



XVII Encontro Nacional de Pesquisa em Ciência da Informação (XVII ENANCIB)

## GT 8 – Informação e Tecnologia

### PLATAFORMA DE APOIO À INOVAÇÃO BASEADA EM AGENTES INTELIGENTES SEMÂNTICOS: A INTERDISCIPLINARIDADE DE MODELOS COMPUTACIONAIS E INFORMACIONAIS

#### *INNOVATION SUPPORT PLATFORM BASED ON SEMANTIC INTELLIGENT AGENTS: INTERDISCIPLINARITY OF COMPUTER AND INFORMATIONAL MODELS*

Marcos L. Mucheroni<sup>1</sup>, Elvis Fusco<sup>2</sup>

**Modalidade da apresentação:** Comunicação Oral

**Resumo:** Com o intuito de compreender a interdisciplinaridade da Ciência da Informação com a Ciência da Computação no que se refere às Tecnologias da Informação, para dar solução às demandas de usuários no cenário contemporâneo de Ambientes Informacionais Digitais de suporte à decisão, em processos de inovação, este artigo tem como objetivo refletir sobre os fundamentos de interdisciplinaridade da Ciência da Informação para estabelecer uma plataforma de apoio a ambientes de inovação baseada em agentes inteligentes semânticos no contexto de *Big Data*, considerando questões de interdisciplinaridade entre modelos computacionais e informacionais. A abordagem metodológica desta pesquisa, parte dos objetivos propostos, aliando-os à análise de conhecimento teórico sobre métodos, modelos e ferramentais consolidados das áreas da Ciência da Informação e da Ciência da Computação para, de forma interdisciplinar, propor uma plataforma de apoio à inovação. Os processos que envolvem a informação se mostram extremamente vinculados às técnicas computacionais, uma vez que para tratar, organizar, representar e recuperar uma informação, algoritmos, aplicações, produtos e serviços informáticos são de extrema importância. A utilização da combinação das tecnologias computacionais, informacionais e semânticas dos espaços contidos na plataforma apresentada, permitem 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 e colaborativa de novos conhecimentos e em ações de incentivo à inovação nas organizações.

**Palavras-chave:** Agentes Inteligentes. Interdisciplinaridade. Inovação. *Big Data*. Web Semântica.

---

<sup>1</sup> Prof. Dr. pela Escola Politécnica da USP, credenciado no PPGCI-USP desde 2010.

<sup>2</sup> Elvis Fusco é doutor em Ciência da Informação e mestre em Ciência da Computação, coordena e leciona nos cursos de Ciência da Computação e Sistemas de Informação do UNIVEM.

**Abstract:** *In order to understand the interdisciplinarity of Information Science in Computer Science with regard to Information Technology to provide solutions to the demands of users, in the contemporary scenario of digital information environments decision support, in innovation processes, this article reflects on the fundamentals of interdisciplinarity of Information Science to establish a platform to support innovation environments based on semantic intelligent agents in the context of Big Data, considering interdisciplinary issues between computational and informational models. The methodological approach of this research, starts from the proposed objectives, combining them to theoretical knowledge of analysis of methods, models and tooling consolidated of the areas of Information Science and Computer Science for, in an interdisciplinary way, propose a platform to support innovation. The processes involving information are shown extremely linked to computational techniques, as to treat, organize, represent and retrieve some information, algorithms, applications, products and IT services are of utmost importance. The use of the combination of computational, informational and semantic technologies of spaces contained in the presented platform, allows the main actors in the field of innovation can have access to product information generated and collaboratively to support in a collective and collaborative construction processes of new knowledge and actions to encourage innovation in organizations.*

**Keywords:** *Intelligent Agents. Interdisciplinarity. Innovation. Big Data. Semantic Web.*

## 1 INTRODUÇÃO

Vivencia-se uma nova disrupção tecnológica pela convergência da colaboração, mobilidade, geração massiva e existência de um grande volume e heterogeneidade de dados. Esse contexto, denominado *Big Data*, apresenta um dos grandes desafios para as pesquisas de desenvolvimento de Ambientes Informativos Digitais para o uso efetivo dessas informações, promover a integração de conceitos, modelos e ferramentais advindos de várias áreas de conhecimento 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 em 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, uma vez que mudanças influenciam instituições e empresas.

Assim, práticas existentes na Inteligência Competitiva figuram como possibilidade de utilizar esse grande volume de informações presentes nos macro-ambientes para apoiar as

ações de recuperação da informação, apoio à decisão e a inovação nas organizações, a partir de processos informacionais sistemáticos para sustentar tais ações.

Dentre estas ações está o monitoramento do ambiente externo que torna-se complexo se realizado de forma manual, devido à amplitude e ao dinamismo informacional dos Ambientes Informacionais Digitais identificados em ambientes *Big Data*. Neste contexto, são utilizados os Agentes Inteligentes de Extração Automática, que fazem o monitoramento e a extração das diversas fontes informacionais da *Web*.

Com o propósito de adicionar significado aos conteúdos buscados em domínios específicos, associam-se conceitos semânticos aos Agentes Inteligentes de Extração Automática, que permitem realizar a busca não mais por palavras chaves num processo de busca textual, mas sim por significado e valor, extraído das páginas e serviços da *Web*, informações relevantes, descartando aquilo que é desnecessário. A partir disto, a questão da *Web Semântica* com o uso de ontologias aparece como solução na busca de inserir semântica neste processo. O uso de ontologias é essencial no processo de desenvolvimento dos Agentes Inteligentes 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 e aderente às necessidades do usuário.

As Tecnologias da Informação e Comunicação (TIC) que têm como base as teorias da Ciência da Computação e o seu uso para processar a informação científica é uma das ideias fundadoras da Ciência da Informação. Com relação às Tecnologias de Informação essas relações com a Ciência da Informação e os problemas com os quais se ocupa, em particular com o problema do acesso e transferência da informação e do conhecimento, Saracevic (1995, p.2, tradução nossa) explicita: “a Ciência da Informação está inexoravelmente conectada à Tecnologia da Informação. Um imperativo tecnológico está compelindo e restringindo a evolução da Ciência da Informação, como é na evolução da sociedade da informação”.

A proximidade da Tecnologia da Informação com a Ciência da Informação, embora pouco clara para muitos autores, e quase consagrada para outros, aos poucos vai se revelando no terreno empírico com uma proximidade interdisciplinar bastante prática.

Não se trata apenas do uso de software e equipamentos em Ambientes Informacionais, como Bibliotecas e ambientes culturais onde a organização da informação é fundamental para funcionamento e uso, o enlace das tecnologias que emergem a partir da *Web Semântica* e uso de ontologias, elaboração estruturada de metadados e certificação de autoria e relevância

tornam necessária uma nova e maior proximidade interdisciplinar entre áreas do conhecimento que estão focadas na Ciência da Informação e na Ciência da Computação.

Desde os planos iniciais da Web Semântica (TIM BERNERS-LEE etc al., 2001) e os primeiros textos sobre agentes inteligentes (HENDLER, 2001) a necessidade de ligar dados, tecnologia hoje conhecida como *linked data* que envolvem projetos como VIAF (*Virtual International Authority File*) (LINKEDDATATOOLS, 2016) para autoridade, ontologias para a ligação entre produções científicas, autores e conteúdos, eram já possíveis de se pensar na interdisciplinaridade que envolviam estes dois campos do saber, um mais ligado a informação no sentido lato do termo (Ciência da Informação) e outra mais ligada ao sentido tecnológico do termo (Ciência da Computação), mas havia a preocupação quase simultânea nestes dois focos de Ambientes Informacionais.

Com o intuito de refletir a interdisciplinaridade da Ciência da Informação com a Tecnologia da Informação para dar solução às demandas de usuários no cenário contemporâneo de Ambientes Informacionais Digitais de suporte à decisão em processos de inovação, este artigo tem como objetivo refletir sobre os fundamentos de interdisciplinaridade da Ciência da Informação para estabelecer uma plataforma de apoio a ambientes de inovação baseada em agentes inteligentes semânticos no contexto de *Big Data* considerando questões de interdisciplinaridade entre modelos computacionais e informacionais.

Especificamente, pretende-se: propor uma arquitetura de apoio ao processo de aquisição, fusão e a persistência das informações baseadas em Ambientes Informacionais de inovação de forma a gerar processos de gestão e tomada de decisão para os atores no cenário da inovação que integre: a) tecnologias computacionais como Agentes de Extração Automática, Fusão de Dados, *Big Data Analytics* e Persistência NoSql; b) tecnologias informacionais semânticas: ontologias, padrões de metadados e Recuperação da Informação.

Sendo assim, baseia-se na afirmação de que a Ciência da Informação se caracteriza por manter uma interdisciplinaridade com outras áreas, como a Ciência da Computação e, esta pesquisa reflete os conceitos que são trabalhados por ambas as áreas na temática abordada, de forma a demonstrar que na problemática apresentada, as pesquisas aplicadas na Ciência da Informação buscam em outras ciências subsídios para aprimorar e estabelecer formas mais eficazes nos processos de representação, persistência, descrição, organização, acesso e recuperação de recursos informacionais, ou seja, o objeto principal da área que é a informação.

Essa interdisciplinaridade busca unir essas áreas para discutir os problemas particulares no ponto de vista de cada área do conhecimento, descobrindo conceitos e

possibilidades que emergem colocando-as juntas. Este é o caso dos agentes inteligentes quando vistos a partir de problemas da Ciência da Informação: correta indexação e relevância e também na Web Semântica no tratamento das representações ontológicas e epistemológicas.

## 2 INTERDISCIPLINARIDADE ENTRE A CIÊNCIA DA INFORMAÇÃO E A CIÊNCIA DA COMPUTAÇÃO

A Ciência da Informação (CI) apresenta em sua essência a interdisciplinaridade, existindo diversas áreas do conhecimento que possuem uma relação intrínseca com esta ciência. Em um dos principais textos tratando da CI intitulado “*Information Science: What is it?*”, Harold Borko (1968, p.5, tradução nossa, grifo nosso) afirma que a CI

[...] é uma **ciência interdisciplinar** que investiga as propriedades e o comportamento da informação, as forças que governam o fluxo e os usos da informação e as técnicas, manuais e mecânicas, de processamento da informação, visando uma ideal armazenagem, recuperação e disseminação.

Tal definição apresentada indica que a natureza interdisciplinar da CI está presente desde a sua concepção. Neste texto, Borko (1968) cita algumas das áreas que a CI se deriva, apresentando assim, laços de interdisciplinaridade, como a Matemática, a Ciência da Computação, a Lógica, a Biblioteconomia, entre outras. Diante disso, há diversas atividades pertencentes a CI que contém uma profunda influência da Ciência da Computação, destacados pelo próprio Borko (1968, p. 3, tradução nossa), afirmando que a CI apresenta o componente de ciência aplicada uma vez que: “[...] desenvolve produtos e serviços.”

As questões levantadas pelo autor conduz a identificar os profundos laços entre a Ciência da Computação e a CI, existindo a necessidade desta primeira levar as suas teorias e a sua prática, especialmente no que condiz a desenvolver aplicações e algoritmos que possibilite a validação das teorias relativos aos fluxos informacionais. A partir disso, a Recuperação da Informação (RI) se mostra como o principal elo entre estas duas ciências.

Em um outro texto da área, Saracevic (1995) aponta as interdisciplinaridades da CI, no texto intitulado de “*Interdisciplinarity Nature of Information Science*”, indicando a RI como um componente central da interdisciplinaridade da CI. Neste sentido, o autor afirma que:

Enquanto isso, a explosão da informação continuou inabalável: textos eletrônicos, bases de dados e redes se disseminaram, a demanda e a busca por informações dispararam; agora as tecnologias trazem mudanças e oportunidades, e os problemas da RI se intensificaram. Estando aí a contínua e a crescente importância das pesquisas de RI que tem atraído profissionais de diversas disciplinas. (SARACEVIC, 1995, p. 3, tradução nossa)

O cenário apresentado por Saracevic, permanece inalterado até o presente momento, havendo um aumento exponencial nos dados disponibilizados digitalmente. Assim, pesquisas multi e interdisciplinares entre a Computação e a CI possibilitam encontrar meios para tratar essas questões, de modo que, a Computação encontre formas de buscar e processar informações, enquanto a CI pode trazer contribuições no que compete as teorias de RI e a melhores formas de organizar e de representar as informações.

Saracevic (1995, p. 5, tradução nossa) declara que: “A Ciência da Computação trata de algoritmos relacionados à informação, enquanto a Ciência de Informação atua sobre a própria natureza da informação e o seu uso pelos seres humanos. Os dois interesses não competem, eles são complementares.”

A partir disso, verifica-se a importância que a Ciência da Computação apresenta para a CI, especialmente no que tange aos algoritmos que possibilitam estudos mais práticos dentro desta segunda área do conhecimento. Assim, as pesquisas de CI que se encontram no âmbito da Informação e Tecnologia, necessitam dialogar com a Ciência da Computação, para compreender a influência e a importância que as Tecnologias de Informação e Comunicação (TIC) possuem no tratamento da informação como um todo.

Os processos que envolvem a informação se mostram extremamente vinculados às técnicas computacionais, uma vez que para tratar, organizar, representar e recuperar uma informação, algoritmos, aplicações, produtos e serviços informáticos são de extrema importância. A CI não é mera utilizadora das TIC, mas sim uma Ciência que estuda e que investiga as tecnologias. Assim, quanto mais a interdisciplinaridade com a Ciência da Computação, mais profundas serão as pesquisas que visam compreender esta relação entre informação e tecnologia.

Saracevic (1996, p. 47) relata que diversos estudos estão sendo realizados dentro da Ciência da Computação, “[...] associado com a representação da informação, sua organização intelectual e encadeamentos; busca e recuperação de informação; a qualidade, o valor e o uso da informação - todos tradicionalmente tratados pela CI.” A partir desta afirmação, verifica-se que existem diversos pontos de confluência entre estas duas áreas do conhecimento, existindo temas de pesquisa que podem agregar significativamente com pesquisas interdisciplinares.

Mais recentemente, Pinheiro (2006) aponta as relações interdisciplinares existentes dentro da CI, a partir das disciplinas que competem a esta área de estudo. Dentre as 17 disciplinas destacadas, a Ciência da Computação foi a que mais apresentou interdisciplinaridades, com oito aparições. Por mais que a autora relativiza este número, ao destacar que por vezes esta relação se encontra na função instrumental de infraestrutura

tecnológica, tal número indica a importância que a Ciência da Computação apresenta dentro da CI.

Os textos estudados demonstram a natureza interdisciplinar da CI, da mesma forma que indicam a influência que a Ciência da Computação tem na CI. A partir deste cenário, a Recuperação da Informação, destacada por Saracevic (1995) como ponto de interdisciplinaridade da CI, assim como, a Representação da Informação, a Modelagem de Dados, a Persistência de Dados, a Interoperabilidade, todas essas áreas possuem ligações com a Computação, estabelecendo a necessidade de pesquisas conjuntas entre as duas áreas do conhecimento.

Neste contexto, esta pesquisa propõe a utilização da interdisciplinaridade de conceitos, métodos, modelos e ferramentais consolidados das áreas da Ciência da Informação e da Ciência da Computação para apoiar processos de inovação nos principais atores dos ambientes de inovação do país.

### **3 INOVAÇÃO E A RECUPERAÇÃO DA INFORMAÇÃO**

A inovação tem como base a utilização da informação como insumo para iniciativas de pesquisas tecnológicas para geração de cenários que gerem melhorias ou novidades em produtos, serviços ou processos. Devido à importância da informação no processo da inovação, propõe-se a reflexão e a investigação do uso de conceitos interdisciplinares da Ciência da Informação com outras ciências de modo a propor soluções para as demandas nos processos de 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.

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 um fluxo informacional é um legado que deve ser construído com esforço colaborativo entre os atores 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.

Os conceitos presentes na Ciência da Informação e na Ciência da Computação exercem um papel importante no desenvolvimento de plataformas que automatizem processos informacionais que dêem suporte aos principais atores da inovação na recuperação de informações que subsidiem ações inovadoras.

Nos processos de inovação, a Recuperação da Informação torna-se alvo de muitos estudos, devido à 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).

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.

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.

Nos processos de inovação, os Sistemas de Recuperação de Informação em ambientes de Big Data devem representar as informações e apresentá-las aos usuários de maneira que estes, por meio daquelas informações recuperadas, consigam satisfazer total ou parcialmente as suas necessidades informacionais em ambientes internos e externos.

A partir deste cenário, 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 Ambientes Informacionais de Inteligência Competitiva.

#### **4 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 as 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.

Neste contexto, 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 (DETERS; ADAIME, 2003).

Na perspectiva desta pesquisa, o termo agente de extração, tem como base o conceito de robô de extração, que é capacitado para extrair conteúdos de páginas *Web*, 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. 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.

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), que apresentam uma plataforma que tem como centro um agente que extrai informações da *Web*, utilizando ontologias para classificar e representar os dados extraídos. Outro estudo é apresentado por Freitas (2002), que em sua tese, 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 da 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 em ambientes da *Web Semântica*.

## **5 WEB SEMÂNTICA**

Em 1989, Tim Berners Lee propões 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* (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.

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*.

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 (BEERNERS-LEE, 2001).

Em ambientes da *Web Semântica*, as ontologias apresentam-se de grande relevância nas investigações sobre a necessidade do uso de agentes que de alguma forma possam garimpar e organizar a informação ali contida.

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. A partir deste cenário, observa-se que com uma ontologia consegue-se 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. Assim, a ontologia aparece como uma possível solução para tratar semanticamente toda a heterogeneidade e complexidade das informações que estão presente nos ambientes de *Big Data*.

## 6 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*.

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.

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).
- **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 natureza heterogênea, pois os dados podem vir de uma variedade de mídias digitais, 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, em que a mudança quantitativa trouxe uma mudança qualitativa das informações (MAYER-SCHÖNBERGER E CUKIER, 2013).

Diante disso, as informações nos ambientes de *Big Data* são importantes insumos nos processos organizacionais como Ambientes e Plataformas Informacionais Digitais de Inteligência Competitiva do Apoio à Decisão e à Inovação e investigações nas áreas da Ciência da Informação e da Ciência da Computação são necessárias para dar conta das atuais demandas informacionais.

Apresentadas as grandes questões de uma plataforma de apoio à inovação utilizando conceitos presentes na Inteligência Competitiva, uma questão que pode ser importante neste contexto é qual seria o procedimento metodológico e implicações para a área da Ciência da Informação. Propõe-se a seguir a articulação computacional e informacional para geração de uma arquitetura sistemática dos fluxos informacionais no cenário da inovação.

## 7 PROCEDIMENTOS METODOLÓGICOS PARA PLATAFORMA DE APOIO À INOVAÇÃO

A abordagem metodológica desta pesquisa, considerando os objetivos, aliados à análise de conhecimento teórico sobre métodos, modelos e ferramentas consolidados das áreas da Ciência da Informação e da Ciência da Computação para, de forma interdisciplinar, propor uma plataforma de apoio à inovação, caracteriza-se por ser uma pesquisa de análise exploratória e qualitativa descritiva do tema.

### 7.1 Arquitetura

Os fluxos informacionais caracterizam-se pelos processos de busca, acesso, recuperação, transformação, processamento, representação, persistência, transferência, apresentação e uso da informação. 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, proposta na Figura. 1.

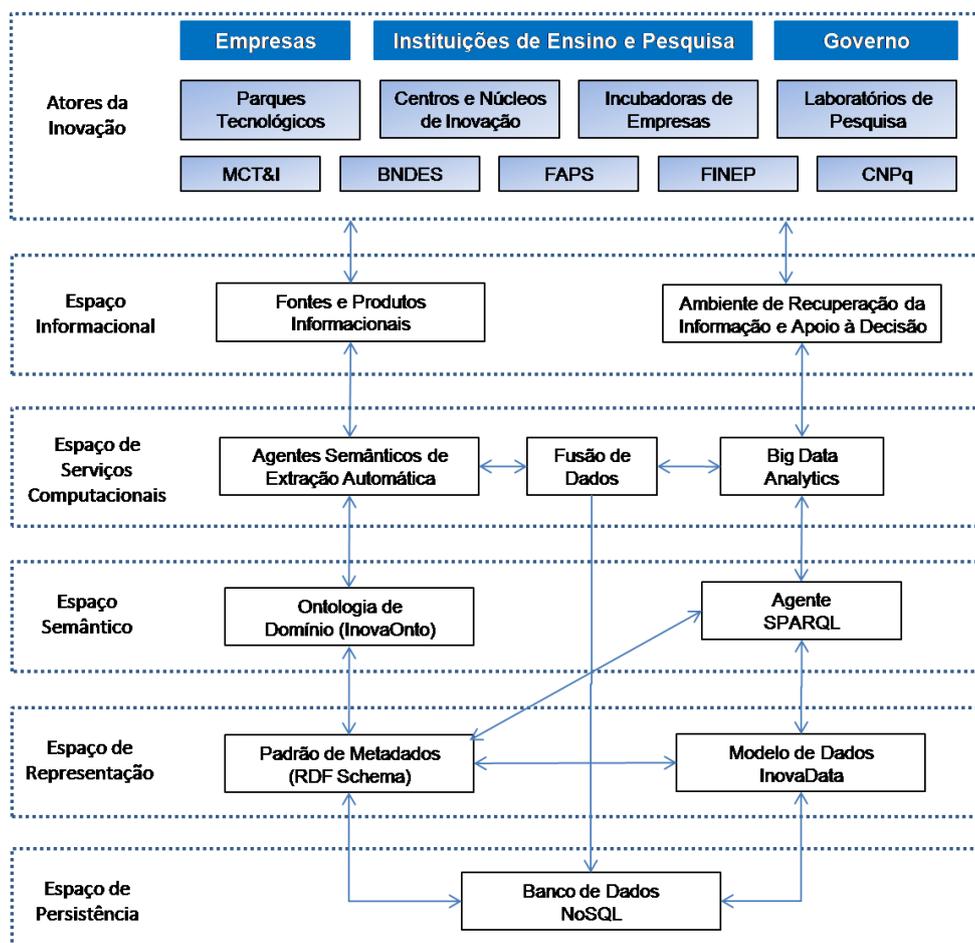


Figura 1: Arquitetura Computacional e Informacional de Apoio à Inovação (Fonte: o autor)

Assim, é proposta uma arquitetura informacional e computacional que utiliza agentes computacionais semânticos para apoiar processos decisórios e inovadores por meio de processos de Recuperação da Informação, como pode ser visto na Figura 1.

O cenário da inovação é composto por diversos atores que disponibilizam e consomem informações dinâmicas e heterogêneas que demandam a utilização de conceitos da: i) Inteligência Competitiva; ii) Ciência da Computação como: agentes computacionais, *Big Data*, Fusão de Dados, *Analytics*, Persistência NoSQL e Modelagem de Dados ; e iii) Ciência da Informação: *Web Semântica*, Ontologia, Representação da Informação, Metadados e Recuperação da Informação.

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 construída interdisciplinarmente se baseia.

## **7.2 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 criam 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.

## **7.3 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.

#### **7.4 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, assim, 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 cenário, 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 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 sobre de um ambiente de interesse (BOTEGA, 2016).

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 de *Big Data 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.

### **7.5 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 (*Query Language for RDF*). Assim essas aplicações podem ter dados persistidos em grafos baseados em RDF nativo e extraí-los de bancos de dados tradicionais.

### **7.6 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 infra-estrutura 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.

### **7.7 Espaço de Persistência**

No espaço de persistência ocorre o armazenamento dos dados que são extraídos pelo agente de busca e pelo processo de fusão. Pode ocorrer tanto em Banco de Dados relacionais quanto em tecnologias NoSQL (*Not Only SQL*). A persistência pode ocorrer destas duas maneiras pois as informações extraídas podem ser dados estruturados ou 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 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.

A plataforma descrita pretende servir de base para o desenvolvimento de serviços informacionais de apoio aos ambientes de inovação a partir da extração de dados das fontes informacionais no contexto de Big Data e da construção de uma base informacional semântica apoiada por modelos ontológicos do domínio da inovação.

## **8 CONSIDERAÇÕES FINAIS**

Espera-se ter evidenciado que os processos informacionais estão vinculados às técnicas computacionais, uma vez que para tratar, organizar, representar e recuperar a informação, algoritmos, aplicações, produtos e serviços informáticos são de extrema importância. Compreende-se que a Ciência da Informação é uma Ciência que estuda e que investiga esses processos e, por esta razão, é essencial a interdisciplinaridade com a Ciência da Computação.

No campo da Inovação e da Inteligência Competitiva, observa-se a importância dos conceitos presentes na CI e na Ciência da Computação no desenvolvimento de plataformas que automatizem processos informacionais que dêem suporte aos principais atores da inovação na recuperação de informações que subsidiem ações inovadoras.

A Web Semântica e o uso de ontologias é essencial no processo de desenvolvimento dos Agentes Inteligentes de busca semântica, sendo aplicados 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 e aderentes às necessidades do usuário.

Acredita-se que este estudo cumpriu o seu objetivo de compreender a interdisciplinaridade entre Ciência da Informação e a Ciência da Computação no que se refere às Tecnologias da Informação nas áreas da Representação e Recuperação da Informação e como essas áreas se articulam para dar solução às demandas de usuários no cenário contemporâneo de Ambientes Informacionais Digitais de suporte à decisão em processos de inovação. Foi proposta uma plataforma de apoio a ambientes de inovação baseada em agentes inteligentes semânticos no contexto de *Big Data* considerando questões de interdisciplinaridade entre modelos computacionais e informacionais.

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

## Referências

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Modern information retrieval**. New York: ACM; Harlow: Addison-Wesley, 1999.

BERNERS-LEE, T. **Information Management: A Proposal. 1989**. Disponível em: <<http://w3.org/History/1989/proposal.html>>. Acesso em 09 de julho de 2015.

BERNERS-LEE, T., LASSILA, O. E HENDLER, J. The semantic web. **Scientific American**, New York, v. 5, 2001.

BORKO, H. Information Science: What is it? **American Documentation**, v.19, n.1, p.3-5, Jan. 1968.

BOTEGA, L. **Modelo de Fusão Dirigido por Humanos e Ciente da Qualidade de Informação**. 260 f. Tese de Doutorado. Programa de pós-graduação em Ciência da Computação, Universidade Federal de São Paulo, 2016.

CONEGLIAN, C. S.; FUSCO, E. Plataforma Semântica de Recuperação da Informação aplicada a Extração de Artigos Científicos. In: **12ª Conferência Internacional sobre Sistemas de Informação e Gestão de Tecnologia - CONTECSI, 2015**, São Paulo: TECSI EAC FEA USP, 2015. v. 12. p. 1-15.

DETERS, J. I., E ADAIME, S. F. Um estudo comparativo dos sistemas de busca na web. Anais do **V Encontro de Estudantes de Informática do Tocantins**. Palmas, TO, 2003. p. 189-200.

DZIEKANIAK, G. V., E KIRINUS, J. B. Web semântica. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, n.18, 2004. p.20-39.

FRANKLIN, S.; GRAESSER, A. 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.], 1996, p.21–36.

FREITAS, F. **Sistemas multiagentes cognitivos para a recuperação e extração integradas de informação da web**. 205 f. Tese de Doutorado, Programa de pós-graduação em Engenharia Elétrica/UFSC, Florianópolis, 2002.

GENESERETH, M. R.; KETCHPEL, S. P. Software agents. **Communications of the ACM**, [S.l.], v.38, n.7, 1994. p.48–53.

GRUBER, T. R. A translation approach to portable ontology specifications. **Knowledge acquisition** 5.2. 1993, p.199-220.

GUARINO, N. Formal ontology in information systems. **Proceedings of the first international conference (FOIS'98)**, June 6-8, Trento, Italy. Vol. 46. IOS press, 1998.

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.

KAKHANI, M. K., KAKHANI, S., E BIRADAR, S. R. **Research Issues in Big Data Analytics**, 2013.

KATAL, A., WAZID, LINKEDDATATOOLS M., E GOUDAR, R. H. Big data: Issues, challenges, tools and Good practices. **Contemporary Computing (IC3)**, 2013 Sixth International Conference on. IEEE, 2013.

LINKEDDATATOOLS. **Introducing Linked Data And The Semantic Web**. Disponível em: <<http://www.linkeddatatools.com/semantic-web-basics>> Acesso em: 28 jul. 2016.

MAYER-SCHÖNBERGER, V., E CUKIER, K. **Big data: A revolution that will transform how we live, work, and think**. Houghton Mifflin Harcourt, 2013.

McAFEE, A., ET AL. Big Data. The management revolution. **Harvard Bus Rev** 90.10, 2012, p.61-67.

MODESTO, L. R. **Representação e Persistência para acesso a Recursos Informacionais Digitais gerados dinamicamente em sítios oficiais do Governo Federal**. 2013. 103 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2013.

PINHEIRO, Lena Vania Ribeiro. Ciência da Informação: desdobramentos disciplinares, interdisciplinaridade e transdisciplinaridade. In: González de Gómez, Maria Nélida (Org.); Dill Orico, Evelyn Goyannes (Org.). **Políticas de memória e informação**. Natal: EDUFRN, 2006. p.111-142.

SANTOS, C.D.; VALENTIM, M. L. P. 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, v. 4, n. 2, 2014. p.19-33, jul./dez.

SARACEVIC, T. Interdisciplinary nature of information science. **Ciência da Informação**. vol. 24, n. 1, 1995.

SARACEVIC, T. Ciência da informação: origem, evolução e relações. **Perspectiva em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996.

SILVA, T. M. S. **Extração de informação para busca semântica na web baseada em ontologias**, 2003, 79f. Programa de pós-graduação em Engenharia Elétrica/UFSC, Florianópolis, 2003.

SOUZA, R. R., E ALVARENGA, L. A Web Semântica e suas contribuições para a ciência da informação. **Ciência da Informação**, Brasília 33.1, 2004. p.132-141.

WOOLDRIDGE, M. Agent-based software engineering. **IEE Proceedings Software Engineering**, London, United Kingdom: [s.n.], 1997.

ZIKOPOULOS, P., C. EATON AND D. DEROOS. **Understanding bigdata: Analytics for enterprise class hadoop and streaming data**. New York: McGraw-Hill, 2012.