

Bibliotecas digitais: um campo de aplicação web semântica

Digital libraries: one semantic web application field

Thabyta Giraldelli Marsulo

Universidade Estadual Paulista – Júlio de Mesquita Filho
thabytagm@hotmail.com

Ângela Maria Grossi de Carvalho

Universidade Estadual Paulista – Júlio de Mesquita Filho
angela@carvalho.jor.br

José Eduardo Santarem Segundo

Universidade de São Paulo
santarem@usp.br

Resumo

O presente estudo tem como tema a web semântica, a web como base para a construção de ontologias e estruturação de bibliotecas digitais. O uso da Web Semântica permite que as bibliotecas digitais prestação de serviços muito mais personalizados gerando um impacto positivo sobre a produção de conhecimento e satisfação de seus usuários, impactando drasticamente a recuperação de conteúdo, obrigá-la a fazer novos ajustes para acomodar o uso das tecnologias ligadas subjacentes o projeto de web semântica. Supomos que as tecnologias da informação e da comunicação não são apenas ferramentas de aplicação no contexto da pesquisa da Ciência da Informação. Eles afetam muito o comportamento dos indivíduos e contribuir para o desenvolvimento da Sociedade da Informação. Ressalta-se que o trabalho corrompido uma biblioteca digital construído semanticamente, tentando relacionar estas iniciativas com a importância de profissional contexto da ciência da informação.

Abstract

The present study has as subject the web semantic,, the web as a basis for ontology construction and structuring of digital libraries. The use of Semantic Web allows digital libraries provides many more personalized services generating a positive impact on the production of knowledge and satisfaction of its users, dramatically impacting the recovery of content, requiring it to make new settings to accommodate the use of the linked underlying technologies the semantic web project. We assume that Information and Communication Technologies are not only application tools in the context of research of Information Science. They greatly affect the behavior of individuals and contribute to the development of the Information Society. We emphasize that the work Corrupteca a digital library built semantically, trying to relate these initiatives with the importance of professional information science context.

Palavras-chave: Web Semântica, Bibliotecas Digitais, Tecnologias de Comunicação e Informação. **Keywords:** *Semantic Web, Digital Libraries, Corruptec, Information and Communication Technologies.*

Introdução

Wurman considera que o produto da Era Digital é a informação e que seu meio de transporte é a Internet, “[...] o que significa mais desinformação, porque a informação errada pode ser transmitida tão facilmente quanto a certa. O apertar de um botão pode espalhar informações errôneas sobre você pelo mundo afora” (WURMAN, 2005, p. 13). Marcondes e Sayão (2002, p. 43) consideram, ainda, que,

As possibilidades abertas pela Internet com seus mecanismos de publicação direta na rede tornam o acesso a um documento digital uma mera questão de conhecer sua URL. No entanto, esta facilidade de acesso tem como contrapartida a grande dificuldade de encontrar informação relevante, as atividades de *information discovery*.

Reis (2012, p.1) completa afirmando que:

O fenômeno das redes sociais, do trabalho colaborativo e a presença dos media *online* em muito contribuem para o crescimento exponencial da informação criada e disponibilizada a cada instante. Deste modo, o ser humano está hoje exposto diariamente a uma enorme quantidade de informação, proveniente das mais diversas fontes, com a qual não consegue lidar. Por outro lado muita da informação gerada não interessa de igual forma a todas as pessoas.

Assim, pode-se considerar que tais fatores podem resultar na denominada ansiedade informacional, definida por Wurman como: “O resultado da distância cada vez maior entre o que compreendemos e o que achamos que deveríamos compreender. É o buraco negro que existe entre dados e conhecimento, e ocorre quando a informação não nos diz o que queremos ou precisamos saber “ (WURMAN, 2003, p. 38).

Segundo Davenport e Prusak (1998) o conhecimento é a informação agregada, pois existe diante do contexto vivenciado por alguém, que deu a aquela informação sua interpretação e significado de modo a tornar-se mais valioso.

A sociedade da informação abarca todas as ferramentas e todas as oportunidades que as tecnologias vem proporcionando as pessoas nos últimos anos, colocando a disposição uma grande quantidade de informações e múltiplas formas de acesso a elas. (Sebastiaán ; Flores, 2004)

A mudança no modo como as pessoas se relacionam com a informação e seu impacto na sociedade se baseia na essência da Internet: sua dinâmica e sua capacidade de extinção das fronteiras para possibilitar o acesso global a diversos tipos de informação e onde encontrar a informação relevante é fundamental para que a mesma possa ser utilizada.

A grande quantidade de informação disponível na *Web* contribuiu para que as pessoas passassem a ter uma diversidade de opções. Esta diversidade torna a pesquisa e até mesmo a escolha uma tarefa difícil e desafiadora. Por isto, para diminuir as dúvidas em relação a tantas alternativas, boa parte dos usuários apoia-se em recomendações, que podem ser dadas por meio de: cartas de recomendação, opiniões de críticos de filmes e livros, opiniões de pessoas famosas em blogs conceituados sobre determinados produtos, entre outros (BORGES; OLIVEIRA, 2010 p.1).

Objetivo

A *Web Semântica* não é uma *Web* separada, mas uma extensão da atual, na qual é dado um significado bem definido para a informação, permitindo que computadores e pessoas trabalhem em cooperação.

No contexto das bibliotecas digitais, as tecnologias da *Web semântica* tem papel importante à medida que possibilitam o acesso eficiente e inteligente aos documentos digitais na *web*.

A construção de ontologia construídas com uso de padrões de metadados para a descrição dos campos da ciência são bem executados a partir de estudos das comunidades usuárias.

Sendo assim busca-se apontar a importância da atuação do profissional da informação na construção das ontologias por meio dos estudos de usuários específicos das comunidades,

para que esta possibilite as bibliotecas digitais se estruturarem de maneira semântica possibilitando a recuperação mais precisa da informação.

Metodologia

A pesquisa se caracterizará como de tipo exploratória, descritiva bibliográfica e analítica, possuindo uma abordagem qualitativa. A pesquisa analítica será realizada a partir da observação direta não participativa no ambiente informacional da Corrupteca, a biblioteca digital do Núcleo de Pesquisas em Políticas Públicas da USP, pautada no projeto web semântica, no qual será observada os aspectos ligados a recuperação da informação, interface e características semântica.

Contextualização

A preocupação com a estruturação, representação e organização de conteúdos informacionais acompanhou o desenvolvimento cultural da humanidade há séculos. Contudo, é na contemporaneidade, com a ruptura das barreiras de tempo e espaço propiciadas pelas Tecnologias de Informação e Comunicação (TIC) que essa preocupação adquire importância ainda maior, principalmente, depois do aparecimento da Internet e do ambiente World Wide Web (Web).

Essas duas tecnologias, cujas interfaces computacionais vem interferindo nos modos de produzir e acessar informação e resultando na constante modificação do sistema híbrido - em que homens e máquinas, pelas suas formas de interação, modificam o fluxo informacional aumentando-o exponencialmente, favorecendo o efetivo acesso à informação e ao conhecimento, tornando-os bens indispensáveis para o desenvolvimento global da humanidade.

Na Web, a informação não está mais presa à estrutura linear característica da obra impressa. A morfologia de informação hoje textual, sonora, imagética estática ou em movimento em um mesmo ambiente multimidiático, permite que cada usuário interaja com o conteúdo informacional de acordo com o seu interesse, construindo seu próprio caminho de navegação pelos documentos e/ou parte deles utilizando-se dos links que ele próprio realiza.

A mudança no modo como às pessoas se relacionam com a informação e seu impacto na sociedade se baseia na essência da Internet: sua dinâmica e sua capacidade de extinção das fronteiras para possibilitar o acesso global a diversos tipos de informação.

Segundo O'Reilly o criador do termo web 2.0, esta se relaciona a revolução industrial da informática causada pela transformação da internet em plataforma e a disponibilização de serviços online que antes eram só prestados por softwares instalados.

Além das questões técnicas, a Web 2.0 possui suas vertentes sociais, através da abertura dos horizontes para uma grande quantidade de informação - o conteúdo passa a ser disponibilizado coletivamente trazendo o foco do usuário para a participação, a auto expressão e a liberdade de criação.

Assim, neste contexto, o usuário deixa a sua função exclusiva de consumidor da informação e passa a se comunicar com outros usuários conectados na Internet tendo a possibilidade de publicar conteúdos na rede, sem necessitar de conhecimentos profundos de programação, valorizando, principalmente, as práticas de cooperação, os diálogos e a participação na construção de conhecimento, de modo a potencializar os processos de trabalho coletivo, troca afetiva, produção e circulação de informações e, sobretudo, a construção social do conhecimento (PRIMO, 2007).

Esse novo papel do usuário, potencializou a criação de páginas e o excesso de conteúdos tornando as buscas menos precisas - o que está muitas vezes ligado a recuperação ilegível por programas, já que esta é uma web documental. Esta recuperação insuficiente está ligada a imersão de problemas como a alta sensibilidade à sintaxe e pouca precisão, onde os mecanismos de busca não levam em consideração a semântica do domínio, causando o retorno de inúmeras ocorrências irrelevantes ligadas a outros contextos.

A sociedade da informação abarca todas as ferramentas e todas as oportunidades que as tecnologias vem proporcionando as pessoas nos últimos anos, colocando a disposição uma grande quantidade de informações e múltiplas formas de acesso a elas. (Sebastiaán ; Flores, 2004)

A mudança no modo como as pessoas se relacionam com a informação e seu impacto na sociedade se baseia na essência da Internet: sua dinâmica e sua capacidade de extinção das fronteiras para possibilitar o acesso global a diversos tipos de informação e onde encontrar a informação relevante é fundamental para que a mesma possa ser utilizada.

A grande quantidade de informação disponível na *Web* contribuiu para que as pessoas passassem a ter uma diversidade de opções. Esta diversidade torna a pesquisa e até mesmo a escolha uma tarefa difícil e desafiadora. Por isto, para diminuir as dúvidas em relação a tantas alternativas, boa parte dos usuários apoia-se em recomendações, que podem ser dadas por meio de: cartas de recomendação, opiniões de críticos de filmes e livros, opiniões de pessoas famosas em blogs conceituados sobre determinados produtos, entre outros (BORGES; OLIVEIRA, 2010 p.1).

Muitas vezes o usuário possui pouca ou quase nenhuma experiência para realizar escolhas entre as várias alternativas que lhe são apresentadas o que aumenta muito a probabilidade desse usuário desistir de sua busca informacional. Neste contexto se o ambiente *online* trouxer a esse usuário um suporte em sua busca, indicando como este deveria procurar tal conteúdo, as chances de uma recuperação precisa aumentam.

Tais problemas afetam principalmente a organização dos recursos relevantes, pois perdidos dentre tantos recursos acabam não se tornando parte da rede lógica de domínio, o que traz grandes prejuízos aos usuários.

As páginas da Web foram construídas com semânticas locais (silos de informações) que não se falam. Este fato se constitui como o maior obstáculo para o seu desenvolvimento. Pensar em organizar o caos informacional disponível na Web, se tornou imperativo para possibilitar novas formas de acesso, e recuperação da informação virtual.

Segundo Saracevic (1995), “a explosão da informação eletrônica exige um novo olhar para o desenvolvimento dos tradicionais serviços de informação: seleção, tratamento e recuperação da informação”.

Entretanto, para que a Web Semântica torne-se uma concreta realidade, é necessário que haja uma padronização de tecnologias. Neste contexto, o W3C (World Wide Web Consortium) e a comunidade acadêmica têm unido esforços no sentido de criar, disponibilizar e padronizar linguagens e tecnologias para garantir a correta e melhor interoperabilidade possível. Sendo assim as tecnologias semânticas propõem avanços na questão da recuperação e processamento semântico de conteúdo, Santos e Alves (2009)

afirmam que a Web Semântica exige a união e o trabalho conjunto das várias ferramentas e consideram o uso dos metadados como essenciais para minimizar os problemas de representação e recuperação

Conforme afirmam Berners-Lee et al. (2001) a Web Semântica se apresenta como uma extensão da Web atual, onde a informação passa a possuir um significado claro e bem definido capaz de interagir sem barreiras com computadores e pessoas.

O projeto Web Semântica constitui-se como uma tentativa inversa de solução para a recuperação da informação, tendo como objetivo desenvolver meios para que as máquinas possam servir aos humanos de maneira mais eficiente, com base em instrumentos que fornecem sentido lógico e semântico aos computadores, como as ontologias.

O termo Ontologia vem da Filosofia e nesta divisão do saber, é empregada para descrever a existência do Ser no mundo. Para outras áreas o termo se liga ao modo de fracionar o mundo em frações compatíveis.

Já no contexto da web, Ramalho (2006, p.97), define a ontologia no contexto da Web Semântica como

Um artefato tecnológico que descreve um modelo conceitual de um determinado domínio em uma linguagem lógica e formal, a partir da descrição dos aspectos semânticos de conteúdos informacionais, possibilitando a realização de inferências automáticas por programas computacionais.

A aplicação das ontologias no universo da Web contribui para que a “interpretação” automática do conteúdo semântico de páginas dos sítios na Internet seja possível e amplia a colaboração entre máquinas e homens.

Para Schiessl e Bräscher (p. 217, 2012)

Ao contrário das ciências experimentais, que objetiva a descoberta e modelagem da realidade sob certa perspectiva, Ontologia se concentra na natureza e estrutura das coisas em si, independente de quaisquer outras considerações, até mesmo se elas realmente existem. Por exemplo, uma ontologia de unicórnios ou qualquer coisa fictícia: embora não existam de fato, a natureza e estrutura delas podem ser descritas em termos de relações e categorias gerais.

As ontologias são planejadas e construídas pelo homem para atender suas necessidades de informação em relação à organização, à estruturação e à recuperação. Dessa forma, elas são

parte integrante de um sistema formal que de alguma maneira necessita de uma forma de representação.

Portanto, a noção de ontologia como forma de particularização é fundamental na elaboração conceitual e na construção da arquitetura da Web semântica, pois o estabelecimento de limites a conceitos específicos e a definição das relações entre estes são essenciais para que a máquina possa “inferir” o significado da informação no contexto do em que o humano se insere.

Além disso, as ontologias são desenvolvidas como propósito bem definido de representar uma visão específica de um domínio por meio de vocabulário e esta representação pode ser contraditória em relação a outros domínios. Portanto, o caráter extensível e a reutilização das ontologias podem sofrer sérias restrições. (Schiessl e Bräscher , p. 219, 2012)

Atualmente, para a Ciência da Informação, o termo ontologia não se refere apenas a uma representação explicativa e sistemática da existência, apresenta-se como a formalização e explicitação de conceitos de determinado dos diversos domínios do atendendo a necessidades dos grupos de usuários.

As tecnologias semânticas estabelecem uma estreita relação com a representação do conhecimento, assim o campo da Ciência da Informação se volta à investigação do atual projeto da Web Semântica contribuindo para a organização e modelagem de domínios de conhecimento.

No âmbito do projeto Web Semântica também se observa uma necessidade de familiarização dos profissionais da informação com as novas tecnologias, para que estes possam desenvolver as atividades ligadas a representação do conhecimento a partir de princípios éticos-sociais e não baseadas única e exclusivamente em conhecimentos e processos puramente técnicos.

Afinal, as tecnologias subjacentes ao projeto Web Semântica e os instrumentos de representação de informações desenvolvidos no âmbito da área de Ciência da Informação possuem como objetivo comum, propiciar meios mais adequados de representar e organizar conteúdos informacionais, possibilitando responder de maneira mais eficiente às buscas realizadas diretamente pelos usuários finais.

A atuação do profissional da informação está pautada, em grande parte, nas habilidades e competências adquiridas por ele no decorrer de sua formação e atuação, sendo assim cabe a ele buscar continuamente a atualização e o aperfeiçoamento, para o desenvolvimento das competências necessárias para as transições do mercado exigidas pela sociedade.

Segundo Garcia Jimenez (2004, p.90,): “As relações entre ontologias e tesouros parecem demonstrar uma tendência evidente: a intenção por parte de diversos especialistas em elaborar determinadas ontologias a partir de um tesouro”. Seguindo esta tendência, verifica-se também o crescente desenvolvimento de projetos no âmbito de bibliotecas digitais que utilizam-se das tecnologias relacionadas ao projeto Web Semântica.

Resultados

A biblioteca digital representa uma evolução do resultado da utilização do computador e da Internet nos serviços da biblioteca e pode ser conceituada como uma coleção organizada de informação, em formato digital, acessível pela rede mundial de computadores.¹ com o objetivo de atender às necessidades de comunidades de usuários (Gonçalves et al. 2004).

A literatura aponta vários benefícios para a implantação de uma biblioteca digital, porém, enfatizamos a possibilidade de acesso remoto a informação, pois “pela primeira vez as bibliotecas dispõem de um meio de difusão e comunicação que ultrapassa as fronteiras da comunidade tradicionalmente servida” (PEREIRA, 2009, p. 13).

O avanço proporcionado pelas tecnologias informacionais e comunicacionais (TIC) influencia a prática dos profissionais da informação no direcionamento e execução de suas atividades, pois muitas vezes, sem a devida reflexão e consciência do uso dos recursos tecnológicos, os profissionais acabam adotando a tecnologia, esquecendo os subsídios teóricos que fundamentam suas produções por meio dos serviços, ferramentas e técnicas existentes há anos. Por consequência, fica a impressão de que serviços tradicionais sumiram ou ficaram obsoletos, quando na realidade ocorreu o contrário, pois apesar da inserção dessas tecnologias afiguram-se como inovação, sabemos que sua natureza está vinculada à tradição e a missão das bibliotecas,

¹ GONÇALVES, M. A. **Streams, Structures, Spaces, Scenarios, and Societies (5S): a formal digital library framework and its applications.** 2004. Tese (Doutorado em Ciência da Computação) - *Virginia Polytechnic Institute and State University.*

como no caso do serviço de Disseminação Seletiva da Informação (DSI) e os recursos utilizados com a tecnologia RSS aplicada ao ambiente de bibliotecas (BASTOS , 2011, p.4).

Inseridas dentre os sistemas de informação idealizados para atender as necessidades do ser humano e proporcionar a transferência de conhecimentos, as bibliotecas digitais trazem consigo muitos empecilhos ao que se refere a sua integração a outros ambientes:

Os usuários finais exigem interfaces simples que os permitam acessar coleção de recursos digitais disponíveis na biblioteca. Além de querer uma informação bibliográfica enriquecida, não se conformando apenas com um simples registro do OPAC, desejando que estes tenham conexões com outras fontes existentes , para obter documentos mais completos ou buscar mais registros do mesmo autor , assunto em outros ambientes web(SEBASTIAÁN ; FLORES, 2004 , p.197 – tradução nossa).

Assim percebemos que a biblioteca digital na *Web* não se cerca apenas das necessidades do usuário em consumir conteúdos, mas também deve promover a troca colaborativa de informações, que permitirá a sua manutenção constante através do *feedback* dos próprios usuários que interagem com a biblioteca, modificando o modelo tradicional, onde o usuário deveria se aproximar da biblioteca, criando a relação inversa que faz com que a biblioteca se aproxime do usuário (BASTOS , 2011).

Considerando a informação como objeto de trabalho do profissional da informação, em especial do bibliotecário, são importantes, diante do fenômeno de crescimento informacional, estudos que abranjam tanto os ambientes informacionais com os quais os usuários interagem, como a forma com a qual é realizada a descrição dessas informações compartilhadas.

Os novos princípios de interação e colaboração sugerem novas necessidades com relação à organização, disponibilização e estruturação dos ambientes informacionais digitais, fazendo necessários estudos acerca da Arquitetura da Informação Digital Colaborativa que busquem o desenvolvimento de sistemas mais personalizados.

Desde meados do século passado, a Ciência da Informação vem aprimorando seus fundamentos teóricos metodológicos de forma a compreender melhor os padrões de comportamento de busca de informação dos seres humanos, afim de elaborar o design de

sistemas de informação de modo a fazer com que tais sistemas atendam melhor seus usuários.

O usuário é fundamental na construção de um ambiente informacional, principalmente para a identificação dos serviços que serão oferecidos no mesmo. Em uma AI ou em qualquer processo de desenvolvimento de ambientes digitais, a identificação do público-alvo é requerida já no início do projeto (CAMARGO; VIDOTTI, 2011, p. 55).

Os estudos de usuários, necessidades e usos da informação da Ciência da Informação, começaram no final da década de 1940 com a apresentação dos trabalhos de Bernal (1948) e Urquhart (1948), que procuravam compreender o modo como cientistas e técnicos buscavam informação e como utilizavam a literatura nas suas respectivas áreas. Desde então, tais estudos tornaram-se um objeto recorrente nessa ciência buscando encontrar uma teoria e metodologia para suportá-los. Silva; Ferreira; Borges (2002, p. 132) citam que “a caracterização da necessidade de informação é um campo clássico da Ciência da Informação e da Biblioteconomia”. Apesar de ser um campo muito estudado, essas autoras afirmam que ainda não existe consenso sobre a forma mais indicada para realizá-los.

Esses estudos atravessaram uma mudança de paradigma ao longo dos últimos 60 anos. Segundo Ferreira (1997, p. 17), durante esse período, os estudos de necessidade e uso da informação evoluíram de uma abordagem centrada no sistema para uma abordagem centrada no usuário, chamadas respectivamente de abordagem tradicional e abordagem alternativa. Essa mudança de paradigma nasceu da preocupação dos novos pesquisadores em compreender os aspectos cognitivos e comportamentais dos usuários quando estão buscando informação (DERVIN; NILAN, 1986, p. 7).

A partir da década de 1970 surgiram críticas aos paradigmas da abordagem tradicional, porque ela não considera as necessidades do usuário ao buscar a informação. Para o usuário, em especial, a informação é uma ferramenta que resolve seus problemas porque lhe transmite o conhecimento que lhe falta para realizar suas necessidades. Compreender estas necessidades é entender a causa que o motivou a buscar informações.

Wilson (2000, p. 11) confirma isto ao afirmar que uma pessoa começa um processo de busca de informação simplesmente para atender suas necessidades, que são motivadas por fatores fisiológicos, emocionais ou cognitivos - todos externos ao sistema de informação.

Esse autor enfatiza essa visão construtivista da informação ao sugerir que o termo “necessidade de informação” seja substituído pelo termo “busca de informação para satisfazer necessidades”.

Para Dervin; Nilan (1986, p. 9), a partir dessa preocupação em entender a necessidade de informação, mudaram o paradigma da abordagem tradicional, trazendo uma nova visão aos pesquisadores. Nessa nova visão, o usuário de um sistema de informação passa a ser observado como um cliente e o seu comportamento e processos cognitivos tornam-se o foco de estudos.

Podemos ver um exemplo de iniciativa semântica na biblioteca do Núcleo de pesquisa de políticas públicas da USP – a Corrupteca.

O NUPPs - Núcleo de Pesquisa de Políticas Públicas da Universidade de São Paulo é a continuação institucional ampliada do antigo Núcleo de Pesquisa do Ensino Superior – NUPES, o qual foi criado em 1989 e se especializou no diagnóstico e análise de políticas para o ensino superior, bem como das estruturas das universidades públicas e dos seus resultados.

Em 2004/2005, por iniciativa da presidente do seu Conselho Diretor, Professora Eunice Ribeiro Durham (FFLCH), e sob a coordenação do Pró-reitor de Pesquisa à época, Professor Luiz Nunes, transformou-se no atual NUPPs, ampliando o escopo de seus objetivos.

O grupo dedica-se ao estudo e análise de programas de políticas públicas sob o ângulo da relação entre governança democrática, cidadania e desigualdades, analisando as etapas de formulação e implementação das políticas públicas e, principalmente, a que se refere aos seus resultados, portanto um de seus objetivos principais, é a análise da qualidade da gestão das políticas públicas no país, priorizando, na atualidade, o estudo das políticas sociais, a exemplo das políticas educacionais e dos programas de transferência de renda, assim como a questão da relação dos cidadãos com as instituições democráticas. Além destas, no âmbito das políticas de Estado, o Núcleo desenvolve também projetos de pesquisa na área de reformas econômicas, de segurança e criminalidade, dos efeitos domésticos da política externa brasileira e da participação cidadã em processos de tomada de decisões públicas.

Em 2013 o Núcleo de Pesquisas em Políticas Públicas (NUPPS) da USP fez uma parceria com o Estadão e o *Open Archives Initiative* (OAI), atualizando sua biblioteca digital para sua versão 2.0, a nova Corrupteca - nome dado ao ambiente informacional digital apresentado na figura 1. Além das inovações tecnológicas, teve seu acervo ampliado com novos casos de corrupção, que passaram a ser contextualizados em linhas do tempo

Figura 1 – ***Linha do tempo com os desdobramentos do caso mensalão***



A Corrupteca foi idealizada pelo pesquisador Giovanni Eldasi, um dos criadores do OAI, que fez a implantação tecnológica do projeto, e pelo professor José Álvaro Moisés, diretor científico do NUPPS. O projeto funciona como uma biblioteca digital, é algo contínuo, a longo prazo.

O objetivo inicial não foi só reunir todas as informações possíveis sobre casos de corrupção no Brasil, mas construir uma plataforma que indique não apenas os fatos e fenômenos da corrupção, trazendo também análises que permitam indicar as consequências dela por meio da ferramenta analítica capaz de fornecer aos estudantes e pesquisadores do tema com material empírico.

A versão da “Corrupteca 2.0” apresenta uma interface diferenciada, como apresentada na Fig. 2 - primeiramente uma área para jornalistas e o público em geral, com ferramentas de acesso rápido às informações, onde os usuários conseguem um resumo de todo o caso, as personagens envolvidas, com a situação atual de cada um dos casos de estudo, além do currículo destes participantes.

Figura 2 – ***Linha do tempo com os desdobramentos do caso mensalão***

The screenshot shows the website interface for Corrupteca. At the top, there is a navigation menu with links: A CORRUPTECA, PARCERIAS, PLANO INSTITUCIONAL, EQUIPE, and CONTATO. Below the menu, the Corrupteca logo is on the left and the NUPPs logo is on the right. A search bar is present with the text "BUSQUE NOS ACERVOS" and options for "Artigos Científicos" and "Acervo de Notícias". The main content area features several article teasers:

- Corrupteca avança em seu Plano Institucional**: Uma nova interface é lançada visando uma maior comunicação com o visitante. São iniciadas várias coleções, com destaque para a "Coleção Casos".
- Coleção Nacional**: Destaques da Pesquisa Brasileira sobre Corrupção. Teaser for "Descontrole nos Tribunais de Contas" with a sub-headline: "Tribunais de Contas fortemente marcados pela politização e pouco criteriosos na admissão de seus conselheiros parecem ser coniventes com irregularidades do Poder Executivo."
- Coleção Didática**: Saiba mais sobre Política e Cidadania. Teaser for "O que é Lavagem de Dinheiro?" with a sub-headline: "A expressão 'lavagem de dinheiro' vem do inglês Money Laundering."
- Conselho Federal de Biblioteconomia noticia a Corrupteca**: O conselho representativo dos bibliotecários brasileiros publicou matéria sobre a Corrupteca.
- Fapesp destaca a Corrupteca**: A Fundação de Amparo à Pesquisa do Estado de São Paulo noticiou a Biblioteca Internacional da

O site traz algo em torno de 100 mil volumes digitais de textos integrais, 48.567 periódicos científicos, provindos de 1.643 universidades e núcleos de pesquisas por 63 países de todo o globo – países que têm consórcio com a OAI.

Uma das inovações importantes na pesquisa dos textos com a Corrupteca é o uso da “Web Semântica”, os conceitos semânticos são levados em conta para a recuperação de artigos que contenham o termo, mas pelo contexto e relevância que a palavra e seu significado tem em todos os textos, como visto na figura 3.

Figura 3 - **As barras determinam o quanto o termo pesquisado se relaciona ao conteúdo geral da notícia e com o termo usado para a busca**

← → ↻ 🏠 nupps.usp.br/corrupteca/?page_id=221

Importado do IE

Acervo de Notícias
Acervo Jornalístico sobre Corrupção

mensalão PESQUISAR

349 itens encontrados em 3.98575 segundos

Histórico
mensalão(349)

Origens
Acervo do Estadão

1. **Jornal "O Estado de São Paulo" de 19/06/2005**
Agência Estado;
NUPPS-USP
NOTÍCIA: PUBLICADO EM 2005. PORTUGUÊS
RELEVÂNCIA NA PESQUISA 29.725%
#CORRUPÇÃO:
...da bandalheira Mensalão, um gigante pela própria natureza "aumentativo em 'ão'" tende a pegar no País, mesmo quando seu significado escandaliza a realidade. O QUE VIRA – Para Giannotti, impossível prever o futuro. Lula ainda pode virar o jogo. Ou deixar para seu sucessor a batata quente de uma reforma na burocracia de governo. Filósofo, professor emérito da USP, pesquisador do Centro Brasileiro de Análise e Planejamento (Cebap) TIAGO QUEIROZ/AF. Mario Sérgio Conti Não foi o deputado Roberto Jefferson quem inventou o "men-salão". Foi a Receita Federal. A palavra se encontra nos manuais do Ministério da Fazenda para facilitar a vida dos que pagam imposto de Renda. Fazendo par com o "carnê-leão", o mensalão é definido como "um recolhimento facultativo que pode ser efetuado pelo contribuinte para antecipar o pagamento do imposto devido na Declaração de Ajuste Anual, no caso de recebimento de duas ou mais fontes pagadoras de pessoas físicas..."
2. **Jornal "O Estado de São Paulo" de 19/06/2005**
Agência Estado;
NUPPS-USP
NOTÍCIA: PUBLICADO EM 2005. PORTUGUÊS
RELEVÂNCIA NA PESQUISA 29.725%
#CORRUPÇÃO:
...da bandalheira Mensalão, um gigante pela própria natureza "aumentativo em 'ão'" tende a pegar no País, mesmo quando seu significado escandaliza a realidade. O QUE VIRA – Para Giannotti, impossível prever o futuro. Lula ainda pode virar o jogo. Ou deixar para seu sucessor a batata quente de uma reforma na burocracia de governo. Filósofo, professor emérito da USP, pesquisador do Centro Brasileiro de Análise e Planejamento (Cebap) TIAGO QUEIROZ/AF. Mario Sérgio Conti Não foi o

Com a parceria e colaboração d'*O Estado de São Paulo*, foi criado, junto com o acervo acadêmico, o *Acervo de Notícias*, que traz, em uma linha do tempo na página principal do site da *Corrupteca*, informações sobre corrupções do *Acervo Digital* do jornal (Fig. 4), que contém todas as notícias, desde a fundação do diário paulistano, em 1875.

Em seus primeiros meses de funcionamento, os principais temas buscados no acervo da biblioteca abrangiam os escândalos de "Collorgate", que gerou o impeachment do ex-presidente, e agora senador, Fernando Collor de Melo e do caso 407, apelidado de Mensalão no governo do ex-presidente Lula. Atualmente os casos conhecidos como o "mensalão mineiro", a compra de votos na reeleição de Fernando Henrique Cardoso e corrupções à época da Ditadura Militar, do assassinato do ex-prefeito de Santo André em São Paulo, Celso Daniel e a recente Operação Satiagraha ganham destaque.

Figura 4 – ***Acervo digital do Jornal Estado de São Paulo integrado a corrupteca.***



6. Conclusões

A era digital vem impactando drasticamente o planejamento das atividades das bibliotecas, exigindo que se efetue novas configurações para acomodar o uso da tecnologia, tanto para a armazenagem quanto para prover acesso às informações digitais disponíveis em redes de computadores, principalmente a Rede Internet. Nesse processo de mudanças, os serviços tradicionais também continuam tendo demandas, requerendo, portanto, uma gestão que permita a coexistência pacífica, entre os vários recursos criados por essas tecnologias avançadas (IFLA, 2001).

Essa valorização das facilidades do ambiente online vem levando as bibliotecas universitárias a disponibilizar serviços a seus usuários pelas chamadas bibliotecas digitais de teses e dissertações, que tem como objetivo facilitar o acesso remoto a essa parte de produção

intelectual das universidades, englobando teses e dissertações nas áreas nas mais diversas áreas do conhecimento.

Calvin Mooers (1951) define três problemas a serem equacionados nesse campo digital : o da representação e organização da informação, o da especificação da busca por informação e o da criação de mecanismos para recuperação.

A inserção da preocupação semântica na elaboração de bibliotecas digitais, juntamente com a criação de uma ontologia específica se traduz em um ambiente informacional digital mais preciso e eficiente quanto a recuperação de conteúdos.

A arquitetura proposta para o estabelecimento da Web Semântica proporciona uma maior estruturação do conteúdo dos recursos, bem como, uma maior definição semântica dos dados representados, criando um ambiente favorável para que agentes de softwares possam realizar atividades sofisticadas como processar e entender o contexto dos dados solicitados na busca, proporcionando assim uma recuperação mais eficiente para os usuários. (Bernes-Lee, Hendler, Lassila, 2001).

Segundo Hendler (2001), a Web Semântica pode ser considerada como a composição de um grande número de pequenos componentes ontológicos que apontam entre si. Dessa forma, companhias, universidades, agências governamentais e grupos de interesses específicos procurarão ter seus recursos web ligados a um conteúdo ontológico, já que ferramentas poderosas serão disponibilizadas para intercambiar e processar essas informações entre aplicações web.

Para essa aplicação eficiente, Degoulet et al., (2011) completam dizendo que a interoperabilidade semântica requer a articulação de três elementos: a terminologia, a explicitação das relações semânticas dos elementos da terminologia e a representação dos conceitos nas mensagens.

Para o sucesso de projetos como o da corrupteca destacamos a importância do profissional da informação que, ao interagir com máquinas sofisticadas e inteligentes, se porta como um agente no processo de tomada de decisão - nesta sociedade onde a utilização eficaz da informação e do conhecimento vem tornando-se fundamental, a sua competência tecnológica

7. Referências Bibliográficas

- ALVES, R. C. V.(2005) *Web semântica: uma análise focada no uso de metadados*. Dissertação de Mestrado, Universidade Estadual Paulista, Marília, SP, Brasil.
- BERNAL, J. D.(1948, july) Priliminary analysis of pilot questionnaire on the use of scientific literature. *The Royal Society Scientific Information Conference*, London, UK.
- BERNERS-LEE, T.; Hendler, J.; Lassila, O. (2001) *The semantic web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities*. *Scientific American*, New York, NY ,USA.
- BLATTMANN, U.; Silva, F. C. C. da(2007, julho 12) Colaboração e interação na web 2.0 e biblioteca 2.0. *Revista ACB: Biblioteconomia*, 12(2), pp 191,215.
- DEGOULET, Patrice; Fieschi, Marius; Attali, Christophe. Les enjeux de l'interopérabilité sémantique dans les systèmes d'information de santé. Disponível em: <<http://www.spim.jussieu.fr/doc/pdg/InfoSante/9-23.pdf>>.
- DERVIN; B.; Nilan, M. (1986) Information needs and uses. *Annual Review of Information Science and Technology (ARIST)*. 21, pp. 3, 33.
- FERREIRA, S. (1997) *Estudos de Necessidades de Informação: dos paradigmas tradicionais à abordagem Sense-Making*. ABEED, 1997.
- NIELSEN, J.; Sano, D. (1995) *Sun Web: User Interface Design for Sun Microsystem's Internal Web*. *Computer Networks and ISDN Systems*, .28(1&2). 1995, pp.179-188.
- PEREIRA, A. S.(2009) *O Advento Digital e a nova missão da Biblioteca Pública*.
- QUEIROZ, J. (2006) *Abordagem Híbrida para a Avaliação da Usabilidade de Interfaces com o Usuário, IHM* [Apostila de Métodos de Avaliação de Interfaces Usuário-Computador] São Paulo.
- REIS, G. A. dos.(2007) *Centrando a Arquitetura de Informação no usuário*. (Dissertação Mestrado) , Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo , SP , Brasil.
- SARACEVIC, T. (1996, julho 12) Ciência da Informação: origem, evolução e relações. *Perspectivas em Ciência da Informação*,. 1(1), pp. 41,62.
- WURMAN, R. S. (2005) *Ansiedade de informação 2: um guia para quem comunica e dá instruções*. São Paulo: Editora de Cultura.