16/09/2015 - Transparência na web: livro de pesquisadores da USP e da Universidade Federal de Alagoas ensina como disponibilizar dados abertos conectados



Obra é a primeira a ser lançada no Brasil sobre o assunto e, além de suprir a carência de informações na área, tem como meta aprimorar o ensino e o treinamento, contribuindo para o aumento da transparência

Imagine se qualquer cidadão pudesse acessar todos os dados sobre o transporte público urbano no Brasil via web, obtendo informações sobre horários, rotas e destinos. Agora pense na relevância desses dados se estivessem ao alcance dos gestores desse sistema, possibilitando que conseguissem otimizar o transporte local, disponibilizando mais veículos em horários com maior demanda e comparando os dados com os de municípios similares para analisar a eficiência do serviço.

Utopia? Pode até ser, mas, na verdade, muitos desses dados já existem na web, a questão é como disponibilizar essas informações de maneira que qualquer pessoa ou computador possa acessá-las, manipulá-las, reutilizá-las e redistribui-las, relacionando-as a outros dados disponíveis sobre o assunto. Explicar como fazer isso é a meta do livro Dados Abertos Conectados, que será lançado no próximo dia 23 de setembro, durante a Conferência Web.br, em São Paulo. A obra é de autoria dos professores Seiji Isotani – do Instituto de Ciências Matemáticas e de Computação da USP, em São Carlos – e de Ig Bittencourt, do Instituto de Computação da Universidade Federal de Alagoas.

Lançado pela Novatec Editora, o livro é o primeiro do Brasil a tratar desse assunto e nasceu a partir de uma parceira dos pesquisadores com o Centro de Estudos sobre Tecnologias Web (Ceweb.br) do Núcleo de Informação e Coordenação do Ponto BR (NIC.br), que resultou também em um curso online. “Como não existiam referências consolidadas sobre o tema no Brasil, construímos o conteúdo com zelo e muita discussão, inclusive consolidando a tradução do termo Open Linked Data para Dados Abertos Conectados”, ressaltam os autores.

Eles afirmam que o conceito de dados abertos conectados é fundamental para ampliar a transparência pública de informações divulgadas pelos órgãos dos governos federal, estadual e pelos municípios, os quais devem atender à Lei de Acesso a Informação (Lei 12.527), que entrou em vigor em março de 2012. O professor Isotani diz que já existem iniciativas importantes para disseminar os dados abertos conectados no Brasil e cita como exemplo o portal dados.gov.br, mas ressalta que ainda há muito a ser aprimorado. “A Inglaterra e os

Estados Unidos estão liderando o movimento em prol da produção dos dados abertos conectados. Nossa publicação é uma forma de fomentar isso no Brasil", diz o pesquisador.
Entre os desafios que permeiam a área, os autores citam a falta de conhecimento técnico sobre como disponibilizar os dados de forma aberta e conectada e também a falta de conhecimento tecnológico sobre as ferramentas existentes para realizar essa tarefa de forma adequada. Neste contexto, o professor Bittencourt ressalta: "A obra é voltada para pessoas de qualquer área do conhecimento interessadas em disponibilizar dados na internet, quer sejam técnicos ou gestores. Por isso, tentamos escrever da forma menos técnica possível".
Tal como os dados abertos conectados, em breve, a obra estará disponível e acessível gratuitamente na Web (http://ceweb.br/publicacao/livro-dados-abertos), sob a licença Creative Commons, com a atribuição uso não comercial (http://creativecommons.org/licenses/by-nc-sa/4.0/). Isso significa que, para fins não comerciais, qualquer pessoa poderá utilizar partes ou todo o livro e redistribui-lo em qualquer suporte ou formato, além de modificá-lo. Bastando, para isso, atribuir o crédito para a obra original e seus autores. "Como recursos educacionais, o livro e os cursos geram capacitação e formam pessoas aptas a utilizar essas tecnologias, que podem produzir novos negócios e novas soluções para a sociedade", destaca Vagner Diniz, gerente do Ceweb.br.

Ranking estrelado – No livro, os pesquisadores destacam um ranking com cinco níveis para classificar os dados disponibilizados na web. O nível mais básico, que ganha somente uma estrela, são aquelas informações que apenas pessoas conseguem visualizar. Isotani cita como exemplo os dados contidos em arquivos de imagens como JPG ou BMP.
Já no nível duas estrelas, o computador consegue processar as informações. Contudo, elas são disponibilizadas em formato proprietário e precisam de software específico para acessá-las. Como exemplo, o professor cita uma planilha construída e salva no padrão Excel.
Subindo um nível na escala e ganhando três estrelas, estão as informações disponibilizadas em um formato aberto, ou seja, mesmo quem não possuir o software proprietário em que o arquivo foi construído poderá acessá-lo (por exemplo, arquivos em formato CSV).
Para chegar às quatro estrelas, além de colocar as informações no formato aberto, é preciso disponibilizá-las usando uma linguagem padronizada e recomendada pelo Consórcio World Wide Web (W3C), voltada especificamente para facilitar o compartilhamento de dados na internet: a Resource Description Framework (RDF). Com o emprego dessa linguagem, os dados ganham a propriedade de se tornarem facilmente acessíveis e processáveis por um computador de forma quase automática. Como vivemos na era do Big data, em que há um grande volume de dados gerados na web, é essencial que os sistemas sejam capazes de processar esses dados de maneira automatizada.
O nível cinco estrelas do ranking é alcançado quando, além de colocar as informações no formato aberto e usar a linguagem RDF, os dados estão relacionados a outras informações disponíveis na web. É aí que a conexão com outras fontes se dá. Para dar um exemplo, voltemos ao início deste texto: o que aconteceria caso todos os dados sobre o transporte público urbano no Brasil estivessem disponíveis na web de forma estruturada e conectada, possibilitando que um computador pudesse processá-los?
Um verdadeiro universo de possibilidades se abriria. Seria viável criar um aplicativo para analisar o transporte público de centenas de municípios e realizar diversas comparações e análises estatísticas, que são fundamentais para obter uma visão geral sobre a cobertura e a

qualidade do serviço realizado. Em vez de demorar dias, tarefas complexas e trabalhosas poderiam ser realizadas em poucos minutos. Não é por acaso que, em busca de incentivar o leitor a pensar sobre as necessidades e potencialidades de produzir e disponibilizar dados livremente na web, Isotani e Bittencourt citam o exemplo do transporte público no livro. Agora imagine poder fazer isso com os dados referentes a todos os serviços públicos do nosso país e, por que não, do mundo.

Texto: Denise Casatti – Assessoria de Comunicação ICMC/USP

A parceria de Ig e Seiji (à frente, da esquerda para a direita) também resultou na criação da startup Meu Tutor (crédito foto: Divulgação)