

Leandro Dal Pizzol

**USO DA *WEB* DE DADOS COMO FONTE DE INFORMAÇÃO
NO PROCESSO DE INTELIGÊNCIA COMPETITIVA
SETORIAL**

Dissertação submetida ao Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento da Universidade Federal de Santa Catarina para a obtenção do Grau de Mestre em Engenharia e Gestão do Conhecimento.

Orientador: Prof. Dr. José Leomar Todesco

Coorientador: Prof. Dr. Denílson Sell

Florianópolis
2014

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Dal Pizzol, Leandro
Uso da Web de Dados Como Fonte de Informação no Processo de Inteligência Competitiva Setorial / Leandro Dal Pizzol ; orientador, José Leomar Todesco ; coorientador, Denilson Sell. - Florianópolis, SC, 214.
138 p.
Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro Tecnológico. Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento.
Inclui referências
1. Engenharia e Gestão do Conhecimento. 2. Engenharia e Gestão do Conhecimento. 3. Inteligência Competitiva. 4. Web de Dados. 5. Web Semântica. I. Leomar Todesco, José. II. Sell, Denilson. III. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento. IV. Título.

Leandro Dal Pizzol

**USO DA WEB DE DADOS COMO FONTE DE INFORMAÇÃO
NO PROCESSO DE INTELIGÊNCIA COMPETITIVA
SETORIAL**

Esta Dissertação foi julgada adequada para obtenção do Título de “Mestre em Engenharia e Gestão do Conhecimento”, e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento da Universidade Federal de Santa Catarina.

Florianópolis, 28 de novembro de 2014.

Prof. Gregório Varvakis, Dr.
Coordenador do Curso

Banca Examinadora:

Prof. José Leomar Todesco, Dr.
Orientador
Universidade Federal de Santa Catarina

Prof. Fernando Ostuni Gauthier, Dr.
Universidade Federal de Santa Catarina

Prof. João Artur de Souza, Dr.
Universidade Federal de Santa Catarina

Prof. Everton Luis Pellizzaro de Lorenzi Cancellier, Dr.
Universidade do Estado de Santa Catarina

Aos meus pais Maria e Stefano, pelo zelo e valores ensinados. A minha esposa Elisângela e a meu filho Arthur pelo amor e pela presença diária em minha vida, que me fortaleceu na conquista deste sonho.

AGRADECIMENTOS

A meus pais, Maria e Stefano a quem muito devo pelo seu carinho pelo zelo e pelo legado ético, moral, intelectual e pelos esforços empregados em minha formação.

Com amor a minha esposa Elisângela e a meu filho Arthur pelo apoio, carinho e compreensão que tiveram durante todo o desenvolvimento deste trabalho.

Ao meu orientador, Prof. Dr. José Leomar Todesco, pelo apoio irrestrito na realização deste trabalho, pela confiança e amizade.

A meus amigos e colegas do EGC pela amizade e parceria ao longo da convivência acadêmica.

Ao Programa de Pós-graduação em Engenharia e Gestão do Conhecimento da Universidade Federal de Santa Catarina, por fornecer o conhecimento necessário para o desenvolvimento desta pesquisa.

Aos professores do referido programa, que compartilharam de seu conhecimento científico e com isso proporcionaram conquistar novos horizontes profissionais.

Por fim, mas não menos importante, a todas as demais pessoas que, mesmo não citadas nominalmente, direta ou indiretamente, contribuíram para a realização dessa dissertação.

"Os resultados provêm do aproveitamento das oportunidades e não da solução dos problemas. A solução de problemas só restaura anormalidade. As oportunidades significam explorar novos caminhos."

(Peter Drucker)

RESUMO

Aproximadamente oitenta por cento da informação necessária em um processo de Inteligência Competitiva (IC) pode ser obtida de fontes abertas. Porém, a falta de semântica desse tipo de fontes dificulta a dedução dos objetos e de seus relacionamentos. Essas dificuldades restringem a tarefa de recuperação de informação, fazendo da captura de conhecimento uma atividade particularmente difícil. A *Web of Data* avança nesse sentido ao possibilitar um espaço global de dados com conexões explícitas entre os conjuntos e com mecanismos padrão para acessar e processar os dados. Assim, este trabalho propõe alinhar o processo de IC à esta fonte de dados. Para tanto, é proposto um modelo composto por tarefas estruturadas de identificação, seleção e classificação da informação baseado em setores econômicos, que objetiva facilitar a recuperação e o uso da informação na etapa de coleta do ciclo de IC. Espera-se com isso que organizações possam explorar novas fontes de conhecimento, diminuir os esforços de coleta devido à estruturação da informação, e conseqüentemente, obter melhor posição estratégica. 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: Engenharia do Conhecimento. Inteligência Competitiva. *Web Semântica*. *Web of Data*.

ABSTRACT

Nearly eighty percent of the information needed for a process of Competitive Intelligence (CI) can be obtained from open sources. However, the lack of semantics of this kind of source complicates the deduction of objects and their relationships. These difficulties restrict the information retrieval task and make knowledge capture a particularly hard activity. The Web of Data moves in this direction, by allowing a global data space with explicit connections between datasets and standard mechanisms to access and process data. So, this paper proposes to align the CI process to this data source. To this end, we propose a model composed for structured identification, selection and classification of information based on economic sectors, which aims to facilitate the retrieval and use of the information in the collection stage of the CI cycle tasks. It is expected that organizations can exploit this new knowledge sources, reduce efforts due to the structuring of information, and hence get better strategic position. The model was validated in the Electricity and Gas sector to identify the requirements of CI and the collection of data belonging to sector.

Keywords: *Knowledge Engineering. Competitive Intelligence. Semantic Web. Web of Data.*

LISTA DE FIGURAS

Figura 1- Processo de Utilização do FCS na determinação de informações.	51
Figura 2 – Exemplo de um motor de busca.	56
Figura 3- Processo de <i>Web Mining</i>	57
Figura 4 - Classificação da <i>Web Mining</i>	58
Figura 5- CommonKADS.	69
Figura 6 - Evolução da <i>Web</i> na visão de Spivac (2007).	71
Figura 7 - Pirâmide de <i>Web Semântica</i>	72
Figura 8 - Tipos de ontologias e relações de especialização.	75
Figura 9 - Classificação cinco estrelas de dados abertos proposta por Berners-Lee.	80
Figura 10 - Crescimento da quantidade de conjunto de dados no formato de <i>Linked Data</i> publicados na <i>Web</i>	83
Figura 11 – Tamanho da nuvem de <i>Linked Data</i> em abril de 2011.	84
Figura 12 – Tamanho da nuvem de <i>Linked Data</i> em abril de 2014.	86
Figura 13 - Modelo conceitual da Proposta.	97
Figura 14 - Classificação da <i>WoD</i> para uso em um processo de IC.	100
Figura 15 - Amostra de um vocabulário setorial.	102
Figura 16 - Interface de Busca da Aplicação.	104
Figura 17 - Interface Data Sources.	104
Figura 18 - Interface Datasets.	105
Figura 19 – Plantas Geradoras da América do Sul.	106
Figura 20 - Planta Geradora de Itaipu.	107
Figura 21 - Itaipu no Global Energy Observatory.	108
Figura 22 – Relatório Quantidade de Triplas.	109
Figura 23 - Relatórios Autores dos <i>datasets</i> e Formatos de dados.	109
Figura 24 - Modelo Relacional.	110
Figura 25 - Data Sources Setor Elétrico.	111

LISTA DE QUADROS

Quadro 1 - Aspectos metodológicos.....	34
Quadro 2- Etapas da pesquisa.....	34
Quadro 3 – Resultado do processo de seleção dos artigos.....	37
Quadro 4- Resultados da revisão bibliográfica.....	37
Quadro 5- Linha do Tempo - Iniciativas de Inteligência na História (2000 a.C. - 1950).....	40
Quadro 6 - Linha do tempo - Iniciativas de Inteligência na História. (1950 - presente).	41
Quadro 7 - Conceitos e Definições a respeito da IC.	45
Quadro 8 - Etapas de estruturação do processo de IC.....	49
Quadro 9 - Classificação das Informações.....	53
Quadro 10 - Produtos de Inteligência.	62
Quadro 11 - Produtos de Inteligência (continuação).....	63
Quadro 12 - Tipos de Ontologias.....	76

LISTA DE TABELAS

Tabela 1- Correspondência entre tamanho dos círculos e o número de triplas ..85	
Tabela 2 - Número de conjuntos de dados, quantidade de triplas e quantidade de ligações RDF por tópico de domínio.87	87
Tabela 3 – Classificação dos tipos de <i>Linked Data</i>90	90
Tabela 4- Principais tarefas de <i>Linked Government Data</i>92	92
Tabela 5 - Distribuição dos <i>Data Sources</i>101	101
Tabela 6 – Necessidades de IC e <i>datasets</i> identificados.112	112

LISTA DE ABREVIATURAS E SIGLAS

a.C. - Antes de Cristo
ANEEL - Agência Nacional de Energia Elétrica
APLs - Arranjos produtivos locais
BI - *Business Intelligence*
BI - *Business Intelligence*
CKAN - *Comprehensive Knowledge Archive Network*
CNAE - Classificação Nacional de Atividades Econômicas
CRIA - Conjunto, Relações, Instâncias e Axiomas
CRM - *Customer Relationship Management*
CSV *comma separated values*
d.C. - Depois de Cristo
EC - Engenharia do Conhecimento
EGC - Engenharia e Gestão do Conhecimento
EIA - *Energy Information Administration*
EUA - Estados Unidos da América
FCS - Fatores Críticos de Sucesso
GC- Gestão do Conhecimento
GDR - *Geothermal Data Repository*
HTML - *HyperText Markup Language*
IA - Inteligência Artificial
IBGE - Instituto Brasileiro de Geografia e Estatística
IC - Inteligência Competitiva
IEEE - *Instituto de Engenheiros Eletricistas e Eletrônicos*
IRI - *Internationalized Resource Identifier*
JSON - JavaScript Object Notation
KADS - *Knowledge Acquisition and Documentation Structuring*
KITs - *Key Intelligence Topics*
LOD - *Linking Open Data*
LOGD - *Linked Government Data*
MPes - Micro e pequenas empresas
OLAP - *Online Analytical Processing*
OML - *Ontology Markup Language*
PPEGC - Programa de Pós-graduação em Engenharia e Gestão do Conhecimento
PSM - *Problem-Solving Method*
RDF - *Resource Description Framework*
SBC - Sistemas Baseados em Conhecimento
SCPI - *Strategic and Competitive Intelligence Professionals*
SDMX - *Statistical Data and Metadata Exchange*

SGML - *Standard Generalized Markup Language*
SIS - *Sistema de Inteligência Setorial*
SKOS - *Simple Knowledge Organization System*
SPARQL - *Simple Protocol and RDF Query Language*
SWOT - *Strengths, Weaknesses, Opportunities and Threats*
UFSC - *Universidade Federal de Santa Catarina*
URI - *Uniform Resource Identifier*
URL - *Uniform Resource Locator*
WCM - *Web Content Mining*
WoD – *Web of Data*
WSM - *Web Structure Mining*
WUM - *Web Usage Mining*
WWW - *World Wide Web*
XLS – *Formato de arquivos do Microsoft Excel*
XML - *eXtensible Markup Language*

SUMÁRIO

1	INTRODUÇÃO	27
1.1	APRESENTAÇÃO DO PROBLEMA DA PESQUISA	28
1.2	OBJETIVOS	29
1.1.1	Objetivo Geral.....	30
1.1.2	Objetivos Específicos	30
1.3	JUSTIFICATIVA	30
1.1.3	Interdisciplinaridade e Aderência ao EGC.....	31
1.4	DELIMITAÇÕES DA DISSERTAÇÃO	32
1.5	PROCEDIMENTOS METODOLÓGICOS	33
1.5.1	Caracterização da Pesquisa	33
1.5.2	Etapas de Elaboração da Pesquisa	34
1.5.3	Busca Sistemática.....	35
1.5.4	Definição dos Termos de Busca	35
1.5.5	Consulta as Bases de Dados	36
1.5.6	Padronização e Análise dos Dados	36
1.5.7	Resultados da Busca Sistêmica	37
1.6	ESTRUTURA DA DISSERTAÇÃO	38
2	INTELIGÊNCIA COMPETITIVA	39
2.1	ORIGENS DA INTELIGÊNCIA COMPETITIVA.....	39
2.2	CONCEITOS	44
2.3	DESMISTIFICAÇÕES A RESPEITO DA IC.....	46
2.4	O PROCESSO DE INTELIGÊNCIA COMPETITIVA	48
2.4.1	Identificação das Necessidades de Informação.....	49
2.4.2	Planejamento.....	52
2.4.3	Coleta	52
2.4.3.1	Técnicas de Coleta.....	54
2.4.3.2	Classificação das ferramentas de coleta.....	54
2.4.3.3	Motores de busca	55
2.4.3.4	<i>Web mining</i>	57
2.4.3.5	Recuperação de informação baseada em ontologia.....	58
2.4.4	Análise.....	59

2.4.4.1	Produtos de IC	60
2.4.5	Disseminação.....	64
2.4.6	Avaliação	64
2.5	CONSIDERAÇÕES SOBRE O CAPÍTULO	65
3	ENGENHARIA DO CONHECIMENTO E WEB DE	
DADOS	67
3.1	ENGENHARIA DO CONHECIMENTO.....	67
3.1.1	CommonKADS	68
3.2	WEB E WEB 2.0	70
3.3	WEB SEMÂNTICA.....	71
3.4	ONTOLOGIAS	73
3.4.1	Uso e benefícios de Ontologias.....	74
3.4.2	Classificação de Ontologias.....	74
3.4.3	Componentes de uma Ontologia.....	77
3.5	DADOS ABERTOS	77
3.5.1	Classificação dos Dados Abertos	79
3.6	WEB OF DATA.....	81
3.6.1	Crescimento da Web de Dados.....	82
3.6.2	Topologia da Web de Dados	87
3.7	LINKED DATA	88
3.7.1	Princípios de <i>Linked Data</i>	88
3.7.2	Classificação de <i>Linked Data</i>.....	89
3.7.2.1	<i>Linked Government Data</i>	91
3.7.2.2	<i>Enterprise Linked Data</i>	92
3.7.2.3	<i>Statistical Linked Data</i>	93
3.7.2.4	Outras classificações de <i>Linked Data</i>	94
3.8	CONSIDERAÇÕES SOBRE O CAPÍTULO	95
4	MODELO PROPOSTO E SUA APLICAÇÃO.....	97
4.1	ETAPA DE CLASSIFICAÇÃO.....	97
4.1.1	Seleção das Fontes	98
4.1.2	Seleção dos Setores Econômicos	99
4.1.3	Análise de Vocabulários e Definição de Palavras-Chave	102

4.2	DESCRIÇÃO DA FERRAMENTA WEB	103
4.2.1	Criação do Repositório de Dados	109
4.3	APLICAÇÃO DO MODELO NO SETOR DE ELETRICIDADE E GÁS	110
4.3.1	Identificação dos Requisitos de IC	111
4.3.2	Coleta dos Dados.....	112
5	CONCLUSÕES E TRABALHOS FUTUROS.....	115
5.1	CONCLUSÕES	115
5.2	TRABALHOS FUTUROS	118
	REFERÊNCIAS.....	119

1 INTRODUÇÃO

A capacidade de lidar com grandes quantidades de informação é um requisito essencial para uma organização. Seja qual for o setor no qual a empresa encontra-se inserida, atividades intensivas em informação estarão presentes, e em alguns casos, constituirão uma das partes mais importante do negócio.

Para Hyland (2010), a arquitetura *Web* começou a se popularizar como fonte de informação em meados da década de 1990, e assim segue até hoje. Essa arquitetura traz uma grande variedade de fontes de dados interligadas que constituem uma teia de informação de livre acesso.

Contudo, acessar conteúdo relevante de forma rápida e precisa em um ambiente com essas características é um desafio que motiva iniciativas de estruturação desde os primórdios da *Web*. Exemplo disso é a *Web Semântica* proposta por (BERNERS-LEE; HENDLER; LASSILA, 2001). A *Web Semântica* acrescentou conteúdos que podem ser compreendidos por computadores a *Web* tradicional. Isso busca solucionar o problema decorrente da falta de significado e de acurácia resultantes da grande quantidade de informação disponível (SHADBOLT; BERNERS-LEE; HALL, 2006; LIU et al., 2011). Assim, ao tornar os dados compreensíveis por máquina faz-se da *Web* um enorme repositório distribuído de dados atuando em paralelo aos sistemas de arquivos tradicionais.

Seguindo o caminho evolutivo das tecnologias *Web*, surge o conceito da *Web de Dados (WoD)* formada por *Linked Data* (BERNERS-LEE, 2006). O termo *Linked Data* refere-se à publicação e ligação de dados estruturados na *Web* de forma escalável e generalizada, vinculando informações que antes não eram conectadas em um espaço de dados global. O pressuposto básico por trás da *Web de Dados* diz que a utilidade e o valor desses dados aumentam a medida que forem acessados e recombinaos uns aos outros (HEATH; BIZER, 2011).

Segundo Berners-Lee (2007), a *Web de Dados* forma um grafo mundial gigante que consiste em bilhões de declarações em formatos estruturados provenientes de várias fontes, cobrindo temas dos mais variados domínios. Além disso, a *Web de Dados* pode ser vista como uma camada adicional fortemente interligada aos documentos da *Web* tradicional tendo muitas das mesmas propriedades (HEATH; BIZER, 2011).

Para Dishman e Calof (2008), um processo que envolva a coleta, análise e comunicação de informações ambientais é essencial para ajudar na tomada de decisão estratégica. Brouard (2006), propoe o

desenvolvimento de um instrumento para medir a capacidade das empresas de varrer sua informação ambiental. Esta informação, no entanto, deve ser parte de um processo ativo e estruturado de inteligência (CALOF; WRIGHT, 2008; SAAYMAN et al., 2008).

Para Lana (2011), com as novas tecnologias e os recursos proporcionados por elas tomando forma velozmente, a capacidade de colher informações é, hoje, um desafio maior do que em qualquer outra época.

Segundo Kolb e Miller (2002), quando, adequadamente pensada como um processo, uma maneira de pensar e não simplesmente como uma atividade de coleta de dados, esta fase da IC ajuda os executivos a administrarem os riscos presentes em seus negócios.

Para melhorar o processo de coleta de informação, organizações desenvolvem esforços para adaptar seus métodos tradicionais de obtenção às características da *Web* de Dados. Um dos processos que pode valer-se desta nova fonte de informação, está a Inteligência Competitiva ou simplesmente IC.

Na visão da *Strategic and Competitive Intelligence Professionals* (SCIP), IC é um processo sistemático e ético que consiste em coletar, analisar, disseminar e gerenciar as informações sobre o ambiente externo que podem afetar os planos, as decisões e as ações da empresa (SCIP, 2013).

Para que a informação realmente atinja as necessidades dos tomadores de decisão, esta precisa ser relevante e confiável. Logo, o emprego de informação estruturada e com valor agregado pode produzir melhores resultados. Esses resultados, de acordo com Kahaner (1996) e Sawka (2006), permitem à gerência realinhar o planejamento estratégico, o posicionamento e a eficácia competitiva.

Apoiado neste cenário, no próximo capítulo é apresentado o problema da pesquisa que norteou este trabalho.

1.1 APRESENTAÇÃO DO PROBLEMA DA PESQUISA

Para MIRANDA (2011), as iniciativas governamentais de transparência e acesso à informação como a Lei nº 12.527 que regulamenta o direito de acesso à informação pública (BRASIL, 2012), estimularam a abertura de dados dos mais diversos setores da administração pública na *Web*, e posteriormente, por setores privados da economia.

Contudo, os dados disponibilizados costumam ser publicados de maneira não estruturada, fragmentada e em diferentes formatos. Isto

dificulta a dedução dos objetos e dos relacionamentos entre eles, já que as páginas não carregam semântica suficiente sobre o conhecimento nelas contido. Ou seja, o foco está centrado no documento, e não nos dados nele contidos (ZHANG; WU; WANG, 2007; NEMRAVA et al., 2008; PASSANT et al., 2010).

Assim, o conteúdo apresenta problemas de ambiguidade e heterogeneidade, restringindo a tarefa de recuperação de informação e fazendo da captura de conhecimento e seu reuso uma atividade particularmente difícil (VANDER WAL, 2007; PASSANT et al., 2010).

O uso de *Linked Data* visa interligar esses dados conduzindo à criação de um espaço global de dados no qual informações de diversas fontes e domínios diferentes são relacionadas (BIZER; HEATH; BERNERS-LEE, 2009).

Com isso, a *Web de Dados (WoD)* formada pela aplicação dos princípios de *Linked Data* surge como uma possível fonte de informação externa para o processo de IC. Contudo é necessário responder a questões tais como: De que forma organizar e classificar a informação presente na *WoD*? De que forma a *WoD* pode facilitar o a etapa de coleta de informação do processo de IC? Somado a isso, quais são as tecnologias necessárias e as dificuldades de se trabalhar com os dados coletados?

Para responder as perguntas anteriores é proposto um modelo conceitual de identificação, coleta e armazenamento de informação proveniente da *Web de Dados*, que possam ser transformadas em produtos pertinentes de IC.

Assim, esta dissertação pretende verificar o emprego da *Web de Dados* em um processo de Inteligência Competitiva, organizando a informação nela contida, para desta forma reduzir os esforços de coleta e armazenamento de informação dentro do processo de IC.

1.2 OBJETIVOS

Partindo do pressuposto de que os dados provenientes da *Web de Dados* são uma fonte de informação relevante para o desenvolvimento de processos de IC, a questão central a ser verificada pelo presente trabalho é averiguar de que forma o processo de coleta e catalogação da informação pode ser melhorado. Nesse contexto, o presente trabalho possui os seguintes objetivos:

1.1.1 Objetivo Geral

Estabelecer uma abordagem baseada em semântica para o uso *Web* de Dados na etapa de coleta de informação do ciclo de IC.

1.1.2 Objetivos Específicos

- Apresentar uma classificação de setores econômicos para catalogação da *Web* de Dados;
- Realizar a catalogação dos *data sources* da *Web* de Dados segundo a classificação definida;
- Propor um modelo para uso de informação da *Web* de Dados na etapa de coleta do ciclo de IC;
- Verificar o modelo proposto em um setor econômico.

1.3 JUSTIFICATIVA

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 aponta para a crescente relevância social do conhecimento científico dentro da sociedade e de seus membros (LANE, 1966).

Passados mais de cinquenta anos, avanços tecnológicos e as facilidades de acesso a informação, fazem da época em que vivemos a verdadeira sociedade do conhecimento. De acordo com Vaitsman (2001) para tomar decisões baseadas em informação é necessário primeiro identificar o que é relevante, as melhores fontes, e também o que pode ser eliminado.

Na visão de Vaitsman (2001), e na de Davenport (1998), é necessário gerenciar e estruturar a informação por meio de um processo que contemple a obtenção, distribuição e o uso dessa informação. Ao relacionarmos esses preceitos com as características da *Web* de Dados, observa-se que esta tem potencial para facilitar o acesso a dados substancialmente mais relevantes em processos informacionais.

Segundo Miller (2002), os profissionais de IC que trabalham em organizações de grande porte podem obter até 80 por cento das informações de inteligência necessária de fontes abertas. A *Web* de Dados possibilita conexões explícitas entre os conjuntos de dados na forma de *data links* (BIZER; HEATH; BERNERS-LEE, 2009). Isto é possível usando formatos de representação e mecanismos padrões de acesso (CYGANIAK; REYNOLDS; TENNISON, 2010). A

representação padronizada e os mecanismos de acesso permitem que ferramentas genéricas, como navegadores da *Web Semântica* e motores de busca, sejam usados para acessar e processar os dados (JENTZSCH et al., 2009).

Portanto, para trabalhar com dados provenientes dessa nuvem, são necessárias não somente ferramentas computacionais, mas meios estruturados para identificar e recuperar a informação. Além disso, é preciso alinhar o processo de IC a este tipo de dado.

Para Prescott e Miller (2002), um ciclo convencional de IC tem três elementos: coleta, análise e disseminação dos dados. Segundo os autores, apenas um terço do tempo é gasto em análise. O tempo restante é empregado na coleta de informações e na disseminação de inteligências pelos usuários, tarefas que sozinhas não acrescentam muito valor a organização. Recomendação comum é gastar um terço do tempo disponível coletando informação, deixando o resto do tempo para análise e ação.

A proposta desenvolvida neste trabalho visa identificar e classificar os *data sources* presentes na *Web* de Dados para que estes sejam utilizados na etapa de coleta de informação do ciclo de IC. Com isso, organizações podem explorar novas fontes de conhecimento, diminuir os esforços de captura devido à estruturação da informação e consequentemente, obter melhor posição estratégica.

1.1.3 Interdisciplinaridade e Aderência ao EGC

Esta dissertação está contextualizada na área de concentração de Engenharia do Conhecimento (EC), do Programa de Pós-graduação em Engenharia e Gestão do Conhecimento (PPEGC) da Universidade Federal de Santa Catarina (UFSC). Sua linha de pesquisa é a Engenharia do Conhecimento Aplicada às Organizações. Esta linha estuda a concepção, desenvolvimento e implantação de soluções da Engenharia do Conhecimento nas organizações (EGC, 2013).

Ao considerarmos o contexto das atividades intensivas em conhecimento dentro de uma organização, a EC deixa de ser uma subárea de Inteligência Artificial (IA) usada para construção de sistemas de conhecimento específicos. Para Schreiber et al. (2002), a EC permite focar as oportunidades e gargalos a respeito de como as organizações desenvolvem, distribuem e aplicam seus recursos de conhecimento.

Processos de IC caracterizam-se por necessitarem de uma grande quantidade de informação que demanda grande esforço para ser coletada. O propósito deste trabalho é disponibilizar uma nova fonte de

informação com a classificação da *Web* de Dados para apoiar a obtenção de conhecimento no processo de Inteligência Competitiva. A interdisciplinaridade é perseguida com a utilização da *Web* de Dados em um processo de Inteligência Competitiva. Para tanto essas duas disciplinas relacionam seus conteúdos com o objetivo de aprofundar o conhecimento.

A aderência deste trabalho ao objetivo do programa pode ser contextualizada a partir dos objetivos de pesquisa e principal do programa:

O objeto de pesquisa do EGC refere-se aos macroprocessos de explicitação, gestão e disseminação do conhecimento. Estes incluem os processos de criação (e.g., inovação de ruptura), descoberta (e.g., redes sociais), aquisição (e.g., inovação evolutiva), formalização/codificação (e.g., ontologias), armazenamento (e.g., memória organizacional), uso (e.g., melhores práticas), compartilhamento (e.g., comunidades de prática), transferência (e.g., educação corporativa) e evolução (e.g., observatório do conhecimento). No EGC conhecimento é pesquisado enquanto fator de produção, gerador de riqueza/valor e de equidade social. Deste modo, o objetivo do EGC consiste em investigar, conceber, desenvolver e aplicar modelos, métodos e técnicas relacionados tanto a processos/bens/serviços como ao seu conteúdo técnico-científico [...].

1.4 DELIMITAÇÕES DA DISSERTAÇÃO

Esta pesquisa propõe o uso da *Web* de Dados como fonte de informação em um processo de Inteligência Competitiva. A proposta integra ferramentas de Engenharia do Conhecimento com o intuito de acessar informação de forma mais rápida e qualificada.

O trabalho propõe o arranjo de tecnologias de EC como *Web* de Dados, *Web* Semântica e vocabulários de dados para facilitar a coleta e o armazenamento de informação dentro do processo de IC.

A validação da proposta consiste da classificação e catalogação das fontes de dados (*data sources*) presentes na *Web* de Dados em um setor econômico pré-definido. Em seguida, os dados resultantes dessa classificação serão usados para responder a necessidades de IC identificadas. Já a abrangência teórica do estudo é composta pela

integração dos conceitos da *Web* de Dados e da Inteligência Competitiva.

1.5 PROCEDIMENTOS METODOLÓGICOS

Segundo Gil (2010), na elaboração de trabalhos científicos é necessário estabelecer procedimentos, definir a sequência de atividades e o caminho a ser seguido para alcançar os objetivos finais.

Estabelecer os métodos e procedimentos que serão adotados é essencial, pois eles orientam o desenvolvimento da pesquisa no que se refere à coleta de dados, à análise e à interpretação dos resultados (MARCONI; LAKATOS, 2002).

O estabelecimento destas etapas é impreterível também para que a pesquisa seja válida dentro do método científico e envolvem tanto o estudo dos conceitos, quanto aplicações práticas (GIL, 2010).

Na sequência, a pesquisa é caracterizada quanto à natureza, objetivo, abordagem e procedimentos técnicos adotados.

1.5.1 Caracterização da Pesquisa

Esta pesquisa caracteriza-se em sua natureza como aplicada, tendo por objetivo gerar conhecimentos práticos e dirigidos à solução de um problema específico (MARCONI; LAKATOS, 2002).

Quanto à finalidade, trata-se de um estudo exploratório. Este tipo de pesquisa tem como objetivo proporcionar maior familiaridade com o problema, pois visa torná-lo mais explícito ou construir hipóteses.

No que se refere à abordagem do problema, esta pode ser definida como qualitativa, não cabendo aqui interpretação por meio de métodos e técnicas estatísticas, pois existe uma subjetividade que não pode ser traduzida em números. Este tipo de pesquisa busca entender o que acontece no mundo real e atribuir significados a esses acontecimentos (SILVA; MENEZES, 2005). As amostras utilizadas visam produzir informações aprofundadas e ilustrativas: sejam pequenas ou grandes, o que importa é que ela seja capaz de produzir novas informações (DESLAURIERS, 1991).

De acordo com Fonseca (2002), a pesquisa científica é o resultado de um inquérito ou exame minucioso realizado com o objetivo de resolver um problema, recorrendo a procedimentos científicos. Para esta pesquisa foram adotados mais de um procedimento científico. O Quadro 1 a seguir apresenta este e os demais aspectos metodológicos.

Quadro 1 - Aspectos metodológicos.

CLASSIFICAÇÃO	CARACTERÍSTICA
Natureza da Pesquisa	Aplicada
Objetivo da Pesquisa	Exploratória
Abordagem	Qualitativa
Procedimentos	Levantamento Bibliográfico Desenvolvimento Referencial teórico Elaboração da Proposta Verificação dos Resultados

1.5.2 Etapas de Elaboração da Pesquisa

Esta seção apresenta as etapas de desenvolvimento da pesquisa. Inicialmente foi definido o tema inicial de pesquisa. Para tanto, os temas Inteligência Competitiva e *Web of Data* foram estudados visando avaliar sua aplicação conjunta. Desta identificação surge o problema de ambiguidade e heterogeneidade de dados apresentado no Item 1.1. Em seguida, foi realizada uma revisão preliminar na tentativa de encontrar respostas às necessidades identificadas observando-se questões como: De que forma assegurar a qualidade da informação presente na *Web*? Ou ainda, quais as tecnologias necessárias e as dificuldades em assegurar a qualidade dos dados coletados. O Quadro 2 apresenta as etapas da pesquisa.

Quadro 2- Etapas da pesquisa.

ETAPAS DA PESQUISA
1 – Definição do Tema
2 – Revisão Preliminar
3 – Formulação do Problema
4 – Revisão da Literatura
5 – Proposição do Modelo
6 – Verificação do modelo

Para a revisão da literatura, foi utilizado o critério histórico que por característica documenta o desenvolvimento da pesquisa em determinada área. Foram delineadas algumas questões a serem seguidas:

- ✓ Identificação de bases de dados que serão consultadas;
- ✓ Definição de palavras-chave que serão utilizadas;
- ✓ Seleção do material encontrado, considerando o foco do estudo;
- ✓ Análise do material selecionado;
- ✓ Posicionamento da pesquisa;

A revisão de literatura permitiu o refinamento de pontos do projeto e a definição de etapas posteriores. A partir destas constatações definiu-se o problema e o modelo conceitual da proposta de pesquisa detalhado no Item 4 MODELO . Neste item serão mostradas tarefas como a classificação e escolha dos *data sources* e a definição do setor econômico que será usado para exemplificar a proposta.

1.5.3 Busca Sistemática

A busca sistemática teve como objetivo sintetizar os resultados de pesquisas primárias sobre os temas que compõem este estudo. Isso serviu para definir estratégias usadas para reduzir viés e possíveis erros de interpretação no momento da junção dos temas. Além disso, a busca sistemática serviu como arcabouço teórico e para identificar o assunto central que compõe o problema de pesquisa.

Esta etapa foi dividida em duas fases em conformidade com os procedimentos metodológicos e recomendações propostas por Macias-Chapula (1998), Vanti (2002), Guedes e Borschiver, (2005) e Kobashi e Santos (2006).

Na primeira fase foi realizada a busca sistemática da literatura com o objetivo de compreender de que forma os termos se relacionam e também de delimitar o problema de pesquisa. Já a segunda fase consiste da análise descritiva dos dados coletados que foram usados na elaboração do documento final.

1.5.4 Definição dos Termos de Busca

Os termos “Inteligência Competitiva” e “Web de Dados” em suas variantes da língua inglesa “*Competitive Intelligence*” e “*web of data*” são o ponto de partida para as buscas. Definidos os termos, o passo seguinte foi a escolha das bases de dados usadas.

1.5.5 Consulta as Bases de Dados

As buscas foram realizadas em quatro bases de dados científicos internacionais: EBSCO¹, IEEE², *Scopus*³ e *Web of Science*⁴. A escolha dessas bases justifica-se por terem caráter multidisciplinar (ALMEIDA, 2006). Ainda por retornarem registros consistentes sobre os termos pesquisados e permitirem a exportação dos dados para *softwares* gerenciadores de referências bibliográficas. As únicas restrições aplicadas aos resultados da pesquisa dizem respeito ao idioma, Inglês ou Português, e que os trabalhos fossem publicados em revistas ou congressos científicos.

1.5.6 Padronização e Análise dos Dados

O resultado das consultas foi exportado e carregado em uma ferramenta de gerenciamento de referências bibliográficas formando um conjunto único com todos os resultados.

A esse conjunto foram aplicados os critérios de seleção anteriormente definidos, excluídos artigos sem autoria, duplicados, ou seja, artigos que estavam indexados em mais de uma base de dados, sem texto completo ou que não possuem acesso gratuito.

Em seguida ocorreu a leitura dos resumos para eliminar trabalhos que não pertencem ao contexto de estudo. Caso esta leitura não fosse suficiente para determinar o escopo do trabalho, prossegue-se com a leitura do texto completo.

Para garantir a consistência das informações recolhidas nas diferentes bases de dados fez-se necessária à padronização dos resultados. Informações como o nome do autor foram revisadas para que não apresentassem cadastros duplicados.

Com os dados padronizados foi possível analisar e identificar os principais autores, as instituições que desenvolvem estudos sobre os temas abordados no trabalho. Outra importante fonte de informação analisada foram as referências citadas nos artigos que possibilitam identificar as bases teóricas em que se fundamentam os trabalhos (GUEDES, 2012; VILAN FILHO; ARRUDA; PERUCCHI, 2012).

¹ <http://search.ebscohost.com>

² <http://ieeexplore.ieee.org/Xplore/guesthome.jsp>

³ <http://www.scopus.com/home.url>

⁴ <http://apps.webofknowledge.com>

1.5.7 Resultados da Busca Sistemática

A busca foi realizada durante o mês de julho de 2013 e inicialmente retornou 206 resultados na base IEEE, 199 na base *Scopus*, 76 resultados na Ebsco e 27 na *Web of Science*. Ao total Foram identificados 508 artigos. O Quadro 3 apresenta o resultado final do processo de seleção dos artigos.

Quadro 3 – Resultado do processo de seleção dos artigos.

Base de dados	Publicações identificadas	Publicações repetidas	Fora de Contexto	Sem texto completo	Total
WoS	27	10	4	3	10
Ebsco	76	5	49	7	15
IEEE	206	5	164	7	30
Scopus	199	16	143	10	30
TOTAL	508	36	360	27	85

Dos 85 artigos selecionados 31 pertencem a periódicos científicos e 54 estão em anais de eventos. 83 são publicados em língua inglesa e dois em Português. Esses trabalhos foram escritos por 230 autores pertencentes a 87 instituições de 22 países diferentes.

Quadro 4- Resultados da revisão bibliográfica.

Informações Bibliométricas	Quantidade
Publicações	85
Autores	230
Fontes de informações	60
Instituições	87
Países	22
Palavras-chave	95
Idiomas	2
Referências	1471

Cabe ressaltar que dos 508 artigos que foram identificados na busca inicial, 360 estavam fora do escopo. Grande parte desse número apresentava como tema principal *Business Intelligence* e não Inteligência Competitiva. Uma possibilidade para isso é a possível confusão na definição dos temas. Outra possibilidade, mais concreta, é o cadastro proposital por parte dos autores dos dois termos nas bases de conhecimento para que seus artigos retornem em buscas sobre os dois termos.

Apesar de serem termos relacionados, os dois apresentam processos bastante distintos. O primeiro é empregado pelas empresas de tecnologia da informação como um conjunto de ferramentas utilizadas para auxiliar nos negócios tais como: *data warehouse*, *data mining*, CRM, ferramentas OLAP e outras. Já a IC é um processo que engloba a obtenção e tratamento de informação, nas quais os produtos resultantes das atividades de BI estão inseridos.

1.6 ESTRUTURA DA DISSERTAÇÃO

Este trabalho é dividido em cinco capítulos que buscam investigar e estruturar o uso do conhecimento proveniente da *Web* de Dados para a construção de um processo de coleta de informação mais consistente dentro do ciclo de IC.

O primeiro Capítulo apresenta o tema, o problema de pesquisa, os objetivos geral e específicos, as justificativas da pesquisa, sua interdisciplinaridade, a aderência da pesquisa ao Programa de Pós-graduação em Engenharia e Gestão do Conhecimento, as demarcações da pesquisa e os procedimentos metodológicos.

O segundo e o terceiro Capítulos destinam-se à revisão teórica, sendo que o segundo aborda a IC em seu contexto histórico e atual, e o terceiro capítulo traz a revisão sobre a *Web* de Dados.

O quarto capítulo é destinado a apresentação do modelo conceitual proposto, da descrição da ferramenta desenvolvida e de suas respectivas verificações. Por último, no quinto Capítulo são apresentadas as conclusões e os trabalhos futuros.

2 INTELIGÊNCIA COMPETITIVA

Este Capítulo aborda os conceitos relacionados à Inteligência Competitiva. Inicialmente é apresentada uma perspectiva histórica sobre a IC e seus conceitos. Em seguida, evidencia-se seu processo e as técnicas de coleta de informação. Por fim são apresentados os produtos de inteligência e as diferenças entre IC e outras áreas de conhecimento.

2.1 ORIGENS DA INTELIGÊNCIA COMPETITIVA

A IC é vista como uma disciplina relativamente nova. No entanto, o uso de inteligência não é um fenômeno empresarial ou atividade acadêmica recente (FLEISHER; BENSOUSSAN, 2003).

Para Juhari e Stephens (2006), a base para a maioria das disciplinas é normalmente encontrada em sua origem e história. Portanto, é importante olhar para trás, para incidências históricas de processos que inspiraram a noção da necessidade de inteligência.

Segundo Kelley (1968), os militares a muito tempo reconhecem a importância da inteligência e, assim, alcançaram um estágio avançado de expertise nessa área. Isso faz dos militares fonte da comprovação do potencial da aplicação dos princípios da administração da informação.

Underwood (2002), sugere que o uso da inteligência remonta de 1000 a.C. Isso é evidenciado por alguns registros escritos de uso da inteligência para fins militares. Juhari e Stephens (2006), vão ainda mais longe apontando atividades datadas de 3 mil anos atrás.

Explicitamente aplicada a negócios, o uso da inteligência data de 1960. A década de 1980 viu a introdução de funções formais de negócios na coleta de inteligência. Já a década de 1990 viu a introdução da tecnologia, mais especificamente a Internet, como um meio para aumentar a profundidade e amplitude da IC. Esta é a década em que a IC se tornou a entidade dinâmica e complexa que é hoje (UNDERWOOD, 2002). Inspirado nas cronologias de Underwood (2002) e Juhari e Stephens (2006) é elaborada uma representação cronológica sobre a incidência de iniciativas de inteligência que retrata sua evolução ao longo da história.

Quadro 5- Linha do Tempo - Iniciativas de Inteligência na História (2000 a.C. - 1950)



2000 a.C. Império Chinês

Envio de eunucos espíões a países com destaque produtivo para descobrir seus segredos.



202 a.C. Cipião

Envia embaixadores sob o pretexto da diplomacia para reunir informações antes da batalha contra o Império Cartaginês.



500 a.C.

Sun Tzu

É escrito "A Arte da Guerra", ensaios sobre a Inteligência militar.



1000-1200 Cruzadas

As cruzadas para o oriente médio usaram estratégias e ferramentas de inteligência



1536-1598

Toyotomi Hideyoshi

Conhecido como Napoleão do Japão, incorpora com sucesso as teorias de Sun Tzu.



1777-1836

Nathan Rothschild

Empregou uma matriz de inteligência que ajudou a Inglaterra a derrotar Napoleão, salvou a bolsa de Londres da quebra e lhe trouxe fortuna.



1926

Charles C. Parlin

Parlin é nomeado o primeiro diretor de marketing. Introdução do Conceito americano de monitoração de competidores e forças competitivas.



1957

Washington Platt

General Americano publica o livro "Strategic Intelligence Production".

2000 a.C. 1000 a.C.

500 a.C. 0 a.C.

500 1000

1500

1700

1800

1900

1950



1000 a.C.

Conquista de Canaã

Moisés envia espíões antes da batalha para identificar vulnerabilidades do inimigo.



220

Zhuge Liang

O mestre estrategista chinês descreve suas estratégias de batalha em um documento com 46 artigos.



1600

R. I. Wickham

O explorador inglês se infiltrou na produção chinesa de chá para contrabandear plantas que posteriormente foram cultivadas na Índia.



1879

Marketing

Nos EUA ocorre a primeira manifestação sobre atividades de marketing.



550

Imperador Justiniano

Monges que trabalhavam para o Imperador bizantino contrabandeavam bichos-da-seda dentro de bambus dando início a indústria da seda neste país.



1940

II Guerra Mundial

Início do uso de tecnologias para inteligência militar.

Quadro 6 - Linha do tempo - Iniciativas de Inteligência na História. (1950 - presente).



1966

Richard M. Greene

Lança o livro "Strategic Intelligence and the Shape of Tomorrow".



1989

Fim da Guerra Fria

Com o fim do conflito profissionais militares de inteligência migraram para as indústrias.



2000

Publicações sobre IC

É registrado um grande crescimento das publicações sobre IC.

1960-1970
Japão

Sistema de produção just-in-time desenvolve importantes estudos de competidores.



1980

Coleta de Informação

Inúmeras empresas introduzem funções de coleta de informação.



1990

Computação/Internet

O uso de TI torna-se popular em processos de inteligência.



2000

Formalização da IC

Formalização de processos, sistemas e produtos de IC.

1960

1970

1980

1990

2000



1959

Burton H. Alden

Competitive Intelligence: Information, Espionage, and Decision-making: A Special Report for Businessmen.



1985

Leonard M. Fuld

Lança o livro "Competitor Intelligence".



1986

SCIP

Fundada a Society of Competitive Intelligence Professionals.



1999

Reconhecimento IC

A IC é defendida como essencial para os negócios.



2006

Open Government

Iniciativas de governo aberto ao redor do mundo. No Brasil, aprovação em 2011 da Lei 12.527 que assegura o direito de acesso à Informação



1961

William M. McGovern

Publica o livro "Strategic Intelligence and the shape of tomorrow".



2000

Complexidade

A inteligência se torna mais complexa, com necessidades baseadas em métodos e ferramentas de IC.

As origens da IC podem ser encontradas três mil anos atrás. O Império Chinês enviava espiões disfarçados de eunucos a países com algum desenvolvimento industrial para obter informações e vantagens competitivas (JUHARI; STEPHENS, 2006).

De acordo com os escritos bíblicos Moisés antes de conquistar Canaã enviou doze espiões, que lá permaneceram durante quarenta dias identificando pontos fortes e vulnerabilidades das pessoas que ocuparam a terra por eles desejada (WALTON, 2010).

O tema inteligência é muito vasto, porém seja qual for sua aplicação, é impossível traçar sua história sem mencionar suas origens militares (PRESCOTT, 1999). Uma das primeiras referências consistentes ligadas à inteligência é a obra “A Arte da Guerra” escrita por Sun Tzu (2006). Constituído por um conjunto de ensaios escritos por volta de 500 a.C., esta obra serviu de base para muitos dos avanços na inteligência militar, sendo popular até hoje com a adaptação de seus dizeres para aplicações no meio empresarial.

No ano de 202 a.C. o general romano Cipião obtém sucesso contra Aníbal de Cartago, na batalha de Zama mesmo em desvantagem numérica. Com 30 mil soldados, Cipião batalha contra 35 mil soldados de infantaria e 10 mil de cavalaria. Sob o pretexto da diplomacia, Cipião enviou embaixadores antes da batalha para reunir informações que seriam posteriormente úteis em seu triunfo. “Estude seu inimigo até que você esteja absolutamente certo de seus hábitos. Em seus hábitos você vai encontrar sua fraqueza” (O’LEARY, 2008).

Em 220 d.C. o mestre estrategista chinês Zhuge Liang descreve suas estratégias de batalha em um documento composto de duas partes, com 46 artigos sobre as qualidades de um general e 16 sobre estratégias de guerra (PHENG; KEONG, 1999). Estes artigos formam um guia de gestão estratégica aplicável tanto no meio militar quanto na administração. Segundo os autores, Zhuge Liang depois se aposentou e fez fortuna nos negócios, praticando as estratégias que ele tinha escrito para uso militar.

Nos anos 1600 o comércio de chá entre china e Grã-Bretanha causava grande desequilíbrio comercial favorável aos chineses. R. L. Wickham foi enviado para a China com o objetivo de reunir informações para tornar a Grã-Bretanha economicamente competitiva neste negócio. Wickham constatou o potencial do chá para enriquecer os cofres britânicos e infiltrou-se na produção chinesa durante 10 anos até conseguir transportar de forma bem sucedida plantas de chá para o Reino Unido. Os britânicos, em seguida, centraram a sua produção de

chá na Índia, e por volta de 1880, o chá indiano torna-se o mais popular suplantando o chá chinês (TOUSSAINT-SAMAT; BELL, 1994).

Ainda na Grã-Bretanha, aproximadamente no final dos anos 1700, Nathan Rothschild empregou uma matriz de inteligência crítica desenvolvida em conjunto com seus irmãos. Esta, além de ajudar o governo britânico a derrotar Napoleão, trouxe fortuna, tanto para a sua família quanto para o Império Britânico. O uso astuto da inteligência o ajudou a alcançar seus muitos sucessos: ele salvou a bolsa de valores de Londres do colapso, e confundiu seus rivais financeiros, construindo a mais antiga instabilidade bancária do mundo (FULD, 2002).

À Inteligência Competitiva, costumeiramente, atribui-se o uso de alguns conceitos provenientes da área de Marketing. Em 1879, nos EUA são registradas as primeiras manifestações na área de marketing, e a conseqüente evolução da IC derivada de objetivos competitivos, elementos e práticas das pesquisas desenvolvidas nessa área (JUHARI; STEPHENS, 2006).

Contudo, a aplicação do marketing como uma função organizada remonta a meados dos anos 1920, quando Charles Cooling Parlin foi reconhecido como o primeiro profissional a ter o título de Diretor de Marketing nomeado pela *Curtis Publishing Company* em 1926 (KELLEY, 1968).

Outra corrente de destaque a respeito das atividades de inteligência diz respeito à segurança nacional como uma questão política (BERKOWITZ; GOODMAN, 1991; WARNER, 2002; FULD, 2010). Este fluxo, especialmente nos EUA, tem suas origens no período da Segunda Guerra Mundial (ANICA-POPA; CUCUI 2009). E a gênese de suas atividades na inteligência militar usada nos campos de batalha da Guerra Fria (DUBEY, 2013).

Já na década de 1980 temos a análise dos competidores por parte da indústria (PRESCOTT, 1999). O foco aqui coloca a organização empresarial no centro do palco (EELLS; NEHEMKIS, 1984). Além disso, nessa década a IC foi impulsionada pelo reconhecimento de que informação de qualidade tem impacto direto sobre os aspectos mais cruciais do negócio (JUHARI; STEPHENS, 2006).

Outro fato de destaque no desenvolvimento da IC foi o fim da Guerra Fria, onde o capitalismo e a globalização fizeram com que os governos procurassem se impor economicamente e não mais militarmente. Os diretores de indústrias e serviços é quem tomam as principais decisões que afetarão o futuro dos cidadãos por meio dos produtos e empregos que elas oferecem (COLBY, 1993).

O advento e uso prolífico do computador impulsionaram a tecnologia e as capacidades analíticas de IC (PRESCOTT; GIBBONS, 1993). Com mais dados podem ser armazenados e gerados, a IC torna-se mais eficiente e seu custo mais aceitável para a alta administração.

Por fim, um grupo de profissionais compreendeu as limitações e problemas do campo da IC. Como resultado, em 1986, a Sociedade dos Profissionais de Inteligência Competitiva (SCIP) foi estabelecida. Além de superar as limitações, então percebidas da IC, a sociedade também abordou as implicações mais amplas e as questões de IC. Em 1989, a SCIP contava com 600 membros e, em 1999, esse número era dez vezes maior com 6.500 membros. Em fevereiro de 2002, uma média de 9.626 usuários por semana estava sendo cadastrados no site da SCIP (SCIP, 2013).

2.2 CONCEITOS

Muitos são os conceitos e definições a respeito da Inteligência Competitiva encontrados na literatura. Para Fleisher e Bensoussan (2007) atualmente existem muitas definições de IC na prática contemporânea e provavelmente nenhuma definição de IC seja precisa e universalmente aceita. Para os autores a IC costuma ser vista como o processo onde organizações buscam informações sobre concorrentes e seu ambiente competitivo para aplicar em seu processo de tomada de decisão.

Assim, a IC pode compreender um simples monitoramento de resultados anuais de uma companhia ou outros documentos públicos, ou tão elaborada quanto à realização do exercício de um jogo de guerra que ocorre no meio digital (SHAKER; GEMBICKI, 1999).

Portanto, na visão de Gilad (2011), ainda que existam diversas tentativas de dar significado à Inteligência Competitiva, o conceito pode ser bastante vago.

Para Fleisher, Wright e Tindale (2007) o campo de IC e a sua gestão sofrem de variedade semântica e da presença da ambiguidade que permanece sem solução após várias décadas de trabalho de investigação acerca do tema.

Assim, sejam definições formais sobre o tema ou sobre aspectos pontuais estas, de maneira geral, destacam que a IC deve apoiar a tomada de decisão. O Quadro 7 apresenta alguns destes conceitos.

Quadro 7 - Conceitos e Definições a respeito da IC.

<i>Conceitos</i>	<i>Autores</i>
A IC deve detectar ameaças competitivas, eliminar ou minimizar surpresas, acrescentar vantagem competitiva minimizando o tempo de reação, e encontrar novas oportunidades.	Combs e Moorhead (1992)
Análise de informação sobre os competidores que tem implicação no processo de tomada de decisão da empresa.	Fuld (1995)
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)
Aponta a IC como uma disciplina formal que consiste de especialistas em inteligência, investigadores de empresa, alimentadores de dados secundários, companhias benchmarking e organizações de pesquisa de mercado.	West (1999)
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)
É um processo sistemático, que visa a descobrir os atores e as forças que regem os negócios, reduzir riscos e conduzir o tomador de decisão a agir antecipadamente, bem como proteger o conhecimento gerado.	Marcial e Grumbach (2008)
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)
Um processo informacional proativo que conduz à melhor tomada de decisão, seja ela estratégica ou operacional. É um processo sistemático que visa descobrir as forças que regem os negócios, reduzir o risco e conduzir o tomador de decisão a agir antecipadamente, bem como proteger o conhecimento gerado.	Abraic (2013)

2.3 DESMISTIFICAÇÕES A RESPEITO DA IC

A Inteligência Competitiva tem sua origem na inteligência governamental. Esta basicamente tinha como foco a busca de informações que seriam aplicadas à defesa nacional. É possível situar cronologicamente a expansão de suas atividades no período entre as duas grandes Guerras Mundiais. As nações perceberam que a informação era uma importante arma na obtenção de informação em cenários de conflito (DAL PIZZOL; TODESCO, 2012).

Prescott e Miller (2001) reforçam esta tese argumentando que a IC herdou práticas e o conhecimento das agências governamentais e grupos militares de inteligência.

As técnicas militares de busca de informações foram adaptadas à realidade empresarial por profissionais de áreas de conhecimento como: (1) administração, marketing, estratégia e gestão; (2) Ciência da Informação, gerenciamento de informações formais; (3) pela Tecnologia da Informação, em ferramentas de gerenciamento de informação e mineração de dados.

Porém, mesmo adaptada e consolidada no meio empresarial, ainda pairam alguns mitos não derrubados sobre a IC. É objeto de discussão o caráter legal das informações levantadas sobre concorrentes durante o processo de Inteligência Competitiva, inclusive com acusações de serem resultados de espionagem. Porém, é princípio básico da IC não violar as normas legais do país onde atua. Espionagem é a busca ou acesso não autorizado a dados, informações e outros conhecimentos de caráter particular caracterizando uma prática ilegal de obtenção de informações (DAL PIZZOL; TODESCO, 2012). No entanto, a maioria da informação necessária a um profissional de IC pode ser encontrada de maneira legal em documentos públicos ou por meio de entrevistas.

Segundo, Dr. Wayne Rosenkrans, ex-presidente da SCIP, 90% das informações sobre seus concorrentes estão disponíveis mediante investigação ética, e os demais 10% podem ser deduzidos por meio de uma boa análise dos dados disponíveis (SCIP, 2013).

Outra área de conhecimento que costuma ser confundida com a Inteligência Competitiva é a Gestão do Conhecimento (GC). A Gestão do Conhecimento teve origem na década de 1980 e tem como principal objetivo o gerenciamento do conhecimento interno acumulado na organização, seja ele proveniente de funcionários ou de processos internos, com o objetivo de torná-lo um ativo da empresa (SHARIQ, 1997; DALKIR, 2013; KIMBLE, 2013).

A GC possibilita que o conhecimento seja documentado, difundido e disseminado dentro da organização o que transforma este conhecimento tácito em explícito. Por sua vez, a IC volta suas atenções principalmente para o conhecimento que é obtido no ambiente externo da empresa (KAHANER, 1996; CALOF; WRIGHT, 2008; NASRI, 2011).

Assim, a implantação de projetos de IC e GC podem acontecer paralelamente completando-se no que se refere a obtenção de informação. Gilad e Gilad (1988) defendem a importância da Inteligência Competitiva perante a gestão do conhecimento da seguinte forma.

“(...) a gestão do conhecimento sempre foi ambígua, reinventando a roda em muitos casos e complicando rodas existentes em outros. Ela tornou-se um título vazio, reduzida a armazenamento de dados e algumas iniciativas de conectividade de intranet. Colocar IC e GC no mesmo lugar é besteira”.

Para Bose (2008) a diferença entre BI e IC é que a primeira se concentra na inteligência interna sobre a própria organização, já a IC é focada na inteligência externa sobre os concorrentes.

Gomes e Braga (2001) afirmam que a diferença é que a IC lida com informações públicas sobre competidores que é usada em decisões estratégicas para que a empresa tenha vantagem competitiva. Já *Business Intelligence* engloba informações que não são necessariamente usadas para análises competitivas da empresa. Segundo os autores, o termo tem sido usado para caracterizar ferramentas tecnológicas que apoiam os Sistemas de IC.

Por fim, a ABRAIC não deixa dúvidas ao afirmar as diferenças entre IC e *Business Intelligence* (BI), este, outro termo comumente confundido com IC.

“Não. Apesar de encontramos algumas citações que utilizam os termos como sinônimos, na maioria das vezes o primeiro está sendo empregado pelas empresas de tecnologia da informação como conjunto de ferramentas utilizadas para auxiliar nos negócios tais como: *data warehouse*, *data mining*, CRM, ferramentas OLAP e outras. Quando falamos de Inteligência Competitiva, estamos nos referindo a um processo muito maior que engloba a obtenção e tratamento

de informações informais advindas das redes mantidas pelos sistemas de IC, nas quais as informações de BI estão inseridas.” (ABRAIC, 2013).

2.4 O PROCESSO DE INTELIGÊNCIA COMPETITIVA

O processo de IC é uma atividade sistemática e ética que consiste em coletar, analisar, disseminar e gerenciar as informações sobre o ambiente externo que podem afetar os planos, as decisões e as ações da empresa (SCIP, 2013).

Este processo resulta em análises de mercado, identificação dos principais e dos novos competidores, alterações na legislação e no mercado, entre outros. Segundo Kahaner (1996) e Sawka (2006), estes resultados permitem a gerência de mudanças no planejamento estratégico, além de maior eficácia organizacional.

Contudo, para que se alcance os resultados esperados, os produtos de IC resultantes desse processo devem estar alinhados às necessidades específicas de cada decisor. Assim, a equipe de IC precisa compreender as necessidades da organização e a natureza das decisões nos diferentes níveis organizacionais.

Rodriguez (2004), destaca que o processo de IC possibilita entender melhor o ambiente de negócio no qual a organização está inserida, constituindo uma maneira de agregar valor ao processo decisório. Contudo, para que o processo de IC represente valor aos gestores é preciso delimitar corretamente as questões a serem respondidas e as necessidades de informação.

Já Abreu; Cobral, Ogliari (2008), destaca a importância de sistematizar estas atividades, tornando o processo de IC algo contínuo dentro da empresa.

Bensoussan e Fleisher (2008) destacam que são relevantes os esforços desenvolvidos em atividades como: seleção criteriosa da informação, verificação da confiabilidade dessa informação, redução dos produtos finais para que estes sejam sucintos e análise, integração e interpretação desses resultados.

Quanto à estruturação do processo de IC, existem diferenças na quantidade de etapas dentro do processo. Também chamado de Ciclo de Inteligência, devido ao seu aspecto cíclico e incremental, o processo de inteligência costuma ter quatro ou cinco etapas. O Quadro 8 a seguir

desenvolvido por Machado (2010) apresenta o número, as etapas e os autores que o desenvolveram.

Quadro 8 - Etapas de estruturação do processo de IC.

<i>Processo de IC estruturado em</i>	<i>Autores</i>
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, atividades de coleta, análises, disseminação, avaliação)	(MILLER, 2009)
5 etapas (planejamento e direção, processamento e armazenamento da informação, coleta, análise e produção, disseminação)	Herring (1998) 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 o Quadro 8 acima, vemos que as etapas de coleta, análise e disseminação são comuns na definição do processo de IC para todos os autores. 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. Esta comumente valida os resultados da aplicação dos produtos resultantes do processo de IC, e se necessário, traz o processo de volta para o início, começando um novo ciclo.

2.4.1 Identificação das Necessidades de Informação

Segundo Herring (1999), o fator crítico de sucesso em qualquer operação de inteligência é satisfazer as necessidades reais dos usuários chave e da alta gestão, de tal forma que a organização passe a atuar com base na inteligência resultante.

Herring (1999), propõe dois modos para atingir este objetivo. O primeiro é o responsivo. Nele o usuário deve tomar a iniciativa e

procurar os gerentes e tomadores de decisão para levantar suas necessidades de informação. Assim, a equipe de inteligência deve estar preparada para gerenciar uma grande quantidade de pedidos dos usuários. Com a grande demanda, algumas solicitações devem ser rejeitadas. Dessa forma, a eficiência do processo está em saber atender as solicitações certas para produzir resultados melhores.

Já o segundo é o modo proativo, onde a própria área de inteligência busca identificar quais as necessidades de informação do usuário a partir do que está disponível. Este modo é chamado por Herring (1999) de *Key Intelligence Topics* (KITs).

Herring (1999), sugere que nos KITs cada solicitação seja tratada como um projeto onde as iniciativas são definidas com base em reuniões com os principais tomadores de decisão da empresa. Dessas reuniões surgem as orientações para as demais etapas do processo como a coleta e a análise dos dados que resultarão nos produtos de inteligência. O autor classifica os KITs em três categorias funcionais:

1. decisões e ações estratégicas, que incluem o desenvolvimento de planos estratégicos.
2. tópicos de alerta antecipado, que compreendem iniciativas dos concorrentes, ameaças e monitoramento do mercado.
3. KITs para atores principais que descreve os principais atores envolvidos como: concorrentes, clientes, fornecedores, órgãos reguladores e parceiros potenciais.

Na prática, o que mais se observa é a própria área de IC tomando a iniciativa de identificar as necessidades informacionais dos usuários. Alguns dos motivos para isso são a falta de iniciativa dos usuários na busca de informações necessárias e a falta de incentivo por parte da organização no intercâmbio de informações.

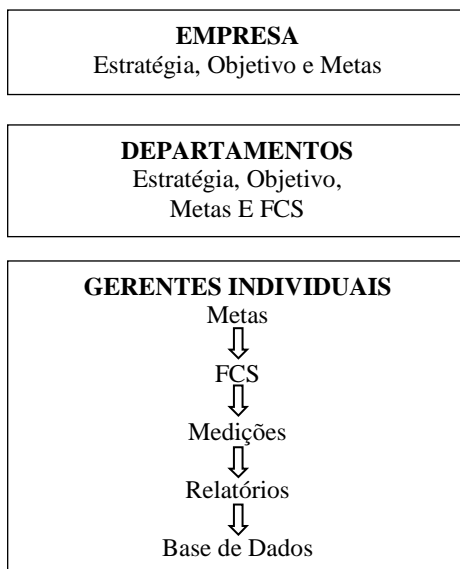
Outro método empregado na identificação de necessidades de informação é o Método dos Fatores Críticos de Sucesso (FCS). Desenvolvido por Rockart (1979), consiste em um método empírico baseado em entrevistas direcionadas a alta administração. Os executivos descrevem o negócio sob seu ponto de vista, as perspectivas de mercado, estratégias e os fatores críticos de sucesso associados. A partir daí, é formatada uma matriz, e os resultados são confrontados para verificação, combinando, eliminando ou incorporando novos fatores críticos até se chegar um consenso.

Os FCS, de acordo com Rockart (1979), sustentam a realização das metas organizacionais, principalmente nas gerências em que um

bom desempenho é necessário para a realização de tais metas, assegurando uma vantagem competitiva de sucesso.

Segundo Rockart e Bullen (1981), o conceito de FCS é usado principalmente para auxiliar no planejamento de sistemas de informação, conforme ilustrado na Figura 1.

Figura 1- Processo de Utilização do FCS na determinação de informações.



Fonte: Bullen e Rockard (1981).

Analisando a Figura 1 vemos que o uso dos fatores críticos concentra-se nos departamentos e gerentes individuais. Assim, os gerentes, seja qual for seu nível hierárquico dentro da organização, devem ter informações apropriadas que permitam determinar se os acontecimentos estão sendo conduzidos com eficiência em áreas consideradas críticas para o sucesso da empresa.

Abreu, Cobral e Ogliari (2008), sugerem o uso do mapa estratégico para identificação de necessidades de informação. Para definir o mapa estratégico de informação os autores recomendam que as necessidades sejam identificadas e organizadas em dois níveis: estratégico e tático-operacional. Em seguida, devem ser atreladas a objetivos estratégicos, fatores críticos de sucesso, oportunidades ou ameaças. Ao fim deste processo temos um mapa que constitui um documento de referência para o processo de IC.

Por fim, cabe ressaltar que independente do instrumento escolhido para identificar as necessidades de informação é imprescindível avaliar se as necessidades de informações estão em conformidade com os produtos que os gestores esperam receber ao final do processo.

2.4.2 Planejamento

A etapa de planejamento é definida por muitos autores como a fase mais importante do ciclo de IC (HERRING et al., 2002; MILLER, 2002; STAREC, 2005; OLIVEIRA; MELO, 2012). Do seu desenvolvimento resulta a missão do projeto de IC, principais objetivos, estratégias e ações que serão tomadas. Definem-se também quais os recursos humanos, tecnológicos, financeiro e o tempo que a empresa dispõe para ser utilizado (MARCEAU; SAWKA, 1999). Esta fase deve levar em conta as metas já traçadas pelo planejamento estratégico da empresa (PRESCOTT, 1999; LANA, 2011).

Na fase de planejamento deve-se definir a missão da IC na organização e as expectativas que o projeto deve cumprir. Conforme Prescott e Miller (2002), esta missão pode ser de três tipos:

1. **Informativa:** desenvolve compreensão geral sobre indústria e competidores;
2. **Ofensiva:** identifica vulnerabilidades dos competidores ou determina o impacto que estratégias adotadas teriam sobre os competidores;
3. **Defensiva:** estabelece as ameaças que o competidor pode oferecer e que poderiam colocar em risco a posição da empresa no mercado.

Ao fim desta fase é possível definir de forma concreta as expectativas do projeto, suas metas, delimitações, abrangência e o ambiente competitivo que será analisado (BOSE, 2008).

Por último, devem ser apontados os caminhos e ações para as demais etapas do ciclo levando em conta também fatores como segurança e planos de contingência (WEISS, 2002; BOSE, 2008).

2.4.3 Coleta

A coleta é caracterizada pela busca de dados e informações necessárias para que a etapa de análise desenvolva 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 no Quadro 9. 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”.

Quadro 9 - Classificação das Informações.

<i>CLASSIFICAÇÃO</i>	<i>TIPO</i>	<i>DESCRIÇÃO</i>	<i>EXEMPLOS</i>
Quanto a 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	Balço 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. No primeiro, denominado primitivo, a coleta é feita sem um objetivo específico. No segundo, *ad hoc*, a coleta não é ativa e costuma ocorrer em resposta a uma necessidade. Já o terceiro, reativo, ocorre somente quando é diagnosticada uma ameaça competitiva concreta a organização. Por fim, o quarto, proativo, caracteriza-se por uma busca ativa de informações. Esta busca tem por característica procurar todo tipo de informações que possam identificar ameaças ou oportunidades para a organização.

Independente de como é definido o processo de coleta, a identificação de boas fontes de informação é essencial. Selecionar boas fontes pode evitar que a organização colete uma grande quantidade de informação sem valor, o que acaba dificultando as tarefas de tratamento e estruturação da informação recuperada.

2.4.3.1 Técnicas de Coleta

Para garantir sua efetiva execução o processo de coleta deve ser rigorosamente planejado. Isso ajuda a definir a estratégia que será usada para identificar as necessidades de informação e os tipos de fontes que contemplam essas necessidades (KRIZAN; COLLEGE, 1999).

Para apoiar o processo de coleta estão disponíveis ferramentas usadas para reunir informações de diferentes fontes na *Web* com base em termos previamente cadastradas. As informações são compiladas em relatórios concisos sobre concorrentes, produtos, serviços, pessoas chave, estratégias, mercados, principais clientes, fornecedores e órgãos reguladores (BOSE, 2008).

2.4.3.2 Classificação das ferramentas de coleta

Programas de inteligência ou coleta de dados necessitam ferramentas que possibilitem o levantamento de domínios de conhecimento e que sejam orientadas a questões específicas, coleta ativa, ou que apoiem às necessidades informacionais que estão em curso, coleta passiva (RAO, 2003; FAN; GORDON; PATHAK, 2006).

Ferramentas de coleta ativa apoiam a coleta pelo desenvolvimento de pesquisas baseadas em terminologias e categorização dos resultados. Exemplo são os motores de busca da *Web* que possibilitam pesquisas e navegação online. Contudo, a estruturação das consultas e a categorização das taxonomias podem ser problemáticas devido à "interação homem-computador", ou a "problemas de vocabulário" que geram dificuldades em usar as palavras certas para interagir com os sistemas computacionais (FAN et al., 2006).

Por fim, instrumentos de coleta ativos não são limitados a tipos específicos de decisões, mas quanto a sua precisão e capacidade de interpretar e apresentar resultados relevantes (BOSE, 2008).

Outro problema dos motores de busca é o fato da informação retornada por eles não ser estruturada, já que estes simplesmente varrem a *Web* em busca de termos. O conteúdo retornado é constituído basicamente por textos simples onde termos relacionados podem não ser retornados. Além disso, se as fontes pré-definidas não forem confiáveis podem retornar informações não verdadeiras.

Outra forma de extração ativa é a que utiliza "*named entities*", que são termos Relevantes que correspondem normalmente a substantivos próprios, onde são feitas conexões entre entidades que podem ser pessoas, organizações, produtos ou lugares (RAO, 2003). São

úteis para avaliar rapidamente resultados de consultas, e ao serem utilizadas em conjunto com tecnologias de visualização de informação, padrões e relações são rapidamente identificados.

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).

Fan et al. (2006) desenvolveram um modelo inteligente de dois estágios para o encaminhamento de informações que incorpora mineração de texto, computação genética e estatística. O modelo foi projetado para superar os problemas associados aos agentes de *software*, abordando tanto a terminologia de consulta quanto os processos de correspondência. Para tanto, realiza consultas de treinamento com base em um conjunto relevante de documentos, que ao final gera uma consulta persistente.

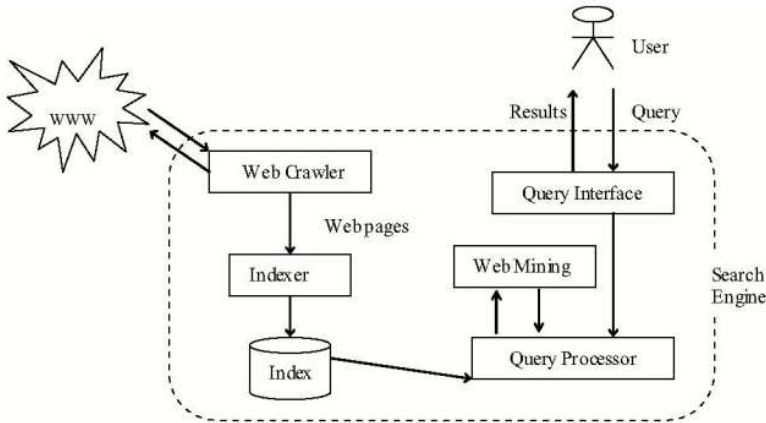
2.4.3.3 Motores de busca

Um motor de busca é um sistema computacional desenvolvido para encontrar o maior número possível de informações armazenadas principalmente na *Web*, mas também em redes corporativas ou em computadores pessoais. A busca é feita a partir de palavras-chave indicadas pelo usuário, ou às vezes, em linguagem natural, reduzindo o tempo necessário para encontrar informações.

O método mais comum para obtenção de informações da *Web* é o uso de motores de busca. Estes processam consultas de um usuário, geralmente expressões constituídas por palavras-chave, para retornar um conjunto de páginas *Web* ou documentos que satisfazem a consulta em algum grau (BONCELLA, 2003).

Um motor de busca da *Web* geralmente é composto pelos seguintes componentes: *Web Crawlers* ou *Spiders*, método de indexação, recuperação e classificação e interface de usuário (BONCELLA, 2003; GORE; PITALE, 2013). A Figura 2 mostra a arquitetura de um motor de busca.

Figura 2 – Exemplo de um motor de busca.



Fonte: (Duhan; Sharma; Bhatia, 2009).

Na Figura 2 é possível identificar componentes importantes da arquitetura de um motor de busca como os *crawlers*, indexadores, e os mecanismos de classificação. Os *crawler* também conhecidos como *Spiders* percorrem a *Web* realizando o *download* de páginas. Essas páginas são enviadas para um módulo de indexação que as analisa e cria um índice com base em palavras-chave contidas nessas páginas. Quando um usuário digita uma consulta usando palavras-chave no motor de busca, o componente que processa as solicitações de consulta irá comparar essas palavras-chave com o índice e retornar as *Uniform Resource Locator* URLs das páginas para o usuário. Antes de apresentar o resultado para o usuário, os motores de busca realizam uma classificação que apresenta as páginas mais relevantes no topo. Isso torna os resultados da busca mais fáceis para o usuário (SINGH; KUMAR, 2009).

a) **Crawlers**

Os *Crawlers* são as peças centrais dos motores de busca. Eles rastreiam continuamente a *Web* para identificar páginas adicionadas ou removidas. Contudo, devido ao crescimento e natureza dinâmica da *Web*, é impossível varrer e lidar com todas as URLs dos documentos.

Uma solução parcial para esse problema é o uso de *crawlers* focados em um determinado tema para visitar e reunir apenas páginas com informação relevante. Esta abordagem recupera documentos que

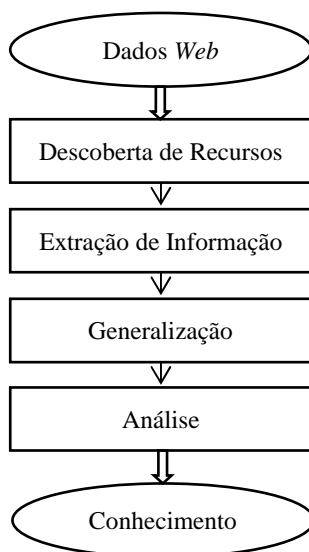
contenham determinada palavra-chave informada na consulta do usuário contudo, a informação continua não sendo estruturada.

2.4.3.4 Web mining

Web mining utiliza técnicas de mineração de dados para descobrir e extrair informações de documentos da *Web* (GORE E PITALE, 2013). Ferramentas para mineração na *Web* permitem aos usuários consultar e combinar os dados com base no seu conteúdo semântico. Aplicações *Web BI* ou *Business Intelligence*, são um tipo emergente de software de apoio à decisão que aproveita o conteúdo da *Web* para extrair conhecimento em um ambiente organizacional (HSU, 2007).

De acordo com Kosala e Blockeel (2000), a *Web Mining* consiste das seguintes tarefas: 1) Descoberta de recursos: é a tarefa de recuperação dos documentos pretendidos. 2) Extração de Informação: tarefa de seleção automática e pré-processamento de informações específicas de recursos recuperados. 3) Generalização: consiste em descobrir padrões gerais tanto em sites individuais como em vários sites. 4) Análise: validação e interpretação dos padrões extraídos.

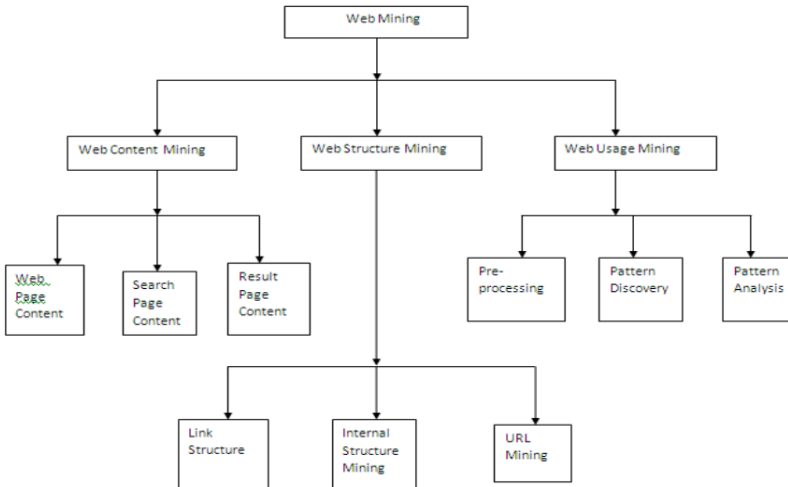
Figura 3- Processo de *Web Mining*.



Fonte: Adaptado de (KOSALA E BLOCKEEL, 2000).

A *Web mining* pode explorar tanto o conteúdo da *Web* quanto sua forma de uso (BOSE, 2008). Existem três áreas de mineração *Web* que são classificadas de acordo com o tipo dos dados utilizados como entrada no processo de mineração. *Web Content Mining* (WCM), baseada em conteúdo, *Web Structure Mining* (WSM), na estrutura e *Web Usage Mining* (WUM), baseada na forma de uso (COOLEY; MOBASHER; SRIVASTAVA, 1997). **A Erro! Fonte de referência não encontrada.** apresenta uma classificação geral da *Web Mining*.

Figura 4 - Classificação da *Web Mining*.



Fonte: (KOSALA; BLOCKEEL, 2000).

As três categorias se concentram no processo de descoberta de conhecimento e informações potencialmente úteis a partir da *Web*. Mesmo que estas sejam três áreas distintas de mineração na *Web*, as diferenças entre elas estão diminuindo devido a interconexão.

2.4.3.5 Recuperação de informação baseada em ontologia

A grande desvantagem da recuperação de informação manual é o tempo gasto nessa tarefa. Dado o tamanho da *World Wide Web*, e sua taxa de crescimento, catalogar, mesmo que uma modesta porcentagem de páginas da *Web* é uma tarefa extremamente complicada.

Robôs *ad-hoc* que tentam reunir informações na *Web* normalmente recuperam apenas a informação semântica que se pode inferir das marcações HTML (*HyperText Markup Language*) existentes.

O processamento de linguagem torna difícil inferir significado semântico a partir do próprio corpo do texto. Além disso, essas abordagens não permitem inferências sobre relações entre páginas *Web* e suas ligações (LUKE et al., 1997).

Para agregar mais informações, é possível incorporar informação diretamente em páginas HTML, tornando simples para usuários e robôs recuperar e armazenar esse conhecimento. Maneira para tanto é fornecer aos autores um conjunto limpo de HTML que adiciona uma sintaxe de marcação, isto é, que lhes permita classificar diretamente suas páginas *Web*, detalhes de relacionamentos e atributos em formato legível por máquina usando HTML.

Esta abordagem baseia-se em ontologias para troca e pesquisa de informações, anotações semânticas para representação de conteúdo de documentos e linguagens formais de conhecimento para ontologias e anotações.

Linguagens de representação de conhecimento que suportem inferências lógicas podem nos ajudar a recuperar conhecimento de forma mais flexível e precisa. A primeira exigência é que a linguagem de metadados seja intuitiva e concisa o suficiente para as pessoas possam utiliza-la facilmente depois de um período de formação de curta duração (MARTIN; EKLUND, 2000).

As linguagens atuais de conhecimento orientado a metadados são construídas com base em *eXtensible Markup Language XML*, como o *Resource Description Framework (RDF)* e a *Ontology Markup Language (OML)*. A escolha do XML como formato subjacente permite que sejam usadas ferramentas XML padrão para trocar e analisar linguagens de metadados. No entanto, como o XML é prolixo, as linguagens de metadados construídos acima dela são extensas e difíceis de usar sem editores especializados.

2.4.4 Análise

Referida como o “coração” ou o “cérebro” do processo de IC, a análise dos dados é a etapa que de fato abriga o processo de transformação de dados e informação em inteligência, pelo que é reconhecida como a fase que mais agrega valor ao ciclo de IC (CORY, 1996; BERGERON; HILLER, 2002; JIN; BOUTHILLIER, 2006; BOSE, 2008; CORREIA; SANTOS, 2010).

Esta etapa geralmente é considerada a mais difícil dentro do processo (KAHANER, 1996; KRIZAN; COLLEGE, 1999; BENSOUSSAN; FLEISHER, 2003; SAWKA, 2006; ABREU;

COBRAL; OGLIARI, 2008). A análise requer habilidades lógicas do analista para avaliar as informações, identificar padrões e visualizar cenários. Contudo, nem sempre todas as informações necessárias estão disponíveis. Nesse caso é preciso trabalhar com suposições para preencher as lacunas (KAHANER, 1996).

Sawka (2006) defende que a análise não é uma função que recebe um pedido e retorna uma resposta. É preciso que usuários de inteligência, e analistas atuem de forma conjunta dentro do processo e direcionem um plano de ação.

O processo de análise é uma tarefa complexa que deve ser dividido em três fases para maximizar suas chances de sucesso. Para Bouthillier e Shearer (2003), é na fase de análise que as informações coletadas ganham significado, sendo necessário classificar, organizar e armazenar as mesmas.

A avaliação é a primeira fase, e diz respeito ao grau de confiabilidade da fonte de dados e da veracidade da informação. Em processos de IC que utilizam motores de busca na *Web* tradicional, esta pode ser uma fase bastante trabalhosa já que a confiabilidade da informação não é garantida. Isto demanda que cada resultado seja avaliado individualmente para confirmar sua confiabilidade.

Catálogoção, a segunda fase compreende o processo de ordenação, catalogação, arquivamento e indexação dos dados levantados. Em dados pouco estruturados, isto demanda a limpeza, a 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.

Para o desenvolvimento da análise, as organizações se valem de técnicas variadas para chegar ao seu objetivo. Algumas destas técnicas são: elaboração de cenários; análise de portfólio; análise do ambiente competitivo (Forças de Porter); análise SWOT; *benchmarking*; e método DELPHI (TARAPANOFF; GREGOLIN, 2002).

Além dessas são passíveis de nota técnicas matemáticas e estatísticas para data mining. São encontradas listas similares das técnicas de análise em (CARDOSO JR., 2003; GOMES; BRAGA, 2004; BOSE, 2008).

2.4.4.1 Produtos de IC

Resultado da fase de análise, os produtos de inteligência é o que propriamente agrega valor para a organização por meio de sua aplicação

dentro de um processo de IC. Os produtos de Inteligência Competitiva visam o suporte às decisões dos níveis estratégico e tático de uma empresa (GOMES; BRAGA, 2004).

Os produtos de IC podem ser trabalhados a curto, médio e longo prazo dentro de uma organização, sua entrega deve ser direcionada para públicos distintos, e modelados usando diferentes modelos analíticos segundo as necessidades de cada público.

Para tanto é importante que a área de IC participe das discussões e com isso fazer chegar aos tomadores de decisão produtos de inteligência mais eficientes. Os profissionais de IC devem interagir com os responsáveis pelo processo decisório, trocando informações sobre as necessidades de inteligência e os produtos de inteligência a serem disponibilizados em tempo oportuno (OLIVEIRA; LACERDA, 2007).

Cabe também a equipe saber a quem essa inteligência é necessária ou relevante, e que tipos de decisão poderão ser tomados com base nela. Além disso, definir a frequência de disseminação é de extrema importância. É importante que os produtos entregues sejam focados e não generalizados, e suas conclusões devem ser priorizadas, permitindo que o leitor localize os pontos relevantes rapidamente.

Existem diversas classificações dos produtos de inteligência gerados. Estas levam em conta características como o tipo de conteúdo, o público alvo, a periodicidade e sua finalidade. A seguir é elaborado um quadro com as características dos principais produtos de inteligência.

Quadro 10 - Produtos de Inteligência.

Autor	Produtos de Inteligência	Características
Dugal (1998)	Inteligência Corrente	Informação é coletada e disseminadas diretamente ao usuário com pouco ou nenhuma análise.
	Inteligência para Negócios	Resultado cotidiano dos analistas de IC. Este produto de inteligência é bem pesquisado, completamente analisado, sistematicamente documentado e voltado para a decisão. Ex. estudo e o perfil do concorrente.
	Inteligência Técnica	Relacionados aos concorrentes incluindo identificação, inovações científicas, tecnológicas e tendências. Predomina o monitoramento de patentes dos concorrentes, localizando nos produtos e tecnologias.
	Inteligência de Advertência	Prever ameaças emergentes ou oportunidades, entregue em forma de alertas de inteligência. Os analistas podem gerar uma lista de indicadores potenciais e monitorar a informação antecipadamente.
	Inteligência Estimativa	Oferece cenários prováveis para clientes internos. Reflete o conhecimento dos analistas com exposição da condição da concorrência durante vários anos. Baseada em análise quantitativa de dados, com visões qualitativas dos analistas.
	Inteligência de Trabalho em Grupo	Oferecida por analistas que trabalham no ambiente interno sendo parte de trabalho em grupos.
	Inteligência Direcionada	Exigências estreitas, específicas de clientes internos. É gerada esporadicamente com base em pedidos fortuitos recebidos pelo sistema de inteligência. Frequentemente, assume a forma de investigações de seguimento.
	Inteligência de Crise	Atividades de IC que ajuda a aliviar ou negar o efeito de uma crise e cumpre as exigências dos responsáveis pela crise.
	Inteligência Internacional	Foca governos, indústrias e concorrentes estrangeiros em termos políticos, industriais, oportunidades e confiança.
Tyson (1998)	Newsletter	Alertas sobre o ambiente, muitas vezes sem análise. Tem como público alvo a área de marketing.
	Minutas de Impacto	Apresenta eventos que podem resultar em impacto estratégico ou tácito. Seu público alvo são as gerências ligadas a área comercial.
	Análises de Situação	Consolidação de informações que resumem assuntos estratégicos e que são chaves para a empresa em um formato mais detalhado. É destinada a alta gerência.
	Evolução do Competidor	Monitora a evolução dos competidores quanto a inovação, mudanças estratégicas e implicações no mercado de atuação.
	Análise de Produto ou Projeto	Atendem demandas sobre a concorrência, novidades de mercado, e ameaças específicas na elaboração de produtos ou projetos.

Quadro 11 - Produtos de Inteligência (continuação).

Autor	Produtos de Inteligência	Características
Stolienwerk (1999)	Relatórios especiais de inteligência	Relatórios de uma a duas páginas sobre uma questão estratégica, resumizando as análises de apoio à decisão e incluindo recomendações para ação.
	Análises situacionais	Sintetizam os aspectos estratégicos de uma determinada situação de impacto para a empresa ou de decisão, incluindo análises detalhadas que suportam os relatórios especiais de inteligência.
	Relatórios mensais de inteligência	Destacam itens estratégicos e notícias relacionadas a esses itens. Incluem sumários executivos de artigos e notas de entrevistas apresentadas em tópicos para os decisores.
	Fichas de impacto estratégico	Semelhantes aos boletins mensais de notícias, porém com uma primeira indicação de sinais de impacto estratégico ou tático do evento sobre a empresa.
	Perfis dos concorrentes	Contêm informação geral sobre concorrentes. Requerem atualização permanente.
	Boletins mensais de notícias	Contêm informações relevantes de fontes internas e externas sobre o ramo de negócio e as forças do macro ambiente que influenciam a atuação das empresas do ramo.
Bernhardt (2003)	Corrente	Avaliam situações específicas, suas consequências a curto-prazo, sinalizando situações perigosas no futuro próximo.
	Estimada	Previsões de caráter estratégico das principais tendências e as implicações para a empresa.
	Alerta	Alertas encaminhados aos decisores classificados de acordo com o nível de prioridade em periódicos e de alerta.
	Pesquisa	Estudos focados e com certo grau de profundidade, geralmente confidenciais. Podem ser de dois tipos, pesquisa básica e suporte operacional.
Dal Pizzol e Todesco (2012)	Científico e tecnológico	Refere a melhorias que o competidor possa oferecer, e que possam impactar diretamente na empresa.
	Inteligência Básica	Monitoração diária de competidores e aspectos chave da indústria. Produtos adaptáveis a qualquer cliente interno e altamente sensíveis e confidenciais.
	Inteligência Antecipada	Antecipa oportunidades, tendências e ameaças, mapeia indicadores e possibilita a construção de cenários.
	Inteligência de Crise	Ajuda a superar crises, composta de grupos formados exclusivamente para esse fim.
	Contra Inteligência	Protegem as atividades de IC contra os concorrentes para que estas fluam somente no seu ambiente.

2.4.5 Disseminação

Nesta fase é feita a disseminação da informação coletada no ambiente da organização. Esta distribuição de conhecimento pode ser feita de maneira simples ou sofisticada, mas em ambas focando em quatro etapas básicas: (1) identificar os personagens alvos da disseminação do conhecimento; (2) o que será disponibilizado para cada alvo; (3) como será feita a disponibilização; (4) o momento que acontecerá.

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.

Costa (2002), expõe a importância da etapa de disseminação ao dizer que se esta não for realizada de forma eficiente, o processo se torna ineficaz. O autor compartilha a ideia que a entrega do produto ao cliente traz o ciclo de volta ao seu início, pois qualquer ação executada como resultado da inteligência vai levar o usuário a novas demandas e necessidades. Essa abordagem cíclica do processo de inteligência é defendida por outros autores e é comumente vista na prática dentro das empresas.

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.

Para que se obtenha o melhor proveito da informação analisada, o formato de entrega deve ser claro e possuir credibilidade decorrente das fontes utilizadas. Entre os mecanismos de disseminação mais populares temos: relatórios, apresentações, boletins, mensagens eletrônicas, clipping, recomendações, entre outros. A linguagem deve obedecer às necessidades e características de cada usuário. 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 (FULD, 1995; KAHANER, 1996; GOMES; BRAGA, 2004).

2.4.6 Avaliação

As atividades desta etapa acontecem após a entrega dos produtos de IC, e são de extrema importância para o aperfeiçoamento e continuidade do processo de IC.

Segundo Correia e Santos (2010), nesta etapa ocorre a avaliação do processo de IC, a identificação dos seus benefícios e a verificação da

sua eficácia para o processo decisório, bem como o retorno dos utilizadores da inteligência, com vista a elaborar os planos futuros de IC, além de rever e reavaliar a estratégia organizacional.

A avaliação pode ser realizada levando-se em conta duas finalidades. Primeiro, em relação ao processo, refere-se a todas as fases, ou seja, se as necessidades foram bem definidas, se as fontes utilizadas foram suficientes e se os melhores métodos de análise foram utilizados. Segundo, sob o aspecto dos produtos, com relação ao conteúdo, e se os mesmos atenderam a necessidade de informação e foram utilizados na prática (GOMES; BRAGA, 2004).

Nesta etapa é importante avaliar se o processo de IC está sendo visto como uma função corporativa. Todos na empresa devem sentir-se parte do processo. Os tomadores de decisão, que efetivamente farão uso dos resultados devem trabalhar em conjunto com a equipe de IC (SAWKA, 2006).

Outra importante função da fase de análise é mensurar os resultados do processo de IC. Pela dificuldade em mensurar o ganho na utilização de um 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 inovação foram alcançados.

2.5 CONSIDERAÇÕES SOBRE O CAPÍTULO

Este capítulo abordou o processo de Inteligência Competitiva, de modo especial a etapa de coleta, suas técnicas e as ferramentas usadas para recuperar a informação. Por fim, foram apresentados os produtos de IC resultantes da análise da informação coletada.

Em um processo de IC, até 80 por cento da informação necessária é oriunda de fontes abertas (MILLER, 2004). Assim, é essencial a identificação de boas fontes, seja qual for o tipo de coleta definido.

Contudo, a maioria da informação de domínio público não apresenta estruturação ou mesmo é contextualizada. Isso torna comum que organizações coletem grande quantidade de informação sem nenhum valor, onerando as etapas seguintes do ciclo de IC.

Segundo Prescott e Miller (2002) apenas um terço do tempo destinado ao processo de IC é gasto com análise, tarefa que propriamente agrega valor a organização. Para os autores a maioria do tempo é gasto equivocadamente na coleta de informações.

A *Web* de Dados pode ser vista como uma camada adicional fortemente interligada com os documentos da *Web* tradicional (HEATH;

BIZER, 2011). O uso de fontes de dados como as presentes na *WoD*, onde a informação é estruturada, contextualizada e cobre temas variados, pode facilitar o trabalho de coleta e recuperação de informação. Como consequência, pode equilibrar os esforços de coleta e análise, fazendo com que o ciclo de IC gere mais valor a organização.

3 ENGENHARIA DO CONHECIMENTO E WEB DE DADOS

Este tópico aborda os conceitos relacionados à Engenharia do Conhecimento e à *Web* de Dados ou *Web of Data*. Primeiramente é feito o resgate dos termos Engenharia do Conhecimento (EC), *Web Semântica*, e Ontologias que são a base para a *Web* de Dados. Em seguida, são abordados Dados Abertos, a *Web* de Dados e *Linked Data*.

3.1 ENGENHARIA DO CONHECIMENTO

Inicialmente pesquisas em Inteligência Artificial (IA) eram focadas no desenvolvimento de formalismos, mecanismos de inferência e ferramentas para operacionalizar Sistemas Baseados em Conhecimento (SBCs) (STUDER; BENJAMINS; FENSEL, 1998).

Embora estes estudos tenham apresentado resultados bastante promissores, o uso comercial desta tecnologia na construção de grandes SBCs falhou em muitos casos (STUDER; BENJAMINS; FENSEL, 1998).

Situação semelhante viveu a construção de *software* tradicional na década de 1960 no que ficou conhecida como “Crise do *Software*” resultando na criação da disciplina de Engenharia de *Software* (MAHONEY, 2004).

SBCs mais complexos deixaram clara a necessidade de abordagens mais metodológicas. Assim, o objetivo da Engenharia do Conhecimento é o mesmo da Engenharia de *Software*, transformar em disciplina de engenharia a construção de sistemas que antes eram feitos de forma artesanal (SHAW; GAINES, 1992).

Segundo Hayes-Roth, Waterman e Lenat (1983), a EC é vista como um processo de transferência e transformação de conhecimento. Para os autores, a resolução de problemas a partir de uma fonte de conhecimento é o coração do processo de desenvolvimento de sistemas especialistas.

Construir um SBC significa construir um modelo de computador com capacidades de resolução de problemas comparáveis a um especialista de domínio (CLANCEY, 1989; MORIK, 1990). Um perito pode conscientemente articular algumas partes do seu conhecimento, contudo ele não estará ciente de uma parte significativa desse conhecimento, uma vez que este está intrínseco em suas habilidades. Este conhecimento não é diretamente acessível mas tem que ser construído e estruturado durante a fase de aquisição de conhecimento.

Portanto, em um SBC, a aquisição de conhecimento não é vista como uma simples transferência de conhecimento para posterior representação em uma ferramenta computacional apropriada, mas como um processo de construção do modelo (CLANCEY, 1989; MORIK, 1990).

Clancey (1989) relata a análise de um conjunto de sistemas especialistas de primeira geração desenvolvidos para resolver diferentes tarefas usando diferentes formalismos de representação. O autor descobriu um comportamento comum de resolução de problemas, abstraiu este comportamento comum para um padrão de inferência genérico chamado classificação heurística, que descreve o comportamento de resolução de problemas destes sistemas em um nível abstrato, o chamado nível de conhecimento.

A descrição em nível de conhecimento da resolução de um problema e dos detalhes relacionados à formação do processo de raciocínio resulta na noção do que conhecemos por Método de Solução de Problemas ou *Problem-Solving Method* (PSM).

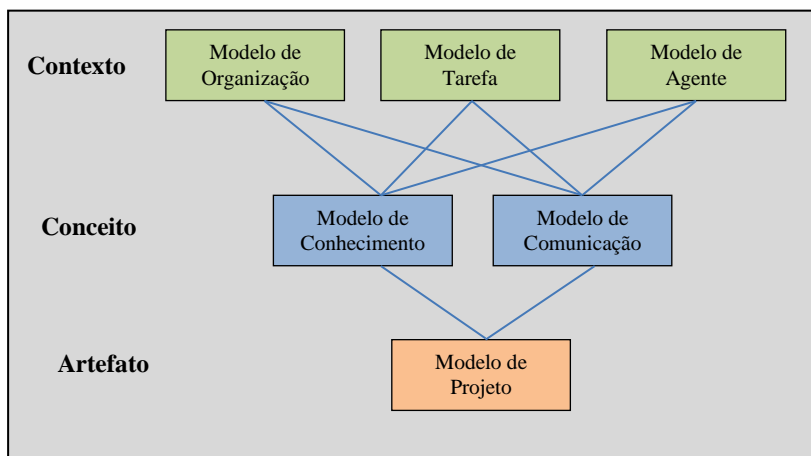
Durante a década de 1980 foram apresentadas algumas abordagens baseadas no Método de Solução de Problemas que tiveram influência significativa no desenvolvimento de técnicas de modelagem da Engenharia do Conhecimento como o Método dos Papéis Limitados ou *Role-Limiting Methods*, apoiado na noção de um método de resolução de problemas reutilizável (Marcus, 1988); e o das Tarefas Genéricas ou *Generic Tasks*, baseado em blocos de construção que podem ser reutilizados para a construção de diferentes SBCs (CHANDRASEKARAN, 1986).

3.1.1 CommonKADS

Uma abordagem de EC de destaque é a KADS (*Knowledge Acquisition and Documentation Structuring*) proposta por Schreiber, Wielinga e Breuker (1993) e seu desenvolvimento para CommonKADS (SCHREIBER et al., 1994). Este framework metodológico nasceu da necessidade de se conceber sistemas de conhecimento com alto grau de qualidade e em grande escala, o que implica em um processo de desenvolvimento estruturado, controlável e repetível (URIONA MALDONADO; COSER, 2010). Comparado ao KADS, o CommonKADS vai ao encontro às necessidades da nova EC, onde é preciso ter uma visão ampla do domínio de aplicação e intimamente ligada às atividades de gestão.

O CommonKADS, é constituído por seis modelos que representam um agrupamento estruturado de conhecimento. Para Marchi (2010), um modelo de conhecimento é uma ferramenta que nos ajuda a esclarecer a estrutura de tarefas de processamento da informação intensiva de conhecimento.

Figura 5- CommonKADS.



Fonte: Schreiber (1999).

O nível de contexto é composto por três modelos diferentes. São eles, organização, tarefa e agente. O modelo de organização descreve a estrutura da organização como, por exemplo, as divisões, e também, o ambiente mais amplo onde o projeto estará situado, ou seja, os objetivos organizacionais (BETANZOS, 2004). Já o modelo de tarefa descreve as tarefas que são executadas no ambiente organizacional, já o modelo de agente especifica restrições, descreve habilidades, normas, preferências e permissões para que os agentes possam executar as tarefas.

O nível de conceito abrange os modelos de conhecimento e comunicação: o modelo de conhecimento proporciona uma descrição do conhecimento requerido para a execução de uma tarefa, independente de como será feita sua implantação futura. Por sua vez, o modelo de comunicação tem como objetivo a especificação detalhada dos processos de transferência de informação e conhecimento entre os distintos agentes do SBC (BETANZOS, 2004).

Por fim, no nível de artefato o modelo de projeto é responsável por descrever a estrutura do sistema que precisa ser construído.

3.2 WEB E WEB 2.0

O termo *Web* oriundo de (*World Wide Web* ou *WWW*) é muitas vezes considerado sinônimo de Internet apesar de ser, tecnicamente, um serviço executado sobre a Internet (HEATH; BIZER, 2011). Para Jacobs e Walsh (2004) a *Web* é “um repositório de informação, no qual itens de interesse, referidos como recursos, são interligados por identificadores globais chamados URI (*Uniform Resource Identifiers*)”.

Proposta por Tim Berners-Lee no início dos anos 80, a *Web* surgiu como solução para o problema de compartilhar dados em diferentes plataformas. Contudo, somente no ano de 1990 depois de inúmeras tentativas de apoio é que foram criadas as bases para o que conhecemos hoje, o protocolo de transferência de hipertexto HTTP, a linguagem de marcação HTML, os primeiros servidores de HTTP e *Web*, as primeiras páginas, e o primeiro navegador (CAILLIAU, 2011).

Segundo Cailliau (2011), em 1994, o tráfego no primeiro servidor *Web* já era mil vezes maior que nos três anos anteriores. Ao final daquele ano foi criado o W3C, (*World Wide Web Consortium*) que viria a se tornar o órgão mais importante da *Web*, responsável por todos os seus padrões e recomendações.

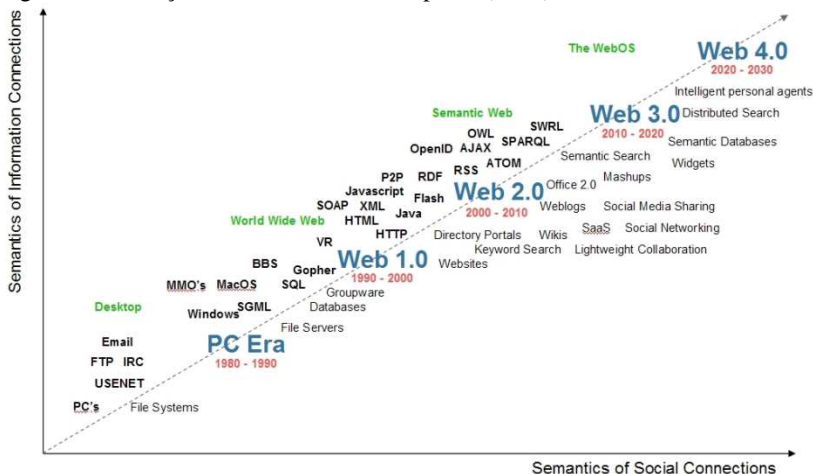
Apesar de ser um termo comum, *Web 2.0* não representa propriamente uma nova versão da *Web*. Nela, não são usadas tecnologias novas, mas sim novas formas de abordagem quanto a possibilidades de interação e compartilhamento de conteúdo.

Segundo Mota (2009), o termo *Web 2.0* apareceu pela primeira vez em 1999 em um artigo publicado na revista “*Print Magazine*” por Darcy DiNucci intitulado “*Fragmented Future*”. O termo foi usado para se referir as mudanças que tornaram a *Web* mais interativa, interconectada e presente no nosso cotidiano (DINUCCI, 2009).

Porém a definição formal do termo é atribuída a Dale Dougherty vice-presidente da *O'Reilly Media, Inc.* na *O'Reilly Media Web 2.0 Conference* de 2004.

Na opinião de Berners-Lee (2011), o termo tem uma conotação mais mercadológica do que uma revolução verdadeira. O termo é reconhecido pela W3C, mas não há recomendações específicas sobre *Web 2.0*.

Figura 6 - Evolução da *Web* na visão de Spivac (2007).



Fonte: Spivack (2007).

O gráfico da Figura 6 proposto por Spivac (2007), apresenta a *Web 2.0*, como a evolução de uma *Web* única. A evolução é caracterizada pelos aspectos mais marcantes do passado e projeção do futuro. Como exemplo, vemos que a *Web 3.0* é caracterizada por seus aspectos semânticos, já a projeção a respeito da *Web 4.0* em 2020 aponta que esta será baseada em Agentes Pessoais Inteligentes e Buscas Distribuídas.

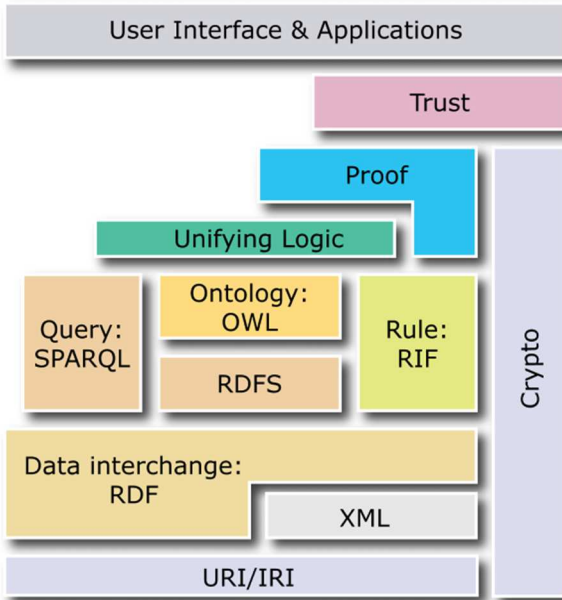
3.3 WEB SEMÂNTICA

Proposta por Berners-Lee, Hendler e Lassila (2001), a *Web Semântica* consiste em uma reestruturação da *Web* baseada em uma série de recomendações sobre a *Web of Data* ou *Web* de Dados. A *Web Semântica* é uma extensão da atual concebida para ligar automaticamente documentos que são semanticamente semelhantes (BONNER, 2002).

Páginas *Web* costumam ser projetadas para serem interpretadas por humanos. Como o conteúdo não é interpretável por máquina e qualquer tipo de automação se torna difícil (LASSILA, 2002). A *Web Semântica* baseia-se em anotações feitas em documentos de tal forma que o seu conteúdo semântico seja acessível e compreensível por ferramentas e agentes de *software* automatizados sem intervenção humana (BONNER, 2002).

A seguir é apresentada a pirâmide composta pelo arranjo das tecnologias que compõe a *Web Semântica*.

Figura 7 - Pirâmide de *Web Semântica*.



Fonte: W3C Elaborado originalmente por (Berners-Lee, 2000).

A compreensão da *Web Semântica* é baseada em cinco principais tecnologias. A primeira, *Extensible Markup Language* (XML) é um formato simples e muito flexível de texto derivado do *Standard Generalized Markup Language* (SGML). XML passou a ser uma recomendação em fevereiro de 1998, com a intenção de se tornar um padrão para a publicação, armazenamento e transferência de documentos por via eletrônica (FILIPPE, 2001; BRAY et al., 2006).

A segunda, URI/IRI, *Uniform Resource Identifier* (URI) é uma sequência compacta de caracteres usada para identificar um recurso abstrato ou físico (BERNERS-LEE; FIELDING; MASINTER, 2005). Já uma *Internationalized Resource Identifier* (IRI) estende a sintaxe de URI a um repertório muito maior de caracteres, internacionalizando-a.

Outra tecnologia é o *Resource Description Framework* (RDF) um meio para o processamento de metadados, que fornece interoperabilidade entre as aplicações para troca de informações

compreensível por máquina na *Web* (LASSILA; SWICK, 1998). RDF aplica-se a situações onde a informação deve ser processada por aplicativos, em vez de ser exibida apenas para as pessoas. Essa característica faz do RDF o padrão para publicação de *Linked Data* que formam a *Web* de Dados (KLYNE; CARROLL, 2004).

Por fim, temos os agentes inteligentes e as ontologias que serão abordadas de forma mais detalhada em seguida.

3.4 ONTOLOGIAS

O termo ontologia é empregado há muito tempo por filósofos como Aristóteles na tentativa de classificar e descrever as coisas do mundo. Contudo, é creditada a Jacob Lorhard, na obra *Ogdoas Scholastica (1606): The Diagraph of Ontology*, a definição do termo ontologia (LAMANNA, 2006). Formado pelas palavras de origem grega “*onto*” que significa “o que é” ou “o que existe” e “*logos*”, que significa “do discurso” ou “de estudo” (GARGOURI; JAZIRI, 2010).

Filosoficamente, o termo trata da natureza do ser e da realidade, principalmente relacionadas a questões metafísicas. Porém, além da perspectiva filosófica tradicional, as ontologias têm sido usadas com diferentes sentidos na área de Inteligência Artificial e Engenharia do Conhecimento (POLI, 2003; VILELA; OLIVEIRA, 2004).

Disciplinas como a IA utilizam o termo ontologia para descrever o que pode ser representado em um determinado domínio de conhecimento por um sistema computacional (GRUBER, 1993; GUARINO, 1998; STUDER; BENJAMINS; FENSEL, 1998).

Dentro dessa visão, a definição mais aceita trata ontologia como: “uma especificação explícita de uma conceitualização” (GRUBER, 1993). Essa definição foi aprimorada por Borst (1997) e diz que: “Uma ontologia é uma especificação formal de uma conceitualização compartilhada”.

Para Gruber (1993), explícita significa que conceitos, relações, instâncias e restrições de uso devem estar explicitamente definidos, e conceitualização refere-se ao modelo abstrato individual de cada pessoa sobre um fenômeno ou domínio determinado.

O termo formal apresentado por Borst (1997), em seu conceito, refere-se aos meios formais de codificação e especificação compreensíveis por agentes humanos e por máquina. Já compartilhada diz respeito ao fato de que uma ontologia deve refletir o conhecimento consensual, aceito por um grupo.

Guizzardi et al. (2009) apontam que uma ontologia é um artefato concreto de engenharia, que representa o modelo de um domínio de conhecimento expresso em uma linguagem de representação. Uma ontologia visa capturar conhecimento consensual de forma genérica e formal para reuso e compartilhamento entre aplicativos ou grupos de pessoas (GUIZZARDI, 2000; CORCHO; FERNÁNDEZ-LÓPEZ; GÓMEZ-PÉREZ, 2003).

Breuker et al. (2009) define ontologias em sua aplicação semântica em uma sigla – CRIA, onde C representa um conjunto de conceitos, R de relações, I de instâncias e A de axiomas, que representam o conhecimento de forma que sejam entendidos e processados por uma máquina.

3.4.1 Uso e benefícios de Ontologias

Ontologias podem ser empregadas em muitas situações com benefícios como facilidades de comunicação e processamento de informação semântica, tanto entre humanos quanto sistemas computacionais, reuso de conhecimento de vocabulários, e abstração de alto nível (GUARINO, 1998).

Ontologias promovem a interoperabilidade entre sistemas ao representarem dados compartilhados por diversas aplicações (USCHOLD; GRUNINGER, 2004). Na *Web Semântica* são peças centrais para estruturar e dar significado ao conteúdo das páginas *Web* (BERNERS-LEE; HENDLER; LASSILA, 2001). Para Blomqvist (2007), ontologias podem ser usadas de diferentes formas, que variam desde agentes autônomos, portais *Web*, intranets corporativas, até sistemas de logística.

Para McGuinness (2003), o caráter lógico das ontologias possibilita que elas sejam utilizadas para descrever vocabulários controlados, sistemas de navegação, personalização de pesquisas, prover a estrutura pela qual se pretende complementar um conteúdo, desambiguação de palavras com base em sentido, checagem de consistência, validação e verificação.

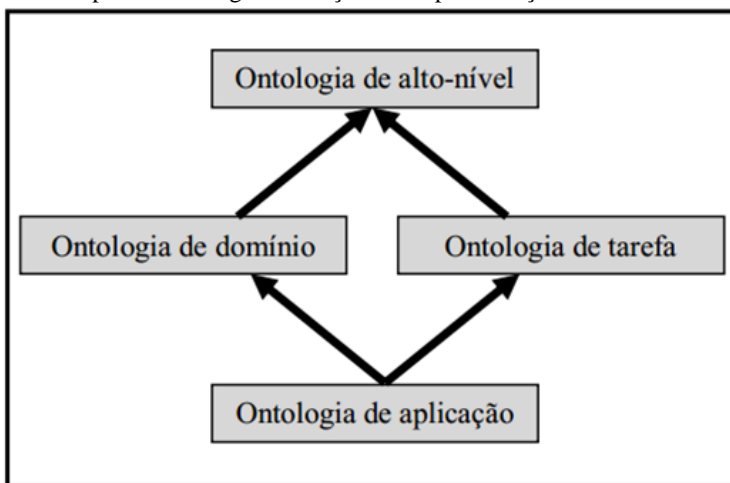
3.4.2 Classificação de Ontologias

De maneira geral, as ontologias podem ser divididas em dois grandes grupos: “ontologias leves” (*lightweight ontologies*) que incluem conceitos, relações e instâncias, e “ontologias pesadas” (*heavyweight ontologies*) que contemplam todos os aspectos de uma ontologia leve

acrescentando-se axiomas e restrições (CORCHO; FERNÁNDEZ-LÓPEZ; GÓMEZ-PÉREZ, 2003).

Guarino (1998) classifica as ontologias de maneira mais detalhada em quatro tipos. A Figura 8 ilustra a relação de especialização entre os tipos de ontologias identificados.

Figura 8 - Tipos de ontologias e relações de especialização.



Fonte: Guarino (1998).

A classificação apresentada na Figura 8 por Guarino (1998) é bastante comum, porém é menos fragmentada e faz menção apenas a quatro tipos de ontologias.

Porém a grande diversidade faz com que as ontologias possam ser classificadas quanto à sua função, grau de formalismo, aplicação, estrutura e conteúdo (ALMEIDA; BAX, 2003). O Quadro 12 apresenta a descrição de cada tipo de ontologia de acordo com a abordagem e suas possíveis classificações.

Quadro 12 - Tipos de Ontologias.

Abordagem	Classificação	Descrição
Quanto à Função Mizoguchi; Vanwelkenhuysen, Ikeda (1995)	Ontologias de Domínio	Reutilizáveis no domínio fornecem vocabulário sobre conceitos, seus relacionamentos, sobre atividades e regras que os governam.
	Ontologias de Tarefa	Fornecem um vocabulário sistematizado de termos, especificando tarefas que podem ou não estar no mesmo domínio.
Quanto ao Grau de Formalismo Uschold e Gruninger (1996)	Ontologias Altamente Informais	Expressa livremente em linguagem natural.
	Ontologias Semi-informais	Expressa em linguagem natural de forma restrita e estruturada.
	Ontologias Semiformais	Expressa em linguagem artificial definida formalmente.
	Ontologia Rigorosamente Formal	Os termos são definidos com semântica formal, teoremas e provas.
Quanto à Aplicação Jasper e Uschold (1999)	Ontologias de Autoria Neutra	Um aplicativo é escrito em uma única língua e depois é convertido para uso em diversos sistemas, reutilizando-se as informações.
	Ontologias Como Especificação	Cria-se uma ontologia para um domínio, a qual é usada para documentação e manutenção no desenvolvimento de softwares.
	Ontologia de Acesso Comum à Informação	Quanto ao vocabulário é inacessível, a ontologia torna a informação inteligível, proporcionando vocabulário compartilhado dos termos.
Quanto à Estrutura Haav e Lubi (2001)	Ontologia de Alto-nível	Descrevem conceitos gerais relacionados a todos os elementos da ontologia (espaço, tempo, matéria, objeto, evento, ação, etc.) os quais são independentes do problema ou domínio.
	Ontologia de Domínio	Escrevem o vocabulário relacionado ao domínio, como, por exemplo, medicina, ou automóveis.
	Ontologia de Tarefa	Descrevem uma tarefa ou atividade, como, por exemplo, diagnósticos ou compras, mediante inserção de termos especializados na ontologia.
Quanto ao Conteúdo (VAN HEIJST; SCHREIBER; WIELINGA, 1997)	Ontologias Terminológicas	Especificam termos que serão usados para representar o conhecimento em um domínio (por exemplo, os léxicos).
	Ontologias de Modelagem do Conhecimento	Especificam conceituações do conhecimento, tem uma estrutura interna semanticamente rica e são refinadas para uso no domínio do conhecimento que descrevem.
	Ontologias de Aplicação	Contém as definições necessárias para modelar o conhecimento em uma aplicação.
	Ontologias de Domínio	Expressam conceituações que são específicas para um determinado domínio do conhecimento.
	Ontologias Genéricas	Similares às ontologias de domínio, mas os conceitos que as definem são considerados genéricos e comuns a vários campos.
	Ontologias de Representação	Explicam as conceituações que estão por trás dos formalismos de representação do conhecimento

Fonte: Almeida e Bax (2003).

Segundo o Quadro 12 proposto por Almeida e Bax (2003), a classificação tipológica de uma ontologia depende diretamente do escopo e objetivo a que esta se propõe. As metodologias de construção de ontologias auxiliam nessa classificação, apontando onde se enquadra determinada ontologia.

3.4.3 Componentes de uma Ontologia

Ontologias possuem componentes básicos como, classes (conceitos), relações, funções, axiomas e instâncias que se referem à especificação de uma conceitualização (GRUBER, 1993; NOY; MCGUINNESS, 2001; GOMEZ-PEREZ; CORCHO, 2002).

✓Classes: podem ser do tipo abstratas ou concretas, simples ou compostas, reais ou fictícias. Em suma, uma classe pode representar “qualquer coisa” a respeito de “algo” que será descrito. Classes possuem atributos que revelam propriedades importantes do domínio. As classes possuem taxonomias que organizam o conhecimento ontológico usando generalização e especialização, onde podem ser aplicadas relações de herança entre as classes.

✓Relações e funções: relações são um tipo de interação entre os conceitos de um domínio e seus atributos. Já as funções são um tipo especial de relação.

✓Axiomas: são utilizados para modelar sentenças verdadeiras, restringindo a interpretação e o uso dos conceitos envolvidos na ontologia. Axiomas são uteis para impor restrições, verificar a correção e realizar dedução de novas informações.

✓Instâncias: as instâncias ou indivíduos de uma ontologia são os dados em si. Representam elementos do domínio associados a um conceito específico. As instâncias possuem atributos que são propriedades relevantes que descrevem as instâncias de um conceito.

3.5 DADOS ABERTOS

Nos últimos anos a publicação de dados abertos registrou grande crescimento. Organizações públicas e privadas têm publicado de forma aberta dados de diferentes domínios 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, a exigência de creditar a sua autoria e compartilhar-los pela mesma licença. Esta última exigência

é geralmente satisfeita pela publicação dos dados em formato aberto e sob uma licença aberta.

Alguns dos formatos de arquivos mais comuns para publicação de dados abertos são o JSON, XML, RDF e CSV. Contudo, é comum que os dados, independente do formato de disponibilização, sejam heterogêneos e não possuam nenhum tipo de tratamento.

Propostas como a *Web Semântica*, onde significado é atribuído aos dados, ou *Linked Data*, que transforma a *Web* em um repositório de dados global composto por diferentes domínios visam minimizar esses problemas por meio do aumento da estruturação dos dados (BIZER; HEATH, BERNERS-LEE, 2009).

David Eaves (2009), especialista em políticas públicas e ativista dos dados abertos propôs três leis a respeito desse tipo de dados:

1. Se o dado não pode ser encontrado e indexado na *Web*, ele não existe;
2. Se não estiver aberto e disponível em formato compreensível por máquina, ele não pode ser reaproveitado; e
3. Se algum dispositivo legal não permitir sua replicação, ele não é útil.

Essas leis foram propostas por Eaves para os Dados Abertos Governamentais, mas é possível aplica-las aos Dados Abertos de forma geral. Já a *Open Government Data Foundation* reuniu-se em dezembro de 2007 na Califórnia e desenvolveu uma lista de recomendações que ficaram conhecidas como *Open Government Data Definition: The 8 Principles of Open Government Data*, onde os dados devem ser:

1. **Completos:** o conjunto completo de dados deve ser publicado, sem limitações válidas de privacidade, segurança ou controle de acesso. Dados, nesse contexto, são informações eletronicamente gravadas, incluindo, mas não se limitando a, documentos, bancos de dados, transcrições e gravações audiovisuais.
2. **Primários:** publicados como coletados na fonte, na menor granularidade possível, sem agregações ou transformações.
3. **Atuais:** disponibilizados o mais rápido possível para preservar o seu valor.
4. **Acessíveis:** disponibilizados para o público mais amplo possível e para os propósitos mais variados.

5. **Processáveis por máquina:** devem ser razoavelmente estruturados para possibilitar o seu processamento automatizado.
6. **Acesso não discriminatório:** disponíveis a todos, sem que seja necessária identificação ou registro.
7. **Formatos não proprietários:** disponíveis em um formato aberto sem controle exclusivo.
8. **Livres de licenças:** não devem ser sujeitos a regulações de direitos autorais, marcas, patentes ou segredo industrial. Restrições razoáveis de privacidade, segurança e controle de acesso podem ser permitidas na forma regulada por estatutos.

Além das características descritas é importante designar uma pessoa como contato responsável pelos dados existentes e novas requisições. Como as três leis propostas por Eaves (2009), estes princípios foram discutidos inicialmente para os Dados Abertos Governamentais, contudo, aplicáveis a dados abertos de modo geral.

3.5.1 Classificação dos Dados Abertos

Tim Berners-Lee sugeriu um esquema de implementação onde os dados abertos classificados em cinco estrelas conforme sua maturidade (BERNERS-LEE, 2006).

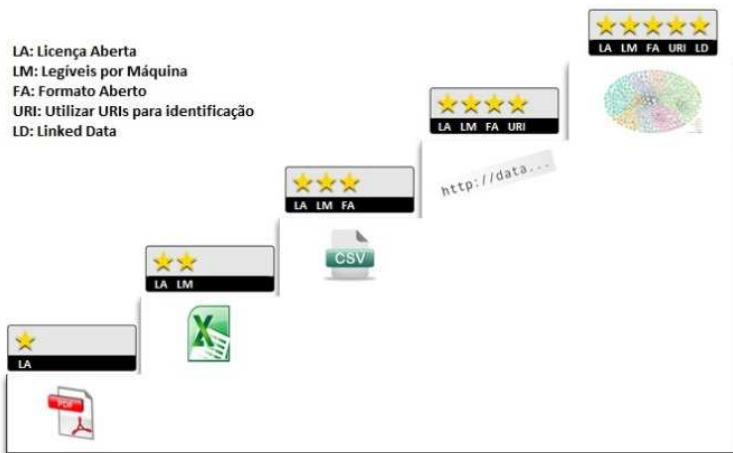
Essencialmente, este sistema de classificação de cinco estrelas, é uma reinterpretação da pirâmide hierárquica de necessidades proposta na obra "*A Theory of Human Motivation*" Maslow's (1943) aplicada ao ambiente de dados abertos. Analogicamente ao primeiro nível proposto por Maslow's, as necessidades fisiológicas, a primeira estrela reflete os requisitos básicos para a publicação de dados. Já a quinta estrela corresponde ao melhor dos cenários como mostra a Figura 9.

Na classificação proposta por Berners-Lee (2006), dados que possuem uma estrela são disponibilizados na *Web* em qualquer formato, mas usando uma licença aberta. Para quem consome, é possível consultar, imprimir, armazenar, usar em outro sistema, compartilhar, ou mesmo alterar os dados que desejar. Já para quem publica é simples uma vez que isso é feito simplesmente disponibilizando os dados em um formato aberto sem tratamento específico.

Dados com duas estrelas são estruturados e legíveis por máquina, por exemplo, uma planilha XLS do Microsoft Excel. Como consumidor você pode fazer o mesmo que é feito com dados uma estrela e,

adicionalmente processá-los com o software proprietário ou exportar os dados para outro formato. Para quem publica ainda é fácil disponibilizar.

Figura 9 - Classificação cinco estrelas de dados abertos proposta por Berners-Lee.



Fonte: Berners-Lee (2006).

Com três estrelas os dados são publicados como em duas estrelas, mas em um formato não proprietário, como CSV em vez de XLS. Para quem consome, são mantidas as possibilidades anteriores e agregada a possibilidade de manipular os dados como desejar, sem se limitar aos recursos de um *software* específico. Já para quem publica, podem ser necessários conversores ou *plug-ins* para exportar os dados do formato proprietário, sendo, porém, ainda bastante simples de publicar.

Com quatro estrelas, são usados padrões W3C como RDF e SPARQL para identificar as coisas, de modo que pessoas possam utilizar seu conteúdo. Para quem consome, isso possibilita ligar este conteúdo em qualquer outro lugar da *Web* ou localmente, fazer marcações, reutilizar partes desses dados ou mesmo combinar os dados de forma segura com outros dados usando URIs. Para quem publica, é preciso atribuir URIs aos itens de dados e pensar em como representar os dados em um padrão previamente definido. Com isso outras fontes de dados agora podem ligar seus dados, promovendo-os a cinco estrelas.

Por fim dados cinco estrelas consiste em ligar seus dados com os dados de outras pessoas para fornecer contexto. Como consumidor você pode descobrir outras informações relacionadas aos dados, ou o próprio esquema de dados. Contudo podem existir *links* de dados quebrados,

erros em páginas da *Web* e ainda enfrentar problemas de confiança já que dados são apresentados a partir de uma ligação arbitrária. Como editor, seus dados serão detectáveis na *Web*, o que aumenta o valor, além de obter os mesmos benefícios dos *links* como os consumidores.

3.6 WEB OF DATA

Um número significativo de indivíduos e organizações têm adotado *Linked Data* como maneira de publicar seus dados, não apenas para colocá-los na *Web*, mas usando *Linked Data* para consolidá-los na *Web* (MENDELSON, 2009). O resultado é um espaço global de dados que chamamos de *Web of Data* ou *Web* de Dados (BIZER; HEATH; BERNERS-LEE, 2009).

A *Web* de Dados forma um grafo mundial gigante Berners-Lee (2007), que consiste em bilhões de declarações RDF provenientes de várias fontes cobrindo temas variados, tais como: locais geográficos, pessoas, empresas, livros, publicações científicas, filmes, música, televisão e programas de rádio, biologia, ensaios clínicos, dados estatísticos, dados governamentais, entre outros.

A *Web* de Dados pode ser vista como uma camada adicional fortemente interligada com os documentos da *Web* tradicional tendo muitas das mesmas propriedades (HEATH; BIZER, 2011):

1. A *Web* de Dados é genérica e pode conter qualquer tipo de dado.
2. Qualquer um pode publicar dados na *Web* de Dados.
3. A *Web* de Dados é capaz de representar desacordos e informações contraditórias sobre uma entidade.
4. Entidades são conectadas por *links* RDF, criando um grafo de dados global, onde as aplicações não precisam ser pensadas para um conjunto fixo de fontes de dados, mas sim para descobrir novas fontes de dados em tempo de execução seguindo os *links* RDF.
5. A publicação de dados não é restrita quanto a escolha do vocabulário que será usado para representar os dados.
6. A *Web* de Dados é auto-descritiva. Se um aplicativo consumindo *Linked Data* encontrar dados descritos com um vocabulário desconhecido, o aplicativo pode dereferenciar os URIs que identificam termos de vocabulário, a fim de encontrar a sua definição.

7. O uso de HTTP como mecanismo padronizado de acesso a dados RDF como que representa um modelo padronizado de dados facilita o acesso aos dados em comparação com APIs *Web*, que dependem de modelos de dados heterogêneos e interfaces de acesso.

3.6.1 Crescimento da *Web* de Dados

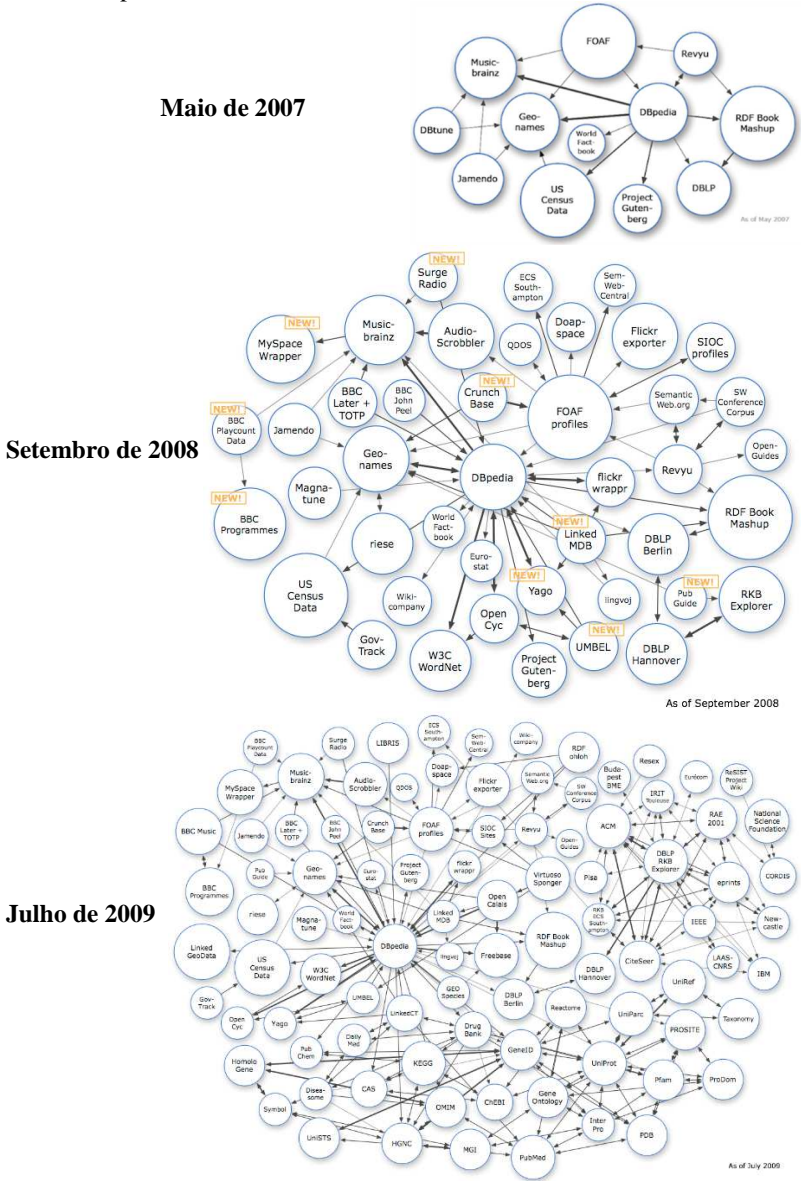
As origens da *Web* de Dados remontam aos esforços da comunidade de pesquisa da W3C sobre *Web* Semântica, particularmente das atividades do *Linking Open Data (LOD) project*,⁵ fundado em janeiro 2007. O objetivo do projeto é estender a *Web* por meio da publicação de conjuntos de dados abertos no formato RDF, estabelecendo *links* RDF entre os dados de diferentes fontes.

É princípio do projeto a abertura para quem desejar publicar dados seguindo os princípios de *Linked Data*. Segundo Heath e Bizer (2011) esta abertura é o provável fator para o sucesso do projeto como ponto de partida da *Web* de Dados.

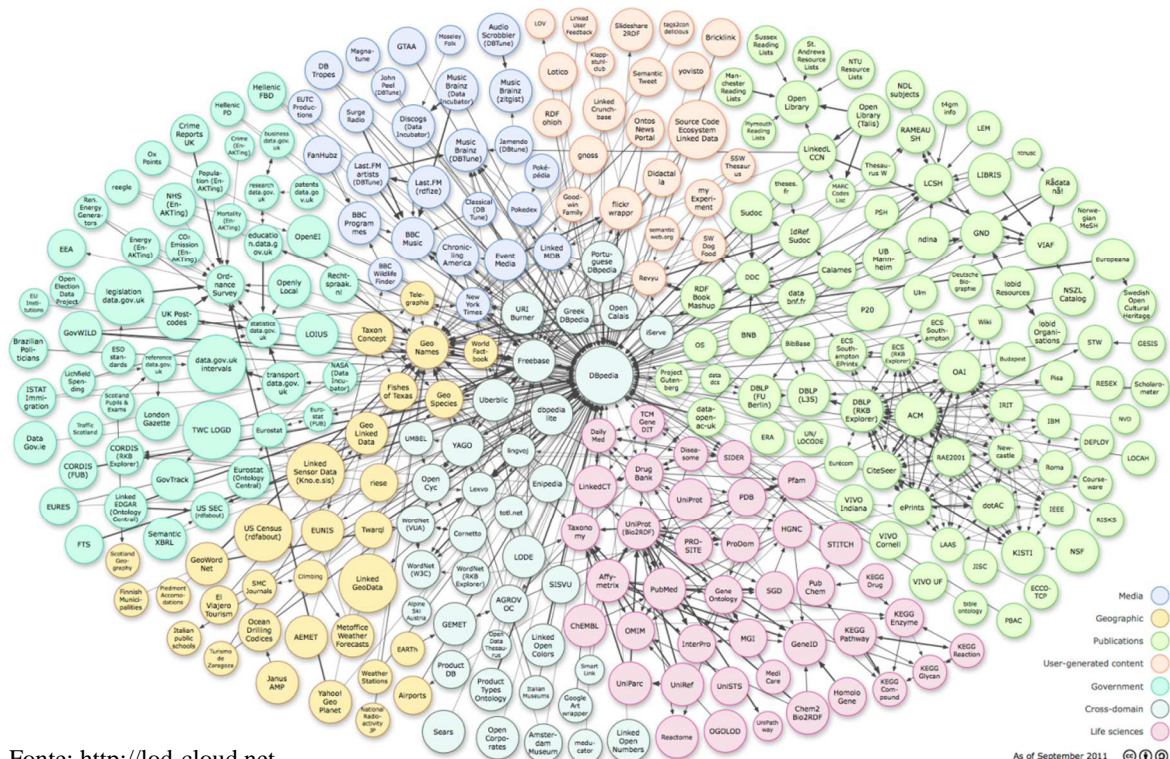
A Figura 10, a Figura 12 e a Figura 12 apresentam os diagramas de crescimento dos conjuntos de dados publicados na *Web* desde o início do *Linking Open Data Project*. Cada nó neste diagrama de nuvem representa um conjunto de dados distinto publicado segundo os princípios de *Linked Data*. Os arcos indicam a existência de *links* RDF entre itens dos dois conjuntos de dados. Arcos maiores correspondem um número maior de ligações, enquanto que os arcos bidirecionais indicam que existem ligações externas em cada conjunto de dados.

⁵ <http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>

Figura 10 - Crescimento da quantidade de conjunto de dados no formato de *Linked Data* publicados na *Web*.



Fonte: Linking Open Data Project (<http://lod-cloud.net/>).

Figura 11 – Tamanho da nuvem de *Linked Data* em abril de 2011.

O tamanho dos círculos corresponde ao número de recursos presentes em cada conjunto de dados. Os números são usualmente fornecidos pelos editores do conjunto e são estimativas por vezes grosseiras.

Tabela 1- Correspondência entre tamanho dos círculos e o número de triplas

Tamanho do Círculo	Número de Triplas
Muito Grande	>1 bilhão
Grande	10 milhões -1 bilhão
Médio	500 mil - 10 milhões
Pequeno	10 mil - 500 mil
Muito Pequeno	<10 mil

Fonte: <http://lod-cloud.net/>

Uma seta no diagrama indica a existência de pelo menos cinquenta ligações entre os dois conjuntos de dados. Uma ligação, neste caso, é uma tripla RDF, onde as URIs que representam sujeito e objeto estão em *namespaces* de diferentes conjuntos de dados.

O diagrama da Figura 11 de setembro de 2011 apresenta 295 conjuntos de dados, que juntos, somam mais de 31 bilhões de triplas RDF interligadas por cerca de 504 milhões de ligações RDF. Já a Figura 12 apresenta a última atualização do diagrama, com dados até abril de 2014 e publicada em 30 de agosto de 2014. Nesta atualização estão presentes 1014 conjuntos de dados e algo em torno de 108 bilhões de triplas.

3.6.2 Topologia da *Web* de Dados

Esta seção apresenta uma visão topológica da *Web* de Dados. Os conjuntos de dados estão classificados nas áreas: geográficas, governo, mídias, publicações, ciência da vida, conteúdo gerado por usuários, *Social Web*, e de domínios variados.

A Tabela 2 apresenta um resumo do número de conjuntos de dados segregados segundo os domínios definidos nos documentos de estado da nuvem de LOD, que regularmente compila resumos estatísticos sobre os conjuntos de dados catalogados no *LOD Cloud Data Catalog on, Comprehensive Knowledge Archive Network* (CKAN).

Tabela 2 - Número de conjuntos de dados, quantidade de triplas e quantidade de ligações RDF por tópico de domínio.

Domínio	Número de datasets (2011)	Número de datasets (2014)	Número Triplas (2011)	Datasets % (2011)	Datasets % (2014)
Media	25	22	1.841.852.061	8,47 %	2,17 %
Geographic	31	21	6.145.532.484	10,50 %	2,07 %
Government	49	183	13.315.009.400	16,61 %	18,05 %
Publications	87	96	2.950.720.693	29,50 %	9,47 %
Cross-domain	41	41	4.184.635.715	13,89 %	4,04 %
Life sciences	41	83	3.036.336.004	13,89 %	8,19 %
User-generated content	20	48	134.127.413	6,77 %	4,73 %
Social Web	-	520	-	-	51,28 %
Total	295	1014	31.634.213.770	100 %	100 %

Fonte: <http://lod-cloud.net/>

As informações presentes na Tabela 2 mostram que em 2011 a maior quantidade de *datasets* era sobre Publicações com 29,50 % do total. Contudo, os dados governamentais apresentavam o maior volume,

com aproximadamente 40 % das triplas publicadas. Para 2014 os dados sobre o número de triplas não estão disponíveis até o momento. Porém, ao analisarmos os dados sobre o número de *datasets* é visível um crescimento considerável nos domínios Governo, ciências da vida e conteúdo gerado por usuários. Além disso, foi incluído um novo domínio, o *Social Web* que já de início apresenta o maior volume de *datasets*, 520 ou 51,28 % do total.

3.7 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*. Como na *Web* formada por *links hypertext*, a rede de dados em *Linked Data* é construída com documentos na *Web*, porém, as ligações são feitas usando *links hyperdata*, onde coisas, ou seja, a informação presente em 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. *Linked Data* difere-se dos demais por ser voltado estritamente para a publicação de dados, e não para apenas adicionar anotações semânticas a textos da *Web* clássica.

3.7.1 Princípios de *Linked Data*

O termo *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). Essas práticas foram introduzidas em nota intitulada “*Linked Data*” e ficaram conhecidas como princípios do *Linked Data* (BERNERS-LEE, 2006):

1. Usar URIs para nomear as coisas.
2. Usar HTTP URIs, assim pessoas podem consultar esses nomes.
3. Quando alguém consultar uma URI, prover informações úteis, usando padrões (RDF, SPARQL).
4. Incluir *links* para outras URIs, então poderá se descobrir mais coisas.

Por traz destes quatro princípios, percebe-se que o objetivo de *Linked Data* é utilizar a arquitetura *World Wide Web*, para compartilhar dados estruturados em uma escala global (JACOBS; WALSH, 2004; HEATH; BIZER, 2011).

Ao publicarmos dados nesse formato estamos adicionando dados em um espaço global, permitindo que estes sejam descobertos e usados por outras aplicações. Publicar um conjunto de dados envolve segundo (BIZER; HEATH; BERNERS-LEE, 2009):

1. Atribuir URIs às entidades descritas pelo conjunto de dados e prever o dereferenciamento dessas URIs sobre o protocolo HTTP em representações RDF;
2. Definir ligações RDF para outras fontes de dados na *Web*, possibilitando que os clientes possam navegar pela rede de dados como um todo, seguindo os *links* RDF;
3. Fornecer metadados sobre os dados publicados, de modo que se possa avaliar a qualidade do que é publicado e escolher entre os diferentes meios de acesso.

Procedimentos similares são encontrados em (BERNERS-LEE, 2006; SAUERMAN; CYGANIAK; VÖLKE, 2011).

O formato padrão para serialização de *Linked Data* é o RDF. O RDF é uma linguagem criada inicialmente para representar e identificar conteúdo semântico em páginas *Web*. Mas, ao generalizarmos o conceito de *Web*, o RDF pode ser utilizado para identificar qualquer informação ou objeto presente na rede.

3.7.2 Classificação de *Linked Data*

A facilidade de publicação, em conjunto com a tendência de abertura de dados, fez surgir uma variada gama de tipos de dados e de termos derivados do conceito inicial de *Linked Data* (BIZER; CYGANIAK; GAUB, 2007).

À medida que dados de domínios diversos são publicados na forma de *Linked Data*, diferentes nomenclaturas são utilizadas na literatura para classificá-los. Para demonstrar algumas dessas novas propostas de *Linked Data* foi elaborada a Tabela 3 que apresenta um comparativo entre as diferentes nomenclaturas, levando-se em conta características como, natureza, origem, iniciativas e utilização dos dados. Em seguida estes termos serão apresentados de forma mais detalhada.

Tabela 3 – Classificação dos tipos de *Linked Data*.

Tipo	Estrutura			Iniciativas	
	Conteúdo	Domínio dos dados	Participantes	Propósito	Referências
Linked Government Data	Dados brutos de Agencias Governamentais	Público	Organizações Governamentais	- Transparência e abertura de dados. - Interligação, exploração e análise dos dados.	Meskeil (2007); TAYLOR (2010); Hyland e Wood (2011); Villazón-Terrazas et al., (2011)
Linked Enterprise Data	Mídias, finanças, produção, negócios, etc.	Geralmente internos a organização	Organizações Privadas	- Criação da informação está intimamente ligado com o ato de compartilhamento. - Ligação de dados internos com fontes externas, ferramenta, e novas formas de visualização.	Servant (2008); HU (2010); Allemang (2011)
Statistical Linked Data	Dados com séries estatísticas dos mais variados domínios	Público e Fechado	Pesquisadores, Organizações governamentais e privadas	- Sistemas OLAP por meio da integração estruturada de <i>Linked Data</i> , utilizando consultas em SPARQL. - Analisar grandes quantidades de dados numéricos de forma exploratória.	Cygniak, Reynolds e Tennison (2010); Kämpgen, O’Rain e Harth (2012)
Geo Linked Data	Dados Geográficos	Público e Fechado	Pesquisadores, Organizações governamentais e privadas	- Enriquecer a Web de dados com dados ligados geoespaciais.	Auer et al., (2009); Vilches-Blázquez et al., (2010) Stadler et al., 2012
Linked Sensor Data	Proveniente de sensores	Público e Fechado	Pesquisadores, Organizações governamentais e privadas	- Dados de sensores e metadados de acesso público armazenados na nuvem de Linked Data.	Patni, Henson e Sheth (2010); Janowicz et al., (2010); Pschorr et al., (2010).

3.7.2.1 *Linked Government Data*

Embora seja indiscutível a importância da abertura dos dados por parte de agentes governamentais, a simples publicação passiva desses dados é insuficiente, já que, somente sua liberação em formatos abertos, estruturados e legíveis torna possível a terceiros utilizar, manipular e visualizar esses dados (FRANTZICH; SULLIVAN, 1996; MESKELL, 2007).

Governos como o do Reino Unido, EUA, Canadá, França, Espanha, Suécia, Itália, Hong Kong, Austrália, Nova Zelândia e Brasil são alguns dos que possuem portais dedicados a tornar acessíveis dados públicos, sendo que, alguns desses já estão disponíveis como *Linked Data*⁶.

A aplicação dos princípios de *Linked Data* a bases de dados governamentais traz um enorme potencial (HEATH; BIZER, 2011). Isso permite um esquema aberto e extensível com informações explicativas sobre um determinado domínio, interligado com dados relacionados de outras fontes (HYLAND; WOOD, 2011).

Contudo, para Cyganiak, Reynolds e Tennison (2010) esse potencial não é completamente explorado, principalmente pela falta de recursos necessários para transformar, em grande escala, dados brutos de alta qualidade em *Linked Data*.

Semelhante a um processo de Engenharia de *Software*, a publicação de *Linked Government Data* (LOGD) necessita de um ciclo de vida (TAYLOR, 2007). Este processo é cíclico incremental, baseado na melhora contínua e expansão dos *Linked Government Data* resultante de várias interações (VILLAZÓN-TERRAZAS et al., 2011). As diretrizes para o processo de publicação de LOGD consistem em cinco atividades principais: (1) Especificação, (2) Modelagem, (3) Geração, (4) Publicação e (5) Exploração.

As atividades são decompostas em uma ou mais tarefas, e algumas técnicas e ferramentas são utilizadas para sua realização. A ordem pode ser alterada conforme necessidades específicas dos órgãos governamentais. Além disso, é importante obter *feedback* contínuo para prover melhoras constantemente. A seguir é apresentada a Tabela 4 com as principais tarefas de cada atividade.

⁶ <http://www.data.gov/community>

Tabela 4- Principais tarefas de *Linked Government Data*.

Atividade	Tarefas
Modelagem	<ol style="list-style-type: none"> 1. Busca recursos governamentais de <i>Web sites</i> altamente confiáveis, sites de domínio relacionados e catálogos de governo; 2. Selecione os recursos governamentais mais adequados; 3. Transformar esses recursos em ontologias.
Geração	<ol style="list-style-type: none"> 1. Transformação dos dados; 2. Limpeza dos dados; 3. Ligação dos dados.
Especificação	<ol style="list-style-type: none"> 1. Identificação e análise das fontes de dados do governo; 2. O projeto URI; 3. Definição de licença.
Publicação	<ol style="list-style-type: none"> 1. Publicação dos conjuntos de dados; 2. Publicação de metadados; 3. Permitir a descoberta efetiva.
Exploração	<ol style="list-style-type: none"> 1. Permitir a transparência; 2. Entregar aplicativos públicos; 3. Incentivar o uso público da informação governamental.

Fonte: Elaborado com base em (VILLAZÓN-TERRAZAS et al., 2011).

3.7.2.2 *Enterprise Linked Data*

Com a demonstração do valor e as histórias de sucesso de projetos acadêmicos e de dados públicos, iniciativas de *Linked Data* passaram a chamar a atenção de grandes companhias (HU; SVENSSON, 2010).

Segundo Allemang (2010), as experiências na *Web* sugerem que existem caminhos para empresas construírem arquiteturas sustentáveis com suas informações empresariais, transformando-se em “*Linked Data enterprises*”, onde o ato de criação de informação está intimamente ligado com o ato de compartilhamento, sendo este, tão importante como sua produção.

As contribuições da comunidade *Linked Open Data* (LOD), como a descrição de melhores práticas e a publicação de grandes conjuntos de dados em RDF, darão início a uma nova era de possibilidades para a criação de *mashups* inovadores usando este tipo de dados. Neles os sistemas de informação de uma companhia podem ser imaginados como um espaço de *Linked Data* (SERVANT, 2008).

Abordagens como as apresentadas por Servant (2008) e Hu e Svensson (2010) descrevem casos de aplicação de *Linked Data* em grandes organizações. Isso ocorre por meio da integração das tecnologias de informação e comunicação existentes, e dos dados internos da organização, com as abordagens de *Linked Data*. O uso de

Linked Data possibilita a ligação de dados a outras fontes externas, o que resulta em novas possibilidades de visualização.

Além da interligação de dados, constitui boa prática a reutilização de vocabulários sempre que possível. Allemang (2010) lembra que as empresas geralmente possuem vocabulários refinados exclusivos para uso interno, além de ontologias ou taxonomias que podem ser úteis a outras organizações, e que devem ser convertidos para padrões internacionais como o SKOS (MILES; BECHHOFFER, 2009) e o void (CYGANIAK et al., 2001).

3.7.2.3 *Statistical Linked Data*

O uso de *Linked Data* permite o acesso a grandes quantidades de dados serializados dos mais variados domínios. Contudo, esses dados costumam ser tipicamente heterogêneos e sem tratamento estatístico integrado (ZAPILKO E MATHIAK, 2011).

Além disso, mecanismos de interação de *Linked Data* como, navegadores, interfaces de busca facetadas e de consultas, não permitem que os usuários possam analisar grandes quantidades de dados numéricos de forma exploratória (KOBILAROV et al., 2009; HARTH, 2010; MILOŠEVIĆ et al., 2012).

Para melhorar a forma como os dados são analisados, uma abordagem comumente utilizada no tratamento de dados estatísticos é a utilização de cubos multidimensionais. Esses cubos foram oficialmente descritos no artigo de 1993 intitulado “*Providing OLAP (Online Analytical Processing) to User-Analysts: An IT Mandate*”, escrito por W.H. Inmon, R. Kimball, and E.F. Codd.

Segundo Cyganiak, Reynolds e Tennison (2010), esses cubos compreendem uma coleção de observações feitas em algum ponto sobre um espaço lógico. Essa coleção é caracterizada por um conjunto de dimensões que definem o escopo de cada observação, juntamente com os metadados descrevendo o que foi medido, como foi medido e de que maneira as observações são expressas.

O uso de operações OLAP (*Online Analytical Processing*) em cubos de dados possibilita a visualização de grandes quantidades de dados estatísticos a partir de diferentes ângulos, granularidades, permitindo a filtragem e comparação de medidas, representando em uma interface de apoio à tomada de decisão (CHAUDHURI; DAYAL, 1997; TRUJILLO, 2008).

As operações típicas do OLAP são de *roll-up* (aumenta o nível de agregação) e *drill-down* (diminui a agregação e aumenta o

detalhamento). Já as operações de *Slice* e *dice* (seleção e projeção respectivamente) trabalham a informação, modificando as posições sempre que houver necessidade de pivoteamento, possibilitando sumarizar e agrupar os dados em diversos formatos (CHAUDHURI; DAYAL, 1997).

Dois desafios são apontados por Kämpgen; O’rain, Harth (2012): OLAP requer um modelo de cubo de dados, dimensões e medidas. Criar automaticamente um esquema multidimensional de Dados Ligados genéricos é difícil (ROMERO; ABELLÓ, 2007; NEBOT et al., 2009; PARDILLO; MAZÓN, 2011). Ainda, consultas OLAP são complexas e requerem modelos de dados especializados como, por exemplo, o modelo estrela usado em bases de dados relacionais (GRAY et al., 1997).

Essa abordagem de *Statistical Linked Data* pode ser observada em Kämpgen e Harth (2011) que utilizam dados de diversas bases disponíveis na *Web* em um sistema OLAP. Os autores apresentam uma forma de interagir com dados estatísticos em um cubo modelado em RDF, possibilitando consultas em SPARQL.

Dentre os vocabulários destinados à representação de dados estatísticos, destacam-se o *Statistical Data and Metadata Exchange* (SDMX) e o *RDF Data Cube Vocabulary* (QB). Segundo Cyganiak; Reynolds, Tennison (2010) a especificação SDMX define ainda um modelo de informação central expresso em RDF chamado SDMX-RDF.

O *RDF Data Cube Vocabulary* (CYGANIAK; REYNOLDS; TENNISON, 2010) tem o propósito de permitir a publicação de dados estatísticos na *Web*, fornecendo um metamodelo para conjuntos de dados multidimensionais, análogo ao modelo relacional utilizando-se do vocabulário SDMX.

3.7.2.4 Outras classificações de *Linked Data*

Além das abordagens anteriormente apresentadas, *Linked Data* abrange uma variada gama de domínios, como pessoas, dados geográficos, dados provenientes de sensores, entre outros. Duas dessas abordagens são aqui apresentadas de forma resumida: *GeoLinked Data* e *Linked Sensor Data*.

GeoLinked Data é uma iniciativa espanhola, cujo objetivo é enriquecer a *Web* de Dados com dados geoespaciais. Esta iniciativa começou com a publicação de fontes de informação pertencentes ao Instituto Geográfico Nacional da Espanha. Essas fontes são disponibilizadas em RDF de acordo com as práticas de *Linked Data*

(VILCHES-BLÁZQUEZ et al., 2010). Outros trabalhos relacionados são descritos em (AUER; LEHMANN; HELLMANN, 2009; STADLER et al., 2012).

Já a ideia por trás da *Linked Sensor Data*, proposta por Patni, Henson e Sheth (2010), é fazer com que os dados de sensores e metadados de acesso público sejam armazenados na nuvem de *Linked Data*. Seguindo a tendência de tornar os dados acessíveis ao público, os dados de sensores foram interligados usando uma ontologia⁷. Janowicz et al. (2010) e Pschorr et al. (2009) apresentam trabalhos relacionados.

3.8 CONSIDERAÇÕES SOBRE O CAPÍTULO

Este capítulo apresentou os conceitos básicos sobre Engenharia do Conhecimento, trouxe uma abordagem evolutiva sobre a *Web* que passou pela *Web 2.0* e *Web Semântica* até chegar a *Web* de Dados.

A principal contribuição trazida pela *Web Semântica* é a atribuição de semântica ao conteúdo. Essa característica acrescenta a *Web* tradicional a possibilidade de ligação automática de documentos que são semanticamente semelhantes (BONNER, 2002). O resultado é um espaço global de dados que chamamos de *Web* de Dados (BIZER; HEATH; BERNERS-LEE, 2009).

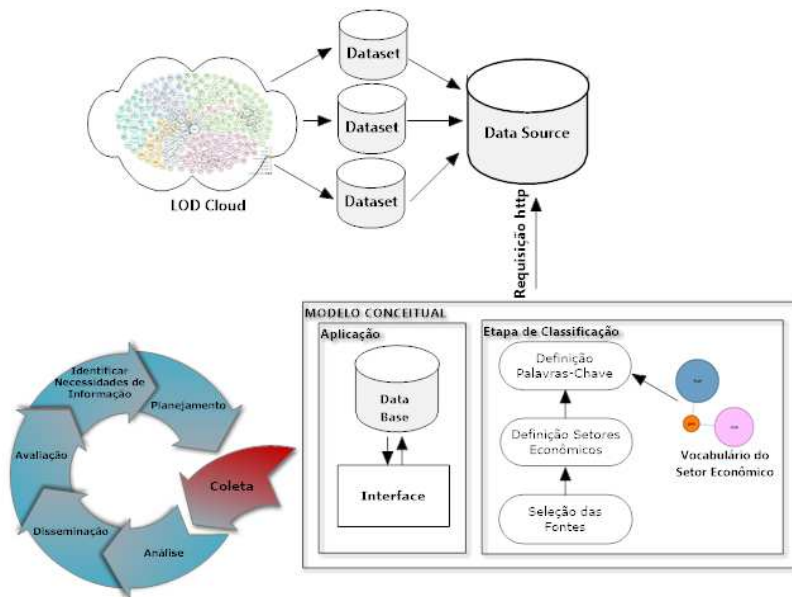
Estas anotações semânticas são a principal característica que faz da *Web* de Dados uma fonte de informação estruturada, mais consistente e de fácil recuperação para a etapa de coleta do processo de IC. O uso da *WoD* pode representar ganho de tempo na etapa de coleta uma vez que os dados possuem tratamento prévio e são ligados a outras fontes. Isso pode auxiliar não somente na coleta mas também na etapa de análise do ciclo de IC.

⁷ http://knoesis.wright.edu/research/semsci/application_domain/sem_sensor/ont/sensor-observation.owl

4 MODELO PROPOSTO E SUA APLICAÇÃO

Este capítulo apresenta o modelo conceitual proposto nesta dissertação. O modelo reúne as etapas e os elementos necessários para realizar a coleta de informações que servirão para apoiar o ciclo de Inteligência Competitiva. A Figura 13 apresenta este modelo.

Figura 13 - Modelo conceitual da Proposta.



O modelo proposto consiste em um arranjo de tecnologias e de 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 *Web* de Dados. A *WoD* proporciona o acesso a informações consistentes, semanticamente estruturadas e interligadas. A seguir será apresentada cada uma das etapas presentes no modelo.

4.1 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 definição dos termos usados para representar a informação. Esses

termos, posteriormente serão utilizados como palavras-chave na busca e recuperação da informação. Nos tópicos seguintes são discutidas cada uma dessas tarefas.

4.1.1 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 2011. O diagrama da *LOD Cloud* de 2014 não foi usado como base devido a sua disponibilização ser posterior ao término deste trabalho. Do conjunto inicial foram selecionados ao fim do processo 135 *data sources*, que seguindo os critérios especificados, possuem alguma relevância para o processo de IC. A seguir são elencados os critérios de seleção utilizados.

1. Presença de *links* ativos: Foram eliminados *data sources* que mesmo representado no diagrama não possuem mais *links* ativos. Esta análise resultou na exclusão de 26 *data sources*;
2. Disponibilidade dos *data sources*: foram excluídos 27 *data sources* que, apesar de estarem representados no diagrama e seus *links* apontarem para os *data sources*, não possuem dados recuperáveis disponíveis;
3. Fontes de dados duplicadas: foram excluídas 36 fontes duplicadas. Estas estão representadas mais de uma vez ou são contempladas como conjuntos de dados dentro de outras fontes;
4. Relevância dos dados: o conteúdo dos *data sources* foi analisado quanto a sua relevância para o processo de IC e os que não apresentam informações pertinentes foram excluídas. Alguns dos temas abordados por fontes eliminadas são: religião, vocabulários de prefixos, vocabulários de unidades métricas e repositórios sobre desenhos animados. Foram excluídos ainda *data sources* que abordam um grupo muito restrito de dados, tais como: dados sobre um determinado bairro, ou um museu. Por fim, foram eliminadas fontes em línguas de difícil compreensão como japonês, chinês, cirílico, entre outros. Nesta etapa foram excluídas 71 fontes, resultando em 135 *data sources* que formam o escopo deste trabalho.

Cabe ressaltar que os 135 *data sources* são compostos, por um ou mais *datasets*. O uso desta abordagem possibilita a separação dos dados

em diferentes assuntos. Por exemplo, um *data source* sobre a indústria farmacêutica pode ser composto por um *dataset* relativo a componentes químicos, um sobre medicamentos e outro que contenha as empresas pertencentes a este setor. Um *dataset*, por sua vez, é um conjunto de informações que são publicadas, mantidas ou agregadas por um único provedor (ALEXANDER et al., 2009).

Estes conjuntos, devido à interligação existente entre eles, podem aparecer em mais de um *data source* simultaneamente. Assim, é difícil apontar de maneira precisa o número total de *datasets*. Contudo, uma estimativa superficial feita com base nas informações presentes nas próprias bases aponta para a existência de aproximadamente 150 mil *datasets* compostos por algo em torno de 50 bilhões de recursos⁸.

4.1.2 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. A Figura 14 apresenta o resultado da classificação dos *data sources* resultantes da etapa de seleção das fontes de dados.

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 5. Como pode ser percebido, o somatório de fontes por seção de setor CNAE é superior a 135, isso se deve ao fato de algumas fontes de dados abrangerem mais de um tema, nesses casos, a fonte pode ser classificada em mais de uma seção CNAE.

⁸ Recurso é a informação presente em um *Dataset*. Estas podem ser uma tripla RDF, um arquivo XML, CSV, uma planilha eletrônica, entre outros.

Tabela 5 - Distribuição dos *Data Sources*.

Setor Econômico	Data Sources
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

Mesmo não sendo a que apresenta o maior número de *data sources*, a seção econômica que apresenta a maior quantidade de dados é “Administração Pública, Defesa e Seguridade Social” com mais de 100 mil *datasets* nos 14 *data sources*. Iniciativas da administração pública, em especial da esfera federal, como o portal <http://www.data.gov/> do Governo dos Estados Unidos e o *data.gov.uk* do governo do Reino Unido englobam respectivamente algo em torno de 87 mil e 20 mil *datasets*. Destaca-se ainda, a existência de aproximadamente 20 mil *datasets* sobre “Indústrias Extrativas” e “Indústrias de Transformação”, e 18 mil sobre “Agricultura, Pecuária, Produção Florestal, Pesca e Aquicultura”. Os demais grupos da classificação econômica CNAE apresentam quantidades substancialmente menores de dados, podendo variar de algumas dezenas a algumas centenas. Contudo, mesmo em número reduzido, estas fontes contemplam grande quantidade de informações. Exemplo disso são fonte de dados de domínio variado como a *DBpedia*, que contempla o conteúdo de uma única fonte de informação, porém muito grande, a *Wikipedia*.

4.1.3 Análise de Vocabulários e Definição de Palavras-Chave

Os *data sources*, em muitos dos casos, apresentam vocabulários que descrevem suas informações. 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. A Figura 15 apresenta parte do vocabulário do setor energético presente em EuroVoc⁹, um tesouro multilinguagem especialmente construído para o tratamento da informação documental das instituições da União Europeia.

Figura 15 - Amostra de um vocabulário setorial.

6606 política energética →	
indústria energética	NT2 reserva estratégica
RT indústria do gás (6616)	NT1 balanço energético
RT indústria elétrica (6621)	RT estatística (1631)
RT indústria nuclear (6621)	NT2 aprovisionamento energético
NT1 combustível	RT abastecimento (2016)
RT recursos energéticos (5211)	NT2 consumo de energia
NT2 combustível fóssil	RT consumo (2026)
RT carvão (6611)	NT2 disponibilidades energéticas
RT gás natural (6616)	RT recursos energéticos (5211)
RT petróleo (6616)	NT2 oferta energética
NT3 turfa	NT2 preço da energia
NT2 gás	NT2 procura energética
RT aparelho a gás (6821)	NT2 produção de energia
RT butano (6616)	RT produção de petróleo (6616)
RT distribuição de gás (2846)	RT produto energético (6606)
RT gás natural (6616)	NT2 rendimento energético
RT propano (6616)	NT2 utilização da energia
NT3 jazigo de gás	RT consumo industrial (2026)
RT jazigo de petróleo (6616)	RT consumo interno (2026)
NT1 conversão de energia	NT1 Carta Europeia da Energia
RT energia não poluente (6626)	NT1 central desativada
NT1 energia poluente	RT central nuclear (6621)
RT energia não poluente (6626)	NT1 combustível de substituição
NT1 produto energético	RT bio-indústria (6411)
RT carburante (6616)	RT carburante (6616)
RT carvão (6611)	RT produto de substituição (6806)
RT energia elétrica (6621)	RT utilização alternativa de produtos agrícolas (5611)
RT gás natural (6616)	

Fonte: EuroVoc (<http://eurovoc.europa.eu/>).

⁹ <http://eurovoc.europa.eu/>

O exemplo acima reúne 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.

4.2 DESCRIÇÃO DA FERRAMENTA WEB

A ferramenta que implementa o modelo é baseada na arquitetura *Web* e tem como objetivo apoiar na recuperação das fontes de dados que serão usadas na etapa de coleta do ciclo de IC. Para facilitar sua recuperação os *data sources* foram cadastrados levando em conta a classificação de atividades econômicas CNAE constituindo uma abordagem *top-down* de busca na aplicação.

Quando um setor econômico é selecionado são apresentados os *data sources* correspondentes a esse setor, seus respectivos *datasets* e um conjunto de palavras-chave correspondentes.

Os *datasets* foram organizados de forma que possam ser recuperados diretamente utilizando-se palavras-chave provenientes dos vocabulários dos *data sources*. Outra maneira de se acessar os *datasets* é pela escolha de uma das fontes de dados. Dessa forma, ao selecionar a fonte de dados, os *datasets* de que ela é composta são apresentados. A Figura 16 apresenta a interface de busca da aplicação onde foi selecionado o setor econômico “Eletricidade e Gás”.

Após escolher o setor desejado, são apresentadas duas possibilidades de acesso à informação como apresentado na Figura 16. Na primeira, ao selecionar o setor, são apresentadas informações sobre as fontes de dados que contemplam o setor selecionado como mostra a Figura 17. Nesta interface são mostradas informações como o nome da fonte de dados, a URI que possibilita a acesso aos dados diretamente no repositório fonte, uma descrição resumida sobre a fonte de dados, o autor e a quantidade aproximada de recursos presentes.

Figura 16 - Interface de Busca da Aplicação.

Classificação da Web de Dados para a IC

Home Data Source Dataset Tipo Dados Relatórios

Buscar Data Sources

SETOR ECONÓMICO ELETRICIDADE E GÁS Busca Setor

Busca Dataset

Palavra Chave
Energia
Energia - Biocombustíveis
Energia - Biomassa
Energia - Carvão
Energia - Combustíveis
Energia - Consumo
Energia - Distribuição
Energia - Eficiência
Energia - Emissões
Energia - Eólica
Energia - Geoespacial
Energia - Geotermal
Energia - Geração
Energia - Gás
Energia - Hidroelétrica

1 - 15

Outra maneira de acessar a informação é utilizando as palavras-chave listadas após a escolha do setor. Estas palavras são termos provenientes dos dicionários de dados que foram cadastrados como metadados utilizados para acessar diretamente os conjuntos de dados que tratam especificamente de um assunto.

Figura 17 - Interface Data Sources.

Data Source	Uri	Descrição	Autor	Quantidade
ENAKTing PSI energy Dataset	http://energy.psi.enaktng.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	2,316,831
Enepedia	http://enipedia.tudeit.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	100,734
Linked dataspaces for Energy Intelligence (LEI)	http://dgsit.deri.ie/?q=node/29	The core enabling technology for our projects is the Linked dataspaces for Energy Intelligence (LEI). LEI serves as an independent layer placed above the existing systems layer that support the three approaches to interoperability (integrated, Unified, Federated) with linked data providing a common syntactic and access protocols. LEI's link at the information-level (data) not the infrastructure-level (system) by focusing more on the conceptual similarities (shared understanding) between information. This is achieved by following an entity-centric approach that focuses on the concepts that exist within the systems, for example, business entities like employees, products, customers, intellectual property, assets, etc. Entities within the dataspaces are enriched with data from multiple systems. This results in a cloud of interlinked resources that reflect virtual or actual entities with links to relevant knowledge and contextual information from across all the information systems that have exposed linked data. Sources may be added in an incremental manner to the cloud where they can be reused. Each entity within the cloud has a dereferenceable URI that returns data in a machine-readable format describing the resource identified.	DERI	1
Open EI	http://en.openei.org/wiki/Main_Page	Open Energy Info is a platform to connect the world's energy data. It is a linked open data platform bringing together energy information to provide improved analyses, unique visualizations, and real-time access to data. OpenEI follows guidelines set by the White House's Open Government Initiative, which is focused on transparency, collaboration, and participation. OpenEI strives to provide open access to this energy information, which will spur creativity and drive innovation in the energy sector. Mais de 2 mil datasets divididos em 6 conjuntos	OpenEI.org	16,384,698

Na Figura 17, o nome da fonte de dados é uma ligação para a interface onde estão cadastrados os conjuntos de dados que formam a fonte de dados. Os dados apresentados são resultantes do processo de seleção das fontes apresentado anteriormente. Os conjuntos resultantes dessa seleção foram tratados e posteriormente carregados para o banco de dados da aplicação. A Figura 18 apresenta o resultado deste trabalho.

Figura 18 - Interface Datasets.

Dataset	URI	Autor
Dados New Zealand	data.govt.nz	Governo Nova Zelândia
Energypedia	Energypedia.info	Energypedia
QDR.OpenEI	QDR.OpenEI.org	OpenEI
OpenEI	OpenEI.org	OpenEI
UK Open Data Portal	data.gov.uk	data.gov.uk
US Open Data Portal	data.gov	Governo EUA

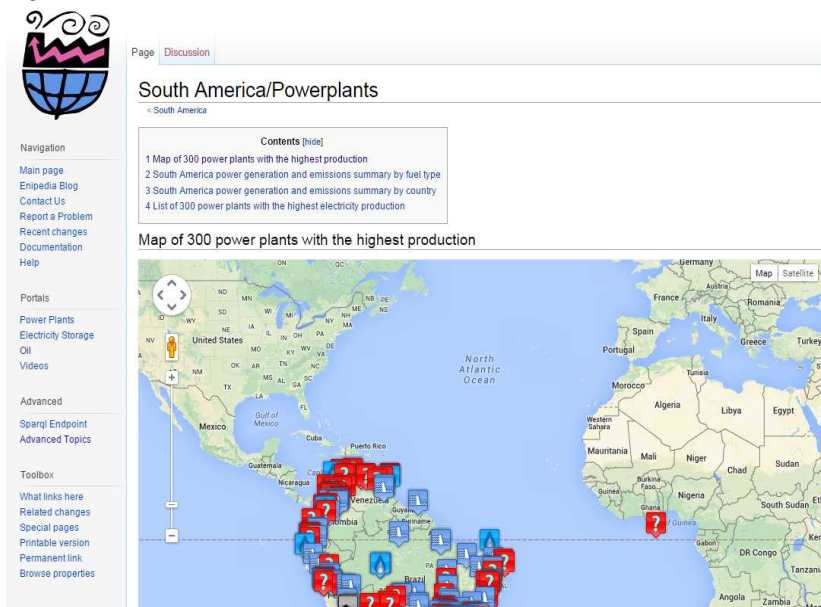
Na Figura 18 são apresentadas informações como nome do conjunto de dados, a URI, que acessa diretamente a informação presente neste conjunto de dados, e o autor. Ao clicar nesta URI o usuário é direcionado para a página inicial do conjunto de dados. A Figura 19 a seguir apresenta a página sobre plantas geradoras de eletricidade da América do Sul¹⁰ presente na Enipedia¹¹, uma fonte de dados sobre o setor de energia estruturada de maneira semelhante a Wikipedia.

Neste conjunto de dados, são apresentadas em um mapa as plantas geradoras de energia da América do Sul. Os ícones usados para apontar a localização no mapa também representam o tipo de fonte utilizada na geração de eletricidade, como por exemplo, hidrelétrica, gás, carvão, óleo ou nuclear.

¹⁰ http://enipedia.tudelft.nl/wiki/South_America/Powerplants

¹¹ http://enipedia.tudelft.nl/wiki/Main_Page

Figura 19 – Plantas Geradoras da América do Sul.



Ao selecionar um desses pontos são apresentadas informações sobre uma planta geradora específica. A Figura 20 apresenta as informações referentes a usina hidroelétrica de Itaipu¹², a maior da América do Sul e segunda maior desse tipo no mundo.

¹² http://enipedia.tudelft.nl/wiki/Itaipu_Powerplant

Figura 20 - Planta Geradora de Itaipu

The screenshot shows the Wikipedia page for 'Itaipu Powerplant'. The page is titled 'Itaipu Powerplant' and has a 'Page Discussion' tab. The main content area includes a 'Contents' table of contents, a 'General' section with information such as 'Operator: Itaipu Binacional', 'Year first built', and 'Other names', and a 'Data Links' section with various external database links. On the right side, there is a 'Maps' section with a Google Map showing the location of the powerplant on the Itaipu Dam, with a red pin and a scale bar.

Nesta interface são apresentadas 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*¹³, uma base colaborativa e de acesso livre com informações sobre energia em escala global. A Figura 21 apresenta a página sobre Itaipu presente no *Global Energy Observatory*.

¹³ <http://globalenergyobservatory.org/geoid/42658>

Figura 21 - Itaipu no Global Energy Observatory

The screenshot shows the Global Energy Observatory (GEO) website interface. At the top left is the GEO logo. A navigation bar contains links for Home, Partners, Help us build GEO, About Us, and Contact Us. Below the navigation bar are tabs for POWER PLANTS, FUELS AND RESOURCES, ENERGY TRANSMISSION, ENERGY CONSUMERS, and MAP DATA IN GEO. The main content area is titled 'ABSTRACT' and contains the following information:

ABSTRACT
 Itaipu Hydroelectric Power Plant Brazil-Paraguay is located at Itaipu, Paraná, Brazil. Location coordinates are Latitude=-25.4082, Longitude=-54.5898. This infrastructure is of TYPE Hydro Power Plant with a design capacity of 14000 MW(e). It has 20 unit(s). The first unit was commissioned in 1984 and the last in 2007. It is operated by Itaipu Binacional.

IDENTIFIERS FOR HYDRO

Name: Itaipu Hydroelectric Power Plant Brazil-Paraguay
 Status of Plant: Operating Fully
 Plant Efficiency and Impact: Please Select
 Plant Overall Rating: State-of-the-art Worth Duplicating Environmentally Responsible
 Country Assigned Identification Number:
 GEO Assigned Identification Number: 42658
 Select Currency for data in this page: Brazilian Real

LOCATION

Latitude: -25.4082 Longitude: -54.5898 Update Placemarks Restore Placemarks
 This location is: Exact Approximate Within 50km Unknown
 Draw Lines: Draw Polygons:
 A satellite map shows the location of the Itaipu Dam on the border between Brazil and Paraguay. The map includes a scale bar and a legend.

On the left side of the page, there is a 'LOGIN TO SUBMIT DATA' section with fields for Username and Password, and buttons for Login and Register. Below this is a 'QUICK ACCESS TOOLS' section with a link to 'Hydro in Brazil'. A sidebar on the left lists various hydroelectric power plants, including Angra 1, Angra 2, Angra 3, Angra 4, Angra 5, Angra 6, Angra 7, Angra 8, Angra 9, Angra 10, Angra 11, Angra 12, Angra 13, Angra 14, Angra 15, Angra 16, Angra 17, Angra 18, Angra 19, Angra 20, Angra 21, Angra 22, Angra 23, Angra 24, Angra 25, Angra 26, Angra 27, Angra 28, Angra 29, Angra 30, Angra 31, Angra 32, Angra 33, Angra 34, Angra 35, Angra 36, Angra 37, Angra 38, Angra 39, Angra 40, Angra 41, Angra 42, Angra 43, Angra 44, Angra 45, Angra 46, Angra 47, Angra 48, Angra 49, Angra 50, Angra 51, Angra 52, Angra 53, Angra 54, Angra 55, Angra 56, Angra 57, Angra 58, Angra 59, Angra 60, Angra 61, Angra 62, Angra 63, Angra 64, Angra 65, Angra 66, Angra 67, Angra 68, Angra 69, Angra 70, Angra 71, Angra 72, Angra 73, Angra 74, Angra 75, Angra 76, Angra 77, Angra 78, Angra 79, Angra 80, Angra 81, Angra 82, Angra 83, Angra 84, Angra 85, Angra 86, Angra 87, Angra 88, Angra 89, Angra 90, Angra 91, Angra 92, Angra 93, Angra 94, Angra 95, Angra 96, Angra 97, Angra 98, Angra 99, Angra 100.

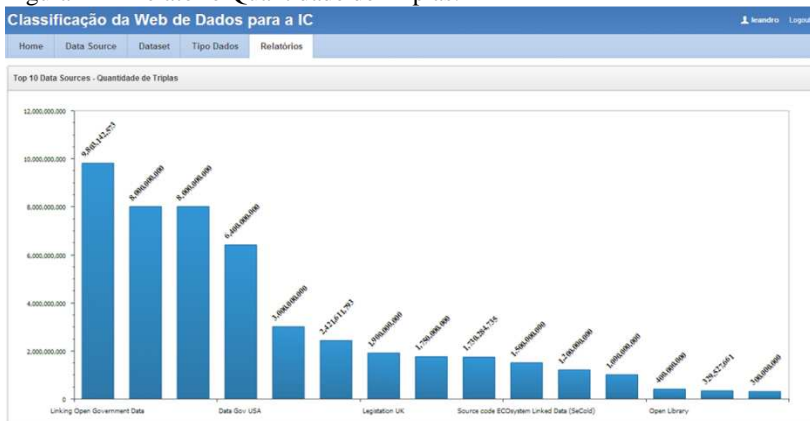
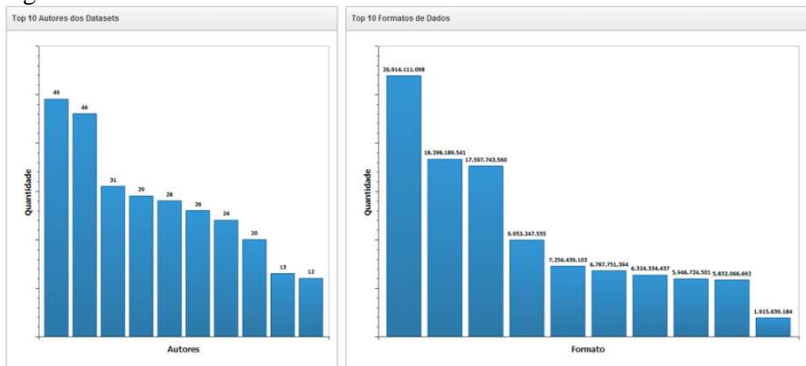
Em *data sources* que possuem milhares de conjuntos de dados, estes costumam ser organizados por assuntos. Dessa forma, para acessar o *dataset*, é necessário primeiramente escolher o assunto. Por exemplo, o *data source* “data.gov”¹⁴ do governo dos Estados Unidos possui 104 mil *datasets* divididos em 21 tópicos.

Como prova de conceito para este trabalho foram cadastrados somente os 21 *datasets* do primeiro nível. Porém para o setor de “Eletricidade e Gás” que foi o escolhido para exemplificar o processo de coleta de dados do processo de IC, todos os *datasets* foram cadastrados.

Por fim na aba “Relatórios” da aplicação representada na Figura 22 e na Figura 23, são apresentados gráficos com estatísticas como: as fontes de dados com maior quantidade de informação, os autores que mais publicaram conjuntos de dados e os formatos de dados mais encontrados nestas fontes de dados.

¹⁴ <http://www.data.gov/>

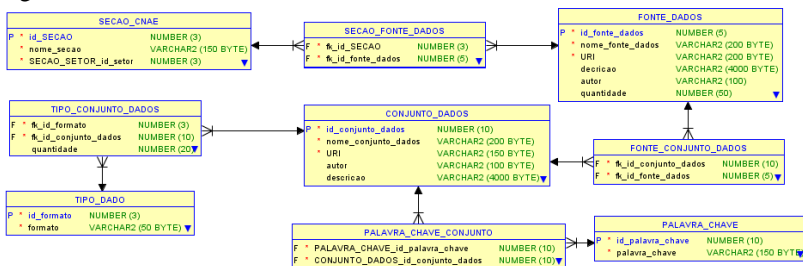
Figura 22 – Relatório Quantidade de Triplas.

Figura 23 - Relatórios Autores dos *datasets* e Formatos de dados.

4.2.1 Criação do Repositório de Dados

As informações foram armazenadas em um modelo de dados relacional. A Figura 24 apresenta este modelo.

Figura 24 - Modelo Relacional.



Neste modelo, os retângulos representam as entidades e as setas os relacionamentos entre elas. Como as entidades apresentam características de cardinalidade de muitos para muitos foi necessária a criação de tabelas associativas. Essas tabelas têm como finalidade eliminar o armazenamento redundante de informação, mantendo somente dados relativos às entidades representadas.

4.3 APLICAÇÃO DO MODELO NO SETOR DE ELETRICIDADE E GÁS

Para a verificação do modelo proposto foi escolhido o setor de Eletricidade e Gás. A escolha se deve pela quantidade e qualidade dos *data sources* presentes na *Web de Dados*, por sua importância social e econômica, e ainda, pelo fato deste setor demandar informações de outras áreas do conhecimento presentes na nuvem de dados como: sustentabilidade, dados geográficos, demográficos, de atividade empresarial entre outros.

Estão presentes na *Web de Dados* seis *data sources* que abordam diretamente o setor elétrico. Essas fontes são compostas por aproximadamente dois mil *datasets* e aproximadamente 23,2 milhões de registros. As informações provenientes da nuvem de dados devem ser capazes de subsidiar estratégias e identificar oportunidades dentro do setor elétrico, de gás e também de outras fontes energéticas. A Figura 25 apresenta os *data sources* do setor elétrico.

Figura 25 - Data Sources Setor Elétrico.

Classificação da Web de Dados para a IC				
Home	Data Source	Dataset	Tipo Dados	Relatórios
<input type="text" value="Q_"/> <input type="button" value="Go"/> <input type="button" value="Actions"/>				
Data Source	URI	Descricao	Autor	Quantidade
EnkTing PSI energy Dataset	http://energy.psi.enatting.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
Enepedia	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, TSM, 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
Linked dataspac for Energy Intelligence (LEI)	http://dgit.teri.ie/?q=node/29	The core enabling technology for our projects is the Linked DataSpace for Energy Intelligence (LEI). LEI serves as an independent layer placed above the existing systems layer that support the three approaches to interoperability (Integrated, Unified, Federated) with linked data providing a common syntactic and access protocols. LEI is link at the information-level (data) not the infrastructure-level (system) by focusing more on the conceptual similarities (shared understanding) between information. This is achieved by following an entity-centric approach that focuses on the concepts that exist within the systems, for example, business entities like employees, products, customers, intellectual property, assets, etc. Entities within the dataspac are enriched with data from multiple systems. This results in a cloud of interlinked resources that reflect virtual or actual entities with links to relevant knowledge and contextual information from across all the information systems that have exposed linked data. Sources may be added in an incremental manner to the cloud where they can be reused. Each entity within the cloud has a dereferenceable URI that returns data in a machine-readable format describing the resource identified.	DERI	1
Open EI	http://en.openei.org/wiki/Main_Page	Open Energy Info is a platform to connect the world's energy data. It is a linked open data platform bringing together energy information to provide improved analyses, unique visualizations, and real-time access to data. OpenEI follows guidelines set by the White House's Open Government Initiative, which is focused on transparency, collaboration, and participation. OpenEI strives to provide open access to this energy information, which will spur creativity and drive innovation in the energy sector. Mais de 2 mil datasets divididos em 6 conjuntos	OpenEI.org	1638468
Reegle	http://data.reegle.info/	Comprehensive set of linked clean energy data including: policy and regulatory country profiles, key stakeholders (organisation profiles), project outcome documents and a thesaurus (SKOS format) on renewables, energy efficiency and climate change for public re-use.	Florian Bauer	330409

O modelo do 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.

Além disso, investimentos no setor requerem análises detalhadas de viabilidade técnica, econômico-financeira e ambiental antes de sua efetiva implantação. Para progredir e tornar-se competitiva, a empresa deve desenvolver políticas de governança corporativa, cumprir seus planos de negócios, seus investimentos e planejar a demanda energética futura.

Por exemplo, o setor elétrico pode ganhar força com o uso eficiente das fontes de energia disponíveis, com o consumo racional e o investimento em fontes alternativas, como a geotérmica, eólica e a solar.

4.3.1 Identificação dos Requisitos de IC

Através de sua experiência, Prescott e Miller (2002, p.279) verificaram que uma empresa tem suas necessidades de informação enquadradas em uma das três categorias funcionais: decisões e ações estratégicas, tópicos de alerta antecipado e descrição dos principais atores do seu mercado.

Baseados nesses conceitos anteriores 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;

4.3.2 Coleta dos Dados

Esta seção apresenta as fontes de informação presentes na *Web* de Dados 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 6 apresenta um apanhado sobre os assuntos e quantidade de *datasets* que contemplam cada uma das necessidades identificadas.

Tabela 6 – 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

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, por exemplo, a demanda e o consumo anual eletricidade¹⁵, gás¹⁶ e carvão¹⁷ 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 como os Estados Unidos. Os dados são disponibilizados em formatos como o CSV ou XLS geralmente por órgãos como o *Energy Information Administration* (EIA).

Ainda sobre a primeira necessidade de IC, no que corresponde a geração de energia temos *datasets* que tratam de geração de energia hidroelétrica e dos aspectos legais da concessão e da geração de eletricidade a partir desta fonte disponibilizado pela Agência Nacional de Energia Elétrica (ANEEL).

São tratados ainda formas de geração que utilizam combustíveis fósseis, como petróleo e carvão divididos por país e disponibilizados segundo os princípios de *Linked Data*, e de fontes renováveis como a eólica presente em *The Wind Power*¹⁸ onde dados como o número de fazendas eólicas e a capacidade de geração em *megawatt* de 102 países é listada. Estão presentes ainda *datasets* sobre a capacidade de geração de plantas ao redor do Mundo, na Europa, na América do Sul e no Brasil, divididas pela categoria energética gerada por essas plantas.

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*¹⁹, 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 Elétrica²⁰ que disponibiliza a média anual do potencial eólico a 50m em território brasileiro.

¹⁵ <http://en.openei.org/datasets/node/877>

¹⁶ <http://en.openei.org/datasets/node/880>

¹⁷ <http://en.openei.org/datasets/node/881>

¹⁸ http://www.thewindpower.net/country_list_fr.php

¹⁹ http://en.openei.org/wiki/Global_Renewable_Resource_Potential

²⁰ <http://en.openei.org/datasets/node/608>

Como fontes energéticas alternativas temos o hidrogênio²¹, 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 energética²², indicadores, estatísticas e tendências do setor elétrico, além de iniciativas como *smart grids*²³ 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.

²¹ <http://en.openei.org/wiki/Gateway:Hydrogen>

²² <http://en.openei.org/datasets/node/468>

²³ <http://en.openei.org/datasets/node/928>

5 CONCLUSÕES E TRABALHOS FUTUROS

Neste capítulo é feita uma conexão entre os objetivos da dissertação e os resultados alcançados. São apresentadas ainda as sugestões de estudos futuros e as limitações do trabalho.

5.1 CONCLUSÕES

O problema apresentado no início deste trabalho trata do uso da *Web* de Dados na etapa de coleta do processo de Inteligência Competitiva Setorial. Para respondê-lo, foi desenvolvido um modelo capaz de identificar, classificar e representar a informação usada com termos de busca que posteriormente são usados para auxiliar a coleta.

O uso desta fonte de dados em um processo de IC é justificado pelo crescente número de indivíduos e organizações que têm adotado esta plataforma para publicar seus dados, não apenas colocando-os na *Web*, mas usando a *Web* de Dados para interligá-los a outras fontes e assim consolidá-los.

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.

Apesar das dificuldades iniciais em encontrar literatura que aborde conjuntamente os temas propostos neste trabalho, a segurança para o andamento da pesquisa veio do embasamento teórico individual de cada um dos termos, e da definição segundo procedimentos metodológicos do modelo apresentado.

O objetivo geral da dissertação, estabelecer uma abordagem baseada em semântica para o uso *Web* de Dados na etapa de coleta de informação do ciclo de IC, foi alcançado com o desenvolvimento do modelo conceitual e da aplicação no setor de eletricidade e gás descrita.

Essa aplicação tem como objetivo catalogar a informação de maneira que facilite a recuperação das fontes de dados no ciclo de IC. Para isso, os *data sources* foram cadastrados levando em conta a classificação de atividades econômicas CNAE. Ou seja, ao selecionar um setor econômico, são apresentadas informações relacionadas exclusivamente a este setor, isso evita o retorno de informação sem relevância.

Quanto ao propósito de atender aos objetivos específicos propostos neste trabalho, conclui-se que:

- *Apresentar uma classificação de setores econômicos para catalogação da Web de Dados.* Foi alcançado com procedimentos de seleção das fontes relevantes contidas na *Web de Dados* (Seção 4.1.1) e a definição dos setores econômicos segundo os procedimentos da Seção 4.1.2.
- *Realizar a catalogação dos data sources da Web de Dados segundo a classificação definida.* Este objetivo foi alcançado com a catalogação dos *data sources* em consonância com os termos definidos e com o desenvolvimento da aplicação onde estas informações foram catalogadas (Seção 4.2).
- *Propor um modelo para uso de informação da Web de Dados na etapa de coleta do ciclo de IC.* Este objetivo foi contemplado com o desenvolvimento da aplicação descrita na Seção 4.2 e com o repositório de dados onde foi armazenada a informação Seção 4.2.1.
- *Verificar o modelo proposto em um setor econômico.* Foi alcançado, primeiramente, com a definição do setor de Eletricidade e Gás usado para verificação do emprego da *Web de Dados* (Seção 4.3). Em seguida foram desenvolvidas as tarefas de identificação de requisitos de IC para o setor (Seção 4.3.1) e por fim a coleta dos dados (Seção 4.3.2).

Como contribuição acadêmica esta dissertação traz uma revisão teórica sobre o tema Inteligência Competitiva, em especial da etapa de coleta, sobre a Engenharia do Conhecimento, e sobre a evolução da *Web* até culminar na *Web de Dados* e nos princípios de *Linked Data*.

Em seguida estes conceitos, foram combinados para formar um modelo usado para catalogação e recuperação de informação. Com o auxílio de ferramentas computacionais, seguindo um procedimento metodológico, este modelo foi transformado em uma ferramenta de Engenharia do Conhecimento.

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 diferentes setores ou mesmos outros processos que demandam informações externas a organização.

As principais dificuldades encontradas do desenvolvimento deste trabalho referem-se a falta de referências bibliográficas que abordem conjuntamente todo o escopo da questão levantada e ao cadastro dos *data sources* na aplicação.

Para esta prova de conceito, a tarefa de cadastro dos *datasets* foi feita manualmente, o que demandaria quantidade significativa de tempo para que todos fossem cadastrados. Por este motivo, somente os *datasets* do setor escolhido para a prova de conceito foram cadastrados em sua totalidade. Contudo, estas dificuldades não acarretaram prejuízos ao resultado final do estudo e em trabalhos futuros esse processo pode ser automatizado.

Sendo este um trabalho de mestrado, conclui-se que a pesquisa alcançou seu objetivo em investigar a possibilidade de uso da *Web* de Dados na etapa de coleta do ciclo de IC ao definir uma metodologia para elaboração de trabalhos seguindo as recomendações da pesquisa exploratória.

As principais limitações metodológicas deste trabalho estão relacionadas ao escopo que abrangeu apenas as ações da Inteligência Competitiva relativas a coleta de informação. 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 a informação presente na *Web* de Dados pode ser utilizada com ganhos significativos para o processo de coleta de IC. Como esta informação 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, confecção de relatórios ou mesmo de aplicações que podem auxiliar não somente na coleta, mas também na etapa de análise do ciclo de IC.

Além disso, o uso de uma ferramenta como a proposta nesta dissertação, facilita o trabalho de recuperação de informações já que o analista responsável pela coleta não tem necessidade de varrer a *Web* de Dados em busca de informação, já que esta se encontra concentrada em um único lugar e previamente dividida por setor econômico e assuntos.

5.2 TRABALHOS FUTUROS

Os trabalhos desenvolvidos e as conclusões obtidas culminaram em recomendações e atividades que podem ter continuidade em trabalhos futuros. Dentre estas atividades temos a automatização do processo de catalogação dos conjuntos contemplando, além da totalidade dos *datasets* para os demais setores econômicos, os dados presentes na versão de 2014 do diagrama da *LOD Cloud*.

Pretende-se ainda, disponibilizar a aplicação em um servidor *Web* para permitir que a catalogação da base possa ser realizada de forma compartilhada por múltiplos usuários da ferramenta.

É importante a aplicação prática do modelo proposto em um programa de IC para validar a consistência da informação recuperada em um processo com demandas reais, sempre que possível mensurando o resultado desta aplicação. Além disso, realizar verificações com especialistas do setor para a análise e validação das fontes selecionadas na etapa de coleta.

Por fim, propor estudos e indicadores capazes de analisar os ganhos decorrentes do uso de fontes de informação semanticamente estruturadas. Primeiramente, na etapa de coleta em pontos como a diminuição do tempo gasto em tarefas de identificação e de recuperação de conteúdo. E por fim, na etapa de análise, em quesitos como, por exemplo, quantidade de informação necessária para responder a demandas de IC de forma satisfatória, e o tempo gasto para análise da informação.

REFERÊNCIAS

- ABRAIC. Associação Brasileira dos Analistas de Inteligência Competitiva. 2013. Disponível em: < <http://www.abraic.org.br/> >. Acesso em: 02/09/2013.
- ABREU, A. F.; COBRAL, E.; OGLIARI, A. Gestão integrada da inovação: estratégia, organização e desenvolvimento de produtos. In: ATLAS (Ed.). Atlas. São Paulo: Atlas, v.1, 2008. p.269. ISBN 978-85-224-4976-7.
- ADIDA, B.; BIRBECK, M. Rdfa primer - bridging the human and data webs. W3C recommendation 2008.
- ALEXANDER, K. Rdf in json. Proceedings of the 4th Workshop on Scripting for the Semantic Web. W3C 2008.
- 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.
- ALLEMANG, D. Semantic Web and the Linked Data Enterprise. In: WOOD, D. (Ed.). Linking Enterprise Data: Springer US, 2010. cap. 1, p.3-23. ISBN 978-1-4419-7664-2.
- ALMEIDA, M. B.; BAX, M. P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. Ciência da Informação, v. 32, p. 7-20, 2003. ISSN 0100-1965. Disponível em: < http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652003000300002&nrm=iso >.
- ANGELE, J. *et al.* Developing Knowledge-Based Systems with MIKE. Automated Software Engg., v. 5, n. 4, p. 389-418, 1998. ISSN 0928-8910.
- ANICA-POPA, I.; CUCUI, G. A Framework for Enhancing Competitive Intelligence Capabilities using Decision Support System based on Web Mining Techniques. Int. J. of Computers, Communications & Control, v. IV, n. 4, p. 9, 2009. ISSN 1841-9836. Disponível em: < <http://www.journal.univagora.ro/download/pdf/378.pdf> >. Acesso em: 02/09/2013.
- AUER, S.; LEHMANN, J.; HELLMANN, S. LinkedGeoData: Adding a Spatial Dimension to the Web of Data. Proceedings of the 8th International Semantic Web Conference. Chantilly, VA: Springer-Verlag: 731-746 p. 2009.
- BAO, J. *et al.* OWL 2 Web Ontology Language Quick Reference Guide (Second Edition). W3C Recommendation 11 December 2012. W3C 2012.

BECKETT, D. RDF/XML Syntax Specification (Revised) W3C Recommendation. W3c 2004.

BECKETT, D.; BERNERS-LEE, T. Turtle - Terse RDF Triple Language. W3C Team Submission: W3C 2011.

BECKETT, D. *et al.* Terse RDF Triple Language. W3C Candidate Recommendation: W3C 2013.

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

BERGERON, P.; HILLER, C. A. Competitive intelligence. *Annual Review of Information Science and Technology*, v. 36, n. 1, p. 353-390, 2002. ISSN 1550-8382. Disponível em: < <http://dx.doi.org/10.1002/aris.1440360109> >.

BERKOWITZ, B. D.; GOODMAN, A. E. *Strategic Intelligence for American National Security*. Princeton University Press, 1991. ISBN 9780691023397. Disponível em: < <http://books.google.com.br/books?id=sy2XMFQwe3EC> >. Acesso em: 03/09/2013.

BERNERS-LEE, T. Semantic Web on XML 2000/12/06. In: W3C, XML 2000 2000, Washington DC.

BERNERS-LEE, T. Linked Data: Design Issues. Online, 29/03/2013 2006. Disponível em: < <http://www.w3.org/DesignIssues/LinkedData.html> >. Acesso em: 02/04/2013.

BERNERS-LEE, T. *Giant Global Graph* 2007.

BERNERS-LEE, T. Pre-W3C Web and Internet Background. World Wide Web Consortium 2011.

BERNERS-LEE, T.; FIELDING, R.; MASINTER, L. Uniform Resource Identifier (URI): Generic Syntax. RFC Editor. 2005. (2070-1721)

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web. *Scientific American*, v. 284, n. 5, p. 6, 2001. ISSN 00368733.

BERNHARDT, D. Competitive intelligence: how to acquire and use corporate intelligence and counter-intelligence. FT Prentice Hall, 2003. Disponível em: < <http://books.google.com.br/books?id=Z7ZaAAAAAYAAJ> >. Acesso em: 04/03/2013.

BETANZOS, A. A. *Ingeniería del conocimiento: aspectos metodológicos*. Pearson-Prentice Hall, 2004. ISBN 9788420541921. Disponível em: < <http://books.google.es/books?id=SrczAQAACAj> >.

BIRON, P. V.; MALHOTRA, A. XML Schema Part 2: Datatypes Second Edition. W3C Recommendation. W3C 2004.

BIZER, C.; CYGANIAK, R.; GAUß, T. The RDF Book Mashup: From Web APIs to a Web of Data. Proceedings of the 3rd Workshop on Scripting for the Semantic Web (SFSW2007). 2007.

BIZER, C.; HEATH, T.; BERNERS-LEE, T. Linked Data - The Story So Far. International Journal on Semantic Web and Information Systems (IJSWIS), v. 5, n. 3, p. 1-22, 33// 2009. ISSN 1552-6283. Disponível em: <<http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>>. Acesso em: 01/04/2013.

BLOMQUIST, E. OntoCase - A Pattern-Based Ontology Construction Approach. In: MEERSMAN, R. e TARI, Z. (Ed.). On the Move to Meaningful Internet Systems 2007: CoopIS, DOA, ODBASE, GADA, and IS: Springer Berlin Heidelberg, v.4803, 2007. cap. 64, p.971-988. (Lecture Notes in Computer Science). ISBN 978-3-540-76846-3.

BONCELLA, R. J. COMPETITIVE INTELLIGENCE AND THE WEB. Communications of the Association for Information Systems, v. 12, p. 14, 2003. Disponível em: < <http://www.washburn.edu/faculty/boncella/COMPETITIVE-INTELLIGENCE.pdf>>. Acesso em: 09/09/2013.

BONNER, P. The Semantic Web. PC Magazine, v. 21, n. 13, p. 2, July 2002 2002.

BORST, W. N. Construction of Engineering Ontologies for Knowledge Sharing and Reuse. 1997. 231, Enschede.

BOSE, R. Competitive intelligence process and tools for intelligence analysis. Industrial Management & Data Systems, v. 108, n. 4, p. 19, 2008. ISSN 0263-5577. Disponível em: < <http://dx.doi.org/10.1108/02635570810868362> >. Acesso em: 06/09/2013.

BOUTHILLIER, F.; SHEARER, K. Assessing Competitive Intelligence Software: A Guide to Evaluating CI Technology. Information Today, 2003. ISBN 9781573871730. Disponível em: < <http://books.google.com.br/books?id=dj4IEWeKCAAC> >.

BRASIL. Lei de Acesso a Informação. 12.527 FEDERAL, G.: Casa Civil. 12.527 2012.

BRAY, T. *et al.* Extensible Markup Language (XML) 1.1 (Second Edition): W3C 2006.

BREUKER, J. *et al.* The Flood, the Channels and the Dykes: Managing Legal Information in a Globalized and Digital World. Proceedings of the 2009 conference

on Law, Ontologies and the Semantic Web: Channelling the Legal Information Flood: IOS Press: 3-18 p. 2009.

BRICKLEY, D.; GUHA, R. V. RDF Vocabulary Description Language 1.0: RDF Schema W3C Recommendation W3C 2004.

BROUARD, F. Development of an Expert System on Environmental Scanning Practices in SME: Tools as a Research Program. *Journal of Competitive Intelligence and Management*, v. 3, n. 4, p. 37-58, 2006. ISSN 1703-5147. Disponível em: < <http://www.scip.org/files/JCIM/3.4.brouard.pdf> >. Acesso em: 30/03/2013.

CAILLIAU, R. A Little History of the World Wide Web. World Wide Web Consortium. . W3C 2011.

CALOF, J. L. A.; WRIGHT, S. B. Competitive intelligence: A practitioner, academic and inter-disciplinary perspective. *European Journal of Marketing*, v. 42, n. 7-8, p. 717-730, 2008. Disponível em: < <http://www.scopus.com/inward/record.url?eid=2-s2.0-48249139412&partnerID=40&md5=9ba1e19188659811670288921760ae23> >. Acesso em: 30/03/2013.

CARDOSO JR., W. F. A Inteligência Competitiva Aplicada nas Organizações do Conhecimento como Modelo de Inteligência Empresarial Estratégica para Implementação e Gestão de Novos Negócios. Florianópolis Tese (Doutorado em Engenharia de Produção) - Universidade Federal de Santa Catarina., 2003.

CHANDRASEKARAN, B. Generic Tasks in Knowledge-Based Reasoning: High-Level Building Blocks for Expert System Design. *IEEE Expert*, v. 1, n. 3, p. 23-30, 1986. ISSN 0885-9000.

CHAUDHURI, S.; DAYAL, U. An overview of data warehousing and OLAP technology. *SIGMOD Rec.*, v. 26, n. 1, p. 65-74, 1997. ISSN 0163-5808.

CLANCEY, W. J. The Knowledge Level Reinterpreted: Modeling How Systems Interact. *Mach. Learn.*, v. 4, n. 3-4, p. 285-291, 1989. ISSN 0885-6125.

COLBY, W. E. Competitive Intelligence in the New World of the 1990s. *Global Perspectives on Competitive Intelligence*, p. 5, 1993.

COMBS, R. E.; MOORHEAD, J. D. *The Competitive Intelligence: Handbook*. Scarecrow Press, Incorporated, 1992. ISBN 9780810826069. Disponível em: < <http://books.google.com.br/books?id=ajpmQgAACAAJ> >. Acesso em: 01/09/2013.

COOLEY, R.; MOBASHER, B.; SRIVASTAVA, J. Web mining: information and pattern discovery on the World Wide Web. *Tools with Artificial Intelligence*, 1997.

Proceedings., Ninth IEEE International Conference on, 1997, 3-8 Nov 1997. p.558-567.

CORCHO, O.; FERNÁNDEZ-LÓPEZ, M.; GÓMEZ-PÉREZ, A. Methodologies, tools and languages for building ontologies: where is their meeting point? *Data Knowl. Eng.*, v. 46, n. 1, p. 41-64, 2003. ISSN 0169-023X. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.100.1223&rep=rep1&type=pdf>>. Acesso em: 07/10/2013.

CORREIA, A. M. R.; SANTOS, M. C. F. Fundamentos teóricos da competitive intelligence como vantagem competitiva. 5ª Conferência Ibérica de Sistemas e Tecnologias de Informação – CISTI 2010. Santiago de Compostela, Espanha: Comunicação. 2: 329-334 p. 2010.

CORRENDO, G. *et al.* SPARQL query rewriting for implementing data integration over linked data. *Proceedings of the 2010 EDBT/ICDT Workshops*. Lausanne, Switzerland: ACM: 1-11 p. 2010.

CORY, H. D. Can competitive intelligence lead to a sustainable competitive advantage? *Competitive Intelligence Review*, v. 7, n. 3, p. 45-55, 1996. ISSN 1520-6386. Disponível em: <<http://dx.doi.org/10.1002/cir.3880070309>>.

COSTA, A. H. Arquitetura básica de um sistema de inteligência competitiva: uma aplicação para o setor de operadoras de telefonia móvel. Florianópolis: Dissertação (Mestrado em Engenharia de Produção) - Universidade Federal de Santa Catarina, Santa Catarina, 2002.

CROCKFORD, D. The application/json Media Type for JavaScript Object Notation (JSON). The Internet Society 2006.

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: 02/04/2012.

CYGANIAK, R. *et al.* Vocabulary of Interlinked Datasets (VoID). DERI Vocabularies 2001.

DAL PIZZOL, L.; TODESCO, J. L. Inteligência Competitiva: Um Paralelo Entre a Teoria e a Prática Organizacional IX Convibra Administração – Congresso Virtual Brasileiro de Administração CONVIBRA. Brasil: Convibra. 2012: 13 p. 2012.

DALKIR, K. *Knowledge Management in Theory and Practice*. Taylor & Francis, 2013. ISBN 9781136389740. Disponível em: <<http://books.google.com.br/books?id=QcC61an8glkC>>.

DEAN, M.; SCHREIBER, G. OWL Web Ontology Language Reference. W3C Recommendation. W3C 2004.

DESLAURIERS, J.-P. Recherche qualitative. Guide pratique. Montreal: McGraw-Hill, 1991.

DINUCCI, D. Fragmented Future. Print Magazine 2009.

DISHMAN, P.; CALOF, J. Competitive intelligence: a multiphasic precedent to marketing strategy. *European Journal of Marketing*, v. 42, n. 7/8, p. 766-785, 30/03/2013 2008. ISSN 00419907. Disponível em:

<<http://www.veilleeconomique.net/>>. Acesso em: 30/03/2013.

DUBEY, R. An Assiduous Study on Linkage between Competitive Intelligence and Field Leadership for Organizational Development. *International Journal of Information, Business and Management*, v. 5, n. 2, p. 16, 2013. ISSN 2076-9202 194. Disponível em:

<http://www.ijibm.elitehall.com/IJIBM_Vol5No2_May2013.pdf#page=195>.

Acesso em: 30/04/2013.

DUGAL, M. CI product line: A tool for enhancing user acceptance of CI. *Competitive Intelligence Review*, v. 9, n. 2, p. 17-25, 1998. ISSN 1520-6386. Disponível em: <[http://dx.doi.org/10.1002/\(SICI\)1520-6386\(199804/06\)9:2<17::AID-CIR5>3.0.CO;2-0](http://dx.doi.org/10.1002/(SICI)1520-6386(199804/06)9:2<17::AID-CIR5>3.0.CO;2-0)>.

DUHAN, N.; SHARMA, A. K.; BHATIA, K. K. Page Ranking Algorithms: A Survey. *Advance Computing Conference, 2009. IACC 2009. IEEE International, 2009, 6-7 March 2009*. p.1530-1537.

EAVES, D. The Three Laws of Open Government Data. *eaves.ca* 2009.

EELLS, R. S. F.; NEHEMKIS, P. R. Corporate intelligence and espionage: a blueprint for executive decision making. Macmillan, 1984. ISBN 9780029092408. Disponível em: <<http://books.google.com.br/books?id=H0TgsiU5qM0C>>. Acesso em: 01/09/2013.

EGC. EGC. Interação das Áreas na Busca do Objeto de Pesquisa do Programa. EGC - Programa de Pós-Graduação Engenharia e Gestão do Conhecimento. 2013. Disponível em: <<http://www.egc.ufsc.br/index.php/pt/pesquisas/linhas-de-pesquisa>>. Acesso em: 15/04/2013.

FAN, W.; GORDON, M. D.; PATHAK, P. An integrated two-stage model for intelligent information routing. *Decision Support Systems*, v. 42, n. 1, p. 362-374, 10// 2006. ISSN 0167-9236. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0167923605000084>>.

FAN, W. *et al.* Tapping the power of text mining. *Commun. ACM*, v. 49, n. 9, p. 76-82, 2006. ISSN 0001-0782.

FIELDING, R. *et al.* Hypertext Transfer Protocol -- HTTP/1.1. W3C/MIT 1999.

FILIFE, P. P. Representação de Conteúdos a Linguagem XML. Instituto Superior de Engenharia de Lisboa Departamento de Engenharia da Electrónica e das Comunicações SECCÃO DE ENGENHARIA DE SISTEMAS 2001.

FLEISHER, C. S.; BENSOUSSAN, B. E. Strategic and Competitive Analysis: Methods and Techniques for Analyzing Business Competition. Pearson Prentice Hall, 2003. ISBN 9780130888525. Disponível em: < <http://books.google.com.br/books?id=ZP2NQgAACAAJ> >.

FLEISHER, C.S.; BENSOUSSAN, B.E. Business and Competitive Analysis: Effective Application of New and Classic Methods. Times Press, 2007. ISBN 9780131873667

FLEISHER, C. S.; BENSOUSSAN, B. Business and competitive analysis - what? so what? now what?. Paper presented at Society of Competitive Intelligence Professionals Annual Conference, San Diego, California, 2008.

FLEISHER, C.; WRIGHT, S.; TINDALE, R. Bibliography and Assessment of Key Competitive Intelligence Scholarship: Part 4 (2003-2006). Journal of Competitive Intelligence and Management, Alexandria, v. 4, n. 1, 2007.

FONSECA, J. J. S. Metodologia da pesquisa científica. Fortaleza:: UEC 2002.

FRANTZICH, S. E.; SULLIVAN, J. The C-span Revolution. University of Oklahoma Press, 1996. ISBN 9780806128702. Disponível em: < http://books.google.com.br/books?id=8aWr_ORJJ_gC >.

FULD, L. M. The New Competitor Intelligence: The Complete Resource for Finding, Analyzing, and Using Information about Your Competitors. Wiley, 1995. ISBN 9780471585091. Disponível em: < http://books.google.com.br/books?id=W_V7QgAACAAJ >. Acesso em: 29/08/2013.

FULD, L. M. "Intelligence Two Centuries Later". Competitive Intelligence Magazine, v. 5, n. 6, p. 2, 2002.

FULD, L. M. The Secret Language of Competitive Intelligence: How to See Through & Stay Ahead of Business Disruptions, Distortions, Rumors & Smoke Screens. DOG EAR PUB LLC, 2010. ISBN 9781608445530. Disponível em: < http://books.google.com.br/books?id=E_-B8G0NgroC >.

GARGOURI, F.; JAZIRI, W. Ontology Theory, Management, and Design: Advanced Tools and Models. Information Science Reference, 2010. ISBN 9781615208609. Disponível em: < <http://books.google.com.br/books?id=8BDcycKldwC> >.

GERBER, A.; MERWE, A.; BARNARD, A. A Functional Semantic Web Architecture. In: BECHHOFFER, S.; HAUSWIRTH, M., *et al* (Ed.). The Semantic Web: Research and Applications: Springer Berlin Heidelberg, v.5021, 2008. cap. 22, p.273-287. (Lecture Notes in Computer Science). ISBN 978-3-540-68233-2.

GIL, A. C. Como elaborar projetos de pesquisa. São Paulo: Atlas, 2010. ISBN 9788522458233. Disponível em: <
<http://books.google.com.br/books?id=HSGHRAAACAAJ> >.

GILAD, B.; GILAD, T. The Business Intelligence System: A New Tool for Competitive Advantage. AMACOM, American Management Association, 1988. ISBN 9780814459294. Disponível em: <
<http://books.google.com.br/books?id=qFNnQgAACAAJ> >.

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

GOLBREICH, C.; WALLACE, E. K. OWL 2 Web Ontology Language New Features and Rationale (Second Edition). W3C Recommendation 11 December 2012. W3C 2012.

GOMES, E.; BRAGA, F. Inteligência competitiva: como transformar informação em um negócio lucrativo. Elsevier, 2004. ISBN 9788535214505. Disponível em: <
<http://books.google.com.br/books?id=R1fXAQAACAAJ> >.

GOMEZ-PEREZ, A.; CORCHO, O. Ontology languages for the Semantic Web. Intelligent Systems, IEEE, v. 17, n. 1, p. 54-60, 2002. ISSN 1541-1672.

GORE, S.; PITALE, R. Web Mining: An Approach towards Information Retrieval From Web with Cloud Mining. International Journal Of Computer Science And Applications, v. 6, n. 2, 2013. ISSN 0974-1011. Disponível em: <
<http://www.researchpublications.org/IJCSA/NCAICN-13/162.pdf> >. Acesso em: 09/09/2013.

GRANT, J.; BECKETT, D. RDF Test Cases. W3C Recommendation W3C 2004.

GRAY, J. *et al*. Data Cube: A Relational Aggregation Operator Generalizing Group-By, Cross-Tab, and Sub-Totals n Data Mining and Knowledge Discovery, v. 1, n. 1, p. 29-53, 1997. Disponível em: <
<http://arxiv.org/ftp/cs/papers/0701/0701155.pdf> >. Acesso em: 10/10/2013.

GRUBER, T. R. A translation approach to portable ontology specifications. Knowledge Acquisition, v. 5, n. 2, p. 199-220, 6// 1993. ISSN 1042-8143. Disponível em: <
<http://www.sciencedirect.com/science/article/pii/S1042814383710083> >.

GUARINO, N. Formal Ontology in Information Systems. Proceedings of FOIS'98 Trento, Italy, 6-8 Amsterdam: IOS Press: 13 p. 1998.

GUEDES, V. L. S. A BIBLIOMETRIA E A GESTÃO DA INFORMAÇÃO E DO CONHECIMENTO CIENTÍFICO E TECNOLÓGICO: uma revisão da literatura. Ponto de Acesso, Salvador, v. 6, n. 2, 2012.

GUIZZARDI, G. Desenvolvimento para e com reuso: Um estudo de caso no domínio de Vídeo sob Demanda. Masters Thesis. Universidade Federal do Espírito Santo. 2000

GUIZZARDI, G. *et al.* Ontologias de Fundamentação e Modelagem Conceitual. II Seminário de Pesquisa em Ontologia no Brasil. Rio de Janeiro 2009.

HAAV, H.-M.; LUBI, T.-L. A Survey of Concept-based Information Retrieval Tools on the Web. 2001, 2001. p.29-41.

HARTH, A. VisiNav: A system for visual search and navigation on web data. Web Semantics: Science, Services and Agents on the World Wide Web, v. 8, n. 4, p. 348-354, 11// 2010. ISSN 1570-8268. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S1570826810000600> >.

HARTIG, O.; BIZER, C.; FREYTAG, J.-C. Executing SPARQL Queries over the Web of Linked Data. Proceedings of the 8th International Semantic Web Conference. Chantilly, VA: Springer-Verlag: 293-309 p. 2009.

HAYES-ROTH, F.; WATERMAN, D. A.; LENAT, D. B. Building expert systems. Addison-Wesley Longman Publishing Co., Inc., 1983. 350 ISBN 0-201-10686-8.

HAYES, P. RDF Semantics. W3C Recommendation, 10 February 2004. W3C 2004.

HEALY, M.; POLI, R.; KAMEAS, A. Theory and Applications of Ontology: Computer Applications. Springer, 2010. ISBN 9789048188475. Disponível em: < <http://books.google.com.br/books?id=IQS6Abf9wzwc> >.

HEATH, T.; BIZER, C. Linked Data: Evolving the Web Into a Global Data Space. Morgan & Claypool, 2011. ISBN 9781608454303. Disponível em: < <http://books.google.com.br/books?id=OFv59Wcfx8C> >.

HERMAN, I. *et al.* RDFa 1.1 Primer - Second Edition`Rich Structured Data Markup for Web Documents. W3C Working Group Note W3C 2013.

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

HERRING, J. P. Key intelligence topics: A process to identify and define intelligence needs. Competitive Intelligence Review, v. 10, n. 2, p. 4-14, 1999.

ISSN 1520-6386. Disponível em: < [http://dx.doi.org/10.1002/\(SICI\)1520-6386\(199932\)10:2<4::AID-CIR3>3.0.CO;2-C](http://dx.doi.org/10.1002/(SICI)1520-6386(199932)10:2<4::AID-CIR3>3.0.CO;2-C) >.

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: 03/09/2013.

HSU, J. Extracting the knowledge entangled in the web: technologies, applications and developments. *International Journal of Innovation and Learning*, v. 4, n. 6, p. 612-630, 01/01/ 2007. Disponível em: < <http://inderscience.metapress.com/content/D6087837KP4L210H> >.

HU, B.; SVENSSON, G. A Case Study of Linked Enterprise Data. In: PATEL-SCHNEIDER, P.; PAN, Y., *et al.* (Ed.). *The Semantic Web – ISWC 2010*: Springer Berlin Heidelberg, v.6497, 2010. cap. 9, p.129-144. (Lecture Notes in Computer Science). ISBN 978-3-642-17748-4.

HYLAND, B. Preparing for a Linked Data Enterprise. In: WOOD, D. (Ed.). *Linking Enterprise Data*: Springer US, 2010. cap. 3, p.51-64. ISBN 978-1-4419-7664-2.

HYLAND, B.; WOOD, D. The Joy of Data - A Cookbook for Publishing Linked Government Data on the Web. In: WOOD, D. (Ed.). *Linking Government Data*: Springer New York, 2011. cap. 1, p.3-26. ISBN 978-1-4614-1766-8.

JACOBS, I.; WALSH, N. *Architecture of the World Wide Web, Volume One*. W3C Recommendation: W#C 2004.

JAIN, S. C. Environmental scanning in U.S. corporations. *Long Range Planning*, v. 17, n. 2, p. 117-128, 4// 1984. ISSN 0024-6301. Disponível em: < <http://www.sciencedirect.com/science/article/pii/0024630184901432> >.

JANOWICZ, K. *et al.* Towards Meaningful URIs for Linked Sensor Data. In *Towards Digital Earth: Search, Discover and Share Geospatial Data*. Workshop at Future Internet Symposium. Berlin, Germany 2010.

JASPER, R.; USCHOLD, M. A framework for understanding and classifying ontology applications. *IJCAI-99, ONTOLOGY WORKSHOP*. Stockholm 1999.

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.

JIN, T.; BOUTHILLIER, F. Understanding information transformation process in the context of competitive intelligence. *Proceedings of the American Society for Information Science and Technology*, v. 43, n. 1, p. 1-14, 2006. ISSN 1550-8390. Disponível em: < <http://dx.doi.org/10.1002/meet.14504301303> >.

JUHARI, A. S.; STEPHENS, D. Tracing the Origins of Competitive Intelligence Throughout History. *Journal of Competitive Intelligence and Management*, v. 3, n. 4, p. 61-82, // 2006. ISSN 1703-5147. Disponível em: < <http://www.ingentaconnect.com/content/scip/jcim/2006/00000003/00000004/art00004> >. Acesso em: 29/08/2013.

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 ISBN 0684844044.

KAMPGEN, B.; HARTH, A. Transforming statistical linked data for use in OLAP systems. *Proceedings of the 7th International Conference on Semantic Systems*. Graz, Austria: ACM: 33-40 p. 2011.

KÄMPGEN, B.; O'RAIN, S.; HARTH, A. Interacting with Statistical Linked Data via OLAP Operations. *Proceedings of the International Workshop on Interacting with Linked Data*: 14 p. 2012.

KELLEY, W. T. *Marketing intelligence: the management of marketing information*. Staples P., 1968. Disponível em: < <http://books.google.com.br/books?id=GwgpAQAAMAAJ> >.

KIFER, M.; BOLEY, H. *RIF Overview (Second Edition)*. W3C Working Group Note 5 February 2013. W3C 2013.

KIMBLE, C. What Cost Knowledge Management? The Example of Infosys. *Global Business and Organizational Excellence*, v. 32, n. 3, p. 6-14, 2013. ISSN 1932-2062. Disponível em: < <http://dx.doi.org/10.1002/joe.21480> >. Acesso em: 02/09/2013.

KLYNE, G.; CARROLL, J. J. *Resource Description Framework (RDF): Concepts and Abstract Syntax*. W3C Recommendation W3C: Brian McBride 2004.

KOBILAROV, G. *et al.* Media Meets Semantic Web --- How the BBC Uses DBpedia and Linked Data to Make Connections. *Proceedings of the 6th European Semantic Web Conference on The Semantic Web: Research and Applications*. Heraklion, Crete, Greece: Springer-Verlag: 723-737 p. 2009.

KOLB, Guy; MILLER, Jerry P. O milênio da inteligência: o futuro. In: MILLER, Jerry P. *O milênio da inteligência competitiva*. Porto Alegre: Bookman, 2002. 265-272 p.

KOSALA, R.; BLOCKEEL, H. Web mining research: a survey. *SIGKDD Explor. Newsl.*, v. 2, n. 1, p. 1-15, 2000. ISSN 1931-0145.

KRIZAN, L.; COLLEGE, J. M. I. Intelligence Essentials for Everyone. Joint Military Intelligence College, 1999. Disponível em: < <http://books.google.com.br/books?id=JdIXAAAAYAAJ> >.

LAMANNA, M. Sulla prima occorrenza del termina "Ontologia". Una nota bibliografica. *Quaestio. Yearbook of the History of Metaphysics*, v. 6, p. 13, 2006.

LANA, R. A. INTELIGÊNCIA COMPETITIVA: FATOR-CHAVE PARA O SUCESSO DAS ORGANIZAÇÕES NO NOVO MILÊNIO. *Revista Inteligência Competitiva*, v. v. 1, n. n. 3, p. 23, 2011. ISSN 2236-210X. Disponível em: < https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0C CoQFjAA&url=http%3A%2F%2Fwww.inteligenciacompetitivarev.com.br%2Fajs%2Findex.php%2Frev%2Farticle%2Fview%2F19%2F38&ei=RWgqUsDVMYjI9gTqy4HYCQ&usq=AFQjCNFfh4wGOSepAzarEVUQ3J49RgfsA&sig2=ILitJGVAs2zc63W_PKm3ww&bvm=bv.51773540,d.eWU >.

LANE, R. E. The Decline of Politics and Ideology in a Knowledgeable Society. *American Sociological Review*, v. 31, n. 5, p. 14, oct/1966 1966. Disponível em: < <http://www.jstor.org/stable/209185> >. Acesso em: 02/04/2013.

LASSILA, O. Towards the Semantic Web. Towards the Semantic Web an Web Services Conference. Helsinke 2002.

LASSILA, O.; SWICK, R. Resource Description Framework (RDF) Model and Syntax. 1998. Disponível em: < <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/> >. Acesso em: 01/10/2013.

LIU, C. *et al.* Web Intelligence Analysis in the Semantic Web Based on Domain Ontology. *Information Technology Journal*, v. 10, n. 12, p. 2343-2349, 2011. Disponível em: < <http://sccialert.net/qredirect.php?doi=itj.2011.2343.2349&linkid=pdf> >. Acesso em: 01/04/2013.

LUKE, S. *et al.* Ontology-based Web agents. Proceedings of the first international conference on Autonomous agents. Marina del Rey, California, USA: ACM: 59-66 p. 1997.

LUNA, A. M. Geração de Interfaces RIA Dirigida por Ontologias. Master Dissertation – Pontifícia Universidade Católica do Rio de Janeiro. Rio de Janeiro 2009.

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.

MAHONEY, M. S. Finding a History for Software Engineering. *Annals of the History of Computing, IEEE*, v. 26, n. 1, p. 8-19, 2004. ISSN 1058-6180.

MALOW'S A. H. A theory of human motivation, Originally Published in *Psychological Review*, 50, 370-396.

MANOLA, F.; MILLER, E. RDF Primer. W3C Recommendation 2004.

MARCEAU, S.; SAWKA, K. Developing a World-Class CI Program in Telecoms. *Competitive Intelligence Review* 10, v. 10, n. 4, p. 11, 1999.

MARCHI, G. D. Um framework para sistemas baseados em conhecimento no contexto da metodologia CommonKADS , . Dissertação de Mestrado - Programa de Pós-Graduação em Ciência da Computação, Universidade Estadual de Maringá, Maringá: 2010.

MARCIAL, E. C.; GRUMBACH, R. J. S. Cenários prospectivos: como construir um futuro melhor. 5. Brasil: FGV Editora, 2008. 228 ISBN 978-85-225-0688-0.

MARCONI, M. A.; LAKATOS, E. M. Técnicas de pesquisa: planejamento e execução de pesquisas, amostragens e técnicas de pesquisa, elaboração, análise e interpretação de dados. São Paulo: Atlas, 2002. ISBN 9788522432639. Disponível em: < <http://books.google.com.br/books?id=zu5LAAAACAAJ> >.

MARCUS, S. Automating Knowledge Acquisition for Expert Systems. Kluwer Academic Publishers, 1988. 270 ISBN 0898382866.

MARTIN, P.; EKLUND, P. W. Knowledge retrieval and the World Wide Web. *Intelligent Systems and their Applications*, IEEE, v. 15, n. 3, p. 18-25, 2000. ISSN 1094-7167.

MCGUINNESS, D. L. Ontologies Come of Age In Dieter Fensel, J im Hendler, Henry Lieberman, and Wolfgang Wahlster, editors. *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*, MIT Press, 2003.

MEI, J.; LIN, Z.; BOLEY, H. An Integration of Description Logic and General Rules. In: MARCHIORI, M.; PAN, J., *et al* (Ed.). *Web Reasoning and Rule Systems*: Springer Berlin Heidelberg, v.4524, 2007. cap. 12, p.163-177. (Lecture Notes in Computer Science). ISBN 978-3-540-72981-5.

MENDELSON, N. The Self-Describing Web. TAG Finding 07 2009.

MESKELL, D. New Opportunities for Involving Citizens in the Democratic Process. *GSA*, v. 20, n. Fall, p. 2-4, 2007.

MILES, A.; BECHHOFFER, S. SKOS Simple Knowledge Organization System Reference. W3C Recommendation 18 August 2009. W3C 2009.

MILLER, J. O milênio da inteligência competitiva. Bookman, 2002. ISBN 9788536300016. Disponível em: < <http://books.google.com.br/books?id=aN-LPgAACAAJ> >.

MILLER, S. H. COMPETITIVE INTELLIGENCE - AN OVERVIEW. Competitive Intelligence Magazine: SCIP 2009.

MILOŠEVIĆ, U. *et al.* Publishing Statistical Data As Linked Open Data. Proceedings of the 2nd International Conference on Information Society Technology Information Society of the Republic of Serbia, 2012, Serbia. p.182–187.

MIRANDA, C. M. C. A Disseminação de Dados Governamentais como Serviço Público – Os Dados Abertos Governamentais e a Experiência Brasileira. CONGRESSO INTERNACIONAL SOFTWARE LIVRE E GOVERNO LETRÔNICO. Brasília: Fundação Alexandre de Gusmão. 1: 11 p. 2011.

MIZOGUCHI, R.; VANWELKENHUYSEN, J.; IKEDA, M. Task ontology for reuse of problem solving knowledge. In: MARS, N. J. I., 1995, 1995. p.46-59.

MORIK, K. Underlying Assumptions of Knowledge Acquisition as a Process of Model Refinement. Knowledge Acquisition, v. 2, n. 1, p. 21-49, 1990.

MOTA, J. Da Web 2.0 ao e-Learning 2.0: Aprender na Rede. Dissertação de Mestrado: Universidade Aberta. 2009.

NASRI, W. Competitive intelligence in Tunisian companies. Journal of Enterprise Information Management, v. 24, n. 1, p. 15, 2011. ISSN 1741-0398. Disponível em: < <http://dx.doi.org/10.1108/17410391111097429> >. Acesso em: 01/09/2013.

NEBOT, V. *et al.* Multidimensional Integrated Ontologies: A Framework for Designing Semantic Data Warehouses. In: SPACCAPIETRA, S.; ZIMÁNYI, E., *et al* (Ed.). Journal on Data Semantics XIII: Springer Berlin Heidelberg, v.5530, 2009. cap. 1, p.1-36. (Lecture Notes in Computer Science). ISBN 978-3-642-03097-0.

NEMRAVA, J. *et al.* Semantic annotation and linking of competitive intelligence reports for business clusters. Proceedings of the first international workshop on Ontology-supported business intelligence. Karlsruhe, Germany: ACM: 1-5 p. 2008.

NEUENDORF, K. A. The Content Analysis Guidebook. SAGE Publications, 2002. ISBN 9780761919780. Disponível em: < <http://books.google.com.br/books?id=huPVtmu4sigC> >.

NOY, N. F.; MCGUINNESS, D. L. Ontology Development 101: A Guide to Creating Your First Ontology. Development. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880 2001.

O'LEARY, J. *The Centurion Principles: Battlefield Lessons for Frontline Leaders*. Thomas Nelson Incorporated, 2008. ISBN 9781418578930. Disponível em: <http://books.google.com.br/books?id=Mz6V_C6Qb6YC>.

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. Disponível em: <<https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&ved=0CDYQFjAB&url=http%3A%2F%2Fwww.inteligenciacompetitivarev.com.br%2Ffojs%2Findex.php%2Frev%2Farticle%2Fview%2F17%2F61&ei=13kqUvvK9NYq88ASDv4GIBQ&usq=AFQjCNEqDRaYJN6HEGxWHom6OVp0RzYZJg&sig2=VTDDJ UdKz6FbPp3GleJcuA&bvm=bv.51773540,d.eWU>>. Acesso em: 07/09/2013.

OLIVEIRA, P.; LACERDA, J. Habilidades e competências desejáveis aos profissionais de inteligência competitiva. *Ci. Inf., Brasília*, v. 36, n. 2, p. 46-53, maio/ago. 2007.

PARDILLO, J.; MAZÓN, J. N. Using Ontologies for the Design of Data Warehouses. *International Journal of Database Management Systems (IJDM)*, v. 3, 2011.

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. ISBN 978-1-4419-7664-2.

PATNI, H.; HENSON, C.; SHETH, A. *Linked sensor data*. 2010.

PÉREZ, J.; ARENAS, M.; GUTIERREZ, C. Semantics and Complexity of SPARQL. In: CRUZ, I.; DECKER, S., *et al* (Ed.). *The Semantic Web - ISWC 2006*: Springer Berlin Heidelberg, v.4273, 2006. cap. 3, p.30-43. (Lecture Notes in Computer Science). ISBN 978-3-540-49029-6.

PHENG, L. S.; KEONG, B. L. S. Developing construction project management skills: lessons from Zhuge Liang's. *Career Development International*, v. 4, n. 1, p. 6, 1999. Disponível em: <<http://www.emeraldinsight.com/journals.htm?issn=1362-0436&volume=4&issue=1&articleid=883168&show=pdf>>.

POLI, R. Descriptive, Formal and Formalized Ontologies. In: (Ed.). *Husserl's Logical Investigations Reconsidered*: Springer Netherlands, v.48, 2003. cap. 12, p.183-210. (Contributions to Phenomenology). ISBN 978-90-481-6324-3.

PRESCOTT, J. E. *The Evolution of Competitive Intelligence: Designing a Process for Action*. PROPOSAL Management, 1999. Disponível em: <<https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0C>>

DgQFjAA&url=http%3A%2F%2Ffiles.paul-medley.webnode.com%2F200000023-97ce398c7e%2FCompetitive%2520Intelligence%2520A-Z.pdf&ei=ISsdUvW8DNG-sASOn4HwCQ&usg=AFQjCNFbIUaJVHltpuU3JUmsZS2xZfqtFAQ&sig2=ZsfUGkwQPgWcg0BNfHYNHg >.

PRESCOTT, J. E.; GIBBONS, P. T. Global competitive intelligence: An overview. In *Global Perspectives on Competitive Intelligence*. Society of Competitive Intelligence Professionals, p. 28, 1993.

PRESCOTT, J. E.; MILLER, S. H. *Inteligência competitiva na prática: técnicas e práticas bem sucedidas para conquistar mercados*. Rio de Janeiro: Campus, 2002.

PRUD'HOMMEAUX, E.; SEABORNE, A. SPARQL Query Language for RDF. W3C Recommendation 15 January 2008. W3C 2008.

PSCHORR, J. *et al.* Sensor Discovery on Linked Data. Kno.e.sis Center, Department of Computer Science and Engineering, Wright State University. Dayton, Ohio, USA 2009.

RAO, R. From unstructured data to actionable intelligence. *IT Professional*, v. 5, n. 6, p. 29-35, 2003. ISSN 1520-9202.

RICHARDSON, M.; AGRAWAL, R.; DOMINGOS, P. Trust Management for the Semantic Web. In: FENSEL, D.; SYCARA, K., *et al.* (Ed.). *The Semantic Web - ISWC 2003*: Springer Berlin Heidelberg, v.2870, 2003. cap. 23, p.351-368. (Lecture Notes in Computer Science). ISBN 978-3-540-20362-9.

ROCKART, F. J.; BULLEN, C. A PRIMER ON CRITICAL SUCCESS FACTORS. *CISR*, v. 69, 1981. ISSN 1220-81. Disponível em: <<http://dspace.mit.edu/bitstream/handle/1721.1/1988/SWP-1220-08368993-CISR-069.pdf?sequence=1>>. Acesso em: 06/09/2013.

ROCKART, J. Chief Executives Define Their Own Data Needs. *Harvard Business Review*, v. 57, n. 2, p. 81-81, // 1979. ISSN 00178012.

RODRIGUEZ, M. V. R. INTELIGÊNCIA COMPETITIVA E PERFORMANCE FINANCEIRA: UMA ASSOCIAÇÃO POSITIVA NOS PEQUENOS NEGÓCIOS EXPORTADORES DE SÓLIDOS DE MADEIRA CERTIFICADOS. Conferência Internacional de Pesquisa em Empreendedorismo na América Latina. CIPEAL: CIPEAL 2004.

ROMERO, O.; ABELLÓ, A. Automating multidimensional design from ontologies. *Proceedings of the ACM tenth international workshop on Data warehousing and OLAP*. Lisbon, Portugal: ACM: 1-8 p. 2007.

SAAYMAN, A. A. *et al.* Competitive intelligence: Construct exploration, validation and equivalence. *Aslib Proceedings: New Information Perspectives*, v. 60, n. 4, p. 383-411, 30/03/2013. Disponível em: < <http://www.scopus.com/inward/record.url?eid=2-s2.0-47949124752&partnerID=40&md5=5b4491f3341c10c95239512d59daa738> >. Acesso em: 30/03/2013.

SAUERMANN, L.; CYGANIAK, R.; VÖLKEL, M. Cool URIs for the semantic web. *W3C Interest Group Note* 03 December 2008, 2011. Disponível em: < <http://www.w3.org/TR/cooluris/> >.

SAWKA, K. Whither Analysis? . *Competitive Intelligence Magazine*, v. 9, n. 2, 2006.

SCHREIBER, A. T.; WIELINGA, B.; BREUKER, J. KADS. A Principled Approach to Knowledge-Based System Development. , *Knowledge-Based Systems*, v. 11, 1993.

SCHREIBER, G. *et al.* Knowledge Engineering and Management: The CommonKADS Methodology. The MIT Press, 2002. 471 ISBN 0262193000.

SCHREIBER, G. *et al.* CommonKADS: a comprehensive methodology for KBS development. *IEEE Expert*, v. 9, n. 6, p. 28-37, 1994. ISSN 0885-9000.

SCIP. Strategic and Competitive Intelligence Professionals. 2013. Disponível em: < www.scip.org >. Acesso em: 30/03/2013.

SERVANT, F.-P. Linking enterprise data. In *Linked Data on Web. Workshop at the 17th World Wide Web Conference* 2008.

SHADBOLT, N.; BERNERS-LEE, T.; HALL, W. The Semantic Web Revisited. *IEEE Intelligent Systems*, v. 21, n. 3, p. 96-101, 2006. ISSN 1541-1672. Acesso em: 29/03/2013.

SHAKER, S. e GEMBICKI, M. *The WarRoom Guide to Competitive Intelligence*. New York: McGraw Hill, 1999.

SHARIQ, S. Z. Knowledge Management: An Emerging Discipline. *Journal of Knowledge Management*, v. 1, n. 1, p. 8, 1997. ISSN 1367-3270. Disponível em: < <http://dx.doi.org/10.1108/EUM000000004582> >.

SHAW, M. G.; GAINES, B. The synthesis of knowledge engineering and software engineering. In: LOUCOPOULOS, P. (Ed.). *Advanced Information Systems Engineering*: Springer Berlin Heidelberg, v.593, 1992. cap. 12, p.208-220. (Lecture Notes in Computer Science). ISBN 978-3-540-55481-3.

SILVA, E. L.; MENEZES, E. M. Metodologia da pesquisa e elaboração de dissertação. Florianópolis: Laboratório de Ensino a Distância, 2005.

SINGH, A. K.; KUMAR, R. P. A Comparative Study of Page Ranking Algorithms for Information Retrieval International Journal of Electrical and Computer Engineering v. 4, n. 7, 2009. Disponível em: < <http://waset.org/journals/ijece/v4/v4-7-69.pdf> >. Acesso em: 10/09/2013.

SIZOV, S. What Makes You Think That? The Semantic Web's Proof Layer. IEEE Intelligent Systems, v. 22, n. 6, p. 94-99, 2007. ISSN 1541-1672.

SPIVAC, N. How the WebOS Evolves? 2007.

SPORNY, M. *et al.* JSON-LD 1.0. A JSON-based Serialization for Linked Data. W3C Candidate Recommendation W3C 2013.

STADLER, C. *et al.* LinkedGeoData: A Core for a Web of Spatial Open Data. Semantic Web Journal., 2012. Disponível em: < http://www.semantic-web-journal.net/sites/default/files/swj173_2.pdf >. Acesso em: 10/10/2013.

STAREC, C. Gestão estratégica da informação e inteligência competitiva. Saraiva, 2005. ISBN 9788502053809. Disponível em: < <http://books.google.com.br/books?id=PfKVAAAACAAJ> >.

STOLIENWERK, M. F. L. Implantação de Sistemas de Inteligencia Competitiva: abordagem corporativa Bibliotecon, Basilia, v. 23/24, n. 4, p. 473-492, 1999.

STUDER, R.; BENJAMINS, V. R.; FENSEL, D. Knowledge engineering: Principles and methods. Data & Knowledge Engineering, v. 25, n. 1-2, p. 161-197, 3// 1998. ISSN 0169-023X. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0169023X97000566> >.

TARAPANOFF, K.; GREGOLIN, J. A. R. Inteligência organizacional e competitiva. Ciência da Informação, v. 31, p. 108-109, 2002. ISSN 0100-1965. Disponível em: < http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652002000300012&nrm=iso >.

TAYLOR, J. Project Scheduling and Cost Control: Planning, Monitoring and Controlling the Baseline. J. Ross Pub., 2007. ISBN 9781932159110. Disponível em: < <http://books.google.com.br/books?id=AnAeKKFGhTwC> >.

TOUSSAINT-SAMAT, M.; BELL, A. A History of Food. Wiley, 1994. ISBN 9780631194972. Disponível em: < <http://books.google.com.br/books?id=i4xuO9TsHf8C> >.

TRUJILLO, J. Bridging the Semantic Gap in OLAP Models: Platform independent Queries Categories and Subject Descriptors. Computing Systems p. 89, 2008.

TYSON, K. W. M. *The Complete Guide to Competitive Intelligence*. Leading Edge Publications, 2002. ISBN 9780966321920. Disponível em: < <http://books.google.com.br/books?id=2AqWAAAACAAJ> >.

TZU, S. *A arte da Guerra*. Brasil: Jardim dos Livros, 2006. ISBN I.S.B.N.: 856001800X.

UNDERWOOD, J. *Competitive Intelligence*. Wiley, 2002. ISBN 9781841122267. Disponível em: < <http://books.google.com.br/books?id=TusLAAAACAAJ> >.

URIONA MALDONADO, M.; COSER, A. ENGENHARIA DO CONHECIMENTO APLICADA AO SERVIÇO DE ATENDIMENTO AO CLIENTE NA INDÚSTRIA DO SOFTWARE. *Ingeniare. Revista chilena de ingeniería*, v. 18, p. 53-63, 2010. ISSN 0718-3305. Disponível em: < http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-33052010000100007&nrm=iso >.

USCHOLD, M.; GRUNINGER, M. Ontologies: principles, methods an applications *Knowledge Engineering Review*, v. 11, n. 2, 1996. Disponível em: < <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.111.5903&rep=rep1&type=pdf> >. Acesso em: 10/10/2013.

USCHOLD, M.; GRUNINGER, M. Ontologies and semantics for seamless connectivity. *SIGMOD Rec.*, v. 33, n. 4, p. 58-64, 2004. ISSN 0163-5808.

VAITSMAN, H. S. *Inteligência empresarial - atacando e defendendo*. Interciência, 2001. 214 ISBN 8571930562.

VAN HEIJST, G.; SCHREIBER, A. T.; WIELINGA, B. J. Using explicit ontologies in KBS development. *Int. J. Hum.-Comput. Stud.*, v. 46, n. 2-3, p. 183-292, 1997. ISSN 1071-5819.

VAN ZUYLEN, C. H. Using Inxight Search Extender for Google with Inxight SmartDiscovery Awareness Server – A First-person Success Story. *Inxight Software Inc.*, Sunnyvale, CA., 2006. Disponível em: < www.inxight.com >.

VANDER WAL, T. *Folksonomy Coinage and Definition*. 2007. Disponível em: < <http://vanderwal.net/folksonomy.html> >.

VILCHES-BLÁZQUEZ, L. M. *et al.* An Approach to Publish Spatial Data on the Web: The GeoLinked Data Case. *Workshop on Linked SpatioTemporal Data 2010* in conjunction with the 6th International Conference on Geographic Information Science GISCIENCE. Zurich, Suíça 2010.

VILELA, R. M.; OLIVEIRA, M. J. Option pricing with fractional volatility. eprint arXiv:cond-mat/0404684, 2004.

VILAN FILHO, Jayme Leiro; ARRUDA, Raíza Veloso; PERUCCHI, Valmira. Análise das citações aos periódicos científicos brasileiros das áreas de informação. Em *Questão: Revista da Faculdade de Biblioteconomia e Comunicação da UFRGS*, Porto Alegre, v. 18, n. 1, p. 115-127, 2012.

VILLAZÓN-TERRAZAS, B. *et al.* Methodological Guidelines for Publishing Government Linked Data. In: WOOD, D. (Ed.). *Linking Government Data*: Springer New York, 2011. cap. 2, p.27-49. ISBN 978-1-4614-1766-8.

W3C OWL WORKING GROUP, W. C. OWL 2 Web Ontology Language Document Overview (Second Edition). W3C Recommendation 11 December 2012. W3C 2012.

WALTON, T. *Challenges in Intelligence Analysis: Lessons from 1300 BCE to the Present*. Cambridge University Press, 2010. ISBN 9780521132657. Disponível em: < http://books.google.com.br/books?id=_cNvJ__P8nAC >.

WARNER, M. Wanted: A Definition of “Intelligence”. *Studies in Intelligence* v. 46, n. 3, p. 3, 2002. Disponível em: < <https://www.cia.gov/library/center-for-the-study-of-intelligence/kent-csi/vol46no3/pdf/v46i3a02p.pdf> >.

WEISS, A. A brief guide to competitive intelligence: how to gather and use information on competitors. *Business Information Review*, v. 19, n. 2, p. 39-47, June 1, 2002. Disponível em: < <http://bir.sagepub.com/content/19/2/39.abstract> >.

WEST, C. Competitive Intelligence in Europe. *Business Information Review*, v. 16, n. 3, p. 143-150, September 1, 1999. Disponível em: < <http://bir.sagepub.com/content/16/3/143.abstract> >.

ZANAZI, A. Text Mining: Competitive and Customer intelligence in real business cases. *IntEmpres2002 Conference Proceedings*. LaHabana 2002.

ZAPILKO, B.; MATHIAK, B. Performing Statistical Methods on Linked Data. *Proc. Int’l Conf. on Dublin Core and Metadata Applications*: 116–125 p. 2011.

ZHANG, Y.; WU, J.; WANG, C. Automatic Competitive Intelligence Collection Based on Semantic Web Mining. *Wireless Communications, Networking and Mobile Computing*, 2007. *WiCom 2007. International Conference on*, 2007, 21-25 Sept. 2007. p.3701-3704.

ZHAO, J.; BOLEY, H. Uncertainty Treatment in the Rule Interchange Format: From Encoding to Extension 2013. Disponível em: < http://ruleml.org/rif/URSW2008_F9_ZhaoBoley.pdf >. Acesso em: 04/10/2013.