



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

Criação e Publicação de Dados Abertos Conectados
sobre os Eventos da Sociedade Brasileira de Computação

PROPOSTA DE TRABALHO DE GRADUAÇÃO

Aluno: Ruan Pablo Gouveia Alves (rpga@cin.ufpe.br)

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

Recife, 12 de Abril de 2016

1. Contextualização

Quando Tim Berners-Lee propôs a *World Wide Web* no final dos anos 80, iniciou uma transformação na maneira com a qual pessoas trocam informações e conduzem seus negócios [1], causando uma mudança significativa na forma de produzir e consumir dados. Com a evolução da Web, dados passaram a ser gerados em grandes quantidades a todo momento ao redor do mundo. Atividades comuns como utilizar o celular, realizar uma compra, pagar faturas, geram dados que podem ser analisados, processados e transformados em informação de valor [2].

Conforme [3], a evolução da Web se dá continuamente. Sua versão 1.0 era prioritariamente informativa (*read-only*), os visitantes não podiam interagir com os *websites*, considerados “estáticos”, tendo como principal objetivo a publicação de informações. Desde o seu estágio inicial, a Web tem como núcleo três tecnologias: o protocolo de rede HTTP (*Hypertext Transfer Protocol*), a linguagem de marcação para páginas Web HTML (*HyperText Markup Language*) e uma forma para identificação de recursos na Web por meio de URI (*Uniform Resource Identifier*).

A possibilidade de interação (*read-write*) foi incluída na Web 2.0, fazendo com que usuários deixassem de ser apenas expectadores e passassem a contribuir com conteúdo [3]. Tecnologias como blogs e *wikis* colaboraram para que a Web se tornasse um ambiente mais dinâmico, melhorando a forma de comunicação e interação entre os usuários [5].

Tal comportamento aumentou significativamente a quantidade de dados disponíveis na Web. Existem alguns mecanismos que possibilitam o acesso a esse grande volume de dados utilizando busca por palavra chave, a qual, mesmo amplamente difundida, possui problemas graves como a alta sensibilidade a sintaxe e a organização ineficiente dos dados [7].

Em seguida, a Web sofre mais uma evolução, chegando à sua versão 3.0, também chamada de Web de dados, ou Web semântica [3]. A Web semântica busca fornecer meios para organizar os dados e possibilitar que eles sejam interpretados por computadores, diminuindo o tempo gasto e realizando um uso mais inteligente tanto dos dados, quanto das máquinas sempre que possível [7].

Algumas tecnologias foram adotadas para apoiar a Web semântica, como o *Resource Description Framework* (RDF), um modelo de dados para descrever recursos na Web [2]; *Web Ontology Language* (OWL), que possibilita a definição formal de um domínio por meio da criação de ontologias [1] e *Extensible Markup Language* (XML), uma linguagem para marcação que permite aos usuários escreverem suas próprias tags [1].

Existe ainda, segundo [1], com apoio das técnicas mencionadas, um conjunto de princípios para publicar e interligar dados estruturados, chamado de *Linked Data* (dados conectados). Além disso, quando os dados estão disponíveis para serem utilizados, reutilizados e redistribuídos por qualquer usuário, sob a condição máxima de citar as fontes originais e respeitar as mesmas licenças,

podem ser denominados de dados abertos. A combinação desses conceitos forma os dados abertos conectados [2].

Os dados abertos conectados tornam mais fácil a descoberta de outros dados vinculados, aumentando seu valor agregado, assim como, permitem a manipulação e reutilização dos dados sem necessariamente depender de algum software em particular [2]. O projeto *Linking Open Data*, um dos mais importantes da área, tem o objeto de estender a Web através da publicação de um conjunto de dados abertos conectados em RDF, utilizando informações públicas e realizando a ligação entre esses dados, são 570 fontes de dados interligadas por 2909 conjuntos de links RDF [8].

Há também esforços no Centro de Informática (CIn) da UFPE para a publicação de dados abertos interligados, como por exemplo, o projeto OpenCIn que oferece informações públicas sobre os docentes do CIn [4] e a publicação de informações relevantes acerca das edições passadas do Simpósio Brasileiro de Banco de Dados (SBBDD) [5]. Nesse contexto este trabalho tem como objetivo a publicação de dados abertos conectados sobre os eventos promovidos pela Sociedade Brasileira de Computação.

A Sociedade Brasileira de Computação (SBC), com sede em Porto Alegre (RS), é uma associação científica, sem fins lucrativos, que incentiva e realiza pesquisas acadêmicas na área de Computação [6]. A SBC é composta por pesquisadores, professores, alunos e profissionais que atuam em diferentes áreas da Computação, como Banco de Dados, Engenharia de Software, Redes de Computação, entre outros. Sua estrutura é formada por delegacias regionais, ligadas à secretaria de diretorias da SBC, as quais possuem regimento específico, e delegacias institucionais, localizadas em diversas instituições de ensino. Tais instituições agem como um canal de comunicação com os associados (fundadores, estudantes, efetivos, institucionais e honorários), como também organizam eventos e diversas atividades. Além disso, possui comissões especiais, relacionadas com subáreas de conhecimento específico, responsáveis pela elaboração de simpósios entre outras atividades.

2. Objetivos

O foco deste trabalho é a criação e publicação de dados abertos conectados sobre dados relevantes dos eventos da SBC, incorporando também alguns elementos da estrutura dessa sociedade, prezando pelo reuso e flexibilidade, a fim de permitir que aplicações possam facilmente manipular tal conjunto de dados.

Como objetivos específicos deste trabalho destacam-se:

- Criação de uma ontologia em OWL para representar os principais conceitos do domínio a ser trabalhado, seguindo os passos da metodologia 101 [9]. Termos de vocabulários já conhecidos serão reusados, sempre que possível, na criação das classes e propriedades.

- Geração e publicação de um conjunto de dados abertos conectados sobre os eventos da SBC. Os dados serão extraídos dos sites dos eventos já realizados e de seus respectivos anais, como também de um banco de dados relacional sobre os eventos a ser disponibilizado pela SBC. Por fim, será disponibilizado um SPARQL *endpoint*, para que possam ser realizadas consultas SPARQL sobre o repositório.
- Realização de experimentos para avaliar a ontologia e o conjunto de dados gerados.

3. Cronograma

A Tabela 1 apresenta o cronograma previsto para este trabalho de graduação.

Atividade	Março				Abril				Maio			Junho				
Formulação da proposta e análise da literatura	X	X	X	X												
Construção da ontologia			X	X	X											
Geração do conjunto de dados						X	X	X	X							
Publicação do conjunto de dados em SPARQL Endpoint										X	X	X				
Elaboração do relatório final											X	X	X	X		
Preparação para apresentação e defesa															X	X

Tabela 1: Cronograma de atividades

4. Possíveis Avaliadores

Os possíveis avaliadores deste trabalho de graduação serão:

- Ana Carolina Salgado
- Bernadette Farias Lóscio

5. Referências

- [1] ANTONIOU, G.; HARMELEN, F. V. A Semantic Web Primer. 2 ed. Cambridge, MA, USA: The MIT Press, 2004. p 264.
- [2] ISOTANI, S.; BITTENCOURT, I. I. Dados Abertos Conectados. 1 ed. São Paulo: Novatec, 2015. p. 176.
- [3] AGHAEI, S.; NEMATBAKHSI, 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] ROCHA, J. L. OpenCIn - *Dados Abertos e Interligados Acerca dos Docentes do Centro de Informática*. 2015. 36 f. Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Pernambuco, Recife - PE.
- [5] 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] SOCIEDADE BRASILEIRA DE COMPUTAÇÃO. Disponível em: <<http://www.sbc.org.br>>. Acesso em: 04 de abr. 2016.
- [7] LÓSCIO, B. F.; Filho, F.W.B.H. Web Semântica: Conceitos e Tecnologias. In: Pedro de Alcantara. (Org.). II Escola Regional de Computação - Ceará, Maranhão e Piauí - ERCEMAPI 2009: SBC, 2009.
- [8] W3C. Linking Open Data. Disponível em: <<http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>>. Acesso em: 05 de abr. 2016.
- [9] NOY, N. F; MCGUINNESS, D. L. *Ontology Development 101: A Guide to Creating Your First Ontology*. Stanford University, Stanford.

6. Assinaturas

Ruan Pablo Gouveia Alves
Orientando

Bernadette Farias Lóscio
Orientadora