

Titulação Automática de Acórdãos Baseado em Ontologia Jurisprudencial

Alberto Heitor Molinari
ahmolinari@gmail.com

Cesar Augusto Tacla
tacla@utfpr.edu.br

Sumário: 1. Introdução 2. Revisão Bibliográfica 3. Metodologia 4. Resultados 5. Conclusão 6. Recomendações de Estudo 7. Referências Bibliográficas.

Este artigo aborda a criação de um sistema computacional de rotulação automática de documentos jurisprudenciais utilizando-se de técnicas de recuperação de informações para popular uma ontologia jurisprudencial, possibilitando assim, possíveis pesquisas semânticas.

Palavras-Chave: Recuperação de Informações; Ontologia; Jurisprudência

1. Introdução

A pesquisa de jurisprudência está presente em praticamente todos os sítios da Internet de tribunais de justiça de nosso país, tratando-se de ferramenta fundamental para a elaboração de teses do Direito, tanto por parte de advogados como de magistrados e ministros da suprema corte. A grande maioria das ferramentas disponibilizadas nos sítios dos tribunais regionais efetua buscas essencialmente textuais, retornando, muitas vezes resultados extensos, tornando a pesquisa pouco produtiva. Uma minoria de tribunais, mais notadamente aqueles situados em nosso Distrito Federal, dispõe de buscas adicionais baseadas na rotulação (ou titulação, no jargão jurídico) prévia dos documentos, propiciando assim resultados muito mais precisos.

É senso comum no setor judiciário que os tribunais não efetuam a rotulação de seus documentos por óbvia escassez de recursos humanos para efetuar esta tarefa tão longa e minuciosa. Assim, este trabalho objetiva construir um sistema que seja capaz de efetuar a rotulação automática de acórdãos, alimentando, assim, uma base de conhecimento adequada e disponível para uma pesquisa estruturada, ou até para uma pesquisa semântica a ser proposta em trabalho futuro.

2. Revisão Bibliográfica

Para embasar as metas deste trabalho é necessária uma revisão bibliográfica dividida em duas frentes de estudo de técnicas ou metodologias:

- rotulação de documentos: é necessário pesquisar técnicas que auxiliem na inferência computacional de rótulos que classifiquem ou definam um documento jurisprudencial. A pesquisa deve se dar dentro do arcabouço de técnicas de Recuperação de Informações;

- desenvolvimento de bases de conhecimento - na sequência é necessária a pesquisa de uma metodologia adequada para sistematizar o conhecimento a ser adquirido em cada documento, em conformidade com a técnica selecionada para a rotulação.

2.1. Técnicas de Recuperação de Informações para Rotulação de documentos

Segundo (TAN; STEINBACH; KUMAR, 2005), a Descoberta de Padrões Sequenciais é uma tarefa que pode ser definida como:

Dado um conjunto de objetos, cada objeto associado com sua própria linha do tempo de eventos, encontrar regras que predigam fortes dependências sequenciais entre diferentes eventos.

$$(A B) (C) \rightarrow (D E)$$

Regras são formadas descobrindo inicialmente padrões. As ocorrências de eventos nos padrões são governadas pelas restrições temporais.

Considerando-se que textos de domínio específico, como a jurisprudência, podem conter sinais ou sentenças típicas dos quais se possam derivar ou extrair informações relevantes, a presente tarefa pode consistir de uma alternativa viável para este trabalho, porém restringindo as técnicas à área de Mineração de Textos e mais especificamente à disciplina de Recuperação de Informações.

Segundo (BERRY, 2004), na extração de sentenças, técnicas clássicas, que consideram ocorrência e localização de palavras bem como palavras de sinalização, podem ser utilizadas para gerar um glossário de palavras-chave a ser utilizado na geração de score para o cálculo da relevância da sentença.

A ocorrência baseia-se na intuição de que termos que aparecem frequentemente em um texto específico, mas não são comuns no conjunto de textos correlatos, podem ser de grande relevância para o primeiro. Assim a densidade da relevância dos termos em uma sentença denotam a importância da sentença para o texto, sendo geralmente calculada através do método TFIDF, definido por (SALTON; BUCKLEY, 1988) pela equação:

$$tfidf(tk,dj) = \#(tk,dj) * \log (Tr / \#Tr(tk))$$

onde $\#(tk,dj)$ denota o número de vezes que o termo tk ocorre no documento dj , e $\#Tr(tk)$ denota a frequência do termo tk nos Tr documentos do corpus, em que tk ocorre (SEBASTIANI, 2002).

A **localização** de termos pode denotar a importância de uma sentença. Segundo (EDMUNDSON, 1968) o título e o cabeçalho são locais importantes de um documento, contendo, assim, termos relevantes. Mas (BERRY, 2004) acrescenta que também seções e seus respectivos títulos podem conter relevantes termos, merecendo então a abrangência em uma captura de termos relevantes por localização.

Por fim, em termos de **sinalização**, (EDMUNDSON, 1968) propôs a hipótese de que a presença de palavras ou frases pragmáticas podem afetar a relevância de uma sentença. Como exposto por (BERRY, 2004), os termos podem ser classificados como *bônus*, que são positivamente relevantes, ou *stigma*, que são negativamente relevantes, sendo dependentes do domínio em questão e devendo, assim, ser levantados manualmente passando a constar de um glossário de termos-chave a ser utilizado na extração das sentenças.

2.2. Metodologias para Desenvolvimento de Bases de Conhecimentos

A Tecnologia da Informação atualmente disponibiliza uma ilimitada sorte de pesquisas, teorias, abordagens e técnicas que estão originando inúmeros artefatos tecnológicos que visam solucionar o problema da Gestão do Conhecimento. Assim, o conhecimento deve ser modelado, apropriadamente estruturado e interligado visando flexibilidade e apresentação personalizada para os usuários (YORK; STEFFEN; STUDER, 2003).

Ontologias têm demonstrado ser a ferramenta adequada para problemas de estruturação e modelagem, provendo uma conceituação de domínios particulares que podem ser compartilhadas por grupos de pessoas dentro de uma organização (GRUBER, 1993), passando então a figurar, neste trabalho, como a solução para o desenvolvimento da base de conhecimento, visto que a técnica escolhida para rotulação de documentos extrairá conceitos de um domínio específico, os quais necessitarão de uma adequada representação, visando a sua reutilização por ferramentas de pesquisa semântica.

2.2.1 Ontologias

De um ponto de vista filosófico, uma ontologia é um sistema de categorias, independente de linguagem, que resulta em uma visão específica do mundo. Em Inteligência Artificial, ontologia consiste de um artefato de engenharia, dependente de linguagem, constituído das seguintes partes (GUARINO, 1998):

- vocabulário, utilizado para descrever uma realidade específica;
- afirmações explícitas, presumindo o significado pretendido para cada vocábulo, geralmente construídas através da teoria da lógica de primeira-ordem;
- relações de subsunção, que descrevem uma hierarquia de conceitos;
- axiomas, informações factuais básicas ou definições de conceitos. Alguns fornecem informações gerais sem definir totalmente um conceito (RUSSELL; NORVIG, 2004).

Tanto o vocabulário, quanto as afirmações e axiomas podem ser descritos através da linguagem Lógica de Descrição (DL).

2.2.2 Metodologias para Gestão de Conhecimento

Existem várias propostas de metodologias para a Gestão de Conhecimento em empresas, geralmente divididas em duas fases (YORK; STEFFEN; STUDER, 2003): **Meta-processo de Conhecimento**, responsável por desenvolver e manter uma solução de conhecimento em uma organização, e **Processo de Conhecimento**, o qual visa a utilização da solução criada na primeira fase.

Como exemplo, podem-se citar as seguintes metodologias ou abordagens, tidas como as mais conhecidas ou difundidas: CommonKADS, DOGMA, Enterprise Ontology, KACTUS, Methontology, On-To-Knowledge Methodology (OTKM), SENSUS e TOVE. Destas pode-se destacar a Methontology, definida por (FERNÁNDEZ; GÓMEZ-PÉREZ; JURISTO, 1997), a qual possui um alto nível de detalhamento e documentação para a fase de Meta-processo de Conhecimento, sendo, assim, bastante adequada para grandes domínios de conhecimento. Outra metodologia bastante difundida é a OTKM, definida por (YORK; STEFFEN; STUDER, 2003), a qual apresenta formalismo e extensão medianos, sendo assim adequada para problemas envolvendo domínios de até médio porte. Um diferencial da OTKM, com relação à Methontology, é a existência da fase Processo de Conhecimento, dando, assim, cobertura para o ciclo de vida completo de sistemas de gestão do conhecimento. Estas características tornam a OTKM uma opção viável para o presente trabalho posto que este envolve um domínio de escopo reduzido e, além disso, demanda do desenvolvimento de uma solução para utilização do conhecimento.

3. Metodologia

Este capítulo descreve a metodologia utilizada na criação e aplicação da ontologia que representa o domínio da jurisprudência, baseado nos conhecimentos adquiridos através dos acórdãos do Tribunal de Justiça do Estado do Paraná (TJPR), denominada assim JurisTJPR. Direcionando as tarefas e técnicas, a seguir descritas, está a metodologia OTKM com diretrizes que guiam todo o processo de desenvolvimento, visando um melhor resultado. Assim são abordadas as fases e passos da referida metodologia conceituando e aplicando-as à solução do problema.

3.1.3 Fase *Refinement* (refinamento)

O objetivo desta fase é a formalização da ontologia semi-formal até a obtenção da ontologia almejada. Primeiramente deve-se formar uma taxonomia com os conceitos levantados, adicionando outros tipos de relações, além das *is-a* (inerentes aos modelos taxonômicos). Além disso, com a ajuda de especialistas nos domínios, devem-se adicionar atributos aos conceitos e traçar novas relações entre os conceitos. Em seguida deve-se formalizar a ontologia almejada, bem como criar um protótipo, baseado nos conceitos e propriedades da taxonomia. Para tal se utilizou o sistema Protégé (INFORMATICS, 2010), através do qual pode-se modelar e criar ontologias, através da linguagem OWL-DL. O Protégé possibilita também a prototipação, através da inserção de instâncias dos conceitos e de valores para as propriedades. O resultado do refinamento pode ser resumido através de uma imagem da taxonomia criada no citado sistema, conforme a figura 2.

3.1.4 Fase *Evaluation* (avaliação)

A OTKM define que, em seguida, devem ocorrer três tipos de avaliação da ontologia, visando garantir a qualidade da mesma:

- **avaliação com enfoque na ontologia:** (YORK; STEFFEN; STUDER, 2003) determina que é necessário que alguém avalie formalmente a ontologia criada. Assim, neste trabalho, adotou-se apresentar a taxonomia a uma banca tecnológica, composta por integrantes do grupo de estudos de Ontologia da UTFPR, para uma avaliação técnica dos conceitos, seus atributos e relações, visando uma melhor adequação lógica e um novo refinamento da ontologia almejada;

- **avaliação com enfoque tecnológico:** com a ontologia mais refinada, deve-se efetuar uma avaliação com o uso de ferramenta que avalie automaticamente propriedades (sintaxe), consistência (semântica) e propriedades tecnológicas (e.g. interoperabilidade, escalabilidade, etc). Para tal, optou-se pela avaliação sintática e semântica através do Pellet (PARSIA, 2010), um programa “raciocinador” que pode ser invocado a partir da interface do Protégé para submeter a ontologia presente na memória às seguintes avaliações: Consistência, Satisfabilidade de Conceitos, Classificação e Realização; nas quais a ontologia aqui almejada obteve sucesso.



Figura 2: Ontologia resultante da taxonomia refinada

- **avaliação com enfoque no usuário:** o objetivo é avaliar a satisfação do usuário com a aplicação de Gestão do Conhecimento, ou seja, se a aplicação, baseada em ontologia, é no mínimo tão boa quanto as aplicações existentes para resolver problemas similares. Inicialmente, a taxonomia resultante do segundo refinamento foi apresentada para o pessoal da Seção de Jurisprudência do TJPR, integrada por profissionais especializados no assunto correlato. Não houve discordâncias e nem tão pouco acréscimos aos conceitos demonstrados na taxonomia. Em seguida foram apresentados aos especialistas, os resultados obtidos na aplicação de população da ontologia (que será abordada na seção 3.2), onde, para cada acórdão processado, é apresentada uma lista de conceitos/propriedades a ser comparada com o texto integral do documento, visando apontar indivíduos julgados inadequados para o acórdão, bem como sugerir outros indivíduos que deveriam ter sido populados na ontologia. O resultado desta avaliação é apresentado no capítulo 4.

3.1.5 Fase *Application and Evolution* (aplicação e evolução)

A aplicação, além de consistir de um passo da metodologia OTKM, deve se utilizar de uma metodologia de desenvolvimento de *software* para a sua construção, sendo, assim, abordado na seção 3.2. A evolução deve consistir de um processo organizacional regido de regras para atualizações, inclusões ou exclusões, sendo bastante recomendável que as mudanças sejam controladas por versionamento da ontologia.

3.2. Aplicativo de Rotulação de Acórdãos e População de Ontologia

O Meta-processo culminou com a criação de uma ontologia jurisprudencial. Assim, é possível se passar para a última fase deste processo, a qual visa aplicar e manter o conhecimento obtido. Neste trabalho optou-se por construir um aplicativo o qual, através de técnicas de rotulação de documentos, extrai conceitos de um corpus de acórdãos, utilizando-os para popular a ontologia. O primeiro requisito para a construção do aplicativo é a sua codificação utilizando a linguagem Java, por ter a sua licença de utilização regida por regras de software livre; por ter se tornado, na presente década, uma linguagem padrão para qualquer aplicação científica; e por facilitar a reutilização do código em trabalhos futuros.

3.2.1 Extração de Conceitos

A extração de conceitos se dá através da metodologia de descoberta de padrões sequenciais, através da qual pode-se lançar mão de várias técnicas, de acordo com os padrões procurados. Uma técnica muito utilizada é a análise de expressões regulares (NETWORK, 2010), a qual visa a identificação de padrões pré-definidos em cadeias de caracteres.

Entretanto, como os padrões mapeados neste trabalho se baseiam, não apenas em sinais mas também, em termos linguísticos é necessário o uso concomitante de uma técnica que padronize os termos. Assim, neste trabalho é utilizada a técnica de *Stemming* (ALVARES, 2005), a qual visa reduzir os termos à sua forma radical.

Por fim, considerando que muitos dos termos oriundos do corpus de acórdãos possuem uma semântica jurídica é importante a utilização de uma técnica que mapeie os termos relevantes no domínio em questão, o que aqui se dá através de pesquisa no Tesouro Jurídico da Justiça Federal, desenvolvido por (CEJ, 1997), ao invés das técnicas mais comuns de cálculo de relevância, como o TFIDF.

3.2.2 População de Ontologia

Os conceitos extraídos de cada acórdão devem, na sequência, ser populados na ontologia, visando a gestão do conhecimento adquirido do corpus.

Conforme já citado, a Lógica de Descrição (DL) é uma linguagem através da qual se pode equacionar uma ontologia. A implementação desta tecnologia foi proposta e regulamentada pela W3C (MCGUINNESS; HARMELEN, 2004) através da especificação da Ontology Web Language, OWL-DL, um padrão de representação e armazenamento de uma ontologia baseado no formato RDF, e da SPARQL, uma especificação de linguagem de *queries* para pesquisas baseadas em ontologias representadas por OWL-DL.

A implementação dos dois padrões propostos pela W3C, OWL-DL e SPARQL, está hoje disponível através do Jena, uma framework implementada em Java para construção de aplicações de Web Semântica (HEWLETT-PACKARD, 2010), que possibilita a gestão de conhecimento em

ontologias representadas pelo padrão OWL, através de uma API para OWL e de um “motor” para queries SPARQL, sendo, assim, utilizada na aplicação aqui postulada, visando a carga em memória e a população da ontologia.

3.2.3 Arquitetura do aplicativo

As metodologias, técnicas e ferramentas até aqui estudadas, por fim promovem o embasamento tecnológico suficiente para se construir a aplicação almejada neste trabalho, a qual foi planejada, para quebrar a sua complexidade, em quatro módulos, conforme a figura 3:

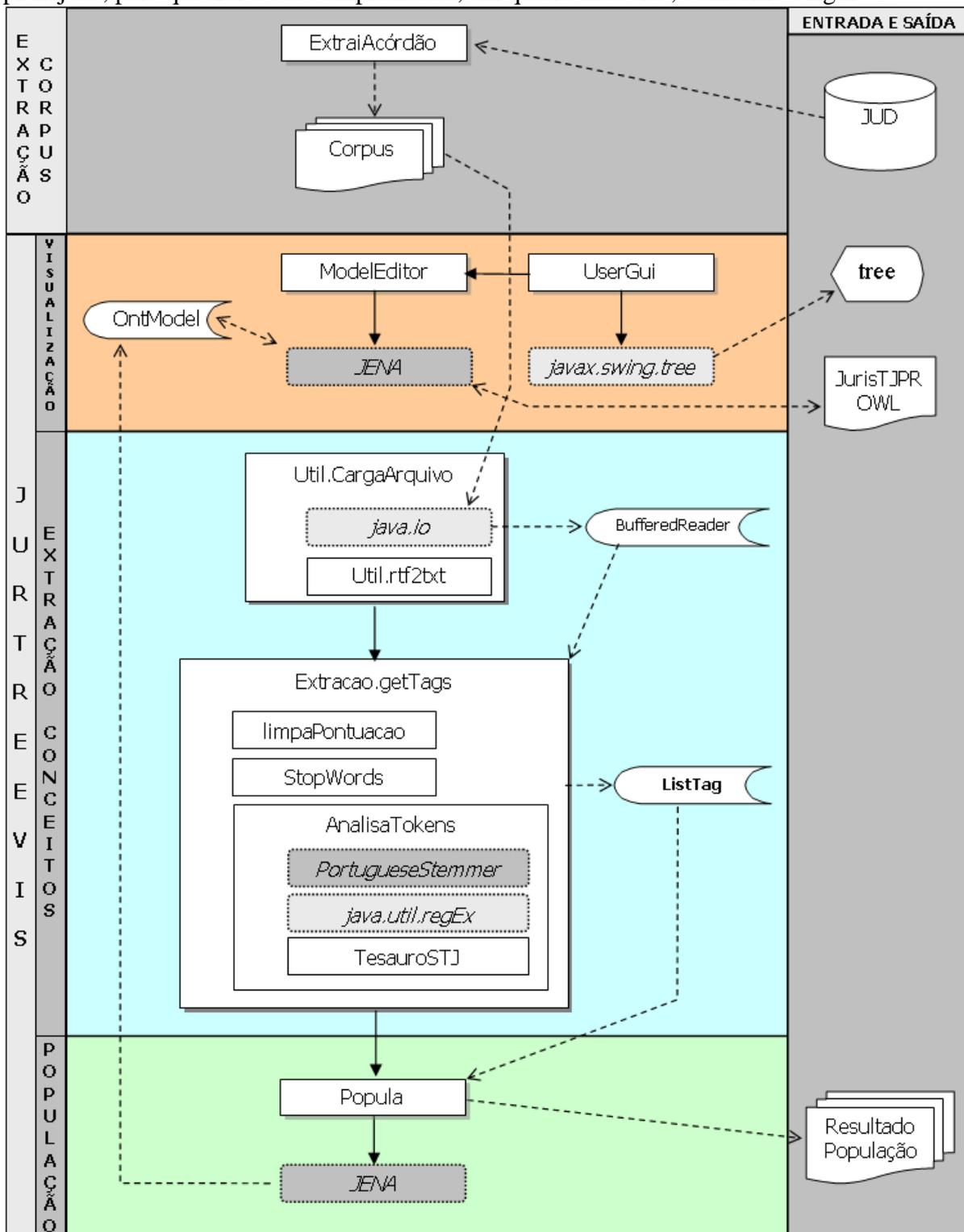


Figura 3: Diagrama da arquitetura do aplicativo

- **Extração de Corpus:** é um módulo, na verdade, externo à aplicação o qual é executado previamente visando a composição do corpus de acórdãos, cujos documentos necessitam ser extraídos de sua base de dados de origem, fornecida pelo TJPR;

- **Visualização:** módulo de entrada e saída do aplicativo, cuja função é proceder à carga da ontologia (do arquivo para a memória), representá-la graficamente na tela do computador e, por fim, gravá-la novamente no arquivo, visando tornar permanentes as populações efetuadas na ontologia em memória. Para a implementação deste módulo foi reaproveitado o sistema TreeVis, de autoria de Anderson Seiti Yamamoto, construído na plataforma Java, sob a orientação do co-autor do presente trabalho, visando justamente a carga, a visualização em forma de árvore e a gravação de ontologias em arquivo no formato OWL;

- **Extração de Conceitos:** módulo onde se efetua uma varredura do corpus, visando a extração de instâncias para os conceitos de jurisprudência mapeados pela ontologia. O resultado deste módulo é gravado em memória em uma lista de conceito-instância;

- **População:** módulo onde a ontologia em memória é populada com as instâncias extraídas do corpus pelo módulo anterior.

Assim, o sistema TreeVis, adicionado dos módulos de extração de conceitos e de população da ontologia jurisprudencial, recebeu um novo nome, visando a diferenciação de sua versão: **jurTreeVis**.

A execução deste aplicativo, tendo como entrada a base de dados de acórdãos do TJPR, resulta, por fim, na ontologia jurisprudencial populada (sendo gravada no arquivo JurisTJPR.OWL) e em uma listagem, para cada acórdão, de conceitos e propriedades populados (representado na figura 3 por “Resultado População”) visando a avaliação com enfoque no usuário, já abordada na sub-seção 3.1.4.

4. Resultados

Foram submetidos ao aplicativo 1.149 documentos, referentes a acórdãos publicados no ano de 2005. Entretanto, a escassez de especialistas disponíveis bem como o alto custo de uma avaliação de domínio, a cada documento, são fatos que exigem a definição de uma amostra de tamanho viável para um ou dois especialistas, mas suficiente para se chegar a conclusões sobre cobertura e precisão dos resultados do aplicativo aqui descrito.

Assim, foram sorteados cinquenta documentos, resultando, a cada um, uma lista de conceitos e propriedades populados na ontologia bem como o texto integral do acórdão, conforme o trecho representado na figura 4.

Por fim, os especialistas foram convidados a efetuar uma comparação cognitiva entre a lista e o respectivo texto de cada acórdão, visando apontar indivíduos que eles julgaram tecnicamente inadequados, bem como sugerir outros indivíduos que deveriam, no seu julgamento, ter sido populados, sendo que o resultado foi tabulado conforme a tabela 1, que demonstra uma amostra da tabulação, bem como a sua totalização.

EXTRAÇÃO
=====

Marque um X nas lacunas das extrações com as quais você não concorda.

[] Apelacao - Cível
[] - Numero: 0231716-0
[] Ator - Nome: Xxxxxx Xxxx XXXXXXXXX
[] - Apelante
[] Ator - Nome: Yyyyyyy Yyyyyyy Cia Seguros
[] - Apelado
[] Ator - Nome: Luiz A. Barry
[] - Relator
[] Fato - Descricao: AÇÃO DE RESSARCIMENTO
[] Assunto - Descricao: ACIDENTE DE TRÂNSITO
[] Assunto - Descricao: AÇÃO REGRESSIVA
[] Fato - Descricao: LEGITIMIDADE DA EMPRESA SUB-ROGADA
[] Fato - Descricao: CULPA DOS RÉUS COMPROVADA
[] Fato - Descricao: VIA PREFERENCIAL
[] Fato - Descricao: FALTA DE CAUTELA DO MOTORISTA AO CRUZAR A VIA
[] Fato - Descricao: FALHA MECÂNICA
[] Assunto - Descricao: CASO FORTUITO
[] Fato - Descricao: CULPA CARACTERIZADA
[] Assunto - Descricao: CULPA CONCORRENTE
[] Fato - Descricao: CAUSA PRIMÁRIA DO ACIDENTE
[] Fato - Descricao: INDENIZAÇÃO DEVIDA
[] Voto - Descricao: APELAÇÃO DESPROVIDA.
[] Legislacao - NumArtigo: 159
[] - Nome: código civil

Anote abaixo extrações que você julga que deveriam ter ocorrido.

ÍNTEGRA DO ACÓRDÃO
=====

APELAÇÃO CÍVEL N° 0231.716-0
VARA CÍVEL DA COMARCA DE GOIOERÊ
APELANTE: XXXXXX XXXX XXXXXXXXX
APELADO: YYYYYYY YYYYYYY - CIA. DE SEGUROS.
RELATOR: LUIZ A. BARRY

AÇÃO DE RESSARCIMENTO - ACIDENTE DE TRÂNSITO - SEGURADOR SUB-ROGADO AÇÃO REGRESSIVA - LEGITIMIDADE DA EMPRESA SUB-ROGADA - CULPA DOS RÉUS COMPROVADA - VIA PREFERENCIAL - FALTA DE CAUTELA DO MOTORISTA AO CRUZAR A VIA - FALHA MECÂNICA - INEXISTÊNCIA DE CASO FORTUITO OU FORÇA MAIOR - CULPA CARACTERIZADA - AUSÊNCIA DE CULPA CONCORRENTE - CAUSA PRIMÁRIA DO ACIDENTE - INDENIZAÇÃO DEVIDA - APELAÇÃO DESPROVIDA.

I - Relatório.
Trata a espécie de Apelação Cível, proposta por Xxxxxx Xxxx XXXXXXXXX inconformado com a procedência da ação de ressarcimento proposta pela empresa seguradora Yyyyyyy Yyyyyyy - Cia. de Seguros.
De se ver, a então autora ingressou com ação que denominou de Ação de Repa-

Figura 4: Trecho do resultado do processamento de um acórdão

Acórdão	Conceitos						Cobertura	Precisão	
	Total	Inadequados		Redundantes		Desejáveis			
		Qtd.	%	Qtd.	%	Qtd.			%
2005/4523	31	2	6,45	0	0,00	10	32,26	74,36	93,55
2005/10515	28	3	10,71	6	21,43	6	21,43	76,00	67,86
2005/10485	42	2	4,76	16	38,10	2	4,76	92,31	57,14
2005/10584	36	3	8,33	13	36,11	2	5,56	90,91	55,56
2005/10641	38	6	15,79	10	26,32	7	18,42	75,86	57,89
2005/10711	28	10	35,71	0	0	12	42,86	60,00	64,29
...
2005/10812	94	7	7,45	44	46,81	2	2,13	95,56	45,74
2005/15555	26	1	3,86	0	0	2	7,69	92,59	96,15
2005/1569	70	3	4,29	33	47,14	0	0	100,00	48,57
2005/1612	50	2	4,00	0	0,00	0	0,00	100,00	96,00
2005/2664	28	1	3,57	4	14,29	13	46,43	63,89	82,14
2005/1469	31	2	6,45	0	0,00	4	12,90	87,88	93,55
TOTAL	1870	148	7,91	277	14,81	246	13,16	85,45	77,27

Tabela 1: Resultado da avaliação com enfoque no usuário

onde:

Total indica a quantidade de conceitos/propriedades populados no acórdão;

Inadequados quantifica os conceitos/propriedades populados que foram julgados inadequados pelo especialista, conforme a sua análise do texto integral do acórdão;

Redundantes quantifica os conceitos populados redundantes, ou seja, aqueles que foram aceitos pelos especialistas, porém apontados como repetidos por terem conteúdo igual ou similar a outros indivíduos do mesmo acórdão;

Desejáveis é a quantidade e o percentual (comparado ao total de conceitos populados) de indivíduos desejáveis, ou seja, sentenças que constam no texto do acórdão e que deveriam ter sido populadas em algum conceito, segundo os especialistas;

Cobertura é a taxa de cobertura da população de conceitos do acórdão a qual demonstra o percentual de conceitos populados corretamente pelo sistema dentro do conjunto esperado. Como neste trabalho não há dados de aprendizado prévio, os quais poderiam determinar o que se espera do resultado, foi adotado o seguinte método para determinar o conjunto esperado: considera-se a quantidade de conceitos esperados subtraindo os indevidos e os redundantes do total populado e somando os desejáveis. Assim, a taxa de cobertura pode ser equacionada da seguinte forma:

$$C = \frac{t - i - r}{t - i - r + d} \cdot 100$$

onde:

C é o percentual de cobertura resultante;

t é o total de conceitos populados pela aplicação;

i é o total de conceitos considerados inadequados;

r é a quantidade de conceitos considerados redundantes;

d é a quantidade de conceitos desejáveis;

Precisão indica a taxa de precisão da população de conceitos do acórdão a qual demonstra o percentual de conceitos populados corretamente dentro do conjunto gerado pelo sistema, sendo definida pela seguinte equação:

$$P = \frac{t - i - r}{t} \cdot 100$$

onde:

P é o percentual de precisão resultante;

t é o total de conceitos populados pela aplicação;

i é o total de conceitos considerados inadequados;

r é a quantidade de conceitos considerados redundantes.

5. Conclusão

Visando a rotulação automática de acórdãos e a consequente disponibilização em uma base de conhecimentos, este trabalho revisou, e selecionou, técnicas de recuperação de textos bem como metodologias para gestão de conhecimento, através das quais criou uma ontologia jurisprudencial e um aplicativo para a sua gestão, o qual é capaz de extrair conceitos e propriedades a partir de documentos de acórdãos, populando-os na ontologia, através da criação de indivíduos.

Uma amostra de cinquenta documentos foi processada pelo aplicativo, sendo que os respectivos resultados foram submetidos à apreciação de especialistas em jurisprudência, visando uma análise de **cobertura**, que totalizou 85,45%, e de precisão, que totalizou 77,27%, resultados estes que demonstram ser muito promissores, principalmente se for considerado que o aplicativo não dispunha de um aprendizado de máquina prévio e nem tão pouco requereu intervenção do usuário para chegar à rotulação final. Assim é possível se afirmar que os objetivos almejados para o presente trabalho foram alcançados.

6. Recomendações de Estudo

A análise dos especialistas demonstrou que alguns indivíduos que por eles foram refutados e outros que foram sugeridos como desejáveis, poderiam ter uma extração e população adequada se houvesse uma análise lingüística mais apurada.

Além destes, foram apontados, pelos especialistas, indivíduos cuja inadequação foi a redundância, ou seja, são indivíduos que, apesar de corretamente populados em conceitos definidos, são similares ou idênticos a indivíduos já populados no mesmo conceito.

Assim, estas não-conformidades sugerem que, uma evolução dos resultados deste trabalho, seria possível através de um aprofundamento em técnicas lingüísticas, visando obter um *stemming* mais apurado, bem como por uma análise de similaridade entre indivíduos, talvez baseada em um enriquecimento do tesouro aqui utilizado, visando reduzir a quantidade de redundâncias.

Por fim, um trabalho futuro poderia partir do aplicativo aqui descrito, bem como da ontologia por ele populada, propondo uma busca semântica jurisprudencial visando resultados mais precisos do que aqueles obtidos atualmente nas pesquisas disponibilizadas pela grande maioria dos tribunais.

7. Referências Bibliográficas

ALVARES, R. V. **Investigação do Processo de Stemming na Língua Portuguesa**. 2005. 83 p.

- BERRY, M.W. Survey of text mining: clustering, classification, and retrieval. **Library of Congress Cataloging-in-Publication Data**, p. 260, 2004.
- BUZAN, T. **Use your Head**. [S.l.]: BBC Books, 1974.
- CEJ. **Tesouro Jurídico da Justiça Federal**. Disponível em: <<http://www.cjf.jus.br/biblioteca/>>. 1997
- EDMUNDSON, H. New methods in automatic extraction. **ACM**, v. 16, p. 264–285, 1968.
- FERNÁNDEZ, M.; GÓMEZ-PÉREZ, A.; JURISTO, N. **Methontology: From ontological art towards ontological engineering**. 1997
- GRUBER, T. R. Toward **Principles for the Design of Ontologies Used for Knowledge Sharing**. Palo Alto: Kluwer Academic Publishers, 1993.
- GUARINO, N. Formal ontology and information systems. p. 3–15, 1998.
- HEWLETT-PACKARD. **Jena - A Semantic Web Framework for Java**. Disponível em: <<http://jena.sourceforge.net/>>. 2010.
- INFORMATICS, S. R. C. F. B. **Protégé**. 2010. Disponível em: <<http://protege.stanford.edu/download/download.html>>.
- MCGUINNESS, D. L.; HARMELEN, F. V. **OWL Web Ontology Language Overview**. 2004. Disponível em: <<http://www.w3.org/TR/2004/REC-owl-features-20040210/#s1.2>>.
- NETWORK, S. D. **Regular Expressions**. Disponível em: <<http://java.sun.com/docs/books/tutorial/essential/regex/index.html>>. 2010.
- PARSIA, L. C. **Pellet**. 2010. Disponível em: <<http://clarkparsia.com/pellet/download>>.
- RUSSELL, S.; NORVIG, P. **Inteligência Artificial**. [S.l.]: Campus, 2004. 1021 p.
- SALTON, G.; BUCKLEY, C. Term weighting approaches in automatic text retrieval. v. 24, 1988.
- SEBASTIANI, F. Machine learning in automated text categorization. **Computing**, v. 34, p. 1–47, 2002.
- TAN, P.-N.; STEINBACH, M.; KUMAR, V. **Introduction to Data Mining**. Addison-Wesley, 2005. Disponível em: <<http://www-users.cs.umn.edu/~kumar/dmbook/index.php>>.
- YORK, S.; STEFFEN, S.; STUDER, R. On-to-knowledge methodology (otkm). 2003.