



UNIVERSIDADE FEDERAL DE PERNAMBUCO
GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO
CENTRO DE INFORMÁTICA
2013.1

Criação e publicação de um *linked lataset* sobre a estratificação da qualidade da produção intelectual dos programas de pós-graduação *strictu sensu* da Capes.

PROPOSTA DE TRABALHO DE GRADUAÇÃO

Aluno: Victor Hugo Carvalho de Amorim (vhca@cin.ufpe.br)

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

Recife, 17 de Julho de 2013.

1. Contextualização

A web, uma das principais ferramentas da atualidade, é composta por uma crescente quantidade de dados que em sua maioria possui apenas a finalidade de apresentação para o usuário final, como as páginas web acessadas diariamente. As tecnologias usadas na criação e formatação dessas páginas conseguem, de fato, atingir o objetivo final de propiciar conteúdo o mais compreensível possível para o usuário, mas em contrapartida o torna de difícil interpretação por parte dos computadores devido à falta de uma semântica agregada a esse conteúdo, impedindo que a web seja usada em sua plenitude.

Para ajudar a resolver esse problema surgiram vários conceitos/tecnologias e um deles foi a web semântica, que seria uma evolução da web atual onde agora a preocupação não seria apenas com a melhor formatação visual, mas também com a melhor estruturação dos dados dispostos buscando que contenham um significado semântico. Dessa forma pessoas e máquinas podem interpretar exatamente a mesma coisa, sempre trabalhando em cooperação. Na web semântica, os dados podem ser acessados pela arquitetura web atual e devem apresentar relacionamentos uns com os outros [1]. O propósito da web semântica pode ser obtido através do uso de tecnologias integradas e do uso de metadados [2] - dados sobre outros dados – provendo informação semântica e permitindo aos computadores interpretarem o conteúdo disponível na web.

Para tornar a web interpretável é preciso que uma enorme quantidade esteja estruturada na web com valor semântico, ou seja, acessíveis e gerenciáveis pelas ferramentas/tecnologias que compõem a web semântica. Além disso, é necessário que esses mesmos dados estejam interligados para que ocorra a criação da web de dados [4]. A partir dessa necessidade surgiu-se o conceito de *Linked Data* ou dados ligados. *Linked Data* faz uso de tecnologias web padrões (HTTP, URI), porém, ao invés de apenas servir páginas web às pessoas, ela permite a leitura dos dados conectados de forma automática por um computador, permitindo, desta forma, que dados de diferentes fontes de dados possam ser conectados e consultados [5].

2. Objetivos

Este trabalho tem como objetivo a criação de um *dataset* com dados relevantes sobre a estratificação da qualidade da produção intelectual dos programas de pós-graduação *strictu sensu* da Capes no triênio de 2011 a 2013 e sua publicação através de um SPARQL *Endpoint*, bem como o desenvolvimento de uma aplicação para a visualização dos respectivos dados. Além do *dataset*, será construída uma ontologia de forma a descrever o domínio das informações das conferências e periódicos, usando sempre que possível termos já definidos em vocabulários conhecidos e adicionando novos sempre que necessário. Dada a importância da ligação com diferentes bases relacionadas para a formação da web semântica, o *dataset* construído utilizará os princípios do *linked data* e será interligado com uma base de publicações bibliográficas em RDF, de modo a prover navegabilidade através de links RDF e o descobrimento de novas informações a partir dos dados já existentes.

3. Cronograma

O cronograma de atividades previsto para o desenvolvimento deste Trabalho de Graduação é apresentado na Tabela 1.

Atividade	Maio		Junho				Julho			Agosto				Setembro			
			X	X	X	X	X	X									
Estudo da Literatura			X	X	X	X	X										
Especificação da solução								X	X								
Implementação da solução									X	X	X	X					
Elaboração do relatório											X	X	X	X	X	X	
Validação e testes da solução													X	X			
Preparação da apresentação																X	X
Defesa																	

Tabela 1: Cronograma de atividades

4. Possíveis Avaliadores

Os possíveis avaliadores desse trabalho de graduação serão:

- Fernando Fonseca
- Bernadette Farias Lóscio

5. Referências

- [1] HERMAN, I. W3C Semantic Web Frequently Asked Questions. W3C, 2001. (atualização 2009). Disponível em: <<http://www.w3.org/2001/sw/SW-FAQ>>. Acesso em: 15 jul. 2013.
- [2] WIKIPÉDIA, A ENCICLOPÉDIA LIVRE. Web semântica. , 2012. Disponível em: <http://pt.wikipedia.org/wiki/Web_semântica>. Acesso em: 15 jul. 2013.
- [3] Berners-Lee, T. (2006, July). Linked Data - Design Issues. Disponível em: <<http://www.w3.org/DesignIssues/LinkedData.html>>. Acesso em: 15 jul. 2013.
- [4] W3C. Linked data. , 2012. Disponível em: <<http://www.w3.org/standards/semanticweb/data>>. Acesso em: 15 jul. 2013.
- [5] WIKIPÉDIA, THE FREE ENCYCLOPEDIA. Linked data. , 2012. Disponível em: <http://en.wikipedia.org/wiki/Linked_data>. Acesso em: 15 jul. 2013.

6. Assinaturas

Recife, 17 de julho de 2013.

Victor Hugo Carvalho de Amorim
Orientando

Bernadette Farias Lóscio
Orientador