

DOI: 10.5748/9788599693124-13CONTECSI/RF-4276

**COMPETITIVE INTELLIGENCE PLATFORM BASED ON INFORMATIONAL SEMANTIC ENVIRONMENTS IN BIG DATA CONTEXT: COMPUTER AND INFORMATIONAL MODEL OF INNOVATION SUPPORT**

Elvis Fusco (Centro Universitário Eurípides de Marília, São Paulo, Brasil) - elvisfusco@univem.edu.br

Marcos Luiz Mucheroni (Universidade de São Paulo, São Paulo, Brasil) - mucheroni.marcosl@gmail.com

Fábio Dacêncio Pereira (Centro Universitário Eurípides de Marília, São Paulo, Brasil) - prof.fabiopereira@gmail.com

The contemporary setting of digital information environments of organizations characterized by dependence on the use of information in the processes of decision-making and innovation, this coupling between management, innovation and information is one of the consequences of the Information Society, context which technology and information constitute the organizational base of relationships in contemporary society. Among the various efforts to use the information to support the actions of decision support and innovation in organizations, the Competitive Intelligence which uses the systematic information processes to support such actions. These processes are represented by information flows that are characterized by the search, access, retrieval, transformation, processing, representation, persistence, transfer, presentation and use of information, and as informational sources, the actors of innovation formed mainly by government, business and educational institutions. In order to establish a Competitive Intelligence platform to support innovation, this paper proposes the combination of: a) computer technology as extraction agents, Data Fusion, Big Data Analytics and NoSQL Databases; b) semantic information technologies: ontologies, metadata standards, Information Retrieval, to build an architecture to support the process of management and decision-making in the innovation scenario. The use of the combination of computational, informational and semantic technologies of the spaces contained in the platform that were described before lets the main actors in the field of innovation (government, business and educational institutions) have access to product information generated and collaboratively support themselves in collective and collaborative construction processes of new knowledge and actions to encourage innovation in organizations.

Keywords: Competitive Intelligence, Semantic Web, Big Data, Extraction agent, Innovation.

## 1. INTRODUÇÃO

A crescente geração massiva de dados está testando a capacidade das mais avançadas tecnologias de armazenamento, tratamento, transformação e análise de informações. As áreas de Gestão e Recuperação da Informação e Apoio à Decisão estão sendo desafiadas pelo volume, variedade e velocidade de uma imensidão de dados semiestruturados e não estruturados de natureza complexa que devem ser encontrados e julgados quanto ao seu valor e veracidade, que também oferece às organizações excelentes oportunidades de terem um aprofundamento no conhecimento mais preciso de seus negócios.

Diversos autores abordam a temática e questionam se as máquinas poderiam tomar decisão. Entre eles, destaca-se David Hilbert, que no VI Congresso Internacional de Matemáticos, em Bolonha, argumentou que o problema é como fazer um conjunto completo tomar uma decisão que está fora dele. Chamado de Problema da Completude mais tarde, mas que Hilbert apresentou da seguinte maneira: não permitir seja “A e não-A” (simbolicamente “ $A \wedge \neg A$ ”).

Kurt Gödel, famoso pela sua participação no círculo de Viena, grupo que se estruturou na década de 1920 e que o nazismo dissolveu, elabora o chamado “paradoxo de Russel” de maneira mais profunda, mostra usando a axiomática da aritmética, que o conjunto A desde que seja consistente, não é completo, chamado então Teorema da Incompletude.

Posteriormente, os estudiosos Norbert Wiener e Vannevar Bush, que lecionavam do MIT passaram a estudar o problema, então chamado de problema da decisão (Entscheidungsproblem). Claude Shannon e Alan Turing dialogaram sobre a questão, sendo que Turing propôs uma solução para ele em seu artigo: "*On computable numbers, with an application to the Entscheidungsproblem*" (TURING, 1936) e chegaram à conclusão que é impossível decidir algorítmicamente se um enunciado na aritmética é verdadeiro ou falso.

Turing reduziu o problema da parada para as Máquinas de Turing ao Entscheidungsproblem, e o seu artigo é considerado a maior influência na construção dos computadores e no problema da Teoria da Decisão subsequente no mundo computacional. Influenciou toda a teoria de Inteligência Artificial (MINSKY, 1985) e mais tarde constituindo uma taxonomia para os agentes inteligentes (FRANKLIN, GRAESSER, 1996) e atualmente influencia a construção de agentes para a Web Semântica (HENDLER, 2001) e em artigo recente Kaplan (2015) mostra os limites do que pode-se chamar realmente de inteligência artificial.

Entretanto, a capacidade de tratar volumes de dados é corrente sendo necessária uma abordagem que torne sustentáveis os volumes de informação.

Neste contexto, surgem inúmeras oportunidades em agregar valor ao negócio com base nessas informações que são geradas tanto no ambiente interno quanto no externo. Porém há a necessidade de uma nova abordagem na estrutura da Ciência de Dados e Tecnologia da Informação das organizações em transformar esses dados em conhecimento para os tomadores de decisão que causará impactos estratégicos sustentáveis.

Para agregar e utilizar as informações que estão espalhadas nos ambientes internos e externos das organizações, surge o conceito da Inteligência Competitiva que segundo a ABRAIC (Associação Brasileira dos Analistas de Inteligência Competitiva), é um processo informacional proativo que conduz à melhor tomada de decisão, seja ela estratégica ou operacional, visando 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 (BRASILIANO, 2002).

No cenário atual destas informações geradas nos ambientes organizacionais, principalmente nos que tem a Internet como plataforma, encontram-se dados que devido às suas características, atualmente classificam-se como Big Data.

Dentre estas características destacam-se: volume - enormes conjuntos de dados que são de magnitude maior do que os dados mantidos em sistemas de armazenamento tradicional; variedade - dados heterogêneos, complexos e variáveis que são gerados em formatos diversos que tem como fonte: e-mails, mídias sociais, vídeos, imagens, blogs e bases da web; velocidade - os dados são gerados em fluxo constante com consultas em tempo real de informações significativas para tomada de decisão; valor - esses dados são potenciais para geração de conhecimentos significativos que oferecem análises preditivas para futuras tendências e padrões, que vão além dos resultados tradicionais de consultas e relatórios de sistemas de informação transacionais.

Na publicação do Journal of Science (GRAHAN-ROWE, 2008) Big Data é definido como a representação do andamento dos processos cognitivos humanos, que geralmente inclui conjuntos de dados com tamanhos além da capacidade da tecnologia atual, métodos e teorias para capturar, gerenciar e processar os dados dentro de um tempo determinado. Beyer e Laney (2012) definem Big Data como o alto volume, alta velocidade e/ou alta variedade de informações que requerem novas formas de processamento para permitir melhor tomada de decisão, nova descoberta do conhecimento e otimização de processos.

Nos ambientes de Big Data, apenas o uso de bancos de dados relacionais não é adequado para a persistência, processamento e recuperação dos dados em ambientes escaláveis e heterogêneos. Para tentar resolver esta questão no âmbito da persistência da informação surgem novos conceitos nas tecnologias de banco de dados, como o NoSQL (*Not Only SQL*) que para De Diana e Gerosa (2010) veio representar soluções alternativas ao modelo relacional, oferecendo maior escalabilidade e velocidade no armazenamento dos dados surgindo como uma opção mais eficaz e barata.

O uso de conceitos de Business Intelligence e Inteligência Competitiva e tecnologias como Data Warehouse, OLAP, Analytics, Data Mining, Fusão de Dados, Ontologias, NoSQL e robôs de busca semântica representam abordagens para capturar, gerenciar e analisar cenários de Big Data. A necessidade da utilização dessas tecnologias no tratamento desses dados massivos e complexos está causando uma mudança de paradigma que está levando as organizações a reexaminar sua infraestrutura de TI (Tecnologia da Informação) e sua capacidade de análise e gestão corporativa da informação.

A gestão eficaz e a análise de dados em larga escala representam um interessante, mas crítico desafio, pois os modelos de gestão baseados na Inteligência Competitiva estão sendo influenciados por esse universo complexo de informações geradas com o conceito de Big Data e novas investigações são necessárias para dar solução a esse desafio de uso eficiente das informações no processo de gestão.

No processo de busca da informação em cenários da Inteligência Competitiva e Big Data são utilizados robôs de extração de dados na Internet, que segundo Deters e Adaime (2003) são sistemas que coletam os dados da Web e montam uma base de dados que é processada para aumentar a rapidez na recuperação de informação e que segundo Silva (2003), a extração de informações relevantes pode classificar uma página segundo um contexto de domínio e também retirar informações estruturando-as e armazenando-as em bases de dados.

Com o propósito de adicionar significado aos conteúdos buscados em domínio

específico associam-se aos robôs de busca na Web conceitos semânticos, que permitem realizar a procura não mais por palavras chaves num processo de busca textual, mas sim por significado e valor, extraindo das páginas e serviços da Web informações de real relevância, descartando aquilo que é desnecessário. A partir disto, a ontologia aparece como solução na busca de inserir semântica neste processo.

A ontologia, no contexto filosófico, é definida por Silva (2003) como a parte da ciência que estuda o ser e seus relacionamentos e neste sentido, o uso de ontologias é essencial no processo de desenvolvimento dos robôs de busca semântica, sendo aplicada na Ciência da Computação e na Ciência da Informação para possibilitar uma busca de maneira mais inteligente e mais próxima do funcionamento do processo cognitivo do usuário de forma que a extração de dados se torne muito mais relevante.

Atualmente vivencia-se uma nova disrupção tecnológica pela convergência da colaboração, mobilidade e grande volume de dados (Big Data). O grande desafio para a pesquisa de ambientes computacionais e informacionais e para a forma de uso das informações nas organizações está em promover a integração destas tecnologias para balancear as necessidades de geração, acesso e controle destas informações, bem como as oportunidades deste comportamento emergente e suas inovações.

Este cenário em que as organizações estão inseridas exige uma evolução dos processos de gestão e de construção contínua de novos conhecimentos na geração da inovação tecnológica, ou seja, de seus produtos e serviços. Como resultado do uso da informação e da construção do conhecimento na promoção da inovação, há a geração de resultados a longo prazo, ou seja, considerando que as inovações são capazes de gerar vantagens competitivas, a médio e longo prazo, inovar torna-se essencial para que as organizações se mantenham estáveis e sustentáveis na sociedade contemporânea.

Neste contexto, com o objetivo de estabelecer uma plataforma de Inteligência Competitiva de apoio à inovação, este artigo propõe a combinação de: a) tecnologias computacionais como Agentes de Extração, Fusão de Dados, Big Data Analytics e Bancos de Dados NoSql; b) tecnologias informacionais semânticas: ontologias, padrões de metadados, Recuperação da Informação, na construção de uma arquitetura de apoio ao processo de gestão e a tomada de decisão no cenário da inovação.

Este artigo está organizado da seguinte forma: inicialmente apresenta-se a conceituação de inteligência competitiva, enfatizando o seu papel na recuperação da informação. Neste contexto são apresentados os agentes computacionais de extração automática de informação e posteriormente conceituada a web semântica, destacando a ontologia, sua relevância e utilidade. Na sequência são explorados os conceitos de Big Data e inovação na era da informação e finalmente, apresenta-se uma plataforma de inteligência competitiva baseada em agentes informacionais semânticos no contexto do Big Data, propondo um modelo computacional informacional de apoio à inovação.

## **2. INTELIGÊNCIA COMPETITIVA**

O cenário organizacional contemporâneo caracteriza-se pela dependência do uso da informação nos processos de tomada de decisão, esse acoplamento entre a gestão e a informação é uma das consequências da Sociedade da Informação, contexto no qual tecnologias e informação constituem a base das relações na sociedade contemporânea, período compreendido a partir dos anos 70 em que se percebe o reconhecimento da Sociedade da Informação, resultado do desenvolvimento das Tecnologias da Informação e Comunicação, é o estado em que a sociedade se encontra em que a informação tem um

valor inestimável e que sua produção, disseminação, acesso e preservação são essenciais para o crescimento, a manutenção e competitividade de uma dada sociedade e suas organizações.

Este contexto caracteriza-se também pela aceleração da inovação científica e tecnológica, rapidez na transmissão de informações em tempo real, gestão de informações não-lineares e ruptura com os conceitos de tempo e espaço.

Com as tecnologias de comunicação dos últimos anos, vivencia-se a Sociedade em Rede, com um conjunto de nós interconectados e flexíveis, potencializados pela Internet, que se transformou da Sociedade da Informação e que agora está toda conectada formando uma espécie de "teia" onde se entrelaçam pessoas, organizações e nações e por meio desses nós é que se produzem e disseminam todo e qualquer tipo de informação.

Em relação à gestão das organizações, o processo decisório é dependente do conhecimento do agente tomador de decisão e esse conhecimento é influenciado e determinado pelo seu acesso à informação de ambientes internos e externos às organizações, assim a informação é um recurso estratégico para obtenção de vantagens competitivas e fortalecimento da sustentabilidade das organizações.

Conforme identificado por Drucker (1995), a falta de acesso à informação gera a obsolescência do conhecimento nas organizações que faz com que haja uma queda na competitividade gerada pela incapacidade de inovar diante das informações intra e extra organizações.

Esse processo de acesso à informação tem sido revisto do ponto de vista dos Sistemas de Informação que historicamente se limitavam a levar em consideração as informações dos ambientes internos da organização e, neste contexto, a Inteligência Competitiva (IC) traz conceitos importantes para a ampliação do domínio de acesso à informação.

A Inteligência Competitiva é definida pela ABRAIC como um processo informacional pró-ativo 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. Esse processo informacional é composto pelas etapas de coleta e busca ética de dados, informes e informações formais e informais tanto do macro-ambiente quanto do ambiente interno, análise de forma filtrada, integrada e respectiva disseminação (ABRAIC, 2015).

O processo de Inteligência Competitiva é realizado por meio da coleta ética e do uso da informação pública e disponível sobre tendências, eventos e atores no ambiente externo da organização, sistematizada para ser utilizada dentro da análise de cada organização.

Para Kahaner (1996), a Inteligência Competitiva é um programa sistemático de coleta e análise da informação sobre atividades dos concorrentes e tendências gerais dos negócios, visando atingir as metas da empresa..

Morais (1999) estende o conceito de IC a um sistema que contempla o processo de planejamento a priori e a utilização, disseminação e avaliação a posteriori, em relação ao processo de coleta e análise de dados.

A IC reforça a necessidade das organizações desenvolverem Competências Informacionais como Gestão da Informação e Conhecimento, Modelagem de Dados, Fluxos Informacionais, Gestão de Dados, Analytics, Processos Informacionais, entre outras.

O processo de construção de um projeto de IC inclui definição dos processos informacionais de busca, seleção, recuperação, tratamento, armazenamento, acesso, análise

e disseminação da informação.

Esse processo é caracterizado pelo dinamismo da informação principalmente no macro ambiente, essa característica gera a necessidade do monitoramento do ambiente interno (registros internos, dados de sistemas legados, etc) e do ambiente externo, identificado pelas forças demográficas, econômicas, naturais, tecnológica, político-legais, sócio-culturais e pela concorrência.

O monitoramento do ambiente externo torna-se complexo se realizado de forma manual e por pessoas devido à amplitude e ao dinamismo informacional dos Ambientes Informacionais Digitais identificados como Big Data. Neste contexto, são utilizados os Agentes de Extração Automática, que fazem o monitoramento e a extração das fontes informacionais da Web.

Em Passos (2005) é apresentado o Modelo Herring de implantação da IC, conforme a Figura 1.



Figura 1: Modelo Herring de implantação da Inteligência Competitiva.

Fonte: Passos (2005)

Neste modelo, o processo começa pela identificação das necessidades informacionais geradas pelo planejamento estratégico da organização, ou seja, as informações são utilizadas para atender as estratégias organizacionais. A partir destas, são identificadas as fontes informacionais, formas de extração, periodicidade de extração, entre outras variáveis que serão utilizadas no processo de coleta de dados que abastecerá o ambiente de análise que transformará o conjunto de informações em produtos informacionais a serem disseminados e apresentados aos atores tomadores de decisão.

## 2.1. Recuperação de Informação em Sistemas de Inteligência Competitiva

A recuperação da informação tem se tornado alvo de muitos estudos, devido à grande quantidade de informações que hoje se encontram espalhados pela rede.

A recuperação da informação lida com a representação, armazenamento, organização e acesso as informações. Devendo prover ao usuário aquilo que ele necessita de uma maneira facilitada (BAENZA-YATES E RIBEIRO-NETO, 1999).

O conceito de recuperação de informação é diferente de recuperação de dados. A recuperação de dados consiste em extrair de um banco de dados qualquer documento que contém uma expressão regular ou os termos ali contidos. Sendo que a recuperação da informação vai além, levando em conta a sintaxe e a semântica daquela informação, buscando satisfazer o que o usuário está pesquisando (BAENZA-YATES E RIBEIRO-NETO, 1999).

Desta maneira a recuperação da informação tem assumido um papel diferenciado na Ciência da Informação e na Ciência da Computação, pois aparece como elo final na busca pela apresentação da informação mais adequada ao usuário no menor tempo possível.

O processo de recuperação da informação não consiste apenas em técnicas e métodos que envolvem o armazenamento e os algoritmos de recuperação, mas também em adaptar os sistemas no comportamento do usuário, entendendo desta maneira, como é a construção da informação e das instruções para a recuperação da informação (SANTAREM SEGUNDO, 2010).

Com o surgimento da Web houve grande aumento no volume das informações eletrônicas, que trouxeram muitas vantagens quanto à possibilidade de troca, difusão e transferência de dados. Entretanto, este crescimento trouxe muitos problemas relacionados ao acesso, busca e recuperação das informações de real valor imerso em grandes volumes de dados (MODESTO, 2013).

Assim, um dos desafios da recuperação da informação é conseguir fazer com os Ambientes Informacionais Digitais entendam o que o usuário está necessitando, de forma que os resultados vindos da busca possam ser de real valor e importância para o usuário.

O termo Recuperação da Informação foi usado pela primeira vez em 1951, por Mooers (1951), quando definiu os problemas que seriam tratados por esta nova disciplina. Desta maneira a Recuperação da Informação trata dos aspectos da descrição e especificação das buscas da informação. Tratando também de qualquer sistema, técnicas e máquinas utilizadas no processo de recuperação da informação.

Desta maneira o processo de Recuperação da Informação, consiste em encontrar em um conjunto de documentos de um sistema, quais são os que atendem às necessidades informacionais do usuário. Assim, o usuário não está interessado em recuperar dados, nem achar documentos que satisfaçam sua expressão de busca, mas em encontrar a informação sobre um determinado assunto (FERNEDA, 2003).

Os sistemas de Recuperação de Informação em ambientes de Inteligência Competitiva devem representar os documentos e apresentá-los aos usuários de maneira que estes, por meio daqueles documentos recuperados, consigam satisfazer total ou parcialmente as suas necessidades informacionais em ambientes internos e externos.

Com o grande aumento do uso da Web como fonte informacional e sua riqueza de informações úteis para as organizações, caracterizadas em Sistemas de Inteligência Competitiva, abre-se um foco de pesquisa relacionada à Recuperação da Informação para conseguir extrair informações de forma automática e sistemática devido à amplitude e ao

dinamismo informacional dos Ambientes Informacionais Digitais identificados no cenário de Big Data.

O grande desafio da recuperação da informação em ambientes de Inteligência Competitiva é o fato que esta foi construída de maneira descentralizada, de forma que muitas estratégias de buscas não conseguem alcançar altos índices de intersecção entre os documentos e as necessidades informacionais do usuário.

Segundo Santarem Segundo (2010, p. 39)

Dentro de uma nova dimensão como a Internet, fica visível o esgotamento de alternativas com relação a esses modelos já conhecidos, visto que existe uma clara mudança do corpus de consulta. Com a introdução da Internet no contexto do usuário, passa-se a ter um depósito de informações muito mais amplo, que carrega consigo a ligação de documentos e informações através de links, criando uma interligação entre os documentos armazenados e disponíveis na rede.

Verifica-se, portanto, a necessidade de buscar novas maneiras de realizar a recuperação da informação, neste novo ambiente, que tem a Web como plataforma, onde as informações são dos mais diversos tipos, onde os motores de busca apresentam uma quantidade muito grande de links e páginas para que o usuário possa encontrar o que atende a sua necessidade.

Neste contexto, a Inteligência Competitiva apresenta-se com propostas de ampliar os resultados da Recuperação da Informação e para isso, os agentes computacionais de extração automática de informação exercem um papel fundamental no fluxo informacional de Sistemas de Inteligência Competitiva.

### 3. AGENTES COMPUTACIONAIS

Os agentes computacionais são elementos que começaram a surgir, com maior frequência, na literatura da área da Ciência da Computação, a partir do final do século XX. Identifica-se tal movimento em bibliografias como Genesereth e Ketchpel (1994), Franklin e Graesser (1996) e Wooldridge (1997).

A princípio verificou-se que o conceito de agentes computacionais estava fortemente vinculado a área de inteligência artificial, em que essa ferramenta se apresentava como um importante requisito na busca de criar aplicações baseadas em inteligência computacional.

Na busca por uma definição de agentes computacionais, encontram-se diversos conceitos utilizados por distintos autores. Uma definição muito citada é dada por Wooldridge (1997), que explicita que agentes computacionais são sistemas capazes de decidir o que deve ser feito em uma determinada situação. Franklin e Graesser (1996) aprofundam na busca pelas definições de agentes computacionais, enumerando diversos conceitos dados pelos principais autores e instituições da época, realizando a partir das mesmas uma definição bastante citada e utilizada na literatura. Os autores relatam que um agente “é um sistema que se encontra dentro de um ambiente, sentindo e agindo sobre o mesmo ao passar do tempo, na busca de cumprir sua agenda, efetuando o que ele sente no futuro” (FRANKLIN; GRAESSER, 1996, p. 25, tradução nossa).

As definições apresentadas contemplam especialmente características envolvendo a autonomia e a capacidade desses sistemas tomarem decisões, verificando o ambiente em que os mesmos se encontram. Cabe ressaltar que as tomadas de decisão



realizadas por agentes podem utilizar-se de diversos tipos de algoritmos computacionais, podendo ser simplesmente sistemas programados para realizarem determinadas tarefas, que se relacione com um ambiente ao se adaptar ao mesmo, ou tomar certa decisão baseada na ocorrência de determinados fatores.

Nessa perspectiva, verifica-se o papel importante dos robôs de extração. Essa tecnologia computacional é definida como uma estrutura que percorre determinadas páginas da Web ou toda a Web recolhendo as informações que apresentam relevância para um determinado sistema. Cada robô apresenta diversos comportamentos, que podem variar conforme cada aplicação, mas em suma, um robô contém uma lista com os websites que devem ser extraídos, podendo executar também uma extração recursiva, em que o robô acessa primeiramente a página inicial, e navega por todos os links que tal página possui. Destaca-se, que a recuperação das páginas ocorre por meio do protocolo HTTP (DETERS; ADAIME, 2003).

Alvarez (2007) contextualiza o conceito de extração de informação, dizendo que esta é uma área focada em extrair de modo automatizado, buscando extrair partes que sejam relevantes, tornando elementos dispostos em linguagem natural, por exemplo, um texto, em sistemas formais para o entendimento de computadores. O autor ressalva ainda que o processo de extração de informação é determinado por um conjunto de regras e padrões, que pode ser definida por especialistas, com graus determinados de automação.

Os dois conceitos apresentados, robô de buscas e extração de informação, estão fortemente interligados. Principalmente em ambientes Web, o processo de extração de informação costuma estar totalmente conectado com robôs de buscas, que realizam a varredura, e extraem, seguindo determinado padrões, as informações que foram previamente estabelecidas como relevantes, em um determinado contexto.

Na perspectiva desse trabalho, o termo agente de extração, tem como base o conceito de robô de extração, que é capacitado para extrair conteúdos de páginas HTML, com as propriedades de agentes computacionais, ao expandir as funções realizadas pelos robôs de busca, podendo o mesmo se adaptar e se integrar com diversos módulos de tomadas de decisão. Assim, os agentes computacionais inteligentes na Internet (não necessariamente na Web) ou simplesmente um agente, é um programa que usa uma informação ou realiza um serviço sem uma presença humana direta em um serviço regular, no âmbito da Ciência da Informação pode-se designar a uma busca de informações por meio de algum vocábulo ou informação simples atendendo a algum critério semântico.

Este vocábulo normalmente é um parâmetro fornecido por um usuário que realiza alguma busca na Internet que está procurando reuniões informações contextual sobre ele, e que pode ser um busca realizada com certa periodicidade, tipicamente um programa para um agente inteligente é chamado bot (ou softbot, uma abreviação para robôs de software).

Normalmente, um programa de agente, utilizando parâmetros fornecidos previamente, procura toda ou alguma parte da Internet, reúne informações relevantes e apresenta em uma base diária ou periódica. Os agentes de extração são chamados também de bot (abreviação de *robot*).

Existem algumas áreas que estão utilizando agentes de extração, no contexto da Web. Uma área que apresenta diversos trabalhos relacionados a agentes de extração é a Web Semântica, em que a extração de informação vem sendo relacionada a ontologias, na busca de inserção de semântica no processo de Recuperação da Informação. Um desses trabalhos é apresentado por Coneglian e Fusco (2015) em que os autores apresentam uma plataforma que tem como centro um agente que extrai informações de páginas Web, utilizando ontologias para classificar e representar os dados extraídos. Outro estudo de

destaque é apresentado por Freitas (2002), que em sua tese, o autor discorre sobre o uso de ontologias para realizar os processos de extração de páginas Web, buscando verificar as informações de maior relevância para um determinado domínio.

Uma área que também apresenta uma forte aproximação com os agentes de extração, é a Inteligência Competitiva, na busca de encontrar as informações que possuam um maior valor às organizações. Nessa perspectiva, o trabalho de Silva (2003) insere a questão do uso de agentes para realizar o monitoramento de fontes informacionais, no intuito de automatizar o processo de extração de informação.

Considerando, assim, o conceito da Inteligência Competitiva, cenário no qual as informações estão num ambiente dinâmico e disperso, os agentes computacionais uma ferramenta capaz de se relacionar com o ambiente, para atender a determinadas tarefas, verifica-se a necessidade de criar agentes computacionais semânticos de extração de informação, que sejam capazes de agregar a função de robôs de extração, com características de agentes computacionais e padrões ontológicos.

#### 4. WEB SEMÂNTICA

Em 1989, Tim Berners Lee propõe a criação da Web, idealizando também posteriormente a criação da primeira versão do “HyperText Markup Language” (HTML), que é a linguagem de formatação de documentos de links de hipertexto, que se tornou o formato básico para a publicação dentro da Web (W3C, 2014) (BERNERS-LEE, 1989).

A partir de então, a Web passa a crescer e disponibilizar informações dos mais variados tipos, sendo estas informações preparadas principalmente para a leitura humana (BERNERS-LEE, 1989)

Desta forma, embora a Web tenha sido criada com o intuito de possibilitar o fácil acesso, intercâmbio e a recuperação da informação, em seu início foi gerada de uma maneira totalmente descentralizada e acabou crescendo de uma maneira exponencial no âmbito da colaboração da Web 2.0. Sendo que hoje existe uma grande quantidade de informações, mas quando há a necessidade de recuperar algo, os resultados obtidos são poucos satisfatórios (SOUZA E ALVARENGA, 2004).

De forma contrária a isto, a Web foi concebida não com a intenção de apenas o ser humano entender o conteúdo ali presente, mas também para que as máquinas conseguissem fazer a leitura e o processamento nas informações ali contidas. Na busca de resolver este problema, surge o termo Web Semântica.

Tim Berners-Lee propõe colocar um sentido nos termos da Web, para que além dos humanos, as máquinas também consigam entender os conteúdos. E assim consigam estruturar as informações, fazendo conjuntos de regras de inferências para automatizar o raciocínio.

A partir de então, surgiram várias representações e maneiras para fazer da Web Semântica uma realidade.

Dziekaniak (2004) diz que a semântica não está apenas relacionada ao conteúdo de um recurso, mas também na relação deste com os outros conteúdos da Web. Logo é necessário que os recursos da Web sejam muito expressivos, para que os agentes e máquinas consigam processar a informação e entender seu valor.

Assim a Web Semântica, deve trazer um significado às páginas, propiciando desta maneira com que os agentes e máquinas encontrem um ambiente que promova buscas e a recuperação da informação (BERNERS-LEE, 2001a).

A Web Semântica não tem a intenção de criar uma nova Web, e sim de trazer

um entendimento sobre a atual Web, onde a informação possa além de ser entendida por pessoas ser entendida por máquinas (PRAZERES, 2004).

Assim, como ambientes da Web Semântica, pode-se imaginar um conjunto de vocabulários e ontologias cuidadosamente construídas por especialistas, ou o que é mais realista, uma teia complexa de inúmeras ontologias e vocabulários construídos por usuários sem uma organização rigorosa e por isto apresentam-se de grande relevância investigações sobre a necessidade do uso de agentes que de alguma forma possam garimpar e organizar a informação ali contida.

#### 4.1.Ontologia

A palavra ontologia vem de *ontos* (ser, ente) e *logos* (saber, doutrina), e de maneira estrita significa o “estudo do ser”. Surgiu do estudo de filósofos, ainda na época de Aristóteles, e era usada neste contexto para fazer uma abordagem do ser enquanto ser, ou seja do ser de uma maneira geral. Mais tarde ainda na filosofia, o termo ontologia passou a ser mais usado para saber aquilo que é fundamental ou irreduzível, comum a todos os seres.

Dentro da Ciência da Computação, Guarino (1998) diz que a ontologia é uma teoria lógica que representa um vocabulário pretendido, ou seja, é uma contextualização de algo particular existente no mundo. Neste sentido observa-se que com uma ontologia você consegue definir contextos e domínios particulares do mundo.

Gruber (1993) diz que em um contexto de múltiplos agentes, a ontologia poderia definir o contexto, o vocabulário daquele domínio, servindo assim de base para a comunicação entre os agentes, e para conseguir fazer suas extrações no conhecimento em que eles estão presentes. Gruber ainda diz que a ontologia é uma especificação explícita de uma conceitualização.

Posteriormente Borst (1997) complementa a definição de Gruber dizendo que a ontologia é uma especificação formal de uma conceitualização compartilhada. Desta maneira, um dos principais objetivos da ontologia é o compartilhamento para o reuso destas informações.

Segundo Santarém Segundo (2010), a Ciência da Computação utilizou a ontologia quando se refere a aquisição de conhecimentos a partir de dados semiestruturados, utilizando da ontologia para aplicar técnicas e métodos, para processar as informações.

Santarém Segundo ainda diz que as ontologias vêm com o principal objetivo de ter um vocabulário compartilhado, onde essas informações possam ser trocadas, e usadas para outros usuários. Sendo que estes usuários são tanto seres humanos quanto agentes inteligentes.

Partindo disto, Guarino (1997) diferencia os tipos de ontologia, de acordo com sua utilização:

- Ontologia de topo (*top-level ontology*): tem uma função de descrever conceitos gerais, como o tempo, objeto, matéria, e que não estão dentro de um problema ou domínio particular. É aplicado na conceitualização de conceitos muito extensos e utilizados em grandes comunidades de usuários;
- Ontologia de domínio (*domain ontology*): já tem uma função de descrever conceitos de um domínio particular. São exemplos disto, áreas do conhecimento, como medicina, ciência da computação, entre outros;

- Ontologia de tarefa (*task ontology*): resolvem uma tarefa (um problema) dentro de um domínio. Ou seja, dentro de um domínio, trata de algo específico, como uma doença dentro da medicina, ou compra e vendas de veículos.
- Ontologia de aplicação (*application ontology*): descrevem conceitos tanto de um domínio específico quanto de uma tarefa, que são especializações de ambas as ontologias. Estes conceitos correspondem a papéis desempenhados por entidades de domínio durante a execução de uma atividade.

Berners-Lee (2001b) diz que para uma semântica dentro da web funcione, é importante que as máquinas tenham acesso a coleções estruturadas de informações e que tenham regras de inferências que conduzam a máquina no processo de busca automatizada.

Dentro deste processo, a ontologia aparece como uma solução, pois a ontologia, conforme visto nos conceitos apresentados, será um conjunto estruturado de informações.

#### 4.1.1. Metodologias de Construção da Ontologia

Várias metodologias foram desenvolvidas para fazer a construção da ontologia, ou seja a engenharia da ontologia.

Falbo (1998) diz que independente do domínio, a construção de uma ontologia é uma tarefa bastante complexa, e a partir disto, alguns mecanismos de decomposição são necessários para facilitar este processo.

É interessante notar que não existe uma metodologia definida de como se deve construir uma ontologia, não existindo um consenso de qual metodologia se deva utilizar, assim, normalmente os desenvolvedores acabam fazendo sua própria metodologia (MARTIMIANO, 2006).

Para a proposta de construção da ontologia deste trabalho, foi utilizada a Metodologia definida por Noy e McGuinness (2001), que explicam uma forma de se realizar a engenharia da ontologia.

#### 4.1.2 Metodologia de Noy

Noy (2001) definiu como deve ser o processo da construção da ontologia, para que esta ontologia não seja falha, e não apresente defeitos durante o seu funcionamento.

Explica os sete passos que são necessários para a construção de uma ontologia, esses passos estão descritos abaixo:

1. Determinar o Domínio e o Escopo da Ontologia: este momento é fundamental para se ter um escopo bem definido do que a ontologia irá representar. Neste passo algumas perguntas devem ser respondidas como: “qual domínio a ontologia irá cobrir?”, “quem irá usar a ontologia?”, “quem irá usar e manter a ontologia?”.
2. Reutilizar Ontologias Existentes: como definido em alguns conceitos, a ontologia se refere ao uso compartilhado de algumas informações. Logo um dos passos para a construção de uma ontologia, é a verificação da existência de alguma ontologia semelhante construída, e reaproveita-la no projeto, apenas a melhorando ou a adaptando. Neste contexto, existem algumas bibliotecas para o compartilhamento de ontologias, como “Ontolingua Ontology Library” (<http://www.ksl.stanford.edu/software/ontolingua/>) ou a “DAML Ontology Library” (<http://www.daml.org/ontologies/>)

3. Levantar termos importantes: é fundamental para a construção da ontologia, fazer um levantamento de vários termos que são importantes para aquele cenário, para ter uma base de quais serão as classes, propriedades e hierarquia da ontologia.
4. Definir classes e sua hierarquia: Uschold e Gruninger (1996) definem que existem diversas maneiras para desenvolver uma hierarquia de classes da ontologia. Como a *top-down*, onde a modelagem começa dos conceitos mais gerais, e posteriormente a construção da hierarquia dos conceitos abaixo. Também existe o *bottom-up*, onde o processo de desenvolvimento começa dos conceitos mais específicos, e depois vai para os conceitos mais genéricos. E existe uma terceira forma, chamado de combinação, onde este processo é uma combinação entre o método de *bottom-up* e *top-down*, sendo definidos os conceitos mais importantes a princípio, e depois, feito uma generalização e uma especificação destes conceitos. O método para a construção deve ser escolhido segundo a necessidade e o conhecimento acerca do domínio.
5. Definir propriedades das classes: Nesta etapa devem-se observar os termos listados no passo três, e então, observar quais termos são propriedades de alguma classe, e assim encaixá-las dentro desta.
6. Restrições das Propriedades: Neste passo devem-se observar as restrições de cada propriedade, como cardinalidade e os tipos de dados das propriedades.
7. Criação de instâncias: aqui, é criado objetos (instâncias) das classes, pois muitas vezes estes valores já devem estar definidos dentro da ontologia. Assim, deve ser criado um objeto da classe e ser preenchido as propriedades da classe, cuja instância foi criada.

A partir destes passos, é possível construir uma ontologia que siga regras, e tenha uma boa consistência.

Ontologias são construídas com base na representação da informação constituída pelo conteúdo e sua semântica. No contexto da Web atual, a complexidade das informações gera a necessidade de estudos sobre as características tanto dos processos de geração, tratamento, processamento e armazenamento, quanto da variedade da natureza dessas informações. Neste contexto, surge o conceito de Big Data, em que elementos como velocidade, volume, variedade, valor e veracidade são variáveis que devem ser estudadas e investigadas como requisitos funcionais para projetos de ambientes informacionais digitais.

## 5. A INFORMAÇÃO NA ERA DO BIG DATA

Com o crescimento exponencial das informações contidas na Web e sua complexidade, ambientes informacionais digitais se deparam com um novo desafio: como conseguir tratar os processos informacionais de forma eficiente e desta maneira, por meio dessas informações, apoiar processos humanos e organizacionais num cenário em que a velocidade, o volume, a variedade das informações cresce a cada dia, e há ainda a necessidade de se levar em conta o valor e a veracidade da informação encontrada. Para entender este processo, é necessário entender a questão do Big Data, e como este processo está mudando a maneira como se lida com as informações no contexto da Web.

## 5.1. Big Data

Vive-se um momento de grande geração e uso das informações geradas de forma online. Esses dados são gerados por e-mails, compartilhamento de informações por redes sociais, transações online, sensores, celulares, GPS, entre vários outros meios.

Para verificar este movimento, foi identificado que até 2003 toda a humanidade criou cerca de 5 exabytes de informações, atualmente esta quantidade de informação é gerada em menos de dois dias (SAGIROGLU E SINANC, 2013).

Zikopoulos e Eaton (2012) definem de forma sintática que Big Data se aplica a informações que não podem ser processadas ou analisadas com as ferramentas e os métodos tradicionais. E diz ainda que a era do Big Data é resultado das mudanças que tem ocorrido no mundo, onde através dos avanços das tecnologias, foi possível que várias pessoas e programas se intercomunicassem não somente num intervalo de tempo, mas sim durante todo o tempo.

O termo Big Data não diz respeito somente ao aspecto de armazenamento de dado, mas também a outros aspectos como a velocidade em que os dados podem ser capturados e processados, quase que em tempo real, dando assim, vantagens competitivas as organizações (MCAFEE, 2012).

A seguir são mostrados os três aspectos que McAfee considera principais neste processo que são o volume, a velocidade e a variedade:

- Volume: O número de dados gerados todos os dias na Web ultrapassam 2,5 exabytes, e este número tem dobrado a cada quarenta meses, em breve a unidade de medida de dados será o zetabytes. Algo que faz com que este volume cresça de forma exponencial, é que a própria Web fornece possibilidades para uma criação de novas informações, como as redes sociais, onde o usuário acaba gerando cada vez mais dados (ZIKOPOULOS E EATON, 2012). Todas as informações geradas são armazenadas, como dados financeiros, médicos, compras realizadas na internet, conversas realizados pelos mensageiros, gerando assim um volume extremamente grande de dados.
- Velocidade: Muitas vezes a velocidade com que o dado é criado e processado é fundamental, pois estes dados podem ser utilizados em tempo real. Esta velocidade não está ligada somente a entrada de dados, mas também a velocidade do fluxo de dados. Ou seja, ter velocidade em conseguir acompanhar a geração e a demanda das requisições das informações.
- Variedade: as formas que os dados estão armazenados são cada vez mais diversas, entre elas textos, músicas, vídeos, imagens. E isto promove que não exista um padrão com que os dados são gerados e/ou armazenados. Segundo Kakhani (2013) os dados podem ser não-estruturados, semiestruturados ou estruturados, sendo de uma natureza heterogênea, pois os dados podem vir de mídias sociais, de blogs, sendo desde textos não-estruturados, à vídeos e fotos, não tendo, assim, uma estrutura fixa e definida.

Posteriormente a McAfee, alguns autores (KAKHANI, 2013) (KATAL, 2013) também incluíram outros dois conceitos que tem importância, para a definição de Big Data, que é a veracidade e o valor:

- Veracidade: todos os dados presentes neste universo podem ser das mais diversas naturezas, portanto é necessário que se tenha dados que sejam verdadeiros, para não trazer informações equivocadas, ao final de um estudo (KAKHANI, 2013).

- Valor: a partir de dados e informações que foram fornecidas ou adquiridas pelos sistemas, pode se chegar a resultados de muito valor, pois podem demonstrar tendências do mercado, que pode levar aos administradores das empresas a tomarem medidas para mudar ou readequar as estratégias comerciais (KATAL, 2013).

O processo do Big Data aparece pelo grande crescimento do uso e da geração da informação, onde a mudança quantitativa (grande crescimento de dados) trouxe uma mudança qualitativa das informações (informações cada vez mais precisas) (MAYER-SCHÖNBERGER E CUKIER, 2013).

Esta mudança quantitativa significa fazer as análises das informações levando em consideração todo o banco de dados existente. Pois antes do atual momento da tecnologia, as análises realizadas, eram feitas quase que totalmente em cima de amostras, o que pode trazer um grande risco, de se ter informações que não são reais, por maior o cuidado que exista para esta amostra representar uma população real, isto é um processo que pode ter falhas.

Esta técnica de amostragem é fruto de um momento, onde não existiam tecnologias para conseguir processar e analisar todas as informações existentes. Hoje este conceito já não faz tanto sentido, pois atualmente é possível processar quantidades enormes de informações em segundos, podendo assim, usar todas as informações do banco de dados, não necessitando pegar uma amostra deste.

Como consequência disto, os resultados destas análises, que agora utilizam o todo, passaram a ser muito mais precisas, e oferecerem dados e informações que antes seria muito difícil de obter.

Verificando todo este movimento, algumas ciências, como a astronomia e a genômica, passaram a utilizar disto e deram o nome de Big Data para este processo.

O uso do Big Data pode ser visualizado, quando se necessita trabalhar com grandes escalas de dados para se extrair novas ideias e criar novas formas de valor que alterem mercados, governos, organizações, entre outros (MAYER-SCHÖNBERGER E CUKIER, 2013).

Este processo pode ser percebido nas grandes corporações da Internet, como Google que processa mais de 24 pentabytes de dados por dia, e o Facebook que recebe mais de dez milhões de fotos a cada hora (MAYER-SCHÖNBERGER E CUKIER, 2013).

Neste contexto, as informações nos ambientes de Big Data são importantes insumos nos processos organizacionais como Sistemas de Inteligência Competitiva no Apoio à Decisão e à Inovação.

Para entendimento dos requisitos desses sistemas, é apresentado o conceito e o cenário da inovação no Brasil, estes requisitos serão considerados para a proposta da plataforma apresentada nesta pesquisa.

## 6. INOVAÇÃO

A inovação é prioridade em países desenvolvidos que têm como estratégia de crescimento e competitividade a diferenciação na pesquisa e desenvolvimento de novas tecnologias. A cultura de inovação de um país ou região depende da articulação dos principais atores da inovação, governo, empresas e instituições de ensino e pesquisa, que tem papéis distintos, porém colaborativos.

A inovação é conceituada por muitos autores, porém, por consenso, estabelece-se que esta somente se materializa quando pode ser explorada com sucesso. A inovação em

suas diversas facetas e aplicações é fomentada em países desenvolvidos e países emergentes de forma contundente.

Segundo World Bank (2015), o Brasil investe discretamente em inovação. Em 2011, o país investiu 1,2% do PIB em inovação, enquanto países como EUA e Coreia do Sul investiram 2,8% e 4%, respectivamente.

Em valores nominais essa diferença pode ser notada com maior evidência. Enquanto o Brasil investiu em bilhões de dólares 19,4, os EUA e Coreia do Sul investiram 405,3 e 65,4 por ano, respectivamente.

A conclusão da pesquisa sobre inovação da Confederação Nacional da Indústria (CNI, 2015) brasileira mostra que 62% dos empresários consideram baixo ou muito baixo o grau de investimento em inovação da própria indústria.

O principal motivo para o Brasil estar atrasado em relação a outros países é o reflexo de defasagem tecnológica acumulada ao longo dos anos. A consequência disso é que a indústria, muitas vezes, acaba por importar ou copiar o que é feito em outros países. De acordo com os entrevistados, falta cultura de inovação nas empresas brasileiras em geral. Eles também elencaram como entraves a falta de políticas de incentivo, a dificuldade de interação entre empresas e universidades e o baixo nível de educação dos profissionais (CNI, 2015).

Esta pesquisa da CNI mostra que a indústria brasileira não desenvolveu em seu DNA a cultura da inovação. No entanto, esta deficiência se estende por esferas governamentais e universidades, dificultando a expansão de iniciativas inovadoras que possam obter sucesso.

Um dos pontos negativos também enfatizados pela CNI foi a dificuldade de interação entre as empresas e as universidades, porém deve ser incluído nessa relação os órgãos governamentais, tornando este cenário ainda mais complexo e desafiador.

No entanto é importante destacar que justamente esta relação bem sucedida entre governo, iniciativa privada e universidades é a base da cultura de inovação de países desenvolvidos. Assim, estes são nomeados neste artigo como os atores de inovação.

Desta forma, promover uma cultura inovadora é um legado que deve ser construído com esforço colaborativo entre os atores de inovação (governo, empresas e instituições de ensino e pesquisa).

Destaca-se o cenário atual de inovação no Brasil e como uma metodologia de fluxo informacional apoiada pela Ciência da Computação e a Ciência da Informação podem contribuir para o crescimento do país em relação a sua cultura de inovação.

O fluxo informacional em uma organização ou entre organizações pode ocorrer empiricamente e/ou estabelecido por regras apoiadas em pessoas, organogramas, hierarquias, dependências e relações. Os fluxos de informação permitem o estabelecimento das etapas de obtenção, tratamento, armazenamento, distribuição, disseminação e uso da informação no contexto organizacional (SANTOS, 2014).

Tratado como elemento estratégico para tomadas de decisões, reorganização e otimização de processos organizacionais, o fluxo informacional pode agregar valor a um conjunto de informações aparentemente correlatas, porém desorganizadas ou não estruturadas.



## 6.1. Cenário Brasileiro de Inovação

Em uma análise superficial, os valores nominais investidos em inovação no Brasil não representam a sua 7ª posição no ranking econômico mundial. No entanto, nos últimos 5 (cinco) anos o país reagiu criando uma infraestrutura e programas para o crescimento da cultura de inovação.

Para fundamentar e contextualizar esse cenário destaca-se a emenda nº 85 de fevereiro de 2015, que inclui na Constituição brasileira a priorização do Estado em relação à ciência, tecnologia e inovação. A seguir, a citação de alguns artigos desta emenda constitucional que estão fortemente relacionados com este projeto:

Art. 213. § 2º As atividades de pesquisa, de extensão e de estímulo e fomento à inovação realizadas por universidades e/ou por instituições de educação profissional e tecnológica poderão receber apoio financeiro do Poder Público.

Art. 218. O Estado promoverá e incentivará o desenvolvimento científico, a pesquisa, a capacitação científica e tecnológica e a inovação. § 6º O Estado, [...] estimulará a articulação entre entes, tanto públicos quanto privados, nas diversas esferas de governo.

"Art. 219. Parágrafo único. O Estado estimulará a formação e o fortalecimento da inovação nas empresas, bem como nos demais entes, públicos ou privados, a constituição e a manutenção de parques e polos tecnológicos e de demais ambientes promotores da inovação, a atuação dos inventores independentes e a criação, absorção, difusão e transferência de tecnologia.

Art. 219-A. A União, os Estados, o Distrito Federal e os Municípios poderão firmar instrumentos de cooperação com órgãos e entidades públicos e com entidades privadas, inclusive para o compartilhamento de recursos humanos especializados e capacidade instalada [...]

Art. 219-B. O Sistema Nacional de Ciência, Tecnologia e Inovação (SNCTI) será organizado em regime de colaboração entre entes, tanto públicos quanto privados, com vistas a promover o desenvolvimento científico e tecnológico e a inovação.

Estes artigos mostram a preocupação do governo em articular os atores de inovação e inclui nessa meta a União, os Estados, e os Municípios, assim como a iniciativa privada e universidades.

Para consolidar a prioridade da inovação, leis, órgãos e programas governamentais estão sendo criados e/ou reestruturados para se submeter a esse propósito. A seguir destacam-se alguns destes:

- Marco da Ciência Tecnologia e Inovação: regulamenta a Emenda Constitucional 85, definindo parcerias de longo prazo entre os setores público e privado, dá maior flexibilidade de atuação às Instituições Científicas, Tecnológicas e de Inovação (ICTs) e respectivas entidades de apoio.
- Programa Nacional de Apoio às Incubadoras de Empresas e aos Parques Tecnológicos (PNI) do Ministério da Ciência, Tecnologia e Inovação (MCTI).
- Financiadora de Estudos e Projetos (FINEP), empresa pública criada em julho de 1967 e vinculada ao MCTI. Tem como missão “Promover o

desenvolvimento econômico e social do Brasil por meio do fomento público à Ciência, Tecnologia e Inovação em empresas, universidades, institutos tecnológicos e outras instituições públicas ou privadas.”

- Política Industrial, Tecnológica e de Comércio Exterior – PITCE - A PITCE379 foi instituída com o objetivo de aumentar a competitividade das empresas brasileiras, mediante elevação dos níveis de eficiência e produtividade, fomento à capacidade inovadora e estímulo às exportações.
- Lei de Inovação e Lei do Bem: (a) incentiva parcerias em P&D entre universidades, instituições de pesquisa e empresas; (b) regula a transferência de tecnologia e a criação de incubadoras; (c) permite compartilhar equipamentos, infraestrutura e pessoal em atividades de desenvolvimento de novas tecnologias; e (d) estabelece subsídios e recursos para tais atividades.
- Conselho Nacional e Agência Brasileira de Desenvolvimento Industrial: Para aprimorar a coordenação institucional e, principalmente, incentivar a inovação e os gastos das empresas privadas em P&D, foram estabelecidos dois novos órgãos: o Conselho Nacional de Desenvolvimento Industrial (CNDI) e a Agência Brasileira de Desenvolvimento Industrial (ABDI).

Já na esfera estadual, os Estados e municípios têm autonomia para criar seus mecanismos de incentivo à inovação. Neste cenário, o governo do Estado de São Paulo, pelo decreto 60.286, de 25 de março de 2014, instituiu e regulamentou o Sistema Paulista de Ambientes de Inovação (SPAI). O SPAI compreende o Sistema Paulista de Parques Tecnológicos (SPTec), a Rede Paulista de Incubadoras de Empresas de Base Tecnológica (RPITec), a Rede Paulista de Centros de Inovação Tecnológica (RPCITec) e a Rede Paulista de Núcleos de Inovação Tecnológica (RPNIT).

A regulamentação do SPAI estabeleceu requisitos para instalação de incubadoras de base tecnológica, centros de inovação, parques tecnológicos e núcleos de inovação tecnológica com apoio do Estado que devem explorar a relação entre os atores de inovação (governo, empresas e universidades) em prol da pesquisa, desenvolvimento e inovação.

O domínio de aplicação da plataforma computacional e informacional desta pesquisa será norteado e apoiado em modelos e estruturas formais, articuladas pelos governos estadual e nacional, como é o caso do SPAI no Estado de São Paulo e do Programa Nacional de Apoio às Incubadoras de Empresas e aos Parques Tecnológicos (PNI) do MCTI. Delimitando a abrangência da aplicação da proposta da plataforma ao escopo de parques tecnológicos, centros e núcleos de inovação tecnológica, incubadoras de empresas e laboratórios de pesquisa e inovação.

O Brasil atribuiu recentemente às suas estratégias de crescimento o investimento em inovação. A criação de leis de incentivos, programas governamentais de tecnologia e inovação, fundos de investimento, programas de formação de pesquisadores, internacionalização das universidades são algumas das ações realizadas, bem como a visão estratégica de que a inovação se dá pela articulação dos atores constituídos pelo governo, empresas e universidades, agregando valor e potencializando as ações e os investimentos realizados.

## 7. PLATAFORMA DE INTELIGÊNCIA COMPETITIVA BASEADA EM AMBIENTES INFORMACIONAIS SÊMANTICOS NO CONTEXTO DE BIG DATA: MODELO INFORMACIONAL DE APOIO À INOVAÇÃO

Apresentados os principais conceitos a serem utilizados na proposta de plataforma de apoio e tomada de decisão no processo de inovação nas organizações, a seguir é apresentada a articulação computacional e informacional para geração de uma arquitetura sistemática dos fluxos informacionais no cenário da inovação.

### 7.1.Arquitetura

Os fluxos informacionais caracterizam-se pelos processos de busca, acesso, recuperação, transformação, processamento, representação, persistência, transferência, apresentação e uso da informação. Neste contexto, com base nas fontes e produtos informacionais gerados e consumidos pelos principais atores da inovação citados anteriormente, são utilizadas tecnologias computacionais, informacionais e ontológicas para apoiar o processo de gestão e a tomada de decisão no cenário da inovação.

Assim, é proposta a criação de uma arquitetura de Inteligência Competitiva de apoio à inovação como pode ser visto na Figura 2.

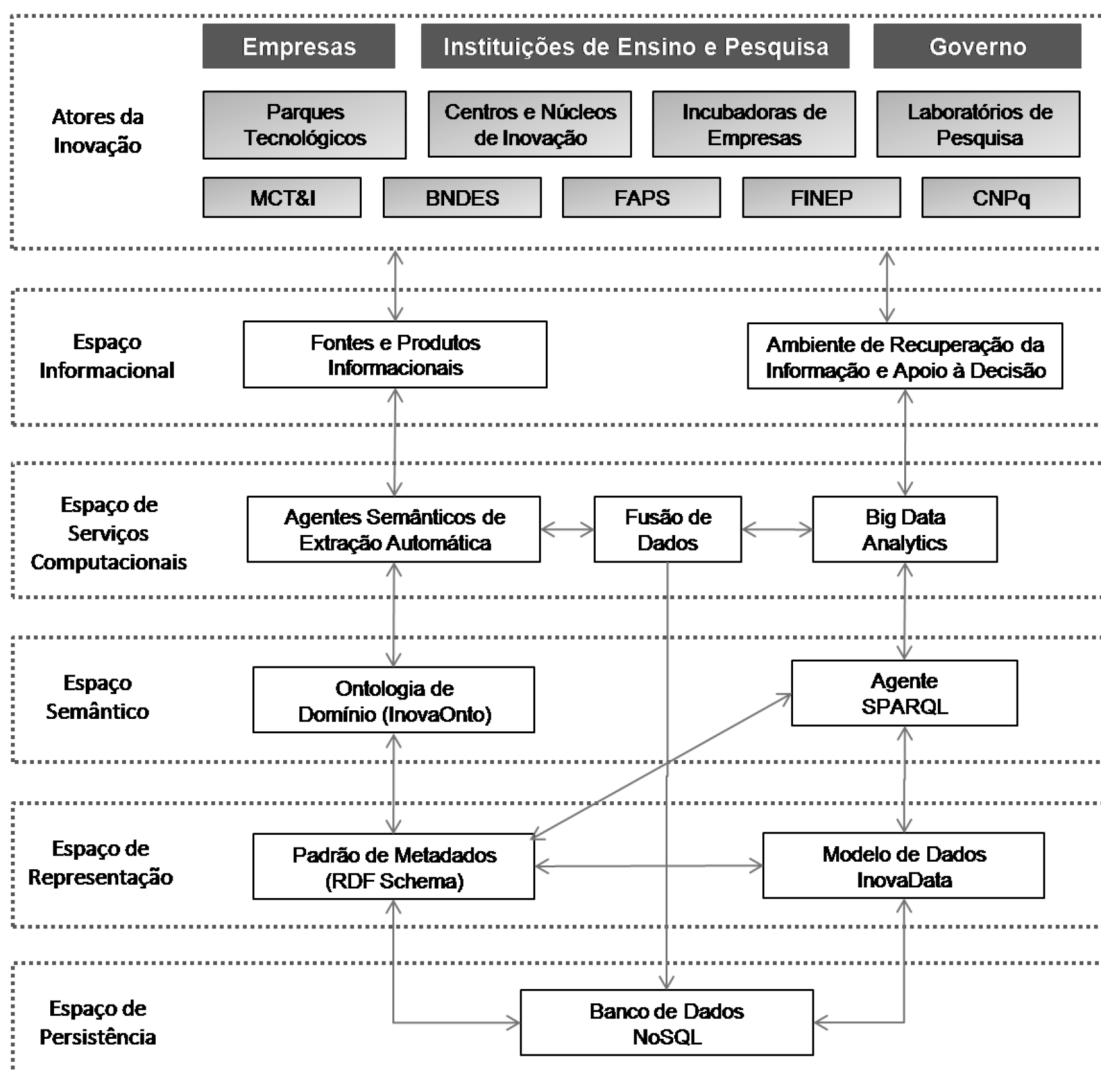


Figura 2: Arquitetura Proposta de Inteligência Competitiva de Apoio à Inovação

O cenário da inovação é composto por diversos atores que disponibilizam e consomem informações dinâmicas e heterogêneas que demandam a utilização de conceitos da Inteligência Competitiva, agentes computacionais, Web Semântica, Big Data, entre outras tecnologias computacionais e informacionais.

A arquitetura proposta considera a premissa da colaboração dos principais atores nos processos da disponibilização da informação para suportar todo o fluxo informacional no qual esta plataforma se baseia.

### **7.1.1. Atores da Inovação**

O cenário utilizado como estudo para a proposta desta pesquisa é o ambiente de inovação das organizações que tem como principais atores o governo que cria as leis de incentivo e as políticas públicas da inovação no país por meio de órgãos de fomento à pesquisa à inovação. As leis de acesso à informação fazem com que o acesso pela plataforma aos produtos de informação seja facilitado e viabilizado para ajudar a abastecer todo o fluxo informacional proposto.

As instituições de ensino e pesquisa também exercem um papel fundamental no processo de apoio à inovação nas organizações por meio das pesquisas científicas e tecnológicas e da formação de mão de obra qualificada que são empregadas nos principais projetos de P&D&I.

### **7.1.2. Espaço Informacional**

Os atores da inovação, empresas, governo e instituições de ensino, geram e consomem produtos de informação tanto no contexto individual quanto colaborativo e servem de retro-alimentação para os processos de inovação nas organizações.

Os produtos de informação gerados e consumidos nos ambientes de inovação pelos principais atores de inovação são: políticas, leis e projetos de lei, editais de fomento e apoio, programas de incentivo, eventos, projetos de P&D&I, patentes, parcerias, transferência tecnológica, relatórios técnicos, produtos, serviços, habilidades e competências.

Neste contexto, é proposta a utilização da Inteligência Competitiva para os processos de entrada da informação e no uso na disponibilização de um sistema de recuperação da informação que apoie os processos de gestão e tomada de decisão dos principais atores da inovação.

### **7.1.3. Espaço de Serviços Computacionais**

Como descritos anteriormente, os Agentes de Extração de Informações exercem um papel estratégico em ambientes de Inteligência Competitiva, devido ao seu poder de processamento computacional para extrair informações da Web.

Para acrescentar relevância ao processo de extração, agregam-se aos agentes computacionais, conceitos semânticos e, neste contexto, os agentes de extração farão uso de uma ontologia de domínio que será utilizada com base na construção do conhecimento sobre conceitos do cenário de inovação.

A implementação consiste na integração dos agentes de extração com a ontologia, ou seja, a comunicação das informações que são extraídas, com o intuito de dar semântica a busca. Desta maneira, o agente extrai um texto de uma página e um algoritmo irá avaliar se aquela informação está dentro do contexto da ontologia e se aquela

informação de fato será útil para o usuário.

Ambientes de Inteligência Competitiva também se caracterizam por diversas fontes informacionais que apresentam produtos de informação heterogêneos e sem padronização. Neste contexto, o uso de processos da Fusão de Dados e Informações permite que a informação seja padronizada antes de ser persistida e utilizada.

A Fusão de Dados e Informações é a rotina de transformação de dados e informações para produzir estimativas e predições de estados de entidades, visando maximizar o valor da informação e estimular a consciência situacional de analistas sobre de um ambiente de interesse (BOTEGA, 2016).

Comumente empregado como sistema dedicado a suportar a avaliação de situações e tomada de decisão em sistemas complexos, os mesmos praticam a redução da dimensionalidade dos dados, agregação de valor à informação, aumento da representatividade e a produção de subsídios para a construção do conhecimento sobre situações (STEINBERG ET AL, 1999) (BLASCH ET AL, 2013).

A Fusão de Dados e Informações envolve um processo multinível e multifacetado de combinação de dados de múltiplas fontes, como objetivo de adquirir e processar informações que devem ser mais significativas do que se adquiridas por fontes de dados individuais, ou seja, sem fusão (BOTEGA, 2016).

Tal área é um importante campo de estudo multidisciplinar, devido principalmente ao aumento do fluxo de dados e informações, a melhoria da infraestrutura de comunicação, computação e tecnologia de sensores. Técnicas para combinar ou fundir dados vem de um diverso conjunto de disciplinas tradicionais: processamento digital de sinais, estimativa estatística, teoria de controle, inteligência artificial e métodos numéricos clássicos.

O resultado da extração e da fusão gera uma ampla e volumosa base de dados contendo uma variedade de tipos de dados. Esse tipo de base de dados dificulta a análise e a geração de produtos de informação que apoiem com eficiência processos de tomada de decisão e inovação.

Assim, faz-se necessário a utilização de um processo computacional que permita examinar o valor e a veracidade desse grande volume e variedade de dados gerados em alta velocidade (Big Data).

Por este motivo, propõe-se o uso do Big Analytics, processo de examinar grandes conjuntos de dados contendo uma alta variedade de tipos de dados, para descobrir padrões escondidos, correlações desconhecidas, tendências de mercado e outras informações úteis sobre as organizações. Os resultados analíticos podem levar a processos mais eficazes de tomada de decisão e de apoio à inovação.

#### **7.1.4. Espaço Semântico**

O espaço semântico tem a função de inserir semânticas nas buscas realizadas pelo agente de busca. Sendo possível que a busca realizada pelo agente, leve em consideração o contexto na qual aquela informação está inserida.

Esta semântica pode ser alcançada por meio do uso de uma estrutura ontológica, que analisa o domínio do contexto que se deseja buscar as informações.

A base de aplicações da Web Semântica reside no RDF (*Resource Description Framework*) para fornecer um meio de ligar dados de múltiplas fontes informacionais com a linguagem de consulta SPARQL, padrão como SQL para consultar dados RDF. Assim essas aplicações podem dados persistidos em grafos baseados em RDF nativo e extraí-los de bancos de dados tradicionais.

### 7.1.5. Espaço de Representação

RDF e OWL (*Web Ontology Language*) são a base das linguagens de representação na Web Semântica, com o RDF servindo como base para aplicações semânticas. O RDF aborda uma questão fundamental na Web Semântica: gerenciamento de dados distribuídos. Todos os outros padrões da Web Semântica são construídos sobre esta base de dados distribuídos. Aplicações construídas sobre o RDF dependem muito da infraestrutura da Web, usando muitas das suas características familiares e comprovadas, enquanto estendendo-os para fornecer uma base para uma rede distribuída de dados.

Enquanto representações em RDF serão utilizadas para persistir informações sobre as informações distribuídas na Web e de relevância ao domínio da área de Inovação representado na ontologia de domínio, um modelo de dados de mais alto nível será utilizado para representar os metadados das informações contidas no espaço de persistência.

### 7.1.6. Espaço de Persistência

No espaço de persistência ocorre o armazenamento das informações que são extraídas pelo agente de busca e pelo processo de fusão. Este armazenamento pode ocorrer tanto em Sistemas Gerenciadores de Banco de Dados relacionais, quanto em Banco de Dados NoSQL (*Not Only SQL* – Não apenas SQL). A persistência pode ocorrer destas duas maneiras pelo fato de que as informações extraídas podem ser, tanto dados estruturados quanto dados não estruturados.

Desta forma os dados estruturados podem ser armazenados em cima de bancos de dados relacionais, que apresentam regras bem definidas, e consegue dar uma integralidade maior aos dados que são armazenados.

Já no contexto da Web, existe uma grande quantidade de dados que são semiestruturados ou não-estruturados, ou seja, não seguem regras, ou não tem uma estrutura exatamente definida. Assim, é necessário que estes dados sejam armazenados em bancos de dados NoSQL, que apresentam características de não ter um esquema totalmente definido, e que permite uma flexibilidade maior ao armazenar estas informações.

Do ponto de vista das questões semânticas, tanto a ontologia quanto os grafos RDF necessitam de tecnologias de persistência que não são atendidas por bancos de dados relacionais, neste contexto, a plataforma necessita de uma tecnologia computacional de persistência que permita o armazenamento de dados semânticos e ontológicos.

A utilização da combinação das tecnologias computacionais, informacionais e semânticas dos espaços contidos na plataforma que foram descritos anteriormente, permite que os principais atores do domínio da inovação (governo, empresas e instituições de ensino) possam ter acesso aos produtos de informações gerados e colaborativamente apoiarem-se em processos de construção coletiva colaborativa de novos conhecimentos e em ações de incentivo à inovação nas organizações.

## 8. CONSIDERAÇÕES FINAIS

Vivencia-se uma nova disrupção tecnológica pela convergência da colaboração, mobilidade e grande volume de dados (Big Data). O grande desafio para a pesquisa de sistemas computacionais e para a forma de uso das informações nas organizações é promover a integração destas tecnologias para balancear as necessidades de geração, acesso e controle destas informações, bem como as oportunidades deste comportamento emergente e suas inovações. A Inteligência Competitiva figura como possibilidade de se utilizar esse grande volume de informações para apoiar as ações de apoio à decisão e a inovação nas organizações, a partir de processos informacionais sistemáticos para sustentar tais ações.

Este cenário em que as organizações estão inseridas exige uma evolução dos processos de gestão e de construção contínua de novos conhecimentos na geração da inovação tecnológica, ou seja, de seus produtos e serviços. Como resultado do uso da informação e da construção do conhecimento na promoção da inovação, há a geração de resultados a longo prazo, ou seja, considerando que as inovações são capazes de gerar vantagens competitivas, a médio e longo prazo, inovar torna-se essencial para que as organizações se mantenham estáveis e sustentáveis na sociedade contemporânea.

O monitoramento do ambiente externo torna-se complexo se realizado de forma manual e por pessoas, devido à amplitude e ao dinamismo informacional dos Ambientes Informacionais Digitais identificados como Big Data. Neste contexto, são utilizados os Agentes de Extração Automática, que fazem o monitoramento e a extração das fontes informacionais da Web.

Este artigo apresentou uma proposta de plataforma de Inteligência Competitiva de apoio à inovação, a partir da combinação das tecnologias computacionais (como Agentes de Extração, Fusão de Dados, Big Data Analytics e Bancos de Dados NoSql), com as tecnologias informacionais semânticas (ontologias, padrões de metadados, Recuperação da Informação), na construção de uma arquitetura de apoio ao processo de gestão e a tomada de decisão no cenário da inovação.

A inovação é prioridade em países desenvolvidos que têm como estratégia de crescimento e competitividade a diferenciação na pesquisa e desenvolvimento de novas tecnologias. A cultura de inovação de um país ou região depende da articulação dos principais atores da inovação, governo, empresas e instituições de ensino e pesquisa, que tem papéis distintos, porém colaborativos.

O Brasil atribuiu recentemente às suas estratégias de crescimento o investimento em inovação. A criação de leis de incentivos, programas governamentais de tecnologia e inovação, fundos de investimento, programas de formação de pesquisadores, internacionalização das universidades são algumas das ações realizadas, bem como a visão estratégica de que a inovação se dá pela articulação dos atores constituídos pelo governo, empresas e universidades, agregando valor e potencializando as ações e os investimentos realizados.

A utilização da combinação das tecnologias computacionais, informacionais e semânticas dos espaços contidos na plataforma que foram descritos anteriormente, permite que os principais atores do domínio da inovação possam ter acesso aos produtos de informações gerados e colaborativamente apoiarem-se em processos de construção coletiva colaborativa de novos conhecimentos e em ações de incentivo à inovação nas organizações.

## 9. REFERÊNCIAS BIBLIOGRÁFICAS

- ABRAIC. Associação dos Analistas de Inteligência Competitiva. Disponível em <<http://www.abraic.org.br>>. Acesso em 25 de maio de 2015.
- ÁLVAREZ, Alberto Cáceres. Extração de informação de artigos científicos: uma abordagem baseada em indução de regras de etiquetagem. 2007. Tese de Doutorado. Universidade de São Paulo.
- BAEZA-YATES, R.; RIBEIRO-NETO, B. (1999) Modern information retrieval. New York: ACM; Harlow: Addison-Wesley.
- BERNERS-LEE, T. (1989) Information Management: A Proposal. Disponível em <<http://www.w3.org/History/1989/proposal.html>>. Acesso em 09 de julho de 2015.
- BERNERS-LEE, T., HENDLER, J. E LASSILA, O. (2001b) The semantic web. Scientific american 284.5. 28-37.
- BERNERS-LEE, T., LASSILA, O. E HENDLER, J. (2001a) The semantic web. Scientific American, New York, v. 5.
- BEYER, M. A., E LANEY, D. (2012) The importance of ‘big data’: a definition. Stamford, CT: Gartner.
- BLASCH, E.; STEINBERG, A.; DAS, S. (2013) Revisiting the JDL model for information Exploitation. Information Fusion, [S.l.].
- BORST, W. N. (1997) Construction of engineering ontologies for knowledge sharing and reuse. 1997. 227 f. Tese (Doutorado). Centre for Telematics for Information Technology, University of Twente, Enschede.
- BOTEGA, L. (2016) Modelo de Fusão Dirigido por Humanos e Ciente da Qualidade de Informação. Tese de Doutorado. Universidade Federal de São Paulo.
- BRASIL. DECRETO Nº 60.286, DE 25 DE MARÇO DE 2014, Institui e regulamenta o Sistema Paulista de Ambientes de Inovação – SPAI e dá providências correlatas. Publicado na Casa Civil, aos 25 de março de 2014.
- BRASIL. Emenda constitucional Nº 85, 26 fevereiro, 2015. Altera e adiciona dispositivos na Constituição Federal para atualizar o tratamento das atividades de ciência, tecnologia e inovação. DOU 27.2.2015 republicado em 03.03.2015.
- BRASILIANO, A. C. R. (2002) A Fuga Involuntária das Informações Estratégicas nas Empresas: Fragilidades nas Redes Humanas. Disponível em <<http://www.abraic.org.br>>. Acesso em 20 de março de 2014.
- CASTRO, J. M., E ABREU, P. G. (2006) Influência da inteligência competitiva em processos decisórios no ciclo de vida das organizações. Ciência da Informação 35.3. 15-29.
- CONEGLIAN, Caio Saraiva; FUSCO, Elvis. (2015) Recuperação da Informação em Ambientes Semânticos: uma ferramenta aplicada à publicações científicas. Journal on Advances in Theoretical and Applied Informatics, v. 1, n. 1, p. 30-37.
- DE DIANA, M., E GEROSA, M. A. (2010) Nosql na web 2.0: Um estudo comparativo de bancos não-relacionais para armazenamento de dados na web 2.0.
- DETERS, J. I., E ADAIME, S. F. (2003) Um estudo comparativo dos sistemas de busca na web. Anais do V Encontro de Estudantes de Informática do Tocantins. Palmas, TO. 189-200.
- DRUCKER, P. A nova sociedade das organizações. In: Administrando em tempos de mudanças. São Paulo: Pioneira, 1995, pp. 43-57.
- DZIEKANIAK, G. V., E KIRINUS, J. B. (2004) Web semântica. Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação, n.18, p.20-39.



- FALBO, R. A. (1998) Integração de Conhecimento em um Ambiente de Desenvolvimento de Software. 1998. 215 f. Tese (Doutorado em Ciências em Engenharia de Sistemas e Computação) – COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro.
- FERNEDA, E. (2003) Recuperação da Informação: Análise sobre a contribuição da Ciência da Computação para a Ciência da Informação. 2003. 147 f. Tese (Doutorado em Ciência da Informação) – Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo.
- FRANKLIN, S.; GRAESSER, A. (1996) Is it an agent or just a program? a taxonomy from autonomous agents. In: Third International Workshop On Agent Theories, Architectures And Languages: Intelligent Agents III. [s.n.], p.21–36.
- FREITAS, Frederico L. G. Sistemas Multiagentes Cognitivos para recuperação, classificação e extração integradas de informação da WEB. 2002. Tese de Doutorado. Universidade Federal de Santa Catarina, Centro Tecnológico. Programa de Pós-Graduação em Engenharia Elétrica.
- GENESERETH, M. R.; KETCHPEL, S. P. (1994) Software agents. Communications of the ACM, [S.l.], v.38, n.7, p.48–53.
- GRAHAM-ROWE, D., et al. (2008) Big data: science in the petabyte era. Nature 455. 1-50.
- GRUBER, T. R. (1993) A translation approach to portable ontology specifications. Knowledge acquisition 5.2. 199-220.
- GUARINO, N. (1997) Understanding, building and using ontologies. International Journal of Human-Computer Studies 46.2. 293-310.
- GUARINO, N. (1998) Formal ontology in information systems. Proceedings of the first international conference (FOIS'98), June 6-8, Trento, Italy. Vol. 46. IOS press.
- HENDLER, Jim. Agents and Web Semantic, IEEE Intelligent System Journal, 2001. Disponível em: <<http://www.cs.rpi.edu/~hendler/AgentWeb.html>>. Acesso em 14 de setembro de 2015.
- KAHANER, L. (1996) Competitive Intelligence: how to gather, analyze, and use information to move your business to the top. New York: Touchstone.
- KAKHANI, M. K., KAKHANI, S., E BIRADAR, S. R. (2013) Research Issues in Big Data Analytics.
- KATAL, A., WAZID, M., E GOUDAR, R. H. (2013) Big data: Issues, challenges, tools and Good practices. Contemporary Computing (IC3), 2013 Sixth International Conference on. IEEE.
- KAPLAN, Jerry (2015). Humans Need Not Apply: A Guide to Wealth and Work in the Age of Artificial Intelligence, Yale: Yale University Press.
- MARTIMIANO, L. A. F. (2006) Sobre a estruturação de informação em sistemas de segurança computacional. 2006. 185 f. Tese (Doutorado em Ciências em Engenharia de Sistemas e Computação) – COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro.
- MAYER-SCHÖNBERGER, V., E CUKIER, K. (2013) Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt.
- McAFEE, A., ET AL. (2012) Big Data. The management revolution. Harvard Bus Rev 90.10. 61-67.
- MODESTO, L. R. (2013) Representação e Persistência para acesso a Recursos Informacionais Digitais gerados dinamicamente em sites oficiais do Governo Federal. 2013. 103 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília.

- MOOERS, C. (1951) Zatocoding applied to mechanical organization of knowledge. American Documentation, Washington, v. 2, n. 1, p.20-32.
- MORAIS, E. F. C. (1999). Inteligência competitiva: estratégias para pequenas empresas. Brasília: Centro de APOIO ao Desenvolvimento Tecnológico, Universidade de Brasília.
- MINSKY, Marvin. (1985) The Society of Mind, New York: Simon and Schuster.
- NOY, N. F., e McGuinness, D. L. (2001) Ontology development 101: A guide to creating your first ontology.
- PASSOS, A. Inteligência Competitiva: como fazer IC acontecer na sua empresa. São Paulo, LCTE Editora, 2005.
- PRAZERES, C. V. S. (2009) Serviços Web Semântica: da modelagem à composição. 2009. 189 f. Tese (Doutorado em Ciência da Computação). – ICMC, Universidade de São Paulo, São Carlos.
- PRESCOTT, J. E. (1995) The evolution of competitive intelligence. International Review of Strategic Management 6. 71-90.
- SAGIROGLU, S., E SINANC, DUYGU. (2013) Big data: A review. Collaboration Technologies and Systems (CTS), 2013 International Conference on. IEEE.
- SANTAREM SEGUNDO, J. E. (2010) Representação Iterativa: um modelo para Repositórios Digitais. 224 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília.
- SANTOS, C.D.; VALENTIM, M. L. P. (2014) As interconexões entre a gestão da informação e a gestão do conhecimento para o gerenciamento dos fluxos informacionais, periódico Perspectivas em Gestão & Conhecimento, João Pessoa, ISSN: 2236-417X ,v. 4, n. 2, p. 19-33, jul./dez.
- SILVA, Helena Pereira. (2003) Inteligência competitiva na Internet: um processo otimizado por agentes inteligentes. Ciência da informação, v. 32, n. 1.
- SILVA, T. M. S. (2003) Extração de informação para busca semântica na web baseada em ontologias.
- SOUZA, R. R., E ALVARENGA, L. (2004) A Web Semântica e suas contribuições para a ciência da informação. Ciência da Informação, Brasília 33.1. 132-141.
- STEINBERG, A.N.; BOWMAN, C.L.; WHITE, F.E. (1999) Revisions to the JDL data fusion model. Proceedings of SPIE, [S.l.], v. 3719, n.1, p. 430–441.
- TURING, Alan (1936). On computable numbers, with an application to the Entscheidungsproblem, Proceedings of the London Mathematical Society, Series 2, 42. pp 230 - 265.
- USCHOLD, M., E GRUNINGER, M. (1996) Ontologies: Principles, methods and applications. The knowledge engineering review 11.02. 93-136.
- WOOLDRIDGE, M. Agent-based software engineering. 1997. Proceedings... London, United Kingdom: [s.n.], 1997.
- WORLD BANK, Indicadores, Pesquisa e desenvolvimento (em% do PIB), disponível em <<http://data.worldbank.org/indicador/GB.XPD.RSDV.GD.ZS>>. Acesso em 04 janeiro de 2016.
- ZIKOPOULOS, P., C. EATON AND D. DERROOS, 2012. Understanding bigdata: Analytics for enterprise class hadoop and streaming data. McGraw-Hill, New York.