



UNIVERSIDADE FEDERAL DE PERNAMBUCO

GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO
CENTRO DE INFORMÁTICA
2015.1

Extensão do Dataset OpenCIn com Dados Referentes
às Notícias e Eventos Relacionados ao
Centro de Informática

PROPOSTA DE TRABALHO DE GRADUAÇÃO

Aluna: Bruna Carolina Baudel de Santana (bcbs@cin.ufpe.br)

Orientadora: Bernadette Farias Lóscio (bfl@cin.ufpe.br)

Recife, 23 de Abril de 2015

1. Contextualização

No início, a Internet surgiu como um meio de transmissão de curtas mensagens de texto entre computadores localizados em algumas universidades dos Estados Unidos. Por volta dos anos 90, Tim Berners-Lee mudou esse cenário criando a Web 2.0, consolidando uma nova maneira de usar a Internet que é usada até os dias atuais [1]. Na Web atual, os usuários podem ter acesso, a partir de navegadores, a documentos os quais são interligados por hiperlinks que levam a outros documentos, tornando essa Web mais dinâmica e amigável, diferente da Web 1.0. Esse dinamismo fez aumentar o interesse dos usuários pela Internet produzindo, fornecendo e compartilhando informações através de fóruns, blogs, sites e redes sociais. O número de usuários aumentou consideravelmente, assim como a quantidade de documentos e conteúdo. Logo, foi necessário motores de busca mais elaborados para encontrar informações relevantes para pesquisas específicas. O conteúdo da Web 2.0 é sintático sendo gerado apenas para o entendimento humano, ou seja, sem uma estrutura semântica os computadores não são capazes de inferir ou concluir informações a partir desses documentos.

Com a imensa quantidade de informação disponível de maneira aleatória e não estruturada na Web, o mesmo criador da Web 2.0 idealizou uma nova forma de organizar esse conteúdo. Como extensão da Web atual, o conceito de Web Semântica [2] iniciou uma revolução na maneira que os computadores consomem as informações. Os documentos são estruturados como dados interligados e podem ser legíveis por máquinas. Os dados recebem significados através de suas interligações e o conteúdo publicado começa a ter sentido não apenas para humanos, mas também para os computadores e, assim, ser consumidos de maneira mais eficiente melhorando o problema da busca por informação relevante na Internet.

Tecnologias como RDF (Resource Description Framework), SPARQL (SPARQL Protocol and RDF Query Language), OWL (Web Ontology Language) e XML (Extensible Markup Language) foram desenvolvidas pela W3C (World Wide Web Consortium) [3] para dar suporte à Web Semântica. O uso dessas tecnologias permite que a Web de documentos se torne uma Web de dados ligados semanticamente, a qual possibilita que as máquinas utilizem melhor o domínio dos dados interligados inferindo conteúdo mais útil para os sistemas. Para estruturar e ligar os dados, o conceito do Linked Data [4] foi introduzido e é parte fundamental da Web 3.0. Esse conceito consiste em um conjunto de boas práticas usadas para expor, compartilhar e conectar dados relacionados não previamente ligados usando tecnologias como URIs (Uniform Resource Identifier) e RDF.

Para os dados estarem disponíveis para todos, existe o conceito de Dados Abertos [5]. A ideia é que qualquer pessoa pode usar e publicar sem restrições de mecanismos de controle, como direitos autorais ou patentes. Atualmente, existe várias fontes, como a área acadêmica, governo e empresas privadas, que fazem a divulgação desses dados através da criação de portais e aplicações que maximiza a acessibilidade de conteúdo e conhecimento para todos.

Atualmente, o Centro de Informática (CIn) [6] possui um portal completo com notícias, eventos, informações sobre o centro, porém esse conteúdo está disponível de maneira sintática e não estruturada. Esta proposta oferece ao CIn um repositório com os dados de notícias, teses e dissertações descritos em RDF, que poderá ser usado para a construção de um portal semântico para o CIn ou outras aplicações que possam usar uma base de dados abertos com essas informações.

2. Objetivos

A finalidade desse projeto é estender o dataset OpenCIn com informações relevantes sobre teses, dissertações e notícias relacionadas ao Centro de Informática. O OpenCIn é um conjunto de dados em RDF que unifica e estrutura informações acerca dos docentes do CIn. Atualmente, o portal do CIn disponibiliza essas informações de forma sintática e estática, apenas para a exibição dos dados em formato HTML. O objetivo, então, é criar um conjunto de dados interligados para estruturar essas informações de maneira semântica facilitando a busca por conteúdo sobre o CIn, suas conquistas, projetos, discentes, docentes e as relações entre eles.

Para descrever os principais conceitos relacionados ao domínio das notícias e eventos disponíveis no portal do CIn, a ontologia do projeto OpenCIn será estendida usando a metodologia 101. Os termos já conhecidos para a criação de classes e propriedades serão reutilizados, mas, se preciso, novos termos serão criados. O conjunto de dados será armazenado no formato de triplas RDF e estará disponível para consultas através da linguagem SPARQL. Esses dados podem servir de base para novas aplicações como o portal semântico do CIn, entre outras.

3. Cronograma

A tabela 1 apresenta o cronograma do Trabalho de Graduação.

Atividade	Março			Abril				Maio			Junho			Julho				
Formulação da proposta e análise da literatura	X	X	X	X	X	X	X	X										
Construção do esquema								X	X	X	X	X	X					
Povoamento do dataset										X	X	X	X	X	X			
Publicação do dataset em SPARQL Endpoint													X	X	X	X	X	
Preparação para apresentação e defesa														X	X	X	X	

Tabela 1: Cronograma do Trabalho de Graduação

4. Possíveis Avaliadores

Os possíveis avaliadores para o trabalho de graduação serão:

- Bernadette Farias Lóscio
- Ana Carolina Salgado

5. Referências

- [1] W3C - WORLD WIDE WEB CONSORTIUM, Tim Berners-Lee. Disponível em: <<http://www.w3.org/People/Berners-Lee/>>. Acesso em: 15 de abril 2015.
- [2] BERNERS-LEE, T, HENDLER, J. and ORA, L. 2001. The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. Scientific American, Maio 2001.
- [3] W3C - WORLD WIDE WEB CONSORTIUM. Disponível em: <<http://www.w3.org/>>. Acesso em: 15 de abril 2015.
- [4] BIZER, C., HEATH, T. and BERNERS-LEE, T. Linked data - the story so far. In Proceedings of the International Journal on Semantic Web and Information Systems. 2009. 5(3), 1-22.
- [5] WIKIPEDIA, A ENCICLOPÉDIA LIVRE, Dados Abertos. Disponível em: <http://pt.wikipedia.org/wiki/Dados_abertos>. Acesso em: 15 de abril 2015.
- [6] CIN – CENTRO DE INFORMÁTICA DA UFPE. Disponível em: <<http://www2.cin.ufpe.br/site/index.php>>. Acesso em: 15 de abril 2015.

6. Assinaturas

Bruna Carolina Baudel de Santana
Orientando

Bernadette Farias Lóscio
Orientador