

SISTEMAS PARA WEB SEMÂNTICA: METODOLOGIAS E FERRAMENTAS PARA DESENVOLVIMENTO

*Isabela da Silva¹ (IC), Jane Adriane Gandra² (PQ), Ronaldo Ferreira da Silva³ (PQ).

e-mail: belamartins2015@gmail.com

Universidade Estadual de Goiás (UEG) – Câmpus Posse. Av. Senhora Santana, QD. 08, LT. Único. Posse-GO. CEP: 73900-000

Resumo: É notório o crescimento exponencial apresentado pela Web, conjuntura já prevista por seu criador, Tim Berners-Lee, que propôs a Web semântica uma nova camada da Web atual, a qual se atribuiria a informação o significado entendível pela máquina, desta forma ela terá a compreensão intuitiva semelhante a cognição humana, evitando a geração de informação não estruturada em massa, a chamada Sobrecarga da informação. Sistemas para Web semântica: metodologias e ferramentas para desenvolvimento é um projeto de pesquisa que possui o objetivo de analisar a representação do conhecimento atribuindo semântica à informação e demonstrar as principais metodologias para o desenvolvimento de sistemas sob as perspectivas da Web semântica. Esta pesquisa é uma revisão bibliográfica sobre a aplicação da Web semântica, na representação do conhecimento e organização da informação, no qual será desenvolvido uma *Proof of Concept* (PoC) para estruturação, armazenamento e recuperação em uma base de dados semântica, acessado por meio de uma aplicação com interface construída sob o protocolo *Hyper Text Transfer Protocol* (HTTP).

Palavras-chave: Compreensão das máquinas. Estruturação da informação. Semântica. Web.

V Introdução

Com o surgimento da Web, a forma de produzir, armazenar e disseminar informação transformou-se, o uso de hipertextos na Web, ganhou dimensões bem além das imagináveis por Ted Nelson em 1963, pioneiro da Tecnologia da Informação, e inventor dos termos hipertexto e hipermídia. A Web em seus primórdios era meramente documental, estática e pouco interativa, após o modelo proposto por Tim Berners-Lee, passando a adotar o modelo conhecido como *World Wide Web* (WWW), ocorre transformações importantes em sua evolução, desta forma possibilitou-se maior interatividade entre o usuário. Nicolau (2011) atribui como característica importante à segunda geração da Web, facilidades de

publicação, por meio de *blogs*, *sites Wikis*, *nuvens de tags*, *folksonomia* e redes sociais, o usuário passa a ser o provedor e consumidor da própria informação.

O aumento exponencial de produção e disponibilização de informação sem estrutura pelo usuário, gerou um grande volume de dados, que tornou a classificação e recuperação de informação extensa e difícil, Silva (2016, p. 5) defini a sobrecarga da informação quando ocorre “informações localizadas que não condizem com as pesquisas”, é comum observar em buscas de sites uma grande variedade de informações irrelevantes para o contexto da busca do usuário. Sobre a contextualização dessa Web Breitman *et al.* (2007), destacou a inercia de interpretação do computador que apenas apresenta a informação, enquanto o homem exigisse avaliar, classificar e selecionar o conhecimento.

A sobrecarga da informação é um grande problema enfrentado hoje pelos usuários, uma solução cabível para esta problemática foi proposta por Berners Lee *et al.* (2001), pela primeira vez no artigo da revista *Scientific American*, Web Semântica: um novo formato de conteúdo para a Web que tem significado para computadores vai iniciar uma revolução de novas possibilidades, ele afirma que a informação estruturada seja compartilhada já com significado, desta forma permitindo a cooperação homogênea entre homens e máquinas. Sobre isto Dziekaniak e Kirinus (2004, p.39), diz “é necessário que agentes percorram a rede, página a página para executar tarefas consideradas sofisticadas para o usuário. Esses agentes serão capazes de identificar o significado exato “.

A Web semântica não é uma nova Web e sim a extensão da Web atual (BERNERS-LEE, 2001), capaz de fazer os motores de busca compreender o significado e processar apenas as informações coerentes com a pesquisa do usuário, ela estrutura todo conteúdo disponível na Web sobre determinado domínio, para esse processo de interpretação do computador é necessário à inferência de vastas coleções de informação e conjuntos de regras que ajudem no processo de dedução automática para que seja administrado o raciocínio automatizado, necessitam das ontologias e linguagens padrões como *Resource Description Framework* (RDF) e *Extensible Markup Language* (XML).

Mas para incorporar a Web semântica nas futuras buscas e desenvolvimento do ciberespaço é necessário explorar as camadas criadas pelo *World Wide Web Consortium* (W3C), no intuito de padronizar as novas linguagens para criação de páginas embasada em conhecimento adquirido através das ontologias. Sobre esta

projeção Lima e Carvalho (2004) afirmam que a Web semântica poderá no futuro permitir que agentes, que não foram propriamente projetados para trabalhar conjuntamente possam transferir dados entre si.

A seguir a representação dessas camadas.

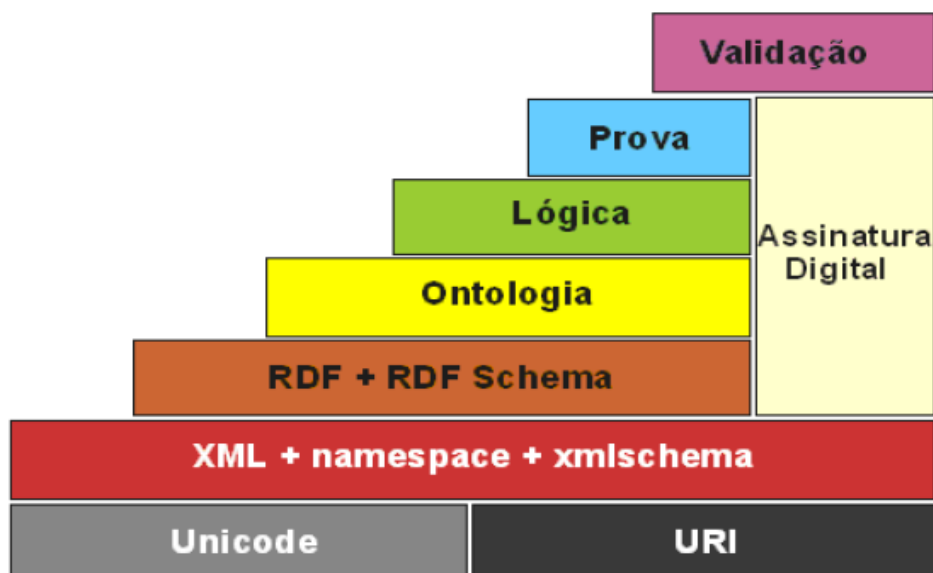


Figura 1: A visão da Web Semântica pelo W3C. Fonte: Uma visão da Web Semântica, 2004

Material e Métodos

A pesquisa apresenta uma abordagem qualitativa e os procedimentos utilizados para responder a questão de pesquisa foram bibliográficos. Com isso foi realizado um extenso levantamento bibliográfico concomitante com a busca das ferramentas e metodologias homologadas pela W3C.

Com a finalidade de demonstrar as implicações práticas em consonância com o levantamento bibliográfico e análise de ferramentas, a segunda parte do estudo será o desenvolvimento de uma prova de conceito, ou PoC (sigla do inglês, *Proof of Concept*). Para isso o aparato tecnológico foi construído contendo uma infraestrutura mínima contemplando um Servidor de Aplicações *Glassfish*, *Java Development Kit* (JDK) contendo um *framework Jena*, e para armazenamento dos dados será utilizado o banco de dados semântico *NoSQL AllegroGraph*. A prova de conceito consiste na criação e publicação de um portal semântico utilizando as ferramentas estudadas.

Sob supervisão e orientação dos coordenadores desta pesquisa, ocorre reuniões com a referida acadêmica-pesquisadora semanalmente, em desenvolvimento das atividades científicas: realização de um *survey* de artigos científicos, com o intuito de avaliar a evolução da Web semântica desde do seu início, análise das principais metodologias e ferramentas de desenvolvimento de soluções na visão da Web semântica, principais soluções utilizando os conceitos da Web semântica e como tem contribuído para organização e disponibilização da informação, elaboração de uma tabela com as principais ferramentas de infraestrutura, desenvolvimento e armazenamento de sistemas baseados na Web semântica, elaboração de uma tabela com os principais projetos desenvolvidos sob as concepções.

Resultados e Discussão

Os resultados parciais obtidos com levantamento bibliográfico pertinente ao tema mostraram que houve uma evolução significativa nas cinco primeiras camadas ilustradas na Figura 1: *URI*, *Unicode*, *XML*, *RDF* e ontologias, avaliação da evolução da Web semântica desde seu surgimento, que possibilitou a elaboração de uma tabela com as principais ferramentas de infraestrutura e outra com os principais projetos desenvolvidos sob as concepções da Web semântica, demonstrou que houve um aumento quantitativo e qualitativo das ferramentas e metodologias de desenvolvimento, considerando as seguintes categorias: *Integrated Development Environment* (IDE), banco de dados, e *frameworks*.

Apurou-se ainda o índice de pesquisas iniciadas desde 2004 a até o momento de desenvolvimento desta pesquisa, a evidente queda de interesse por projetos relacionados a Web semântica, ao menos é o que transparece quando as buscas são realizadas pelas palavras “Web semântica”. Na busca por ferramentas já utilizadas para atribuir significado a documentos da Web, as que aparecem com mais evidência são: os *frameworks Jena* e *Sesame* respectivamente utiliza linguagem Java e fornece uma API para extrair dados e projetar gráficos *RDF*. O seguinte oferece uma API para o usuário ter acesso a seus repositórios, bem como uma interface HTTP que suporta o protocolo *SPARQL*. Ambos com licenças gratuitas.

As IDEs *Semantic Java* tem a finalidade de oferecer um ambiente de desenvolvimento utilizando recursos da Web semântica com a linguagem Java. O

banco de dados *AllegroGraph* é um repositório de dados baseado em triplas RDF, utiliza uma eficiente combinação de armazenamento e desempenho. *OWLIM* Sistema de gerenciamento de banco de dados RDF, serve para organizar, armazenar e gerenciar conteúdo na forma de dados semânticos.

No levantamento de projetos utilizando a Web semântica destacou-se, a tecnologia *Friend of a Friend* (FOAF), que permite compartilhar e usar informações sobre os usuários pretende constituir páginas que descrevam o cotidiano humano que os computadores possam interpretar. *WolframAlpha* site *wiki*, suas respostas são formuladas por meio de uma interpretação do texto digitado pelo usuário, associado a um enorme banco de dados atualizado constantemente.

Ressalta-se que ainda faz parte do escopo deste projeto, a criação de um *Wiki* contendo todos os passos e ferramentas utilizados no desenvolvimento da prova de conceito.

Considerações Finais

O intuito deste projeto é mostrar as tecnologias englobando ferramentas e metodologias que surgem com a Web Semântica e como está o desenvolvimento atualmente, no qual no ano de 2006, teve seu ápice em pesquisas pertinentes a *Semantic Web*, e hoje tem o número de estudos relacionados apresentam-se reduzidos.

Constata-se uma grande dificuldade dos pesquisadores de evoluir no processo de desenvolvimento da Web, a partir das três últimas camadas da arquitetura da Web semântica, deve-se a essa dificuldade a ausência de estrutura definida para implementação, havendo a possibilidade de diferentes desenvolvedores usarem *tags* diferentes para descrever a mesma coisa.

A Web semântica é a ferramenta mais indicada para solucionar o problema da falta de semântica no conteúdo compartilhado da Web, o seu crescimento modesto ao longo desses anos é provocado indiretamente pela falta de interesse comercial das empresas, devido os custos de investimentos em novos projetos, pois é necessário profissionais qualificados, mas é evidente a essa nova realidade que se caminha a Web atual, pois a evolução tecnológica é constante, para se ter qualidade a oferecer ao usuário é imprescindível novos mecanismos e ferramentas como a Web semântica para acompanhar este desenvolvimento.

Agradecimentos

Expresso aqui os meus agradecimentos sinceros às muitas pessoas que direta e indiretamente colaboraram na elaboração deste projeto. A Deus por me proporcionar saúde e inteligência para superar todas as dificuldades. Ao meu orientador Ronaldo Silva, por todo o tempo que dedicou a me ajudar durante o processo de realização deste trabalho e a minha família pelo incentivo e apoio.

Referências

- BERNERS-LEE, T.; HENDLER, J.; LASSILA, O.; **The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities.** Disponível em: http://www-sop.inria.fr/acacia/fabien/lecture/licence_travaux_etude2002/TheSemanticWeb/. Acesso em 23 de jul. 2017.
- BREITMAN, K.K.; CASANOVA, M.A.; TRUSZKOWSKI W. **Semantic Web: Concepts, Technologies and Applications.** NASA Monographs in Systems and Software Engineering. Springer-Verlag London Limited, 2007.
- DZIEKANIAK, G. V.; KIRINIUS, J. B. **WEB SEMÂNTICA.** Florianópolis: Bibliotecon, 2004, n.18.
- LIMA, J. C.; CARVALHO, C. L. **Uma visão da Web Semântica.** Instituto de Informática. Universidade Federal de Goiás. 2004, 15 p.
- NICOLAU, M. A busca por uma web semântica cognitiva. **Revista de Comunicação e Epistemologia da Universidade Católica de Brasília.** n. 6, 2010. p. 138-154.
- SILVA, R. F. **A contribuição da web semântica em sistemas de inteligência coletiva.** P.16. 2016. Posse.