

2

Descrição do problema e motivação

Diversos novos desafios surgem devido ao crescimento da web-semântica. Uma vez que a informação começa a ser disponibilizada através da internet de forma semi-estruturada, surge a necessidade de criação de novas ferramentas e de adaptação das ferramentas existentes para que se torne possível o aproveitamento da semântica contida nas informações.

Um dos problemas que surgem diz respeito ao acesso a informações disponibilizadas através de bases que seguem o modelo RDF. Existem alguns padrões para o armazenamento e a distribuição das bases de dados representados em RDF. Os mais comuns são através dos padrões N3 [Berners-Lee, 1998], N-Triples² e o RDF/XML [Beckett et al, 2004], que utiliza o padrão XML [Bray et al, 2000]. O consumo das informações contidas em uma base RDF por seres-humanos, em qualquer um desses padrões, não é eficiente. São bases extensas e de difícil compreensão. É necessário um processamento por parte dos computadores para que a informação desejada seja exibida de forma fácil e prática.

Existem algumas ferramentas que se propõem a exibir dados de bases em RDF. A grande maioria realiza a exibição de forma relativamente simples, aonde o usuário acessa um recurso da base e todas as triplas que possuem aquele recurso na posição do sujeito são exibidas. A maioria dos visualizadores também permite a navegação pelos recursos da base. Outra forma comum de visualização é através de um grafo, que pode ser montado através das triplas que relacionam os recursos. Ferramentas que utilizam essas abordagens serão apresentadas e discutidas no próximo capítulo. Tanto a exibição via grafo quanto a exibição via navegação entre os recursos podem ser satisfatórias dependendo da tarefa que o usuário deseja realizar em cima dos dados. Porém, ambas as soluções possuem algumas limitações por não permitir consultas avançadas que relacionem diversos dados da base ou até de diversas bases.

² <http://www.w3.org/TR/rdf-testcases/#ntriples>

Também é fácil perceber que qualquer auxílio por parte do computador será limitado se não for desenvolvido especificamente para as ontologias utilizadas na base RDF em questão. Uma solução genérica de navegação a bases RDF não pode prover o nível de auxílio de uma solução que conheça a semântica dos dados que estão sendo exibidos. É importante lembrar que uma das motivações por trás da web semântica é que os dados passem a ser semi estruturados de forma que a máquina possa auxiliar os seres humanos a realizarem suas tarefas com maior praticidade.

Para aumentar o aproveitamento dos dados semi-estruturados disponibilizados pela web-semântica é necessário prover ferramentas adequadas para que os usuários possam realizar consultas que possuam maior complexidade. Consultas com cruzamento de dados e utilização de filtros são necessárias para resolver diversas tarefas de forma satisfatória. Por exemplo, podemos considerar a tarefa de encontrar, utilizando uma base que possua dados sobre conferências na área de web semântica, trabalhos realizados em conjunto por pesquisadores de dois institutos de pesquisa. Para realizar a tarefa de forma satisfatória, é necessário conhecer a ontologia utilizada na base, realizar cruzamento de dados e utilizar filtros. Trata-se de uma tarefa difícil de ser realizada com as ferramentas atuais.

Uma opção para a realização de consultas avançadas é a utilização de uma linguagem de consultas específica para bases RDF, definida como padrão pela W3C chamada de SPARQL [Prud'hommeaux et al, 2008]. Porém a utilização de SPARQL requer noções de programação, conhecimento da linguagem e conhecimento prévio da ontologia utilizada na base. Esses fatores tornam consultas em SPARQL muito custosas para serem utilizadas diretamente pela maioria dos usuários.

Outro fator importante que deve ser considerado por uma ferramenta que auxilie a consulta a bases RDF é a facilidade do reaproveitamento das consultas realizadas. Bases de dados contendo informações sobre o mesmo domínio devem ou seguir as mesmas ontologias ou ao menos seguir ontologias compatíveis. Ontologias podem ser relacionadas através de extensões existentes no modelo RDF (a extensão OWL provê meios de definir equivalências entre ontologias). Relacionando as ontologias de um mesmo domínio, espera-se que grande parte dos dados semânticos disponibilizados de um mesmo domínio siga a mesma ontologia ou uma ontologia compatível. Por isso é importante pensar no

reaproveitamento das consultas desenvolvidas, de forma que as ferramentas forneçam mecanismos para que grande parte das consultas possa ser reutilizada em novos dados provenientes de outras bases RDF.

Da mesma forma que a reutilização das consultas deve ser prevista, também é interessante a possibilidade do compartilhamento das consultas realizadas por um usuário. Dessa forma é possível prover para outros usuários diversas formas de consultas a informações contidas em bases de um mesmo domínio. Caberá a cada usuário escolher a forma de consulta e exibição que melhor atenda a tarefa que ele deseja realizar.

É interessante prover esse tipo de compartilhamento na ferramenta que auxilia o acesso a bases RDF porque ele permite que usuários assumam ora o papel de desenvolvedor de uma solução (quando ele está montando uma consulta própria que será compartilhada) e em outros momentos o papel de um usuário consumidor, que utiliza soluções montadas por outros usuários. Ferramentas que possuem esse tipo de interação com o usuário possuem traços de meta-design [Fischer et al, 2006]. Nesse tipo de colaboração usuários com mais experiência podem prover consultas mais complexas, que seriam utilizadas por usuários com menos conhecimento do modelo RDF.

A compreensão do modelo RDF é um obstáculo para usuários que não possuem conhecimentos na área de web-semântica conseguirem realizar consultas a bases RDF. Portanto, além dos problemas e idéias já apresentados, que fazem parte da motivação desse trabalho, existe também a preocupação de tornar mais factível para esses usuários o acesso aos dados contidos em bases RDF. A idéia principal é permitir o compartilhamento das consultas geradas através de uma visualização que abstraia por completo o modelo RDF no qual os dados são armazenados. Nesse tipo de compartilhamento, usuários que possuem conhecimentos suficientes do modelo de dados formariam consultas e a poderiam, através da ferramenta de acesso à base, gerar uma forma de compartilhamento para permitir que outros usuários pudessem ter acesso aos dados contidos na base através das consultas pré-montadas. Considerando que as consultas seriam montadas na grande maioria dos casos em cima das ontologias utilizadas na base RDF, essa forma de acesso ainda poderia ser re-utilizada por outros usuários em qualquer base que seguisse as mesmas ontologias.

Os problemas e idéias apresentados nesse capítulo motivaram a pesquisa realizada para propor uma contribuição que permita a usuários criarem consultas complexas em cima de bases RDF com o objetivo de realização de diversas tarefas. Através do compartilhamento, essas consultas poderiam ser reutilizadas e evoluídas em um ambiente cooperativo. Finalmente, o compartilhamento poderia ser estendido para possibilitar a utilização das consultas em uma forma de exibição que abstraia o modelo de representação RDF, gerando uma pequena aplicação web de consultas a dados semânticos.

No próximo capítulo são apresentadas soluções relacionadas ao trabalho proposto, bem como algumas formas alternativas de obter os resultados almejados pelos objetivos dessa pesquisa.