

Chamada a Serviços descritos por Interfaces Semânticas a partir da navegação semântica

José Renato Villela Dantas
Universidade de Fortaleza, SERPRO
Av. Washington Soares, 1321 J-30
CEP: 60.811-341, Fortaleza, CE, Brasil
jrdantas@edu.unifor.br

Pedro Porfírio Muniz Farias
Universidade de Fortaleza
Av. Washington Soares, 1321 J-30
CEP: 60.811-341, Fortaleza, CE, Brasil
porfírio@unifor.br

ABSTRACT

Navigation in the Web can be improved when the user can receive information instead of only documents. The development of aggregated searches using semantic Web resources may contribute to a better navigation. The adoption of Web services may be a support technology to supply data to such navigation. Web services have become a standard of fact and law in the whole computational media, since its concepts are simple and functional. Along with this model, several proposals have been built to add semantics to Web services. However this specification still suffers from a lack of standardization in order to describe their syntactic and semantic services. Furthermore, despite their evolution, the discover and composition of Web services to execute a task is a field still open. In this paper, we present our proposal for a research about an architecture for semantic navigation in the Web of data. Additionally, we will propose a model for semantic Web services discovery adopting SERIN semantic interfaces. SERIN, an extension of Active OWL model proposed for semantic Web services SOAP, applying it in the context of RESTful model. SERIN defines conceptual models that describe Web services syntactically and semantically. Finally, we intend to evaluate the proposal architecture in a government open data environment.

Categories and Subject Descriptors

H.3.3 [Information Systems]: Information Search and Retrieval—*Selection process*

; H.3.5 [Information Systems]: Online Information Services—*Web-based services*

; H.5.4 [Information Systems]: Hypertext/Hypermedia—*Navigation*

General Terms

Standardization

Keywords

navegação semântica, serviços web semânticos, descoberta de serviços web, dados abertos governamentais

1. CONTEXTO TEÓRICO

O advento de serviços Web permitiu a construção de um ambiente computacional distribuído com capacidade de fornecer dados e serviços que podem ser acessados por máquinas. As dificuldades encontradas pelas máquinas no uso dos serviços disponíveis está na descoberta e composição automáticas dos serviços de modo que possam executar uma determinada tarefa. As pesquisas voltadas para essa área[16, 18, 8], em geral, procuram desenvolver algoritmos para encontrar um casamento de padrões (*matching*) a partir da descrição do serviço em comparação ao objetivo da tarefa solicitada. A descrição do serviço é analisada em função de suas entradas, saídas e critérios de pré e pós condições.

O avanço das pesquisas em Web semântica tem contribuído para a evolução nas pesquisas sobre descoberta de serviços ao inserir uma camada semântica para descrição dos serviços. Essa camada adicional na descrição dos serviços permite que se faça inferências sobre o serviço expandindo o conjunto de possibilidades de construção dos algoritmos de descoberta. Atualmente, muitas pesquisas adotam padrões da web semântica para descrição de serviços Web[13, 1, 11, 12] e para construção de algoritmos de descoberta desses serviços[16, 18, 8].

A proposta da Web semântica[3] é anotar semanticamente os dados disponíveis na web de forma que sejam reconhecíveis por máquinas. Uma tecnologia chave para a adoção da Web semântica é o uso de ontologias. A definição mais comumente aceita na literatura define ontologia como uma especificação explícita e formal de uma conceitualização compartilhada[9]. Na prática, são definidos conceitos através da especificação de classes, propriedades e indivíduos que os representam. Ao serem compartilhados, esses conceitos passam a representar um conhecimento comum. Dentre as diversas tecnologias adotadas na Web semântica, destacam-se o RDF¹ e o OWL², linguagens usadas para representação dos dados, e o SPARQL³, linguagem utilizada para executar consultas aos dados armazenados em formato RDF.

A Web semântica forma uma Web de dados. As tecnologias

¹<http://www.w3.org/RDF/>

²<http://www.w3.org/TR/owl-features/>

³<http://www.w3.org/TR/rdf-sparql-query/>

que a implementam formam um ambiente no qual é possível acessar e fazer inferências por esses dados. Além de estabelecer um formato padrão para os dados, para construir essa Web de dados, é necessário estabelecer relacionamentos entre eles. Assim, define-se como *Linked Data*⁴ a coleção de dados interrelacionados na Web.

Com o crescimento dos dados estruturados e disponibilizados em formatos aderentes à Web semântica, surgiu a necessidade de melhorar a apresentação desses dados para os usuários que navegam pela Web. Uma forma bastante comum de navegação na Web é fazer uso de ferramentas de pesquisa textual para encontrar documentos de interesse do usuário. Atualmente, essas ferramentas de pesquisa estão sendo evoluídas para apresentar resultados mais elaborados que uma simples lista de documentos. A intenção é passar a fornecer respostas aos usuários em vez de apenas documentos. Firmou-se então o conceito de consulta agregada, que direciona a tarefa de pesquisar e formatar a informação obtida de diversas fontes e exibi-las em uma interface única[15]. Exemplos de ferramentas que adotam o paradigma de consulta agregada são Yahoo! Alpha⁵, Wolfram Alpha⁶, Google Knowledge Graph⁷ e Sig.ma[20].

Todas essas tecnologias encontram uma aplicação bastante relevante na construção e disponibilização de dados abertos governamentais⁸. Segundo definição da Open Knowledge Definition⁹, para serem considerados abertos, os dados precisam estar disponíveis para que as pessoas possam usar, reusar e redistribuir livremente, tendo, no máximo, a obrigação de creditar a autoria dos dados e de compartilhar segundo as mesmas regras. Para isso, um grupo de trabalho¹⁰ definiu um conjunto de princípios que regem os dados abertos.

2. IDENTIFICAÇÃO DO PROBLEMA

Com a popularização da Web 2.0, modificou-se a forma de interação do usuário com o ambiente Web. Atualmente, a navegação pela Web vai muito além da simples recuperação de documentos a partir de uma lista de resultados. Mais do que fornecer documentos, a Web deve passar a fornecer dados e informações a partir de questionamentos dos usuários. Esse questionamento pode surgir a partir da navegação do usuário pelas páginas da Web. Assim, a partir da leitura de um documento na Web, o usuário pode desejar obter informações sobre os conceitos encontrados no documento. Apesar dos avanços nessa área, a construção de consultas agregadas, associadas à navegação pelos documentos Web, que tragam informações relevantes aos usuários ainda é uma área de pesquisa em aberto.

Um passo importante para contribuir com a construção dessa navegação é a identificação de serviços Web para compô-la. A descoberta de serviços web relevantes que possam compor

⁴<http://www.w3.org/standards/semanticweb/data>

⁵<http://au.alpha.yahoo.com>

⁶<http://www.wolframalpha.com/>

⁷<http://www.google.com/insidesearch/features/search/knowledge.html>

⁸<http://www.w3.org/TR/2009/WD-gov-data-20090908/>

⁹<http://opendefinition.org/>

¹⁰<http://www.opengovdata.org/>

um conjunto de serviços para executar determinada tarefa é uma área de intensa pesquisa. Uma dificuldade na seleção dos serviços é encontrar métricas ou algoritmos que façam o casamento entre a descrição do serviço Web e a especificação da tarefa a ser executada. Esse casamento deve ter um grau de precisão suficiente para que uma máquina possa executar uma tarefa a partir dos serviços selecionados fornecendo a resposta esperada pelo usuário. A maioria das pesquisas relacionadas a descoberta de serviços procura identificar métricas e algoritmos que aumentem essa precisão.

Outra dificuldade em relação aos serviços Web é a falta de um padrão que seja compartilhado ou uma descrição efetiva que possa ser reconhecida pelos agentes de software. Apesar da intensa pesquisa sobre a descrição dos serviços Web, ainda há necessidade de se estabelecer padrões para facilitar a descoberta dos serviços.

Especificamente na área de dados governamentais, apesar dos esforços de alguns grupos, os dados abertos disponibilizados não adotam integralmente os padrões da Web semântica e carecem de um padrão compartilhado. A construção de mecanismos de padronização pode contribuir para a disseminação dos dados abertos e para a intensificação do desenvolvimento de sistemas que levem esses dados para a sociedade.

3. OBJETIVOS

A pesquisa tem como objetivo geral definir e implementar uma arquitetura de navegação semântica na Web, utilizando a descoberta de serviços semânticos para recuperar dados disponíveis em fontes diversas e compor uma consulta agregada para o usuário. O desenvolvimento dessa arquitetura pretende evoluir a forma de navegação na Web de uma simples recuperação e apresentação de documentos Web para recuperação e exibição de informações.

A partir do objetivo geral, pretende-se alcançar os seguintes objetivos específicos:

- Estender a proposta de navegação semântica, desenvolvida durante o mestrado, para possibilitar a construção de uma consulta agregada de dados provenientes de diversas fontes a partir da navegação do usuário nas páginas Web.
- Definir um mecanismo de descoberta de serviços web semânticos que fornecerão os dados para a consulta agregada.
- Construir uma interface semântica que represente os dados abertos governamentais oferecendo um mecanismo para disponibilizar e padronizar os dados publicados pelas diversas fontes governamentais.
- Avaliar a aplicação da arquitetura proposta em um cenário de governo eletrônico de forma a comprovar a utilidade da navegação semântica para usuários que busquem informações sobre o governo.

4. CONTRIBUIÇÕES ESPERADAS

As contribuições esperadas deste trabalho são: a construção de uma interface semântica para os dados governamentais; uma proposta de padronização de serviços web semânticos para divulgação dos dados governamentais; um modelo de descoberta de serviços web semânticos; a construção de uma arquitetura para navegação na Web utilizando recursos semânticos para prover resultados significativos que vão além da simples recuperação de documentos; a avaliação da utilidade de se ter os dados governamentais, fornecidos por um conjunto heterogêneo de órgãos, distribuídos na Web de forma padronizada e adotando um vocabulário comum.

5. METODOLOGIA

Inicialmente foi feito o levantamento do referencial teórico através de pesquisa na literatura disponível sobre os assuntos a serem trabalhados. Esse levantamento subsidiou uma série de debates do grupo de pesquisa para definir uma proposta.

Em seguida, foi feita a especificação da interface semântica e da proposta de interface SERIN. Foi modelada uma ontologia para representar a interface SERIN e construiu-se uma programa cliente e um programa servidor que foram usados como protótipos para avaliar o comportamento da interface.

A seguir, será definida a especificação da arquitetura de descoberta de serviços Web semânticos e da navegação semântica. Para essa especificação, será feito um levantamento e escolha das tecnologias disponíveis que possam dar suporte à implementação da arquitetura.

Após a definição da arquitetura, serão modelados os cenários de uso nos quais a arquitetura proposta será avaliada. Esses cenários serão criados a partir do levantamento de cenários reais de uso dos dados governamentais disponíveis. O conjunto de dados abertos disponíveis será coletado e mapeado para construção de ontologias que possam representar os dados e de serviços Web que irão disponibilizá-los na Web. Esses dados serão utilizados para construção de aplicações exemplo que simulem a existência de diversos provedores de serviços Web semânticos que forneçam os dados abertos.

Nos cenários propostos, serão avaliadas a efetividade da construção das consultas agregadas, com dados obtidos a partir dos serviços disponíveis, e a percepção de sua utilidade pelos usuários. Também será medido o ganho em relação à coleta dos dados com a adoção das interfaces semânticas.

6. SITUAÇÃO ATUAL DO TRABALHO

Até o momento, foi realizado um estudo da literatura a respeito de interfaces semânticas e descoberta de serviços web semânticos.

A interface semântica SERIN foi definida, proposta e apresentada[7]. Para esse trabalho, foi construída uma ontologia e uma aplicação de exemplo que faz uso de interfaces semânticas. Além da especificação da interface semântica, o SERIN propõe também uma padronização de endereçamento para os serviços Web semânticos.

Foi também construído e apresentado um trabalho preliminar abordando a proposta de descoberta de serviços semânticos[4] adotando a interface semântica SERIN. Essa parte da pesquisa encontra-se, no momento, em andamento.

A pesquisa sobre navegação semântica foi objeto da dissertação de mestrado e foi apresentada em diversas publicações[5, 21, 6]. A pesquisa atual pretende evoluir essa proposta de navegação semântica. A arquitetura está sendo revista para contemplar o uso de interfaces semânticas, a descoberta de serviços Web semânticos e a forma de apresentação dos resultados para o usuário.

A construção das aplicações e a aplicação delas no cenário de uso de dados governamentais irá se iniciar após a construção da arquitetura. A utilização de dados governamentais está em processo de negociação junto à entidades do governo.

7. TRABALHOS RELACIONADOS

A área de pesquisa sobre descoberta de serviços Web semânticos é bastante extensa. Além dos trabalhos citados na seção 1, podemos citar um conjunto grande de outras fontes de pesquisa, dentre as quais menciona-se apenas algumas [14, 19, 17]. A arquitetura proposta nesta pesquisa propõe o uso de interfaces semânticas como alternativa para evitar a necessidade de algoritmos de *matching* normalmente encontrados nas demais pesquisas sobre o assunto.

A proposta de navegação semântica encontra paralelo em pesquisas que adotam tecnologias relacionadas a Linked Data e a navegação sobre grafos RDF além de ferramentas que se propõe a exibir consultas agregadas. Exemplos de ferramentas que navegam por grafos RDF são Visinav[10] e Tabulator[2]. Google knowledge Search, Wolfram Alpha e Sig.ma, citados na seção 1 são alguns exemplos de ferramentas que se propõe a construir consultas agregadas. Algumas ferramentas de consulta agregada, a exemplo da Wolfram Alpha, não adotam semântica para construir seus resultados. A ferramentas pesquisadas partem de uma interface de pesquisa por uma palavra-chave para construir uma consulta agregada como resultado da pesquisa. Em nossa proposta, pretende-se associar a navegação pelas páginas Web, para identificar conceitos de uma ontologia nessas páginas, à navegação pelos dados disponíveis em grafos RDF para construir as consultas agregadas buscando esses conceitos e dados nos serviços web semânticos.

Em relação à pesquisa sobre dados abertos governamentais, os Governos de diversos países tem despendido esforços para fomentar a construção e adoção de recursos de governo eletrônico, isto é, a disponibilização dos dados governamentais na Web para consulta e uso por seus cidadãos. Essa proposta é fortemente apoiada por diversos governos, a exemplo da Inglaterra¹¹, dos Estados Unidos¹² e do Brasil¹³.

8. AGRADECIMENTOS

Os autores agradecem ao SERPRO, Serviço Federal de Processamento de Dados, pelo apoio financeiro, em forma de bolsa, para a pesquisa.

9. REFERENCES

- [1] A. Alowisheq and D. E. Millard. EXPRESS: EXPressing REstful Semantic Services. *Web Intelligence and Intelligent Agent Technology*,

¹¹<http://data.gov.uk/>

¹²<http://www.data.gov/>

¹³<http://www.governoeletronico.gov.br/>

- IEEE/WIC/ACM International Conference on*, 3:453–456, 2009.
- [2] T. Berners-lee, Y. Chen, L. Chilton, D. Connolly, R. Dhanaraj, J. Hollenbach, A. Lerer, and D. Sheets. Tabulator: Exploring and analyzing linked data on the semantic web. In *In Proceedings of the 3rd International Semantic Web User Interaction Workshop*, 2006.
- [3] T. Berners-Lee, J. Hendler, and O. Lassila. The semantic Web. *Scientific American*, 284(5):34–43, 2001.
- [4] L. M. Chaves, J. R. V. Dantas, B. de Azevedo Muniz, J. C. C. Neto, and P. P. M. Farias. Semantic Interfaces Discovery Server. pages 80–87, Hotel Sofitel, Redwood City, San Francisco Bay,, 2012. Knowledge Systems Institute Graduate School.
- [5] J. R. V. Dantas and P. P. M. Farias. Navegação conceitual em ambientes de gestão do conhecimento usando NavCon. In *Proceedings of the 14th Brazilian Symposium on Multimedia and the Web, WebMedia '08*, pages 250–257, New York, NY, USA, 2008. ACM.
- [6] J. R. V. Dantas and P. P. M. Farias. Using NavCon for conceptual navigation in web documents. *Int. J. Adv. Media Commun.*, 3(4):421–439, Sept. 2009.
- [7] B. de Azevedo Muniz, L. M. Chaves, J. C. C. Neto, J. R. V. Dantas, and P. P. M. Farias. SERIN - SEMANTIC RESTFUL INTERFACES. In *Proceedings of the IADIS International Conference on WWW/Internet*, pages 463–467, Rio de Janeiro, Brazil, 2011. IADIS Press.
- [8] T. A. Farrag, A. I. Saleh, and H. Ali. Semantic web services matchmaking: Semantic distance-based approach. *Computers & Electrical Engineering*, 39(2):497–511, 2013.
- [9] T. R. Gruber. A Translation Approach to Portable Ontology Specification. Technical Report KSL 92-71, Stanford University, Stanford, 1993.
- [10] A. Harth. Visinav: A system for visual search and navigation on web data. *Web Semantics: Science, Services and Agents on the World Wide Web*, 8(4), 2010.
- [11] J. Kopecky, T. Vitvar, and D. Fensel. MICROWSMO: SEMANTIC DESCRIPTION OF RESTFUL SERVICES. Technical Report D38v0.1 MicroWSMO, 2008.
- [12] Y.-J. Lee and K. Chang-Su. Building semantic ontologies for RESTful web services. pages 383–386, Krakow, Poland, 2010.
- [13] D. Martin, M. Burstein, J. Hobbs, O. Lassila, D. McDermott, S. McIlraith, S. Narayanan, M. Paolucci, B. Parsia, T. R. Payne, E. Sirin, N. Srinivasan, and K. Sycara. OWL-S: Semantic Markup for Web Services. 2004.
- [14] E. Meshkova, J. Riihijärvi, M. Petrova, and P. Mähönen. A survey on resource discovery mechanisms, peer-to-peer and service discovery frameworks. *Computer Networks*, 52(11):2097–2128, 2008.
- [15] V. Murdock and M. Lalmas. Workshop on aggregated search. *SIGIR Forum*, 42(2):80–83, Nov. 2008.
- [16] M. Paolucci, T. Kawamura, T. Payne, and K. Sycara. Semantic Matching of Web Services Capabilities. In I. Horrocks and J. Hendler, editors, *The Semantic Web — ISWC 2002*, volume 2342 of *Lecture Notes in Computer Science*, pages 333–347. Springer Berlin / Heidelberg, 2002. 10.1007/3-540-48005-6_26.
- [17] J. Sangers, F. Frasincar, F. Hogenboom, and V. Chepegin. Semantic web service discovery using natural language processing techniques. *Expert Systems with Applications*, 40(11):4660–4671, 2013.
- [18] M. L. Sbodio, D. Martin, and C. Moulin. Discovering Semantic Web services using SPARQL and intelligent agents. *Web Semant.*, 8(4):310–328, Nov. 2010.
- [19] H. N. Talantikite, D. Aissani, and N. Boudjlida. Semantic annotations for web services discovery and composition. *Computer Standards & Interfaces*, 31(6):1108–1117, 2009.
- [20] G. Tummarello, S. Danielczyk, R. Cyganiak, R. Delbru, M. Catasta, E. P. Federale, and S. Decker. Sig.ma: Live views on the Web of data. In *In Proc. WWW-2010*, pages 1301–1304. ACM Press, 2010.
- [21] J. R. Villela Dantas and P. P. Muniz Farias. Conceptual navigation in knowledge management environments using NavCon. *Inf. Process. Manage.*, 46(4):413–425, July 2010.