

1.03.99 – Ciência da Computação.

ANÁLISE DE METODOLOGIA PARA MINERAÇÃO DE DADOS EM SENTENÇAS JUDICIAIS UTILIZANDO ONTOLOGIA

Antônio P. Castro Jr^{1*}, Wesley P. Calixto², Viviane M. Gomes¹, Laís F. A. Silva¹, Ernesto F. Veiga³

1. Estudantes da Escola de Engenharia Elétrica, Mecânica e de Computação (EMC) da UFG

2. Professor e pesquisador da EMC/UFG e do Instituto Federal de Goiás (IFG) / Orientador

3. Estudante do Instituto de Informática (IC) da UFG

Resumo:

O propósito deste trabalho é apresentar metodologia semântica, utilizando Ontologia, no intuito de melhorar os resultados na Mineração de Dados em Banco de Dados de Sentenças/Decisões Judiciais. Objetiva-se apresentar método inteligente e automático para buscar sentenças em processos judiciais correlatos ao que está em julgamento, utilizando Ontologia relacionada ao negócio. Para isto, foi construída Ontologia Jurídica com e sem regras do especialista. Acredita-se que este método proporcionará agilidade no Poder Judiciário Goiano, procurando resolver rapidamente o anseio da sociedade, pois poderá simular o trabalho dos Assessores Jurídicos na preparação das sentenças/decisões judiciais.

Palavras-chave: Mineração de Dados, Ontologia, Gestão do Conhecimento.

Introdução:

O número de processos no judiciário brasileiro tem aumentado de forma exponencial e superlotado as serventias, prejudicando o rápido atendimento aos direitos da sociedade. Para que essa realidade seja transformada, faz-se necessário que os recursos disponíveis supram as necessidades sociais, garantindo maior qualidade, dinamicidade e celeridade na prestação de serviços públicos, como os ofertados pelo judiciário, visando satisfazer os anseios da coletividade [1]. Estes aspectos são tratados na Emenda Constitucional nº 19/98, que inseri eficiência como um princípio da Administração Pública [2].

Dessa forma, busca-se a inserção de novas ferramentas de apoio a operação do Direito, objetivando promover o aperfeiçoamento da gestão da informação e do conhecimento no judiciário [3] [4].

A pesquisa em decisões proferidas é etapa essencial no fluxo processual, pois ajuda na elaboração de sentenças e decisões, realizada basicamente por meio dos sítios dos tribunais de justiça. A maioria das ferramentas disponíveis no Brasil para esse tipo de filtro e busca são substancialmente textuais, tornando

a pesquisa morosa, trabalhosa e pouco produtiva [5]. O trabalho de estudar o processo, seja físico ou digital, e buscar por decisões similares é feito basicamente pelos Assessores dos Magistrados, são eles que elaboram as minutas de sentenças para ajudar os Juizes na lida diária.

Estes Assessores Judiciários consomem tempo significativo no processo de estudar o processo e encontrar, por meio de busca textual, decisões semelhantes ao caso em tela. A busca textual depende muito da interpretação do funcionário, bem como da estruturação do banco de dados disponibilizado, podendo resultar em problemas na precisão da informação e na qualidade da recuperação, influenciando diretamente na minuta da decisão para estudo pelo Juiz [5].

Desse modo, torna-se imprescindível a construção de mecanismos rápidos, automáticos e inteligentes para mineração de informações nas bases de dados.

Assim, objetiva-se nesse trabalho apresentar metodologia inteligente e automática para mineração de dados em banco de dados de sentenças em processos judiciais correlatos ao que está em julgamento, utilizando pesquisa semântica com Ontologia. A intenção é melhorar a precisão e a recuperação da informação.

Metodologia:

O projeto teve início com o levantamento dos sistemas do Judiciário e as estruturas de dados (diagramas de entidades e relacionamentos) dos ambientes que permitem a pesquisa textual no banco de dados de sentenças e decisões na Internet. Foi construído ambiente tecnológico para simulação e testes de metodologia para mineração de dados usando *Ruby-on-Rails* com *Elasticsearch* e *PostgreSql*. Neste ambiente foram migrados tantos os dados do inventário de processos do 1º grau de Jurisdição, quanto a base de dados de sentenças/decisões.

Com a estrutura para simulação criada foi iniciada a próxima etapa: construção da

Ontologia Jurídica. Para essa definição semântica foi necessário definir o escopo de atuação, visto o grande volume de dados e naturezas processuais existentes no Judiciário Goiano. Para testes, ficou estabelecido criar a Ontologia para as naturezas processuais de Revisional e Consignatória. Essa decisão foi baseada no volume de processos judiciais existentes dessas classes processuais.

As Ontologias Jurídicas construídas foram implementadas no software de simulação desenvolvido, denominado Autosent, objetivando analisar a capacidade de mineração no banco de dados de decisões monocráticas.

Para verificar e comparar os resultados dos experimentos foram aplicados mecanismos para medir a capacidade do algoritmo em minerar os dados com e sem a Ontologia. As medidas utilizadas foram Precisão (*Precision*) e Sensitividade (*Recall*), conforme equações 1 e 2.

$$\text{Precisão} = \left\| \frac{\text{Verdadeiro Positivo}}{\text{Verdadeiro Positivo} + \text{Falso Positivo}} \right\|, \quad (1)$$

$$\text{Sensitividade} = \left\| \frac{\text{Verdadeiro Positivo}}{\text{Falso Negativo}} \right\|, \quad (2)$$

O Autosent foi alimentado com 100.000 sentenças/decisões e 1.757.132 processos judiciais (inventário) para realizar a mineração de dados, conforme Figura 1.

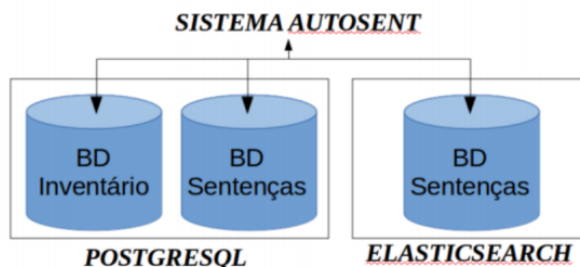


Figura 1. Infraestrutura de Banco de Dados utilizado na Simulação.

Para validação da metodologia decidiu-se utilizar uma abordagem sem Ontologia e três abordagens com Ontologia, sendo: com regras do especialista; sem regras do especialista e híbrida.

O termo "com regras do especialista" faz uso do conhecimento humano, de um especialista na área jurídica. O termo "sem regras do especialista" faz uso de modelo matemático e computacional para estruturar matriz de similaridade de palavras, objetivando encontrar a relação de palavras que ocorrem com maior frequência para busca pretendida, dentro do universo de decisões/sentenças similares ao processo em estudo. O termo

"híbrido" é quando emprega a Ontologia com e sem regras do especialista.

Para criar a Ontologia sem regras do especialista foram contabilizadas as incidências de todas as palavras com mais de 5 letras em todos os 100.000 documentos (sentenças/decisões). Com essa relação foi criada uma matriz termo a termo com as 10 palavras de maior incidência. Essas 10 palavras foram combinadas, duas a duas, aplicando a equação em 3, em cada um dos documentos, criando assim a matriz de similaridade. Na matriz de similaridade, as combinações dos termos/palavras que ficaram com o percentual de similaridade superior a 50% na matriz foram utilizadas na Ontologia sem as regras do especialista.

$$\frac{\sum_{i=1}^n w_{i,k} w_{i,h}}{\sum_{i=1}^n (w_{i,k})^2 + \sum_{i=1}^n (w_{i,h})^2 - \sum_{i=1}^n w_{i,k} w_{i,h}} \quad (3)$$

Na equação 3, o i é o número do documento de um total de n documentos, W é o termo/palavra no documento, sendo que Wk é o termo k e o Wh é o termo h .

Resultados e Discussão:

Realizadas as simulações no Autosent, sem e com a aplicação da Ontologia, tem-se os resultados da Tab. 1.

Tabela I. Resultados das Simulações

	Precisão	Sensitividade
Sem Ontologia	0.37	0.34
Com Ontologia Humana	0.64	0.54
Com Ontologia Automática	0.46	0.30
Híbrido	0.64	0.63

Observa-se que utilizando Ontologia há ganhos significativos frente a não utilização, e que a Ontologia Humana (com regras do especialista) apresentou resultados melhores que a Ontologia automática (sem regras do especialista). É relevante informar que a metodologia de mineração de dados usada para estruturar a Ontologia automática, apesar de conseguir precisão melhor que a não utilização da Ontologia, não conseguiu superar a Ontologia com regras do especialista, porém acredita-se que pode-se, ainda, utilizar outras técnicas de mineração para conseguir resultados superiores para gerar Ontologia de forma automática.

Os testes das simulações da solução chamada híbrida, com relação a métrica sensibilidade foi melhor, porém com pequenas diferenças da Ontologia gerada por regras do especialista. Acredita-se que a metodologia de mineração identificando a combinação de palavras de maior incidência conseguiu buscar

quantidade maior de sentenças/decisões no conjunto das verdadeiramente positivas.

Os resultados das simulações, com o uso da Ontologia, indicam que há melhorias tanto nas decisões verdadeiramente positivas retornadas frente aos falsos positivos retornados, quanto na quantidade de decisões verdadeiramente positivas retornadas frente todas as que são verdadeiramente corretas existentes no banco de dados, conforme Figura 2.

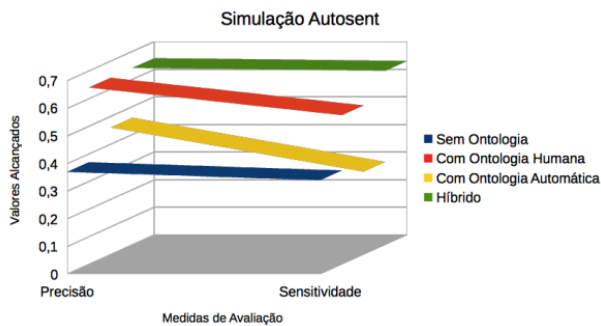


Figura 2. Valores alcançados pelas medidas de avaliação na simulação.

Conclusões:

As simulações e testes demonstraram que a aplicação da Ontologia é viável para mineração de dados em Decisões Judiciais, tanto na precisão quanto na recuperação das informações.

Outro ponto relevante é poder estabelecer a uniformização das pesquisas no universo do inteiro teor das decisões monocráticas, evitando que assuntos correlatos em processos possam ter decisões divergentes.

Considerando que a base de dados utilizada para os testes possa haver erros no protocolo/cadastro dos processos, 64% na precisão e 63% na recuperação, conseguidos na solução híbrida, são números promissores.

O trabalho ainda está em desenvolvimento e como todos os testes realizados foram para os processos de naturezas Revisional e Consignatória, há a intenção de trabalhar outras naturezas correlatas, que podem trazer resultados melhores, bem como utilizar outras metodologias em mineração de dados, tentando melhorar os indicadores das métricas, com o objetivo de superar a Ontologia com as regras do especialista.

O objetivo final deste trabalho foi alcançado, pois o modelo híbrido proposto demonstrou que a aplicação da semântica de forma automática, por meio da Ontologia, resultou conhecimento mais acertivo na mineração de dados.

Referências bibliográficas

- 1 - S. J., EGON; F. S., PAULO; R. G., EDSON. Gestão de Conhecimento para Administração Judiciária: Levantamento de Demandas de Conhecimento e Estabelecimento de Ontologias. Revista Democracia Digital e Governo Eletrônico, n. 05, 2011.
- 2 - J. R., AIRES; S. R. J., HÉLIO. O Ato Administrativo Eletrônico sob a Ótica do Princípio da Eficiência. II Conferência Sul-Americana de Ciência e Tecnologia Aplicada ao Governo Eletrônico-CONEGOV, p. 33-44, 2005.
- 3 - P. C. J., ANTÔNIO; P. C., WESLEY; F. F., BEATRIZ. Gestão da Informação em Grandes Volumes de Dados no Poder Judiciário. V Coletânea Luso-Brasileira - Gestão da Informação, Cooperação em Redes e Competitividade, Cap. 2, p. 61-78, 2014.
- 4 - S., FERNANDA. Informática e Organização do Poder Judiciário. Gestão E Conhecimento, v.3, n.2, p. 12-24, jul/dez. 2005.
- 5 - R. G. S., EDSON. Representatino of Legal Knowledge Through Ontologies: Exercise in Electronic Government. International Conference on Information Systems and Technology Management, 2009.
- 6 - C. C. M., MARIA. Ontologia legal-Estudo sober a modelagem do conhecimento legal no contexto do direito tributário. Prêmio Schontag, 2006.
- 7 - H. M., ALBERTO; C. T., AUGUSTO. Titulação Automática de Acordãos Baseado em Ontologia Jurisprudencial. Revista Democracia Digital e Governo Eletrônico, v.2, n.3, 2010.
- 8 - H., MATTHEW. A Pratical Guide to Building OWL Ontologies using PROTÉGÉ and CO-ODE Tools. The University of Manchester, 2011.
- 9 - E. R. G. S., EDSON. Representation of Legal Knowledge Through Ontologies: Exercise in Electronic Government. International Conference on Information Systems and Technology Management, 2009.
- 10 - W., BORST. Construction of Engineering Ontologies for Knowledge Sharing and Reuse. PhD thesis, University of Twente, P.O. Box 217 - 7500 AE Enschede - The Netherlands, 1997.

11 - T., GRUBER. What is an ontology. <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>, acessado em jan/2017., 2005.

12 - M., ALMEIDA and M., BAX. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. *Revista Ciência da Informação*, 32(3), 2003.

13 - E. A. M., MORAIS and A. P. L., AMBRÓSIO. Ontologias: conceitos, usos, tipos, metodologias, ferramentas e linguagens. Relatório Técnico, INF, UFG, 2007.

14 - G., SALTON and M., MCGILL. *Introduction to Modern Information Retrieval*. Computer Science Series, USA. McGraw-Hill, 1983.