

## INTRODUÇÃO

**Identificação do Órgão / Unidade / Cargo / Função:** Tribunal Regional Eleitoral do Distrito Federal / Secretaria de Tecnologia da Informação / Analista Judiciário

**E-mail para contato e envio de informações:** [symball@tre-df.gov.br](mailto:symball@tre-df.gov.br), [rrodrigues@tre-df.gov.br](mailto:rrodrigues@tre-df.gov.br), [sandra@tre-df.gov.br](mailto:sandra@tre-df.gov.br)

**Nome do trabalho / projeto:** Consulta de Jurisprudência Eleitoral na Web Baseada em Casos

**Nome do responsável:** Symball Rufino de Oliveira

**Equipe:** Rodrigo Rodrigues Barbosa, Sandra Mary dos Santos Lima

**Delimitação da ação:** Este trabalho apresenta aplicação de um modelo de consulta de jurisprudências eleitorais na web que utiliza técnica de inteligência artificial conhecida como Raciocínio Baseado em Casos. Neste modelo, as jurisprudências são organizadas sob a forma de casos jurídicos concretos. O raciocínio baseado em casos tem como princípio a idéia de que um caso jurídico passado pode ser útil para resolver um problema atual, desde que exista entre eles algum grau de semelhança. Para estabelecer semelhanças entre casos atuais e passados o modelo estudado propõe o uso de cálculo de similaridade que é realizado com base na comparação de índices temáticos obtidos a partir do processo de indexação realizado por especialistas utilizando-se como apoio um tesouro jurídico. A consulta de jurisprudências eleitorais na Web foi avaliada quanto à medida de precisão nos resultados das buscas, após a aplicação do novo modelo, obtendo resultados mais expressivos que os do modelo atual de consulta que é baseado no uso de operadores lógicos, tais como: ou (or), e (and), próximo (adj), entre outros, contribuindo com a melhora no atendimento prestado pelo TRE-DF aos profissionais do direito que utilizam este tipo de informação jurídica em suas tarefas diárias. Para avaliar a precisão no resultado das buscas por jurisprudências foram utilizados métodos consagrados na conferência de recuperação de informação textual realizada no ano de 2007.

**Objetivos e Metas:** Contribuir para a consolidação de um modelo inteligente de recuperação de jurisprudência que se estabeleça como um novo paradigma para a construção de sistemas de consulta de jurisprudências na Web que apresentem resultados de busca mais relevantes.

Melhorar o nível de satisfação dos usuários em relação à recuperação de jurisprudência eleitoral, elevando o nível das taxas atuais de precisão, diminuindo o esforço do usuário nas consultas realizadas na base de jurisprudência eleitoral do Distrito Federal.

Possibilitar que magistrados, advogados, estudantes do direito e outros interessados tenham acesso mais preciso aos acórdãos que mais se assemelhem às suas necessidades de argumentação jurídica.

## DESENVOLVIMENTO

### IDENTIFICAÇÃO DO PROBLEMA, ANÁLISE DAS PRINCIPAIS CAUSAS E PLANO DE AÇÕES DE MELHORIAS E RESULTADO ESPERADO

A fim de prover serviço de acesso às bases de jurisprudência eleitoral, o Tribunal Regional Eleitoral do DF desenvolveu um sistema de Recuperação de Informação (RI) de jurisprudência eleitoral, disponibilizado no seu sítio na Internet, que visa recuperar decisões judiciais (sentenças, acórdãos, súmulas) que sejam úteis ou relevantes para apoiar as atividades de magistrados, juristas, advogados, estudantes e cidadãos (usuários), pertencentes ou não ao quadro de pessoal do Tribunal.

De maneira geral, os usuários de sistemas de recuperação da informação formulam suas sentenças em linguagem natural, por meio de uma consulta que pode ser convertida em um conjunto de palavras-chaves formadas a partir da retirada dos termos de ligação (stopwords<sup>1</sup>) utilizados na sentença original e, ainda, pela redução das palavras à sua raiz gramatical (stemming<sup>2</sup>) (RIJSBERGEN, 1979). No caso do Tribunal Regional Eleitoral do Distrito Federal, para busca de casos jurídicos relevantes, tais palavras-chaves são então comparadas com termos índices criados a partir do processo de indexação das jurisprudências eleitorais.

Uma inconveniência desta abordagem é que na recuperação baseada em palavras-chaves geralmente há uma distância semântica entre a necessidade do usuário e o conjunto de jurisprudências que são recuperados, ou seja, nem sempre as jurisprudências recuperadas são de interesse dos usuários. Baeza-Yates e Ribeiro-Neto (1999) definem esse sistema de recuperação de informação como modelo clássico. São exemplos, o modelo booleano, o modelo vetorial, o modelo *fuzzy*. O Tribunal Regional Eleitoral do Distrito Federal utiliza-se do modelo booleano para recuperar jurisprudências eleitorais.

Maron e Blair (1985), em uma pesquisa utilizando uma coleção chamada STAIRS, acrônimo de Storage And Information Retrieval System, concluíram que modelo de sistema de recuperação de informação booleano apresentam índice de precisão máximo de 25%, ou seja, será necessário ao profissional do direito que utilize o serviço de consulta a jurisprudência eleitoral do Tribunal Regional Eleitoral do Distrito Federal ler cem casos jurídicos para poder encontrar vinte e cinco entre eles que lhe seja relevante.

Como possível solução, estudos realizados no Brasil apontam a aplicação do raciocínio baseado em casos na recuperação de jurisprudências. Um dos trabalhos mais marcantes é a tese de doutorado Weber Lee (1998) intitulada Pesquisa Jurisprudencial Inteligente. Nesse trabalho, Weber-Lee (1998) propõe o desenvolvimento de um sistema para pesquisa jurisprudencial que seja mais eficiente que os sistemas de banco de dados de texto disponíveis no mercado em termos da utilidade das decisões judiciais recuperadas.

Merece destaque também a dissertação de mestrado Bueno (1999) com o título “Recuperação da Informação Jurídica em Sistemas Baseados em Casos”. Bueno (1999), nessa dissertação, apresenta elementos que buscam estabelecer um modelo de recuperação de jurisprudência utilizando o RBC.

Mais recentemente, destaca-se a dissertação de mestrado intitulada “Proposta de Modelo RBC para a Recuperação Inteligente de Jurisprudência na Justiça Federal” de Mario Sena Braga Júnior (2001) que pretende estabelecer um modelo para construção de

---

<sup>1</sup> Palavras comuns como preposições e artigos que são removidas para facilitar o processamento da linguagem natural

<sup>2</sup> Processo de redução de uma palavra à sua parte essencial do item lexical, semelhante ao radical

sistemas de recuperação de jurisprudências no âmbito do Tribunal Regional Federal – TRF.

A introdução de uma ferramenta de recuperação de informação que utiliza o raciocínio baseado em casos apresenta-se como solução para melhorar a qualidade do resultado da pesquisa em bases de jurisprudência eleitoral, aproximando o resultado da consulta da necessidade de informação dos usuários. Contudo, a seguinte questão ainda precisa ser respondida: o uso de um sistema de recuperação de informação que utiliza o raciocínio baseado em casos, pautado no modelo de recuperação inteligente de jurisprudência (WEBER-LEE, 1998; BUENO, 1999; BRAGA JÚNIOR, 2001), produzirá melhores índices de precisão, comparados aos do modelo clássico (BLAIR e MARON, 1985), aprimorando, dessa forma, o sistema de recuperação de jurisprudência eleitoral do Tribunal Regional Eleitoral do Distrito Federal?

Com base nos estudos de Weber-Lee (1998), Bueno (1999) e Braga Júnior (2001), esse trabalho apresenta um estudo da avaliação do modelo de recuperação de informação baseado em casos a partir da sua aplicação no ambiente do Tribunal Regional Eleitoral do Distrito Federal e posterior verificação do seu índice de precisão no resultado de busca por informações jurídicas. Foi sua intenção comprovar que o modelo aplicado melhora a relevância dos resultados das buscas na base de dados textual, aproximando o resultado de uma consulta à necessidade de informação do usuário.

## MÉTODOS E TÉCNICAS

### **1. Método Aplicado à Pesquisa**

O método adotado para realização do experimento de avaliação de sistema de recuperação de informação baseado em casos baseou-se nas experiências desenvolvidas pela *Legal Track* ocorrida na *Text REtrieval Conference* do ano de 2007, conforme descrito por Voorhees (2007) e Tomlinson et al. (2007).

Foram utilizados, para este experimento: uma base de documentos, que nessa pesquisa foi identificada como uma base de casos jurídicos concretos (jurisprudências eleitorais); um grupo de especialistas participantes, que nesta pesquisa foi formado por cinco especialistas em direito eleitoral; um conjunto de consultas e/ou perguntas que estabeleceram que tipo de informação foi procurada, ou seja, quais as reais necessidades de informação dos especialistas participantes; um conjunto de critérios que constituem o julgamento de relevância de um documento recuperado (caso jurídico concreto) estabelecidos pelos especialistas participantes; um processo de avaliação da medida de precisão obtida pelo sistema de recuperação de informação baseado em casos para cada conjunto de documentos recuperados; índices de medida de precisão dos resultados da busca visando identificar o grau de acerto de cada resultado apresentado e a forma de ordenação dos documentos nestes resultados.

Nesta pesquisa apenas os dez primeiros documentos de cada resultado de recuperação serão avaliados quanto à relevância e o grau de precisão. Esta escolha foi baseada na experiência profissional dos participantes da pesquisa em relação à quantidade de documentos normalmente lidos por eles em cada resultado de busca por informações jurídicas que realizam em suas tarefas cotidianas; O valor do índice de precisão final será obtido a partir da média aritmética da precisão calculada em cada tópico para cada usuário; por fim, um relatório apresenta as considerações do pesquisador sobre o sistema de recuperação de informação avaliado.

### **2. Fases da pesquisa**

- A definição da coleção de jurisprudências que compõem o universo pesquisado;

- A definição do grupo de participantes;
- A definição das necessidades de informação dos participantes (tópicos);
- A definição da amostra que formará a base de casos jurídicos utilizada na avaliação do sistema de recuperação de informação baseado em casos;
- A construção do protótipo de recuperação de informação baseado em casos;
- Coleta de dados de precisão obtidos a partir da avaliação do protótipo de recuperação de informação;
- Análise dos resultados e discussão sobre a avaliação realizada;

### **3. Universo**

O universo inicial estudado é representado por aproximadamente 4.500 (quatro mil e quinhentos) acórdãos representando as decisões jurídicas do colegiado pleno do Tribunal Regional Eleitoral do Distrito Federal, armazenadas em banco de dados textual. Não são contempladas neste universo as decisões administrativas proferidas pelos membros do Tribunal, já que essas não foram consideradas objeto de estudo por não serem de interesse do processo eleitoral do Distrito Federal.

Em anos eleitorais, há um aumento na quantidade de decisões colegiadas proferidas pelos tribunais eleitorais e, por consequência, a procura por jurisprudências eleitorais é mais freqüente. Sendo assim, essa pesquisa optou por fazer um recorte no universo estudado que permita representar o fenômeno da recuperação de informação jurídica no âmbito do Tribunal Regional Eleitoral do Distrito Federal ocorrido em anos de eleições gerais, quando efetivamente há pleito eleitoral. Para representar esse cenário fez-se a escolha pelas jurisprudências publicadas nas últimas eleições gerais realizadas no ano de 2006. Portanto, o universo final estudado será composto por todos os acórdãos produzidos no ano de 2006, em número aproximado de 800 casos jurídicos concretos.

### **4. Especialistas participantes**

Para realização das tarefas de julgamento de relevância, definição das necessidades de informação e formulação das consultas a serem submetidas ao sistema de recuperação de informação jurídica, cinco especialistas em direito eleitoral foram convidados a participar dessa pesquisa. O critério de seleção dos especialistas considerou o desempenho de atividades diretamente relacionadas com a organização, armazenamento e recuperação de jurisprudência eleitoral.

### **5. Necessidades de informação**

Na nomenclatura da Conferência de Recuperação de Informação Textual a necessidade de informação declarada é chamada de tópico. Os tópicos permitem a utilização de diversos métodos de construção de consulta, além de permitir uma declaração clara de qual critério torna um documento relevante. O formato padrão de um tópico é: uma identificação, um título, uma descrição e uma narrativa.

Cada especialista participante da pesquisa formulou cinco necessidades de informação diferentes relacionadas com as suas reais atividades profissionais, dessa forma, a pesquisa teve 25 tópicos diferentes, que representam necessidades de informação reais dos especialistas participantes, que foram utilizados na construção da amostra de casos jurídicos concretos e também na avaliação do sistema de recuperação de informação baseado em casos que é objeto de estudo da pesquisa.

### **6. Amostra**

A amostra utilizada foi formada por um conjunto de jurisprudências obtidas a partir do universo pesquisado utilizando-se a técnica de avaliação de relevância chamada

método de *pooling*<sup>3</sup> (VOORHEES, 2007), já consagrado nas *Text Retrieval Conferences* e que foi adaptado para o caso em estudo.

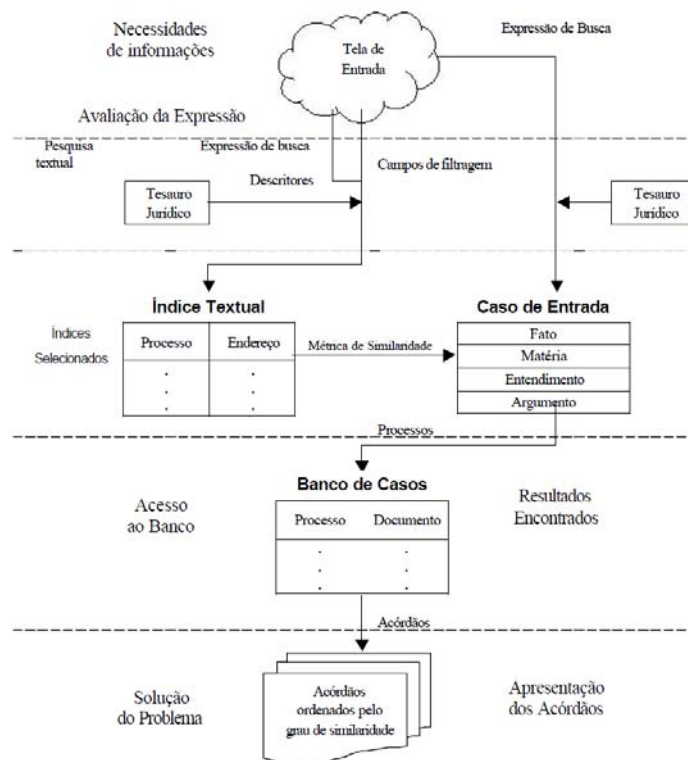
Para se obter as jurisprudências que representaram a amostra desse estudo. Foram utilizados os 25 tópicos elaborados previamente pelos 5 especialistas participantes da pesquisa, representando um padrão das suas reais necessidades de informação. Cada tópico foi convertido em consulta (query) e submetido à busca na base de jurisprudências que representava o universo da pesquisa. De cada resultado são selecionadas as 10 primeiras jurisprudências (k=10, onde k é o ponto de corte para seleção de jurisprudências que formaram a amostra). As jurisprudências selecionadas foram, então, submetidas à avaliação dos especialistas que decidem sobre a relevância de cada uma, registrando formulário específico a identificação das jurisprudências consideradas relevantes para cada tópico. Dessa forma, um especialista identifica 10 documentos para cada tópico, totalizando uma amostra de 250 jurisprudências eleitorais.

## 7. Protótipo de recuperação de informação baseada em casos

O protótipo de sistema de recuperação de informação jurídica baseado em casos, construído a partir do modelo preconizado por Braga Júnior (2001), propõe um modelo de organização e recuperação de casos jurídicos onde o cálculo de similaridade é acrescentado ao processo de recuperação. Braga Júnior (2001) propõe a criação de um motor de recuperação de jurisprudência formado por: um método de indexação que transforma jurisprudências em casos jurídicos, pelo processamento de pesquisa textual em campos de ementa dos casos jurídicos e pelo uso do cálculo de similaridade para identificação de casos jurídicos semelhantes.

Figura 1 – Modelo de Recuperação de Casos Jurídicos Baseado em Casos

Fonte: Braga Júnior (2001)



<sup>3</sup> Método utilizado para formar um conjunto de documentos utilizáveis em um experimento

## 8. Método de indexação de jurisprudências

A estrutura de uma jurisprudência é caracterizada pela ocorrência de dois grupos de elementos: descritivos e temáticos. O primeiro tem o papel de identificar fisicamente do documento, diferenciando-o dos demais. Nele encontramos elementos como: nome do relator, tipo, número e data da decisão. No segundo, estão os elementos que dizem respeito ao conteúdo do acórdão e que de acordo com os Art. 458 e 563 do Código do Processo Civil, são a ementa, o relatório, a motivação (ou fundamentação) e o dispositivo.

O método de indexação adotado utiliza quatro grupos de elementos. O primeiro e o segundo são formados por elementos de descrição física da jurisprudência. O primeiro grupo distingue-se do segundo por possuir um único elemento descritivo, o número do processo, cuja característica é a identificação única da ocorrência de um acórdão em uma coleção. A finalidade do segundo grupo é servir como filtro na seleção de casos jurídicos, permitindo um refinamento dos casos pesquisados. O terceiro grupo é formado pela ementa, que descreve o conteúdo (elemento temático) de uma jurisprudência sob a forma de resumo. Este grupo foi utilizado na pesquisa textual como um filtro que faz a seleção inicial dos casos jurídicos similares. O quarto grupo descreve o conteúdo de uma jurisprudência por meio da definição de suas categorias de análise. Guimarães (1994) define categorias de análise como elementos temáticos que identificam: o fato, a matéria (instituto jurídico), o entendimento e o argumento, contidos em uma jurisprudência. Por possuírem características fundamentais para determinação de similaridade entre os casos jurídicos, o grupo quatro foi utilizado como parâmetros na função de cálculo de similaridade e por esse motivo recebeu atribuição de pesos específicos.

Os pesos atribuídos foram assim distribuídos: 0,35 para o fato, 0,35 para a matéria, 0,20 para entendimento e 0,10 para o argumento, atribuindo um peso total igual a 1 ou 100%. No caso de não haver termos relacionados a uma determinada categoria de análise, o peso será atribuído ao fato, sendo essa a única categoria obrigada a possuir termo(s). O critério de distribuição dos pesos baseou-se na experiência do dia-a-dia de funcionários do Tribunal Regional Eleitoral do Distrito Federal que utilizam a pesquisa de jurisprudência. Eles constataram que as consultas se concentram, em sua maioria, no fato e na matéria jurídica e menos no entendimento e argumentos proferidos nos acórdãos.

## 9. O processamento da pesquisa textual

Elaborado pelo Tribunal Superior Eleitoral, esteve disponível para o uso dos especialistas durante o processo de conversão de suas necessidades de informação em consultas descritas em linguagem controladas.

O tesouro utilizado representa o vocabulário controlado da Justiça Eleitoral por meio de termos organizados em ordem alfabética, sem rigorosa relação de hierarquia entre eles.

Cada descritor do tesouro é identificado e classificado pela categoria de análise correspondente: fato, matéria, entendimento ou argumento.

O tesouro foi utilizado no processo de formulação do caso de entrada construído a partir da consulta formulada pelos especialistas participantes. O protótipo de recuperação de informação jurídica baseado em casos utilizou a consulta formulada por cada especialista no seu ato de busca por informação para convertê-la em um caso de entrada. Durante este processo são eliminadas as stopwords<sup>4</sup> e, a seguir, comparam-se as palavras da consulta com os descritores do tesouro jurídico para montagem do caso de entrada.

---

<sup>4</sup> Conhecidas também como Noise Words, são palavras retiradas do texto de uma consulta formulado em linguagem natural por um usuário em uma busca antes do seu processamento.

## 10. Medida de Similaridade

Formalmente, a medida de similaridade entre um conjunto de casos novos N e um conjunto de casos passados P pertencente ao universo U é uma função do tipo:  $\text{sim}(x,y):(N \times P \rightarrow [0,1])$ , tal que  $x \in N$  e  $y \in P$ . Se  $\text{sim}(x,y)=1$ , então a similaridade entre o caso x (novo) e o caso y (passado) é máxima, se  $\text{sim}(x,y)=0$  não há similaridade entre x e y, e se  $0 < \text{sim}(x,y) < 1$ , então existe similaridade parcial entre x e y.

### 10.1. Similaridade Local

Similaridade local determina a semelhança entre elementos temáticos, ou seja, a similaridade local está preocupada com a semelhança entre um determinado elemento temático do caso de entrada e o seu correspondente em um caso existente na base de casos. Como visto anteriormente os elementos temáticos são descritos de forma textual, contendo termos informados pelos usuários que são ajustados com o uso do tesauro jurídico. Por esse motivo, o cálculo de similaridade local foi realizado utilizando a técnica da contagem de palavras. Portanto, a fórmula da similaridade utilizada foi assim formulada:

$$\text{Sim}(Q,C) = \left[ \frac{QPI}{QPC} \right]$$

Onde: QPI representa o total de palavras idênticas encontradas na relação entre a categoria de análise do caso de entrada e a categoria de análise do caso recuperado; QPC representa a quantidade de palavras contidas na categoria de análise do caso recuperado.

### 10.2. Similaridade Global

A similaridade global determina a similaridade entre o caso de entrada e um caso existente na base de casos, permitindo identificar a utilidade de um caso existente na resolução de um determinado problema (caso de entrada). Para modelar a similaridade global utilizou-se a técnica do vizinho-mais-próximo (nearest neighbour), segunda a qual as ocorrências em uma base de casos podem ser vistas como pontos em um espaço multidimensional.

O cálculo de medida de similaridade global utilizado nessa pesquisa foi o seguinte:

$$\text{Sim}(C_e, C_i) = \sum (f_j (C_{aej}, C_{aij}) \times W_j)$$

Onde:  $C_e$  é o caso de entrada,  $C_i$  é o caso i da base de casos jurídicos,  $f_j$  é a medida de similaridade local entre  $C_{aej}$  e  $C_{aij}$ ,  $C_{aej}$  é a categoria de análise j-ésimo do caso de entrada,  $C_{aij}$  é a categoria de análise j-ésimo do caso i da base de casos,  $W_j$  é o peso atribuído à categoria de análise j-ésimo, J é o número de categorias de análise de cada caso (fato, matéria, entendimento e argumento), I é o número de casos utilizados no cálculo de similaridade.

## 11. Cálculo de Precisão

Para o cálculo da precisão no resultado da busca, utilizou-se o método da Média Principal da Precisão (MPP) e o método da Precisão-R (Prec-R) (VOORHEES, 2007). Ambos os métodos produzem como resultado a medida da precisão entre o caso jurídico recuperado e a necessidade de informação do participante. A ideia do método MPP é gerar um único valor que represente a precisão média obtida por cada um dos documentos identificados como relevantes no resultado da busca considerando-se a sua posição relativa. No método Prec-R a precisão de uma busca é calculada com base no número de documentos do resultado que serão considerados para o cálculo. Nesta pesquisa, as dez primeiras jurisprudências foram consideradas para o cálculo da precisão-R em cada resultado de busca.

## RESULTADOS E BENEFÍCIOS ALCANÇADOS

Os resultados obtidos pela pesquisa experimental são apresentados a seguir sob a forma de gráficos comparativos das medidas de precisão adotadas neste estudo.

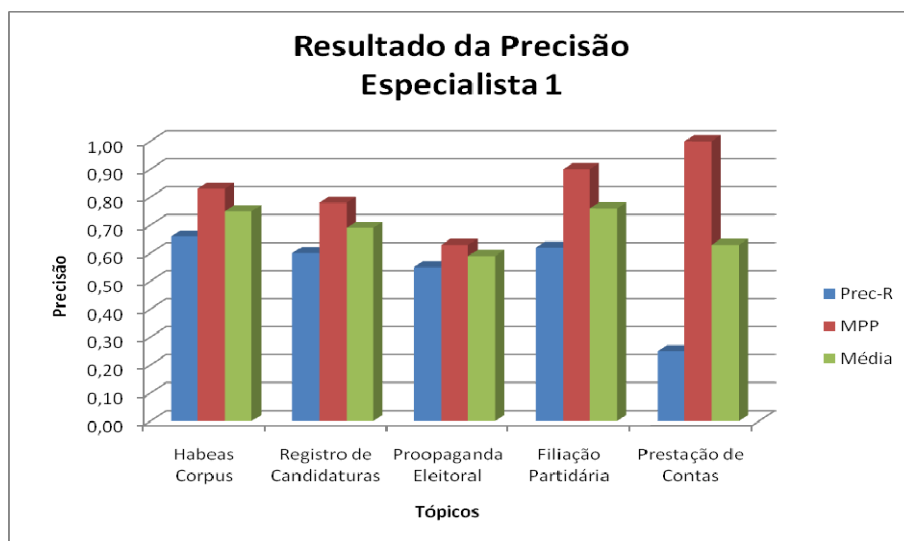
Os dados utilizados para a produção dos gráficos são resultado de um conjunto de ações de busca, realizadas utilizando-se um protótipo de sistema de recuperação de informação jurídica baseada em casos, refletindo o tipo de questionamento que os participantes da pesquisa fazem em suas práticas profissionais diárias.

Os gráficos mostram os resultados obtidos utilizando-se como base os métodos Precisão-R e Média Principal de Precisão, além de exibirem a média de precisão de cada tópico.

### 1. Especialista 1

A Figura 2 revela que a precisão baseada no método de Média Principal da Precisão foi sempre superior ao método de Precisão-R. Isso pode ser explicado pelo fato do método MPP considerar não apenas a relação entre os documentos recuperados e os documentos recuperados relevantes, mas também a posição relativa do documento relevante no resultado apresentado, ou seja, esse índice valoriza o método de ordenação utilizado. No caso desta pesquisa os resultados são ordenados de forma decrescente de similaridade entre o caso existente e o caso de entrada.

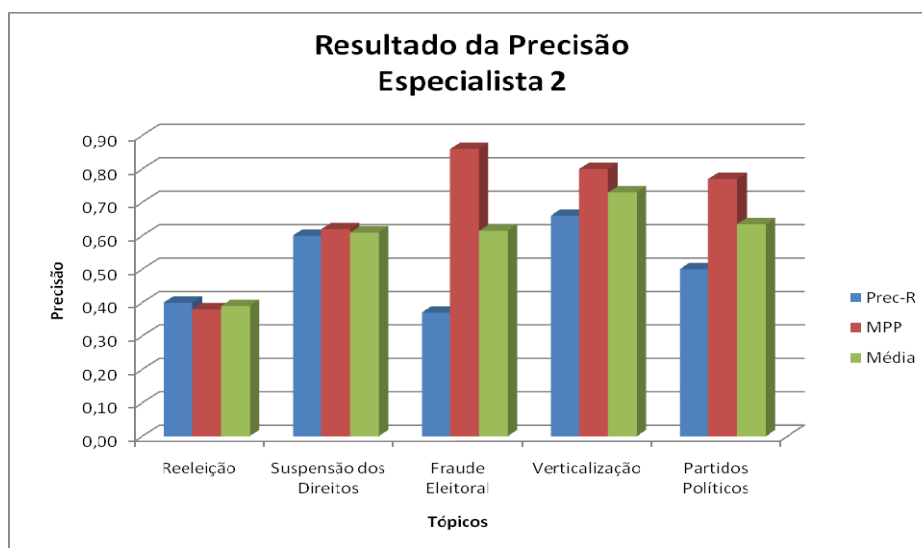
Figura 2 – Gráfico do resultado da precisão para o especialista 1



### 2. Especialista 2

A Figura 2, relativo ao índice de precisão para o especialista 3, confirma a tendência observada no gráfico anterior, quando deixa explícita a superioridade do resultado obtido pelo método Média Principal de Precisão, o que reforça a importância da ordenação do resultado da busca pelo grau de similaridade utilizado no modelo de recuperação de informação jurídica baseada em casos.

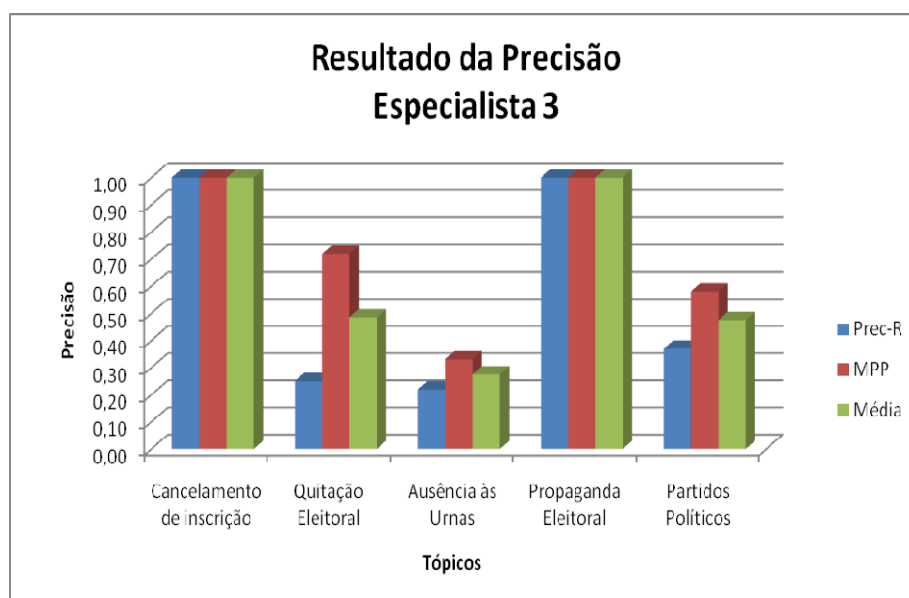
Figura 3 – Gráfico do resultado da precisão para o especialista 2



### 3. Especialista 3

Observou-se que o índice de precisão obtido pelo especialista 3 na Figura4, apresenta a ocorrência do índice máximo de precisão nos tópicos *Cancelamento de Inscrição* e *Propaganda Eleitoral*. Esse fenômeno foi classificado por esta pesquisa como definição insuficiente de documentos relevantes para um tópico. Tal fenômeno descaracteriza a aparente efetividade obtida nessa busca. Por esse motivo e para não comprometer o resultado, os índices de Precisão-R e Média Principal de Precisão desse tópico serão desconsiderados no cálculo da média final da precisão obtida pela pesquisa.

Figura 4 – Gráfico do resultado da precisão para o especialista 3

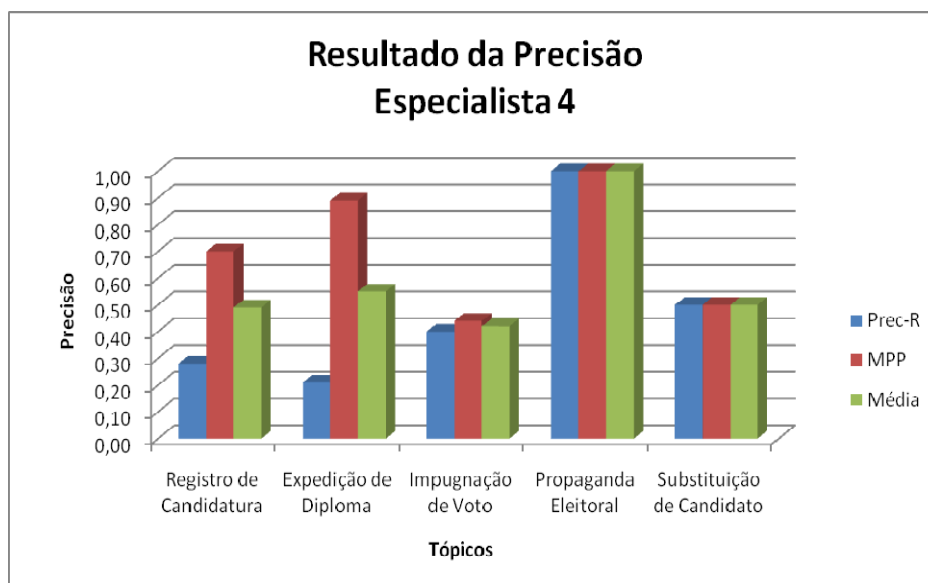


### 4. Especialista 4

Na Figura 5, observou-se valores de precisão coletados a partir dos testes realizados pelo especialista 4, onde é possível perceber que a precisão Precisão-R apresenta-se com valor inferior ou igual ao valor de precisão Média Principal de Precisão em todos os tópicos analisados. Isso caracteriza uma vantagem dos sistemas de recuperação de informação fortemente focados na ordenação dos seus resultados em

relação aos sistemas preocupados apenas em relacionar documentos relevantes em seus resultados.

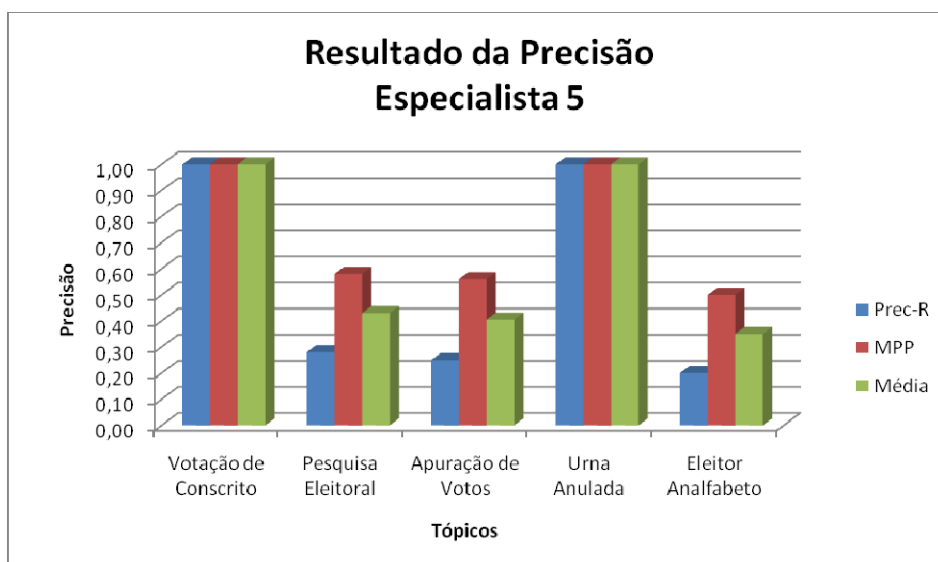
Figura 5 – Gráfico do resultado da precisão para o especialista 4



## 5. Especialista 5

Na Figura 6, ao analisarmos os valores de precisão coletados a partir dos testes realizados pelo especialista 5, podemos perceber que a combinação de similaridade com relevância permitiu produzir um modelo de recuperação de informação que apresenta uma média geral para o especialista 5 superior a 60% de precisão nos resultados obtidos.

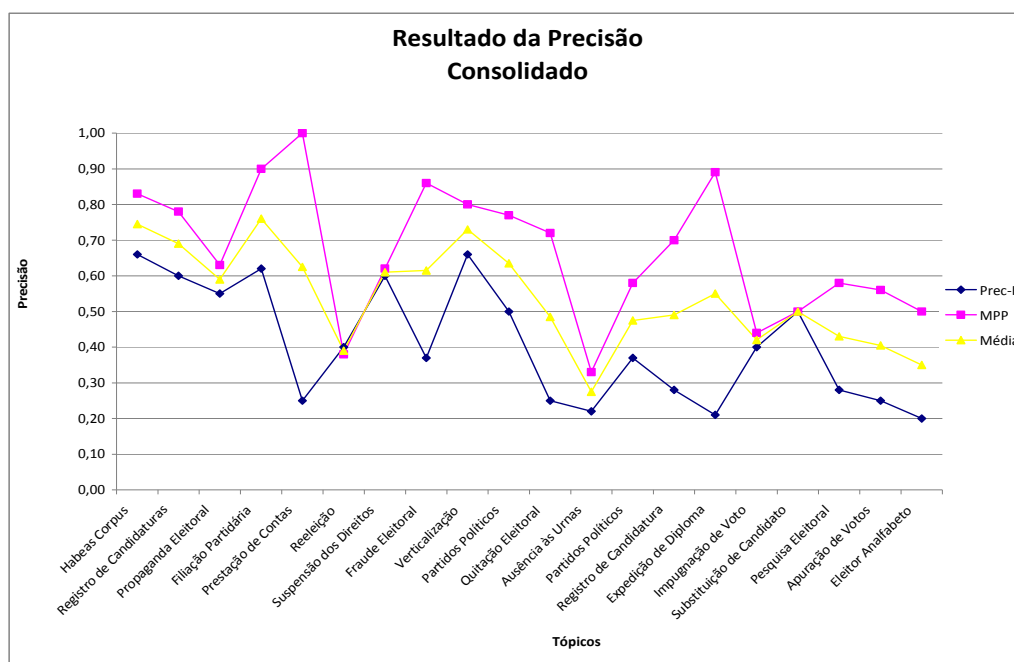
Figura 6 – Gráfico do resultado da precisão para o especialista 5



## 6. Resultado consolidado

Observando a Figura 7, é possível constatar que o menor valor registrado para a Precisão-R ocorreu no tópico *Eleitor Analfabeto* cuja recuperação foi realizada pelo especialista 5, tendo atingido 20% de precisão no resultado. Ainda observando o método de Precisão-R, os tópicos *Habeas Corpus* utilizado pelo especialista 1 e *Verticalização* pelo especialista 2, tiveram o maior valor de precisão, 66%. Com isso a média final da Precisão-R foi de 41%.

Figura 7 – Gráfico do resultado consolidado da precisão



A Média Principal da Precisão encontrou no tópico *Ausência às Urnas* recuperado pelo especialista 3 seu menor índice de precisão, que foi de 33%, tendo obtido para o especialista 1 no tópico *Prestação de Contas* o seu melhor índice, 100%. No Gráfico 6, comparando a Precisão-R com a Média Principal da Precisão para o tópico *Prestação de Cotas* percebe-se a importância da organização hierárquica do resultado obedecendo a ordem de similaridade entre os documentos. Para esse tópico, o Gráfico 6 apresenta uma Precisão-R de 25% e uma Média Principal da Precisão de 100%, caracterizando a maior distância entre esses dois índices.

## DISCUSSÃO

A Média Principal da Precisão representou melhor o pensamento do profissional do direito quando considerou mais importante encontrar os documentos relevantes no início do resultado da busca, ainda que a razão entre os documentos relevantes e o total de documentos recuperados seja pouco expressiva. Como média final, a Média Principal da Precisão obteve resultado de 67% na recuperação de documentos relevantes.

A média global da pesquisa, considerando os dois índices foi de 54%, demonstrando que as buscas por informações jurídicas realizadas pelos especialistas participantes utilizando um protótipo baseado em casos apresentaram em seus resultados, em média, mais da metade dos documentos recuperados classificados como relevantes.

Portanto, o modelo de recuperação baseado em casos utilizado nessa pesquisa apresentou-se como um mecanismo eficiente na recuperação de jurisprudência eleitoral na medida em que obteve uma média global de 54%. Trata-se de um resultado que supera os 25% apresentados por Maron e Blair (1985) nos estudos da coleção Storage and Information Retrieval System e de 29% apresentados por Voorhees (2007) nas avaliações da Legal Track 2007, ambos os resultados referentes a avaliação do modelo booleano de recuperação da informação.

## CONCLUSÃO

Os números obtidos a partir da avaliação do modelo de recuperação baseado em casos mostram que a utilização do método de organização das informações jurídicas baseados nas categorias de análise: fato, matéria, argumento e entendimento, propostas por Guimarães (1994), associado a um motor de busca baseado em casos apresentam boas perspectivas para a utilização de mecanismos busca por jurisprudências na Web, contribuindo para: definição de um novo paradigma para os sistemas de recuperação de informação jurídica; melhorar o índice de precisão no resultado da busca quando comparado ao modelo booleano de recuperação de informação jurídica; construção de um sistema de recuperação de informação jurídica que aproxime o resultado da busca da necessidade de informação dos profissionais do direito.

## REFERÊNCIAS

- Baeza-Yates, R.; Ribeiro-Neto, B. Modern Information Retrieval. Addison-Wesley, 1999.
- Braga Júnior, Mário de Sena. Proposta de Modelo RBC para a Recuperação Inteligente de Jurisprudência na Justiça Federal. Florianópolis, 2001. Dissertação (Mestrado em Engenharia de Produção) - Programa de Pós-Graduação em Engenharia de Produção, UFSC, 2001.
- Bueno, Tânia C. D. et al. Uso da teoria jurídica para recuperação em amplas bases de textos jurídicos. Programa de Pós-Graduação em Engenharia de Produção - Universidade Federal de Santa Catarina, Brasil, 1999.
- Câmara Júnior, Auto. Indexação Automática de Acórdãos por meio de processamento de linguagem natural. Brasília, 2007. Dissertação (Mestrado em Ciência da Informação) - Programa de Pós-Graduação em Ciência da Informação, UnB, 2007.
- Guimarães, José Augusto Chaves. Análise documentária em jurisprudência: subsídios para uma metodologia de indexação de acórdãos trabalhistas brasileiros. 1994. Tese (Doutorado em Ciência da Comunicação – área de Biblioteconomia) - Escola de Comunicação e Artes da USP, São Paulo.
- Maron, M. E.; Blair, David C. An Avaluation of Retrieval Efectiveness for a Full-Text Document Retrieval. Working Paper No. 364. 1985.
- Tomlinson. Stephen. OARD, Douglas W., BARON, Jason R., THOMPSON, Paul. Overview of the TREC 2007 Legal Track. 2007. Disponível em: <http://trec.nist.gov/pubs/trec16/papers/LEGAL.OVERVIEW.pdf>. Acessado em novembro de 2008.
- Voorhees, Ellen M. The Text Retrieval Conference. In Proceedings. of the 16th Text REtrieval Conference, TREC 2007, at the National Institute of Standards and Technology (NIST) November 6–9, 2007
- Weber-Lee, Rosina. Pesquisa jurisprudencial inteligente. Florianópolis, 1998. Tese (Doutorado em Engenharia de Produção) - Programa de Pós-Graduação em Engenharia de Produção, UFSC, 1998.