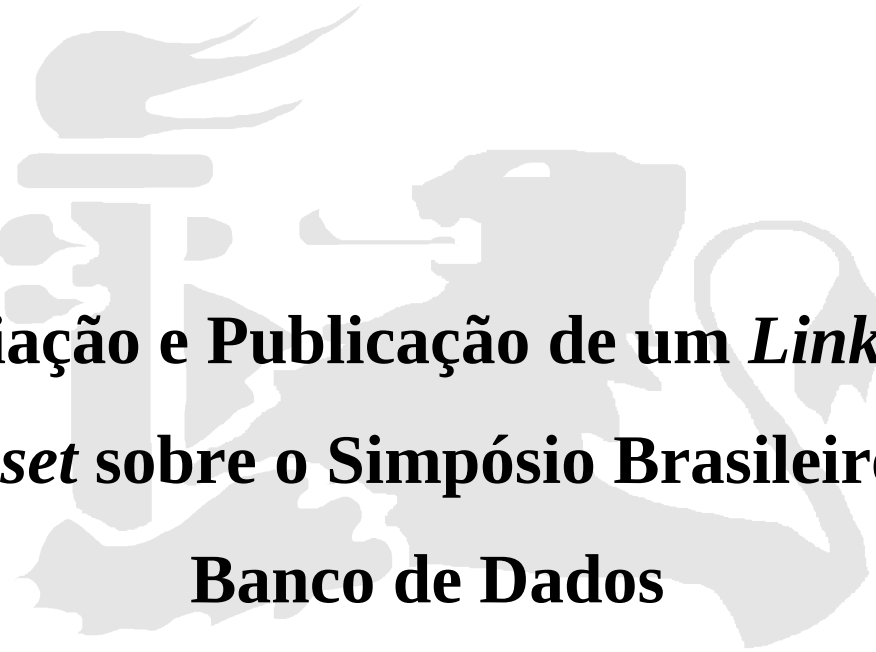
Universidade Federal de Pernambuco

GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

CENTRO DE INFORMÁTICA



Criação e Publicação de um *Linked Dataset* sobre o Simpósio Brasileiro de Banco de Dados

Aluno: Mateus Gondim Romão Batista {mgrb@cin.ufpe.br}

Orientadora: Dra. Bernadette Farias Lóscio {bfl@cin.ufpe.br}

Abril, 2013

Universidade Federal de Pernambuco

CENTRO DE INFORMÁTICA

Mateus Gondim Romão Batista

Criação e Publicação de um *Linked Dataset* sobre o Simpósio Brasileiro de Banco de Dados

Trabalho apresentado à disciplina de Trabalho de Graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Bacharel em Ciência da Computação.

Orientadora:*Profa. Dra. Bernadette Farias Lóscio*

Recife, Abril de 2013.

Dedico este trabalho à minha família.

Agradecimentos

Agradeço a toda minha família por ter me acompanhado e me apoiado durante todos estes anos. Em especial, agradeço à minha mãe, que sempre se sacrificou para ter condições de pagar uma educação de qualidade para mim e meus irmãos.

Agradeço a todos os professores que tive durante toda a minha vida. Professores da escolinha Tia Beth, IESE, Colégio Santa Emília, também fazem parte desta conquista, e contribuíram para a minha formação.

Agradeço ao Centro de Informática (CIn) da UFPE. Neste centro, aprendi a aprender, e ganhei uma confiança que carregarei para toda a vida: a confiança de que podemos contribuir de forma relevante na sociedade. Em especial, agradeço ao PET pelas lições aprendidas sobre espírito de equipe e responsabilidade social.

Agradeço à Profª Berna, que me orientou e me apoiou durante todo o decorrer deste trabalho. Suas orientações foram muito importantes e me ensinaram bastante sobre trabalhos científicos.

Agradeço à minha namorada Vanessa, que sempre está do meu lado.

Agradeço aos meus amigos, pela amizade verdadeira e momentos de descontração. Muitas destas amizades nasceram aqui no CIn, e tenho certeza que irão muito além da graduação.

Resumo

A Web Semântica foi projetada para expandir a Web que conhecemos atualmente, possibilitando que a imensa quantidade de dados disponíveis na Web possa ser compreendida não só por pessoas, mas também por máquinas. Neste processo, a Web Semântica agregou novas tecnologias, como RDF e OWL, que permitem a modelagem e descrição de dados com definições semânticas associadas.

Em conjunto com a adoção dessas tecnologias, surgiu o conceito de *Linked Data*, um conjunto de princípios e técnicas criado com o objetivo de possibilitar a interligação de dados de fontes de dados distintas. Desta forma, cria-se uma grande rede de dados de diferentes domínios, permitindo que aplicações tenham acesso a um maior volume de dados, e possam utilizá-los de forma mais simples e eficiente. O movimento *Linked Data* tem ganhado destaque entre a comunidade acadêmica nos últimos anos, com uma crescente adoção de seus princípios.

O objetivo deste trabalho é criar um conjunto de dados, modelado em RDF, sobre os metadados das edições passadas do Simpósio Brasileiro de Banco de Dados. Além disto, um *SPARQL Endpoint* foi disponibilizado para a execução de consultas SPARQL, e uma aplicação Web foi desenvolvida com o intuito de oferecer visões mais amigáveis destes dados, a partir de gráficos e tabelas. Além do conjunto de dados ter flexibilidade para ser reusado por outros trabalhos, ele terá seus dados ligados a outras fontes de dados, aprimorando sua integração com dados já publicados na Web Semântica.

Palavras-chave: Web Semântica, *Linked Data*.

Abstract

The Semantic Web was designed to expand the current Web we know, enabling the large amount of data available on the Web to be processed not only by humans, but also by machines. Along this process, the Semantic Web gathered new technologies, such as RDF and OWL. These technologies allow the modeling and description of data embebbed with semantic definitions.

In addition to the adoption of these technologies, the concept of Linked Data emerged. Linked Data is a set of principles and techniques created with the purpose of interlinking data from distinct data sources. With a large data network linking data from different domains, applications can access a wider range of data and use them in an easier and more efficient way. The Linked Data Movement has gained prominence in the academia over the last years, with an increasing adoption of its principles.

The goal of this work is to create a dataset, descripted in RDF, with metadata about the past editions of the Brazilian Symposium on Databases. Furthermore, a SPARQL Endpoint was made available to be queried using SPARQL, and a Web application was developed with the purpose of providing more user-friendly views about this dataset, through charts and tables. Moreover, the dataset has flexibility to be reused in other future work, and is interlinked with other datasets, improving its integration with data already available on the Semantic Web.

Keywords: Semantic Web, Linked Data.

Lista de Figuras

Lista de Tabelas

Sumário

Capítulo 1 - Introdução

Neste capítulo, uma breve introdução do trabalho será feita, mostrando qual foi a motivação para o seu desenvolvimento, quais os seus objetivos e contribuições e como o documento está estruturado.

1.1 Motivação

Desde sua criação, a Web passa por um processo contínuo de transformação. Nos seus primeiros anos, usuários podiam visualizar páginas Web e navegar entre as mesmas através de *hyperlinks*, mas não tinham a possibilidade de contribuir com o conteúdo das respectivas páginas. A interação dos usuários era, portanto, limitada à visualização passiva de conteúdo.

Com a chegada da Web 2.0, usuários passaram a ter a experiência de interagir e colaborar entre si por meio de um diálogo em mídia social, como criadores de conteúdo em uma comunidade virtual [30]. Consequentemente, o volume de dados disponíveis na Web passou a crescer em quantidade exponencial [11]. Aspectos como interatividade, colaboração e padronização ganharam destaque. Além das motivações diretamente relacionadas à experiência dos usuários, a quantidade de aplicações desenvolvidas para lidar de alguma forma com os dados que circulam nesse ambiente Web cresceu consideravelmente. No entanto, ainda existem sérios problemas em lidar com esses dados por uma série de motivos, como formato, mas principalmente pela ausência de semântica entre os dados.

Neste contexto, surgiu o conceito de Web Semântica. Segundo [23], a Web Semântica é a extensão da *World Wide Web* que permite às pessoas compartilharem conteúdo além dos limites de aplicações e *websites*. Encorajando a inclusão de conteúdo semântico em páginas Web, a Web Semântica visa converter a Web atual dominada por documentos estruturados e semi-estruturados em uma "Web de dados" [29]. A implementação deste conceito possibilita que aplicações tirem mais proveito dos dados disponibilizados, ao dotá-los de mais semântica.

Para tornar a Web Semântica uma realidade, uma série de novas tecnologias foram adotadas, como *Resource Description Framework* (RDF), *Web Ontology Language* (OWL), Extensible Markup Language (XML), entre outros. Conectado com estes novos atores da Web,

surgiu um conjunto de princípios e tecnologias chamado *Linked Data*. De uma maneira geral, *Linked Data* refere-se a empregar o *Resource Description Framework* (RDF) e o *Hypertext Transfer Protocol* (HTTP) para publicar dados estruturados na Web, além da interligação de dados de diferentes fontes de dados, efetivamente permitindo que dados em uma fonte sejam ligados a dados em outra fonte de dados [9]. Este conceito tem ganhado crescente destaque e adesão da comunidade acadêmica e, por isso, já existem grandes conjuntos de dados na Web implementando o mesmo. Alguns exemplos destes *datasets* são a *DBPedia*[1], *Freebase*[2] e *MusicBrainz*[3]. A *DBPedia* extrai dados da Wikipédia e já tem cerca de 2,6 milhões de conceitos descritos em mais de 247 milhões de registros. O *Freebase*, por sua vez, extrai dados de outras fontes além da Wikipédia, como IMDB, Flickr, entre outros. Já o *MusicBrainz* armazena metadados sobre música em geral.

Como exemplo do uso de Linked Data, na World Wide Web Conference 2012 - WWW2012, houve uma iniciativa com o objetivo de traduzir dados e metadados sobre a conferência. Especificamente, desejava-se representar informações sobre artigos aceitos, autores, comitês, programação e locais de vários formatos para RDF. Com isso, aplicações existentes ou futuras que têm interesse em dados sobre conferências poderiam integrar essa porção extra de conhecimento para aprimorar seu serviço, aperfeiçoar seu *mashup* e melhorar suas estatísticas [31].

De maneira semelhante, este trabalho aborda o problema de criação de uma base de metadados sobre o Simpósio Brasileiro de Banco de Dados (SBBD). Promovido anualmente pela Comissão Especial de Banco de Dados da Sociedade Brasileira de Computação (SBC), o SBBD reúne pesquisadores, estudantes e profissionais do Brasil e do exterior, que apresentam e discutem temas relacionados aos últimos avanços da área. Tradicionalmente, este evento tem reunido em torno de 500 participantes e é o maior evento na América Latina para a apresentação e discussão de resultados de pesquisa relacionados à área de banco de dados. Além de sessões técnicas e tutoriais, o simpósio também inclui tutoriais e palestras convidadas apresentadas por pesquisadores de renome da comunidade nacional e internacional [25].

[1] http://dbpedia.org/About
[2] http://www.freebase.com/
[3] http://musicbrainz.org/

Com este trabalho, a comunidade acadêmica passa a ter disponível uma base de dados interligados sobre um evento acadêmico de renome, a qual pode gerar uma série de trabalhos futuros envolvendo o SBDD. O volume de dados sobre as mais de 25 edições deste simpósio possibilita análises sobre o comportamento de vários fatores como, por exemplo, o número de publicações de uma certa instituição varia com o tempo. Com isto, padrões e tendências neste domínio podem ser descobertos de forma mais eficiente.

1.2 Objetivos e Contribuições

Este trabalho tem como objetivo a criação de um *dataset* com dados relevantes sobre todas as edições passadas do Simpósio Brasileiro de Banco de Dados (SBBD) e sua publicação através de um SPARQL *Endpoint*, bem como o desenvolvimento de uma aplicação para a visualização dos respectivos dados. O conteúdo abrange desde dados sobre a estrutura e organização de cada evento, como comissões e palestras, até dados dos artigos apresentados, autores e suas respectivas afiliações. Utilizando os princípios de *Linked Data*, os dados são ligados com outros *datasets* já estabelecidos. Além disso, vários vocabulários existentes e amplamente conhecidos são reusados.

Como principais contribuições deste trabalho, destacam-se:

- A criação de um conjunto de dados seguindo os princípios de *Linked Data*, descrito em RDF, com alto potencial para ser reutilizado por futuros trabalhos e aplicações envolvendo dados e metadados sobre o SBBD.
- A criação de uma ontologia com novos termos sobre o domínio de conferências.
- Disponibilização de um SPARQL *Endpoint* para a realização de consultas SPARQL sobre o *dataset* criado.
- Visões históricas dos dados do SBBD em formatos amigáveis, como gráficos e tabelas.
- Apoio na disseminação do movimento *Linked Data* na comunidade acadêmica.

O conjunto de dados interligados sobre o SBBD apresenta um alto grau de flexibilidade e reusabilidade, podendo ser utilizado por demais aplicações interessadas em extrair dados

históricos do evento. Além disso, com a aplicação para visualização dos dados do *dataset*, é possível ter um panorama histórico mais elaborado do evento.

1.3 Estrutura do documento

Os capítulos seguintes estão estruturados da seguinte forma: O Capítulo 2 aborda a contextualização deste trabalho, apresentando os conceitos de Web Semântica e *Linked Data*, além das principais tecnologias e técnicas que os envolvem. O Capítulo 3, por sua vez, descreve todo o trabalho realizado, desde a definição do esquema do *dataset* e sua criação, até a disponibilização de um SPARQL Endpoint e o desenvolvimento de uma aplicação para a visualização dos dados. Por fim, o Capítulos 4 expõe a conclusão deste trabalho, além de sugestões para projetos futuros.

Capítulo 2 - Contextualização

Neste capítulo serão abordados conceitos essenciais para o entendimento deste trabalho. Na seção 2.1, apresentamos os conceitos de Web Semântica e *Linked Data*, e seus princípios e componentes. A seção 2.1.1 descreve os conceitos básicos de Web Semântica, assim como suas principais tecnologias, como RDF, OWL e SPARQL. Já na seção 2.1.2, introduzimos os princípios básicos de *Linked Data*. A seção 2.2 descreve as considerações que precisam ser analisadas em um projeto de *Linked Datasets*, além de uma breve descrição dos vocabulários que foram reusados neste trabalho. Por fim, a seção 2.3 discorre sobre a geração de bases RDF a partir de bancos de dados relacionais, bem como as ferramentas existentes com este objetivo.

2.1 Web Semântica

2.1.1 Conceitos gerais

A maioria do conteúdo Web produzido atualmente é projetado para ser interpretado por humanos, mas não por máquinas. Apesar de ser possível desenvolver programas que busquem informação em páginas HTML, a ausência de semântica nestas últimas limita consideravelmente o que computadores podem fazer com os dados disponíveis.

Em 2001, Tim Berners-Lee introduziu os primeiros conceitos da Web Semântica. A Web Semântica não é uma nova Web, mas sim uma extensão da atual, na qual informações ganham um significado bem definido, permitindo que computadores e pessoas trabalhem melhor em cooperação [4]. O objetivo é permitir que comunidades possam colocar dados compreensíveis por maquinas na Web, que podem ser compartilhados e processados tanto por ferramentas automáticas como por pessoas [26].

A Web Semântica trouxe consigo uma série de conceitos auxiliares essenciais para a implementação da mesma, mas também tirou proveito de tecnologias já existentes. Duas dessas tecnologias que merecem destaque são: XML (eXtensible Markup Language) e RDF (Resource Description Framework). O XML é uma arquitetura que não possui elementos e marcadores (*tags*) predefinidas, dando total liberdade ao autor para definir suas próprias *tags*. Esta

característica possibilita melhorias significativas em processos de recuperação e disseminação de informação [2]. Dada a complexidade de modelagem do mundo real, é importante o uso de uma linguagem flexível para retratar a realidade de forma mais aproximada possível. Entretanto, o problema relacionado à ausência de significado dos dados ainda não é completamente resolvido. Para solucionar essa questão, faz-se uso de RDF (abordado com mais detalhes na próxima seção), que permite adicionar semântica aos dados.

Com o emprego da semântica, é possível diferenciar significados distintos para um mesmo termo. Além disso, permite-se que termos localizados em fontes de dados diferentes possam ser associados a um mesmo significado. Para que seja possível identificar essas relações, utilizam-se ontologias, outro componente fundamental da Web Semântica [4]. Ontologias são esquemas de (meta) dados, que fornecem um vocabulário controlado de conceitos, cada qual com uma semântica explicitamente definida e processável por máquinas [14]. São compostas essencialmente por classes e propriedades. Uma classe representa um conceito sobre algo, a qual indivíduos podem pertencer. Já uma propriedade representa a relação entre estes indivíduos e dados, ou outros indivíduos. A partir destes esquemas, é possível reusar seus termos na criação de novos *datasets* e ontologias, o que facilita a ação e adaptação automática de agentes de software que trabalhem ou venham a trabalhar em um momento futuro com estas ontologias.

O W3C (*World Wide Web Consortium*) é uma comunidade internacional fundada com o objetivo de desenvolver padrões para a Web e garantir seu crescimento a longo prazo. É composto por organizações membros, como o Google, Facebook, IBM, Oracle, entre outras. Liderado por Tim Berners-Lee, criador da *World Wide Web*, o W3C tem desempenhado um papel importante no que diz respeito aos padrões da Web Semântica e na construção de um conjunto de tecnologias que possam dar suporte à Web de Dados [28].

2.1.2 RDF

RDF (*Resource Description Framework*) é um modelo de dados que possibilita a definição de sentenças sobre um recurso. Foi proposto com o objetivo de ser uma possível solução para a limitação do XML em descrever semântica [11].

A infraestrutura oferecida pelo RDF permite a interoperabilidade de dados por meio do projeto de mecanismos que oferecem suporte para convenções comuns de semântica, sintaxe e

estrutura. Este modelo não estabelece a semântica da descrição de recursos de comunidades, mas permite que as próprias comunidades possam definir seus elementos [16].

No modelo RDF, um recurso pode ser "qualquer coisa" sobre a qual se quer expressar um significado e que possa ser identificado unicamente por um URI (*Uniform Resource Identifier*) [16]. URI é um identificador único atribuído a cada recurso, permitindo que este último possa ser referenciado em sentenças por esquemas ou *datasets*.

Com o uso de sentenças, é possível relacionar estes recursos com dados ou com outros recursos. Uma sentença é estruturada no formato sujeito + predicado + objeto onde [11]:

- Sujeito: Tem como valor o recurso do qual se quer escrever uma sentença.
- Predicado: Especifica um relacionamento entre sujeito e objeto. O predicado é especificado por meio de propriedades, que são relações binárias, geralmente nomeadas por um verbo e permitem relacionar um recurso a dados ou a outros recursos.
- Objeto: Denomina o dado ou recurso que se relaciona ao sujeito.

2.1.3 RDF *Schema*

Documentos definidos em RDF *Schema* são usados para declarar vocabulários, os conjuntos de tipos de classes e propriedades semânticas de cada comunidade. Dada uma descrição RDF, RDF *Schemas* definem suas propriedades, assim como as demais características ou restrições de valores destas. O mecanismo de *namespace* de XML é utilizado para identificar RDF *Schemas* [16].

Uma descrição de um RDF *Schema* compreensível tanto por máquinas quanto por humanos pode ser acessada através de uma URI dereferenciável. Caso o esquema seja processável por máquinas, é possível que uma aplicação possa extrair a semântica de propriedades definidas no esquema. A estrutura de RDF *Schema* é baseada no modelo RDF. Portanto, mesmo que uma aplicação projetada para lidar com RDF não possua entendimento de um esquema (descrito em RDF *Schema*) particular, ela ainda assim estará apta a traduzir a descrição deste esquema em propriedades e valores correspondentes [16].

2.1.4 OWL

A Web Semântica tem como um de seus principais objetivos a definição de significados explícitos para conceitos do mundo real, tornando a tarefa de integrar e processar automaticamente dados disponíveis na Web mais simples para as máquinas. No contexto prático, a Web Semântica baseia-se na possibilidade de customizar esquemas de *tags*, proveniente do XML, e na forma flexível de representar dados em RDF. É necessário, no entanto, que sobre a camada RDF, haja uma linguagem para criação de ontologias que possa descrever o significado da terminologia utilizada em documentos Web [15].

A *Web Ontology Language* (OWL) foi projetada para o uso de aplicações que necessitam processar o conteúdo de informação ao invés de apenas apresentar informações a humanos. OWL agrega uma melhor interpretabilidade de conteúdo Web para as máquinas do que XML, RDF e RDF *Schema*, provendo vocabulário adicional com uma semântica formal. Assim como estes dois últimos, é uma recomendação do W3C no que diz respeito a padrões para a Web Semântica [15].

Com o intuito de prover aspectos diferentes de uma linguagem de ontologias para comunidades específicas de usuários e programadores, o W3C definiu três sublinguagens de OWL [3]:

- *OWL Full* - É a linguagem completa, incluindo todas as primitivas de OWL. Permite combinar estas primitivas de formas arbitrárias com RDF e RDF *Schema*, possibilitando, por exemplo, alterar o significado de primitivas pré-definidas (RDF ou OWL), aplicando primitivas de uma linguagem em outra. A grande vantagem de OWL *Full* é que ela é totalmente compatível com RDF, tanto em sintaxe quanto em semântica: qualquer documento RDF válido é também um documento OWL *Full* válido.

- *OWL DL* - A fim de alcançar eficiência computacional, OWL *DL* (abreviação de *Description Logic*) restringe a forma como construtores de OWL e RDF podem ser usados. Isto significa que não é permitido aplicar construtores OWL uns sobre os outros, de forma a garantir que a linguagem corresponda a uma lógica de descrição bem estudada. A vantagem desta sublinguagem é permitir um suporte a raciocínio eficiente. A desvantagem, por sua vez, é a perda de compatibilidade com RDF: um documento RDF

geralmente precisa sofrer restrições para ser um documento OWL *DL* válido. Reciprocamente, todo documento OWL *DL* válido também é um documento RDF válido.

- *OWL Lite* - Esta sublinguagem é uma aplicação de restrições a OWL *DL*, tomando um subconjunto de seus construtores. Traz como vantagem o fato de ser uma linguagem tanto simples de compreender (para usuários) como de implementar (para desenvolvedores de ferramentas). A desvantagem certamente é a expressividade restringida.

2.1.5 SPARQL

Como visto anteriormente, RDF é um modelo de dados que permite armazenar informações com conteúdo semântico agregado. Com o seu lançamento, no entanto, surgiu naturalmente o problema de como consultar dados RDF, de forma que aplicações pudessem utilizar esses dados de maneira eficiente. Depois de várias propostas de *design* e implementação, o W3C adotou a linguagem SPARQL como padrão para recuperação de informações em documentos RDF.

SPARQL é uma linguagem e protocolo para consulta em RDF, baseada em casamento de padrões. De forma análoga a SQL, possui uma estrutura SELECT-FROM-WHERE onde [11]:

- SELECT: Especifica uma projeção sobre os dados como a ordem e a quantidade de atributos e/ou instâncias que serão retornados.
- FROM: Declara as fontes que serão consultadas. Esta cláusula é opcional. Quando não especificada, assumimos que a busca será feita em um documento RDF particular.
- WHERE: Determina restrições na consulta. Os registros retornados pela consulta deverão satisfazer as restrições impostas por essa cláusula.

Uma consulta SPARQL consiste basicamente de três partes [18]:

- Casamento de padrões - Inclui características interessantes de casamento de padrões de grafos, como união de padrões, aninhamento, filtragem e restrição de valores, entre outros.
- Modificadores da solução - Permite modificar os valores da saída do casamento de padrões, aplicando operadores clássicos, como *distinct*,*order*, *limit* e *offset.*
- Saída - A saída de uma consulta SPARQL pode ser de diferentes tipos: consultas sim/não, seleção de valores das variáveis que casaram com os padrões, construção de novas triplas a partir desses valores, e descrições sobre consultas de recursos.

2.2 Linked Data

2.2.1 Conceitos gerais

Com o surgimento do RDF, e da identificação única de recursos via URI, surgiu o *Linked Data*, um conjunto de conceitos e técnicas que possibilita que dados de um *dataset* publicado na Web possam ser conectados a dados de outras fontes de dados por meio de links RDF. Um dos principais objetivos do *Linked Data* é estender a Web que conhecemos, onde documentos HTML estão interconectados, para uma Web onde os dados possam estar diretamente ligados, sem necessidade da intervenção de alguma aplicação que faça essa ligação. Esta extensão da Web também é conhecida por Web de Dados.

Em 2006, Tim Berners-Lee esboçou os princípios do *Linked Data*, que provê orientações gerais nas quais editores e publicadores de dados se basearam para começar a tornar a Web de Dados uma realidade [9]. As regras básicas esboçadas foram as seguintes [8]:

1 Usar URIs como nomes para as coisas.
2 Usar URIs HTTP para que as pessoas possam buscar por esses nomes.
3 Quando alguém pesquisar por uma URI, prover informação útil, utilizando os padrões recomendados (RDF, SPARQL).
4 Incluir *links* para outras URIs, para que mais coisas possam ser descobertas.

O exemplo mais visível de adoção e aplicação dos princípios de Linked Data tem sido o projeto *Linking Open Data*, fundado em 2007 e apoiado pelo *W3C Semantic Web Education and Outreach Group* (SWEO). O objetivo do projeto é dar um grande impulso inicial à Web de Dados, identificando *datasets* existentes que estão disponíveis sob licenças abertas, convertendo-os para RDF de acordo com os princípios de Linked Data, e publicando-os na Web [8].

2.2.2 Considerações de Projeto de *Linked Datasets*

Os quatro princípios definidos por Tim Berners-Lee em 2006 serviram como guia para o surgimento de vários projetos de *Linked Datasets.* No processo de publicação de *Linked Data*, no entanto, existem mais aspectos que devem ser levados em conta para que o projeto possa se tornar eficiente e alcançar seus objetivos.

Como visto previamente, URIs identificam unicamente recursos na Web. Dada essa importância, é preciso dedicar tempo na escolha destas URIs. Primeiramente, elas devem ser nomes que outros publicadores de dados possam usar de forma confiável para criar ligações entre as duas fontes de dados. Além disso, é necessário ter infraestrutura técnica para tornar estas URIs dereferenciáveis, ou seja, prover conteúdo quando as URIs forem acessadas [7].

Algumas outras recomendações na escolha de URIs são:

- Utilizar URIs HTTP, pois o esquema “http://” é o único esquema de URIs que é amplamente suportado pelas ferramentas e infraestrutura dos dias atuais.
- Definir URIs em um *namespace* HTTP sob controle, onde se pode implementar o que for necessário para torná-las dereferenciáveis.
- Tentar manter as URIs estáveis e persistentes. Trocar as URIs em um momento posterior irá quebrar todos os *links* existentes.

Um dos elementos mais importantes no que diz respeito à criação de *Linked Data* são os vocabulários. Vocabulários são descrições de domínios gerais ou específicos, materializadas em um conjunto de termos. Utilizando ontologias e seus componentes (classes e propriedades),

definem-se termos que, assim como os outros recursos, possuem uma URI, o que possibilita o seu reuso.

Com o objetivo de tornar a tarefa das aplicações de processar dados o mais simples possível, o reuso de termos de vocabulários mais conhecidos é fortemente recomendado. Publicadores de dados devem criar novos termos apenas caso os vocabulários já existentes não forneçam os termos desejados [8]. Se existe a necessidade de URIs para lugares geográficos, por exemplo, fontes de dados conhecidas como *Geonames* e *DBPedia* podem ser utilizadas. Um dos maiores benefícios do reuso dessas fontes de dados é que suas URIs já estão ligadas com URIs de outras fontes de dados. Consequentemente, utilizando esses *datasets*, um novo *dataset* pode se interligar com uma rede rica e crescente de outras fontes de dados [7].

Vocabulários geralmente descrevem um domínio em particular, por isso é uma prática comum misturar termos de diferentes vocabulários na criação de uma nova ontologia, para descrever todos os conceitos relacionados a um conjunto específico de dados.

Caso realmente seja necessário definir novos termos, pode-se fazê-lo utilizando RDF ou OWL. Como classes e propriedades também são recursos, as regras de escolha de URIs também se aplicam a eles.

Algumas recomendações importantes para a criação de novos termos são [7]:

- Não definir novos vocabulários a partir de um rascunho, mas complementar vocabulários existentes com termos adicionais para representar dados da maneira desejada.
- Utilizar termos de outros vocabulários. Ao prover mapeamentos para termos existentes, o nível de troca de informações na Web de Dados cresce. Exemplos de propriedades comumente utilizadas para realizar mapeamentos são *rdfs:subClassOf* e *rdfs:subPropertyOf*.
- Definir toda informação importante explicitamente.

2.2.3 Vocabulários

Seguindo os princípios e recomendações sobre *Linked Data*, neste trabalho procurou-se ao máximo reutilizar termos de vocabulários existentes e amplamente conhecidos. Entre os vocabulários reusados, alguns são de cunho mais geral, como os seguintes:

- **FOAF** - O projeto *Friend of a Friend* (FOAF) é um dos maiores projetos da Web Semântica. FOAF se tornou um vocabulário padrão largamente aceito para representar redes sociais entre pessoas, e muitos websites de redes sociais o utilizam para produzir perfis de seus usuários na Web Semântica. Com milhões de perfis *online*, FOAF contém termos que descrevem informações pessoais, participações em grupos, conexões sociais, entre outros [12].

- **DUBLIN CORE**- A *Dublin Core Metadata Initiative* (DCMI) é uma organização aberta sem fins lucrativos, que mantém um conjunto de termos de metadados. Criada em 2000, a comunidade da DCMI focou na ideia de que registros de metadados poderiam utilizar os termos da *Dublin Core* com outros vocabulários especializados para satisfazer requisitos particulares de implementação. Nesta mesma época, o W3C trabalhava em um modelo de dados genérico para metadados, que viria a se tornar o RDF. Com o passar do tempo, *Dublin Core* se tornou um dos vocabulários mais populares para uso com RDF, mais recentemente no contexto do movimento *Linked Data* [10]. O *namespace* deste vocabulário é o *dcterms*.

- **EVENT** - A *Event Ontology* é uma ontologia sobre eventos em geral, desenvolvida no *Centre for Digital Music* em *Queen Mary, University of London*. Graças a sua simplicidade e usabilidade, tem sido usada em uma ampla gama de contextos: desde palestras em conferências, até descrições de um show musical, ou acordes tocados em uma música de jazz, festivais, etc [20].

- **GEONAMES** - O *Geonames* é uma ontologia sobre dados geográficos, e já contém em seu *dataset* mais de 10 milhões de nomes geográficos. O *Geonames* está integrando esses tipos de dados, como nomes de lugares em vários idiomas, altitude, população e outros, de várias outras fontes. A base pode ser acessada tanto via interface Web como via Web Services [27]. O seu *namespace* é o *gn.*

Além destes vocabulários, também destacamos alguns outros que são de um domínio mais específico e relacionado a este trabalho. Estes incluem, por exemplo, vocabulários que definem termos sobre trabalhos acadêmicos, eventos de comunidades de pesquisadores, entre outros. Os vocabulários reusados neste trabalho e que têm este perfil foram os seguintes:

- **BIBO** - A *Bibliographic Ontology* (BIBO) descreve termos bibliográficos na Web Semântica em RDF. Pode ser utilizada como uma ontologia de citações, uma ontologia de classificação de documentos, ou simplesmente como uma forma de descobrir qualquer tipo de documento em RDF. Foi inspirada por vários formatos de descrição de metadados de documentos. Ao invés de cobrir todos os tópicos relacionados a dados bibliográficos, descreve os tópicos básicos, e permite que se tire vantagem utilizando-a em conjunto com outras ontologias com vocabulários mais específicos [5].

- **AKT** - A AKT *Reference Ontology* modela o domínio do meio acadêmico, e contém representações de pessoas, conferências, projetos, organizações, publicações, etc. Escrita em OWL, foi desenvolvida durante um período de seis meses por diversos parceiros do projeto AKT. Este último é uma colaboração de pesquisas interdisciplinares (IRCs) financiadas pela *Engineering and Physical Sciences Research Council* (EPSRC) para ajudar a identificar e resolver os problemas de TI do futuro [1].

- **SWC** - A *Semantic Web Conference Ontology* (SWC) é uma ontologia para descrever conferências acadêmicas. Foi projetada inicialmente para dar suporte à *European Semantic Web Conference*(ESWC 2007), e posteriormente estendida para uma série maior de conferências. SWC é principalmente uma convenção sobre como usar classes e propriedades de outras ontologias, principalmente da FOAF e SWRC, mas também de outros vocabulários, como o SIOC e *Dublin Core*. Para unir todos estes termos, usa um conjunto de termos próprios [22].

- **SWRC** - A ontologia SWRC (*Semantic Web for Research Communities*) modela de forma genérica entidades chave em uma típica comunidade de pesquisa e representa um dos primeiros esforços para pôr em prática as tecnologias da Web Semântica no meio

acadêmico. Lançada em OWL em suas versões mais recentes, possui representações de pesquisadores, grupos de pesquisa, suas publicações e atividades assim como suas inter-relações mútuas [26].

2.3 Geração de triplas RDF a partir de banco de dados relacional

Um requisito crítico para a evolução da atual Web de Documentos para uma Web de Dados (e finalmente uma Web Semântica) é a inclusão de vastas quantidades de dados armazenadas em bancos de dados relacionais [21]. Considerando a quantidade enorme de dados que são armazenados em modelos relacionais atualmente, a possibilidade dessa conversão traz grandes benefícios no que diz respeito ao aproveitamento de dados existentes para a criação de *datasets* RDF. Com essa ação, a tarefa de publicar dados nos padrões da Web Semântica e Linked Data torna-se mais simples, principalmente para aqueles interessados em fazê-lo reaproveitando dados relacionais.

O mapeamento entre estes dois modelos de dados tem sido alvo de um amplo corpo de pesquisa em diversos domínios, e tem levado à implementação tanto de ferramentas genéricas de mapeamento, como de aplicações de domínio específico. Além disso, o papel de RDF como uma plataforma de integração para dados de múltiplas fontes, principalmente as relacionais, é uma das principais motivações que tem incentivado pesquisas sobre mapeamentos de dados relacionais para RDF [21].

Atualmente, existem ferramentas que realizam a conversão entre estes formatos de dados. Entre as mais difundidas na comunidade de publicadores de dados na Web de Dados, podemos destacar [21]:

- **Virtuoso RDF *View*** - O Virtuoso RDF *View* utiliza uma estratégia baseada em conversão de tabela para classe e coluna para predicado, e leva em consideração casos especiais, como quando uma coluna é parte de uma chave primária ou estrangeira. O relacionamento de chave estrangeira entre tabelas é tornado explícito entre as classes

relevantes que representam as tabelas. Os dados relacionais são representados como grafos RDF virtuais, sem criação física dos *datasets* RDF. Através de uma linguagem de meta-esquemas do Virtuoso, mapeamentos são definidos a partir de um conjunto de colunas para triplas.

- ***Triplify*** - Utiliza uma abordagem simplista para publicar RDF e *Linked Data* a partir de bancos de dados relacionais. *Triplify* é baseado no mapeamento de requisições de URIs HTTP em consultas a bases relacionais, expressas em SQL, com algumas funcionalidades adicionais. Esta ferramenta transforma as relações resultantes da consulta em RDF, e publica os dados na Web em várias serializações RDF, em particular como *Linked Data*. É um componente de software que pode ser facilmente integrado com várias aplicações Web existentes, porém não suporta SPARQL.

- **R2O** - É uma linguagem declarativa baseada em XML projetada para expressar mapeamentos entre conceitos de bancos relacionais e uma ontologia. Mapeamentos R2O podem ser utilizados para detectar inconsistências e ambiguidades em definições de mapeamentos. Através de uma ferramenta chamada ODEMapster, documentos R2O são utilizados para duas ações: traduzir os resultados de uma consulta, ou criar um dump RDF com todos os indivíduos do banco de dados. É independente do sistema de gerenciamento de banco de dados.

- **D2RQ** - A plataforma D2RQ fornece um ambiente integrado com múltiplas opções para acessar dados relacionais como grafos RDF virtuais de leitura (*read-only*). Os métodos de acesso possíveis são [6]:
    - Dumps RDF - Nos formatos *RDF/XML* ou *N-Triples*
    - APIs RDF - D2RQ pode ser incorporado em aplicações Java para prover acesso a dados relacionais a partir de APIs Jena e Sesame. Chamadas às APIs são reescritas em comandos SQL e executados no banco de dados.
    - SPARQL *Endpoint* - O D2R *Server* provê acesso remoto a uma base mapeada pelo D2RQ a partir do protocolo SPARQL.

- *Linked Data* - Descrições RDF dos recursos armazenados no banco podem ser acessados ao dereferenciar suas URIs.
- HTML *View* - D2R *Server* fornece uma página HTML simples para auxiliar no processo de escrita e *debugging* dos mapeamentos.

A arquitetura da plataforma D2RQ pode ser visualizada na Figura 1. Os mapeamentos entre o esquema relacional e os vocabulários RDFS ou ontologia OWL são expressos em uma linguagem declarativa de mapeamentos.

O D2RQ foi escolhido para ser utilizado neste trabalho por uma série de fatores, dentre os quais podemos destacar: flexibilidade da linguagem de mapeamentos, comandos simples e geração de dumps RDF, tornando possível reutilizar este *dataset* em conjunto com outras ferramentas de publicação de dados na Web Semântica, além desta plataforma. Como será mostrado com mais detalhes no próximo capítulo, esta ferramenta teve um impacto positivo relevante no povoamento do *dataset* criado.

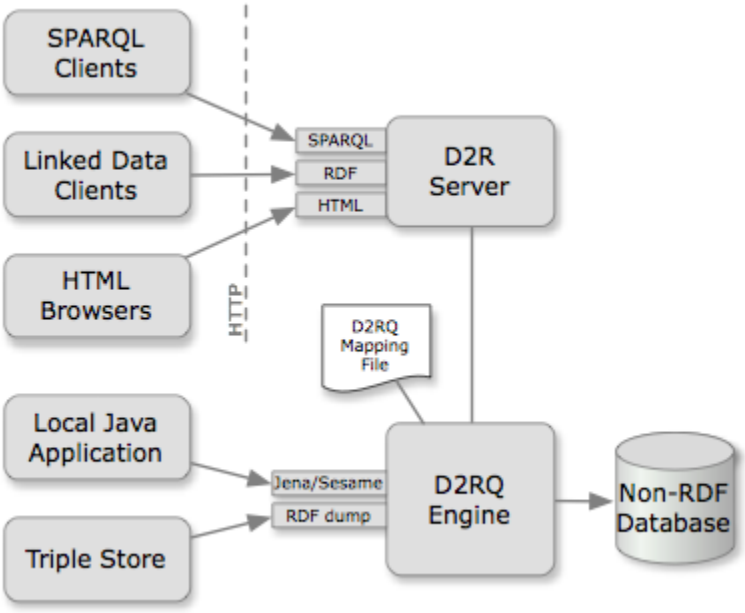


Figura 1. Arquitetura da plataforma D2RQ

2.4 Considerações

Neste capítulo, foram abordados os dois conceitos principais que envolvem este projeto: Web Semântica e *Linked Data*. Sobre o primeiro, suas principais tecnologias foram introduzidas. Todas elas (RDF, RDF *Schema*, OWL e SPARQL) são utilizadas neste trabalho. No que diz respeito a *Linked Data*, foram mostrados que fatores devem ser analisados no projeto de um *Linked Dataset*. Estas considerações refletiram neste trabalho, uma vez que este último inclui a criação de um conjunto de dados interligado. Além disso, os vocabulários reusados neste trabalho foram brevemente apresentados.

A técnica de conversão de dados relacionais para triplas RDF recebeu destaque, pois como será detalhado no próximo capítulo, a principal fonte de dados para a criação do *dataset* foi um banco de dados relacional.

Capítulo 3 - Criação e Publicação do *dataset* SBBD

3.1 Visão geral

Este trabalho dividiu-se essencialmente em 4 fases. A primeira etapa foi dedicada à criação da ontologia que serviu de base para o *dataset*. Neste processo, através de uma análise de quais dados do domínio são relevantes para criar um *dataset* interessante, foi definido quais termos de vocabulários conhecidos seriam reaproveitados. Para conceitos cujas representações não foram contempladas por nenhum destes vocabulários, novos termos foram criados. Na segunda fase, foi criado um *dataset* com os dados do SBBD. A extração deu-se em parte de forma semiautomática, convertendo dados de uma base relacional para RDF, e em parte de forma manual, consultando os *websites* e anais dos eventos passados. Na terceira fase, um SPARQL *Endpoint* foi criado para a realização de consultas sobre o conjunto de dados. Por fim, na quarta fase do projeto, uma aplicação para a visualização de dados através de gráficos foi criada.

3.2 Criação do esquema do *dataset*

O primeiro passo para a criação do esquema foi avaliar quais conceitos, dados e metadados referentes ao SBBD são importantes para uma modelagem coerente do domínio. Além disso, analisou-se quais destes elementos podem fornecer conteúdo para que aplicações possam extrair informações relevantes do *dataset*.

Neste contexto, os conceitos deste domínio selecionados para integrar o esquema da ontologia foram os seguintes:

- **Pessoa -** Pessoas podem participar de simpósios atuando em diferentes papéis. Dentre estas ocupações, elas podem ser autoras/co-autoras de *papers*, membros ou coordenadores de comitês, palestrantes, entre outros.

- **Organização -** Seja qual for a forma de participação de uma pessoa no SBBD, ela sempre estará afiliada a alguma organização. Estas organizações podem ser de cunho acadêmico, como universidades e institutos de tecnologia, mas também podem ser empresas ou até mesmo órgãos do governo, como já foi o caso do Ministério de Ciência e Tecnologia, por exemplo. É importante destacar que pessoas podem ter afiliações diferentes em edições diferentes do simpósio.

- **Conferência -** Representam o simpósio em si. São os eventos de maior escala nesta ontologia, que possuem subeventos como sessões e workshops, e mantém algum tipo de relacionamento com a maioria dos conceitos deste esquema.

- **Papers -** São os artigos apresentados no simpósio e em seus eventos satélites. Podem ser artigos completos (*full papers*), artigos resumidos (*short papers*), teses ou dissertações. Não são atrelados diretamente a uma organização, mas sim a um ou mais autores, que por sua vez possuem uma afiliação.

- **Eventos satélites -** São eventos de menor porte, realizados em conjunto com o SBBD. Os mais comuns são o Workshop de Teses e Dissertações e a Sessão de Demos. Apesar destes eventos possuírem seus próprios comitês, também são considerados parte do domínio do simpósio.

- **Comitês -** Um comitê é um grupo de pessoas responsável por alguma tarefa específica na organização de um evento. Possuem um ou mais membros e um coordenador. Neste trabalho, os tipos de comitês descritos são dois: Comitê Diretivo e Comitê de Programa.

  O comitê diretivo funciona como a "direção" do SBBD. É composto por 6 professores da comunidade que são responsáveis por definir as regras que norteiam o simpósio. Tradicionalmente, o comitê diretivo é formado por professores mais experientes da comunidade, os quais se revezam com o passar dos anos. Uma das funções do comitê diretivo é definir os palestrantes do SBBD, por exemplo. Apenas os simpósios possuem um comitê diretivo.

O comitê de programa, por sua vez, é formado por um número maior de professores, que podem ser experientes ou menos experientes. Os professores do comitê de programa são responsáveis por avaliar os artigos que são submetidos ao simpósio. Cada evento satélite também tem seu próprio comitê de programa. O coordenador do comitê de programa é responsável por gerenciar as atividades de submissão e avaliação dos artigos, bem como é responsável pela comunicação com os demais membros do comitê de programa.

- **Sessões técnicas -** São as sessões em que os artigos são apresentados. Geralmente agrupam trabalhos que compartilham um tema em comum. Fazem parte tanto da programação do simpósio como dos eventos satélites.

- **Palestras, Minicursos e Tutoriais -** São eventos distintos, mas que possuem a mesma estrutura. São ministrados por uma ou mais pessoas em um simpósio específico.

3.2.1 Reuso de vocabulários

Após a definição dos conceitos incluídos no esquema da ontologia, buscou-se analisar vocabulários conhecidos e estáveis com o objetivo de reaproveitar o maior número de termos possível. Além da procura de classes existentes com definições semânticas equivalentes a estes conceitos, houve também uma pesquisa por propriedades que representem bem as relações entre estas classes.

Consultando tanto vocabulários de caráter geral como vocabulários com um foco específico no mundo acadêmico (eventos, publicações, etc.), um total de 13 classes foram reutilizadas. A Tabela 1 mostra quais termos foram reusados (prefixados com o *namespace* do respectivo vocabulário).

Conceito	**Termo reusado**
Pessoa	*foaf:Person*
Organização	*akt:Organization*
Empresa	*akt:Company* (subclasse de *akt:Organization*)
Organização focada em aprendizado	*akt:Learning-Centred Organization* (subclasse de *akt:Organization*)
Paper	*bibo:Article*
Conferência/Simpósio	*bibo:Conference*
Comitê de Programa	*swc:ProgrammeCommitteeMember*
Workshop (Evento satélite)	*swc:WorkshopEvent*
Chair do Simpósio	*swc:Chair*
Sessão técnica	*swc:SessionEvent*
Palestra	*swc:TalkEvent*
Tutorial	*swc:TutorialEvent*
Minicurso	*swrc:Lecture*

Tabela 1. Termos reusados para classes

Além destas classes, várias propriedades também foram reutilizadas nesse trabalho Algumas delas têm, em sua ontologia original, seus *domain* e *range* definidos com classes mais genéricas, mas para facilitar o entendimento da aplicação destas propriedades no esquema deste trabalho, listaremos como *domain* e *range* as classes presentes neste esquema que se conectam com a propriedades em questão. As *Datatype properties* estão listadas na Tabela 2, enquanto as *Object properties* estão listadas na Tabela 3.

Termo reusado	***Domain***	***Range***
foaf:name	*foaf:Person, sbbd:Committee*	*String*
foaf:gender	*foaf:Person*	*String*
dcterms:title	*bibo:Article, bibo:Conference, swc:TalkEvent, swc:TutorialEvent, swrc:Lecture, swc:SessionEvent, swc:WorkshopEvent, sbbd:DemoSession*	*String*
dcterms:type	*bibo:Article*	*String*
dcterms:date	*bibo:Conference*	*Date*
swrc:startDate	*bibo:Conference*	*Date*
swrc:endDate	*bibo:Conference*	*Date*

Tabela 2. Termos reusados para Datatype Properties

Termo reusado	***Domain***	***Range***
foaf:based_near	*foaf:Person*	*gn:Feature*
dcterms:language	*bibo:Article*	*dcterms:LinguisticSystem*
event:place	*bibo:Conference*	*gn:Feature*
dcterms:creator	*bibo:Article*	*foaf:Person*
bibo:presented_at	*bibo:Article*	*bibo:Conference*
swc:isSubEventOf	*swc:TalkEvent, swc:TutorialEvent, swrc:Lecture, swc:SessionEvent, swc:WorkshopEvent, sbbd:DemoSession*	*bibo:Conference*
swc:hasRelatedDocument	*swc:SessioEvent*	*bibo:Article*

akt:has-speaker	*swc:TalkEvent, swc:TutorialEvent, swrc:Lecture*	*foaf:Person*
swc:heldBy	*swc:Chair, swc:ProgrammeCommitteeMember, sbbd:SteeringCommitteeMember*	*foaf:Person*
swc:isRoleAt	*swc:Chair, swc:ProgrammeCommitteeMember, sbbd:SteeringCommitteeMember*	*bibo:Conference*

Tabela 3. Termos reusados para Object Properties

3.2.2 Criação de novos termos

Apesar dos vocabulários existentes cobrirem uma ampla parte dos termos necessários para a criação de uma ontologia, ainda foi necessário definir alguns novos termos, já que em alguns casos não foi encontrado nenhum termo existente refletindo a semântica do conceito que se desejava representar.

Desta forma, um conjunto de termos foi criado com o objetivo de complementar o esquema da ontologia. O *namespace* deste novo vocabulário é o *sbbd*, que durante este trabalho ficou provisoriamente hospedado no endereço http://www.cin.ufpe.br/~mgrb/ontology/sbbd. Ao invés de criar termos para todo o domínio em questão, nesta fase o foco foi na criação apenas dos termos essenciais ao esquema projetado e, até então, inexistentes.

A ferramenta utilizada para a criação do vocabulário foi o *Protégé*. O *Protégé* é uma ferramenta que permite construir ontologias, personalizar formulários de entrada de dados, inserir e editar dados, possibilitando, então, a criação de bases de conhecimento guiadas por uma ontologia. Sua interface gráfica provê acesso à barra de menus e à barra de ferramentas, além de apresentar cinco áreas de visualização (*views*) que funcionam como módulos de navegação e

edição de classes, atributos, formulários, instâncias e pesquisas na base de conhecimento, propiciando a entrada de dados e a recuperação das informações [24].

Utilizando esta ferramenta, a ontologia foi criada em OWL. Todos os novos termos foram ligados a outros vocabulários, utilizando *tags* de RDF *Schema* como *rdfs:range*, *rdfs:domain* e *rdfs:subClassOf*. Entre as classes criadas, por exemplo, duas foram definidas como subclasses de uma classe da SWC. Por questão de uniformização, o nome destas classes seguiram o mesmo padrão dos termos da SWC. Uma outra classe criada que merece destaque é a *sbbd:Participation*. Foi criada para representar um relacionamento entre três conceitos: Pessoa, Organização e Conferência (equivalente a um relacionamento ternário em modelagens relacionais). Alguns vocabulários possuem propriedades que ligam uma organização a uma pessoa, e uma pessoa a uma conferência. No entanto, a afiliação de participantes em simpósios não é definitiva. Ao longo de diferentes edições do SBBD, autores, por exemplo, podem estar afiliados a diferentes universidades. Como esta questão impacta em métricas importantes, tal como o *ranking* das instituições que mais publicam artigos, a modelagem da ontologia deveria oferecer suporte à representação deste tipo de situação de forma consistente. Seguindo uma recomendação do W3C [17], uma classe (*sbbd:Participation*) foi criada para representar essa relação ternária. Com isso, uma instância desta classe representa uma instância da relação entre indivíduos das 3 classes relacionadas.

A Tabela 4 exibe as classes criadas. Já as Tabelas 5 e 6 contêm as propriedades criadas, tendo a primeira as *Datatype Properties* e a segunda as *Object Properties*.

Classe	**Descrição**	**Mapeamentos com outros vocabulários**
sbbd:SteeringCommitteMember	Comitê de Programa	É subclasse (*rdfs:subClassOf*) de *swc:Role*
sbbd:DemoSessionEvent	Sessão de Demos (Evento satélite do SBBD)	É subclasse (*rdfs:subClassOf*) de *swc:AcademicEvent*
sbbd:Participation	Participação de uma pessoa, afiliado a uma organização, em um simpósio.	Não houve

Tabela 4. Termos criados para classes

Propriedade	***Domain***	***Range***
sbbd:acronym	*akt:University*	*String*
sbbd:amountAcceptedPapers	*bibo:Conference, swc:AcademicEvent*	*int*
sbbd:amountSubmittedPapers	*bibo:Conference, swc:AcademicEvent*	*int*

Tabela 5. Termos criados para Datatype Properties

Propriedade	***Domain***	***Range***
sbbd:bestPaper	*bibo:Article*	*bibo:Conference*
sbbd:participationPerson	*sbbd:Participation*	*foaf:Person*
sbbd:participationOrganization	*sbbd:Participation*	*akt:Organization*
sbbd:participationConference	*sbbd:Participation*	*bibo:Conference*
sbbd:state	*akt:Organization*	*gn:Feature*
sbbd:country	*akt:Organization*	*gn:Feature*

Tabela 6. Termos criados para Object Properties

Com a criação destes termos, o esquema da ontologia foi finalizado. Este esquema serviu como base para a criação do *dataset* RDF (próxima seção). A Figura 2 mostra o esquema final. Por questões de legibilidade, as *Datatype properties* foram omitidas da figura.

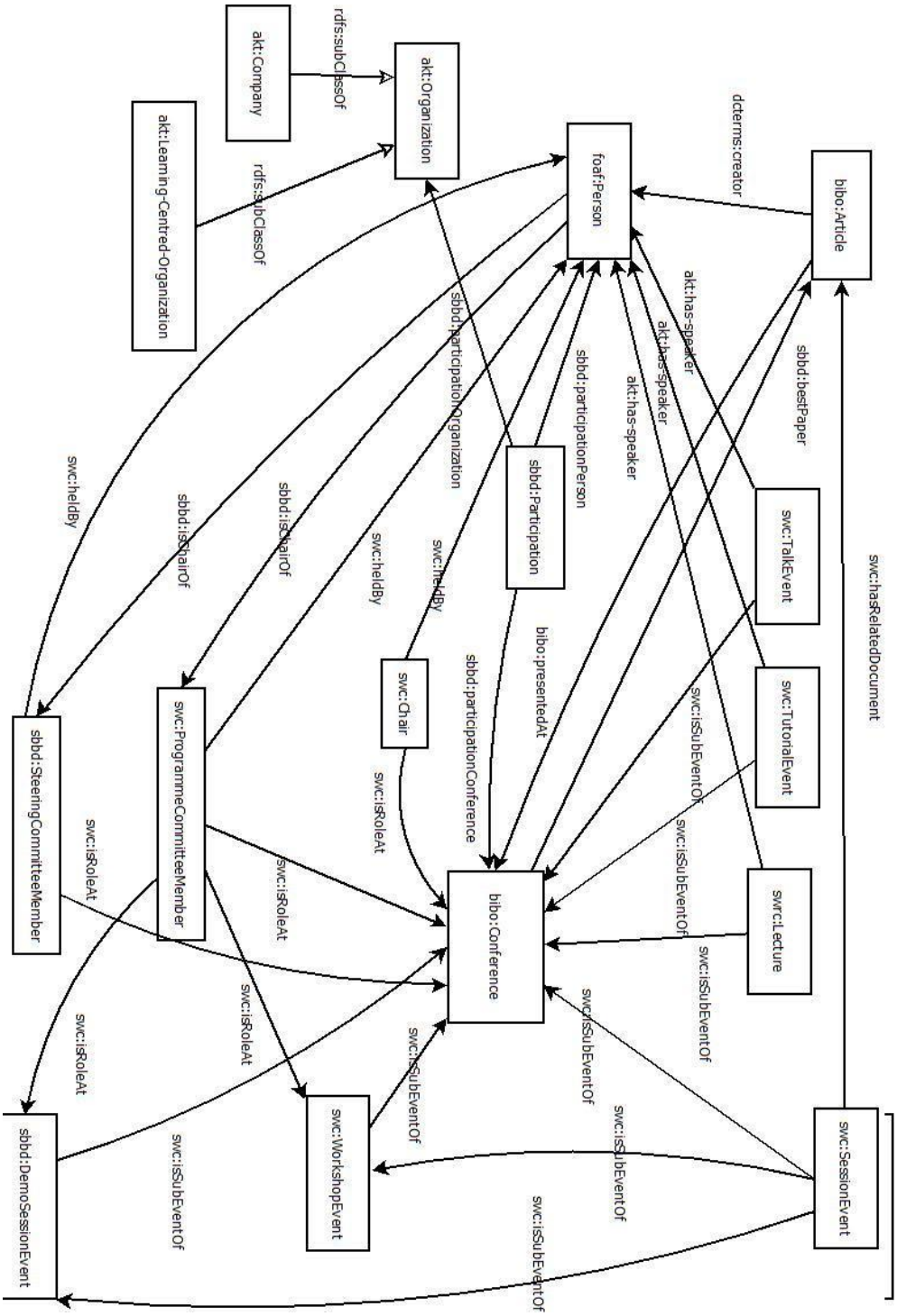


Figura 2. Esquema do dataset

3.3 Criação do *Dataset* SBBD

A criação do *dataset* baseou-se em duas linhas de ações: a conversão de dados em bases relacionais para o modelo RDF e a extração manual de dados dos *websites* e anais das edições passadas.

Em 2010 o SBBD completou 25 anos, e aproveitando esta ocasião, um portal Web foi criado com dados e estatísticas sobre todas as edições decorridas até então. Neste mesmo ano, também foi desenvolvido um trabalho [19] sobre uma rede de coautoria baseado em artigos publicados no SBBD, discutindo tanto suas características estruturais quanto sua evolução temporal. Para este estudo, houve uma coleta de dados envolvendo fontes digitais e não digitais. Os dados das edições de 1986 a 1998 foram coletados manualmente, diretamente dos anais dessas edições, enquanto que os das edições de 1999 a 2010 foram extraídos automaticamente da DBLP [13]. Os dados oriundos destas duas fontes foram reunidos em um banco de dados relacional *MySQL*[4]. Este banco reuniu dados de autores, artigos, conferências, afiliações, sessões técnicas, entre outros.

Após a análise desse banco, concluiu-se que a maioria de seu conteúdo era relevante no *dataset* deste trabalho, e então foi realizada uma pesquisa sobre ferramentas de geração de dados RDF a partir de bases relacionais. Após uma avaliação dos resultados da pesquisa, a plataforma D2RQ foi escolhida para ser utilizada. Um dos componentes desta plataforma é o *generate-mapping*. Esta ferramenta cria um arquivo de mapeamento no formato *Turtle*[5] a partir da análise do esquema do banco de dados. Para utilizar esta ferramenta, executa-se um comando, passando como parâmetros as credenciais do banco, para que a aplicação tenha acesso a seu esquema, e o nome do arquivo de saída.

A partir deste arquivo, é possível configurar os mapeamentos de tabelas e colunas do banco para classes e propriedades de ontologias. Por padrão, o *generate-mapping* cria termos locais para cada elemento do banco, mas é possível editar esses campos para reaproveitar termos existentes.

Para ilustrar como este mapeamento funciona, a Figura 3 mostra uma porção de um arquivo de mapeamentos gerado pelo *generate-mapping* a partir do banco utilizado no trabalho

[4] http://www.mysql.com/

[5] http://www.w3.org/TR/turtle/

de coautoria. Esta parte refere-se a como o D2RQ traduziria a tabela *Author*, que armazena os dados dos autores de artigos.

```
map:author a d2rq:ClassMap;
        d2rq:dataStorage map:database;
        d2rq:uriPattern "author/@@author.id@@";
        d2rq:class vocab:author;
        d2rq:classDefinitionLabel "author";
        .
```

Figura 3. Mapeamento da tabela Author para uma classe.

A primeira linha define que o mapeamento *map:author* será o mapeamento de uma classe (*d2rq:ClassMap*). Na linha seguinte, é definido que este mapeamento será feito consultando o banco que está configurado em *map:database*, uma configuração no início do arquivo que armazena o caminho do banco e suas respectivas credenciais. As próximas linhas definem o padrão da URI deste recurso (*d2rq:uriPattern*), a classe que ele representa (*d2rq:Class*), e um rótulo de definição opcional (*d2rq:classDefinitionLabel*). O padrão da URI é definido em relação ao seu endereço relativo, assumindo como URI base o endereço do servidor onde o D2R *Server* (outra ferramenta da plataforma, detalhada posteriormente) estará em execução. A expressão "*@@author_id@@*" é utilizada para referenciar o campo *id* da tabela *author*. A classe definida por padrão é a *vocab:author*, termo criado a partir do nome da tabela, em um vocabulário padrão local (*vocab*).

Já a Figura 4 mostra o mapeamento de propriedades (*d2rq:PropertyBridge*) a partir das colunas desta mesma tabela *Author*. O nome das propriedades (*d2rq:property*) também é prefixado com o vocabulário padrão. A coluna que será consultada para gerar cada propriedade e o seu respectivo tipo são definidos pelas *d2rq:column* e *d2rq:datatype*.

```
map:author_id a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:author;
        d2rq:property vocab:author_id;
        d2rq:propertyDefinitionLabel "author id";
        d2rq:column "author.id";
        d2rq:datatype xsd:integer;
        .
map:author_name a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:author;
        d2rq:property vocab:author_name;
        d2rq:propertyDefinitionLabel "author name";
        d2rq:column "author.name";
        .
map:author_gender a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:author;
        d2rq:property vocab:author_gender;
        d2rq:propertyDefinitionLabel "author gender";
        d2rq:column "author.gender";
        .
map:author_country a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:author;
        d2rq:property vocab:author_country;
        d2rq:propertyDefinitionLabel "author country";
        d2rq:column "author.country";
        .
```

Figura 4. Mapeamento das colunas da tabela Author para propriedades.

A partir deste arquivo, um dump RDF do banco de dados pode ser gerado através de outro componente da plataforma D2RQ, a ferramenta *dump-rdf*. Para utilizá-la, basta executar um comando, passando como parâmetro o tipo da serialização do dump (*Turtle, RDF/XML, RDF/XML-Abbrev, N3 ou N-Triple*), o caminho do arquivo de saída, e o arquivo de mapeamentos que será usado. Neste trabalho, a serialização utilizada foi *RDF/XML*. Após executar o *dump-rdf* com os mapeamentos das Figura 3 e 4, um dump RDF é criado. Um exemplo de recurso RDF gerado para um autor nesse processo está ilustrado na Figura 5. Percebe-se que este recurso de fato descreve um autor, porém não há reuso de vocabulários.

```
<rdf:Description rdf:about="#author/374">
  <vocab:author_country>Brazil</vocab:author_country>
  <vocab:author_gender>F</vocab:author_gender>
  <vocab:author_name>Bernadette Farias Lóscio</vocab:author_name>
  <vocab:author_id rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">374</vocab:author_id>
  <rdf:type rdf:resource="vocab/author"/>
</rdf:Description>
```

Figura 5. Recurso que representa um autor.

A Figura 6 mostra a seção do arquivo de mapeamentos referente aos autores modificada. A primeria linha representa a definição do prefixo (*@prefix*) de um vocabulário, neste caso o FOAF. O nome do mapeamento foi alterado para *map:person*, a classe utilizada para a representação passou a ser *foaf:Person* e o padrão da URI também foi modificado. A razão da substituição do nome *Author* por *Person* deve-se ao fato de que estas pessoas participam do simpósio, mas não necessariamente como autores. Podem ser membros de comitês, palestrantes, entre outros. Por isso, a definição genérica de *Person* foi a mais adequada. Desta forma, a partir das propriedades que um indivíduo da classe *foaf:Person* possui em relação a uma certa edição do SBDD, pode-se descobrir que papéis ele desempenhou nesta edição em particular.

```
@prefix foaf: <http://xmlns.com/foaf/spec/> .


map:person a d2rq:ClassMap;
        d2rq:dataStorage map:database;
        d2rq:uriPattern "person/@@author.id@@";
        d2rq:class foaf:Person;
        .
```

Figura 6. Mapeamento da tabela Author para uma classe com reuso de termos.

Os mapeamentos de propriedades referentes às colunas *name*, *gender* e *country* da tabela *Author* foram respectivamente atualizadas para *foaf:name*, *foaf:gender* e *foaf:based_near* (Figura 7). O mapeamento para a coluna *id* foi retirado, uma vez que esta tem a função apenas de chave primária no banco, e não agrega nenhuma semântica a respeito de um autor. No entanto, ela foi

utilizada no padrão das URIs, para garantir que cada recurso tenha um identificador único. Outro tratamento necessário foi a substituição dos valores literais da propriedade *foaf:gender*, visto que esses dados estavam no padrão "M" e "F", mas depois de pesquisas, concluiu-se que o padrão mais largamente utilizado é *male* e *female*.

O mapeamento da coluna *country* para a propriedade *foaf:based_near*, contudo, teve uma modificação mais significativa, pois passou a representar uma *Object Property*. Com isso, foi adicionado um padrão de URI absoluto, concatenando a URL base do *geonames* (http://www.geonames.org/) com o *GeoNameId* do local. Para isso, foi adicionada uma coluna na tabela *Author* chamada *geoname_id*, e via updates SQL, ela foi preenchida com os valores corretos que identificam os países determinados na coluna *country*. Feito isso, estas URIs geradas passaram a ser os objetos do predicado *foaf:based_near*. Os recursos identificados por estas URIs são indivíduos da classe *gn:Feature*, armazenados na base de dados do Geonames. Desta forma, os dados deste *dataset* passaram a estar interligados com o *dataset* do Geonames. A Figura 8 mostra como o mesmo recurso da Figura 5 passou a ser descrito.

```
@prefix foaf: <http://xmlns.com/foaf/spec/> .
map:person_name a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:person;
        d2rq:property foaf:name;
        d2rq:column "author.name";
        .
map:person_gender a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:person;
        d2rq:property foaf:gender;
        d2rq:column "author.gender";
        .
map:person_country a d2rq:PropertyBridge;
        d2rq:belongsToClassMap map:person;
        d2rq:property foaf:based_near;
        d2rq:uriPattern "http://www.geonames.org/@@author.geoname_id@@";
        .
```

Figura 7. Mapeamento das colunas da tabela Author para propriedades, com reuso de termos.

```
<rdf:Description rdf:about="#person/374">
  <foaf:based_near rdf:resource="http://www.geonames.org/3469034"/>
  <foaf:gender>female</foaf:gender>
  <foaf:name>Bernadette Farias Lóscio</foaf:name>
  <rdf:type rdf:resource="http://xmlns.com/foaf/spec/Person"/>
</rdf:Description>
```

Figura 8. Recurso que representa um autor, com reuso de termos.

Este mesmo processo descrito neste exemplo também foi aplicado de forma semelhante a todas as tabelas do banco. O banco de dados original sofreu mudanças ao longo deste trabalho, na maior parte das vezes ligadas à padronização de valores, e a colunas auxiliares para o mapeamento (caso da coluna *geoname_id*). A partir disso, foi gerado um arquivo RDF com todos esses dados armazenados de acordo com o esquema definido na Seção 3.2.

Além dos dados disponíveis neste banco, outros dados foram adicionados ao *dataset* manualmente. Atraveś dos *websites* e anais dos eventos passados, foram criados *scripts* SQL para a inserção desses dados no banco que, por sua vez, foram convertidos para RDF através do D2RQ, como descrito no exemplo anterior.

3.4 SPARQL *Endpoint*

A plataforma D2RQ possui uma ferramenta chamada *D2R-Server*, que permite a publicação do conteúdo de bases relacionais na Web Semântica. Esta ferramenta provê uma interface para navegar nos dados RDF (Figuras 9 e 10), de forma semelhante a um *browser* semântico. Além disso, ela disponibiliza um SPARQL *Endpoint* para a submissão de consultas SPARQL. Ao receber requisições Web via este *endpoint*, o *D2R-Server* reescreve estas consultas em SQL e submete-as ao banco. A Figura 11 mostra a interface desta ferramenta para realizar consultas no *endpoint*. Os resultados podem ser vistos em formato HTML, XML ou JSON.

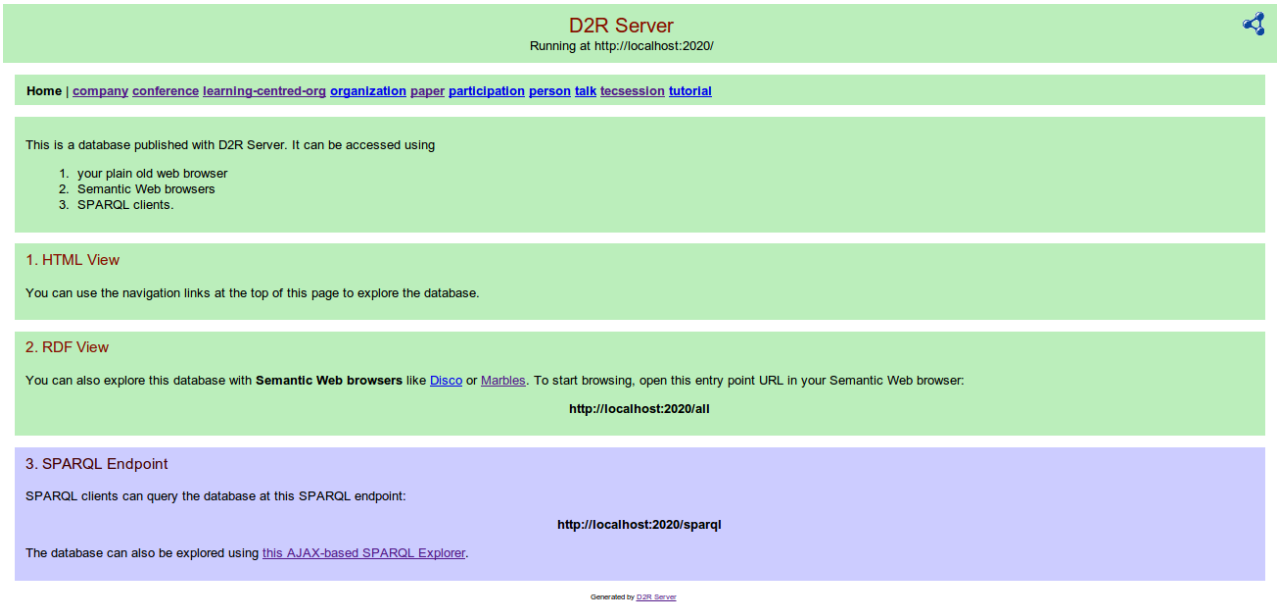


Figura 9. Interface do D2R-Server - Página Inicial.

Figura 10. Interface do D2R-Server - Navegação em um paper.

Figura 11. Interface do D2R-Server para consultas no SPARQL Endpoint.

A partir desta interface, foram realizadas consultas explorando as classes e propriedades definidas no mapeamento. A visão do *dataset* através desta tela foi importante para garantir a possibilidade de consultas relevantes e checar se os mapeamentos estavam sendo realizados corretamente. A Tabela 7 mostra os prefixos utilizados na composição das consultas.

Prefixo	**Vocabulário**
foaf	http://xmlns.com/foaf/spec/
dcterms	http://purl.org/dc/terms/
sbbd	http://cin.ufpe.br/~mgrb/ontology/sbbd#
event	http://purl.org/NET/c4dm/event.owl#
akt	http://www.aktors.org/publications/ontology/portal#

bibo	http://purl.org/ontology/bibo/
swc	http://data.semanticweb.org/ns/swc/ontology#
swrc	http://ontoware.org/swrc/swrc/SWRCOWL/swrc_updated_v0.7.1.owl#

Tabela 7. Prefixos dos vocabulários utilizados nas consultas SPARQL.

Seguem abaixo algumas consultas realizadas:

Consulta 1: Quais palestras foram ministradas por palestrantes do sexo feminino?

```
SELECT DISTINCT ?name WHERE {
  ?person a foaf:Person.
  ?person foaf:name ?name.
  ?person foaf:gender "female".
  ?talk a swc:TalkEvent.
  ?talk akt:has-speaker ?person.
}
```

A Tabela 8 ilustra o resultado desta consulta.

name
Anastassia Ailamaki
Claudia Bauzer Medeiros
Marta Lima de Queirós Mattoso
Tova Milo

Tabela 8. Resultado da consulta 1.

Consulta 2: Quais artigos apresentados por autores da UFPE?

```
SELECT DISTINCT ?title ?authorName WHERE {
  ?article a bibo:Article.
  ?article dcterms:title ?title.
  ?person a foaf:Person.
  ?person foaf:name ?authorName.
  ?article dcterms:creator ?person.
  ?participation sbbd:participationPerson ?person.
  ?participation sbbd:participationOrganization ?institution.
  ?institution sbbd:acronym  "UFPE".
}
```

A Tabela 9 exibe o resultado desta consulta. Como muitas linhas foram retornadas, a tabela exibe apenas as 10 primeiras.

title	**authorName**
Uma Linguagem para um Gerenciador de Banco de Dados Relacional em Micro-computador	Sonia Schechtman Sette
Uma Linguagem para um Gerenciador de Banco de Dados Relacional em Micro-computador	Danilo Florisi
Proposta de uma Linguagem de Restrição de Dados	Patrícia Gomes Soares
Proposta de um Subsistema de Restrições para um Gerenciador de Banco de Dados	Anderson Andrade De Menezes Filho
Uma Ferramenta de Apoio a Projeto Lógico em Banco de Dados	Andréa Zisman
MODOC - Uma Metodologia para Modelagem Integrada de Documentos	Décio Fonseca
Instanciating Integrated Schemas	Regina Motz

GET: Um Gerenciador Dinâmico de Tipos de Dados	Ana Carolina Salgado
Querying Geographical Data Warehouses With GeoMDQL.	Valéria Cesário Times
Towards a Formal Object-Oriented Data Modeling: The KBZ+ Approach	Fernando da Fonseca de Souza

Tabela 9. Resultado da consulta 2.

Consulta 3: Quais instituições de ensino já publicaram no SBBD e quantos autores publicaram por cada uma delas (ordem descendente)?

```
SELECT DISTINCT ?name (COUNT(?person) as ?authors) WHERE {
  ?institution a akt:Learning-Centred-Organization.
  ?institution foaf:name ?name.
  ?person a foaf:Person.
  ?article dcterms:creator ?person.
  ?participation sbbd:participationPerson ?person.
  ?participation sbbd:participationOrganization ?institution.
  ?participation sbbd:participationConference ?conference.
  ?article bibo:presentedAt ?conference.
}
GROUP BY ?institution ?name
ORDER BY DESC(?authors)
```

A Tabela 10 mostra as 10 primeiras linhas do resultado desta consulta.

name	**authors**
Universidade Federal do Rio Grande do Sul	100
Universidade Federal do Rio de Janeiro	76

Pontifícia Universidade Católica (RJ)	71
Universidade Federal de Pernambuco	64
Universidade Estadual de Campinas	47
Universidade Federal da Paraíba	45
Universidade Federal de Minas Gerais	43
Universidade de São Paulo	28
Instituto Militar de Engenharia	26
Universidade Federal de São Carlos	13

Tabela 10. Resultado da consulta 3.

3.5 Aplicação para visualização de dados

Uma aplicação de visualização de dados foi desenvolvida com o objetivo de fornecer informações relevantes sobre o SBBD, através de gráficos e tabelas. Também é possível que o usuário selecione alguns parâmetros na geração de gráficos dinâmicos, oferecendo maior interatividade.

As principais tecnologias utilizadas foram a linguagem de programação *Python*[6], e o framework para desenvolvimento Web *Django*[7]. Para a criação dos gráficos, foi utilizada a API *Google Chart Tools*[8], que oferece opções de visualização em diversos formatos, como gráficos em barra, pizza, bolha, entre outros. Já as consultas SPARQL foram submetidas ao *Endpoint* do *D2R-Server* através da API *SPARQL Endpoint interface to Python*. Esta API agrega como grande vantagem sua simplicidade, bastando apenas definir o *Endpoint* que será consultado e a

[6] http://www.python.org/
[7] https://www.djangoproject.com/
[8] https://developers.google.com/chart/

respectiva consulta. As consultas podem ser retornadas em formato XML, JSON, ou em uma estrutura de dados própria de *Python* chamada *Dictionary*[9].

As Figuras 12 e 13 ilustram exemplos de gráficos gerados na aplicação consultando o *dataset* criado.

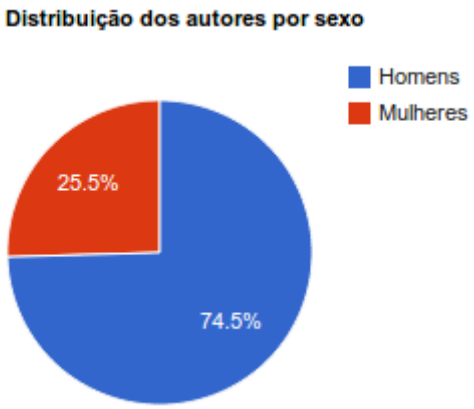


Figura 12. Gráfico de distribuição de autores por sexo.

[9] http://docs.python.org/2/tutorial/datastructures.html#dictionaries

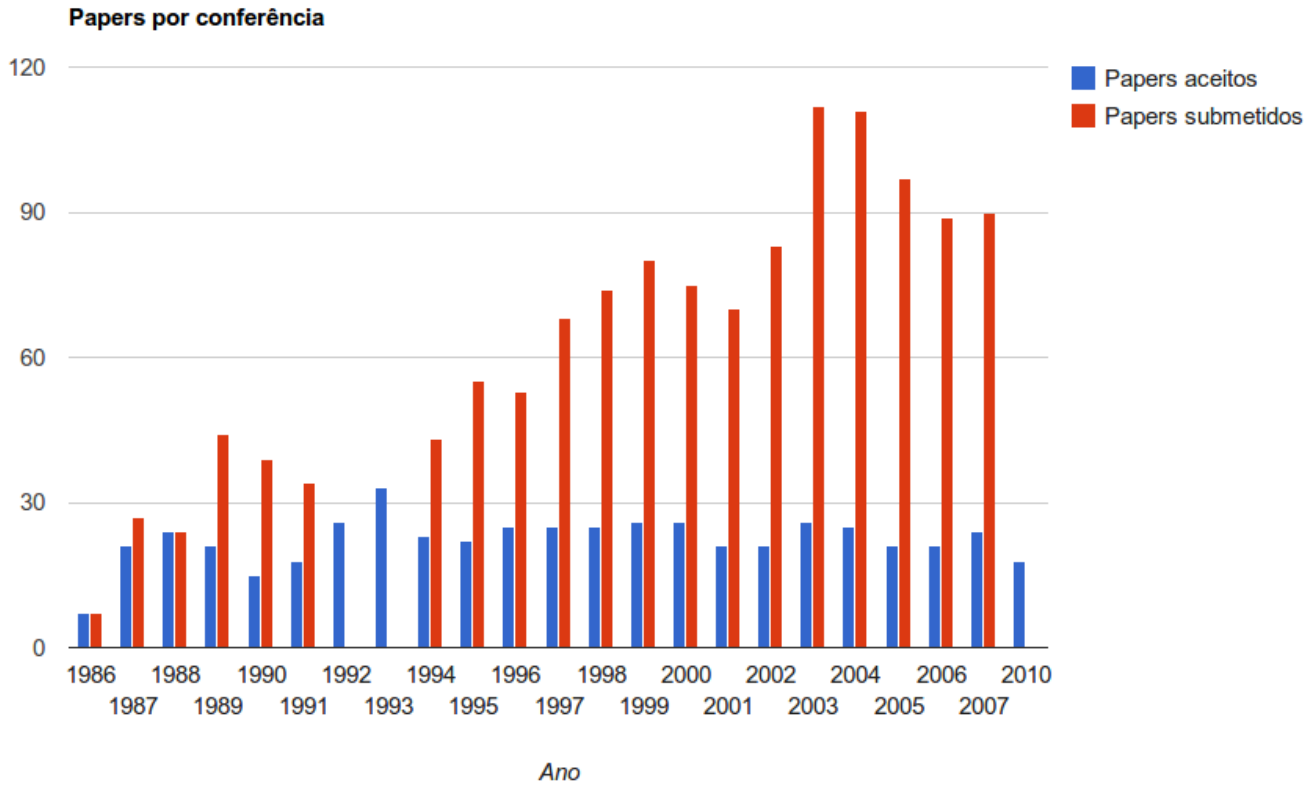


Figura 13. Gráfico de papers por conferência ao longo das edições do SBBD.

Na Figura 13, é possível notar que o *dataset* não possui a quantidade de *papers* submetidos para algumas edições do SBBD, como os anos 1992 e 1993. Além disso, as edições de 2011 e 2012 não possuem os valores quantitativos de *papers* aceitos e submetidos.

3.6 Considerações

Este capítulo abordou o desenvolvimento deste trabalho, dividido em 4 etapas. Estas fases, representadas por seções neste capítulo, foram desenvolvidas de forma incremental, sendo a fase subsequente dependente da fase anterior.

Capítulo 4 - Conclusão

Este trabalho dividiu-se essencialmente em três partes: a criação de um conjunto de dados em RDF, interligado com grandes *datasets* conhecidos; a publicação deste conjunto de dados em um *SPARQL Endpoint*, disponível para a execução de consultas; o desenvolvimento de uma aplicação Web para a visualização de gráficos e tabelas gerados a partir do *dataset* criado.

O esquema foi composto por 16 classes e 26 propriedades. Deste total, 13 classes (81%) e 17 propriedades (65%) tiveram seus termos reusados de outros vocabulários. Desta forma, houve êxito no reaproveitamento de vocabulários existentes. Com este *dataset* RDF criado sobre os dados e metadados do SBBD, as informações das edições passadas deste grande evento passam a fazer parte da Web de dados. Além disso, aplicações relacionadas ao SBBD agora têm a possibilidade de utilizar este conjunto de dados como uma de suas fontes de dados. Este material também pode incentivar futuros trabalhos na área de Web Semântica e *Linked Data* sobre este simpósio, ou de forma mais genérica, sobre eventos acadêmicos em geral.

Além da criação do *dataset*, também foi criado um vocabulário com termos adicionais sobre o SBBD, mas que podem ser aplicados no esquema de outros conjuntos de dados sobre conferências. A ligação destes termos com termos de outros vocabulários facilita sua integração a aplicações existentes.

O *SPARQL Endpoint* oferece ao usuário a possibilidade de escrever suas próprias consultas SPARQL. A partir do esquema do *dataset*, disponibilizado na aplicação desenvolvida, é possível explorar o conjunto de dados utilizando todos os recursos e flexibilidade do SPARQL. A aplicação de visualização de dados traz duas contribuições principais: a visualização de informações em um formato mais amigável, e a comprovação de que é possível construir aplicações consistentes a partir deste *dataset* que foi criado.

Com a realização deste trabalho, algumas linhas de ação futuras podem ser traçadas. Dentre elas, podemos destacar as seguintes:

- **Ampliação do *Dataset* -** Ampliar o conjunto de dados com mais dados que não foram coletados durante a realização deste trabalho. Um exemplo de ação com este foco pode ser a análise com mais detalhes dos dados da DBLP, para tentar desenvolver uma forma

automatizada de recuperar mais dados dela sobre o SBBD. Com mais dados, mais informações relevantes podem ser extraídas do *dataset*.

- **Mudança de ferramenta para disponibilização do *SPARQL Endpoint* -** Uma desvantagem de utilizar o *SPARQL Endpoint* oferecido pelo *D2R-Server* é que um de seus pré-requisitos é a existência de uma base relacional, para que os mapeamentos sejam feitos em tempo real. Em algumas situações, é interessante que o *Endpoint* não possua esta dependência, necessitando apenas de um conjunto de dados RDF para executar consultas. Desta forma, uma possível melhoria para este trabalho seria adotar uma ferramenta independente de banco de dados relacional para a disponibilização do *SPARQL Endpoint*.

- **Enriquecer aplicação de visualização de dados -** À medida que o *dataset* é ampliado, surgem novas possibilidades de representar estes dados em formatos mais amigáveis para o usuário final. Desta forma, de maneira incremental, a visualização das informações do SBBD pode ser melhorada com mais variedade de gráficos, tornando a aplicação mais dinâmica e aprimorando a interatividade com o usuário.

Em suma, este trabalho contribui para a disseminação das práticas de Web Semântica e *Linked Data* na comunidade acadêmica. As contribuições em várias camadas, desde uma fonte de dados até uma aplicação para o usuário final, demonstram as possibilidades que a Web de Dados está proporcionando com sua expansão.

Referências bibliográficas

[1] ALANI, H.; HARRIS, S.; O'NEILL, B. "OntologyWinnowing: A Case Study on the AKT Reference Ontology." *Computational Intelligence for Modelling, Control and Automation, 2005 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, International Conference on*. Vol. 2. IEEE, 2005.

[2] ALMEIDA, M. B. "Uma introdução ao XML, sua utilização na Internet e alguns conceitos complementares." *Ciência da informação* 31.2 (2002): 5-13.

[3] ANTONIOU, G.; VAN HARMELEN, F. "Web ontology language: Owl."*Handbook on ontologies*. Springer Berlin Heidelberg, 2009. 91-110.

[4] BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. "The semantic web."*Scientific american* 284.5 (2001): 28-37.

[5] BIBLIOGRAPHIC ONTOLOGY SPECIFICATION. Disponível em: <http://bibliontology.com/specification>. Acesso em: 8 abr. 2013.

[6] BIZER, C.; CYGANIAK, R. "D2RQ-lessons learned." *W3C Workshop on RDF Access to Relational Databases*. 2007.

[7] BIZER, C.; CYGANIAK, R.; HEATH, T. "How to publish linked data on the web." (2008).

[8] BIZER, C.; HEATH, T.; BERNERS-LEE, T. "Linked data-the story so far." *International Journal on Semantic Web and Information Systems (IJSWIS)* 5.3 (2009): 1-22.

[9] BIZER, C.; HEATH, T.; IDEHEN, K.; BERNERS-LEE, T. Abril, 2008. Linked data on the web (LDOW2008). In *Proceeding of the 17th international conference on World Wide Web* (pp. 1265-1266). ACM.

[10] DUBLIN CORE METADATA INITIATIVE. Disponível em: <http://dublincore.org/metadata-basics/>. Acesso em: 8 abr. 2013.

[11] FILHO, F.W; LÓSCIO, B. F (2010) Web Semântica: Conceitos e Tecnologias. Minicurso apresentado na Escola Regional de Computação Ceará - Maranhão – Piauí 2010.

[12] GOLBECK, J.; ROTHSTEIN, M. "Linking social networks on the web with FOAF: a semantic web case study." *Proceedings of the 23rd national conference on Artificial intelligence*. Vol. 2. 2008.

[13] LEY, M. "The DBLP computer science bibliography: Evolution, research issues, perspectives." *String Processing and Information Retrieval*. Springer Berlin Heidelberg, 2002.

[14] MAEDCHE, A.; STAAB, S. "Ontology learning for the semantic web". *Intelligent Systems, IEEE* 16.2 (2001): 72-79.

[15] MCGUINNESS, D. L.; VAN HARMELEN, F. "OWL web ontology language overview." *W3C recommendation* 10.2004-03 (2004): 10.

[16] MILLER, E. "An introduction to the resource description framework." *Bulletin of the American Society for Information Science and Technology* 25.1 (1998): 15-19.

[17] NOY, N.; et al. "Defining n-ary relations on the semantic web." *W3C Working Group Note* 12 (2006): 4.

[18] PÉREZ, J.; ARENAS, M; GUTIERREZ, C. "Semantics and Complexity of SPARQL." *The Semantic Web-ISWC 2006*. Springer Berlin Heidelberg, 2006. 30-43.

[19] PROCÓPIO, P. S.; LAENDER, A. H. F.; MORO, M. M. "Análise da rede de coautoria do simpósio brasileiro de bancos de dados." *SIMPÓSIO BRASILEIRO DE BANCO DE DADOS, Florianópolis, 2011. Proceedings... Florianópolis* (2011).

[20] RAIMOND, Y.; ABDALLAH, S. *The event ontology*. Technical report, 2007. http://motools. sourceforge. net/event, 2007.

[21] SAHOO, S. S.; et al. "A survey of current approaches for mapping of relational databases to rdf." *W3C RDB2RDF Incubator Group Report* (2009).

[22] SEMANTIC WEB CONFERENCE ONTOLOGY. Disponível em: <http://data.semanticweb.org/ns/swc/swc_2009-05-09.html>. Acesso em: 8 abr. 2013.

[23] SEMANTIC WEB WIKI, 2013. Disponível em: <http://semanticweb.org/wiki/Main_Page>. Acesso em: 9 jan. 2013.

[24] SEMPREBOM, T.; CAMADA, M. Y.; MENDONÇA, I. "Ontologias e Protégé." *Documento extraído em* 21.07 (2007): 2007.

[25] SIMPÓSIO BRASILEIRO DE BANCO DE DADOS. SBBD, 2012. Disponível em: <http://sws2012.ime.usp.br/sbbd/index.php>. Acesso em: 25 jan. 2013.

[26] SURE, Y.; et al. "The SWRC ontology–Semantic Web for research communities." *Progress in Artificial Intelligence*. Springer Berlin Heidelberg, 2005. 218-231.

[27] VATANT, B.; WICK, M. "Geonames ontology." (2012).

[28] W3C. Semantic Web, 2013. Disponível em: <http://www.w3.org/standards/semanticweb/>. Acesso em: 8 abr. 2013.

[29] WIKIPEDIA, THE FREE ENCYCLOPEDIA. Semantic Web, 2013. Disponível em: <http://en.wikipedia.org/wiki/Semantic_Web>. Acesso em: 9 jan. 2013.

[30] WIKIPEDIA, THE FREE ENCYCLOPEDIA. Web 2.0, 2013. Disponível em: <http://en.wikipedia.org/wiki/Web_2.0>. Acesso em: 17 jan. 2013.

[31] WORLD WIDE WEB CONFERENCE 2012. Metadata Initiative, 2012. Disponível em: <http://www2012.wwwconference.org/committee/metadata-initiative/>. Acesso em: 25 jan. 2013.