



UNIVERSIDADE FEDERAL DE
PERNAMBUCO

GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO
CENTRO DE INFORMÁTICA

2014.2

OpenCIn – Dados Abertos e Interligados Acerca dos
Docentes do Centro de Informática
PROPOSTA DE TRABALHO DE GRADUAÇÃO

Aluno: Jônatas de Lira Rocha (jlr@cin.ufpe.br)

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

Recife, 14 de Outubro de 2014

1.Contextualização

A Web é uma das ferramentas mais usadas no mundo para estudo, trabalho e diversão. No início, a Web era composta apenas por documentos ligados através de hiperlinks mas com o tempo ela foi evoluindo para uma plataforma de conteúdo dinâmico gerado pelos usuários [1]. O fenômeno da Internet como meio de interação, colaboração e compartilhamento é denominada de Web 2.0 e foi impulsionada pelas redes sociais, blogs, fóruns e wikis contrapondo a Web 1.0 onde os usuários não geravam conteúdo, eram apenas expectadores. Por conta dessa evolução, o uso da Internet cresceu significativamente e logo o volume de dados na Web também, tornando-se necessário o aprimoramento das ferramentas de busca para que sejam encontrados conteúdos relevantes. As tecnologias usadas na Web 2.0 permitem que sejam gerados conteúdos para que humanos entendam, mas os computadores não conseguem inferir nem concluir nada acerca dos documentos, pois eles não possuem nenhuma estrutura semântica.

Para superar esse problema e permitir a melhor interação entre os documentos da Web tornando-os legíveis para as máquinas, foi criado o conceito de Web Semântica, que é uma extensão da Web atual e fornece um conjunto de tecnologias que permitem a organização dos dados para que máquinas possam entendê-los [2]. A intenção é que seja incluída em cada documento da Web seu respectivo conteúdo semântico, permitindo que os computadores entendam o documento e possam fazer inferências e conexões acerca do conteúdo apresentado, enriquecendo assim a busca pela informação desejada.

Várias tecnologias foram desenvolvidas pela W3C (Consórcio *World Wide Web*) para tornar a Web de documentos em uma Web de dados ligados, como por exemplo, o RDF (*Resource Description Framework*), SPARQL (*SPARQL Protocol and RDF Query Language*), OWL (*Web Ontology Language*) e XML (*Extensible Markup Language*). O uso destas tecnologias possibilita que computadores façam coisas mais úteis por causa da interação entre os dados disponíveis na Web e já estão sendo usadas por diversos sistemas. O coração da Web semântica é o *Linked Data*, que é um conjunto de boas práticas para publicação de dados estruturados, usando a estrutura geral da Web [4]. Essas práticas incluem dar nome as coisas usando URIs, usar URIs HTTP para outros possam demandar mais informações sobre essas coisas e quando as informações forem solicitadas através das URIs devem ser retornadas informações úteis em RDF.

Além dos dados serem entendíveis por máquinas e ligados com outros dados relevantes, há uma outra tendência de tornar os dados disponíveis para que todos usem e publiquem sem mecanismos de controle [5], chamado dados abertos. Atualmente, várias organizações governamentais e não-governamentais estão criando seus portais de dados abertos e disponibilizando informações úteis para qualquer pessoa sem restrições de direitos autorais ou patentes.

Apesar do Centro de Informática ser considerado um dos mais competentes centros de informática em toda a América Latina [3], ele ainda

não possui um portal com o dados abertos com as informações descritas em RDF, por isso meu trabalho propõe um repositório de dados abertos sobre os docentes do centro. Quando concluído, esse repositório irá trazer várias informações relevantes sobre os professores e permitir que sejam produzidas várias aplicações usando essa base de dados.

2.Objetivos

O objetivo desse trabalho é criar um conjunto de dados com informações relevantes acerca dos docentes do Centro de Informática, chamado OpenCIn, unindo informações de diversas fontes de dados. Atualmente, as informações importantes acerca dos docentes estão distribuídas em diversos provedores de conteúdo, por exemplo: Os dados acerca das publicações dos professores estão na plataforma Lattes, enquanto os dados do professor na universidade estão armazenados no site do Centro de Informática. O OpenCIn agrega as informações desses sites em um só banco de dados interligados, que aborda desde os dados básicos do professor, como nome, telefone e email até as publicações do docente.

Será criada uma ontologia em OWL usando a metodologia 101, a fim de descrever os principais conceitos relacionados ao domínio de docentes. Na criação das classes e propriedades, sempre que possível, serão reutilizados termos já existentes nos vocabulários já conhecidos. Os dados serão armazenados em triplas RDF e colocados num banco de dados para consultas em SPARQL. Além disso, será criada uma aplicação Web em HTML para acessar os dados do *dataset* por pessoas que não tem conhecimento em SPARQL.

3.Cronograma

A tabela 1 apresenta o cronograma do Trabalho de Graduação.

Atividade	Outubro				Novembro				Dezembro				Janeiro			
Formulação da proposta e análise da literatura	X	X	X	X												
Construção do esquema			X	X	X											
Povoamento do dataset						X	X									
Publicação do dataset em SPARQL Endpoint								X	X							
Desenvolvimento da aplicação e visualização de dados									X	X	X	X				
Preparação para apresentação e defesa													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 BRASIL – WORLD WIDE WEB CONSORTIUM ESCRITÓRIO BRASIL, 2011. Disponível em: <<http://www.w3c.br/palestras/internet-web-jun-jul-2008/internetWeb-02.html>> Acesso em: 8 de out. 2014.

[2] W3C BRASIL – WORLD WIDE WEB CONSORTIUM ESCRITÓRIO BRASIL, 2011. Disponível em: <<http://www.w3c.br/Padroes/WebSemantica>>. Acesso em: 8 de out. 2014.

[3] CIN – CENTRO DE INFORMÁTICA DA UFPE, 2014. Disponível em: <<http://www2.cin.ufpe.br/site/secao.php?s=1&c=1>>. Acesso em: 13 de out. 2014.

[4] DADOS LIGADOS (LINKED DATA) - CONSEGI, 2011. Disponível em: <<http://pt.slideshare.net/pablomendes/dados-ligados-linked-data-consegi-2011>>. Acesso em: 20 de out. 2014.

[5] WIKIPEDIA, A ENCICLOPÉDIA LIVRE, Dados Abertos, 2014. Disponível em: <http://pt.wikipedia.org/wiki/Dados_abertos>. Acesso em: 20 de out. 2014.

6. Assinaturas

Jônatas de Lira Rocha
Orientando

Bernadette Farias Lóscio
Orientador