

OpenCIn: Usando Dados Abertos e Conectados para a Publicação de dados sobre o CIn/UFPE

Jônatas de Lira Rocha¹, Bernadette Farias Lóscio

¹Centro de Informática – Universidade Federal de Pernambuco (UFPE)
Av. Jornalista Anibal Fernandes, s/n - Cidade Universitária– 50.740-560 – Recife – PE
– Brasil

{jlr, bfl}@cin.ufpe.br

***Abstract.** In general, the currently available information on the Web is produced to be comprehended by humans and not by machines, bringing difficulties both for extraction and for processing such information by computers. In this context, the Semantic Web has been proposed as a solution to facilitate the sharing and processing of data available on the Web because it makes the information readable by machines adding meaning to Web documents. This work aims to create an environment that facilitates access and manipulation of information about the teachers at Centro de Informática - UFPE, using the technologies that the Semantic Web proposes.*

***Resumo.** Em geral, as informações atualmente disponíveis na Web são produzidas para serem entendidas por humanos e não por máquinas, trazendo dificuldades tanto para a extração quanto para o processamento de tais informações por parte de computadores. Neste contexto, a Web Semântica foi proposta como uma solução para facilitar o compartilhamento e processamento de dados disponíveis na Web, pois torna as informações legíveis para máquinas, adicionando sentido aos dados publicados na Web. Neste contexto, este trabalho tem como objetivo a criação de um ambiente que faz uso das tecnologias propostas pela Web Semântica a fim de facilitar o acesso e a manipulação de informações sobre os docentes do Centro de Informática da UFPE.*

1. Introdução

A partir da década de 90, a Web tornou-se uma peça de extrema importância na vida das pessoas. Todos os dias, milhões de pessoas usam a Internet tanto como ferramenta de trabalho, como também para diversão e busca de informações. Dessa forma, com o aumento do uso, também aumentou a quantidade de informações disponíveis na Web. Porém, como as informações nem sempre são estruturadas, diversos desafios precisam ser enfrentados quando se deseja encontrar informações precisas nos buscadores de dados, bem como no momento da integração de dados distribuídos entre fontes de dados distintas. Isso acontece, principalmente, em consequência das limitações dos computadores em “entender” os dados disponíveis da Web. A fim de sobrepor estas limitações, foi proposta a Web Semântica [Berners-Lee e Hendler e Lassila 2001], que tem o objetivo de fazer com que humanos e computadores possam trabalhar em conjunto, permitindo que dados possam ser melhor compreendidos, compartilhados e

reutilizados. A Web Semântica, conseguindo realizar seus objetivos, irá trazer diversas melhorias para o dia-a-dia dos usuários da Web.

Atualmente, as tecnologias da Web Semântica já têm sido amplamente usadas para facilitar a integração e compartilhamento de dados na Web. De acordo com o W3C¹, empresas como BBC e Volkswagen são algumas das instituições que se beneficiaram com o uso da Web Semântica nas suas aplicações. Porém, o benefício com o uso dessas tecnologias pode ir além da área corporativa. É importante notar que a área acadêmica também pode se beneficiar dessas soluções, uma vez que os dados sobre instituições de ensino e pesquisa, geralmente estão disponíveis em fontes de dados distintas, o que dificulta obter uma visão integrada das informações sobre a instituição. Com o uso das tecnologias da Web Semântica é possível conectar os dados que estão distribuídos em diversas fontes, criando um conjunto de dados conectados e facilitando o compartilhamento de informações.

Neste contexto, este trabalho tem como um de seus principais objetivos a aplicação de tecnologias da Web Semântica para a publicação de dados sobre os docentes do Centro de Informática da UFPE (CIn/UFPE). Atualmente, existem diversas informações acerca dos docentes do CIn distribuídos em fontes de dados distintas, como: o curriculum vitae de um professor está disponível na plataforma Lattes², enquanto que outros dados mais específicos estão disponíveis apenas no site do Centro de Informática³. Com o objetivo de integrar informações acerca dos docentes do CIn e outras entidades relacionadas, foi criado o OpenCIn, um ambiente que permite fácil acesso as informações sobre os docentes do CIn/UFPE. A partir do OpenCIn, é possível extrair conhecimentos diversos sobre os docentes, incluindo: suas áreas de interesse, as disciplinas lecionadas por ele, suas publicações, participações em eventos e os projetos que ele faz parte. Como principais componentes do OpenCIn, destacam-se a ontologia OpenCInOnto, que descreve o domínio da aplicação, e uma série de conjuntos de dados abertos conectados, denominados OpenCInData.

O restante deste trabalho está organizado como se segue. Na Seção 2 são apresentados conceitos importantes para o entendimento deste trabalho a partir de uma breve revisão sobre Web Semântica. Na Seção 3 é descrito o processo de criação do OpenCIn, desde a criação da ontologia até o consumo dos dados. Na Seção 4 é apresentado o processo de avaliação da abordagem proposta. Na Seção 5 são expostos os trabalhos relacionados e, por último, são apresentadas as conclusões acerca do trabalho na Seção 6.

2. Conceitos Básicos

De acordo com seu criador, Tim Berners-Lee, a Web Semântica é uma extensão da Web atual onde as informações ganham um significado que pode ser processado por uma máquina e permite que humanos e computadores trabalhem em cooperação [Berners-Lee e Hendler e Lassila 2001]. Mas, para que a Web Semântica faça parte do dia-a-dia das pessoas, é necessário que existam padrões para as tecnologias envolvidas nesse

¹ <http://www.w3.org/2001/sw/sweo/public/UseCases/>

² <http://lattes.cnpq.br/>

³ <http://www.cin.ufpe.br>

conceito. Dessa forma, o W3C (*World Wide Web Consortium*) tem buscado desenvolver padrões para garantir a interoperabilidade de dados no contexto da Web Semântica, como o modelo de dados RDF⁴ e a linguagem de consultas SPARQL⁵.

Um componente fundamental da Web Semântica são as ontologias, pois elas tornam possível a correta interpretação do significado dos dados pelo computador. De maneira geral, uma ontologia específica, de forma explícita, os conceitos de um determinado domínio e permite que essa semântica seja processável por máquina [Antoniou e Harmelen 2008].

A fim de que a publicação de dados na Web seja mais completa e efetiva, é importante que, além de fornecer a semântica dos dados, os dados das diferentes fontes ou conjuntos de dados estejam conectados entre si. Com os dados conectados, os computadores podem navegar entre diferentes fontes, permitindo que o usuário tenha uma visão global das informações sobre um determinado domínio. Para isso, foram propostos os princípios de Linked Data [Heath e Bizer 2011] que buscam estender a Web atual e prover orientações para que os dados fiquem conectados sem necessidade de intervenção por meio de um agente externo, tornando a Web de Documentos em uma Web de Dados.

Outro conceito importante para este trabalho é o de Dados Abertos. De acordo com a *Open Knowledge Foundation*⁶, dados são abertos quando qualquer pessoa pode livremente usá-los, reutilizá-los e redistribuí-los, estando sujeito a, no máximo, a exigência de creditar a sua autoria e compartilhar pela mesma licença [The Open Definition 2015]. Além disso, é importante que os dados abertos sejam publicados em formatos estruturados e facilmente compreensíveis por máquina, a fim de facilitar seu reuso e manipulação de forma automática. O site *OpenGovData.org*, por exemplo, define que um dos princípios para os dados abertos é que eles sejam compreensíveis por máquinas para possibilitar o processamento automatizado. Sendo assim, existe um crescente interesse em utilizar tecnologias da Web Semântica para publicação de Dados Abertos. Como exemplo de iniciativa, destaca-se a de publicação de Dados Abertos do Governo do Reino Unido⁷, onde busca-se utilizar RDF, ontologias e os princípios de *Linked Data* na disponibilização de dados abertos sobre o governo.

3. O Ambiente OpenCIn

Este trabalho tem como principal objetivo a criação de um ambiente para facilitar o acesso a informações acerca dos docentes do CIn/UFPE. Para que esse ambiente fosse criado, foi necessário, primeiramente, elaborar uma ontologia para descrever o domínio, a OpenCInOnto e, após a criação da ontologia, foi gerado um conjunto de *datasets* com os dados dos docentes do CIn, denominado OpenCInData. Por último, foi desenvolvida uma interface amigável para o consumo dos dados do OpenCIn.

⁴ <http://www.w3.org/RDF/>

⁵ <http://www.w3.org/TR/rdf-sparql-query/>

⁶ <https://okfn.org/>

⁷ <http://data.gov.uk/>

A OpenCInOnto foi criada usando a metodologia 101 [Noy e McGuinness 2001] e foi implementada na linguagem OWL. Para criação da OpenCInOnto, inicialmente, foi necessário analisar quais os conceitos e dados referentes aos docentes deveriam ser representados na ontologia. Como resultado, foram obtidas as classes e as propriedades da ontologia, sempre que possível associando os termos criados aos termos de vocabulários já conhecidos, para que os dados pudessem estar conectados. Depois de criada a ontologia, iniciou-se a fase de identificação das fontes de dados relevantes, ou seja, fontes de dados a partir das quais poderiam ser obtidas as informações para a criação dos conjuntos de dados.

A primeira fonte de dados identificada foi o site do CIn/UFPE⁸, que contém as informações básicas acerca dos docentes, como: nome, email, telefone, sala entre outros. A outra fonte de dados foi a Plataforma *Lattes*⁹ do CNPQ, onde todos os professores do CIn possuem cadastro. A última fonte foram alguns documentos internos do centro com informações acerca da alocação dos professores nas disciplinas. Após identificadas as fontes, foram realizadas as etapas de extração e transformação dos dados.

Para a transformação dos dados em triplas RDF, foram construídos alguns *scripts* que acessam o site do CIn e a plataforma *Lattes*. No site do CIn são capturadas informações básicas sobre o professor, como o nome, e-mail, página pessoal, telefone, sala e as áreas de interesse. Já no site da plataforma *Lattes*, foram capturados os currículos dos professores em XML com informações sobre publicações, participações em eventos, áreas de interesse e atuação.

Depois de serem capturados, os dados foram transformados para o formato RDF com a utilização do *framework* Apache Jena. Com os dados em triplas RDF, o *dataset* foi colocado em um servidor Virtuoso, que possibilita fazer consultas SPARQL. Após essa etapa, o consumo de dados pode ser feito através de uma página Web criada para que o usuário possa ter acesso aos dados de forma amigável ou por meio de um *endpoint* SPARQL, onde podem ser realizadas consultas.

Nas próximas seções são apresentados detalhes sobre a OpenCInOnto, o OpenCInData e as opções para o consumo dos dados.

3.1. OpenCInOnto

Para construir o esquema dos dados foi necessário analisar os principais conceitos que poderiam ser aplicados no contexto dos docentes do CIn e que poderiam ser úteis para análise posterior. É importante notar que, não existe uma única forma correta de modelar um domínio, mas para facilitar o processo de criação de uma ontologia foram propostas metodologias que tem a função de sistematizar esse processo e contribuir para a geração de ontologias de boa qualidade. Para a criação da OpenCInOnto foi utilizada a metodologia 101 [Noy e McGuinness 2001].

Os principais conceitos que compõem a OpenCInOnto estão apresentados a seguir:

⁸ <http://www.cin.ufpe.br>

⁹ <http://lattes.cnpq.br>

- **Pessoa:** representa um docente.
- **Publicação:** representa uma publicação de um docente. Uma publicação pode ser um artigo, um capítulo ou um livro.
- **Área de Conhecimento:** representa uma área de especialização ou de atuação de um professor. Uma área de conhecimento também pode estar relacionada com uma Disciplina. Uma área de conhecimento pode ser ou uma área de atuação, como uma das grandes áreas da Computação, como Banco de Dados, Engenharia de Software entre outros, ou uma área de interesse, que é uma área específica dentro de uma área de atuação.
- **Evento Temporário:** é um evento que ocorre durante um intervalo de tempo definido. Pode ser uma alocação, que é uma designação de um ou mais docentes para uma disciplina a ser ministrada durante um intervalo de tempo, ou um cargo, que é uma função pública exercida por uma pessoa dentro de uma organização durante um intervalo de tempo.
- **Agrupamento Acadêmico:** representa uma coleção de recursos acadêmicos com finalidade educacional ou de pesquisa num escopo bem definido dentro de uma organização. Um agrupamento acadêmico pode ser uma Disciplina ou um Projeto.
- **Disciplina:** é um componente curricular oferecido por uma organização que possui associação com uma ou mais Áreas de Conhecimento. Pode ser uma disciplina obrigatória ou eletiva de algum curso de graduação ou pós-graduação.

O reuso de termos de vocabulários é muito recomendado para uma aplicação que segue os princípios de *Linked Data*, uma vez que facilita a integração de dados de diferentes fontes, facilitando o seu processamento. Porém, caso não seja possível encontrar um vocabulário que consiga descrever os conceitos sendo representados, torna-se necessária a criação de novos termos.

Durante o processo de criação da OpenCInOnto, foram encontrados quatro vocabulários que puderam ser reusados na ontologia: o FOAF (*Friend of a Friend*)¹⁰, o AIISO (*Academic Institution Internal Structure Ontology*)¹¹, o AKT (*Advanced Knowledge Technologies – AKT Portal Ontology*)¹² e o BIBO (*Bibliographic Ontology*)¹³.

Na Tabela 1 podem-se encontrar alguns dos principais conceitos reutilizados para a criação das classes da OpenCInOnto.

¹⁰ <http://www.foaf-project.org>

¹¹ <http://vocab.org/aiiso/schema>

¹² <http://www.aktors.org>

¹³ <http://bibliontology.com>

Tabela 1. Termos reusados para classes

Conceito	Termo reusado
Pessoa	<i>foaf:Person</i>
Agrupamento Acadêmico	<i>aiiso:KnowledgeGrouping</i>
Área de Conhecimento	<i>akt:Research-Area</i>
Docente	<i>akt:Academic</i>
Disciplina	<i>aiiso:Course</i>
Centro	<i>aiiso:Center</i>
Grupo de Pesquisa	<i>aiiso:ResearchGroup</i>
Artigo	<i>bibo:Article</i>
Capítulo	<i>bibo:Chapter</i>
Livro	<i>bibo:Book</i>

Apesar dos vocabulários existentes já preencherem uma boa parte dos termos necessários para criação da ontologia, ainda foi necessário criar novos conceitos e propriedades. O principal motivo para isso é a especificidade das novas classes e propriedades no contexto da OpenCInOnto, não sendo possível encontrar um conceito ou propriedade já existente que tivesse a mesma semântica.

O diagrama com a hierarquia de classes da ontologia e com as principais propriedades de objetos que fazem o relacionamento entre as classes está disponível em <http://www.cin.ufpe.br/~jlr/openCIn/>. A ferramenta utilizada para criação da ontologia foi o *Protégé*¹⁴, um editor de ontologias de código aberto e gratuito, considerada uma das mais usadas para a construção de ontologias atualmente. O *Protégé* permite, através de uma interface gráfica, a criação de classes e todas as propriedades de uma ontologia, além de oferecer meios de visualizar a ontologia por meio de diagramas.

3.2. OpenCInData

Após a construção da ontologia, fez-se necessário extrair os dados necessários das fontes de dados e transformá-los em RDF. Como dito anteriormente, os dados foram extraídos de três fontes de dados diferentes, o site do Centro de Informática, o site da plataforma Lattes e alguns documentos internos do CIn. Juntando os dados dessas três fontes é possível obter o conteúdo necessário para montar um ambiente com informações relevantes acerca dos docentes do CIn/UFPE. A Figura 2 ilustra o processo desde a extração e transformação dos dados até a exibição deles para o usuário.

¹⁴ <http://protege.stanford.edu>

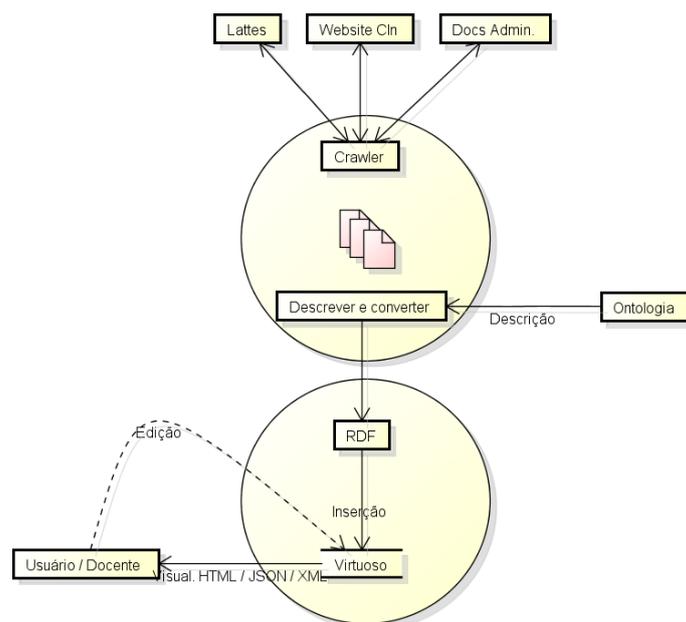


Figura 2. Processo de extração e descrição dos dados

Para extrair os dados do site do CIn, foi desenvolvido um *crawler* capaz de acessar todas as páginas referenciadas pelo website que descrevem algum docente e convertê-las em arquivos XML. Como a corretude desses dados é essencial para promover a devida identificação única de cada docente e, conseqüentemente, evitar inconsistências, as informações essenciais de cada professor (como *login*, *email* e *lattes*) foram checadas e validadas manualmente.

A partir do link de cada professor para o seu respectivo currículo na plataforma Lattes, foi possível fazer o download dos dados em formato XML, os quais são gerados pela própria plataforma. Mas como a plataforma não autoriza o *download* em lotes de currículos lattes sem a devida autorização expressa, foi necessário baixar cada arquivo separadamente.

Além disso, também foram extraídos dados advindos de fontes internas e/ou administrativas, como documentos de alocação professor-disciplinas, orientações de alunos e cargos administrativos ocupados. Esses documentos geralmente estão em formato XLS ou CSV, pois esses formatos são bem comuns nos ambientes administrativos.

De posse dos dados obtidos a partir das diversas fontes, foi feita a transformação dos dados para o formato RDF e de acordo com a descrição do domínio fornecida pela OpenCInOnto. Para converter os documentos XML gerados na fase de extração para RDF, foi desenvolvido um descritor em Java que reconhece os atributos do arquivo XML, a partir de uma gramática pré-definida para a extração.

Após a geração dos dados em RDF, foi feita sua inclusão em um repositório de dados que provê suporte a RDF e que oferece uma interface de consulta para acesso aos dados. Como solução para repositório de dados, foi adotado o Virtuoso¹⁵ que oferece

¹⁵ <http://semanticweb.org/wiki/Virtuoso>

suporte para RDF e disponibiliza um *endpoint* para consultas via SPARQL. A inserção foi feita via repositório WebDav¹⁶ em diretórios internos.

3.3. Consumo de Dados do OpenCIn

Para que os dados do OpenCInData pudessem ser consumidos foi necessário que os conjuntos de dados RDF fossem armazenados em um banco de dados nativos para triplas RDF, denominados repositórios RDF. Para que possam ser feitas consultas SPARQL é necessário a disponibilização de um *endpoint* que se comunica com o repositório RDF¹⁷.

Com o endpoint disponível, é possível realizar consultas na linguagem SPARQL, como: “*select distinct ?name ?lattes where { ?x cin:name ?name . ?x cin:lattes ?lattes}*”. A Figura 4 mostra o resultado dessa consulta, que são os nomes de todos os professores cadastrados e o *link* para o currículo Lattes desses professores.

name	lattes
Cristiano Coêlho de Araújo	http://lattes.cnpq.br/9638500605562489
Daniel Carvalho da Cunha	http://lattes.cnpq.br/8971986984647323
Silvio Romero de Lemos Meira	http://lattes.cnpq.br/4973731190814126
Nelson Souto Rosa	http://lattes.cnpq.br/4220236737158909
Carlos André Guimarães Ferraz	http://lattes.cnpq.br/7716805104151473
Francisco de Assis Tenorio de Carvalho	http://lattes.cnpq.br/3909162572623711
Roberto Souto Maior de Barros	http://lattes.cnpq.br/2153962690732683
Judith Kelner	http://lattes.cnpq.br/7532050172035129
Frederico Luiz Gonçalves de Freitas	http://lattes.cnpq.br/6195215666638965
Ricardo Martins de Abreu Silva	http://lattes.cnpq.br/6234141900588267

Figura 4. Dados retornados em HTML após consulta ao Virtuoso via SPARQL

Além do endpoint SPARQL, a fim de permitir que um número maior de usuários pudesse consumir os dados do OpenCIn, ou seja, aqueles que não conhecem SPARQL, foi construída uma aplicação¹⁸, denominada OpenCInExplorer, onde é possível visualizar de forma amigável as informações contidas nos *datasets* RDF.

4. Avaliação

Depois da criação da ontologia, da construção do conjunto de *datasets* e da implementação da aplicação para consumo de dados, foi feita uma avaliação do trabalho. O critério para avaliação consistiu em identificar se o OpenCIn permite a obtenção de uma visão integrada para os dados sobre os docentes do CIn, ou seja, se uma consulta a diversas fontes resulta numa resposta unificada.

Por um lado, na aplicação OpenCInExplorer, pode-se ver os dados do perfil do docente, as publicações e as áreas de interesse na mesma interface, sem que seja

¹⁶ <http://shadok.enst.fr:8890/doc/html/qswebdav.html>

¹⁷ <http://webproj04.cin.ufpe.br/sparql/>

¹⁸ <http://www.cin.ufpe.br/~fmca/cinexplorer/>

necessário acessar as diferentes fontes onde os dados se encontram. Por outro lado, no *endpoint* SPARQL é permitido fazer consultas a todas as informações de diferentes fontes em um mesmo local, oferecendo uma interface única de acesso aos dados. Por exemplo, a consulta a seguir retorna o número da sala, que está disponível no site do CIn, e todos os títulos das publicações dos professores, que estão disponíveis no *Lattes*: “*select distinct ?name ?office ?publication where { ?x cin:name ?name . ?y cin:idProfessor ?x . ?y cin:title ?publication . ?x cin:office ?office}*”. O resultado desta consulta é mostrado na figura 5.

name	office	publication
Djamel Fawzi Hadj Sadok	C132	Third IEEE International Conference on Mobile and Wireless Communications - MWCN'2001
Djamel Fawzi Hadj Sadok	C132	ProNet: a Policy Networks Environment
Djamel Fawzi Hadj Sadok	C132	Self-Addressing for Autonomous Networking Systems
Djamel Fawzi Hadj Sadok	C132	MC01: Construção de Redes de Voz Sobre IP
Djamel Fawzi Hadj Sadok	C132	Estimating Video Quality over ADSL2+ under Impulsive Line Disturbance.
Djamel Fawzi Hadj Sadok	C132	Resource Allocation in Clouds: Concepts, Tools and Research Challenges
Djamel Fawzi Hadj Sadok	C132	Knowledge Sharing to Improve Routing and Future 4G Networks
Djamel Fawzi Hadj Sadok	C132	Self-organization of Addresses Protocol: Supporting Autoconfiguration in Autonomous Networks
Djamel Fawzi Hadj Sadok	C132	Challenges for the Brazilian Green and Yellow Internet
Djamel Fawzi Hadj Sadok	C132	Better network traffic identification through the independent combination of techniques?
Djamel Fawzi Hadj Sadok	C132	Managing the Future Internet: Services, Policies and Peers
Djamel Fawzi Hadj Sadok	C132	Tráfego Internet não Desejado: Conceitos, Caracterização e Soluções.
Djamel Fawzi Hadj Sadok	C132	Construção de Redes de Voz sobre IP

Figura 5. Dados de diversas fontes sendo exibido em um só local

5. Trabalhos Relacionados

Existe um grande esforço das universidades, principalmente no exterior, para tornar seus dados públicos disponíveis para acesso em *Linked Data*. Na Europa, foi criado um portal chamado *LinkedUniversities.org*¹⁹ que reúne as universidades européias que têm seus dados disponibilizados usando os princípios do *Linked Data*. A *University of Munster*, *University of Bristol*, *The Open University* e mais sete outras fazem parte desse projeto que visa facilitar a criação de dados para Web Semântica por meio da colaboração entre as universidades. Além das instituições de ensino, outros setores da sociedade estão interessados em disponibilizar os seus dados usando as tecnologias da Web Semântica. O esforço mais visível é do grupo *Linked Open Data* [Bizer e Heath e Berners-Lee 2009] que tem como objetivo converter dados abertos disponíveis na Web, como dados da Wikipédia²⁰, para RDF de acordo com os princípios do *Linked Data*.

6. Conclusão

A Web Semântica tem o poder para evoluir a Web atual, oferecendo mecanismos e ferramentas para armazenar e recuperar informações que são extremamente eficientes,

¹⁹ <http://linkeduniversities.org/index.html>

²⁰ <https://www.wikipedia.org/>

além de oferecer maior capacidade de inferência. Para que tudo isso ocorra, é necessário que os dados estejam organizados por meio de formatos bem definidos, como RDF, e OWL, permitindo que os dados sejam manipulados de forma eficaz.

Como resultado deste trabalho, os dados de várias fontes de dados distintas e sem comunicação a respeito dos docentes do CIn/UFPE foram integrados, gerando um conjunto de dados abertos e conectados. Além disso, também foi criada uma ontologia, a OpenCInOnto, para a representação dos principais conceitos relacionados aos docentes CIn.

É importante destacar que o OpenCIn pode ser uma ferramenta muito útil tanto para alunos, quanto professores e para a administração do CIn. A partir do OpenCIn é possível determinar facilmente quais docentes estão alocados em disciplinas da sua área, como também permitir que um aluno possa identificar professores que são mais indicados para uma possível orientação. Além disso, pode auxiliar a formação de parcerias entre professores, uma vez que todos os docentes terão acesso facilitado as informações de seus pares. Como próxima atividade, será realizado um *hackaton* para estimular outros alunos a fazerem aplicações usando os dados do OpenCIn. Esse tipo de evento é bastante valorizado atualmente e sempre gera bons resultados, pois as aplicações geram informações que nem sempre foram pensadas pelos idealizadores, expandindo a utilidade do *dataset*. Além disso, como trabalhos futuros destacam-se a extensão da ontologia para incluir classes e propriedades referentes ao portal de notícias do CIn, para integrar os dados do site no nosso *dataset*.

Referências

- DuCharme, B. (2011), “Learning SPARQL”, O’Reilly Media, First Edition.
- Dodds, L. e Davis, I. (2012), “Linked Data Patterns: A pattern catalogue for modelling, publishing, and consuming Linked Data”, <http://patterns.dataincubator.org/book/linked-data-patterns.pdf>, Abril.
- Berners-Lee, T. e Hendler, J. e Lassila O. (2001), “The Semantic Web. ScientificAmerican”, v. 284, n. 5.
- The Open Definition (2015), “The Open Definition - Open Definition - Defining Open in Open Data, Open Content and Open Knowledge”, <http://opendefinition.org>, Fevereiro.
- Noy, N. F e McGuinness, D. L. (2001), “Ontology Development 101: A Guide to Creating Your First Ontology”, Stanford University, Stanford
- Antoniou, G. e Harmelen, F. V. (2008), “A Semantic Web Primer”, Massachusetts Institute of Technology, Cambridge.
- Heath, T. e Bizer, C. (2011), “Linked data: Evolving the web into a global data space”, Morgan & Claypool Publishers.
- Bizer, C. e Heath, T. e Berners-Lee, T. (2009), “Linked data-the story so far”, <http://eprints.soton.ac.uk/271285/1/bizer-heath-berners-lee-ijswis-linked-data.pdf>, Abril.