

InfoNorma: Um Sistema de Recomendação baseado em Tecnologias da Web Semântica

LUCAS RÊGO DRUMOND¹

ALISSON NERES LINDOSO²

ROSARIO GIRARDI³

UFMA - Universidade Federal do Maranhão

DEINF - Departamento de Informática

CEP 65085-580 São Luís (MA)

¹ ldrumond@gmail.com

² alissonlindoso@uol.com.br

³ rgirardi@deinf.ufma.br

Resumo. A enorme quantidade de dados disponível na Web, fenômeno conhecido como sobrecarga de informação, cria uma demanda por aplicações que possam selecionar, de maneira automatizada, itens relevantes para os usuários. Para responder a essa demanda, os sistemas de recomendação utilizam algoritmos de filtragem de informação para suprir as necessidades de informação a longo prazo dos usuários. Entretanto, a falta de estruturação semântica dos dados disponíveis na Web compromete seriamente a eficiência de tais algoritmos. A Web Semântica apresenta-se como uma solução viável para este problema, visando estruturar a informação na Web de modo que esta possa ser compreendida e processada pelas entidades de software. É neste contexto que o InfoNorma, um sistema multiagente de recomendações para a área jurídica que se utiliza de tecnologias da Web Semântica, foi desenvolvido. A utilização de uma fonte baseada na Web tradicional acarretou em muitas dificuldades relativas ao processamento da fonte de informação. Tais dificuldades foram superadas graças ao uso de tecnologias da Web Semântica. Este artigo apresenta o InfoNorma e como o uso das tecnologias da Web Semântica contribuiu para a melhora da efetividade do sistema no acesso à informação.

Palavras-Chave: Web Semântica, Filtragem de Informação, Sistemas de Recomendação.

InfoNorma: A Recommender System based on Semantic Web Technologies

Abstract. The huge amount of data available on the Web, a phenomenon known as information overload, generates a demand for applications able to select, in an automated way, items that are relevant to users. To answer this demand, recommender systems use information filtering algorithms to satisfy the long term information needs of users. However, the poor semantic structure of data available on the Web shortens seriously such algorithm efficiency. The Semantic Web figures as a reasonable solution to this problem aiming at structuring the information on the Web in a way that it can be understood and processed by software entities. It was in this context that InfoNorma, a multi-agent system for providing recommendations in the juridical area using Semantic Web technologies, was developed. The use of a traditional Web-based source resulted in many difficulties related to information source processing. Such difficulties were overcome through the use of Semantic Web technologies. This paper introduces the InfoNorma system and discusses how the use of Semantic Web technologies contributes to improve the system effectiveness on information access.

Keywords. Semantic Web, Information Filtering, Recommender Systems.

(Received January 27, 2006 / Accepted May 24, 2006)

1 Introdução

O acesso à grande quantidade de informação disponível na Web de maneira efetiva está se tornando um desafio cada vez maior para as aplicações de software. Isso acontece porque a Web tradicional foi desenvolvida com enfoque na maneira como a informação é exibida e pouca atenção foi dispensada à maneira como ela é estruturada. Em outras palavras, os dados na Web tradicional podem ser perfeitamente lidos e compreendidos por humanos, mas as entidades de software encontram grandes dificuldades para captar a semântica dos recursos na Web. Para solucionar este problema uma nova geração Web encontra-se em desenvolvimento: a Web Semântica [2] [3] [6].

Uma das implicações mais importantes da Web Semântica é o aumento da efetividade no acesso à informação. Uma das modalidades de acesso à informação é a filtragem, um processo utilizado para satisfazer as necessidades de informação a longo prazo de um usuário a partir de fontes de informação dinâmicas [1]. A falta de estruturação da informação na Web é uma das grandes dificuldades enfrentadas pelos mecanismos de filtragem. Problemas como a ambigüidade da linguagem natural dificultam o cálculo da similaridade entre um dado elemento de informação e um perfil de usuário, o que pode resultar em itens filtrados pouco relevantes.

Uma das modalidades de filtragem de informação é a filtragem baseada em conteúdo [4]. Neste processo, são feitas comparações entre as representações dos itens de informação e os perfis de usuário. Pode-se ter ainda perfis de usuário sendo atualizados com base na avaliação que o usuário faz de elementos de informação. Desta maneira, são filtrados itens similares a outros que o usuário tenha avaliado positivamente no passado.

O InfoNorma [12], um sistema multiagente de recomendações para a área jurídica que utiliza filtragem baseada em conteúdo foi originalmente desenvolvido com base nas tecnologias da Web tradicional. Este artigo discute as dificuldades encontradas ao se trabalhar com uma fonte de informação desestruturada e como as mesmas podem ser superadas com a introdução de tecnologias da Web Semântica, como OWL [2] [14], XML [2] [6] e as respectivas ferramentas de apoio como o JENA [13] e o Protégé [8] [14].

O artigo é organizado da seguinte forma. A seção 2 apresenta brevemente a Web Semântica. Na seção 3 é introduzida a modelagem explícita de usuários. A seção 4 faz uma breve apresentação da filtragem baseada em conteúdo. A seção 5 descreve a arquitetura e o funcionamento da aplicação. Na seção 6 são apresentadas algumas dificuldades encontradas

ao se trabalhar com tecnologias da Web Tradicional e como o uso de tecnologias da Web Semântica proporcionou a superação de alguns obstáculos. A seção 7 expõe as considerações finais e as conclusões decorrentes do desenvolvimento deste trabalho.

2 Web Semântica

É notória a incapacidade por parte das entidades de software de captar a semântica das páginas da Web pelos mecanismos de representação atualmente utilizados. A Web Semântica é uma extensão da Web atual onde a informação é estruturada de maneira que os componentes de software possam compreender o significado da informação.

Isso será viabilizado através de técnicas de representação do conhecimento como as ontologias, que são especificações explícitas de conceitualizações [9]. Com isso a Web se transformará em uma grande base de conhecimento onde serão realizadas inferências. Deste modo a Web deixará de ser apenas um meio de circulação da informação, se tornando também um ambiente para processá-la.

Com o intuito de padronizar a estruturação semântica dos dados na Web, no ano 2000 o W3C (World Wide Web Consortium) [18] definiu uma série de novas camadas para a Web. Tais camadas, mostradas na Figura 1, representam um padrão sobre o qual as aplicações da Web Semântica devem ser desenvolvidas.

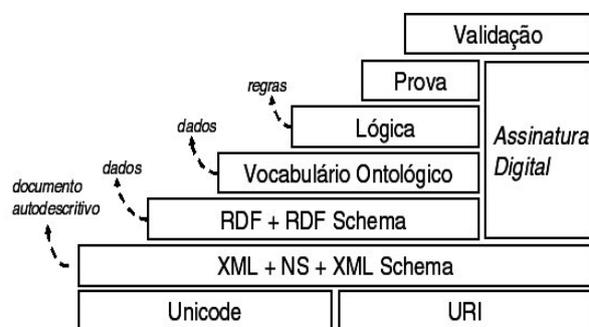


Figura 1: Camadas da Web Semântica

A camada Unicode define que os documentos Web usarão o conjunto de caracteres Unicode e os recursos na Web serão referenciados através de URIs.

Subindo um nível, a camada XML define a estrutura básica dos documentos da Web Semântica. O XML ("eXtensible Markup Language") é uma linguagem de marcação utilizada para descrever estruturalmente o conteúdo dos documentos de maneira que seja de fácil processamento pelas entidades de software e entendimento pelos humanos.

Todas as tecnologias das camadas superiores apresentam uma sintaxe baseada em XML. Assim sendo, documentos RDF e OWL podem tirar proveito dos recursos XML como transformações XSLT [6].

Devido a algumas restrições do XML quanto ao seu poder expressivo, a camada RDF [2] [5] ("Resource Description Framework") foi adicionada para descrever como os recursos se relacionam, em outras palavras, descrições RDF serializadas em XML são adicionadas aos documentos. RDF é um modelo para escrever metadados. Este modelo visualiza os dados como recursos, identificados através de URIs, e descreve as relações entre eles através de triplas do tipo <objeto, atributo, valor>.

Acima da camada RDF encontra-se a camada das ontologias. Segundo [9], uma ontologia é uma explicitação formal dos objetos e conceitos existentes em uma área de interesse, onde são expressas suas relações e restrições, através de axiomas, definindo um vocabulário comum para o compartilhamento do conhecimento. O padrão adotado pelo W3C para a representação de ontologias é o OWL ("Ontology Web Language"), que adiciona ainda mais semântica aos documentos utilizando algumas primitivas básicas do RDF e adicionando outras para que desse modo possamos escrever ontologias. O OWL, quando escrito em uma sintaxe baseada em XML, utiliza os tipos de dados do XML Schema e, utilizando recursos do XML, é possível transformar ontologias OWL através do XSL para exibi-las de maneira amigável para o usuário.

A camada de lógica define regras, escritas em XML, que serão usadas para anotar informações que seriam inviáveis de expressar em OWL. Além disso, estas regras serão utilizadas por aplicações dotadas de capacidade de raciocínio para realizar inferências sobre as ontologias OWL.

3 Modelagem Explícita de Usuários

A modelagem de usuários é uma linha de pesquisa da Engenharia de Software que estuda as formas através das quais as informações dos usuários podem ser adquiridas, representadas e usadas por sistemas computacionais.

O modelo de usuário é uma fonte de conhecimento que contém informações, explícita ou implicitamente adquiridas, de todos os aspectos relevantes do usuário para serem utilizadas por uma aplicação de software.

O conhecimento do modelo de usuário é importante, pois através dele se pode facilitar e melhorar a interação do usuário com o sistema, por exemplo, através da adaptação de interfaces ou da geração de recomendações.

Várias técnicas têm sido investigadas para abordar os problemas da modelagem explícita de usuários, algumas parcialmente tomadas de outras áreas de pesquisa e outras desenvolvidas no próprio campo da modelagem de usuários [10].

Essas técnicas são aplicadas nas três fases do processo de modelagem de usuários: a aquisição; a representação; e a manutenção de modelos de usuário [7] [11].

Na fase de aquisição, o sistema coleta informações através de entradas fornecidas explicitamente pelo usuário, sendo utilizadas para isso técnicas como questionamentos através de formulários.

As informações coletadas são então processadas na fase de representação, onde são utilizadas técnicas para a representação formal das informações adquiridas sobre o usuário, ou seja, a construção do modelo do usuário.

A fase de manutenção contempla as técnicas para a incorporação de novas informações e atualização das informações existentes no modelo, bem como técnicas para fazer inferências a partir dessas informações.

4 Filtragem Baseada em Conteúdo

A filtragem de informação é um processo utilizado para satisfazer as necessidades de informação de um usuário a longo prazo a partir de fontes de informação dinâmicas e desestruturadas [1] [4].

No processo de filtragem de informação baseado em conteúdo são feitas comparações entre as descrições dos itens de informação e os modelos de usuário. Pode-se ter ainda modelos de usuário sendo atualizados com base na avaliação que o usuário faz de elementos de informação recebidos. Desta maneira, são filtrados itens similares a outros que o usuário tenha avaliado positivamente no passado.

Uma das principais aplicações da filtragem de informação são os sistemas de recomendação. O funcionamento básico de um sistema de recomendação baseado em filtragem pode ser descrito da seguinte forma: primeiramente, são criados modelos de usuário a partir de suas preferências, que podem ser capturadas de forma implícita, observando-se o seu comportamento, ou de forma explícita, especificados através de formulários; em seguida, são criadas as representações internas dos novos elementos de informação assim que são detectados; por fim, os elementos de informação são comparados com os modelos de usuário e aqueles com maior índice de similaridade são recomendados.

Existem várias técnicas para a representação interna dos elementos de informação, tais como: o modelo de espaço vetorial; as ontologias; as árvores de decisão etc, e para cada um existem técnicas para calcular a similaridade, como o coeficiente de Pearson e a distância angular entre vetores [1].

Os elementos de informação são representados no InfoNorma através de ontologias, mais especificamente por uma ontologia de normas legais, como será visto adiante.

5 O Sistema de Recomendação InfoNorma

5.1 Especificação do InfoNorma

O InfoNorma é um sistema que visa promover a filtragem de informações relativas a instrumentos jurídico-normativos e a entrega de maneira personalizada para cada usuário, conforme os perfis de interesse que eles apresentem. Deste modo pode-se dizer que o InfoNorma tem dois requisitos principais. Um diz respeito à forma como os usuários especificam explicitamente seus interesses por certos tipos e categorias de instrumentos jurídico-normativos. O outro diz respeito ao monitoramento de uma fonte de informação e à filtragem de novos elementos. Os objetivos do InfoNorma, explicitados na Figura 2, são:

- Objetivo Geral:
 - Prover informações jurídico-normativas personalizadas
- Objetivos Específicos
 - Satisfazer necessidades de usuários legais
 - Processar novas informações legais

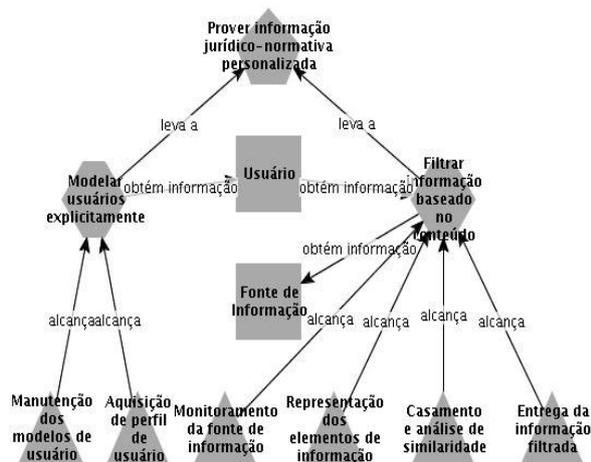


Figura 2: Modelo de Objetivos do InfoNorma

O InfoNorma é estruturado em uma arquitetura multiagente em duas camadas conforme ilustra a Figura 3: uma camada responsável pelo tratamento das informações dos usuários que utiliza os serviços da camada responsável pelo tratamento das informações vindas da fonte de informação, um conjunto de páginas Web em formato OWL contendo informações sobre instrumentos jurídico-normativos.

Além da arquitetura multiagente, o InfoNorma possui três aplicações auxiliares. A primeira delas é uma aplicação desenvolvida em Java utilizada para manusear a fonte de informação. Outra aplicação é composta por formulários que os usuários poderão preencher de acordo com seus perfis de interesse. E, por último, uma aplicação que permite ao mantenedor

do sistema, manter atualizadas as informações sobre os Ramos Jurídicos.

Os agentes que compõem as camadas são:

- Camada de processamento do usuário
 - Interfaceador: responsável por informar ao sistema quando novos perfis de usuário são especificados e por realizar a entrega das recomendações aos usuários.
 - Modelador: cria e mantém os modelos de usuário. Responsável também por extrair destes modelos as necessidades de informação para que a filtragem seja realizada.
- Camada de processamento da informação
 - Monitor: responsável por monitorar a fonte de informação e informar ao sistema quando novos elementos de informação encontram-se disponíveis.
 - Construtor: cria a representação interna dos novos elementos de informação informados pelo monitor.
 - Filtrador: realiza a filtragem em si. Através das necessidades de cada usuário fornecidas pelo modelador e dos elementos de informação fornecidos pelo construtor, o filtrador informa ao interfaceador quais elementos devem ser recomendados a quais usuários.

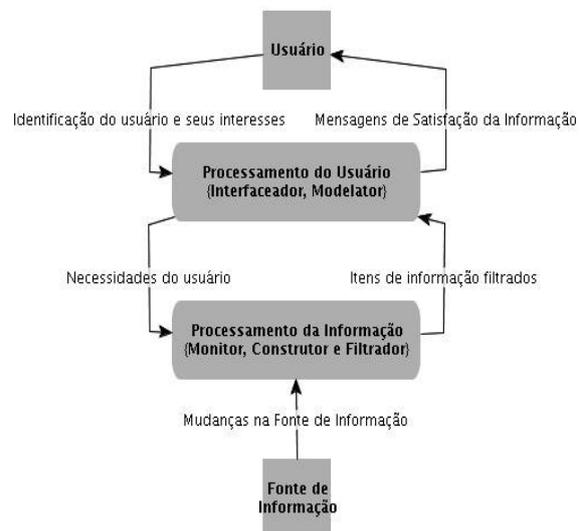


Figura 3: Modelo dos Mecanismos de Cooperação e Coordenação do InfoNorma

5.2 O Funcionamento do InfoNorma

Quanto ao funcionamento do InfoNorma, tem-se uma aplicação acessória suportada pela Web, composta por formulários que os usuários podem preencher – tanto para inserir quanto para atualizar, remover e exibir seus dados – de acordo com seus perfis de interesse,

ficando estes armazenados em um banco de dados e acessíveis à aplicação, para que ela possa modelá-los e deles inferir necessidades específicas a serem confrontadas com novas informações detectadas.

Através de tais formulários é realizada a modelagem explícita dos usuários, onde estes especificam seus interesses em determinados ramos jurídicos e tipos de instrumentos normativos. Estas informações são armazenadas em uma base de dados monitorada pelo agente modelador que utiliza esses dados para construção da representação interna dos modelos de usuário. Os modelos de usuários são comparados pelo agente filtrador com as representações internas dos elementos de informação para determinar se estes pertencem a alguma categoria de interesse do usuário em questão.

De fato, diretamente, tais registros de dados -- bem como, indiretamente, tais páginas na Internet -- representam uma face dos usuários, assim considerados como entidades que interagem com a aplicação para lhe fornecer dados de entrada ao seu processamento, posto que é por intermédio daqueles que esta pode receber a identificação e os interesses de cada usuário, através de consultas periódicas à base de dados, cujo interfaceamento ocorre com ambos, aplicação e usuário, uma independentemente do outro.

Por outro lado, para entrega dos resultados do processo de filtragem aos usuários é usado o correio eletrônico, o que retira do conjunto de requisitos concernentes à aplicação questões tais como segurança, autenticação etc, já contempladas pelo próprio serviço de e-mail.

Na verdade, as caixas de mensagens representam outra face dos usuários, assim encarados como sujeitos com que a aplicação interage para lhes oferecer dados de saída do seu processamento, visto que é através daquelas que esta é capaz de entregar as satisfações de informação a cada usuário, a qualquer instante, mesmo que ele não esteja acessando suas mensagens eletrônicas no momento.

Também de forma auxiliar, outra aplicação apóia o funcionamento do InfoNorma, no tocante à fonte de informação, que se encontra na seção de Legislação do website da Presidência da República Federativa do Brasil [15]. Isso porque nessa fonte tanto os instrumentos jurídico-normativos quanto os índices que os listam são documentos da Web tradicional, escritos na linguagem HTML, sendo, pois, bem exibidos, mas completamente desestruturados, o que dificulta o seu processamento pela aplicação.

Então, para contornar essa dificuldade, tais elementos e índices de informação foram convertidos para a Web semântica, através da construção de uma "fonte-espelho", ficando os diversos tipos de documentos normativos representados na OntoLegis --

Uma Ontologia de Normas Legais. Tal ontologia está escrita na linguagem OWL, sendo por sua vez suportada pelo modelo RDF e pela linguagem XML, resultando em documentos bem descritos e adequadamente estruturados, o que facilita o seu tratamento pela aplicação.

Contudo, para a organização dos arquivos contendo os instrumentos jurídico-normativos e seus respectivos índices, é preservada, na migração para a abordagem semântica, a hierarquia de diretórios tradicional. Assim, a decorrência dessa passagem é apenas o ganho quanto à manipulação automatizada de tais arquivos (com extensão) e diretórios (entre parênteses), que estão organizados conforme exemplificado a seguir:

- (legislacao)
 - index.owl
 - (constituicao)
 - index.owl
 - * (federal)
 - cf_1988.owl
 - cf_1967.owl
 - cf_1946.owl
 - * (estaduais)
 - ce_ma.owl
 - ce_df.owl
 - ce_sp.owl
 - (leis)
 - index.owl
 - * (ordinarias)
 - index.owl
 - (2006)
 - index.owl
 - lei_12345.owl
 - lei_12344.owl
 - lei_12343.owl
 - (2005)
 - (2004)
 - * (complementares)
 - * (delegadas)
 - (decretos)
 - (resoluções)
 - (medidas_provisorias)

Portanto, de acordo com o andamento do processo legislativo, são variáveis somente os diretórios anuais e arquivos indexatórios e instrumentais, restando fixos todos os demais. E o sistema permite, através de uma interface própria para tanto, que o mantenedor da fonte de informação instancie a ontologia para cada

norma legal, sendo que os correspondentes índices são automaticamente criados se não existirem ainda ou apenas atualizados, caso contrário. O mesmo ocorre em relação aos diretórios, sendo criados quando o documento for o primeiro do ano, ou somente utilizados quando já houver outros.

Além disso, para tornar os índices de legislação disponíveis para o monitoramento pela aplicação, há que se representar a fonte de informação como uma entidade – precisamente um objeto, cujo único atributo é um conjunto de índices, cada um composto por um endereço, uma data e um ponteiro – com a qual a aplicação possa interagir em busca de mudanças que disparem o processo de filtragem de informação.

Um detalhe adicional é que tal objeto deve ser persistente, preservando seu estado atual, mesmo que a aplicação tenha sido finalizada, até que ela seja inicializada novamente. Para isso, é usada a serialização, através da qual um objeto pode ser armazenado em arquivo e recuperado no futuro, feita com o JENA - Semantic Web Framework for Java. Além do mais, é nesse objeto serializado que são propagadas as criações e as atualizações de índices de informação.

Por fim, há uma aplicação adicional para permitir o gerenciamento de outra parte do InfoNorma, dessa vez pelo especialista no domínio jurídico, o qual associa – e mantém atualizados – conjuntos de termos-chave ponderados a cada uma das variadas categorias do Direito representadas na OntoJuris – Uma Ontologia de Ramos Jurídicos.

Considerando que tais categorias jurídicas devem estar sempre acessíveis à aplicação para a filtragem, é necessário que elas sejam representadas na forma de um objeto serializável, o qual não poderia ser transiente, posto que precisa se conservar independentemente do funcionamento da aplicação.

6 Do HTML ao OWL

Na primeira versão do InfoNorma sua fonte de informação encontrava-se em formato HTML. Isto acarretava uma série de dificuldades para recuperar a informação desejada.

A seguinte situação é usada como exemplo destas dificuldades: a aplicação necessita saber o tipo do instrumento normativo (e.g. Lei Ordinária, Decreto, Medida Provisória, etc). A página Web referente ao documento utilizado neste exemplo é mostrada na Figura 4. Esta consulta em particular é importante para o processo de filtragem, pois o usuário especifica explicitamente (através do formulário) em quais tipos de instrumentos normativos ele está interessado. Portanto, esta é uma informação muito relevante para a filtragem.

Para tanto, o componente de software responsável por obter essa informação teria que procurar, o trecho de código mostrado abaixo, e identificar que a informação desejada encontra-se logo após a abertura do elemento *STRONG*. A precariedade desse método pode ser evidenciada se, de repente, decide-se mudar a forma como a página é apresentada, ou então a simples decisão de escrever o nome do elemento com letras minúsculas como em *strong*. Em ambos os casos o código do componente de software que descobre o tipo de instrumento normativo em questão teria de ser reescrito.

```
<P align=center>
  <A href=
    "http://legislacao.planalto.gov.br/
    legislacao.nsf/View_Identificacao/
    lei\%2010.860-2004?OpenDocument">
    <FONT face=Arial color=#000080>
      <SMALL>
        <STRONG>
          LEI N<SUP>o</SUP> 10.860,
          DE 14 DE ABRIL DE 2004.
        </STRONG>
      </SMALL>
    </FONT>
  </A>
</P>
```

Se ao invés de estar disponível somente a página HTML, toda a informação estivesse devidamente estruturada, problemas como estes seriam facilmente contornáveis. Ao introduzir tecnologias da Web Semântica para estruturar a fonte de informação, a compreensão do conteúdo das páginas se tornou um problema mais fácil de se tratar.

A solução adotada para tratar o problema da falta de estruturação semântica foi reescrever as páginas HTML que constituem a fonte de informação em formato OWL numa sintaxe baseada em XML/RDF. Boa parte deste processo pode ser automatizada com o auxílio do editor de ontologias Protégé [8].

Uma das principais vantagens decorrentes do uso do OWL para estruturar a fonte de informação é que desta forma a fonte está descrita por uma ontologia semelhante à usada para representar o conhecimento dos agentes do sistema, facilitando assim a compreensão por parte dos agentes do conteúdo da fonte.

A escolha por uma sintaxe baseada em XML/RDF se deu em consequência de vários motivos. Em primeiro lugar, esta sintaxe é uma recomendação do W3C, por estar de acordo com a idéia de camadas da Web Semântica. Outro forte motivo é o fato de que esta sintaxe é suportada por inúmeras ferramentas de suporte à Web Semântica. Em terceiro lugar, a sintaxe XML/RDF possibilita que muitas consultas que não

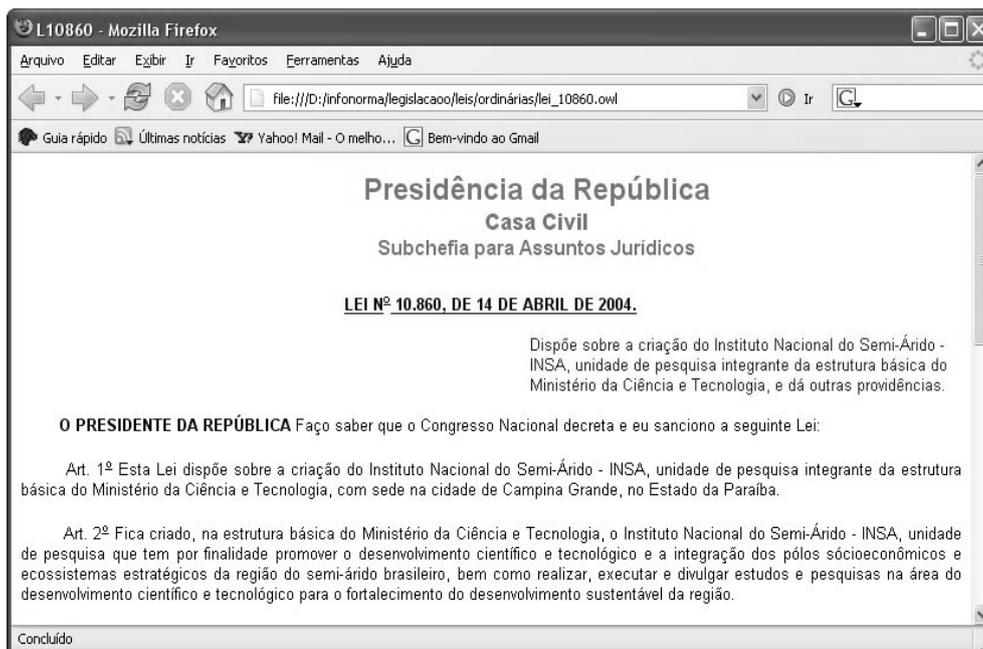


Figura 4: Exemplo de exibição de um elemento de informação

necessitem de todo o poder expressivo do OWL, sejam feitas através de linguagens destinadas ao RDF (mais simples de manipular) como o RDQL [17] e o SPARQL [16], ambas suportadas pelo JENA. Por último, serializar o OWL em XML permite que se aplique transformações XSL às ontologias OWL, ou seja, permite que as ontologias sejam exibidas de maneira amigável, sendo possível até mesmo a exibição idêntica à versão em HTML.

Na abordagem do problema enunciado no início desta seção utilizando as tecnologias aqui descritas, é possível perceber melhor as vantagens das mesmas. A recuperação do tipo do instrumento normativo em questão pode ser feita com uma simples consulta em RDQL à ontologia, como mostrado a seguir.

```
SELECT ?tipo
WHERE (inst:in_10860 legis:tipo ?tipo)
USING legis FOR
<http://www.dominio.com.br/ontologias/
ontolegis#>
inst FOR
<http://www.dominio.com.br/legislacao/
leis/ordinarias/2004/lei_10860#>
```

De modo análogo, qualquer informação disponível na ontologia, como o número do instrumento normativo, sua ementa, a quantidade de artigos ou parágrafos, entre outras, pode ser facilmente obtida através de tais consultas. Este é um dos benefícios advindos do uso das tecnologias da Web Semântica: a facilidade em recuperar a informação desejada. Entretanto, para tirar proveito dessas facilidades é preciso que se disponha de uma ontologia descrevendo o domínio da aplicação, como a

OntoLegis [12] e a OntoJuris [12] no caso do InfoNorma.

Assim fica fácil para o sistema determinar se um dado instrumento jurídico normativo é uma lei, um decreto ou uma medida provisória, bem como determinar seu número, a data de publicação, o presidente da república que o sancionou, etc.

É importante ressaltar que essas consultas não dependem da forma como a informação é exibida. Qualquer alteração na maneira de exibição do documento é feita apenas na folha de estilo XSL, não afetando em nada a estrutura da fonte e nem das aplicações que trabalham com ela.

7 Conclusão

Este trabalho apresentou o InfoNorma, um sistema de recomendações para a área jurídica.

O InfoNorma é um sistema originalmente desenvolvido para a Web tradicional que foi melhorado através do uso de tecnologias da Web Semântica para estruturar os dados da sua fonte de informação. Esta característica torna muito mais simples e intuitivo o tratamento da informação, já que esta passa a estar semanticamente estruturada.

Uma característica relevante do sistema, possibilitada pela Web Semântica, é a independência entre a maneira como a informação é estruturada e a maneira como a mesma é exibida. Isto possibilita que alterações sejam feitas no formato em que os dados são exibidos ao usuário sem provocar nenhuma alteração na estrutura dos mesmos e, em consequência disto, sem haver a necessidade de alterações nas aplicações que trabalham diretamente com a fonte.

Outro ponto importante observado no desenvolvimento deste trabalho é a disponibilidade de ferramentas para o desenvolvimento de aplicações para a Web Semântica. O amplo suporte do Protégé ao OWL [14]; o JENA, que possibilita que programas Java escrevam, leiam e consultem documentos OWL e RDF foram de fundamental importância para a realização do trabalho.

Evidenciando as dificuldades encontradas pelo InfoNorma ao se trabalhar apenas com tecnologias da Web tradicional, percebe-se as dificuldades enfrentadas pelos sistemas de acesso à informação em geral. De modo análogo, as soluções para alguns destes problemas evidenciam como a Web Semântica contribui para melhorar a efetividade do acesso à informação na Web.

Contudo, as técnicas de filtragem de informação utilizadas pelo InfoNorma não tiram total proveito das tecnologias da Web Semântica. O desenvolvimento e aplicação de técnicas de filtragem que exploram a camada lógica da Web Semântica para inferir interesses dos usuários em determinados elementos de informação é um trabalho a ser realizado no futuro.

8 Agradecimentos

Este trabalho tem o apoio do CNPq.

Referências

- [1] Adomavicius, G. and Tuzhilin, A. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(7), Jun. 2005.
- [2] Antoniou, G. and Harmelen, F. V. *A Semantic Web Primer*. MIT Press, 2004.
- [3] Berners-Lee, T., Hendler, J., and Lassila, O. The semantic Web. *Scientific American*, 284(5):34-43, 2001.
- [4] Burke, R. Knowledge-based Recommender Systems. *Encyclopedia of Library and Information Systems*, 69, 2000.
- [5] Decker, S., Mitra, P., and Melnik, S. Framework for the Semantic Web: an RDF Tutorial. *IEEE Internet Computing*, 4(6):68-73, Nov./Dez. 2000.
- [6] Drumond, L. and Girardi, R. Uma Análise das Técnicas e Ferramentas para o Desenvolvimento de Aplicações para a Web Semântica. *REIC – Revista Eletrônica de Iniciação Científica*, 6(1), Mar. 2006.
- [7] Fleming, M. and Cohen, R. User Modeling in the Design of Interactive Interface Agents. *Proceedings of UM99 (User Modeling)*, 1999.
- [8] Gennari, J., Musen, M. A., Ferguson, R. W., and et al. The Evolution of Protégé: An Environment for Knowledge-Based Systems Development. Technical report, SMI-2002-0943., 2002.
- [9] Gruber, T. R. Toward Principles for the Design of Ontologies used for Knowledge Sharing. *International Journal of Human-Computer Studies*, (43):907-928, 1995.
- [10] Kobsa, A. User Modeling and User-Adapted Interaction. *Conference Companion on Human Factors in Computing Systems*, pages 415-416, Abr. 1994.
- [11] Langley, P. User Modeling in Adaptive Interfaces. In *Proceedings of the Seventh International Conference on User Modeling*, pages 357-370, Banff, Canada, Jun. 1999. Springer.
- [12] Lindoso, A. N. Uma Metodologia Baseada em Ontologias para a Engenharia de Aplicações Multiagente. Master's thesis, Mestrado em Engenharia de Eletricidade - Área de Ciência da Computação, Universidade Federal do Maranhão (CPGEE/ UFMA), 2006.
- [13] McBride, B. Jena: a Semantic Web toolkit. *Internet Computing IEEE*, 6:55-59, 2002.
- [14] Orridge, M., Knublauch, H., Rector, A., Stevens, R., and Wroe, C. A Practical Guide to Building OWL Ontologies using the Protégé-OWL Plugin and CO-ODE Tools Edition 1.0. page 117, 2004.
- [15] Presidência da República Federativa do Brasil. Disponível em: <<http://www.planalto.gov.br>>. Acesso em: 27 jan. 2006.
- [16] Prud'hommeaux, E. and Seaborne, A. SPARQL Query Language for RDF. Disponível em: <<http://www.w3.org/TR/rdf-sparql-query/>>. Acesso em: 19 dez. 2005.
- [17] Seaborne, A. RDQL Query Language for RDF. Disponível em: <<http://www.w3.org/Submission/2004/SUBMRDQL-20040109/>>. Acesso em: 19 dez. 2005.
- [18] W3C (World Wide Web Consortium). Disponível em: <<http://www.w3.org/>>. Acesso em: 19 dez. 2005.