de ocorrência nos textos.

A construção de ontologias é alvo de pesquisas que objetivam utilizá-las em sistemas de informação e o grande desafio está na prospecção de conhecimento e no investimento de recursos como tempo e custo para construí-la e integrá-las a outros domínios e aplicações. Para desenvolver uma ontologia um dos principais processos é a entrada de dados, ou seja, textos, os quais são fundamentais para que a ontologia possa ser aprendida via extração de termos, seus respectivos conceitos e as possíveis relações entre eles (SCHIELL; BRASCHER, 2011). De acordo com os autores, apoiados em Buitelaar *et al.* (2005), as atividades para desenvolver uma ontologia devem possibilitar a automatização dos processos necessários para sua construção.

Para caracterização das etapas utilizou-se o software *Bizagi Process Modeler*<sup>9</sup> com o objetivo de modelar as atividades que compõe o processo, conforme figura 5:

**Figura 5** - Percurso metodológico para construção de ontologias jurídicas.

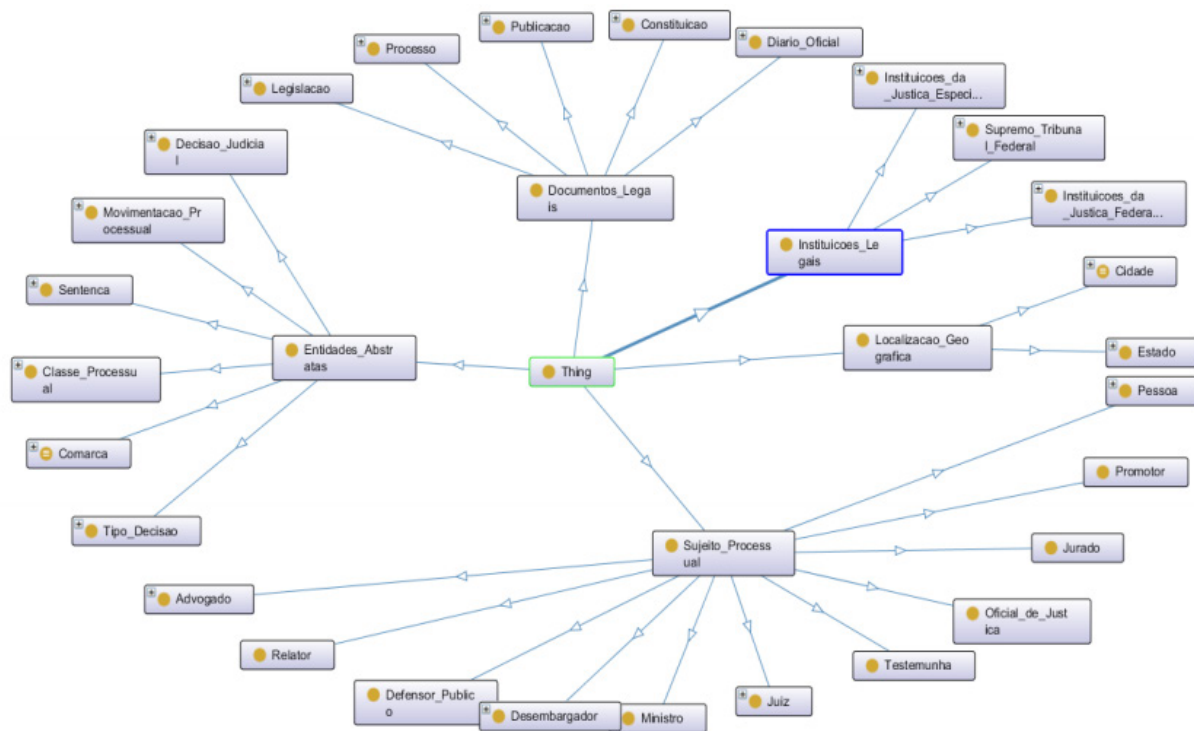


Fonte: elaborado pelos autores.

O aprendizado de um sistema como ontologia está baseado em camadas que integram sua construção, tais como: a camada de termos e de conceitos, bem como a de sinônimos, a de hierarquização dos conceitos, a das relações do domínio e a das regras da ontologia, conforme ilustrado na figura 2 apresentada na seção “Metodologia”, baseada em Schiessl e Brascher (2011).

A manifestação dos elementos ontológicos e as técnicas de mineração de textos podem ser simuladas em uma ontologia real. Em princípio é relevante entender como está formado um processo jurídico, essa representação está ilustrada na figura 6, pela ontologia jurídica apresentada na pesquisa de Nascimento e Siebra (2015):

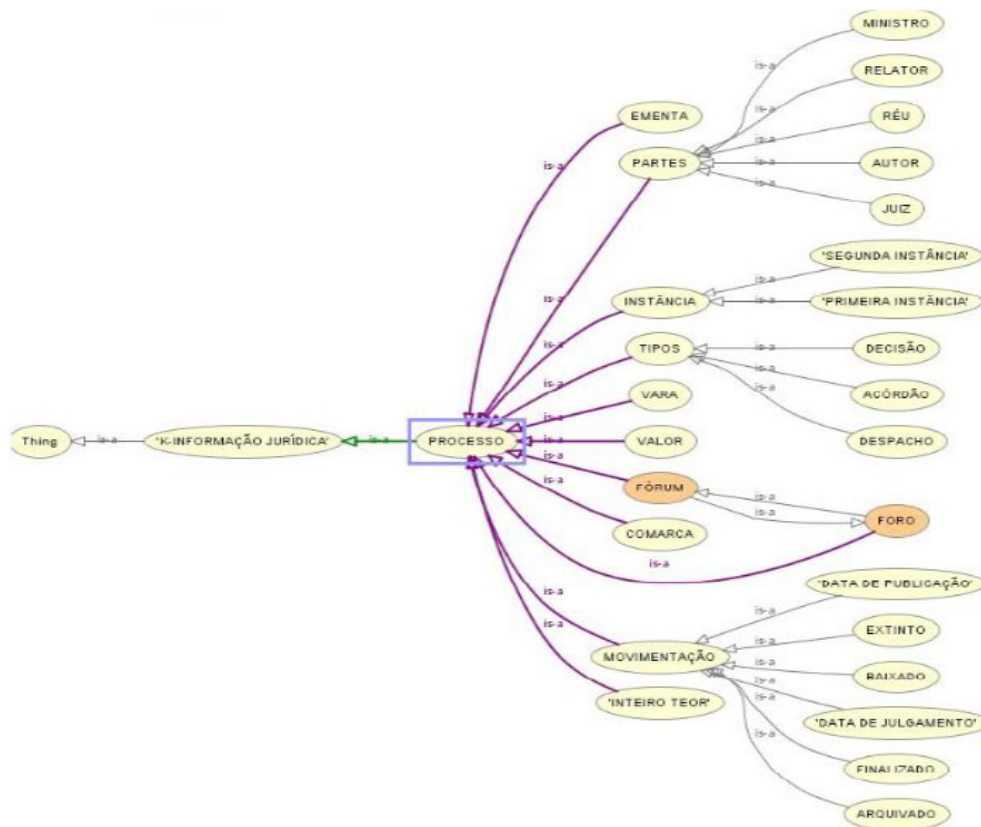
**Figura 6** - Integração de classes de uma ontologia jurídica.



**Fonte:** Nascimento e Siebra (2015, p. 463).

Ao aplicar um filtro para compreender especificamente as propriedades do processo jurídico tem-se o esquema a seguir:

**Figura 7** - Esquema de montagem da ontologia relacionada ao processo judicial.



Fonte: Nascimento (2015, p. 58).

Como visto na imagem acima, os elementos de um processo são basicamente: *ementa*, os *sujeitos* envolvidos, ou seja, as *partes de um processo*, a *instância* em que esse processo tramita, o *tipo de processo* relacionado, ou seja, a *temática jurisdicional* a qual pertence, o *ambiente de distribuição processual*, como vara, comarca e tribunal e a *situação processual* em que o documento se encontra.

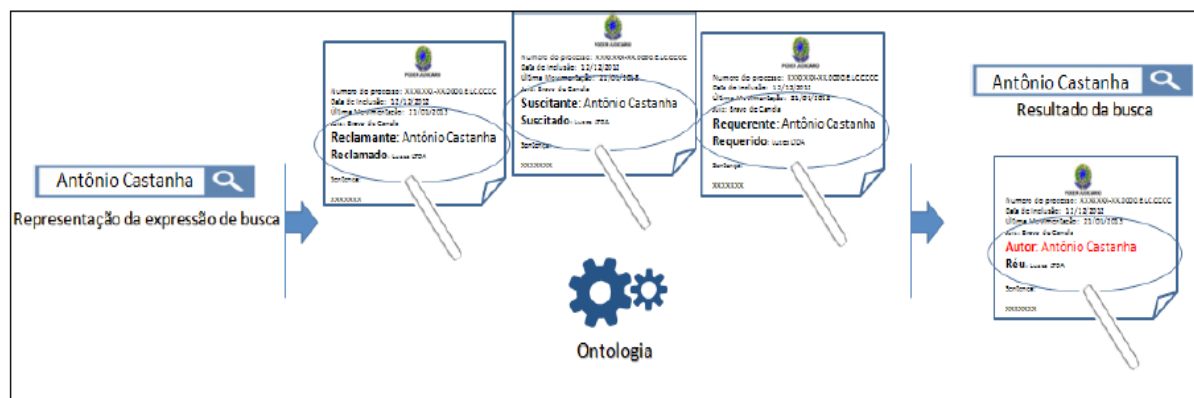
Nesse sentido, as etapas de aprendizado de ontologias, aliada às técnicas de mineração de textos, podem ser apontadas abaixo (SCHIESSL; BRASCHER, 2011, p. 308 – 309).

- **Camada de termos:** Como exemplo de suporte para atender esta necessidade temos as ferramentas de Processamento de Linguagem Natural (PLN) atuais, ao qual apresentam resultados satisfatórios na identificação de termos a partir de textos. Essas ferramentas são a combinação de processamentos estatísticos e análises linguísticas. As técnicas mais utilizadas são: *parsers* para extração de termos, *POS tagger* (etiquetadores), algoritmos de identificação de palavras-chave, distribuição estatística de termos e testes estatísticos para identificação de termos compostos. No

documento jurídico processual tem-se como estratégia de etiquetagem os termos mais buscados pelos usuários em uma solução jurídica de recuperação de informação. Essa estratégia tem como propósito reconhecer as necessidades informacionais mais latentes dos usuários do sistema, neste processo pode-se considerar os métodos de indexação automática.

- **Camada de sinônimos:** As principais técnicas utilizadas para descoberta de sinônimos são de agrupamento (*clustering*), *Latent Semantic Indexing (LSI)* e a utilização de dicionários especializados de apoio, como tesouros, por exemplo. Para esta a descoberta de sinônimos no processo jurídico, a própria ontologia coopera diretamente para a mineração como, por exemplo, a simulação apresentada por Nascimento (2018, p. 79) em que são apresentados os termos e seus sinônimos para os sujeitos processuais como autor (reclamante, suscitante, impetrante, requerente, recorrente, embargante, embargado, agravante, exequente, liquidante, excipiente, reconvinte) e réu (reclamado, suscitado, impetrado, requerido, recorrido, embargado, agravado, executado, liquidado, exceto, reconvindo). Na simulação, Nascimento (2018, p. 81), exemplifica o processo de busca de um usuário que utiliza como expressão de busca os termos mais gerais, neste caso apenas “autor”, conforme figura 8:

**Figura 8** - Simulação de busca do sujeito “autor” em um sistema baseado em ontologia jurídica.



Fonte: Nascimento (2018, p. 81).

Ainda na camada de sinônimos podem ser utilizados os instrumentos de tesouro de domínio jurídico como, por exemplo: Tesouro do Supremo Tribunal Federal (TSTF)<sup>10</sup>, o Tesouro Jurídico da Justiça Federal<sup>11</sup>, o Vocabulário Jurídico do Supremo Tribunal Judiciário<sup>12</sup>.

10 <http://www.stf.jus.br/portal/jurisprudencia/pesquisarVocabularioJuridico.asp>

11 <http://www.cjf.jus.br/cjf/biblioteca/tesouro-juridico>

12 <http://www.stj.jus.br/SCON/thesaurus/>

- **Camadas de conceito e hierarquização de conceitos:** considera-se como elementos chaves a intensão e extensão do conceito. Para esta camada podem ser inferidas como propriedades de intensão do conceito, adotando como exemplo o próprio conceito de “processo”: o processo é um documento jurídico, que tramita em comarcas, possui partes como autor e réu, pode pertencer às esferas federais, estaduais, eleitorais, militares ou trabalhistas. Como propriedades da extensão do conceito tem-se, por exemplo, a característica que torna o processo como único, neste caso, o número processual unificado (NPU). Para a camada de hierarquização de conceitos, adota-se a taxonomia que pode ser construída por meio da mineração de texto como, por exemplo as movimentações processuais simuladas pelo ordenamento: “Processo” tem “Movimentação” que pode estar “Baixado”, “Extinto”, “Arquivado” ou “Encerrado”.
- **Camadas de relações e regras:** As principais técnicas se apoiam em análise linguísticas, algoritmos de descoberta de agrupamentos e de recorrências de termos. Nesse sentido as regras ou os axiomas são os elementos que dão sentido à formulação de questões de busca, realizando o exercício para validação das relações entre os conceitos. Aqui podem ser tidas questões em potencial que contribuem para o processo de tomada de decisão de um operador de Direito, como: “os advogados que mais ganharam processos referentes às empresas de telecomunicações”, “os juízes mais favoráveis às causas de feminicídio no município de Jaboaão”, “as organizações que mais sofrem processos por descumprir o direito de garantia do consumidor no ramo de eletrodomésticos”.

Ainda na camada de relações e regras, citamos a subcamada hiperonímia e hiponímia, em que nesse processo as principais técnicas se apoiam em análise linguísticas, algoritmos de descoberta de agrupamentos hierárquicos e de ocorrência de termos. Como exemplo do processo jurídico tem-se “Tipos” como hiperônimo de “decisão”, “acórdão” e “despacho”. Tal como “juiz” e “relator” são hipônimos de “Partes”.

Fundamental compreender que “quando um usuário aplica uma ferramenta de mineração de texto, ele não realiza só uma busca, ele analisa o texto para identificar padrões e tendências que possam lhe interessar” e que “o processo de mineração não retorna o conhecimento explícito, sendo necessário realizar uma análise e contextualização do conteúdo para que resulte em conhecimento útil” (REATEGUI; CAMPELO; OLIVEIRA, 2017, p. 40).



## 6 CONSIDERAÇÕES FINAIS

Pesquisas que se aportam do empirismo e pragmatismo são fundamentais para o desenvolvimento de uma área. Logo, estudar o sistema de ontologia como pretendido neste trabalho legitimou uma forte contribuição para a Ciência da Informação, principalmente ao que contribui para os estudos relacionados ao desenvolvimento de ontologias jurídicas frente às tarefas de mineração de textos.

O objetivo da pesquisa contentou-se à medida que foram explorados os conceitos de ontologia e mineração de texto, sendo solidificado na seção última em que os autores evidenciaram um percurso metodológico para construção de ontologias do domínio jurídico brasileiro, que por sua vez foram identificadas de acordo com as camadas de construção do sistema: termos, sinônimos, conceitos, hierarquia de conceitos, relações e regras.

Espera-se, com o desenvolvimento desta pesquisa, contribuir para a Ciência da Informação e, como apontado por García Gutiérrez, fazer caminhar a área de Organização do Conhecimento, ao que o Gutiérrez considera estar em um estado de inércia justificado por uma carência de um posicionamento ético evidenciado pelo escasso desenvolvimento de métodos, de técnicas e de ferramentas existentes para o exercício da organização do conhecimento em toda sua pluralidade, no caso deste estudo a organização do conhecimento por meio do sistema ontologia, explorando como representação metodológica a proposta de uma ontologia jurídica baseada no Processo Judicial Eletrônico.

## REFERÊNCIAS

ALMEIDA, M. B. Uma abordagem integrada sobre ontologias: Ciência da Informação, Ciência da Computação e Filosofia. **Perspectivas em Ciência da Informação**, v. 19, n. 3, p. 242-258, 2014. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/1736>. Acesso em: 7 jul. 2019.

BATRES, E. J. Q. *et al.* Uso de ontologias para a extração de informações em atos jurídicos em uma instituição pública. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, v. 10, n. 19, p. 73-88, 2005. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2005v10n19p73/5501>. Acesso em: 7 jul. 2019.

BECALLI, C. L. P. G. **UFO-L**: uma ontologia núcleo de aspectos jurídicos construídas sob a perspectiva das relações jurídicas. 2018. 308 f. Tese (Doutorado em Informática) – Universidade Federal do Espírito Santo. 2018. Disponível em: <http://repositorio.ufes.br/handle/10/9875>. Acesso em 7 jul. 2019.

- BEZERRA, C. A.; GUIMARÃES, A. J. R. Mineração de texto aplicada às publicações científicas sobre gestão do conhecimento no período de 2003 a 2012. **Perspectivas em Ciência da Informação**, v. 19, n. 2, p. 131-146, 2014. Disponível em: <http://dx.doi.org/10.1590/1981-5344/1834> . Acesso em: 7 jul. 2019.
- BRAGA, F. R. Extração semiautomática de taxonomia para domínios especializados usando técnicas de mineração de textos. **Ciência da Informação**, v. 45, n. 3, 2016. Disponível em: <http://revista.ibict.br/ciinf/article/view/4056>. Acesso em: 7 jul. 2019.
- BUITELAAR, P.; CIMIANO, P.; MAGNINI, B. Ontology Learning from Text: An Overview. **Ontology Learning from Text: Methods, Applications and Evaluation**. 3. ed. Amsterdam; New York; Oxford: IOS Press, 2005. Cap. 1, p. 3–12.
- CIMIANO, P. **Ontology learning and population from text: algorithms, evaluation and applications**. Karlsruhe, Germany: Springer, 2006. 375 p.
- CONSTÂNCIO, A. S. **Ontologia para motor de busca semântica para recuperação jurisprudencial no Brasil**. 2017. 202 f. Dissertação (Mestrado em Ciência, Gestão e Tecnologia da Informação) – Universidade Federal do Paraná, 2017. Disponível em: <https://acervodigital.ufpr.br/handle/1884/46276>. Acesso em: 7 jul. 2019.
- FARIAS, K. M.; PINHO, F. A. Ontologias como ferramenta de organização e representação do conhecimento: um olhar sobre os laudos médico-legais. **Informação em Pauta**, v. 1, n. 2, p. 41-65, 2016. Disponível em: <https://doi.org/10.32810/2525-3468.ip.v1i2.2016.5431>. Acesso em: 7 jul. 2019
- FERNEDA, E. *et al.* Potential of ontology for interoperability in e-government: discussing international initiatives and the Brazilian case. **Brazilian Journal of Information Science: Research Trends**, v. 10, n. 2, 2016. Disponível em: <https://doi.org/10.5016/brajis.v10i2.4812>. Acesso em: 7 jul. 2019.
- GAJZER, M. Text and data mining techniques in aspect of knowledge acquisition for decision support system in construction industry. **Technological and Economic Development of Economy**, v. 16, n. 2, p. 219-232, 2010.
- GUIMARÃES, J. A. C. La información jurídica. In: BARITÉ, Mario; GUIMARÃES, José Augusto Chaves. **Guia metodológica para el acceso, el análisis y la organización temática de documentos jurídicos**. Montevideo: Universidad de La República, 1999. p. 11-18.
- HOTH, A.; NÜRNBERGER, A.; PAAß, G. A brief survey of text mining. In: **Ldv Forum**. 2005. p. 19-62.
- JANAITE NETO, J.; FERNEDA, E. Ontologia como recurso de padronização terminológica no processo de recuperação de informação. **Informação em Pauta**, v. 1, n. 1, p. 30-45, 2016. Disponível em: <https://doi.org/10.32810/2525-3468.ip.v1i1.2016.2967>. Acesso em: 7 jul. 2019.
- KAUR, R.; AGGARWAL, S. Techniques for mining text documents. **International Journal of Computer Applications**, v. 66, n. 18, 2013. Disponível em: <https://www.ijcaonline.org/archives/volume66/number18/11184-6318>. Acesso em: 7 jul. 2019.

MINGHELLI, D. T. **A relação de meronímia em uma ontologia jurídica**. 2011. 126p. Dissertação (Mestrado em Linguística Aplicada) – Universidade de Vale do Rio dos Sinos, 2011. Disponível em: <http://www.repositorio.jesuita.org.br/handle/UNISINOS/3520?show=full>. Acesso em: 7 jul. 2019.

MIRANDA, A. C. C. de. A política de desenvolvimento de coleções no âmbito da informação jurídica. In: PASSOS, E. (Org.). **Informação jurídica: teoria e prática**. Brasília: Thesaurus, 2004. p. 137-152.

NASCIMENTO, F. M. S.; SIEBRA, S. A. Organização da Informação: a ontologia no contexto do *business intelligence*. **Revista Eletrônica de Gestão Organizacional**, v. 13, edição especial, p. 456-465, 2015. Disponível em: <https://periodicos.ufpe.br/revistas/gestaoorg/article/download/22124/18489> . Acesso em: 7 jul. 2019.

\_\_\_\_\_. A organização da Informação Jurídica: a ontologia no contexto do business intelligence. 2015. 82f. Monografia (Bacharelado em Gestão da Informação) – Universidade Federal de Pernambuco, 2015.

\_\_\_\_\_; PINHO, F. A. Ontologia na gestão de conhecimento jurídico. **Revista P2P e INOVAÇÃO**, v. 4, n. 2, p. 41-52, 2018(a). Disponível em: <https://doi.org/10.21721/p2p.2018v4n2.p41-52> . Acesso em: 7 jul. 2019.

\_\_\_\_\_; PINHO, F. A. Uso estratégico de ontologia para organização e gestão da informação da informação jurídica. In: Encontro Nacional de Pesquisa em Ciência da Informação, 19., 2018b, Londrina. **Anais...** Londrina: ANCIB, 2018. Disponível em: <http://enancib.marilia.unesp.br/index.php/XIXENANCIB/xixenancib/paper/view/931> . Acesso em: 7 jul. 2019.

\_\_\_\_\_; BRAZ, M. I.; PINHO, F. A. Os mapas conceituais para representação da informação no domínio jurídico. **Cadernos de Informação Jurídica**, Brasília, v. 5, n. 1, p. 5-21, 2018. Disponível em: <https://www.cajur.com.br/index.php/cajur/article/view/172>. Acesso em: 7 jul. 2019.

\_\_\_\_\_; **Uso estratégico da ontologia para organização e gestão da informação jurídica**. 2018. 105 f. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal de Pernambuco, 2018. Disponível em: <https://repositorio.ufpe.br/handle/123456789/29646> . Acesso em: 7 jul. 2019.

PASSOS, E.; BARROS, L. V. **Fontes de informação para pesquisa em direito**. Brasília, DF: Briquet de Lemos, 2009.

REATEGUI, E. B.; BARCELLOS, P. S. C. C.; OLIVEIRA, S. O apoio de uma ferramenta com base na mineração de texto para escrita acadêmica. **Informática na educação: teoria & prática**. Porto Alegre. Vol. 20, n. 1, p. 36-45, 2017 Disponível em: <https://doi.org/10.22456/1982-1654.70063> . Acesso em: 7 jul. 2019.

SANGA, D. A. O. **Mineração de textos para o tratamento automático em sistemas de atendimento ao usuário**. 2017. 93 f. Dissertação (Mestrado em Computação Aplicada) – Universidade Tecnológica Federal do Paraná, 2017. Disponível em: <http://repositorio.utfpr.edu.br:8080/jspui/handle/1/2850>. Acesso em: 7 jul. 2019.

SOUZA NETO, M. P. Da Automatização à Virtualização: Apontamentos Arquivísticos sobre Processo Judicial Eletrônico. **Revista Memorare**, v. 2, n. 1, p. 102-124, 2014. Disponível em: [www.portaldeperiodicos.unisul.br/index.php/memorare\\_grupep/article/view/2645](http://www.portaldeperiodicos.unisul.br/index.php/memorare_grupep/article/view/2645) . Acesso em: 7 jul. 2019.

SILVA, A. G. **Fontes de informação jurídica**. Rio de Janeiro: Interciência, 2010.

SCHIESSL, M.; BRÄSCHER, M. Do texto às ontologias: uma perspectiva para a ciência da informação. **Ciência da Informação**, v. 40, n. 2, 2011. Disponível em: <http://revista.ibict.br/ciinf/article/view/1318>. Acesso em: 7 jul. 2019.

TOMOYOSE, K.; JÚNIOR, F. J. G.; SIMIONATO, A. C.; ROCHA, E. S. S. Indicadores bibliométricos no estudo da ontologia. **Biblionline**, v. 14, n. 1, p. 29-40, 2018.

ZAHRA, F. M.; *et al.* Ferramentas para aprendizagem de ontologias a partir de textos. **Perspectivas em Ciência da Informação**, v. 19, n. 1, p. 3-21, 2014. Disponível em: <http://www.scielo.br/pdf/pci/v19n1/02.pdf>. Acesso em 7 jul. 2019.

ZAMUR FILHO, J. **Processo judicial eletrônico**: alcance e efetividade sob a égide da Lei nº 11.419, de 19.12.2006. 2011. 152 f. Dissertação (Mestrado em Direito) – Faculdade de Direito da Universidade de São Paulo, 2011. Disponível em: <http://www.teses.usp.br/teses/disponiveis/2/2137/tde-02052012-105409/es.php>. Acesso em: 7 jul. 2019.