

Navegação e Visualização Semânticas de Linked Data na Web 2.0

André Luiz Carlomagno Rocha, Cássio V. S. Prazeres
Universidade Federal da Bahia – Departamento de Ciência da Computação
Av. Adhemar de Barros, Ondina, Salvador-BA, Brasil
{andreluiz, prazeres}@dcc.ufba.br

RESUMO

A Web atual (2.0), se estendida pela semântica de *Linked Data* (RDF), pode oferecer novas possibilidades, não apenas para as máquinas. Pessoas podem se beneficiar da semântica explícita de dados embutidos em páginas HTML por meio de padrões de marcação (RDFa, Microdata e Microformats). Entretanto, poucas ferramentas, baseadas nos princípios *Linked Data*, estão disponíveis para auxiliar usuários sem expertise na realização dessa tarefa. A navegação e visualização de dados semânticos mostra-se um problema recorrente em tais soluções. Esse trabalho pretende propor um modelo, com o objetivo de estender a Web 2.0, para prover navegação e visualização de *Linked Data*, possibilitando que pessoas comuns tirem proveito do conteúdo estruturado implícito em páginas HTML no formato RDFa.

Palavras-Chave

Web de Dados, Linked Data, RDFa, Modelo Dexter, SLN, Navegação, Visualização.

1. FUNDAMENTAÇÃO TEÓRICA

Na proposta inicial da Web Semântica, articulada por Berners-Lee et al. [3], a ideia principal foi a extensão da Web atual mediante técnicas e padrões de estruturação. Desde então, a quantidade de dados estruturados na Web, embutidos em páginas HTML por meio de diferentes padrões de marcação (e.g. RDFa), vem crescendo consideravelmente [5]. Em contraste com a Web atual, que relaciona, por meio de *hyperlinks*, documentos completos como páginas Web, a Web Semântica relaciona dados baseada em um modelo de dados uniforme. Esse modelo consiste de declarações elementares na forma de “sujeito-predicado-objeto” expressados por meio do padrão RDF (*Resource Description Framework*) [14]. Além disso, o modelo permite a utilização de vocabulários definidos como ontologias RDFS (*RDF Schema*) [6] e OWL (*Ontology Web Language*) [1].

Linked Data [2] refere-se a princípios para publicação, consumo, reutilização e interconexão de dados RDF na Web, de tal forma que possam ser lidos e entendidos por máquinas. O significado desses dados é explicitamente definido e os mesmos podem ser ligados a outros dados contidos em fontes externas. RDFa [16] é uma das formas de serializar dados RDF. Especificamente, RDFa permite embutir triplas RDF em documentos HTML, usando os próprios atributos da linguagem de marcação. Ao incorporar RDFa em páginas HTML, autores são capazes de indicar aos agentes de software o significado de informações contidas nessas páginas, transformando-as em dados legíveis por máquinas. Marcações RDFa podem fornecer conhecimento além daqueles exibidos explicitamente para as pessoas no *layout* das páginas

HTML. Na maioria das vezes, o conteúdo dos atributos RDFa é acessível apenas por agentes de software.

A Web de *Linked Data* (Web de Dados [15]), assim como a Web atual 2.0 (Web de Documentos [4]), pode ser considerada um sistema hipertexto (nesse contexto hipertexto não está restrito apenas a texto), onde componentes do sistema estão relacionados, dando origem a uma rede de nós e *links*. O modelo Dexter [12] é um modelo de referência que captura, formal e informalmente, abstrações importantes encontradas em uma ampla variedade de sistemas hipertexto. Portanto, esse novo ambiente de informação, a Web de Documentos estendida pela Web de Dados (ambiente semântico), pode ser mapeado para o modelo Dexter com o objetivo de padronizar elementos, atividades e operações e, conseqüentemente, resolver problemas característicos de sua heterogeneidade.

O modelo Dexter divide o hipertexto em três camadas: camada de tempo de execução (*runtime layer*); camada de armazenamento (*storage layer*) e a camada que trata o interior do componente (*within-component layer*). O principal foco do modelo é na camada de armazenamento, que modela a estrutura de rede básica de nós e *links*. O modelo Dexter também se preocupa com o mecanismo de ancoragem e especificação de apresentação, que formam as interfaces entre a camada de armazenamento e as camadas de interior do componente e de tempo de execução, respectivamente. Por outro lado, o modelo, propositadamente, não cobre a camada de interior do componente (tentar criar um modelo genérico para todos os tipos de conteúdo seria inviável) e apenas sugere uma estrutura básica simplificada para a camada de tempo de execução.

A camada de tempo de execução do modelo Dexter trata das questões de apresentação dos elementos da rede hipertexto para o usuário do sistema. Ou seja, essa camada se preocupa especificamente com a maneira como os nós e *links* serão apresentados para os usuários finais. Dessa forma, o mapeamento do ambiente semântico para o modelo Dexter pode trazer benefícios valiosos. Tais benefícios estão diretamente ligados à resolução dos problemas relacionados à navegação e apresentação de conteúdo semântico para usuários não técnicos, considerando o aperfeiçoamento da camada de tempo de execução do modelo para lidar com as particularidades encontradas na Web de Dados.

Para aprimorar a maneira como o sistema interage com o usuário em termos de apresentação e navegação, é necessário utilizar técnicas de inferência baseadas em regras de raciocínio. Essas técnicas permitem explorar a Web de Dados oferecendo possibilidades de novas descobertas orientadas, por exemplo, às preferências do usuário e ao contexto em que ele se encontra. A

Rede de Links Semânticos (*Semantic Link Network* - SLN) [17] é um modelo flexível de dados semânticos para gerenciar redes de recursos. Uma SLN consiste de três componentes: um conjunto de nós semânticos; um conjunto de *links* semânticos e um conjunto de regras de raciocínio. Embora o mecanismo de lógica de descrição (OWL) combinado com a linguagem de regras da Web Semântica (SWRL) forneçam regras de raciocínio entre ontologias, SLN possui regras de raciocínio especiais que podem suportar raciocínio analógico e indutivo, entre outros. Além disso, SLN possui um mecanismo de representação completo e uma forte capacidade de expressividade semântica. Portanto, SLN pode ser usada como peça do mecanismo de inferência e raciocínio no mapeamento do ambiente semântico para o modelo Dexter, a fim de refinar tal mecanismo.

2. IDENTIFICAÇÃO DO PROBLEMA

De acordo com o projeto *Web Data Commons* [5], em seu relatório de extração de agosto de 2012, existem mais de 168 milhões de páginas Web contendo triplas RDFa, com uma média de 2 mil triplas por domínio. Embora esses dados tenham as máquinas como seus principais usuários, pessoas podem se beneficiar desse ambiente para explorar conhecimento relacionado, tirando proveito de conteúdo *Linked Data* implícito em páginas da Web 2.0. No entanto, a recuperação e uso adequado de tal conhecimento, representado formalmente, é um processo não trivial e propenso a erros para a maioria das pessoas sem experiência com Web Semântica.

Para Camarda et al. [7], apesar das centenas de milhares de triplas RDF contidas na Web de Dados, é difícil encontrar ferramentas de apresentação e navegação fáceis de usar, que sejam realmente baseadas em padrões RDF e capazes de demonstrar a eficácia do modelo *Linked Data*. Davies et al. [10] argumentam que: se a finalidade é a de que os usuários finais possam utilizar a Web de Dados da mesma forma que eles utilizam a Web de Documentos, então eles precisam de ferramentas, adequadas à sua condição de usuários não técnicos, que lhes permitam fazê-lo.

Portanto, percebe-se que, embora as pessoas também se beneficiem do significado dos dados, elas devem ser munidas de meios adequados para obter tais benefícios. Dessa forma, esses usuários inexperientes podem satisfazer suas necessidades de informação de maneira mais orientada e eficaz. Eles podem explorar conceitos desconhecidos e novos relacionamentos; obter acesso personalizado e contextualizado a recursos e serviços relevantes; e, conseqüentemente, contribuir para o crescimento da Web de Dados [18]. Isso cria uma necessidade por modelos, métodos e ferramentas que permitam a tais usuários interagirem adequadamente com os dados contidos no ambiente semântico.

3. OBJETIVO

O objetivo dessa pesquisa é possibilitar que pessoas comuns possam tirar proveito de conteúdo *Linked Data* implicitamente incorporado em páginas HTML por meio do padrão RDFa. Ao final do trabalho, espera-se propor um modelo para tratar as questões de navegação e visualização nesse ambiente. Para demonstrar o modelo, deverá ser construída uma ferramenta que apoie o usuário durante a navegação e visualização de dados estruturados em páginas Web. A ferramenta deve prover o contato adequado do usuário diante dos conjuntos de dados

semânticos disponíveis na Web de Dados. Ela deve ser capaz de fornecer recursos para habilitar as pessoas a navegar e visualizar o espaço global de dados de forma apropriada a sua condição de usuário não técnico, aprimorando sua experiência de interação enquanto lida com *Linked Data*.

4. CONTRIBUIÇÕES ESPERADAS

A principal contribuição que esse trabalho espera deixar é um modelo de navegação e apresentação de dados semânticos, focado no usuário final. O modelo proposto deve estender a camada de tempo de execução do modelo Dexter para definir propriedades, atividades, operações e transações, que serão usadas para padronizar a construção de sistemas desse domínio. O modelo ainda deve suportar regras de raciocínio baseadas nas SLN para permitir inferências, suprindo o usuário com novas descobertas de recursos e serviços, até então, desconhecidos.

Como contribuições secundárias, o trabalho pretende desenvolver uma API de desenvolvimento, baseada no modelo proposto e, conseqüentemente, uma ferramenta, utilizando essa API, com foco na navegação e apresentação de dados semânticos para o usuário final. A ferramenta, inicialmente, será desenvolvida como uma extensão (*add-on*) para navegadores Web, com o objetivo de processar dados semânticos embutidos em páginas Web por meio de RDFa. A ferramenta poderá utilizar um serviço Web ou um *plugin* para executar ações de maior complexidade.

5. METODOLOGIA

A sequência lógica de macroatividades, a princípio, está ordenada da seguinte forma: (1) revisão sistemática da literatura acerca de estudos voltados para interfaces de usuário, priorizando navegação e apresentação, no âmbito da Web Semântica e levantamento de modelos de sistemas hipertexto; (2) criação de um modelo, baseado nos conhecimentos adquiridos na revisão sistemática, para tratar tais questões; e (3) para demonstrar e validar o modelo, criação de uma ferramenta para apoiar o usuário comum na tarefa da navegação e visualização de RDFa.

Além das macroatividades, microatividades estão previstas para serem realizadas durante e ao término de algumas dessas atividades maiores, são elas: (1a) uso e experimentação de ferramentas de apresentação e navegação; (1b) determinação de funções e recursos chave, bem como, definição dos problemas principais de apresentação e navegação encontrados em tais ferramentas; (2a) definição de operações elementares para guiar a construção do modelo; (2b) definição de uma API baseada nas operações elementares; (3a) definição do ambiente de testes e validação; (3b) definição dos requisitos de usabilidade a serem aferidos na avaliação. Dentro de cada macroatividade ainda podem aparecer outras microatividades específicas.

6. ESTADO ATUAL DO TRABALHO

Considerando as etapas descritas na Seção 5, foram realizadas, até então, as atividades (1), (1a), (1b) e (2a). Um artigo científico oriundo dessas atividades foi aceito recentemente, como artigo completo, no 19º Simpósio Brasileiro de Sistemas Web e Multimídia (WebMedia 2013) [8]. Além disso, uma microatividade não especificada, anterior a microatividade (2a), foi realizada: mapeamento do ambiente semântico para a camada de armazenamento do modelo Dexter.

No momento atual, os conhecimentos adquiridos na macroatividade (1) estão sendo explorados na definição do modelo de interação com o usuário. Uma proposta inicial do modelo pode ser vista na Seção 7. Paralelamente, estudos preliminares em ambientes de desenvolvimentos de extensões para navegadores estão sendo realizados com o objetivo de definir plataforma, linguagem de programação, arquitetura, entre outros, que serão utilizados no desenvolvimento da ferramenta que demonstrará o modelo (macroatividade (3)).

7. PROPOSTA INICIAL DE SOLUÇÃO

O modelo proposto nesse trabalho é baseado no modelo Dexter (Seção 1). O foco do trabalho é na camada de tempo de execução. No entanto, mostrou-se necessário o mapeamento do ambiente semântico para a camada de armazenamento, a fim de entender todos os aspectos envolvendo a modelagem. A Figura 1 apresenta o esquema desse mapeamento.

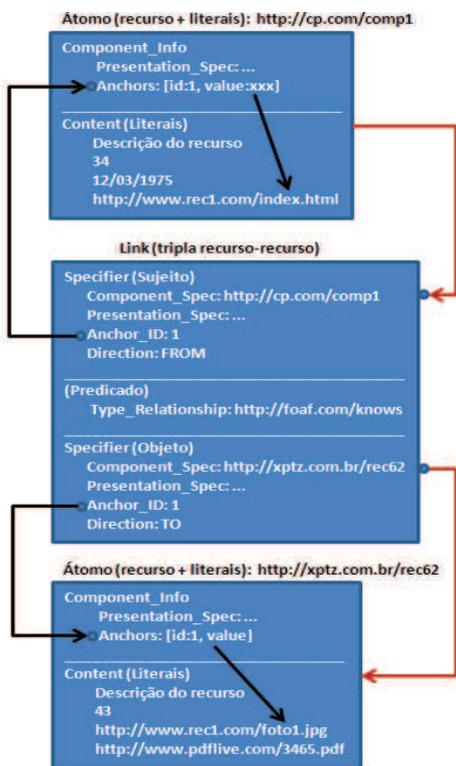


Figura 1. Mapeamento para a camada de armazenamento.

No mapeamento, todo recurso da Web de Dados é um componente (átomo) do modelo Dexter. Os átomos possuem um *Component_Info* (meta informação) e um *Content*. O *Component_Info* é formado de um *Presentation_Spec* – usado pela camada de tempo de execução para especificar como o átomo será apresentado para o usuário – e de um *Anchors*, uma coleção de pares chave-valor com *id* e *value* das ancoras definidas no átomo. O *Content* é formado pelos literais diretamente ligados ao recurso (e.g. descrição, número, data, URL, entre outros tipos de dados). Triplas que relacionam recursos (recurso-recurso), também são componentes (*links*). Por outro lado, triplas que ligam recursos a literais (recurso-literal), não são consideradas *links*. Um *link* possui dois *Specifiers* (sujeito e objeto) e um predicado usado para relacioná-los. Cada

Specifier possui um *Component_Spec*, para especificar o átomo; um *Presentation_Spec*, para definir como o *Specifier* será apresentado; um *Achor_ID*, para identificar a ancora (se houver) e um *Direction*, definindo a direção das ligações entre as ancoras. O predicado possui apenas o *Type_Relationship*, para identificar o tipo de relacionamento entre os *Specifiers*.

Na Figura 1 o átomo *http://cp.com/comp1*, possui uma descrição, idade, data de nascimento e uma *homepage*. Ele está relacionado com o átomo *http://xptz.com.br/rec62*, que possui descrição, idade, foto e um arquivo PDF. A relação é definida pelo *link* por meio do predicado *http://foaf.com/knows*. Ainda existe uma ligação, definida por ancoras, entre um ponto no interior da *homepage* do recurso *comp1* (página HTML) e a foto do recurso *rec62* (arquivo JPG), conhecida como ligação *span-to-span*.

Levando em conta o mapeamento do ambiente semântico para o modelo Dexter, um modelo simplificado (preliminar) foi definido. Embora o mapeamento tenha sido realizado, o modelo proposto aqui não se preocupa com as questões de armazenamento da rede hipertexto. A preocupação é especificamente com a navegação e apresentação dos dados para o usuário final, definidos na camada Dexter de tempo de execução e na interface dessa camada com a camada de armazenamento. Ou seja, a proposta desse trabalho consiste em um modelo que residirá dentro da *Runtime Layer* do modelo Dexter. A Figura 2 deixa essa questão bem clara.

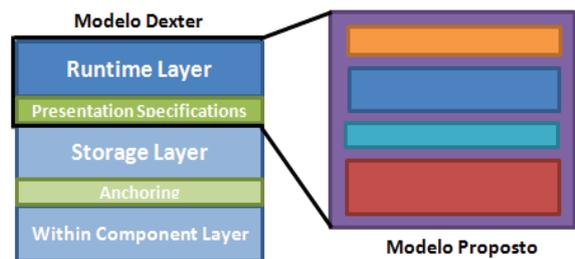


Figura 2. Escopo do trabalho (Modelo Proposto).

O esboço do modelo proposto foi definido em termos de operações elementares. Essas operações são categorizadas em Operações de Mapeamento e Carga, Operações de Apresentação e Navegação e Operações de Descoberta e Aconselhamento. As operações mais relevantes são descritas aqui de forma resumida:

1 – **Define Apresentação:** Permite ao usuário definir o tipo de apresentação. Os tipos predeterminados são baseados em filtros, permitindo mais ou menos detalhes durante a apresentação. Por exemplo, o usuário pode definir um nível mínimo de detalhes, onde apenas as propriedades mais importantes sobre uma pessoa (segundo o contexto) são apresentadas (e.g. dados pessoais, pessoas conhecidas, organizações que tem ligação, entre outros). Ao mesmo tempo pode ser definido um nível de salto (*hop*) para delimitar a quantidade de resultados ao percorrer a rede (A conhece B, B conhece C, C conhece D. Limite de 3 saltos).

2 – **Apresenta Átomo:** Usa: (i) a especificação de apresentação definida no próprio átomo; (ii) a apresentação definida pelo usuário; e (iii) a apresentação inferida com base no contexto e preferências do usuário. Um átomo pode ser apresentado, por exemplo, em uma caixa flutuante no evento de movimento do mouse sobre um elemento (*tooltip*). O *tooltip* mostra os literais diretamente ligados ao recurso que o átomo representa.

3 – **Apresenta Link**: Usa: (i) a especificação de apresentação definida nos especificadores; (ii) a apresentação definida pelo usuário; e (iii) a apresentação inferida com base no contexto e preferências do usuário. A apresentação de um link consiste na apresentação de dois átomos e a ligação que existe entre eles (usa Apresenta Átomo). Essa ligação pode, por exemplo, estar em forma de lista contendo os átomos relacionados e o principal.

4 – **Apresenta Rede**: Disponibiliza a rede de átomos e *links* usando: (i) a apresentação definida pelo usuário; e (ii) a apresentação inferida com base no contexto e preferências do usuário (usa Apresenta Link). Uma rede pode ser apresentada e/ou navegada, por exemplo, por meio de facetas onde os recursos são categorizados por tipo (*person, organization, product, event*). O nível de detalhes e quantidade de recursos depende da operação Define Apresentação.

5 – **Descobre Conhecimento**: Descobre novos conhecimentos a partir de um átomo(s), link(s) ou rede, consultando a Web de Dados com base em inferências, regras de raciocínio, contexto e preferências. Os novos átomos recém-descobertos são montados usando Operações de Construção e Carga e anexados a rede. O processo de descoberta pode usar, como base, ontologias OWL e RDFS, além de Rede de Links Semânticos (SLN). O resultado dessa operação pode ser, por exemplo, a obtenção de acesso personalizado e contextualizado a recursos e serviços relevantes.

6 – **Aconselha Navegação**: Aconselha o usuário, baseado em regras de raciocínio (usa Descobre Conhecimento), com relação à navegação, direcionando-o para a descoberta de novos conhecimentos com o objetivo de satisfazer suas necessidades. Pode, por exemplo, disponibilizar navegações facetadas contendo novos recursos, voltados para as necessidades do usuário.

8. TRABALHOS RELACIONADOS

RDF2RDFa [13] parte do princípio de que muitos usuários da Web Semântica têm dificuldades técnicas para publicar conteúdo estruturado (e.g. usuários de Sistema de Gerenciamento de Conteúdo). O foco da ferramenta, diferente da proposta desse artigo, é disponibilizar essas informações para agentes de software. Utilizando o modelo proposto, seria possível compor uma camada de tradução para disponibilizar os dados RDFa gerados por RDF2RDFa de forma amigável para usuários finais.

NautiLOD [11] é uma linguagem declarativa para enfrentar o desafio da navegação ao longo dos nós da Web de Dados. NautiLOD se mostrou uma poderosa ferramenta de consulta para navegação em grafos RDF. No entanto, nos testes realizados, a linguagem apresentou limitações em termos de descoberta de novos conhecimentos, usando SPARQL. O modelo proposto pretende usar SLNs, além de OWL e RDFS, para tratar esse problema, consequentemente, obtendo melhores resultados.

MyView [9] permite que pessoas consultem *Linked Data* por navegação. MyView possui um excelente mecanismo de raciocínio e recursos de personalização. Entretanto, a ferramenta deixa a desejar na apresentação dos resultados para o usuário final, tornando a visualização confusa quando usa representação centrada em objetos. Com o modelo proposto aqui, que usa representação centrada em sujeito, esse problema pode ser

reduzido. Ainda assim, a arquitetura de MyView pode ser de grande ajuda para o estudo e definição do modelo proposto.

9. REFERÊNCIAS

- [1] Bechhofer, S. et al. 2004. OWL Web Ontology Language Reference. *W3C Recommendation 10 February 2004*. Series editor: Brian McBride and Guus Schreiber.
- [2] Berners-Lee, T. 2006. Design Issues: Linked Data. <http://www.w3.org/DesignIssues/LinkedData.html> (last viewed April 28, 2013).
- [3] Berners-Lee, T., Hendler, J., and Lassila, O. 2001. The Semantic Web. *Scientific American*. Mai 2001. <http://dx.doi.org/10.1038/scientificamerican0501-34DOI>.
- [4] Bizer, C., Heath, T., and Berners-Lee, T. 2009. Linked Data – The Story So Far. *Int. J. Semantic Web Inf. Syst.*
- [5] Bizer, C. et al. Web Data Commons Project. <http://webdatacommons.org/> (last viewed July 17, 2013).
- [6] Brickley, D. and Guha, R. 2004. RDF Vocabulary Description Language 1.0: RDF Schema. *W3C Recommendation 10 February 2004*. Brian McBride.
- [7] Camarda, D., Mazzini, S., and Antonuccio A. 2012. LodLive, exploring the web of data. In *Proceedings of the 8th International Conference on Semantic Systems*.
- [8] Carlomagno, A., Gardel, T., Cerqueira, M. and Prazeres, C. 2013. jRDFa: navegação e visualização de Linked Data na Web. In *Proceedings of the 19th Brazilian Symposium on Multimedia and the Web. (WebMedia 2013)*.
- [9] Cheng, G., Wu, H., Gong, S., Zhang, H. and Qu, Y. 2011. Browsing Linked Data with MyView. In *Proceedings of the 10th International Semantic Web Conference (ISWC2011)*.
- [10] Davies, S., Hatfield, J., Donaher, C. and Zeitz, J. 2010. User Interface Design Considerations for Linked Data Authoring Environments. In *Proceedings of LDOW*.
- [11] Fionda, V., Gutierrez, C. and Pirró, G. 2012. Navigation - Semantic Navigation on the Web of Data. In *Proceedings of the 21st international conference on World Wide Web*.
- [12] Halasz, F. and Schwartz, M. 1990. The Dexter Hypertext Reference Model. In *Proceedings of the NIST Hypertext Standardization Workshop*, Gaithersburg, MD.
- [13] Hepp, M., García, R. and Radinger, A. 2009. RDF2RDFa: Turning RDF into Snippets for Copy-and-Paste. *8th International Semantic Web Conference (ISWC)*.
- [14] Klyne, G. and Carroll, J. 2004. Resource Description Framework (RDF): Concepts and Abstract Syntax. *W3C Recommendation*. Series editor: Brian McBride.
- [15] Kossman, D. 2005. A web of data: new architectures for new technology? In *the 7th annual ACM international workshop on Web information and data management*.
- [16] McCarron, S. et al. 2013. HTML+RDFa 1.1: Support for RDFa in HTML4 and HTML5, <http://www.w3.org/TR/rdfa-in-html/> (last viewed July 07, 2013).
- [17] Wang, S. et al. 2012. Advances of the Semantic Link Network. In *the Eighth International Conference on Semantics, Knowledge and Grids*.
- [18] Ziegler, J. 2011. Semantic web meets UI: context-adaptive interaction with semantic data. In *Proceedings of the 29th Annual European Conference on Cognitive Ergonomics*.