

# ObserveUnB - Um portal Web de rede social científica

Célia G. Ralha<sup>1,3</sup>, Gabriela N. Guerra<sup>1</sup>, Gedalias H. O. Valentim<sup>2</sup>, Murilo S. Sousa<sup>1</sup>, Marcelo F. M. Persegona<sup>2</sup>, Marcelly L. S. Godinho<sup>1</sup>, André L. Persegona<sup>2</sup>, Raquel N. Blumenschein<sup>2</sup>, Maria V. D. F. Tomé<sup>2</sup>, Denise B. B. Carvalho<sup>3</sup>

<sup>1</sup>Departamento de Ciência da Computação – Universidade de Brasília (UnB)  
Caixa Postal 4466 – 70910-900 – Brasília/DF – Brasil

<sup>2</sup>Laboratório do Ambiente Construído, Inclusão e Sustentabilidade (LACIS)  
Universidade de Brasília (UnB), Caixa Postal 04431 – 70904-970 – Brasília/DF – Brasil

<sup>3</sup>Decanato de Pesquisa e Pós-Graduação – Universidade de Brasília (UnB)  
CEP – 70910-970 – Brasília/DF – Brasil

ghedini@cic.unb.br, gabrielag.guerra@gmail.com, gedalias@gmail.com  
murlsousa@gmail.com, marcelo.persegona@gmail.com  
marcellygodinho@gmail.com, persegona@gmail.com, raquelblum@terra.com.br  
viferrari@gmail.com, denisebomtempo@terra.com.br

**Abstract.** *This paper presents the ObserveUnB, a social network on the Web, developed with the scientific information of the University of Brasília (UnB) academic community. The essence of ObserveUnB is based in the construction and analysis of the relationships between the researchers of the network, which was build with the aid of a reputation model and an ontology. This is the first Web portal for scientific social network developed in the scope of a Federal Institution of Higher Education in Brazil. The ObserveUnB's big challenge is the availability of scientific knowledge, allowing universal and participatory access to Brazilian citizens, as well as to manage the integration of a large volume of data from different repositories.*

**Resumo.** *Este artigo apresenta o ObserveUnB, um portal Web de rede social científica desenvolvido com base nas informações da comunidade acadêmica da Universidade de Brasília (UnB). A essência do ObserveUnB encontra-se na construção e análise dos relacionamentos existentes entre os pesquisadores da rede, com auxílio de um modelo de reputação e uma ontologia. Trata-se do primeiro portal Web de rede social científica desenvolvido no escopo de uma Instituição Federal de Ensino Superior no Brasil. O grande desafio do ObserveUnB está na disponibilização do conhecimento científico, permitindo o acesso participativo e universal do cidadão brasileiro, bem como gerenciar a integração de um grande volume de dados de diferentes repositórios.*

## 1. Introdução

A comunidade acadêmica das Instituições Federais de Ensino Superior (IFES) no Brasil é composta por um grande volume de pesquisadores, cujos dados encontram-se distribuídos de forma heterogênea em diferentes bases na *Web* e em aplicações institucionais. Dessa forma, surge a necessidade de se criar veículos ou instrumentos onde a comunidade acadêmica científica possa visualizar e trocar informações pertinentes à produção científica e aos projetos de pesquisa. Como estudo de caso, podemos citar a comunidade acadêmica da Universidade de Brasília (UnB), que é composta por mais de 38 mil indivíduos oficialmente registrados nas bases de dados dos

sistemas acadêmicos institucionais, além das bases de Currículo *Lattes*<sup>1</sup> e de Diretório de Grupos de Pesquisa do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), bem como nas bases de dados dos programas de pós-graduação da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Além da distribuição dos dados em diferentes repositórios, os esforços colaborativos envolvendo pesquisadores de diferentes grupos de pesquisa, áreas de conhecimento, instituições ou até mesmo de diferentes países constituem a motivação para o desenvolvimento do portal *Web* denominado *ObserveUnB*<sup>2</sup>. O objetivo do projeto consiste em fomentar a colaboração científica, maximizando o potencial da produção, através da integração e disponibilização de informações e do conhecimento que permitem a ampliação das parcerias entre os grupos intra e inter-institucionais de ensino e pesquisa.

Com a finalidade de adicionar aspectos semânticos ao repositório do *ObserveUnB* está sendo utilizada uma ontologia estendida do Grupo *Conscientias*, base do Diretório de Grupos de Pesquisa e dos Currículos *Lattes* [Conscientias 2002, 2006]. O uso dessa ontologia, além de permitir a organização do domínio de conhecimento de forma adequada através de conceitos que integram os dados das bases heterogêneas, possibilita a recuperação da informação por meio da realização de inferências realizadas a partir dos relacionamentos definidos entre os conceitos, permitindo a execução de consultas não suportadas em um banco de dados relacional. O uso de ontologias no *ObserveUnB* também se mostra adequado nos aspectos de desambiguação, onde uma pessoa possui diferentes nomes de citação, bem como no problema de deduplicação, na qual duas pessoas têm o mesmo nome, aspectos importantes referenciados em outros portais na *Web* como o CiênciaBrasil [Laender *et al.* 2011].

Além da ontologia, o repositório faz uso de um modelo de reputação aplicado aos membros da rede social científica. De acordo com Resnick *et al.* [2000], um sistema com modelo de reputação tem o objetivo de coletar, distribuir e agregar informações sobre o comportamento de seus membros, de maneira a ajudá-los a decidir suas relações de confiança, encorajando comportamentos colaborativos. Nesse contexto, o objetivo do modelo de reputação aplicado ao *ObserveUnB* pretende auxiliar na busca por potenciais parcerias de pesquisa e na recomendação de pesquisadores seniores a seus pares, bem como às entidades de fomento à pesquisa.

Para viabilizar o intercâmbio de informações e disponibilizar o conhecimento produzido na UnB, o projeto do *ObserveUnB* enfrenta dois dos Grandes Desafios da Pesquisa em Computação no Brasil elencados pela Sociedade Brasileira de Computação<sup>3</sup> (SBC). O primeiro desafio “Gestão da informação em grandes volumes de dados multimídia distribuídos” relaciona-se ao projeto visto que um grande volume de dados é extraído de bases de dados heterogêneas e integrado em um repositório de informação científica na *Web*. Outro grande desafio “Acesso participativo e universal do cidadão brasileiro ao conhecimento” é abordado no projeto desde que o *ObserveUnB* oportuniza a geração do conhecimento, a extensibilidade, a interatividade e o acesso universal, por meio da disponibilização de informações e análise dos relacionamentos e do uso da ontologia.

---

<sup>1</sup> <http://lattes.cnpq.br/>

<sup>2</sup> <http://www.observeunb.unb.br/>

<sup>3</sup> <http://www.sbc.org.br/>

O projeto do *ObserveUnB* foi desenvolvido por equipe formada por pesquisadores de diferentes áreas da UnB, tendo sido iniciado em março de 2011 e disponibilizado à comunidade acadêmica em 28/11/11. Como questões básicas de pesquisa o projeto *ObserveUnB* pretende responder quatro perguntas: (i) quem são os pesquisadores; (ii) o que os pesquisadores estão produzindo; (iii) onde os pesquisadores estão localizados; e (iv) com quem os pesquisadores estão pesquisando. O restante deste artigo está organizado da seguinte forma: na Seção 2 são descritas iniciativas de rede de colaboração científica com técnicas computacionais; na Seção 3 são apresentados os principais módulos da arquitetura do *ObserveUnB*; na Seção 4 são apresentados os resultados obtidos; na Seção 5 são discutidos as conclusões e os trabalhos futuros.

## 2. Trabalhos Relacionados

Recentemente, vários projetos de pesquisa têm viabilizado a criação de redes sociais científicas na *Web*, possibilitando uma melhor integração dos pesquisadores através de seus relacionamentos científicos. Podemos citar como iniciativas de pesquisa a ferramenta *ArnetMiner*<sup>4</sup> de Tang *et al.* (2008), a qual fornece serviços para busca e extração de informações baseada na comunidade de pesquisadores na *Web*. Para isso a ferramenta estende a ontologia FOAF<sup>5</sup> para unificar os perfis dos pesquisadores e permite a pesquisa de especialistas por tópico ou palavras-chave. O *ArnetMiner* utiliza informações contidas nas páginas Web dos pesquisadores e de diferentes bibliotecas digitais internacionais, como DBLP<sup>6</sup>, *CiteSeer*<sup>7</sup> e ACM<sup>8</sup>.

Outro projeto interessante é o *Microsoft Academic Search* (MAS)<sup>9</sup>, uma ferramenta de livre acesso destinada à pesquisa aplicada a diferentes áreas de conhecimento. O MAS disponibiliza consultas sobre publicações, autores, conferências, periódicos, instituições e palavras-chave, mediante um mecanismo de busca que coleta informações na *Web* de bases internacionais (e.g. DBLP, *CiteSeer*, ACM).

O Portal Brasileiro de Ciência e Tecnologia (CiênciaBrasil)<sup>10</sup> de Laender *et al.* (2011) disponibiliza um mapa da produção científica e constrói a rede de co-autoria dos atuais pesquisadores vinculados aos Institutos Nacionais de Ciência e Tecnologia (INCT)<sup>11</sup>. Para tal, utiliza os dados individuais dos pesquisadores coletados da plataforma *Lattes* do CNPq por meio de técnicas de extração, integração, desambiguação, deduplicação e análise. Esta iniciativa está no âmbito do *InWeb*<sup>12</sup>, um dos 126 INCTs do país, que tem como objetivo desenvolver modelos, algoritmos e novas tecnologias que permitam aumentar a integração da *Web* com a sociedade.

Dentre os trabalhos que utilizam cálculo de reputação, o algoritmo *PageRank* [Page *et al.* 1999] é o mais popular [Hogan *et al.* 2006, Matsuo *et al.* 2004, Haveliwala 2003]. Em Ding *et al.* (2009) o algoritmo é utilizado para classificar o autor em uma rede de co-citação. O trabalho de Singh *et al.* (2011) que tem como foco classificar uma

---

<sup>4</sup> <http://arnetminer.org/>

<sup>5</sup> <http://www.foaf-project.org/>

<sup>6</sup> <http://dblp.uni-trier.de/>

<sup>7</sup> <http://citeseer.ist.psu.edu/>

<sup>8</sup> <http://www.acm.org/>

<sup>9</sup> <http://academic.research.microsoft.com/>

<sup>10</sup> <http://pbct.inweb.org.br/>

<sup>11</sup> [http://www.cnpq.br/programas/inct/\\_apresentacao](http://www.cnpq.br/programas/inct/_apresentacao)

<sup>12</sup> <http://www.inweb.org.br/>

rede de citação, utiliza uma versão modificada do *PageRank* que considera a qualidade do artigo citado. Gollapalli *et al.* (2011) propõe um modelo do *PageRank* baseado em grafo que classifica especialistas em determinados tópicos. Diferentemente dos métodos anteriores, os quais consideram a rede de citação para o cálculo da reputação, o modelo proposto por Cervi *et al.* (2011) considera toda a trajetória acadêmica do pesquisador para aferir a sua reputação, utilizando um modelo de perfil adaptativo.

Conforme exposto, as iniciativas de construção de rede social científica com modelo de reputação e aspectos semânticos para suporte à comunidade acadêmica brasileira, ainda não se apresentam de forma totalmente explorada. Nesta direção, o projeto do *ObserveUnB* constrói uma rede social de pesquisadores vinculados a UnB, à partir da integração de dados provenientes de diferentes repositórios nacionais utilizando uma ontologia estendida do Grupo *Conscientias*, e apresenta um modelo de reputação do pesquisador adaptado de Cervi *et al.* (2011), características ainda inexistentes em um repositório de informação científica no âmbito das IFES.

### 3. Arquitetura do *ObserveUnB*

A Figura 1 apresenta a arquitetura do *ObserveUnB* com seus módulos básicos: extração, integração, repositório e interface de visualização, descritos nas Seções 3.1, 3.2, 3.3 e 3.4. A versão atual utiliza os dados dos sistemas institucionais da UnB, do currículo *Lattes* dos docentes e as informações dos programas de pós-graduação disponíveis na CAPES, bem como o modelo de reputação dos membros da rede social científica.

#### 3.1 Extração

Esse módulo abrange o processo de extração de dados dos docentes e discentes utilizados e armazenados no repositório, os quais são obtidos das bases de dados de sistemas institucionais da UnB<sup>13</sup>: Sistema de Pessoal - SIPES, Sistema de Informação Acadêmica da Graduação – SIGRA, Sistema de Informação Acadêmica da Pós-Graduação - SIPPOS. Os dados dos 36.199 discentes e 2.181 docentes são fornecidos pelo Centro de Informática da UnB em arquivos texto com o formato CSV (*Comma Separated Values File*).

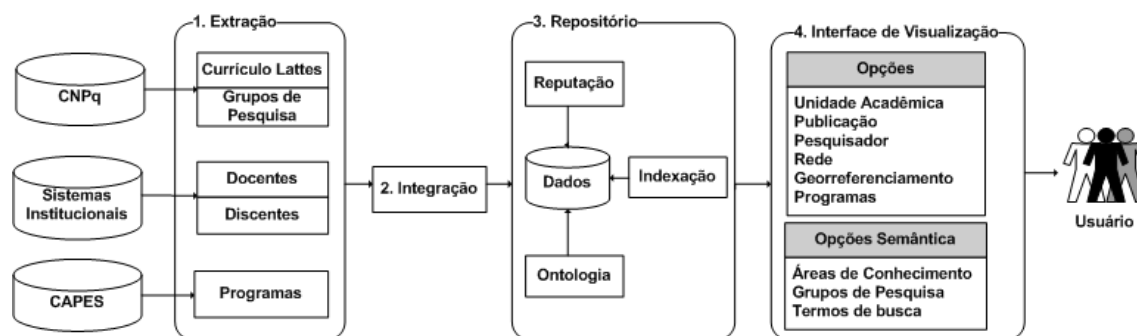


Figura 1. Arquitetura do *ObserveUnB*.

Os dados dos programas de pós-graduação da UnB, envolvendo 1.368 pesquisadores vinculados aos programas de pós, coletados através do aplicativo Coleta de Dados da CAPES, são baixados utilizando o código da instituição e gerando arquivos do tipo texto convertidos para o formato CSV.

<sup>13</sup>[http://www.cpd.unb.br/index.php?option=com\\_content&view=article&id=67&Itemid=71](http://www.cpd.unb.br/index.php?option=com_content&view=article&id=67&Itemid=71)

Para executar a extração dos CV-Lattes do CNPq foi implementado um robô na linguagem Java. O Robô-CV utiliza os dados dos docentes e discentes da UnB obtidos dos sistemas institucionais (SIPES, SIGRA e SIPPOS), acessa a plataforma Lattes do CNPq através de um endereço IP pré-definido e procede à coleta dos arquivos de CV-Lattes na linguagem XML (*Extensible Markup Language*). O processo de extração enfrentou vários desafios do grande volume de dados, seja pelo limite de quota para baixa diária definido pelo CNPq às instituições parceiras, ou pelo próprio processo de tratamento dos dados no repositório do *ObserveUnB*. Desta forma, a extração dos CV-Lattes está programada para ocorrer uma vez por mês, sendo que a última coleta realizada em abril totalizou 14.369 currículos de uma população total de 38.380 indivíduos. Destes, 2.100 referem-se aos docentes, correspondendo a um percentual de 96% do total de docentes; enquanto 12.269 discentes com CV-Lattes corresponde a um percentual de 33% do total de discentes.

### **3.2 Integração**

Os dados dos arquivos XML e dos arquivos texto com o formato CSV gerados durante o módulo de extração são filtrados através do módulo de integração, o qual foi desenvolvido na linguagem PHP, convertendo os dados armazenados em uma base de dados relacional MySQL, cuja estrutura foi modelada de acordo com as ontologias do Grupo *Conscientias*. Durante a integração foram detectados alguns problemas como a inconsistência, a ambiguidade e a duplicação de dados: (i) pesquisadores vinculados a mais de uma unidade acadêmica na instituição; (ii) pesquisadores sem CV-Lattes; (iii) pesquisadores diferentes com nomes iguais; (iv) pesquisadores com mais de um nome para citação; (v) publicação vinculada a mais de um autor e (vi) duplicação de pesquisadores na base. No caso de dados incorretos, a correção foi realizada manualmente e na inexistência de CV-Lattes, apenas os dados institucionais do pesquisador foram integrados. Enquanto que no caso de duplicação e ambigüidade de dados foram utilizados os recursos da estrutura de classes e relacionamentos entre conceitos fornecidos pela ontologia estendida do Grupo *Conscientias*.

### **3.3 Repositório**

Na versão atual do *ObserveUnB*, o repositório consiste em uma base de dados relacional composta por dados pessoais (nome, nacionalidade e vinculação acadêmica) e CV-Lattes de pesquisadores, obtidos no módulo de extração, além dos dados de reputação, cujo tamanho total da base é de 970 Mb. Além desses dados, o repositório agrega uma ontologia estendida do Diretório de Grupo de Pesquisa e do Currículo *Lattes* do Grupo *Conscientias*, além de um modelo de reputação, os quais serão detalhados nas Seções 3.3.1 e 3.3.2.

#### **3.3.1 Ontologia**

O propósito da utilização da ontologia neste trabalho consiste em auxiliar na representação do domínio de conhecimento da área de pesquisa de uma IFES, bem como na realização de consultas semânticas. Para isso, uma ferramenta ontológica faz-se necessário, o qual faça uso de um mecanismo de inferência, visando explorar a estrutura de conceitos, relacionamentos e restrições do tipo funcional, funcional inverso, transitivo, simétrico, assimétrico, reflexivo e irreflexivo, com o objetivo de permitir a realização de consultas ainda não suportadas pelo paradigma de um banco de dados

relacional [Horridge 2011]. Outra funcionalidade da ontologia no *ObserveUnB* encontra-se na deduplicação e na desambiguação dos dados, as quais inicialmente foram feitas de forma manual durante o processo de instanciação da ontologia. Para auxiliar no problema da desambiguação estão sendo pesquisados recursos diversos, tais como o uso de funções de similaridade para o cálculo da proximidade de duas entidades, bem como o uso de esquemas de integração ontológica entre várias ontologias [Morris *et al.* 2008, Hakimpour and Geppert 2001].

Para o tratamento ontológico encontram-se disponíveis diversas ferramentas, como *Altova Semantic Works*, *TopBraid Composer*, *Onto Edit Professional*, *OilEd* e *Protégé* [Bergman 2011]. Considerando as metodologias ontológicas, Corcho *et al.* (2003) apresentam alguns métodos: *Methontology*, *On-To-Knowledge*, *TOVE* e *Método101*. Para a extensão da ontologia do Grupo *Conscientias* optou-se por utilizar a ferramenta *Protégé* [Protégé 2011], versão 4.1 beta, por ser uma ferramenta *open source* de ampla utilização e oferecer diversos *plug-ins*, ampliando as funcionalidades no processo de extensão. Além do *Protégé*, optou-se por seguir o Método 101 [Noy and McGuinness 2001] por ser uma metodologia simples e intuitiva proporcionando maior facilidade no processo de extensão das ontologias envolvidas no projeto, onde foram utilizadas todas as fases do método que consistem em: (i) determinar o domínio e escopo da ontologia; (ii) considerar o reuso de ontologias existentes; (iii) enumerar os termos importantes da ontologia; (iv) definir classes e hierarquia de classes; (v) definir as propriedades das classes; (vi) definir os valores das propriedades e (vii) criar instâncias.

As ontologias de Diretório de Grupo de Pesquisa e de Currículo *Lattes* do Grupo *Conscientias* encontram-se disponíveis em esquemas de dados definidos na linguagem DTD (*Document Type Definiton*) e XML. Com o uso da ferramenta *TopBraid Composer*<sup>14</sup>, esses esquemas de dados foram importados e migrados para a linguagem OWL (*Web Ontology Language*) [McGuinness and van Harmelen 2004], a qual permite descrever mais propriedades, restrições, relacionamentos e classes que outras linguagens ontológicas, além de possibilitar o raciocínio automático através do mecanismo de inferência.

O processo de conversão das ontologias de XML e DTD para OWL gerou classes com erros devido à definição de tipos de dados não suportados na linguagem OWL. Após a correção dos erros, por meio da ferramenta *Protégé* [Protégé 2011] foi realizado um *merge* das duas ontologias do Grupo *Conscientias*, ocasionando novos problemas como a duplicação de tipos de dados e de classes, e.g., as classes “Autor” e “Pesquisador” que representam o mesmo conceito. Dessa forma, manteve-se na ontologia final apenas uma classe representando um determinado conceito. Para viabilizar a integração da ontologia ao repositório de informações científicas, foi implementado um módulo na linguagem Java utilizando a API *Jena*<sup>15</sup> - um *Framework open source* em Java para *Web Semântica*, utilizado para criar protótipos, manipular classes, instanciar e recuperar os conceitos da ontologia.

Tendo em vista a complexidade da ontologia estendida que engloba mais de 200 classes, bem como o escopo do projeto que representa a produção científica dos

---

<sup>14</sup>[http://www.topquadrant.com/products/TB\\_Composer.html](http://www.topquadrant.com/products/TB_Composer.html)

<sup>15</sup><http://jena.sourceforge.net/license.html>

pesquisadores da rede da UnB, foram seleccionados os seguintes conceitos para integrar a ontologia: (i) autores; (ii) grupo de pesquisa; (iii) grande área de conhecimento, área de conhecimento, sub-área e especialidade; (iv) produção bibliográfica (artigos publicados, livros e capítulos de livro, textos publicados em jornais ou revistas, trabalhos em eventos); (v) produção técnica (registro ou patente, software, trabalhos técnicos, processos ou técnicas e demais produções técnicas); (vi) outra produção (produção artística e cultural, orientações concluídas); (vii) palavras-chave e (viii) instituição (conselhos universitários, unidades acadêmicas). A Figura 2 apresenta um substrato da ontologia do *ObserveUnB*, onde a primeira e a última classe, *GrupoDePesquisa* e *CurriculoVitae*, são provenientes de duas ontologias diferentes do Grupo *Conscientias*, sendo a primeira do Diretório de Grupos de Pesquisa e a última do Currículo *Lattes*.



**Figura 2. Exemplos de conceitos da ontologia definida no *ObserveUnB*.**

### 3.3.2 Modelo de Reputação

O repositório do *ObserveUnB* armazena também as informações de reputação dos integrantes da rede social científica. O objetivo do modelo de reputação proposto neste trabalho é qualificar pesquisadores em suas áreas de pesquisa, servindo como um instrumento de colaboração e fomento à produção científica. Dentre os diversos métodos e algoritmos de reputação estudados (citados na Seção 2), o modelo proposto por Cervi *et al.* (2011), para a área específica de Ciência da Computação, foi selecionado para servir de base no projeto do *ObserveUnB*, uma vez que ele considera toda a trajetória acadêmica do pesquisador fazendo uso de um modelo de perfil adaptativo.

Dessa forma, nosso modelo de reputação utiliza os dados ou elementos extraídos do CV-Lattes do pesquisador (e.g., doutorado na titulação), distribuindo em categorias ponderadas seus elementos. No modelo proposto, os pesos foram definidos de tal forma que representem a grande maioria das áreas de conhecimento, e.g., foi atribuído maior peso à categoria produção científica por possuir grande relevância em todas as áreas de ciência. A Tabela 1 exibe estas categorias com os respectivos elementos e pesos. Note que o maior peso foi atribuído à categoria produção (40), sendo que o somatório dos pesos das categorias é igual a 100.

Outra diferença entre o modelo de Cervi *et al.* (2011) e esta proposta está na fórmula para calcular a reputação apresentada na Equação 1. Na métrica do modelo de Cervi *et al.* (2011) é considerado o maior valor da reputação entre os membros da rede, tornando-se necessário o cálculo deste valor durante o cômputo da reputação de cada membro. Na métrica desta proposta, consideramos o somatório da multiplicação do número de ocorrências de determinado elemento pelo peso do próprio elemento, dividido pelo somatório de elementos de todas as categorias, simplificando a métrica do modelo original. Como na primeira categoria apenas um dos elementos é considerado para cálculo, também será considerado apenas um elemento no quantitativo.

$$\text{Reputação}_{\text{Pesquisador}} = \sum (\text{Quantidade}_{\text{Elemento}} * \text{Peso}_{\text{Elemento}}) / \text{Total}_{\text{Elemento}} \quad (1)$$

Tabela 1. Categorias, elementos e pesos [adaptado de Cerviet *al.* 2011].

<b>Categoria</b>	<b>Peso Máximo</b>	<b>Elemento</b>	<b>Peso</b>
<b>1. Formação</b>	15	Pós-doutorado	15
		Doutorado	12
		Mestrado	8
		Graduação	4
<b>2. Orientações Concluídas</b>	14	Pós-doutorado	5
		Doutorado	4
		Mestrado	3
		Iniciação científica	2
<b>3. Banca Examinadora</b>	7	Doutorado	4
		Mestrado	3
<b>4. Membro</b>	7	Editorial	4
		Revisor de periódico	2
		Projeto de pesquisa	1
<b>5. Bolsista</b>	12	Bolsista do CNPq	12
<b>6. Prêmios e Títulos</b>	3	Prêmios e títulos	3
<b>7. Produção</b>	40	Artigos publicados	10
		Livros publicados ou organizados	8
		Capítulos de livros	6
		Trabalhos em eventos	3
		Texto em jornal ou revista	2
		Demais tipos de produção bibliográfica	1
		Técnica	5
		Artística	5
<b>8. Co-autoria</b>	2	Pesquisador	2

Para facilitar a interpretação do resultado do modelo de reputação, assim como no modelo de Cervi *et al.* (2011), o índice de reputação foi classificado em uma tabela de níveis, que varia de 1 a 5 conforme apresentado na Tabela 2.

Tabela 2. Classificação da reputação.

<b>Nível</b>	<b>Reputação</b>	<b>Classificação</b>
<b>1</b>	$\geq 0$ e $< 20$	Muito baixa
<b>2</b>	$\geq 20$ e $< 40$	Baixa
<b>3</b>	$\geq 40$ e $< 60$	Média
<b>4</b>	$\geq 60$ e $< 80$	Alta
<b>5</b>	$\geq 80$	Muito alta

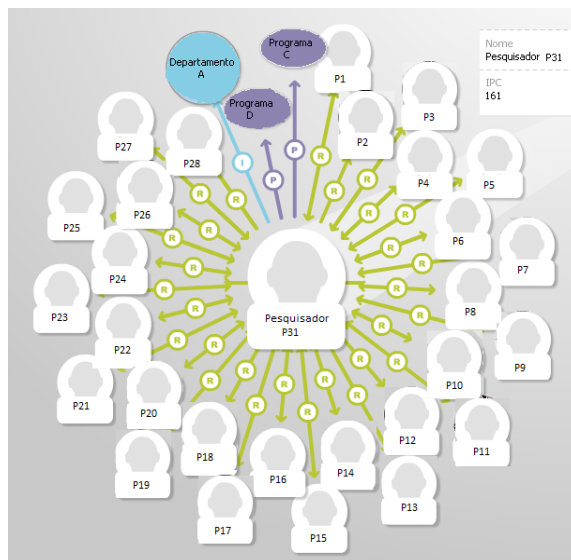
### 3.4 Interface de Visualização

Para responder as quatro perguntas básicas de pesquisa apresentadas na Seção 1, o *ObserveUnB* disponibiliza ao usuário como opções de pesquisa: (i) dados da UnB (o total da produção científica da UnB e o total por unidade acadêmica); (ii) georreferenciamento (a localização dos pesquisadores e das unidades acadêmicas dentro dos quatro *Campi* da UnB); (iii) rede (a rede de vinculação das unidades acadêmicas); (iv) programas (os pesquisadores vinculados aos programas de pós-graduação e o total de produção científica correspondente ao programa); (v) publicação (dado o nome de uma publicação, o sistema retorna os autores) e (vi) pesquisador (dado o nome de um pesquisador, o sistema retorna informações de seu currículo, sua produção científica e os relacionamentos com outros pesquisadores ou rede de pesquisa).

Como nas demais redes sociais, o sistema permite a visualização de conexões entre pesquisadores, possibilitando ao usuário identificar as parcerias existentes, conforme apresentado na Figura 3. Foram definidos três tipos de relacionamentos na rede, a saber: (i) “I” – representa a integração do pesquisador a uma Unidade Acadêmica da UnB; (ii) “P” – representa a participação do pesquisador em determinado



programa de pós-graduação da UnB; e (iii) “R” - representa a colaboração de pesquisa entre pesquisadores via produção científica. Para a visualização das redes de relacionamento foi utilizada a biblioteca *Relation Browser*<sup>16</sup>, destinada a apresentar estruturas de rede social a partir de redes de relacionamento complexas.



**Figura 3. Rede de relacionamentos de um pesquisador da UnB.**

Além das opções de pesquisa por pesquisador ou publicação, a partir da ontologia é possível a realização de consultas mais elaboradas por meio de relacionamentos definidos entre os conceitos da ontologia e do mecanismo de inferência disponibilizado pelo Protégé (*Pellet reasoner*). Para exemplificar uma busca por grande área e área de conhecimento, autores vinculados a uma unidade acadêmica e produção bibliográfica serão recuperados por relacionamentos de transitividade as produções e os pesquisadores que melhor atenderem ao critério de busca, conforme apresentado na Figura 4.



**Figura 4. Opção de pesquisa por área de conhecimento no ObserveUnB.**

<sup>16</sup><http://moritz.stefaner.eu/projects/relation-browser/>

#### 4. Análise dos Resultados

A partir do portal do *ObserveUnB* pode-se obter dados estatísticos relacionados à produção científica dos 2.181 pesquisadores da UnB, com 115.291 unidades de produção, correspondendo ao total da produção acadêmica dos pesquisadores vinculados a UnB, incluindo artigos em periódicos, trabalhos em eventos, artigos aceitos para publicação, livros publicados ou organizados, capítulos de livros publicados, textos em jornais ou revistas e demais tipos de produção bibliográfica.

Com essas informações, pode-se visualizar o gráfico quantitativo de produção científica de cada unidade acadêmica e de cada programa de pós-graduação. Baseado no CV-Lattes é apresentado a produção científica de cada pesquisador, incluindo a área de conhecimento, palavras-chave e ano da publicação. Com a rede de pesquisa obtêm-se todos os pesquisadores com os quais um determinado docente publicou. Enquanto que, na rede de relacionamentos em forma de grafo é possível visualizar os docentes da instituição com os quais o pesquisador produziu, conforme apresentado na Figura 3.

A Figura 5 (A) exibe um gráfico dos resultados do cálculo de reputação conforme as categorias apresentadas na Tabela 1, em uma população composta por 2.101 pesquisadores da UnB. Enquanto a Figura 5 (B) apresenta os resultados de uma avaliação parcial do modelo da Tabela 1, o qual está disponível no *ObserveUnB*, incluindo apenas os dados das categorias 1, 7 e 8 (titulação, produção e co-autoria). Os dados dos pesquisadores foram coletados dos CV-Lattes, a reputação calculada conforme Equação 1 e classificada segundo a Tabela 2. Visando amenizar o impacto da divulgação do cálculo de reputação do pesquisador integrante da rede social científica da UnB, o valor da reputação do pesquisador está sendo apresentado no portal do *ObserveUnB* como Índice de Produção Científica – IPC (Equação 1). A Figura 3 apresenta a rede de um pesquisador “P31” e o valor 161 que indica o valor de sua reputação ou IPC.

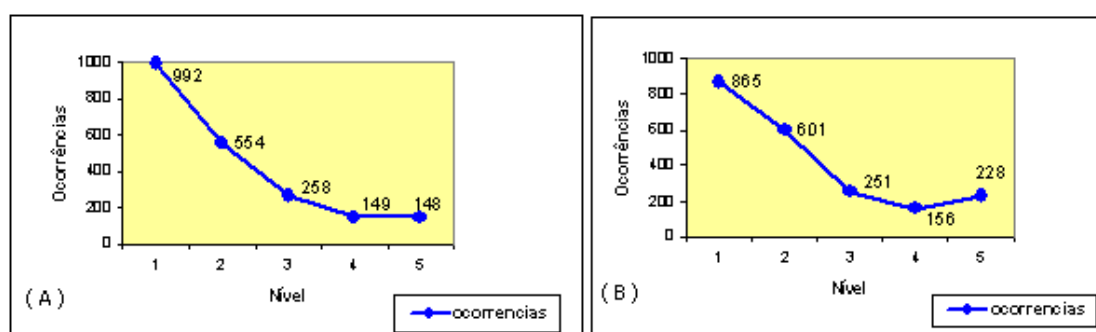
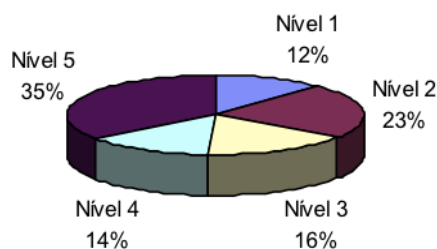


Figura 5. Classificação da reputação: A) Modelo Completo e B) Modelo Parcial.

Note que os dois gráficos apresentam curvas semelhantes, sendo que o Nível 1 tem a maior concentração de pesquisadores. Isso se deve em parte à incompletude e inconsistência dos dados nos CV-Lattes, bem como o fato de que muitos desses pesquisadores apresentarem pouco tempo de carreira acadêmica. A Figura 6 apresenta o percentual do quantitativo de produção bibliográfica referente a cada nível da Tabela 2, resultante da implementação do modelo no *ObserveUnB*. Percebe-se que apesar do Nível 1 concentrar a maior quantidade de docentes (865), ele representa o menor percentual de produção bibliográfica da UnB (12%), enquanto o Nível 5 com 228 docentes representar 35% da produção científica da instituição.



**Figura 6. Produção Bibliográfica referente aos níveis de reputação no *ObserveUnB*.**

## 5. Conclusões e Trabalhos Futuros

Este trabalho apresentou o portal *Web* de rede social científica da UnB, o qual possui como desafio a integração de um grande volume de dados provenientes de diferentes repositórios, bem como o gerenciamento e a disponibilização do conhecimento científico de forma a promover o acesso participativo e universal do cidadão brasileiro no fomento a produção científica. Nesse contexto, o *ObserveUnB* consiste no primeiro trabalho de uma rede social científica no âmbito das IFES, que permite a visualização e análise dos relacionamentos formados entre os pesquisadores membros da rede por meio de um repositório de informação científica, fazendo uso de um modelo de reputação e de uma ontologia estendida do Grupo *Conscientias*.

O projeto do *ObserveUnB* pode ser integrado a outras IFES ou a outros portais de Ciência e Tecnologia, como o CiênciaBrasil, e com isso permitir análises comparativas de produção científica dos pesquisadores vinculados às instituições de ensino e pesquisa do país. Conforme citado, opções semânticas de consulta da ferramenta estão em desenvolvimento fazendo uso dos conceitos e do mecanismo de inferência da ontologia definida. Falta ainda validar o modelo de reputação utilizado com uso de modelos estatísticos para avaliação paramétrica dos resultados apresentados, bem como uma avaliação da usabilidade do portal do ponto de vista do usuário.

Como trabalhos futuros, novas funcionalidades podem ser incorporadas ao *ObserveUnB* para permitir a visualização da evolução da reputação e da produção científica dos pesquisadores ao longo do tempo, como por exemplo nos triênios de avaliação dos programas de pós-graduação pela CAPES. Para viabilizar uma maior interação do cidadão brasileiro com os pesquisadores no portal, uma rede social *online* pode ser viabilizada por meio da troca de mensagens eletrônicas como *e-mails* e *twitters*. A indexação da base de produção científica do *ObserveUnB* utilizando os dados dos repositórios internacionais (e.g., Scopus, DBLP e ACM) viabiliza uma avaliação qualitativa da produção através de índices de impacto, o que associado ao cálculo de reputação retrataria a realidade acadêmica dos pesquisadores do país, conforme parâmetros estabelecidos internacionais, e requeridos pelos órgãos de fomento, tais como a CAPES, CNPq e Finep. Conforme exposto existem diversas sugestões de trabalhos futuros, uma vez que o *ObserveUnB* trata do desafio de gerenciamento e da disponibilização do conhecimento científico para promover o acesso universal do cidadão.

## Referências

- Bergman, M. K. (2010). "Listing of 185 Ontology Building Tools", Disponível em: <http://www.mkbergman.com/904/listing-of-185-ontology-building-tools/>. Acesso Jan.2012
- Cervi, C. R.; Galante, R. & Oliveira, J. P. M. (2011). "Identificando a Reputação de Pesquisadores Usando um Modelo de Perfil Adaptativo", In 38 SEMISH, RN, Brasil.
- Conscientias (2002). "Padronização XML: Grupos de Pesquisa", Disponível em: <http://lmp.cnpq.br/lmp/?go=grupo.jsp>. Acesso Abr. 2012.
- Conscientias (2006). "Padronização XML: Curriculum Lattes", Disponível em: <http://lmp.cnpq.br/lmp/?go=cv.jsp>. Acesso Abr. 2012.
- Corcho, O.; Fernández-López, M. & Gómez-Pérez, A. (2003). "Methodologies, Tools and Languages for Building Ontologies: Where is Their Meeting Point?", *Data & Knowledge Engineering*, 46:1, 41-64.
- Ding, Y.*et al.* (2009). "PageRank for Ranking Authors in Co-citation Networks", *Journal of the American Society for Inf. Science and Technology*, 60:11, 2229-2243.
- Gollapalli, S. D.; Mitra, P. & Giles, C. L. (2011). "Ranking authors in digital libraries", In Proc. 11<sup>th</sup> Int. ACM/IEEE Joint Conf. on Digital Libraries, 251-254, NY, USA.
- Hakimpour, F. & Geppert, A. (2001). "Resolving semantic heterogeneity in schema integration", In Proc. Int. Conf. on Formal Ontology in Inf. Systems, 297-308, USA.
- Haveliwala, T. H. (2003). "Topic-Sensitive PageRank: A Context-Sensitive Ranking Algorithm for Web Search", Tech. Report, IEEE Transactions on Knowledge and Data Engineering, Stanford InfoLab.
- Hogan, A.; Harth, A. & Decker, S. (2006). "Reconrank: A scalable ranking method for semantic web data with context". In Proc 2<sup>nd</sup> Workshop on Scalable Semantic Web Knowledge Base Systems.
- Horridge, M.*et al.* (2011). "A Practical Guide to Building OWL Ontologies Using The Protégé 4 and CO-ODE Tools", Edition 1.3, The University of Manchester, UK.
- Laender, A. *et al.* (2011). "CiênciaBrasil-The Brazilian Portal of Science and Technology", In 38 SEMISH, RN, Brasil.
- Matsuo, Y. *et al.* (2004). "Finding Social Network for Trust Calculation", In Proc. 16<sup>th</sup> European Conference on Artificial Intelligence, 510 - 514.
- McGuinness, D. L. & van Harmelen, F. (2004). "Owl web ontology language overview" Disponível em: <http://www.w3.org/TR/2004/REC-owl-features-20040210/>. Acesso em Marc. 2012.
- Morris, A.; Velegrakis Y. & Bouquet P. (2008). "Entity Identification on the Semantic Web", In Proc. 5<sup>th</sup> Workshop on Semantic Web Applications and Perspectives, Italy.
- Noy, N. F. & McGuinness, D. L. (2001). "Ontology Development 101: A Guide to Creating your First Ontology", Tech. Report, Stanford University, Stanford, USA.
- Page, L.*et al.* (1999). "The PageRank CitationRanking: Bringing Order to the Web", Technical Report 1999-66, Stanford InfoLab, USA.
- Protégé (2011). "Ontology Editor and Knowledge Acquisition System", v. 4.1 beta, Disponível em: <http://protege.stanford.edu>. Acesso em Dez. 2011.
- Resnick, P. *et al.* (2000). "Reputation Systems", *Commun. ACM*, 43, 12, 45-48.
- Singh, A. P.; Shubhankar, K. & Pudi, V. (2011). "An Efficient Algorithm for Ranking Research Papers Based on Citation Network", In Proc. 3<sup>rd</sup> Conf. on Data Mining and Optimization (DMO), 88-95, Putrajaya, Malaysia.
- Tang, J.*et al.* (2008). "Arnetminer: Extraction and Mining of Academic Social Networks", In Proc. 14<sup>th</sup> ACM Int. Conf. on Knowledge Discovery and Data Mining, 990-998, New York, USA.