

Publicação e consumo de dados abertos conectados acadêmicos

André de Alencar ^[1], Douglas Xavier ^[2], Luiz Carlos Chaves ^[3], Damires Souza ^[4]

[1] andrealencar@outlook.com.br. [2] douglasf.xavier@gmail.com. [3] lucachaves@gmail.com. [4] damires@ifpb.edu.br. Instituto Federal da Paraíba /Unidade Acadêmica de Informática.

RESUMO

A Web atualmente é considerada um ecossistema propício à publicação e consumo de dados abertos. Os dados abertos podem estar também conectados, o que implica no uso de tecnologias semânticas, como vocabulários referendados, e na existência de links que permitam sua ligação com outros conjuntos de dados. Considerando um escopo de dados da Unidade Acadêmica de Informática do IFPB-Campus João Pessoa, um conjunto de dados abertos e conectados foi criado e publicado para consumo. Ele inclui informações oriundas da Plataforma Lattes e de dados internos da Unidade sobre professores, projetos, cursos e áreas de atuação. Para isso, os dados passaram por um processo de extração, transformação e carga com base no uso de uma ontologia chamada *Ontology for University and Academic Institutions* (OUAI), desenvolvida neste trabalho. Como resultado, os dados foram publicados usando o modelo RDF e disponibilizados para consumo, por meio de um endpoint. Por intermédio de seu consumo, foi desenvolvida uma aplicação de visualização dos dados, denominada de OpenUAI, que objetiva ser uma vitrine de pessoas e atividades da referida Unidade. Este trabalho, portanto, apresenta o processo instanciado de publicação e consumo dos dados, a ontologia OUAI criada, a aplicação OpenUAI e os resultados da avaliação realizada.

Palavras-chave: Dados Acadêmicos Abertos e Conectados. Ontologia. Publicação e Consumo de Dados. Visualização de dados.

ABSTRACT

Nowadays, the Web may be considered an adequate ecosystem for publication and open data consumption. Published datasets may provide open and, additionally, linked data, which results in the use of semantic technologies such as recommended vocabularies and their connection with other datasets. Taking into account a data scope from the Academic Unit of Informatics at IFPB-Campus João Pessoa, a set of open and linked data was created and published for consumption. This dataset includes information obtained from the Lattes Platform and from some internal data regarding teachers, projects, courses and areas of expertise. Source data went through a process of extraction, transformation and load based on the use of an ontology, named "Ontology for University and Academic Institutions" (OUAI), which was developed in this work. As a result, the dataset was published in the RDF model and was made available for consumption through an endpoint. Based on data consumption, the OpenUAI application was developed as a means to provide data visualizations on the unit activities and people. This work presents the process regarding the data publication and consumption, the ontology created to help matters, the OpenUAI application and some results obtained with the performed evaluation.

Keywords: *Academic Open Linked Data. Ontology. Data Publication and Consumption. Data Visualization.*

1 Introdução

A Web atualmente é considerada um ecossistema propício à produção e consumo de dados abertos. Segundo a *Open Knowledge Foundation Brasil*¹, os dados estão abertos quando qualquer pessoa pode livremente usá-los, reutilizá-los e redistribuí-los, estando estes sujeitos, no máximo, à exigência de creditar a sua autoria e a seu compartilhamento pela mesma licença. Além de estarem em formatos abertos, como, por exemplo, em CSV (*Comma-separated Values*)² ou XML (*Extensible Markup Language*)³, eles podem estar conectados a outros conjuntos de dados. Para isso, os conjuntos de dados devem ser construídos tendo como base os princípios de *linked data* (dados conectados), utilizando tecnologias da Web Semântica. Isso significa que os conjuntos de dados estarão publicados em modelo de dados RDF (*Resource Description Framework*), em que cada recurso é identificado unicamente por meio de URI (*Uniform Resource Identifier*), e podem existir links nomeados entre esses recursos (HEATH; BIZER, 2011). Além disso, é necessário utilizar vocabulários de referência de termos, para que a semântica dos dados esteja bem definida e padronizada (LÓSCIO; BURLE; CALEGARI, 2017).

Na realidade de uma instituição de ensino, há vários conjuntos de dados como, por exemplo, currículos de professores, ementas de disciplinas, cursos e projetos, que podem ser disponibilizados de forma aberta. Se, além de abertos, eles estiverem estruturados semanticamente e conectados por meio de serializações RDF, fazendo uso de vocabulários referenciados e links, tornam-se mais compreensíveis e processáveis tanto por humanos quanto por máquinas (LÓSCIO; BURLE; CALEGARI, 2017). Como resultado, esses conjuntos de dados se tornam mais reutilizáveis pela própria instituição ou por outras, além de favorecerem a criação de aplicativos que ajudam a sociedade a conhecer melhor a instituição, suas atividades, membros e informações importantes.

Com base nesse panorama, este trabalho aborda o problema de publicação e consumo de dados abertos conectados na Web, considerando como escopo

dados acadêmicos da Unidade Acadêmica de Informática (UAI) do IFPB (Instituto Federal da Paraíba)-Campus João Pessoa. Para isso, foi instanciado um processo de Extração, Transformação e Carga (ETL) baseado no uso de um vocabulário especificamente criado. O vocabulário foi desenvolvido por meio de uma ontologia, denominada de OUI (*Ontology for Universities and Academic Information*), que reusa termos de vocabulários recomendados e acrescenta outros específicos. O vocabulário OUI é utilizado para referenciar os termos no processo de ETL e gerar uma base de dados integrados, provenientes da Plataforma Lattes e de dados internos da Unidade. O conjunto de dados criado em RDF possui um SPARQL endpoint que permite seu consumo. A partir do consumo do conjunto de dados publicado, foi criada uma aplicação de visualização dos dados, denominada de OpenUI, que objetiva apresentar atividades e pessoas da unidade citada de forma interativa e simples. Foi realizada uma avaliação da OpenUI com base em um questionário aplicado a potenciais usuários da aplicação e dos dados. As contribuições mencionadas são apresentadas neste artigo.

Este artigo está organizado da seguinte forma: a Seção 2 introduz alguns conceitos e tecnologias utilizados no trabalho; a Seção 3 apresenta as contribuições; a Seção 4 mostra uma parte do conjunto de dados como ilustração e a avaliação realizada sobre a aplicação OpenUI; a Seção 5 discute trabalhos relacionados e os compara a este. Na Seção 6, são tecidas considerações e apontados trabalhos futuros.

2 Referencial teórico

Com os conceitos da Web Semântica e, em particular, dos princípios de dados abertos conectados (*open linked data*), está sendo possível estender a Web atual para uma Web de Dados (BERNERS-LEE; HENDLER; LASSILIA, 2001). O objetivo da Web Semântica é a extensão da Web para um ambiente onde documentos são publicados não apenas com seus dados, mas também com os metadados padronizados semanticamente, o que permite uma maior compreensão e processabilidade acerca dos dados (HEATH; BIZER, 2011). A ideia é que tanto humanos quanto máquinas possam entender e manipular mais facilmente os dados (RUSSO, 2015).

Nesse contexto, os dados conectados podem ser publicados seguindo preceitos de dados abertos. Os dados abertos devem ser completos, atuais, acessíveis, processáveis por máquina e devem estar em

1 <https://okfn.org/opendata/>

2 <https://www.w3.org/TR/tabular-data-primer/>

3 <https://www.w3.org/XML/>

formatos não proprietários e sob licença que permita seu reuso (OLIVEIRA *et al.*, 2016; ISOTANI, BITTEN-COURT, 2015).

O modelo de dados RDF permite que os dados estejam abertos e possuam os metadados referenciados semanticamente, de modo que sua compreensão e processabilidade se tornam mais simples. Dados em RDF são organizados em estrutura de grafo composto de triplas (S, P, O), em que S representa um recurso sujeito, P indica um predicado e O um valor de objeto que pode ser um literal ou um recurso (HEATH; BIZER, 2011). Cada recurso é identificado por meio de URI. Um documento RDF codifica um conjunto de dados em uma serialização RDF como *Turtle*⁴, *Resource Description Framework in Attributes* (RDFa)⁵ ou *JavaScript Object Notation for Linked Data* (JSON-LD)⁶ (CYGANIAK; WOOD; LANTHALER, 2014).

Para prover o referencial semântico dos metadados, é necessário adotar vocabulários que são conjuntos de termos de um determinado domínio de conhecimento normalmente publicados e acessíveis na Web (HEATH; BIZER, 2011). Os vocabulários podem ser construídos por meio de ontologias (GRUBER, 2009). Os termos de vocabulários (e.g., *foaf:name*) são usados para caracterizar propriedades, relacionamentos ou recursos.

Para prover consultas a conjuntos de dados RDF, é utilizada a linguagem *Protocol and RDF Query Language* (SPARQL)⁷. Ela fornece opções que permitem consultas e operações de manipulação. A forma mais comum de realizar consultas é usando a cláusula *SELECT*, que retorna variáveis presentes em padrões de triplas passados dentro da cláusula *WHERE*.

Como ilustração dos conceitos introduzidos, considere a Figura 1 e a Figura 2. A primeira mostra um fragmento de um documento RDF serializado em Turtle. Os dados descrevem uma publicação com suas propriedades (predicados) e seus valores. A publicação identificada pela URI <http://roma.rkbexplorer.com/id/publications/Baldoni06> possui título (*akt:has-title*) com valor "Publish/Subscribe on the top of Overlay networks: when subscription dynamics meet churn."

A Figura 2 provê uma consulta SPARQL, que busca o título da publicação cujo recurso é <http://roma.rkbexplorer.com/id/publications/Baldoni06>. O resultado mostra o título obtido que se encontra no RDF exibido na Figura 2.

Figura 1 – Um fragmento de documento RDF em RDF/Turtle

```
<http://roma.rkbexplorer.com/id/publications/Baldoni06>
  a <http://www.aktors.org/ontology/portal#Proceedings-Paper-Reference>;
  akt:has-title "Publish/Subscribe on the top of Overlay networks: when subscription dynamics meet churn." ;
  akt:has-date <http://www.aktors.org/ontology/date#2006>;
  akt:has-author <http://roma.rkbexplorer.com/id/person-00001>.
...
<http://roma.rkbexplorer.com/id/person-00001>
  a akt:Person;
  akt:full-name "Roberto Baldoni".
```

Fonte: Repositório Roma RKB Explorer.⁸

Figura 2 – Exemplo de consulta SPARQL

SPARQL Query Interface

This interface permits queries to be made over the information held within the repository, u

Result format: HTML Table

Query:

```
PREFIX id: <http://roma.rkbexplorer.com/id/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX akt: <http://www.aktors.org/ontology/portal#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX akts: <http://www.aktors.org/ontology/support#>

SELECT ?titulo
WHERE {<http://roma.rkbexplorer.com/id/publications/Baldoni06> akt:has-title ?titulo}
```

Results:

Result	Binding	Value
1	?titulo	Publish/Subscribe on the top of Overlay networks: when subscription dynamics meet churn.

1 result(s) found in 0.0 seconds.

Fonte: Autoria própria.

3 Abordagem OpenUIA

A abordagem definida neste trabalho inclui a instanciamento de um processo de construção do conjunto de dados abertos conectados, como mostra a Figura 3 (ALENCAR *et al.*, 2017). Conforme o processo apresentado, os dados são extraídos e selecionados dos currículos Lattes dos professores, de dados institucionais do sistema acadêmico interno do IFPB e de referências como a tabela de áreas de conhecimento fornecida pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes). Para prover

4 <https://www.w3.org/TR/turtle/>

5 <https://www.w3.org/TR/rdfa-primer/>

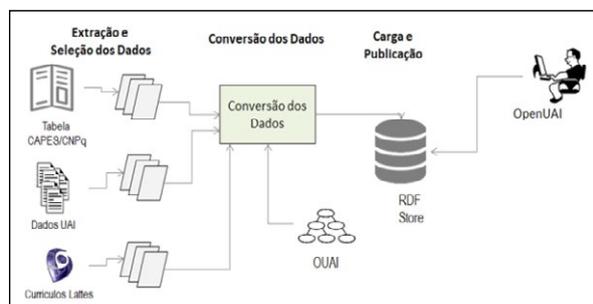
6 <https://www.w3.org/TR/json-ld/>

7 <https://www.w3.org/TR/rdf-sparql-query/>

8 <http://roma.rkbexplorer.com/>

a conversão dos dados coletados para RDF, foi necessário criar um vocabulário que pudesse viabilizar seu referenciamento semântico. Isso foi realizado por meio da criação da ontologia *Ontology for University and Academic Institutions* (OUAI), a ser descrita na Seção 3.2. Após a triplicação dos dados para RDF, estes foram tornados acessíveis por meio de um serviço a partir do *RDF Store*. A aplicação OpenUAI foi desenvolvida como a primeira a consumir os dados publicados de modo que eles se tornem mais simples e acessíveis ao público.

Figura 3 – Processo de publicação e consumo de dados abertos conectados da UAI



Fonte: Autoria própria.

Todo o processo instanciado é baseado em princípios de Extração, Transformação e Carga (ETL), porém com etapas adicionais associadas ao referenciamento semântico dos dados e metadados. Os dados gerados compreendem dados de professores (perfis), projetos de pesquisa, extensão e inovação, áreas de atuação e dados sobre cursos e disciplinas da UAI. Esta é a primeira iniciativa de disponibilizar e usar dados abertos conectados de uma unidade acadêmica do IFPB com o propósito de facilitar seu acesso e visualização.

Cada etapa do processo será descrita a seguir.

3.1 Coleta e extração dos dados

Nesta etapa, foram coletados os currículos Lattes dos professores lotados na UAI. Os dados extraídos eram originalmente arquivos XML obtidos através de um *scraper* codificado na linguagem Python. A aplicação teve como finalidade acessar individualmente perfis do Lattes de acordo com o ID de cada professor registrado em uma planilha CSV previamente criada. Assim, foram extraídos 32 currículos Lattes para a conversão.

Os dados institucionais foram obtidos em formato CSV a partir do sistema acadêmico do IFPB e incluíram dados sobre turmas, disciplinas e cursos.

Também foi coletada a tabela de áreas de conhecimento da CAPES, que contém a taxonomia de áreas usadas pela Plataforma Lattes.

3.2 Criação da ontologia OUAI

Uma vez que não foi encontrada uma ontologia completa adequada para o domínio de dados deste trabalho, isto é, para dados acadêmicos e de pesquisa, foi necessário criar uma nova ontologia, a chamada OUAI. A OUAI considera o contexto de dados acadêmicos e de pesquisa do Brasil e, em especial, oferece cobertura de termos tanto para os dados dos currículos Lattes quanto para os dados internos da UAI.

Para compor a OUAI, fez-se necessário conhecer a taxonomia do esquema do XML do currículo Lattes, assim como os dados fornecidos pela UAI (sobre cursos, disciplinas e turmas). Com base nessa compreensão, foi instanciado um processo iterativo e incremental de construção da ontologia, observando-se metodologias de desenvolvimento de ontologias (SURE; STAAB; STUDER, 2009). O processo foi composto das seguintes etapas: (i) determinação do domínio do conhecimento e seu escopo (neste caso, dados acadêmicos e de pesquisa); (ii) enumeração e definição de termos; (iii) pesquisa de vocabulários existentes e relevantes para permitir a reutilização de alguns termos; (iv) definição de classes, hierarquias e propriedades; (v) validação dos termos ontológicos por especialistas no domínio. No nosso caso, professores e pesquisadores da área forneceram tal validação.

Durante o processo, uma pesquisa foi realizada no sentido de analisar e comparar vocabulários e ontologias dentro do mesmo domínio de conhecimento. A principal fonte para essa investigação foi o site LOV (*Linked Open Vocabularies*)⁹, um repositório de vocabulários abertos com *search engine* que oferece busca por termos. Além dele, foram consultados o swoogle 10 e o falcons 11. Na escolha dos vocabulários, alguns requisitos foram ponderados: i) cobertura

9 <http://lov.okfn.org>

10 <http://swoogle.umbc.edu/>

11 <http://ws.nju.edu.cn/falcons/ontologysearch/>

de termos compatíveis com o escopo dos dados; e ii) estabilidade e disponibilidade do vocabulário, apresentando URIs ativos.

Dessa forma, ao final das análises, elegeram-se para reuso os seguintes vocabulários: *bibtex*, *dbpedia*, *dcterms*, *Friend of a Friend (foaf)* e *Semantic Web Portal Ontology (swpo)*. Alguns dos termos desses vocabulários foram reutilizados, como mostra o Quadro 1. Outros termos não encontrados foram originalmente criados na OUAJ, como é retratado na última linha do Quadro 1. Este quadro apresenta uma parte dos termos reusados e criados. No total a OUAJ possui 129 termos entre os criados e reusados de outros vocabulários.

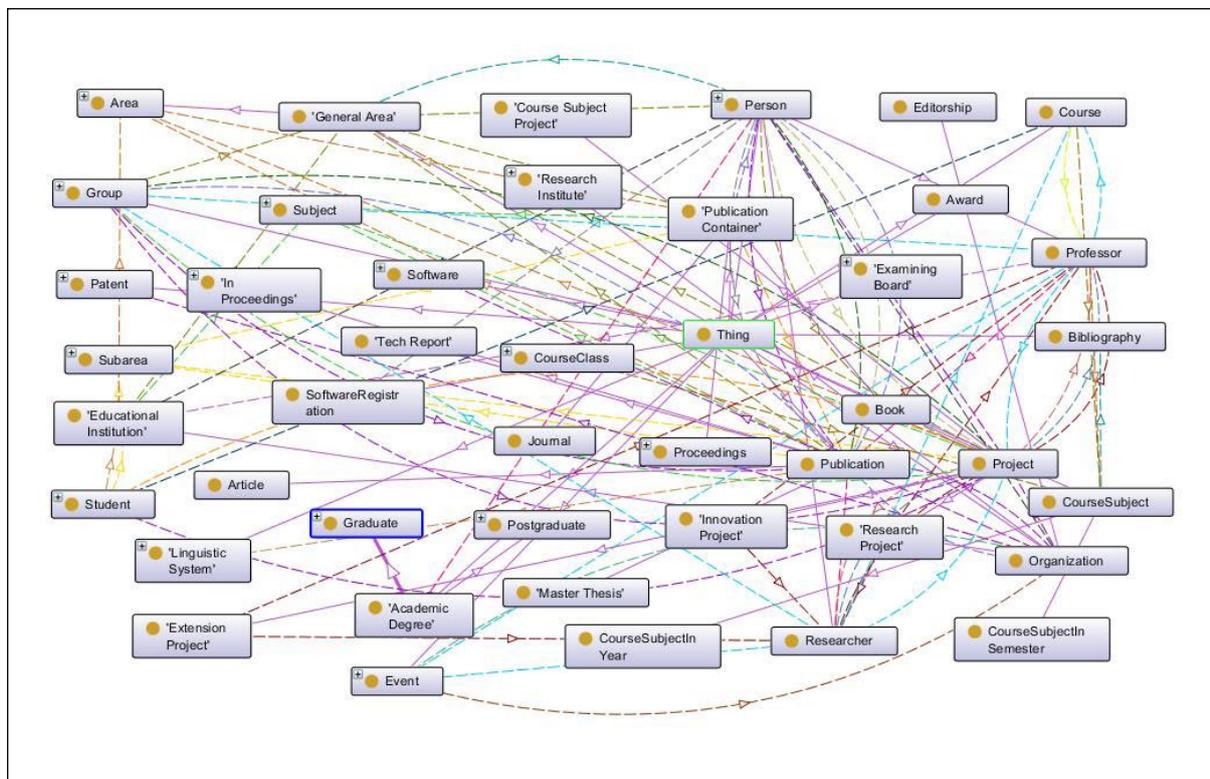
A Figura 4 mostra um fragmento da OUAJ, no qual é possível observar termos representados pelos retângulos com seus respectivos rótulos. Na figura, é possível ver também relacionamentos entre eles, representados pelas linhas tracejadas.

Quadro 1 – Vocabulário e termos da OUAJ

VOCABULÁRIO	TERMOS REUSADOS OU CRIADOS
FOAF	Person, name
SWPO	Researcher, Topic, Publication, Article, TechnicalReport, Inproceedings, PublicationContainer, Journal, Proceedings, PublishingCompany, hasAuthors
DBPEDIA	EducationalInstitution, Occupation, Project, ResearchProject
VIVO	Department, AcademicDepartment
DCTerms	Language
OUAJ	User, EducationalLevel, GeneralKnowledgeArea, KnowledgeArea, Keyword, Language, DevelopmentProject, ExtensionProject, InnovationProject, lattesURL

Fonte: Autoria Própria.

Figura 4 – Fragmento da OUAJ



Fonte: Autoria própria.

Os termos criados foram necessários para o escopo de dados do trabalho e estavam ausentes nos vocabulários pesquisados. Abrangem um escopo associado a dados acadêmicos, de pesquisa, inovação e extensão. Foram definidos tanto em português quanto em inglês, para prover maior interoperabilidade.

Como exemplos, o termo *CourseSubject* representa a classe de uma disciplina de um curso, *CourseClass*, por sua vez, reflete a turma de um curso e *InnovationProject* indica um projeto de inovação. Usando o prefixo definido para a ontologia (de acordo com um namespace), esses termos passam a ser *ouai:CourseSubject*, *ouai:CourseClass* e *ouai:InnovationProject*, respectivamente.

3.3 Conversão dos dados

Uma vez que a coleta dos dados e o vocabulário OUAL estavam prontos, a próxima etapa consistiu na conversão dos dados, originalmente em CSV e XML, para o formato RDF. Para isso, foi desenvolvida uma aplicação, codificada em Java, que permitiu o tratamento dos dados extraídos e sua conversão para grafos em RDF.

Nessa etapa é necessário que a ligação dos dados, no grafo de recursos (classes e propriedades), respeite a estrutura semântica definida no vocabulário OUAL. Isso significa que a conversão tem de considerar os relacionamentos entre professor, projeto, curso, turma, entre outros e criar as triplas de acordo com o que está definido na OUAL. Assim, por exemplo, uma instância de Professor (recurso) passa a se conectar com as instâncias de seus projetos e suas turmas. Essa identificação é parte do processo de conversão e é uma maneira de integrar os dados das diversas fontes de origem.

No caso dos arquivos CSV, cada linha é reconhecida como a instância de um recurso de uma determinada classe correspondente. Quando se trata do CSV da lista de disciplinas, por exemplo, cada linha é convertida em uma instância da classe *ouai:CourseSubject*, que terá suas propriedades geradas a partir dos campos do CSV. Para os arquivos XML, o mesmo princípio é usado.

O resultado da conversão é que todos os dados são relacionados entre si de acordo com o contexto semântico dos dados acadêmicos da UAI juntamente com os dados do Lattes e das definições de semântica e de termos da OUAL. Isso permite uma maior facilidade na consulta aos dados, pois, uma vez que os dados foram mapeados, todos os atributos

correspondentes podem ser usados. Exemplos de mapeamentos são mostrados no Quadro 2. Ao final da conversão, são montadas as triplas identificando o sujeito (recurso em questão), suas propriedades e objetos. O conjunto de triplas forma o grafo RDF.

Quadro 2 – Exemplos de mapeamentos entre campos e termos da OUAL

CAMPO DO CSV	TERMO DA OUAL
Nome da Disciplina	dc:title
Período	ouai:courseSemester
Carga Horária	time:hours
Ementa	ouai:courseContent
Bibliografia básica	ouai:hasBibliography

Fonte: Autoria própria.

3.4 Carga e publicação

O grafo RDF gerado precisa ser persistido para permitir seu uso. Entre as possíveis formas de persistir dados em RDF, existe o uso de bancos de dados específicos, também conhecidos como RDF *stores*. Para este trabalho, o banco escolhido foi o Virtuoso (OPENLINK, 2015), por ser estável e largamente utilizado para propósitos semelhantes.

O grafo RDF gerado foi, então, persistido no Virtuoso. Além do armazenamento, o Virtuoso trouxe recursos de administração e visualização dos dados, como a possibilidade de publicar dados para aplicações Web por meio do seu *SPARQL endpoint*, que funciona como um *web service*. Isso possibilita o consumo dos dados coletados da UAI de forma pública, preservando, ao mesmo tempo, a conexão e a semântica próprias do modelo RDF. Portanto, com consultas adequadas, é possível gerar as informações acadêmicas desejadas.

3.5 Consumo de dados

Em consequência da publicação dos dados, a última parte do processo consiste em realizar o consumo dos dados disponibilizados. Para isso, requisições por meio do *endpoint* são realizadas.

Como forma de testar essa utilização dos dados por possíveis aplicações, foi idealizada a criação da primeira aplicação para consumir os dados da UAI. A aplicação em questão será descrita na próxima seção.

3.6 A aplicação OpenUAI

A OpenUAI surgiu com o intuito de atender antigas demandas sobre informações acerca de perfis dos professores, projetos executados, áreas de atuação da Unidade, conforme a padronização da C, e cursos ofertados. A ideia é montar uma vitrine sobre as atividades e membros que compõem a UAI.

Os requisitos funcionais da OpenUAI foram definidos como: (i) apresentar o perfil dos professores da UAI, com dados do Lattes integrados a dados da UAI; (ii) mostrar informações sobre os cursos ofertados por meio de suas matrizes; (iii) indicar as áreas de atuação dos professores; e (iv) listar projetos de pesquisa, extensão e inovação e seus professores responsáveis.

A OpenUAI é uma aplicação Web dinâmica desenvolvida em Node.js¹² que consome dados por meio do SPARQL endpoint (Seção 3.4), para obter os dados e gerar as telas com a visualização dos dados. A Figura 5 é um exemplo dessa geração dinâmica, que resgata algumas triplas sobre informações de professores da Unidade, como nome, link do Lattes e área de atuação.

Figura 5 – Fragmento da lista de professores com link para seus perfis



Fonte: Autoria própria.

A Figura 6 mostra a matriz do Curso de Sistemas para Internet (CSTSI). A matriz curricular é gerada a partir dos dados sobre as disciplinas do CSTSI. Ela foi montada de forma interativa, em que o usuário, ao mover o mouse sobre as disciplinas, recebe em destaque seus pré-requisitos e pós-requisitos. Ao clicar nas disciplinas, é possível ver suas ementas e carga horária.

Figura 6 – Matriz interativa do Curso Superior de Tecnologia em Sistemas para Internet (CSTSI)

1º Período	2º Período	3º Período	4º Período	5º Período	6º Período
Algoritmo e Programação Estruturada	Bancos de Dados I	Banco de Dados II	Análise e Projeto de Sistema	Comércio Eletrônico	Desenvolvimento Ágil Para Web
Fundamentos da Computação	Ciência Tecnologia e Meio Ambiente	Interação Humano-Computador	Ética e Direitos Humanos	Empreendedorismo	Desenvolvimento e Execução de Projetos de Software
Fundamentos de Redes de Computadores	Estrutura de Dados	Língua Inglesa	Legislação Social	Gerência de Projetos de Software	Inteligência Empresarial
Língua Portuguesa	Fundamentos da Metodologia Científica	Probabilidade e Estatística	Métodos e Técnicas de Pesquisa	Padrões de Projeto de Software	Programação Distribuída
Linguagens de Marcação	Linguagens de Script	Programação Orientada a Objetos	Persistência de Objetos	Programação Para Dispositivos Móveis	Tópicos Especiais
Matemática Aplicada a Sistemas Para Internet	Protocolos e Interconexão de Redes de Computadores	Psicologia do Trabalho	Programação Para Web I	Programação Para Web II	
	Sistemas Operacionais		Segurança de Dados		

Fonte: Autoria própria.

4 O conjunto de dados e a avaliação da aplicação

O conjunto de dados RDF criado é composto de informações relativas a docentes, projetos, áreas de atuação, cursos e disciplinas da UAI. Um pequeno fragmento do conjunto de dados é mostrado na Figura 7.

Figura 7 – Fragmento do conjunto de dados RDF gerado

```
<http://openuai.ifpb.edu.br/ouai#subarea12>
a <http://openuai.ifpb.edu.br/ouai#Subarea> ;
rdfs:subClassOf <http://openuai.ifpb.edu.br/ouai#area1> ;
<http://purl.org/dc/terms/identifier> 10303030 ;
<http://purl.org/dc/terms/title> "Banco de Dados" .

<http://openuai.ifpb.edu.br/ouai#professor10>
a <http://dbpedia.org/ontology/Professor> ;
<http://openuai.ifpb.edu.br/ouai#citationName>
"CAVALCANTI, V. M. B." ;
<http://openuai.ifpb.edu.br/ouai#hasSubarea>
<http://openuai.ifpb.edu.br/ouai#subarea9> ;
<http://openuai.ifpb.edu.br/ouai#isCoordinatorIn>
<http://openuai.ifpb.edu.br/ouai#researchproject34> ;
<http://openuai.ifpb.edu.br/ouai#lattesURL>
"http://lattes.cnpq.br/2868420260808800"^^xsd:anyURI ;
<http://purl.org/dc/terms/identifier> "K477082J4" ;
foaf:name "Valéria Maria Bezerra Cavalcanti" .
```

Fonte: Autoria própria.

Na Figura 7 é apresentada uma descrição de área de conhecimento ("Banco de Dados"), cujo sujeito é `http://openuai.ifpb.edu.br/ouai#subarea12`, e de uma professora, com dados como a url do Lattes e sua associação com um determinado projeto (`http://openuai.ifpb.edu.br/ouai#researchproject34`).

O conjunto de dados está publicado em distribuição RDF (serializado em turtle), sob licença aberta

12 <https://nodejs.org/en/>

de acordo com a especificação “Licença Aberta para Bases de Dados (ODbL) do Open Data Commons”¹³. O conjunto é composto de 3.325 triplas e possui um tamanho de 333 KB.

Para avaliar a utilidade dos dados integrados e da aplicação final de consumo por meio das visualizações apresentadas, foi realizada uma avaliação da OpenUAI. Para isso, discentes e docentes da Instituição, além de pessoas externas, foram convidados a usar a aplicação e responder um formulário disponibilizado na Web 14.

Participou da avaliação um total de 37 pessoas. Desse montante, 81,1% são estudantes, 8,1% são professores do IFPB, enquanto os outros 10,8% são pessoas que não possuem vínculo com a instituição. Trata-se de uma amostra da população relativa aos potenciais usuários da OpenUAI.

As questões realizadas disseram respeito às visualizações dos dados, bem como a aspectos de usabilidade. As respostas foram organizadas por meio da escala Likert (1 a 5, em que 1 significa discordância total e 5 significa concordância total com a afirmação) (LIKERT, 1932). Também foram solicitadas respostas dissertativas para, assim, o público poder dar sugestões e feedbacks mais precisos e pessoais.

Em relação às visões de dados apresentadas, um dos pontos-alvo da avaliação, como ilustração, a Figura 8 mostra que as informações são facilmente compreendidas pelo público. A Figura 9 apresenta o nível de satisfação das informações quanto à sua utilidade. Percebe-se uma resposta positiva do público quanto à aplicação e às visões produzidas.

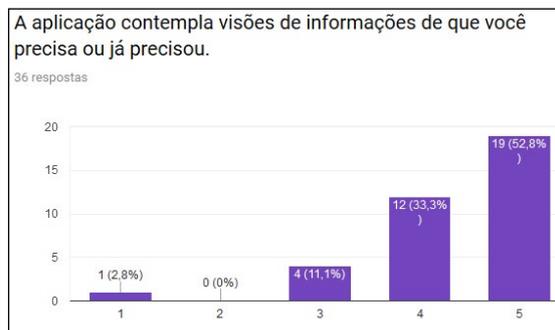
Figura 8 – Facilidade de entendimento das informações



13 <http://opendefinition.org/licenses/odc-odbl/>

14 Formulário disponível em <https://goo.gl/forms/FlnK4Q8jtldz4E9u1>

Figura 9 – Utilidade das informações



Fonte: Autoria própria.

Por meio dos resultados obtidos pelas avaliações, foi possível melhorar a experiência para o usuário, já que algumas das respostas indicaram que existia uma carência de aperfeiçoamento no design responsivo da aplicação para a navegação em dispositivos móveis. Também foram sugeridas otimizações na navegação e usabilidade da aplicação como, por exemplo, a fixação de cabeçalho com opções na parte superior de todas as páginas.

5 Trabalhos relacionados

No contexto nacional, existem algumas aplicações que fazem uso de dados abertos conectados para prover algum serviço à população, como são os casos da OpenIn (ROCHA; LÓSCIO, 2015) e do OpenSBB (BATISTA; LÓSCIO, 2013). O primeiro é um trabalho semelhante ao nosso, tendo como resultado uma aplicação para facilitar o acesso às informações referentes aos docentes do Centro Acadêmico de Informática da UFPE. Este trabalho também produziu uma ontologia, denominada de OpenInOnto, e um conjunto de dados chamado de OpenInData. Já o segundo é um trabalho que envolve informações referentes a edições do Simpósio Brasileiro de Bancos de Dados (SBB), que também incluiu a construção de uma ontologia, chamada de SBCEVENT, o desenvolvimento de um conjunto de dados em RDF e uma aplicação para visualização dessas informações.

Estes trabalhos serviram como inspiração para o desenvolvimento da abordagem da OpenUAI e, em algumas etapas, foi necessário realizar tarefas em comum, como o desenvolvimento de uma ontologia para referência semântica dos dados, como também o desenvolvimento de conjuntos de dados. Todos os trabalhos também desenvolveram aplicações para consumo. Mesmo com todas as semelhanças, esses trabalhos possuem suas diferenças: este trabalho, por

exemplo, teve que lidar com dados provenientes de fontes diferentes e em formatos diferentes, a fim de gerar um conjunto de dados com informações diversificadas, informações que até então se encontravam dispersas. A ontologia gerada por este trabalho compreende termos da área acadêmica e de pesquisa, além de informações provenientes de redes sociais acadêmicas. A OpenCInOnto é composta apenas de termos do ambiente acadêmico em questão. Já com relação à SBCEVENT, seu domínio é completamente diferente e diz respeito a eventos.

Além destes, o portal *data.open.ac.uk* (ENRICO *et al.*, 2015) reúne dados abertos de vários repositórios da *Open University (OU)*. Ele disponibiliza conjuntos de dados abertos e conectados sobre publicações, qualificações, cursos e recursos educacionais, além de resultados de pesquisas da OU. Os recursos publicados seguem os princípios das cinco estrelas dos dados abertos (de dados conectados), como, por exemplo, disponibilização dos conjuntos em modelo RDF e uso de URLs para identificar recursos (BERNERS-LEE; HENDLER; LASSILIA, 2001). Além disso, para o acesso aos dados, consumo destes e descoberta de outros conteúdos relevantes, o portal oferece um *SPARQL endpoint*.

Outro trabalho relacionado é o HAL (RAFES; GERMAIN, 2015), um repositório de documentos de pesquisas científicas, criado para o fim de gestão do conhecimento da comunidade científica, permitindo o compartilhamento de dados abertos conectados e facilitando a interoperabilidade entre diferentes ontologias. Nesse trabalho foi usado o *Data Catalog Vocabulary*¹⁵ (DCAT), como o vocabulário de referência semântica para descrição dos recursos encontrados no portal. Para isso, foram usadas as palavras-chave de cada documento para a sua descrição.

6 Conclusões e trabalhos futuros

Este trabalho gerou resultados importantes na área de Web Semântica, por tratar de escopo acadêmico e de pesquisa. Um deles diz respeito à ontologia OUIAI, capaz de representar informações acadêmicas, de pesquisa e de perfis profissionais, levando em consideração informações voltadas para a realidade de instituições acadêmicas brasileiras. O seu uso permite a referência semântica em dados de instituições de ensino de todo país. Esta tarefa é

considerada árdua usando outras ontologias, devido à ausência de termos próprios do contexto nacional. Um exemplo disso é a Plataforma Lattes, em que boa parte dos dados é específica ao contexto brasileiro.

O processo de publicação e consumo de dados definido e instanciado atende às boas práticas do *World Wide Web Consortium (W3C)* (LÓSCIO; BURLE; CALEGARI, 2017) e pode ser utilizado em outros escopos de dados. Ele ajuda a integrar dados de diversas fontes e produz como resultado conjuntos de dados que podem ser acessados via *web service*.

A aplicação OpenUAI mostra que é possível consumir os dados abertos conectados e tornar o que seriam dados brutos em visões úteis e interativas. O desenvolvimento da aplicação é também uma forma de aproximar a Instituição (UAI) da comunidade interna e externa. Sua disponibilização provê meios para que mais pessoas tenham acesso às atividades, cursos, projetos, membros e áreas que compõem a UAI.

Como trabalho futuro, é necessário definir políticas de atualização dos dados, bem como organizar mais conjuntos de dados abertos por meio de um catálogo ou portal. Outra tarefa planejada diz respeito à expansão das informações contidas no conjunto. Entre elas, informações sobre projetos realizados nas disciplinas dos cursos serão adicionadas.

REFERÊNCIAS

- ALENCAR, A. *et al.* Publicando e Consumindo um Conjunto de Dados Abertos Conectados da UAI. In: SIMPÓSIO BRASILEIRO DE BANCO DE DADOS, 32.. 2017, Uberlândia. **Anais...** Porto Alegre: SBC, 2017. p. 267-277.
- BATISTA, M, G. R.; LÓSCIO, B, F. OpenSBBDD: Usando Linked Data para Publicação de Dados Abertos sobre o SBBDD. In: SIMPÓSIO BRASILEIRO DE BANCO DE DADOS, 28., 2013, Recife. **Anais...** Recife: UFPE, 2013. p. 10:1-10-6.
- BERNERS-LEE T.; HENDLER J.; LASSILIA O. The semantic web. **Scientific American**, v. 284, n. 5, p. 34-44, maio 2001.
- ENRICO, D. *et al.* The Open University Linked Data - data.open.ac.uk. **Semantic Web Journal**, v. 7, n. 2, p. 183-191, 2015.
- GRUBER, T. Ontology. In: Liu, L.; Ozsu, M. T. **Encyclopedia of Database Systems**. Springer, Berlin, Heidelberg, 2009.

¹⁵ <https://www.w3.org/TR/vocab-dcat/>

CYGANIAK, R.; WOOD, D.; LANTHALER, M. **RDF 1.1 Concepts and Abstract Syntax**, 2014. Disponível em: <<https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>>. Acesso em: 21 fev. 2018.

HEATH, T.; BIZER, C. **Linked Data - Evolving the Web into a Global Data Space**. Ed. Morgan & Claypool. Canadá, 2011.

ISOTANI, S.; BITTENCOURT, I. **Dados Abertos Conectados: Em busca da Web do Conhecimento**. Editora Novatec, 2015.

LIKERT R. A technique for the measurements of attitudes. **Archives of psychology**, v. 22, p. 5-55, 1932.

LÓSCIO, B. F.; BURLE, C.; CALEGARI, N. **Data on the Web Best Practices**, 2017. Disponível em: <<http://w3c.github.io/dwbp/bp.html>>. Acesso em: 11 fev. 2018.

OLIVEIRA, M. *et al.* Enabling a Unified View of Open Data Catalogs. In: Proceedings of the 18th International Conference on Enterprise Information Systems - Volume 2: ICEIS, 2016.

OPENLINK. **About OpenLink Virtuoso**, 2015. Disponível em: <<https://virtuoso.openlinksw.com/>>. Acesso em: 10 fev. 2018.

RAFES, K., GERMAIN, C. **A platform for scientific data sharing**. BDA2015 - Bases de Données Avancées, Sep. 2015.

ROCHA, J.; LÓSCIO, B. F. **OpenCIn: Usando Dados Abertos e Conectados para a Publicação de dados sobre o CIn/UFPE**. In: CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO. 35., 2015. Recife: UFPE, 2015.

RUSSO, V. Semantic Web: Metadata, Linked Data, Open Data. Science & Philosophy. **Journal of Epistemology, Science and Philosophy**, v. 3, n. 2, p. 37-46, dez. 2015.

SURE, Y.; STAAB, S.; STUDER, R. Ontology Engineering Methodology. In: **Handbook on Ontologies**. Springer-Verlag, Berlin, Heidelberg, 2009.