



COMO A WEB DE DADOS PODE APOIAR O PROCESSO DE INTELIGÊNCIA COMPETITIVA

Leandro Dal Pizzol

Mestrando em Engenharia e Gestão do Conhecimento pela
Universidade Federal de Santa Catarina, Brasil.

E-mail: leandro@egc.ufsc.br

José Leomar Todesco

Doutor em Engenharia de Produção pela
Universidade Federal de Santa Catarina, Brasil. Professor da
Universidade Federal de Santa Catarina, Brasil.

E-mail: jose.todesco@ufsc.br

Bernardo Portela Reboredo Todesco

Bacharel em Relações Internacionais pela
Universidade Federal de Santa Catarina, Brasil.

E-mail: bernardotodesco@gmail.com

Resumo

A maioria da informação necessária em um processo de Inteligência Competitiva (IC) pode ser obtida de fontes abertas. Contudo, a falta de semântica dessas fontes restringe a tarefa de recuperação de informação, fazendo da captura de conhecimento uma atividade particularmente difícil. A Web of Data (WoD) avança nesse sentido possibilitando um espaço global de dados com conexões entre as fontes e mecanismos padrão para acessar e processar os dados. Assim, este trabalho propõe alinhar o processo de IC à Web of Data. Para tanto, é proposto um modelo para identificação, seleção e classificação da informação baseado em setores econômicos que facilita a recuperação e uso dos dados na etapa de coleta do ciclo de IC. Espera-se com isso que organizações possam explorar novas fontes de conhecimento, minimize esforços de coleta de informação, e consequentemente, de análise. A verificação do modelo se deu pela sua aplicação no setor de Eletricidade e Gás, pela identificação dos requisitos de IC e pela coleta dos dados pertencentes ao setor escolhido.

Palavras-chave: Inteligência Competitiva. Engenharia do Conhecimento. Linked Data. Web of Data.

HOW CAN DATA WEB SUPPORT INTELLIGENCE COMPETITIVE PROCESS

Abstract

Most of the information required in a process of Competitive Intelligence (CI) can be obtained from open sources. However, the lack of semantics of these sources restricts the information retrieval task and makes knowledge access a particularly difficult activity. The Web of Data (WoD) progresses in this direction, enabling a global data space with connections between data sources and standard mechanisms to access and process the data. Thus, this work proposes align the CI process to the WoD. For this end, it is proposed a model for identification, selection and classification of information based on economic sectors, which facilitates the retrieval and use of data in the collection step of the CI cycle. It is expected, therefore, that organizations could be able to explore new sources of knowledge, minimize information collection efforts and, consequently analysis efforts. The verification of the model was due to its application in the "Electricity and Gas" sector, the identification of CI goals and the collection of data sets to the chosen sector.

Keywords: Competitive Intelligence. Knowledge Engineering. Linked Data. Web of Data.

1 INTRODUÇÃO

Um dos primeiros autores a usar o termo sociedade do conhecimento foi Robert E. Lane em 1966 na obra “The Decline of Politics and Ideology in a Knowledgeable Society”. O autor apontava para a crescente importância do conhecimento para a sociedade e seus membros (LANE, 1966).

A Web começou a se popularizar como fonte de informação em meados da década de 1990, e assim segue até hoje (HYLAND, 2010). Com as tecnologias Web tomando forma cada vez mais rápido, a capacidade de colher informações, é atualmente, um desafio maior do que em qualquer outra época (LANA, 2011). Esse tipo de ambiente traz grande variedade de fontes de informação de livre acesso. Contudo, os dados não costumam ser estruturados, apresentam problemas de ambiguidade e heterogeneidade o que restringe a tarefa de recuperação de informação e faz da captura de conhecimento e seu reuso atividades particularmente difíceis (PASSANT et al., 2010).

Uma forma de contornar estes problemas é a Web of Data. A WoD é formada por um grafo gigante com bilhões de declarações em formatos estruturados provenientes de várias fontes, cobrindo temas dos mais variados domínios. O pressuposto básico por trás da WoD é que a utilidade e o valor desses dados aumentam à medida que forem acessados e recombinações uns aos outros (HEATH; BIZER, 2011).

Assim, adaptando os métodos tradicionais de coleta de informação, a WoD pode ser uma possibilidade interessante para a melhoria de processos organizacionais como a IC. Para Prescott e Miller (2002), em um processo de IC, dois terços do tempo é gasto na coleta de informação por vezes desnecessária. Tarefa que sem a análise do que foi coletado não acrescenta valor a organização.

A Web of Data possibilita conexões explícitas entre os conjuntos de dados usando formatos padrões de representação e acesso (CYGANIAK; REYNOLDS; TENNISON, 2010). Além disso, ferramentas genéricas como navegadores e motores de busca podem ser usados para acessar e processar os dados (JENTZSCH et al., 2009).

Diante disto, a proposta deste trabalho visa identificar e classificar os dados presentes na WoD para que estes sejam utilizados no processo de IC. É esperado que as organizações possam explorar novas fontes de conhecimento, diminuir esforços de captura e consequentemente de análise devido à estruturação da informação.

Nas seções seguintes deste artigo serão apresentados os trabalhos relacionados, os conceitos teóricos sobre IC e WoD, o modelo proposto e sua verificação, e por fim as conclusões obtidas.

2 TRABALHOS RELACIONADOS

Segundo Allemang (2011), as experiências da *Web* sugerem que há um caminho para empresas construírem uma arquitetura sustentável de informações empresariais, transformando-se em “*Linked Data Enterprises*”, na qual o ato de criação da informação está intimamente ligado com o de compartilhamento, sendo este, tão importante como sua produção.

Servant (2008) e Hu e Svensson (2010) descrevem casos de aplicação da *WoD* em grandes organizações, por meio da integração de tecnologias de informação, comunicação existente e dados internos, com as abordagens de *Linked Data*, promovendo a ligação destes com fontes externas, resultando em novas possibilidades de visualização dos mesmos. Já (JENTZSCH et al. 2009) apresenta os dados sobre medicamentos e a indústria farmacêutica presentes na *WoD*.

No trabalho de Eisenberg e Kanza (2011), são usadas as tecnologias da *Web Semântica* para integrar a informação organizacional, ou de uma comunidade fechada, como um banco de dados global, criando um espaço corporativo. Segundo os autores atualmente a atividade empresarial é baseada em enorme quantidade de informações e extrair informação certa na hora certa é uma tarefa difícil e tediosa, problema este que norteou este trabalho.

3 INTELIGÊNCIA COMPETITIVA

Para Juhari e Stephens (2006), a base para a maioria das disciplinas é encontrada em sua origem e história. Os autores apontam atividades de inteligência para fins militares datadas de três mil anos atrás. Inspirado nas cronologias de Underwood (2002) e Juhari e Stephens (2006) a seguir, são retratadas cronologicamente as iniciativas de inteligência e sua evolução ao longo da história.

Figura 1 – Linha do Tempo – Iniciativas de inteligência (2000 a.C. - presente)



Fonte: Autoria própria

Explicitamente aplicada a negócios, o uso de inteligência data de 1960. Já a década de 1980 viu a introdução de funções formais de coleta e análise de informações sobre competidores (PRESCOTT, 1999). Outros fatores de destaque no desenvolvimento da IC foram o fim da Guerra Fria e o surgimento do capitalismo onde diretores de indústrias e serviços afetam diretamente o futuro dos cidadãos por meio dos produtos e empregos que elas oferecem. Por fim, a década de 1990 viu a introdução da tecnologia, mais especificamente a Internet, como um meio para aumentar a

profundidade e amplitude da IC. É nesta década que a IC torna-se a entidade dinâmica e complexa que é hoje (UNDERWOOD, 2002).

3.1 Conceitos

Para Fleisher e Bensoussan (2007), existem muitas definições de IC e provavelmente nenhuma delas seja precisa e universalmente aceita. Na visão de Gilad (2011) o conceito é bastante vago apesar de diversas tentativas de atribuição de significado. Assim, de maneira geral, definições formais sobre o tema ou sobre aspectos pontuais destacam que a IC deve apoiar a tomada de decisão. A Tabela 1 apresenta alguns destes conceitos.

Tabela 1: Conceitos e Definições a respeito da IC

Conceitos	Autores
Programa sistemático de coleta e análise da informação sobre a atividade dos concorrentes e tendências de negócios, para auxiliar nos objetivos da empresa.	Kahaner (1996)
Processo sistemático que transforma dados e partes de informações competitivas em conhecimento estratégico para a tomada de decisão.	Tyson (2002)
IC é uma estratégia para a empresa descobrir o que se passa no ambiente de negócios do seu setor, e esse conhecimento dá aos executivos condições de tomar atitudes que forneçam à empresa uma vantagem sobre seus concorrentes.	Miller (2002)
Processo ético de identificação, coleta, tratamento, análise e disseminação da informação estratégica para a organização, viabilizando seu uso no processo decisório.	Gomes e Braga (2004)
Processo sistemático e ético de coleta, análise, disseminação e gerenciamento das informações sobre o ambiente externo, que podem afetar os planos, as decisões e a operação da organização. IC efetivo é um processo contínuo envolvendo coleta ética e legal de informação, análise que não evita conclusões negativas, e disseminação controlada de inteligência acionável aos tomadores de decisão.	Scip (2013)

Fonte: Autoria própria

3.2 Processo de Inteligência Competitiva

O processo de IC possibilita entender melhor o ambiente no qual a organização está inserida, agregando valor ao processo decisório. Segundo Kahaner (1996) e Sawka (2006), os resultados desse processo permitem gerenciar mudanças no planejamento estratégico e maior eficácia organizacional.

Para Abreu, Cobral e Ogliari (2008), é importante sistematizar estas atividades, tornando o processo de IC algo contínuo que compreenda as necessidades da organização e de cada decisor. Para tanto, Bensoussan e Fleisher (2003) destacam que são relevantes os esforços em atividades como: seleção criteriosa da informação, verificação da confiabilidade, análise, integração e interpretação desses resultados. A Tabela 2 a seguir desenvolvido por Machado (2010) apresenta o número, as etapas e os autores que o desenvolveram.

Tabela 2: Etapas de estruturação do processo de IC

Processo de IC estruturado em	Autores
4 etapas (identificação das necessidades de informação, coleta, análise, disseminação)	Miller (2002) Abreu; Cobral, Ogliari (2008)
4 etapas (planejamento e direção, coleta, análise, disseminação)	Kahaner (1996)
5 etapas (planejamento e direção, coleta, processamento e exploração, análise e produção, disseminação)	Bernhardt (2003)
5 etapas (planejamento e direção, processamento e armazenamento da informação, coleta, análise e produção,	Herring (1998)

disseminação)	Bensoussan e Fleisher (2003)
5 etapas (identificação das necessidades de informação, coleta, análise, disseminação, avaliação)	Gomes e Braga (2004)
5 etapas (planejamento e direção, coleta, análise, disseminação e avaliação)	Bose (2008)

Fonte: Machado (2010).

Analisando a Tabela 2, vemos que as etapas de coleta, análise e disseminação são comuns a todos os autores no processo de IC. A identificação das necessidades de informação é apresentada como uma etapa por (GOMES; BRAGA, 2004). Já para os demais autores esta constitui um passo dentro da etapa de planejamento. Por fim alguns autores incluem uma etapa de avaliação ao fim do processo que comumente valida os resultados, e se necessário, traz o processo de volta para o início, começando um novo ciclo.

3.2.1 Identificação das Necessidades de Informação

O fator crítico de sucesso em qualquer operação de inteligência é satisfazer as necessidades de usuários chave e da alta gestão, de forma que a organização atue com base na inteligência resultante. Para tanto, (HERRING, 1998) propõe os Key Intelligence Topics (KITs) onde cada solicitação deve ser tratada como um projeto e iniciativas definidas com base em reuniões com os principais tomadores de decisão.

Outros métodos empregados na identificação de necessidades de informação são o Método dos Fatores Críticos de Sucesso, baseado em descrições do negócio, perspectivas, estratégias e fatores críticos de sucesso associados (Rockart 1979). O mapa estratégico que organiza a informação em dois níveis: estratégico e tático-operacional. Em seguida, estas são atreladas a objetivos estratégicos, fatores críticos de sucesso, oportunidades ou ameaças (ABREU, COBRAL e OGLIARI 2008).

3.2.2 Planejamento

A etapa de planejamento é definida por muitos como a fase mais importante do ciclo de IC (HERRING et al., 2002; MILLER, 2002; OLIVEIRA; MELO, 2012). Do seu desenvolvimento resulta a missão do projeto, objetivos, estratégias e ações que serão tomadas. Definem-se também os recursos humanos, tecnológicos, financeiros e o tempo que a empresa dispõe para ser utilizado (BOSE, 2008). Por último, devem ser apontados os caminhos e ações para as demais etapas do ciclo levando em conta fatores como segurança e planos de contingência (BOSE, 2008).

3.2.4 Coleta

A coleta caracteriza-se pela busca de dados e informações necessárias para desenvolver conhecimento sobre o ambiente competitivo. A tarefa de coleta é essencialmente prática dividindo-se na identificação das fontes, captação, tratamento, e o armazenamento de informações. Kahaner (1996) classifica as informações quanto à origem, domínio e tipo como descrito na Tabela 3. O autor defende que “não se determina nenhum fator que indique que um tipo de informação supera a outra em termos de consistência, veracidade ou aplicabilidade”.

Tabela 3 – Classificação das Informações

Classificação	Tipo	Descrição	Exemplos
Quanto à origem	Primária	Tem origem no próprio concorrente	Relatórios anuais, discursos, entrevistas.

	Secundária	Tem origem em outras fontes que observam os concorrentes	Jornais, revistas, relatórios de analistas.
Quanto ao domínio	Público	Informações tornadas públicas pelos concorrentes	Balanço de S.A., publicações na <i>Web</i> .
	Não-público	Informações que não são publicadas pelos concorrentes	Pesquisas em feiras e com forças de vendas.
Quanto ao tipo	<i>Hard</i>	Informações baseadas em dados quantitativos	Relatórios estatísticos, relatórios financeiros.
	<i>Soft</i>	Informações baseadas em dados qualitativos	Entrevistas, discursos, boatos e rumores.

Fonte: adaptado de Kahaner (1996)

Segundo Jain (1984), os processos de coleta são divididos em quatro tipos: Primitivo, feito sem um objetivo específico. Ad hoc, costuma ocorrer em resposta a uma necessidade. Reativo, quando é diagnosticada uma ameaça competitiva concreta. E por fim, proativo, que caracteriza-se pela busca ativa de todo tipo de informações que possam identificar ameaças ou oportunidades para a organização.

Ferramentas de coleta costumam reunir informações de diferentes fontes na Web com base em termos previamente cadastradas e podem ser ativas ou passivas. As primeiras apoiam a coleta com pesquisas baseadas em terminologias e categorização dos resultados. Exemplo são os motores de busca da Web que possibilitam pesquisas e navegação online. A restrição está na estruturação das consultas que podem ser problemáticas devido à "interação homem-computador", ou a "problemas de vocabulário" ao usar as palavras certas na busca. Já ferramentas de coleta passiva são direcionadas a apoiar a inteligência informacional em curso. É comum que os tomadores de decisão e profissionais de IC utilizem software para fornecer atualizações diárias sobre notícias, atividades e mudanças em sites de concorrentes (VAN ZUYLEN, 2006).

3.2.4 Análise

Referida como o "cérebro" do processo de IC, a análise dos dados é a etapa que de fato transforma dados e informação em inteligência, pelo que é reconhecida como a fase que mais agrega valor ao ciclo de IC (BOSE, 2008; CORREIA; SANTOS, 2010). Para Bouthillier e Shearer (2003) o processo de análise é uma tarefa complexa que deve ser dividido em três fases: A avaliação, diz respeito ao grau de confiabilidade da fonte de dados e veracidade da informação. A catalogação, que compreende o processo de ordenação, catalogação, arquivamento e indexação dos dados levantados. Em dados pouco estruturados, isto demanda limpeza, estruturação e um pré-processamento que pode incluir a conversão desses dados para um formato específico antes de sua catalogação. Por fim, a última fase corresponde à interpretação, onde a informação transforma-se em conhecimento.

3.2.5 Disseminação

Herring et al. (2002) mostra que nesta etapa, são apresentadas ao tomador de decisão as respostas às suas questões e sugeridos possíveis cursos de ação. O produto final da inteligência é quem guiará a tomada de decisão, estratégica, tática ou operacional. Segundo Gomes e Braga (2004), a disseminação pode ser focada quando se trata de informação específica, ou geral quando a informação é divulgada para toda a empresa. Mecanismos de disseminação mais populares são: relatórios, apresentações, boletins, mensagens eletrônicas, recomendações, entre outros. É uma boa prática focar o resultado da análise quando a informação é dirigida a um ou a um grupo de usuários que solicitaram aquela informação (KAHANER, 1996; GOMES; BRAGA, 2004).

3.2.6 Avaliação

Segundo Correia e Santos (2010), nesta etapa ocorre a avaliação do processo de IC, são identificados benefícios e a eficácia, são revistas e reavaliadas a estratégia organizacional e se o processo de IC está sendo visto como uma função corporativa. Outra importante função da fase de análise é mensurar os resultados dos produtos do processo de IC. Gomes e Braga (2004), sugerem a criação da memória do processo e a definição de indicadores para contabilizar os resultados reais em termos de produtividade, redução de custos, investimentos e o retorno com a inovação que foi alcançada.

4 WEB OF DATA

Nos últimos anos organizações públicas e privadas têm disponibilizado seus dados de forma aberta para estabelecer uma construção coletiva de conhecimento. Segundo a definição da *Open Knowledge Foundation*, dados são abertos quando, qualquer pessoa pode livremente usá-los, reutilizá-los e redistribuí-los, estando sujeito a, no máximo, creditar a sua autoria e compartilhá-los pela mesma licença.

O resultado disso é um espaço global que chamamos de *Web of Data* (BIZER; HEATH; BERNERS-LEE, 2009). A WoD forma um grafo mundial gigante que consiste em bilhões de declarações *Resource Description Framework* (RDF) provenientes de várias fontes cobrindo temas variados, tais como: geografia, pessoas, empresas, publicações científicas, mídia, estatística, governo, entre outros.

A WoD pode ser vista como uma camada adicional fortemente interligada com os documentos da Web tradicional tendo muitas das mesmas propriedades como o caráter genérico, a facilidade de publicação, representação, conexão de informação, além de ser auto-descritiva e possuir mecanismos padronizados de publicação e acesso aos dados (HEATH; BIZER, 2011).

4.1 Linked Data

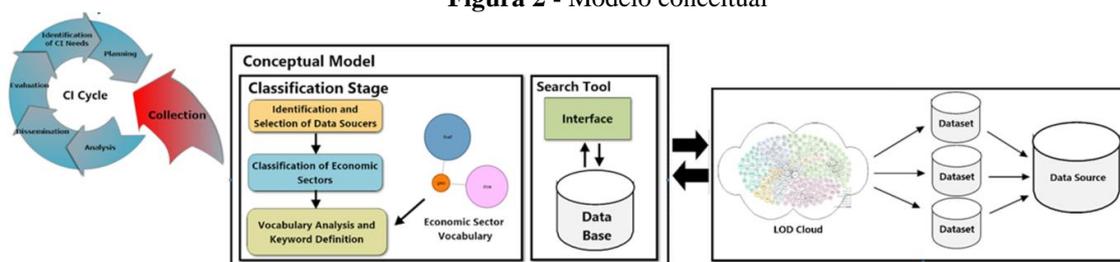
Proposto por Sir. Tim Bernes-Lee, em 2006, o termo *Linked Data* refere-se a um estilo de publicar e interligar dados estruturados na Web. *Linked Data* não representa propriamente uma nova tecnologia, mas sim um conjunto de melhores práticas para publicação e interligação de dados estruturados na Web (HEATH; BIZER, 2011).

O objetivo de *Linked Data* é utilizar a arquitetura Web, para compartilhar dados estruturados em uma escala global (HEATH; BIZER, 2011). Assim como a Web formada por *links hypertext*, *Linked Data* é construída com documentos na Web, porém, as ligações são feitas usando links hyperdata, onde a informação presente nos documentos pode ser ligada (BERNERS-LEE, 2006). Essa forma de ligação permite que as pessoas compartilhem dados estruturados na Web de forma tão fácil quanto documentos são atualmente compartilhados.

5 MODELO PROPOSTO

O modelo proposto reúne as etapas e os elementos necessários para realizar a coleta de informações da WoD para apoiar o ciclo de Inteligência Competitiva. A Figura 2 apresenta este modelo.

Figura 2 - Modelo conceitual



Fonte: Autoria própria

O modelo consiste no arranjo de tecnologias e conceitos em uma ferramenta de engenharia do conhecimento com o objetivo final de alimentar a fase de coleta do processo de IC com informações oriundas da WoD. O modelo conceitual contempla questões centrais de IC e está focado na coleta de dados na WoD de forma genérica. Informações mais detalhadas sobre um aspecto específico ou até mesmo proprietárias, tais como: rentabilidade, clientes, propriedade intelectual, segredos comerciais, planos estratégicos ou procedimentos de gestão interna devem ser inferidos na fase de análise. A seguir são descritas cada uma das etapas do modelo.

5.1 Etapa de Classificação

A etapa de classificação compreende os passos iniciais para que a WoD seja usada no processo de IC. Nesta etapa foram desenvolvidas tarefas de seleção dos data sources, definição dos setores econômicos e dos termos usados para representar e retornar a informação. Nos tópicos seguintes são discutidas cada uma dessas tarefas.

5.1.1 Seleção das Fontes

O ponto de partida para a etapa de classificação foram os 295 conjuntos de dados que formam o diagrama da LOD *Cloud* de 2011ⁱ. Desse conjunto inicial foram selecionados ao fim do processo 135 fontes de dados, que segundo os critérios especificados a seguir, possuem relevância para o processo de IC:

1. Presença de links ativos: Foram eliminados *data sources* que mesmo representados no diagrama não possuem links ativos. Esta análise resultou na exclusão de 26 *data sources*;
2. Disponibilidade: foram excluídos 27 *data sources* que não possuem dados recuperáveis disponíveis;
3. Duplicidade: foram excluídas 36 fontes duplicadas ou contempladas como conjuntos de dados dentro de outras fontes;
4. Relevância: o conteúdo dos *data sources* foi analisado quanto a sua relevância para o processo de IC e os que não apresentam informações pertinentes foram excluídos como, por exemplo: religião e repositórios sobre desenhos animados. Por fim, foram eliminadas fontes em línguas de difícil compreensão como japonês, chinês, cirílico, entre outros. Nesta etapa foram excluídas 71 fontes.

Cabe ressaltar que o conjunto resultante é composto, por um ou mais *datasets*. Um *dataset* é um conjunto de informação publicada, mantida ou agregada por um único provedor (ALEXANDER et al., 2009). Devido à interligação existente entre eles, um *dataset* pode aparecer em mais de um *data source* simultaneamente, o que torna difícil apontar de maneira precisa sua quantidade. Contudo, uma estimativa feita com base nas informações presentes nas próprias bases identificou aproximadamente 150 mil *datasets* compostos por algo em torno de 50 bilhões de recursosⁱⁱ.

5.1.2 Seleção dos Setores Econômicos

Os dados foram classificados com base na CNAE (Classificação Nacional de Atividades Econômicas) elaborada pela Comissão Nacional de Classificação e disponibilizada pelo IBGE (Instituto Brasileiro de Geografia e Estatística). A classificação levou em conta também as necessidades de informação mais comuns dentro do processo de IC.

Dentro da estrutura CNAE, as atividades econômicas são classificadas em 21 seções e 99 divisões. Para este trabalho foi usada somente a classificação por seção, já que, o elevado número de divisões dificultaria o agrupamento dos data sources sob um mesmo tema. As seções e a respectiva quantidade de fontes presente na WoD são representadas na Tabela 4.

Tabela 4: Classificação da WoD segundo o CNAE

<i>Setor Econômico</i>	<i>Data Sources</i>
Agricultura, Pecuária, Produção Florestal, Pesca e Aquicultura	16
Informação e Comunicação; Atividades Financeiras, de Seguros e Serviços Relacionados	16
Administração Pública, Defesa e Seguridade Social	14
Artes, Cultura, Esporte e Recreação; Outras Atividades de Serviços	13
Organismos Internacionais e Outras Instituições Extraterritoriais	13
Água e Esgoto, Atividades de Gestão de Resíduos e Descontaminação	11
Atividades Profissionais, Científicas e Técnicas	11
Educação	10
Saúde Humana e Serviços Sociais	10
Atividades Administrativas e Serviços Complementares	8
Indústrias Extrativas	6
Indústrias de Transformação	6
Eletricidade e Gás	6
Comércio, Reparação de Veículos Automotores e Motocicletas	6
Transporte, Armazenagem e Correio	6
Alojamento e Alimentação	3
Construção	0
Serviços Domésticos	0

Fonte: Autoria própria

5.1.3 Análise de Vocabulários e Definição de Palavras-chave

Os *data sources*, em muitos dos casos, apresentam vocabulários para descrever seu conteúdo. Alguns termos presentes nesses vocabulários foram utilizados como palavras-chave para representar a informação contida nos *datasets*. Quando a fonte de dados não possui um vocabulário associado, o conteúdo presente na descrição da fonte foi utilizado para representar os termos de busca na aplicação. Esses vocabulários reúnem termos em uma estrutura hierárquica destinada à indexação e à recuperação de informações em um determinado setor. Além de organizar o conhecimento, um vocabulário possui uma hierarquia que auxilia na organização e compreensão do domínio.

5.1 Ferramenta de Busca

A ferramenta desenvolvida visa apoiar a recuperação das fontes de dados usadas na etapa de coleta do ciclo de IC. Assim, ao selecionar um setor econômico são apresentadas somente as fontes de dados correspondentes ao setor. Por sua vez os conjuntos de dados foram organizados de forma

que possam ser recuperados diretamente utilizando-se palavras-chave. Dessa forma, ao selecionar a fonte de dados, os *datasets* de que ela é composta são apresentados.

Figura 3 – Exemplo conjunto de dados

Data Source	URI	Description	Autor	Quantity
EnAKTing PSI energy Dataset	http://energy.psi.enakt.org/	Data extracted from the statistics for road transport consumption compiled by the UK Department for Business, Enterprise and Regulatory Reform (BERR). The data covers the whole UK territory from 2002 to 2007.	University of Southampton	2316831
Enipedia	http://enipedia.tudelft.nl/wiki/Main_Page	Enipedia is an active exploration into the applications of wikis and the semantic web for energy and industry issues. Through this we seek to create a collaborative environment for discussion, while also providing the tools that allow for data from different sources to be connected, queried, and visualized from different perspectives.	Enipedia Team Energy and Industry Section, TBM, Delft University of Technology	180734
Energypedia	https://energypedia.info/wiki/Main_Page	Energypedia is a wiki-based platform for collaborative knowledge exchange on renewable energy and energy access issues in the context of development cooperation.	Energypedia	1623

6 APLICAÇÃO DO MODELO

Para a verificação do modelo proposto foi escolhido o setor de Eletricidade e Gás. O setor elétrico, em especial o brasileiro, caracteriza-se por ser altamente regulado pelo governo, que promove a competitividade entre os agentes setoriais. São práticas comuns leilões de novas linhas de transmissão, usinas de geração e processos licitatórios para outorga de concessões. Investimentos requerem análises detalhadas de viabilidade técnica, econômico-financeira e ambiental antes de sua efetiva implantação.

A escolha se deve pela quantidade e qualidade dos conjuntos de dados presentes na *Web of Data*, pela importância social e econômica do setor, e ainda, pelo fato desse demandar informações de outras áreas presentes na nuvem de dados como: sustentabilidade, dados geográficos, demográficos, de atividade empresarial entre outros.

Estão presentes na *Web of Data* seis *data sources* que abordam diretamente o setor elétrico. Essas fontes são compostas por pouco mais de dois mil *datasets* e aproximadamente 23,2 milhões de registros. A Figura 4 apresenta um dos *data sources* do setor elétrico.

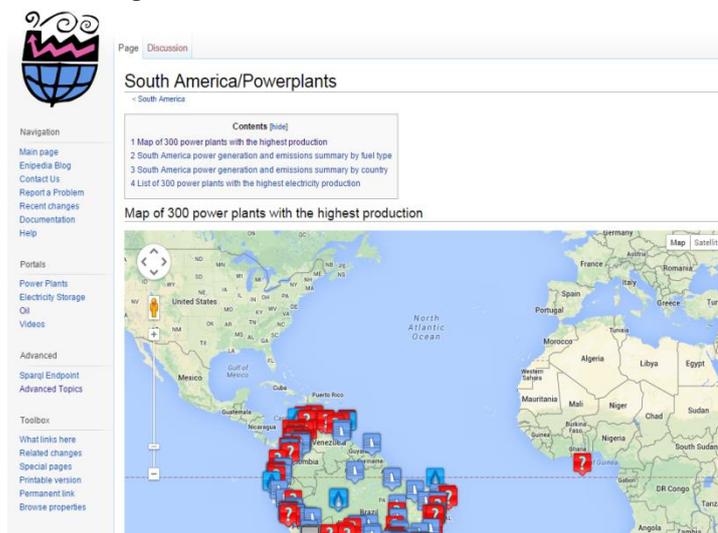
Figura 4 - Data Sources Setor Elétrico

Data Source	URI	Description	Autor	Quantity
EnAKTing PSI energy Dataset	http://energy.psi.enakt.org/	Data extracted from the statistics for road transport consumption compiled by the UK Department for Business, Enterprise and Regulatory Reform (BERR). The data covers the whole UK territory from 2002 to 2007.	University of Southampton	2316831

Fonte: Autoria própria

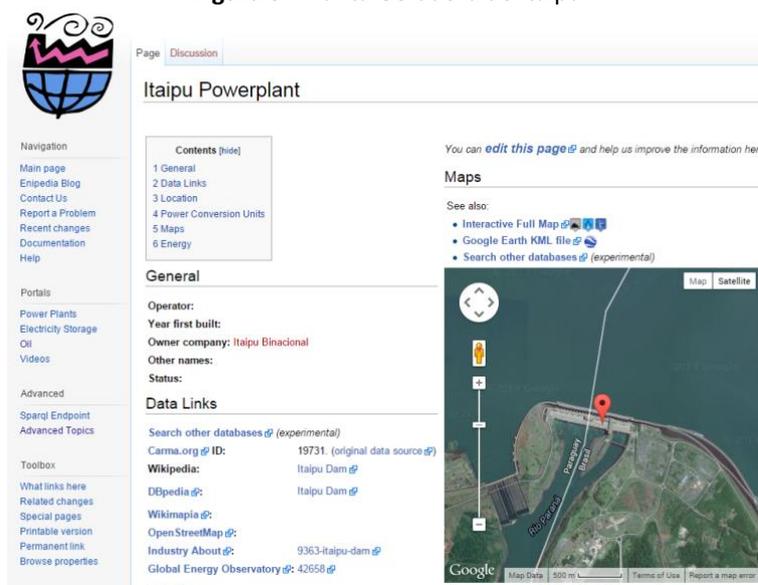
Na Figura **Erro! Fonte de referência não encontrada.**4 a coluna que contém o nome da fonte de dados aponta para a interface onde estão cadastrados os conjuntos de dados que a compõe. Já ao clicar na URI o usuário é direcionado para a página inicial do conjunto de dados. A Figura 5 apresenta a página sobre plantas geradoras de eletricidade da América do Sulⁱⁱⁱ presente na Enipedia^{iv}, uma fonte de dados sobre o setor de energia estruturada de maneira semelhante a Wikipedia.

Figura 5 – Plantas Geradoras da América do Sul



Os ícones usados para apontar a localização das plantas geradoras também representam o tipo de fonte utilizada na geração de eletricidade. Ao selecionar um desses pontos são apresentadas informações sobre a planta geradora específica. A Figura 6 apresenta as informações sobre a usina hidroelétrica de Itaipu^v, a maior da América do Sul e segunda maior desse tipo no mundo.

Figura 6 - Planta Geradora de Itaipu



Nesta interface estão presentes informações detalhadas sobre Itaipu como: mapas com coordenadas geográficas, tipo de combustível usado a capacidade de geração e a quantidade de

poluentes emitidos, entre outros. São disponibilizados ainda ligações para bases externas sobre a mesma planta geradora como a do *Global Energy Observatory*^{vi}, uma base colaborativa e de acesso livre com informações sobre energia em escala global.

6.1 Identificação dos Requisitos de IC

Através de sua experiência, Prescott e Miller (2002) verificaram que uma empresa tem suas necessidades de informação enquadradas em três categorias funcionais: decisões e ações estratégicas, tópicos de alerta antecipado e descrição dos principais atores do seu mercado. Baseados nesses conceitos foram elencadas três necessidades informacionais essenciais para o andamento de um projeto de inteligência no setor elétrico:

1. Identificar necessidades de expansão da produção e da demanda por energia;
2. Identificar fontes energéticas alternativas;
3. Identificar investimentos, sua destinação e os montantes envolvidos, fusões e participações acionárias.

6.2 Coleta dos Dados

Esta seção apresenta as fontes de informação identificadas para cada uma das necessidades levantadas anteriormente. Será discutido como a informação é disponibilizada, o formato dos dados, a origem e o seu grau de confiabilidade e relevância. A Tabela 5 apresenta um apanhado sobre os assuntos e quantidade de datasets que contemplam cada uma das necessidades identificadas.

Tabela 5 – Necessidades de IC e datasets identificados

Necessidade de IC	Assunto	Qtde.
Identificar necessidades de expansão da produção e da demanda por energia	Geração	109
	Consumo	81
	Demanda	10
Identificar fontes energéticas alternativas	Geotérmica	185
	Solar	85
	Renovável	42
	Eólica	32
	Sustentável	26
	Hidrogênio	8
	Biomassa	4
Identificar alternativas para melhorar a eficiência, mitigar perdas e estimar tendências do setor elétrico.	Indicadores	45
	Bioenergia	23
	Eficiência	14

Fonte: Autoria própria

Ao todo foram identificados 664 datasets com informação que atendem aos três requisitos de IC propostos. Para responder a necessidade “*Identificar necessidades de expansão da produção e da demanda por energia*” foram encontrados 200 datasets. Estes abordam assuntos como: demanda e o consumo anual eletricidade^{vii}, gás^{viii} e carvão^{ix} de cada país entre os anos de 1980 e 2009.

Estão disponíveis também dados sobre a demanda por energias renováveis no período de 2005 a 2009 e sobre o consumo setorial de energia de um determinado país. Os dados são disponibilizados em formatos como o CSV ou XLS geralmente por órgãos oficiais.

Ainda sobre a primeira necessidade de IC, no que corresponde a geração de energia temos datasets que tratam da geração de energia hidroelétrica no Brasil disponibilizado pela Agência

Nacional de Energia Elétrica (ANEEL), geração utilizando combustíveis fósseis, como petróleo e carvão por país, e de fontes renováveis como a eólica presente em *The Wind Power*^x onde dados como o número de fazendas eólicas e a capacidade de geração em megawatt de 102 países é listada.

Para responder a necessidade “Identificar fontes energéticas alternativas” estão catalogados 405 *datasets*. Estes abordam de maneira geral o uso de energias renováveis e as tecnologias envolvidas, o potencial dos recursos renováveis mundiais disponibilizados pelo *Nacional Renewable Energy Laboratory*^{xi}, órgão ligado ao governo americano dedicado a pesquisas na área de energias renováveis. Existem ainda *datasets* que abordam o uso de recursos sustentáveis como bioenergia, biomassa e eólica. Por exemplo, o Centro de Pesquisas de Energia Elétrica^{xii} que disponibiliza a média anual do potencial eólico a 50m em território brasileiro.

Como fontes energéticas alternativas temos o hidrogênio^{xiii}, em aplicações *Web* que possibilitam a visualização e download dos dados relativos a sua utilização, além de grande quantidade de fontes de informações sobre energia geotérmica pertencentes ao projeto *Geothermal Data Repository* (GDR).

A última pergunta a ser respondida “Identificar alternativas para melhorar a eficiência, mitigar perdas e estimar tendências do setor elétrico” contempla *datasets* como o sumário mundial sobre eficiência energética^{xiv}, indicadores, estatísticas e tendências do setor elétrico, além de iniciativas como *Smart Grids*^{xv} que buscam tornar a cadeia de produção e distribuição de energia mais inteligente.

Por fim, cabe ressaltar que todos os *datasets* que respondem as necessidades identificadas podem ser recuperados facilmente por meio de palavras-chave ou por busca textual na aplicação.

7 CONCLUSÕES

Para responder o problema apresentado no início deste trabalho foi desenvolvido um modelo conceitual capaz de identificar, classificar e representar a informação presente na *WoD* que deu origem a uma ferramenta de auxílio a coleta de informação. A ferramenta cataloga a informação e facilita a recuperação das fontes de dados no ciclo de IC. Para esta prova de conceito, a tarefa de cadastro dos *datasets* foi feita manualmente. Por este motivo, somente os *datasets* do setor de Eletricidade de Gás foram cadastrados em sua totalidade. Contudo, isso não acarreta prejuízo ao resultado final do estudo e futuramente esse processo pode ser automatizado.

O mérito deste trabalho está no uso da *WoD* como fonte de informação externa, estruturada e de fácil recuperação para o processo de IC. A *Web* de Dados acrescenta uma camada semântica adicional fortemente interligada com os documentos da *Web* tradicional mantendo muitas das suas propriedades como a interligação, comunicação e a facilidade de acesso.

Entre as contribuições para os profissionais da área de IC, destacam-se, o modelo proposto para a coleta de informação e a ferramenta resultante deste. Como os dados catalogados contemplam informações de diferentes setores, estes podem ser adaptados para ambientes distintos ou mesmo outros processos que demandam informações externas a organização.

Conforme escopo inicial, não era objetivo o desenvolvimento de todo o ciclo de IC, haja vista que as informações presentes na *Web* de Dados são de interesse direto do processo de coleta, e que posteriormente estas informações seguem o ciclo comum ao das outras fontes.

Conclui-se que como a informação presente na *WoD* é disponibilizada em formatos estruturados e provém de fontes confiáveis e especializadas no assunto, seu uso pode representar ganho de tempo na coleta e na etapa de análise. Constituem ainda benefícios identificados, o tratamento prévio dos dados, sua ligação com outras fontes, e presença de relatórios e aplicações que podem auxiliar diretamente na etapa de análise. Além disso, usando a ferramenta o analista responsável pela coleta não tem necessidade de varrer a *Web of Data* em busca de informação, já que esta se encontra concentrada em um único lugar e previamente dividida por setor econômico e assuntos.

Finalmente, como trabalhos futuros pretende-se discutir e esclarecer questões como: quais atributos de dados presentes nestes conjuntos podem ser combinados em diferentes fontes; A aderência, cobertura e a precisão dos dados de origem ao escopo do projeto de IC; e quais são os desafios da automatização do processo de IC com precisão aceitável, e como estes problemas podem ser resolvidos através de abordagens de dados vinculados?

REFERÊNCIAS

ABREU, A. F.; COBRAL, E.; OGLIARI, A. **Gestão integrada da inovação**: estratégia, organização e desenvolvimento de produtos. V.1. São Paulo: Atlas, 2008.

ALLEMANG, D. (2011) Semantic Web and the Linked Data Enterprise. In: WOOD, David. (Ed.). **Linked Enterprise Data**. [S.l.]: Springer, 2011.

ALEXANDER, K. et al. **Describing Linked Datasets On the Design and Usage of void**, the “Vocabulary Of Interlinked Datasets”. Linked Data Workshop at WWW09 2009.

BERNHARDT, D. **Competitive intelligence**: how to acquire and use corporate intelligence and counter-intelligence. [S.l.]: FT Prentice Hall, 2003.

BENSOUSSAN, B.; FLEISHER, C. S. Strategic group analysis. **CI Magazine**, v.6, n.1, p.5, 2003.

BERNERS-LEE, T. **Linked Data**: Design Issues. Disponível em: <<http://www.w3.org/DesignIssues/LinkedData.html>>. Acesso em: 02 fev. 2015.

BIZER, C.; HEATH, T.; BERNERS-LEE, T. Linked Data - The Story So Far. **IJSWIS**, v. 5, n. 3, p. 1-22, 33, 2009.

BOSE, R. Competitive intelligence process and tools for intelligence analysis. **Industrial Management & Data Systems**, v. 108, n. 4, p. 19, 2008.

BOUTHILLIER, F.; SHEARER, K. **Assessing Competitive Intelligence Software**: A Guide to Evaluating CI Technology. [S.l.]: Information Today, 2003.

CORREIA, A. M. R.; SANTOS, M. C. F. Fundamentos teóricos da competitive intelligence como vantagem competitiva. **CISTI 2010**. Santiago de Compostela, Espanha: Comunicação. 2. 2010. p. 329-334.

CYGANIAK, R.; REYNOLDS, D.; TENNISON, J. **The RDF Data Cube vocabulary**. Online, 12 March 2013 2010. Disponível em: <<http://www.w3.org/TR/vocab-data-cube>>. Acesso em: 02 abr. 2012.

EISENBERG, V.; KANZA, Y. Ruby on semantic web. In: ABITEBOUL, S. et al. (Ed.). **ICDE**. [S.l.]: IEEE Computer Society, 2011. p. 1324–1327.

FLEISHER, C.S.; BENSOUSSAN, B.E. **Business and Competitive Analysis**: Effective Application of New and Classic Methods. [S.l.]: Times Press, 2007.

GILAD, B. Strategy Without Intelligence, Intelligence Without Strategy. **Business Strategy Series**, v. 12, n. 1, p. 4-11, 2011.

GOMES, E.; BRAGA, F. **Inteligência competitiva**: como transformar informação em um negócio lucrativo. [S.l.]: Elsevier, 2004.

HEATH, T.; BIZER, C. **Linked Data**: Evolving the Web Into a Global Data Space. [S.l.]: Morgan & Claypool, 2011.

HERRING, J. P. What is intelligence analysis? **Competitive Intelligence Magazine**, v. 1, n. 2, p. 4, 1998. Disponível em: <<http://www.scip.org/>>. Acesso em: 22 jan. 2015.

- HERRING, J. P. et al. Corporate culture as a tool for anticipating the competition. **Competitive Intelligence Magazine**, v. 5, n. 4, p. 5, 2002. Disponível em: <<http://abmmanage.com/article.pdf>>. Acesso em: 03 set. 2013.
- HU, B., SVENSSON, G. A Case Study of Linked Enterprise Data. In: **ISCW2010**, Shanghai, China, 2010.
- HYLAND, B. Preparing for a Linked Data Enterprise. In: WOOD, D. (Ed.). **Linking Enterprise Data: [S.l.]**: Springer US, 2010. cap. 3, p.51-64.
- JAIN, S. C. Environmental scanning in U.S. corporations. **Long Range Planning**, v. 17, n. 2/4, 1984. p. 117-128,
- JENTZSCH, A. et al. Enabling Tailored Therapeutics with Linked Data. **Proceedings of the WWW2009 workshop on Linked Data on the Web (LDOW2009)**, 2009, Madrid, Spain.
- JUHARI, A. S.; STEPHENS, D. Tracing the Origins of Competitive Intelligence Throughout History. **Journal of Competitive Intelligence and Management**, v. 3, n. 4, 2006. p. 61-82.
- KAHANER, L. The basics of competitive intelligence. In: **Competitive Intelligence: How to Gather Analyze and Use Information to Move Your Business to the Top**. 1. New York: Simon & Shuste, 1996. 300.
- LANA, R. A. Inteligência competitiva: fator-chave para o sucesso das organizações no novo milênio. **Revista Inteligência Competitiva**, v. 1, n. 3, p. 23, 2011.
- LANE, R. E. The Decline of Politics and Ideology in a Knowledgeable Society. **American Sociological Review**, v. 31, n. 5, p. 14, 1966.
- MACHADO, C. R. Análise estratégica baseada em processos de Inteligência Competitiva (IC) e Gestão do Conhecimento (GC): proposta de um modelo. 2010. 273. Universidade Federal de Santa Catarina, Florianópolis: 2010.
- MILLER, J. **O milênio da inteligência competitiva**. Bookman, 2002.
- OLIVEIRA, P. H.; MELO, S. C. O. A etapa de planejamento no processo de inteligência competitiva: desafios e algumas reflexões. **Revista Inteligência Competitiva**, v. 2, n. 3, p. 14, 2012. ISSN 2236-210X.
- PASSANT, A. et al. Enhancing Enterprise 2.0 Ecosystems Using Semantic Web and Linked Data Technologies: The SemSLATES Approach. In: WOOD, D. (Ed.). **Linking Enterprise Data**. Springer US, 2010. cap. 5, p.79-102.
- PRESCOTT, J. E. **The Evolution of Competitive Intelligence: Designing a Process for Action**. PROPOSAL Management, 1999.
- PRESCOTT, J. E.; MILLER, S. H. **Inteligência competitiva na prática: técnicas e práticas bem sucedidas para conquistar mercados**. Rio de Janeiro: Campus, 2002.
- ROCKART, J. **Chief Executives Define Their Own Data Needs**. Harvard Business Review, v. 57, n. 2, p. 81-81, 1979. ISSN 00178012.
- SAWKA, K. Whither Analysis? **Competitive Intelligence Magazine**, v. 9, n. 2, 2006.
- SCIP. Strategic and Competitive Intelligence Professionals. 2013. Disponível em: <www.scip.org>. Acesso em: 30 mar. 2013.
- SERVANT, François-Paul. "Linking enterprise data. In Linked Data on Web" in Workshop at the 17th World Wide Web Conference. 2008.
- TYSON, K. W. M. **The Complete Guide to Competitive Intelligence**. Leading Edge Publications, 2002. ISBN 9780966321920.

UNDERWOOD, J. **Competitive Intelligence**. Wiley, 2002.

VAN ZUYLEN, C. H. **Using Inxight Search Extender for Google with Inxight SmartDiscovery Awareness Server**. Inxight Software Inc., Sunnyvale, CA, 2006.

Artigo recebido em 05/09/2015 e aceito para publicação em 28/12/2015

ⁱ <http://lod-cloud.net/state/>

ⁱⁱ Recurso é a informação presente em um Dataset. Estas podem ser uma tripla RDF, um arquivo XML, CSV, uma planilha eletrônica, entre outros.

ⁱⁱⁱ http://enipedia.tudelft.nl/wiki/South_America/Powerplants

^{iv} http://enipedia.tudelft.nl/wiki/Main_Page

^v http://enipedia.tudelft.nl/wiki/Itaipu_Powerplant

^{vi} <http://globalenergyobservatory.org/geoid/42658>

^{vii} <http://en.openei.org/datasets/node/877>

^{viii} <http://en.openei.org/datasets/node/880>

^{ix} <http://en.openei.org/datasets/node/881>

^x http://www.thewindpower.net/country_list_fr.php

^{xi} http://en.openei.org/wiki/Global_Renewable_Resource_Potential

^{xii} <http://en.openei.org/datasets/node/608>

^{xiii} <http://en.openei.org/wiki/Gateway:Hydrogen>

^{xiv} <http://en.openei.org/datasets/node/468>

^{xv} <http://en.openei.org/datasets/node/928>