



UNIVERSIDADE FEDERAL DE PERNAMBUCO
GRADUAÇÃO EM SISTEMAS DE INFORMAÇÃO
CENTRO DE INFORMÁTICA
2016.1

Uma Aplicação para Consumo de Dados Abertos Conectados Sobre os
Eventos da Sociedade Brasileira de Computação

PROPOSTA DE TRABALHO DE GRADUAÇÃO

Aluno: Rayelle Ingrid Vera Cruz Silva (rivcs@cin.ufpe.br)

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

Recife, 12 de Abril de 2016

1. Contextualização

A Web surgiu a partir da dificuldade existente de se compartilhar informações. Na época, para ter acesso a informações que estivessem em diferentes computadores, era necessário conectar-se a cada um deles, pois não havia um meio que tornasse essa troca de informações mais rápida ou, até mesmo, simultânea [1]. Quando Tim Berners-Lee propôs a World Wide Web, em sua versão 1.0, considerada a Web read-only, no final dos anos 80, a forma de acessar dados mudou completamente; quando antes era necessário se deslocar para adquirir informações, agora, a informação poderia ser acessada de maneira mais rápida.

Berners-Lee definiu três tecnologias fundamentais ainda na versão 1.0, e que até hoje constituem os fundamentos da Web, são elas: o *HyperText Markup Language* (HTML), linguagem de marcação para Web, *Uniform Resource Identifier* (URI), identificador de recursos na Web, e o *HyperText Transfer Protocol*, protocolo de aplicação utilizado para requisições na Web [1]. A motivação da Web 1.0 era simplesmente uma maneira de disponibilizar conteúdo, uma Web informativa, ou seja, o usuário tinha acesso ao dado, mas não era capaz de gerar conteúdo.

Tendo em vista esta limitação, a Web partiu para sua versão 2.0, considerada a Web *read-write* ou Web de comunicação [3], onde o usuário é o principal gerador de conteúdo, seja por meio de *blogs*, redes sociais e em diversos outros meios.

Atualmente, as pessoas têm acesso a muito mais dados em um único dia, do que a maioria das pessoas teve acesso em toda a vida nas últimas décadas. Em 2011 havia 2,27 bilhões de usuários, consumindo cerca de 30 PB/mês [2]. Em 2015 a quantidade de usuários subiu para cerca de 3,2 bilhões, com um consumo de dados ainda maior. Os dados estão presentes no cotidiano das pessoas, sejam em forma de mensagem de texto, imagens com *links* na Web, e-mails, transações online, ou até mesmo rotas realizadas por veículos que dispõem de GPS, entre outros. Entretanto, muitas vezes, não é possível ou torna-se muito complexo gerar informações de valor a

partir desses dados. Em geral, os dados estão disponíveis de uma maneira em que só as pessoas conseguem compreender, ou seja, o processamento automático dos dados pelas máquinas pode ser bastante caro e complexo.

A Web 2.0 tinha a grande dificuldade de não conseguir lidar com a quantidade de dados gerados diariamente, nem possuir tecnologias necessárias para filtragem e análise desses dados. Tendo isso como contexto, surgiu a Web 3.0, conhecida como Web de Dados ou Web Semântica, que se propõe a decrementar a tarefa manual de análise e processamento de dados, através da criação de *links* entre páginas que conectam dados relacionados, de forma que esses dados possam ser lidos por máquinas e não apenas por pessoas [3].

A Web Semântica faz uso de algumas tecnologias, das quais podemos destacar o *Resource Description Framework* (RDF), modelo que possibilita a definição de sentenças sobre um recurso; a *Web Ontology Language* (OWL), linguagem recomendada pelo W3C para especificações de ontologias, que são capazes de descrever um domínio; *eXtensible Markup Language* (XML), linguagem que possibilita que o usuário crie *tags* personalizadas de maneira a facilitar a organização e estruturação de dados existentes em determinada aplicação.

O termo *Linked Data* surgiu a partir do desenvolvimento deste conceito de Web Semântica, referindo-se a um conjunto de melhores práticas para publicar e conectar dados estruturados na Web [5]. Existe também o conceito de dados abertos que são dados disponíveis para utilização de qualquer usuário, estando sujeito, no máximo, à exigência de creditar a sua autoria e compartilhar pela mesma licença [6]. A integração desses dois conceitos forma os dados abertos conectados [7], tornando mais fácil a manipulação e reutilização dos dados, agregando valor e possibilitando a descoberta de novos dados vinculados.

Atualmente várias instituições, sejam elas governamentais ou não, estão estimulando a publicação de dados abertos conectados através de portais que disponibilizam informações úteis para qualquer tipo de usuário. Nessa perspectiva, o

Governo Brasileiro foi um dos fundadores da Open Government Partnership, iniciativa para tornar o governo mais aberto aos seus cidadãos [7], onde dados sobre gastos e despesas estão disponíveis para qualquer cidadão, a qualquer hora na internet.

Além do governo e outras grandes corporações (como Google, Facebook, entre outras), o Centro de Informática (CIn) da UFPE tem incentivado a publicação de dados abertos conectados, onde podemos destacar projetos como o OpenCIn, um portal de dados abertos contendo informações sobre docentes, cursos de graduação, notícias relacionadas ao centro [8], e a publicação de um *dataset* com dados relevantes sobre as últimas dez edições do Simpósio Brasileiro de Banco de Dados (SBBD) [9]. Sendo motivado por esses projetos, este trabalho de graduação tem como objetivo o consumo de dados abertos conectados sobre eventos promovidos pela Sociedade Brasileira de Computação.

Com sede em Porto Alegre (RS), a Sociedade Brasileira de Computação é uma associação que fomenta e desenvolve pesquisas científicas na área de Computação [4]. A SBC está organizada em Comissões Especiais, responsáveis por cada subárea de conhecimento específico, que promovem diversas atividades e eventos, além de simpósios periódicos; além dessas comissões, a SBC é composta por pesquisadores, professores, alunos e profissionais.

2. Objetivos

O objetivo deste trabalho é o desenvolvimento de uma aplicação para o consumo de dados abertos conectados sobre os eventos da SBC. A aplicação proposta deverá consumir dados que serão gerados a partir de outro trabalho de graduação atualmente em desenvolvimento.

A aplicação criada deverá ser capaz de gerar indicadores que possam auxiliar gestores na tomada de decisão, contendo conhecimentos relevantes sobre os eventos passados promovidos pela SBC, onde não apenas os gestores, mas todo o usuário poderá ter acesso a esses dados.

Como objetivos específicos deste trabalho destacam-se:

- Desenvolvimento da aplicação para consumo dos dados abertos conectados sobre eventos da SBC.
- Realização de experimentos para avaliação da aplicação proposta.
- Geração de indicadores para facilitar a análise dos dados sobre eventos da SBC.
- Geração de regras a serem usadas para realização de inferências sobre uma determinada ontologia (a ser desenvolvida em conjunto com o trabalho de graduação já mencionado) a fim de auxiliar a tomada de decisão por parte dos gestores da SBC, em especial da Diretoria de Eventos.

3. Cronograma

A Tabela 1 descreve o cronograma previsto para o desenvolvimento deste Trabalho de Graduação.

Atividade	Março			Abril				Maio			Junho			Julho			
Formulação da proposta	X	X	X	X	X												
Análise da literatura	X	X	X	X	X	X	X	X									
Especificação da aplicação							X	X									
Desenvolvimento da aplicação									X	X	X	X	X				
Realização de testes e implementação de melhorias											X	X	X	X			
Elaboração do relatório												X	X	X	X	X	X
Preparação para apresentação e defesa															X	X	

Tabela 1: Cronograma de atividades

4. Possíveis Avaliadores

Os possíveis avaliadores desse trabalho de graduação serão:

- Ana Carolina Salgado
- Bernadette Farias Lóscio

5. Referências

- [1] WORLD WIDE WEB FOUNDATION. History Of The Web. Disponível em: <<http://webfoundation.org/about/vision/history-of-the-web>>. Acesso em: 6 mar. 2016.
- [2] THE EVOLUTION OF THE WEB. Disponível em: <<http://www.evolutionoftheweb.com/?>>. Acesso em: 6 mar. 2016
- [3] AGHAEI, S.; NEMATBAKHS, M; FARSANI, H. Evolution of the Word Wide Web: From Web 1.0 to Web 4.0. *International Journal of Web & Semantic Technology (IJWesT)*, Vol.3, No.1, jan. 2012.
- [4] SOCIEDADE BRASILEIRA DE COMPUTAÇÃO. Disponível em: <<http://www.sbc.org.br>>. Acesso em: 09 de mar. 2016.
- [5] CUNHA, D. R.; SOUZA, D.; LÓSCIO, B. F.. Linked Data: da Web de Documentos para a Web de Dados. Escola Regional de Computação - Ceará - Maranhão - Piauí 2011.
- [6] THE OPEN DEFINITION. Disponível em: <<http://opendefinition.org/>>. Acesso em: 25 mar. 2016.
- [7] ISOTANI, S.; BITTENCOURT, I. I. Dados Abertos Conectados. 1 ed. São Paulo: Novatec, 2015. p. 176.
- [8] ROCHA, J. L. OpenCin – *Dados Abertos Interligados Acerca dos Docentes do Centro de Informática*. 2015. 36 f. Trabalho de Conclusão de Curso (Graduação) – Universidade Federal de Pernambuco, Recife – PE.
- [9] BATISTA, M. G. R. *Criação e Publicação de um Linked Dataset sobre o Simpósio Brasileiro de Banco de Dados*. 2013. 57 f. Trabalho de Conclusão de Curso (Graduação) – Universidade Federal de Pernambuco, Recife – PE.

6. Assinaturas

Rayelle Ingrid Vera Cruz Silva
Orientando

Bernadette Farias Lóscio
Orientador