Universidade Federal de Pernambuco

Graduação em Ciência da Computação

Centro de Informática

2013.2

Criação e publicação de um dataset de dados interligados para alocações de docentes do CIn-UFPE

**Proposta de Trabalho de Graduação**

**Aluna:** Thaís Mota dos Santos (tms2@cin.ufpe.br)

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

**Recife, 27 de novembro de 2013**

Índice

[1. Contexto 3](#_Toc373619183)

[2. Objetivo 3](#_Toc373619184)

[3. Cronograma 4](#_Toc373619185)

[4. Possíveis avaliadores 4](#_Toc373619186)

[5. Referências 4](#_Toc373619187)

[6. Assinaturas 5](#_Toc373619188)

# Contexto

A web, desde que começou a ser difundida, vem ocupando um espaço cada vez maior na vida das pessoas. A quantidade de dados dispersos nela é assustadoramente grande e vem crescendo mais e mais. As páginas web são a maneira mais comum de apresentação desses dados para as pessoas. Além de dados, uma página dessas é também composta por linguagens que definem, para o computador, como a informação deve estar estruturada e o comportamento que a página deve apresentar (respostas a cliques, etc). Essas tecnologias são utilizadas, principalmente, com a finalidade de deixar o conteúdo compreensível para quem a lê. No entanto, esse conteúdo é de difícil interpretação para uma máquina, pois ela não consegue identificar a semântica relacionada a ele, o que faz como que esses dados não sejam usados de forma integral.

Na tentativa de oferecer um significado interpretável aos dados disponíveis na web por parte das máquinas, uma série de novas tecnologias e conceitos surgiu, e um deles é o conceito de web semântica. A web semântica ou, também, web de dados vem como forma de evolução à web atual e tem por objetivo dar significado aos dados de forma que humanos e máquinas possam trabalhar em conjunto [1]. A web semântica conta com várias tecnologias utilizadas de forma integradas, tais como Resource Description Framework (RDF), Web Ontology Language (OWL), Extensible Markup Language (XML), entre outros, e também metadados, que são dados que descrevem outros dados [2]. Tudo isso em conjunto provê dados que terão sentido tanto para pessoas quanto para computadores.

A web semântica só será possível se os dados presentes na web forem aos poucos postos em formatos padrões que possam ser acessíveis e gerenciáveis pelas tecnologias usadas por ela. Mais ainda, é preciso que esses dados estejam ligados entre si [3]. Para atender às necessidades de criação de padrões de publicação de dados e de interligação entre eles, aparece o conceito de dados ligados (*Linked Data*, em inglês). Dados ligados fazem uso de tecnologias já difundidas na web como HTTP e URI, porém, ao invés de apenas servir páginas web às pessoas, permitem 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 [4].

# Objetivo

O objetivo deste trabalho de graduação é o de criar e publicar um dataset de dados ligados com informações relevantes a respeito das alocações dos docentes do CIn-UFPE. Como forma de alcançar esse objetivo maior, ações como a criação de uma nova ontologia com novos termos sobre o domínio de alocações, a criação de um conjunto de dados seguindo o princípio de *Linked Data* para futura utilização e a disponibilização de um SPARQL *Endpoint* para eventuais consultas deverão ser feitas.

# Cronograma

O cronograma de atividades previsto para o desenvolvimento deste trabalho de graduação é apresentado na tabela a seguir (Tabela 1). O período de cada mês está divido em colunas que representam o período de uma semana.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Atividade | Novembro | Dezembro | Janeiro | Fevereiro |
| Formulação da proposta e análise da literatura | **X** | **X** |  |  |  |  |  |  |  |  |  |  |  |
| Construção do esquema |  | **X** | **X** |  |  |  |  |  |  |  |  |  |  |
| Povoamento do dataset |  |  |  | **X** | **X** | **X** | **X** |  |  |  |  |  |  |
| Elaboração do relatório final |  |  |  | **X** | **X** | **X** | **X** | **X** | **X** |  |  |  |  |
| Preparação para apresentação e defesa |  |  |  |  |  |  |  |  |  | **X** | **X** |  |  |

Tabela 1: Cronograma de atividades

# Possíveis avaliadores

Os possíveis avaliadores desse trabalho de graduação são Bernadette Farias Lóscio (bfl@cin.ufpe.br) e Ana Carolina Brandão Salgado (acs@cin.ufpe.br).

# 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: 27 nov. 2013.

[2] WIKIPÉDIA, A ENCICLOPÉDIA LIVRE. Web semântica. , 2012. Disponível em: <[http://pt.wikipedia.org/wiki/Web\_semântica](http://pt.wikipedia.org/wiki/Criptografia_qu%C3%A2ntica)>. Acesso em: 27 nov. 2013.

[3] W3C. Linked data. , 2012. Disponível em: <http://www.w3.org/standards/semanticweb/data>. Acesso em: 27 nov. 2013.

[4] WIKIPÉDIA, THE FREE ENCYCLOPEDIA. Linked data. , 2012. Disponível em: <http://en.wikipedia.org/wiki/Linked\_data>. Acesso em: 27 nov. 2013.

# Assinaturas

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Thaís Mota dos Santos

**Aluna**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Bernadette Farias Lóscio

**Orientadora**