



UNIVERSIDADE FEDERAL DE PERNAMBUCO  
CENTRO DE ARTES E COMUNICAÇÃO  
DEPARTAMENTO DE CIÊNCIA DA INFORMAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO  
MESTRADO EM CIÊNCIA DA INFORMAÇÃO

Dominique de Lira Vieira Corrêa

**A INTERPRETAÇÃO SEMÂNTICA DE TEXTOS CIENTÍFICOS EM  
PORTUGUÊS NA PERSPECTIVA DA CIÊNCIA DA INFORMAÇÃO:**  
procedimentos e aplicação à área de Ciências Agrárias

RECIFE/PE  
2016

DOMINIQUE DE LIRA VIEIRA CORRÊA

**A INTERPRETAÇÃO SEMÂNTICA DE TEXTOS CIENTÍFICOS EM PORTUGUÊS  
NA PERSPECTIVA DA CIÊNCIA DA INFORMAÇÃO: procedimentos e aplicação à  
área de Ciências Agrárias**

Dissertação apresentada ao Programa de Pós-Graduação em  
Ciência da Informação da Universidade Federal de  
Pernambuco como requisito final para a obtenção do título de  
Mestre em Ciência da Informação.

**Área de concentração:** Informação, Memória e Tecnologia.

**Linha de Pesquisa:** Comunicação e visualização da memória.

**Orientador:** Prof.º Dr.º Raimundo Nonato Macedo dos Santos

RECIFE/PE  
2016

Catálogo na fonte

Bibliotecário Jonas Lucas Vieira, CRB4-1204

C824i Corrêa, Dominique de Lira Vieira

A interpretação semântica de textos científicos em português: procedimentos e aplicações à área de Ciências Agrárias na perspectiva da Ciência da Informação / Dominique de Lira Vieira Corrêa. – 2016.

127 f.: il., fig.

Orientador: Raimundo Nonato Macedo dos Santos.

Dissertação (Mestrado) – Universidade Federal de Pernambuco, Centro de Artes e Comunicação. Ciência da Informação, 2016.

Inclui referências e apêndices.

1. Ciência da informação. 2. Recuperação da informação. 3. Semântica. 4. Redação técnica. 5. Redação acadêmica. I. Santos, Raimundo Nonato Macedo dos (Orientador). II. Título.

020 CDD (22. ed.)

UFPE (CAC 2016-110)



Serviço Público Federal  
Universidade Federal de Pernambuco  
Programa de Pós-graduação em Ciência da Informação - PPGCI

DOMINIQUE DE LIRA VIEIRA CORRÊA

*A interpretação semântica de textos científicos em português: procedimentos e aplicação  
à área de Ciências Agrárias na perspectiva da Ciência da Informação*

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de mestre em Ciência da Informação.

Aprovada em: 29/02/2016

**BANCA EXAMINADORA**

---

Prof. Dr. Raimundo Nonato Macedo dos Santos (Orientador)  
Universidade Federal de Pernambuco

---

Prof. Dr. Fábio Mascarenhas e Silva (Examinador Interno)  
Universidade Federal de Pernambuco

---

Prof. Dr. Piotr Trzesniak (Examinador Externo)  
Universidade Federal de Itajubá



## **DEDICATÓRIA**

A Deus, que me conduziu com muito amor em cada passo dessa caminhada permitindo assim a conclusão de mais uma etapa em minha vida. A meu esposo Renato, a minha filha Milena, a meu filho Miguel, minha mãe Aparecida e a minha irmã Jéssica pelo amor, carinho e apoio.

## AGRADECIMENTOS

A Deus por me conduzir em cada passo do meu caminho permitindo que todas as coisas fossem favoráveis à conclusão desse estudo.

Ao meu esposo Renato pela compreensão, companheirismo, orientação, apoio e amizade. Aos meus filhos Milena e Miguel, mesmo tão pequenos foram compreensíveis alegrando e movimentando meus dias. Gratidão também à minha mãe Aparecida e a minha irmã Jessica pelo carinho e apoio constantes contribuindo de diversas maneiras para que esse trabalho se concretizasse.

Ao meu professor e orientador Raimundo Nonato Macedo dos Santos pelas orientações, direcionamentos e contribuições. Grata pela sua compreensão e paciência. Ao professor e também “orientador” Piotr Trzesniak por ter dedicado seu tempo incentivando, aconselhando e contribuindo no aprimoramento da pesquisa.

Gratidão também ao professor Fábio Mascarenhas pela contribuição como membro da banca de qualificação desta pesquisa e pelos ensinamentos ao longo de toda a minha graduação na Universidade Federal de Pernambuco.

Agradecimento aos professores do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Pernambuco (PPGCI/UFPE) que se dedicaram e transmitiram seus nobres conhecimentos contribuindo assim para minha vida pessoal, acadêmica e profissional.

De forma equivalente, agradeço aos meus colegas de mestrado da turma 2014.1 pela parceria, companheirismo e crescimento mútuo. Foram dois anos de muitas experiências vividas e de enriquecimento de ideias.

À FACEPE pela concessão da bolsa de Pós-Graduação.

E por fim, meus sinceros agradecimentos a todos que contribuíram direta ou indiretamente para a conclusão de mais essa etapa na minha vida.

*“Sim, coisas grandiosas fez o Senhor por nós,  
por isso estamos alegres”  
(Salmo 126:3)*

## RESUMO

A presente pesquisa se desenvolveu no âmbito do Observatório Temático e Laboratório – Ensino, Tecnologia, Ciência e Informação (OtletCI) com a intenção de avançar na questão de como extrair informação relevante e de como representá-la para fins de recuperação semântica da informação, em particular no caso de textos de publicações científicas em português. Para tanto, como metodologia, investigou-se a tecnologia da busca semântica quanto aos fundamentos teóricos, sua utilidade no contexto do *OtletCI* e requisitos para aplicação em textos científicos em português. Como experimento, buscou-se explicitar os requisitos da busca semântica para a aplicação em textos científicos, através da análise da extração de relacionamentos semânticos do tipo “causa e efeito” em 60 resumos, em português, de artigos científicos da área de Ciências Agrárias. O estudo apresentou, por meio de considerações de ordem qualitativa e quantitativa, uma comparação entre o processo manual e automático de extração de sentenças de causa e efeito. Esses documentos foram previamente analisados de forma manual, e as sentenças de causa e efeito foram extraídas através da leitura dos resumos. Para o processo automático, com os dados transferidos do *software PALAVRAS* para a planilha do *Excel*, foi possível realizar uma programação para localizar sentenças de causa e efeito automaticamente. O objetivo foi comparar as sentenças identificadas diretamente pelo pesquisador e as sentenças reconstruídas automaticamente a partir do conjunto de células programadas. Conclui-se enfatizando que a possibilidade de usar técnicas automáticas acelera o processo de criação e extração de relações de causa e efeito e pode ser usada como alternativa ao processo custoso de identificação manual de informações semânticas. Porém, mais importante que propor uma estrutura de relações de causa e efeito para a construção de sistemas de busca, o que pode-se apontar como o resultado mais expressivo da presente pesquisa é o estabelecimento preliminar de rotinas para a versão automatizada.

**Palavras-chave:** Recuperação da Informação. Busca semântica. Observatório de Ciência, Tecnologia & Inovação. Software Goldfire. Tecnologia de mineração de texto. Verbos causa e efeito.

## ABSTRACT

This research is developed within the Thematic Observatory and Laboratory - Education, Technology, Science and Information (OtletCI) with the intention to move forward on the question of how to extract relevant information and how to represent it for purposes of semantic retrieval of information, particularly in the case of texts of scientific publications in Portuguese. Therefore, as a methodology, we investigated the semantic search technology based on the theoretical foundations, its usefulness in the context of OtletCI and requirements for application in scientific texts in Portuguese. As an experiment, we tried to clarify the semantic search requirements for the application of scientific texts by analyzing the extraction of semantic relationships such as "cause and effect" in 60 abstracts, in Portuguese, of scientific articles in the area of Agricultural Sciences. The study shows, through qualitative and quantitative considerations, a comparison between manual and automatic extraction process of cause and effect sentences. These documents were previously analyzed manually, and the sentences of cause and effect were extracted by reading the summaries. For automatic process, with data transferred from PALAVRAS software to the Excel spreadsheet, it was possible to carry out a program to find cause and effect sentences automatically. The goal was to buy the sentences identified directly by the researcher and sentences automatically reconstructed from the set of programmed cells. The research concludes emphasizing that the possibility of using automatic techniques accelerates the process of creating and extracting of cause and effect relationship and may be used as an alternative to costly manual process of identifying semantic information. However, more important than to propose a structure of cause and effect relationships for building search engines, we can point out as the most significant result of this research the preliminary establishment of routines for automated version.

**Keywords:** Information Retrieval. Semantic Search. Observatory of Science, Technology & Innovation. Goldfire Software. Text mining technology. Verbs cause and effect.

## LISTA DE FIGURAS

Figura 1- Arquitetura da Web Semântica.....	38
Figura 2- Critérios da busca semântica.....	45
Figura 3- Processo de KDD.....	55
Figura 4- Etapas do processo de mineração de texto.....	57
Figura 5- Tecnologia semântica do Goldfire.....	67
Figura 6- Tela do software Goldfire .....	69
Figura 7- Interface do software <i>PALAVRAS</i> .....	87
Figura 8- Resultado de um texto submetido ao processador <i>PALAVRAS</i> .....	88
Figura 9- Arquivo no formato .csv .....	89
Figura 10- Marcação da relação causa e efeito.....	90
Figura 11- Procedimentos das células programadas no <i>Excel</i> .....	95
Figura 12- Diagrama dos resultados quantitativos.....	98

## LISTA DE TABELAS

Tabela 1- Marcação com o software gerada no Excel.....	89
Tabela 2- Relacionamentos semânticos mais comuns.....	92
Tabela 3: lista dos verbos encontrados nas sentenças marcadas.....	94

## LISTA DE QUADROS

Quadro 1- Aspectos referentes aos observatórios de C,T&I .....	26
Quadro 2- Formas de atuação dos observatórios.....	28
Quadro 3- Etapas do Processamento de Linguagem Natural.....	51
Quadro 4- Tipos de abordagem dos dados .....	57
Quadro 5- Possíveis análises tecnológicas através do uso de <i>Tech Mining</i> .....	59
Quadro 6- Fases de decisão do processo Tech Mining.....	60
Quadro 7- Algumas funções do software Goldfire.....	64
Quadro 8- Categorias e subcategorias do software Goldfire.....	67
Quadro 9- Quantidade de sentenças recuperadas.....	97

## LISTA DE SIGLAS

CAPES- Coordenação de Aperfeiçoamento de Pessoal de Nível Superior  
CGEE- Centro de Gestão e Estudos Estratégicos  
CI –Ciência da Informação  
CNPq- Conselho Nacional de Desenvolvimento Científico e Tecnológico  
CRN- Conselho Europeu para a Investigação Nuclear  
CT&I – Ciência, Tecnologia e Inovação  
HTML- HyperText Markup Language  
HTTP- Hypertext Transfer Protocol  
IA- Inteligência Artificial  
IEA- Instituto de Estudos Avançados  
IIB- Institut International de Bibliographie  
IPT- Instituto de Pesquisas Tecnológicas do Estado de São Paulo  
KDD- knowledge-discovery in databases  
MCT- Ministério de Ciência e Tecnologia  
MOT- Management of Technology  
OC- Organização do Conhecimento  
OTI- Observatório de Tecnologia e Inovação  
OTLETCI- Observatório Temático e Laboratório – Ensino, Tecnologia, Ciência e Inovação  
P&D- Pesquisa e Desenvolvimento  
PLN- Processamento de Linguagem Natural  
PPGCI- Programa de Pós-Graduação em Ciência da Informação  
RDF- Resource Description Framework  
SAO- Sujeito, ação, objeto  
SOC- Sistema de Organização do Conhecimento  
TRIZ- Teoria da Solução Inventiva de Problemas  
UFPE- Universidade Federal de Pernambuco  
URI- Identificador Uniforme de Recurs  
VISL- Virtual Interactive Syntax Learning  
W3C- World Wide Web Consortium  
WWW- World Wide Web  
XML- eXtensible Markup Language

## SUMÁRIO

1	<b>INTRODUÇÃO.....</b>	<b>13</b>
1.1	<b>Objetivos e estrutura da dissertação.....</b>	<b>19</b>
1.1.1	<i>Objetivo Geral.....</i>	<i>19</i>
1.1.2	<i>Objetivos específicos.....</i>	<i>19</i>
2	<b>FUNDAMENTAÇÃO TEÓRICA.....</b>	<b>21</b>
2.1	<b>Observatórios de C, T &amp; I.....</b>	<b>21</b>
2.1.1	<i>Formas de atuação dos observatórios.....</i>	<i>26</i>
2.2	<b>Recuperação Semântica da Informação.....</b>	<b>28</b>
2.2.1	<i>Evolução da tecnologia: contexto histórico.....</i>	<i>28</i>
2.2.2	<i>Web Semântica.....</i>	<i>32</i>
2.2.2.1	<i>A estrutura da Web Semântica.....</i>	<i>34</i>
2.2.3	<i>A semântica e os sistemas de organização do conhecimento.....</i>	<i>38</i>
2.2.4	<i>Semântica e tesouros, taxinomias, ontologias.....</i>	<i>39</i>
2.2.5	<i>Motor de busca semântico.....</i>	<i>40</i>
2.2.6	<i>As barreiras linguísticas.....</i>	<i>43</i>
2.3	<b>Processamento de Linguagem Natural.....</b>	<b>45</b>
2.4	<b>Tecnologias de mineração.....</b>	<b>51</b>
2.4.1	<i>Mineração de dados.....</i>	<i>51</i>
2.4.2	<i>Mineração de texto.....</i>	<i>53</i>
2.4.3	<i>Metodologia Tech Mining ou mineração de tecnologia.....</i>	<i>55</i>
2.5	<b>Teoria de Solução Inventiva de Problemas – TRIZ.....</b>	<b>58</b>
2.6	<b>Tecnologia de busca semântica – Goldfire.....</b>	<b>60</b>
2.7	<b>Revisão sistemática da literatura científica sobre o Software Goldfire.....</b>	<b>66</b>
2.7.1	<i>Síntese das publicações analisadas.....</i>	<i>67</i>
2.7.1.1	<i>Impacto.....</i>	<i>67</i>
2.7.1.2	<i>Abordagem teórica.....</i>	<i>68</i>
2.7.1.3	<i>Aplicado em estudo de caso.....</i>	<i>69</i>
2.7.1.4	<i>Mencionado no texto.....</i>	<i>74</i>
2.7.2	<i>Considerações sobre as análises.....</i>	<i>79</i>
3	<b>METODOLOGIA.....</b>	<b>81</b>
3.1	<b>Tipo de pesquisa.....</b>	<b>81</b>
3.2	<b>Etapas da pesquisa.....</b>	<b>81</b>
3.2.1	<i>Revisão sistemática da literatura científica sobre o Goldfire.....</i>	<i>81</i>

3.2.2	<i>Extração de relacionamentos semânticos de causa e efeito.....</i>	82
3.2.2.1	Escolha do corpus.....	82
3.2.2.2	Transferência do formato de arquivo para texto simples.....	83
3.2.2.3	Análise morfossintática .....	83
3.2.2.4	Análise semântica- processo manual.....	87
3.2.2.5	Análise semântica- processo automático.....	88
4	<b>RESULTADOS E DISCUSSÕES.....</b>	89
4.1	<b>Relacionamentos semânticos de causa e efeito.....</b>	89
4.2	<b>Verbos localizados nas sentenças de causa e efeito.....</b>	90
4.3	<b>Procedimentos para extração automática de sentenças de causa e efeito.....</b>	92
4.4	<b>Comparação entre as sentenças identificadas manualmente e automaticamente .....</b>	94
5	<b>CONSIDERAÇÕES FINAIS.....</b>	97
	<b>REFERÊNCIAS.....</b>	101
	<b>APÊNDICE A- CLASSE DE PALAVRAS (FORMA MORFOLÓGICA.....</b>	113
	<b>APÊNDICE B -SENTENÇAS DE CAUSA E EFEITO RECUPERADAS PELO PROCESSO AUTOMÁTICO E PELO PROCESSO MANUAL.....</b>	114

## 1 INTRODUÇÃO

Apesar da democratização, acessibilidade e uso da informação, a *internet* trouxe novas dificuldades, que limitam sua eficácia, especialmente no território da ciência, tecnologia e inovação. O enorme volume de informações disponíveis (e em constante crescimento) torna difícil não apenas recuperá-las, mas também gerenciar a recuperação e seus resultados de modo inteligente. Diante desse contexto, a gestão de recursos eficazes para organizar e dar acesso ao conhecimento tem se tornado importante para solucionar os problemas relacionados à informação e sua recuperação. Nesse âmbito, uma das áreas preocupadas e envolvidas com essa questão é a Ciência da Informação (CI) que, acompanhada de outras, é uma integrante ativa na evolução da sociedade da informação (SARACEVIC, 1996).

Com o advento das tecnologias digitais, observa-se que a verdadeira ruptura tecnológica reside na ideia da utilização, cada vez mais intensiva, dos computadores como uma espécie de “ferramenta universal”, capaz de armazenar e manipular representações simbólicas de forma automática. Tais tecnologias indicam perspectivas inovadoras a partir da desmaterialização dos suportes tradicionais e da constituição de um crescente “acervo universal” de informações virtuais (RAMALHO; OUCHI, 2011).

Em função dessa modificação ocorrida por meio da tecnologia digital, pesquisadores e profissionais da informação, atualmente, se preocupam em criar meios para recuperar e dar visibilidade aos documentos e ao conhecimento neles registrado. Dada a importância do conhecimento dentro da economia atual e, por outro lado, a incapacidade humana de trabalhar com tamanha quantidade de informações, infere-se que o desenvolvimento de ferramentas que “amplifique” a capacidade humana de procurar por informações relevantes para o seu negócio, dentro da enorme quantidade e variedade de publicações existentes atualmente, é estratégica para um novo modo de criação de tecnologia e inovação (MARCONDES, 2011).

Nesse sentido, tem se ajustado os “Observatórios de Ciência, Tecnologia e Inovação (CT&I)”, funcionando como facilitadores do processo de desenvolvimento de competências, métodos e técnicas para a realização de monitoramento e elaboração de diagnósticos (SANTOS, 2013). Tendo como missão principal viabilizar a agregação, sistematização e tratamento “inteligente” e coordenado de uma enorme gama de dados, oriundos de fontes diversas (nacionais e internacionais), de forma a garantir um maior grau de compatibilidade, complementaridade e comparabilidade entre eles (GUSMÃO, 2005).

Essa pesquisa se desenvolveu no âmbito do *Observatório Temático e Laboratório – Ensino, Tecnologia, Ciência e Informação (OtletCI)*, que desde 2010 vem sendo instituído no ambiente do Programa de Pós-Graduação em Ciência da Informação (PPGCI) da Universidade Federal de Pernambuco (UFPE), pela equipe do Grupo de Pesquisa SCIENTIA<sup>1</sup>. O *OtletCI* é um observatório científico e tecnológico que vem sendo projetado como um sistema de informações voltado para o fornecimento amplo, correlacionado e exaustivo de informações acerca de ciência, tecnologia, pesquisa, desenvolvimento e inovação. O nome *OtletCI* também é uma homenagem a Paul Otlet, um dos mais destacados precursores da Ciência da Informação. A visão ampla de Paul Otlet revolucionou não só o modo de trabalhar com a informação no seu tempo, mas também teve impactos, que perduram até hoje.

De acordo com Trzesniak (2014), o *OtletCI* é um sistema de informação de alto desempenho, funcionando como:

- laboratório, quanto à identificação e à descrição de informações de todas as naturezas, disponíveis formal ou informalmente, na *web* ou não, e na criação e desenvolvimento de relatórios e indicadores que permitam agrupar e correlacionar essas informações de múltiplas maneiras, podendo, assim, responder às mais variadas questões sobre competências científicas e tecnológicas, oportunidades de pesquisa, áreas do conhecimento emergentes e produção, no seu sentido amplo (incluindo livros e capítulos);
- meta-acervo (acervo sobre acervos), apontando a natureza e a localização de informações sobre qualquer microtema de pesquisa que tenha, em algum momento, aparecido como palavra-chave do autor em artigos de revistas, patentes ou teses de doutorado.

O *OtletCI* irá contribuir de modo efetivo, eficaz e significativo em várias dimensões: definição de políticas de pesquisa, estabelecimento de áreas prioritárias, sincronização com tecnologias emergentes e solução de problemas complexos. Trata-se de uma primeira aproximação, que possibilitará a projeção de um modelo organizacional, metodológico, operacional da estrutura de funcionamento de um observatório, projetado como um sistema de informações de alto desempenho, voltado para o fornecimento de informações acerca de ciência, tecnologia, pesquisa, desenvolvimento e inovação.

Quando se trata da informação disponível na *web*, verifica-se que as tecnologias desenvolvidas nesse ambiente têm provocado uma série de avanços e melhorias significativas no âmbito da representação de informações em ambientes digitais. A estrutura da *web* é

---

<sup>1</sup> Grupo de Pesquisa sob a liderança do Professor Doutor Raimundo Nonato Macedo dos Santos, certificado pela UFPE e cadastrado desde 2006 na Plataforma de Grupo de Pesquisa do CNPq

composta de linguagens de marcação com *hiperlinks*, que geram uma relação dinâmica entre documentos e as informações neles contidas. No entanto, essas linguagens descrevem unicamente *a forma* como a informação deve ser apresentada para o usuário pelo navegador, mas não expressam nada sobre o seu significado.

Neste contexto, é fácil para os computadores navegar nos dados da *web*, porém, dificilmente, se obtêm automaticamente uma interpretação semântica dos mesmos. Isto é equivalente a dizer que a *web* atual (2016) está no nível sintático (lido por máquinas) e não no nível semântico (compreendido por máquinas), esta é uma observação importante, pois as informações representadas na forma de texto livre não permitem estabelecer, de imediato, correlações entre os conceitos, explorar relações e evidenciar fatos que estão expressos nos textos.

Para García Sánchez (2008) o maior problema não é a quantidade de informações disponível no ambiente da *web*, uma vez que o usuário pode refinar a busca através de uma consulta mais detalhada, o problema surge quando se trata de obter informação sobre uma relação concreta entre dois conceitos. Para o autor acima citado, a principal limitação da *web* é o fato de que as tecnologias que a constituem não são capazes de capturar, ou representar formalmente, a semântica do conteúdo apresentado.

Tecnologias vêm sendo exploradas no contexto da Web Semântica, com base no projeto e na implementação de padrões de metadados que adicionam aos dados informações relevantes sobre seus contextos, marcando-os semanticamente; e com base em mecanismos de busca que levem em conta esses dados marcados (BERNERS-LEE et al., 2001).

Dentre as tecnologias da Web Semântica, encontra-se a busca semântica, permitindo que o conhecimento seja organizado em universos conceituais de acordo com seu significado. Nesse contexto, as palavras deixam de ser simples palavras para converterem-se em conceitos, e os buscadores evoluem de motores de buscas à máquinas de aprendizagem. Esse tipo de organização permite que ferramentas de busca por informação sejam capazes de fazer a seleção e a filtragem dessa informação baseadas na semântica dos termos de busca e dos itens pesquisados.

A busca semântica tenta compreender a intenção do usuário e o significado contextual dos termos usados na busca, como eles aparecem no espaço de dados pesquisáveis, seja na *web* ou dentro de um sistema fechado, para gerar resultados mais relevantes e precisos (AMANQUI, 2014).

Para Souza (2005), a utilização da semântica embutida nos próprios documentos deveria ser mais explorada, ou seja, as potencialidades intratextuais da linguagem natural para automatizar e melhorar as tarefas de indexação, organização e recuperação de informações. Segundo o autor, palavras isoladas, embora sirvam de forma bastante razoável aos propósitos de recuperação de informações, falham, em grande parte, por não considerarem o contexto informacional implícito em toda a consulta. Falta, porém, um preparo para lidar com a forma com que estas palavras ou conceitos estão relacionados, dessa forma, perdem-se informações fundamentais sobre o escopo em que as palavras estejam sendo utilizadas e, em consequência, a pertinência da pesquisa diminui. Esses relacionamentos, na prática, determinam as minúcias e especificidades dos assuntos pesquisados.

Muitas pesquisas desenvolvidas nos últimos anos procuram encontrar padrões que possibilitem agregar valor para superar os obstáculos da linguagem natural<sup>2</sup>. O objetivo destas pesquisas é “aumentar a eficiência dos mecanismos de busca e de outros tipos de ferramentas de processamento automático de documentos, através da utilização de linguagens que permitam definir dados e regras para o raciocínio sobre esses dados” (FERNEDA, 2003, p. 110).

Porém a complexidade associada a linguagem natural exige uma atenção especial quando se trata de recuperar a informação textual que satisfaça a necessidade de informação de um usuário. É nesse cenário que a automação da informação documentária tem proporcionado aos cientistas da informação uma aproximação com os estudos linguísticos, em especial os estudos semânticos, como consequência da necessidade de analisar o conteúdo de textos para deles se extrair elementos que permitam sua eficaz recuperação.

Quando um mecanismo qualquer atende uma consulta, devolvendo centenas de milhares de páginas e documentos de dez em dez, e hierarquiza essa devolução conforme critérios de popularidade e buscas anteriores do consulente, pode até atender demandas comerciais e de interesse geral, porém, dificilmente, priorizará uma informação científico-tecnológica: a resposta procurada, por ser especializada, aparecerá na 98756ª posição, quem terá paciência de ir até lá? Recursos como o *Google Scholar* tentam contornar esse problema, mas acabam caindo no outro extremo, de devolver documentos a menos, por não cobrir integralmente todas as possíveis fontes de interesse.

Os "mecanismos quaisquer", a que nos referimos, baseiam-se preferencialmente em presença, frequência e proximidade das palavras inseridas como critério de busca. É um método de força bruta, sem muita inteligência semântica, desconsiderando o significado das palavras e

---

<sup>2</sup> Linguagem natural é aquela falada espontaneamente por um grupo humano ou aquela escrita na obra por seu autor. Não possui funções específicas, funciona em muitos contextos diferentes e para propósitos diversos (NAKAYAMA, 1986).

o sentido de sua sequência, ou seja, vê as palavras como figuras vazias e não atribui qualquer significado à sua combinação (por exemplo, não distingue “aumento de juvenis em crimes” de “aumento de crimes em juvenis”). Isso o impede de ser científica e tecnologicamente mais eficaz: esse método apenas retorna o que recebe, não infere, não deduz, não relaciona.

Uma primeira alternativa de melhorar esse contexto é o uso de metadados, mas esses são em número limitado; dependem do discernimento do autor, que pode não destacar algum conteúdo do documento que seja exatamente o interesse do consulente; e não estão disponíveis para todos os tipos de informação.

Há, porém, um caminho adicional, muito mais promissor: e se além de localizar palavras, o *software* também entender *sentidos*, sendo capaz de identificar relações de causa-efeito, de comparar e fornecer vantagens e desvantagens das várias soluções de um problema específico, de localizar os *experts* num dado tema, enfim, de ir além dos *inputs*, de fato *acrescentando inteligência semântica às palavras ingressadas na consulta*? Essa tecnologia existe e está substanciada no IHS *GoldFire*, desenvolvido a partir de 1992, inicialmente pela empresa *Invention Machine* e, posteriormente, pela IHS. A tecnologia utilizada no presente *software* utiliza a indexação e busca semântica com a finalidade de realizar pesquisas de conhecimento.

O método é baseado em um processador semântico, que realiza preformatação de documentos de texto; análise linguística de documentos de texto/consultas de usuário (por exemplo: análise léxica, morfossintática, sintática e análise semântica); bem como a sua rotulagem semântica, usando termos (marcadores) de tipos de conhecimentos básicos (como objetos / classes de objetos, fatos e as regras que refletem regularidades do mundo/domínio de conhecimento, sob a forma de relações de causa-efeito), seus componentes e atributos (TODHUNTER; SOVPEL; PASTANOHAU, 2014).

É possível observar que o potencial semântico do *software Goldfire* e seus recursos avançados são capazes de manipular e realizar inferências, reconhecendo relações no conteúdo dos textos, não apenas localizando informações para o usuário, mas também sendo capaz de compreender o conteúdo de textos em linguagem natural.

O problema é que esse *software* não abrange nenhum idioma néo-latino, excluindo da base mundial de conhecimento tudo o que é produzido em português e espanhol, entre outros idiomas. Então, surge a questão: como estabelecer uma rotina computacional capaz de ler, interpretar e marcar textos científicos e tecnológicos em português, de modo a possibilitar sua inclusão em buscas semânticas inteligentes?

O presente trabalho dá os primeiros passos nessa iniciativa. Trabalhando diretamente com resumos da área de Ciências Agrárias, localizam-se sentenças que envolvam relações de

causa-efeito. Por outro lado, empregando recursos computacionais de identificação morfológica e sintática, decompõem-se e se recompõem os textos, igualmente destacando-se sentenças que se presume atendam as mesmas condições (relações de causa-efeito). A relativa convergência dos resultados obtidos por uma e outra via, e o estabelecimento preliminar de rotinas para a versão automatizada são o resultado aqui alcançado, que se encontra ilustrado para alguns dos resumos estudados.

A escolha pelo estudo dos elementos causais presentes nos textos de artigos científicos da área de Ciências Agrárias em detrimento de outros recursos da linguagem se deu pela percepção de que formulações de relações causais marcadas nos textos poderia tornar-se mecanismo central no processo de construção e recuperação do conhecimento científico. Assim, tratou-se de privilegiar a observação da causalidade expressa por meio de verbos.

Para tanto, investiga-se os fundamentos teóricos e a aplicação da tecnologia da busca semântica como recurso na recuperação precisa e eficaz da informação; com embasamento na epistemologia da Ciência da Informação; com aportes da linguística computacional, ou Processamento de Linguagem Natural, e das tecnologias oriundas da Web Semântica e mineração de texto.

Pouco explorada, entretanto, é a utilização da semântica embutida nos próprios documentos, ou seja, as potencialidades intratextuais da linguagem natural, para automatizar e melhorar as tarefas de indexação, organização e recuperação de informações. Nesse sentido, partiu-se do pressuposto de que a extração e representação computacional de relações de causa e efeito, pelo maior grau de informação semântica embutida, podem vir a se tornar mais eficazes do que as palavras-chave usualmente extraídas e utilizadas como descritores em outros processos automatizados de representação de documentos.

Aprofundar no conhecimento da tecnologia de busca semântica pode ajudar pesquisadores e gestores de desenvolvimento e inovação a realizar melhores planos, projetos e decisões, permitindo, deste modo, ganhos significativos em vantagem competitiva. Daí a importância de projetos que apliquem essa tecnologia ao problema de busca e demonstrem seus resultados através de experimentos. Espera-se, com esta pesquisa, viabilizar a identificação de indícios de novas descobertas científicas no âmbito do *Observatório Temático e Laboratório – Ensino, Tecnologia, Ciência e Informação (OtletCI)*.

## 1.2 Objetivos e estrutura da dissertação

### 1.2.1 Objetivo Geral

Investigar a tecnologia da busca semântica quanto aos fundamentos teóricos, sua utilidade no contexto do *Observatório Temático e Laboratório – Ensino, Tecnologia, Ciência e Informação (OiletCI)* e requisitos para aplicação em textos científicos em português.

### 1.2.2 Objetivos específicos

- Descrever a tecnologia da busca semântica quanto aos fundamentos teóricos;
- Apontar a utilidade da busca semântica no contexto do observatório;
- Apresentar, mediante uma revisão sistemática da literatura científica sobre o *Software Goldfire* o que vem sendo estudado pelos pesquisadores acerca dos experimentos práticos para os quais o software foi utilizado;
- Explicitar os requisitos da busca semântica para a aplicação em textos científicos, em português, através da análise da extração de relacionamentos semânticos do tipo “causa e efeito” em resumos, em português, de artigos científicos da área de ciências agrárias.

Dessa forma, a presente pesquisa está organizada em seis seções. Nesta *Introdução* foram apresentados o contexto da pesquisa; o problema; a questão de pesquisa; os objetivos e a forma em que foram encadeadas as temáticas a serem tratadas ao longo do trabalho.

Na seção dois, intitulada *Fundamentação Teórica*, como forma de fundamentar e embasar teoricamente os assuntos tratados neste trabalho, bem como os conceitos pertinentes para uma melhor compreensão do estudo, apresenta, em suas subseções, reflexões sobre os Observatórios de Ciência e Tecnologia de um modo geral. Depois, trata da recuperação semântica da informação, fazendo um relato histórico sobre a evolução da tecnologia desde Gutenberg até a Web Semântica. Enfatiza a busca semântica, dando ênfase aos motores de busca semântica e discute as barreiras linguísticas, principalmente no que tange a recuperação semântica de documentos na língua portuguesa do Brasil. Em seguida, trata sobre o

Processamento de Linguagem Natural, contextualizando a interpretação/automação da semântica dos textos, dando continuidade com as tecnologias de mineração de texto e enfatizando a extração automática do conhecimento em documentos textuais. Finaliza-se a seção com uma breve explanação sobre a Teoria da Solução Inventiva de Problemas – TRIZ.

A segunda seção apresenta também a revisão sistemática sobre o *Software Goldfire* (a qual se constitui como um dos resultados da pesquisa), seus recursos e seu potencial semântico. As vantagens e benefícios que podem ser conquistados com o uso do *Goldfire* foram descritos e estabelecidos através da análise da produção científica sobre o *software*, com o objetivo de revelar os experimentos práticos para os quais o *Goldfire* foi utilizado, enfatizando como os pesquisadores têm usado o *software* em termos de ferramenta para resolver problemas, ou seja, em quais aplicações este tipo de tecnologia foi utilizada.

A seção três, *Metodologia*, apresenta os procedimentos metodológicos utilizados para realização da pesquisa, isto é, a metodologia empregada para a descrição e análise dos resultados, sendo traçados, assim, os caminhos percorridos para construção de cada etapa.

Na seção quatro, *Análise dos Resultados*, é apresentada a análise do trabalho realizado com os resumos da área de Ciências Agrárias, através da marcação das sentenças que envolvem relações de causa-efeito. Empregando recursos computacionais de identificação morfológica e sintática, decompõem-se e se recompõem os textos, igualmente destacando-se sentenças que se presume atendam as mesmas condições (relações de causa-efeito).

A seção cinco diz respeito às *Considerações Finais* do trabalho, onde são apontadas algumas reflexões acerca da pesquisa realizada e das contribuições identificadas. Este trabalho apresenta também apêndices que complementam informações necessárias ao entendimento de alguns aspectos apresentados nesta pesquisa.

## 2 FUNDAMENTAÇÃO TEÓRICA

Com o objetivo de fundamentar e embasar teoricamente os assuntos tratados nesse trabalho, bem como os conceitos pertinentes, são apresentados nesta seção os marcos teóricos, necessários para uma melhor compreensão do estudo. Visa também apresentar, mediante uma revisão sistemática da literatura científica sobre o *Software Goldfire* o que vem sendo estudado pelos pesquisadores acerca dos experimentos práticos para os quais o software foi utilizado.

### 2.1 Observatórios de C,T&I

Diferentes empresas do segmento de tecnologia da informação competem pela inovação, buscando tornarem-se as mais rápidas, as de maiores bases catalogadas, as que oferecem melhores consultas às necessidades do usuário etc. A medida que as empresas inovam, a busca por essa diferenciação cria diferentes visões sobre a forma pela qual a informação é tratada (COUTO, 2011).

No Brasil, nos anos recentes, as políticas públicas voltadas à área de Ciência, Tecnologia e Inovação (C,T&I) têm adquirido relevância. Muitas das ações do Ministério da Ciência, Tecnologia e Inovação (MCTI)<sup>3</sup> vão ao encontro da ideia de que o Brasil precisa de um modelo de desenvolvimento que faça a aliança entre o conhecimento científico e a economia, nos quais a ciência realmente seja projetada nas atividades econômicas e leve benefícios mais direta e rapidamente à sociedade. Para isso, o MCTI considera imprescindível o fortalecimento dos institutos de pesquisa e, naturalmente, das universidades, que são parte importante do sistema, que contempla a ciência como fator de geração de riqueza. Segundo Raupp (2010), o papel fundamental da universidade é a formação de profissionais qualificados, para satisfazer às diversas demandas da sociedade, além da realização de uma pesquisa científica que contribua para a evolução do conhecimento em suas diferentes áreas.

Conceitualmente, essas instituições de pesquisas podem ser entendidas como um conjunto de organizações, públicas e privadas, diferentes e interconectadas que, individualmente ou em conjunto, contribuem para a geração, estoque, transferência e difusão de informações, conhecimento, competências e tecnologias, proporcionando a estrutura pelas quais governos

---

<sup>3</sup> <http://www.mcti.gov.br/>

podem criar e implementar políticas para guiarem o processo de inovação na sociedade (YU ABRAHAM et al., 2006).

Nesse sentido, os observatórios nascem numa sociedade onde a informação e o conhecimento são os principais “motores” do desenvolvimento político, social, cultural e econômico de um país ou de uma região e adquirem um papel essencial na tomada de decisões, no projeto e na evolução de tais políticas.

O termo “observatório”, com base nos dicionários Houaiss, Aurélio e Michaelis, significa “lugar de onde se observa qualquer coisa; mirante ou miradouro; Ato ou efeito de observar; observação; Instituição ou serviço de observações astronômicas ou meteorológicas”.

Os sinônimos de observar denotam a delimitação do conceito de observatório: olhar, contemplar, considerar, vigiar etc. Ainda que sua origem esteja vinculada aos observatórios astronômicos, os observatórios atuais, dirigidos a temas socioeconômicos, começam a surgir, especialmente, a partir do final da década de 80, como ressaltava Gusmão (2006, p. 3): o conceito de observatório deixa de ser exclusivo para fatores ligados à natureza, passando a abranger também os fenômenos sociais. Segundo esta autora, “do exame dos fatos relativos aos temas que acontecem naquele cenário (dos fenômenos observados), as pessoas ligadas ao observatório passam a compilar informações que serão úteis para apoiar processos relativos à tomada de decisão”.

Em um estudo realizado por Martins (2007), foi possível observar que, a partir do início da década de 90, com a generalização da internet e da World Wide Web, o “movimento observador” aumenta e se desenvolve em vários domínios: são referenciadas mais de uma dezena de observatórios, em áreas tão diversas quanto a luta contra a pobreza em Lisboa, as pequenas e as médias empresas, a droga e a toxicodependência, o emprego e a formação profissional, a imprensa ou a justiça. Ao desenvolvimento global da informática, em particular das bases de dados programadas na *web*, correspondeu o enorme crescimento – verificado entre 2000 e 2007.

Uma das circunstâncias fundamentais que favorecem a criação de observatórios, segundo Ortega e del Valle (2010), é a necessidade de sistematizar as diferentes fontes de informação existentes de caráter parcial, com a finalidade de oferecer uma fonte de informação global. Os observatórios surgem para a satisfação desta necessidade, implantando sistemas de informação de dados completos, integrais, confiáveis e disponíveis, a partir de diferentes fontes.

Diante do exposto, depreende-se que o fato de existirem múltiplas visões e concepções acerca dos observatórios não impede que se estabeleçam algumas linhas-mestre de direcionamento e orientação. Em um estudo feito por Silva et al. (2013), é possível observar diretrizes norteadoras que, adequadas aos diferentes modelos, podem funcionar como verdadeiras “boas práticas” na condução destes organismos. A seguir, serão destacadas algumas diretrizes, conforme o estudo do autor acima citado:

- **Organização de (e para o) conhecimento:** observatórios devem atuar como organismos voltados ao conhecimento e à sua prospecção, criação e descoberta, ou seja, ao

conhecimento novo. Isso significa a ruptura, tanto da “síndrome da mera informação” quanto da “síndrome do repositório”, verificadas em vários observatórios. A primeira, diz respeito à excessiva valorização e difusão de abordagens que apenas informam, e não geram ou propiciam conhecimento e, portanto, aprendizagem. A “síndrome do repositório” se refere à adoção de um perfil limitado ao registro, à reprodução e ao armazenamento de dados e/ou às informações pré-existentes.

- Olhar dirigido ao futuro: a análise contínua do observatório sobre a realidade importa, não só para compreender o passado, mas, sobretudo, para permitir uma análise prospectiva, de forma a ‘antecipar’ os problemas que surgirão.
- Espaço para interatividade e participação: a efetividade dos observatórios está diretamente relacionada à sua capacidade de atender às necessidades e expectativas de seu público-alvo, formado tanto pela rede de colaboradores quanto pelos usuários. A adoção de uma política de interface com o cidadão deve ser um ponto focal dos observatórios.
- Produtos customizados e contextualizados: é imperativo que se busque agregar valor e qualidade à informação e ao conhecimento gerados, mas não apenas retransmiti-los.
- Articulação em redes: os observatórios catalisam a formação de redes de pessoas e de instituições parceiras, bem como se estruturam em torno de redes temáticas de conhecimento com outros observatórios, fortalecendo o seu poder de ação e de influência, por meio da troca de experiências, *expertise* e informações.
- Agente de mediação do conhecimento: os observatórios precisam consolidar seu espaço na sociedade do conhecimento, firmando-se como elementos diferenciados e, como consequência, necessários. Evoluindo para a condição de órgãos de mediação do conhecimento.
- Avaliação permanente, acreditação e certificação: os observatórios necessitam se submeter a programas de avaliação como condição básica para alcançarem efetividade e desfrutarem de confiabilidade, seja pública e/ou de seus órgãos mantenedores. É preciso institucionalizar uma política de avaliação permanente, dispondo de ferramentas e critérios que questionem objetivos, metodologias, resultados e impacto social. A política de avaliação terá efeito ainda maior, se associada a processos de acreditação e/ou certificação formatados em moldes técnicos e executados por organizações idôneas

e habilitadas para tal, permitindo a devida visibilidade aos padrões de qualidade das entidades que são, de fato, observatórios.

Tendo em vista que o campo de observação atravessa múltiplos interesses, métodos e dispositivos que se influenciam mutuamente para atingir diversos fins, vale ressaltar que o foco da presente pesquisa não são os observatórios relacionados às áreas de clima, astronomia, saúde, educação, violência, trabalho e direitos humanos, mas aos que estão relacionados à área de (C,T&I).

Gusmão (2005) enfatiza uma das características principais dos observatórios de (C,T&I), onde seu caráter – de certa forma, inovador – reside no fato de não estar associado à produção de dados primários. Trata-se de identificar e processar dados já existentes. Tendo como principal missão viabilizar a agregação, a sistematização e o tratamento “inteligente” e coordenado por uma enorme gama de dados, oriundos de fontes diversas (nacionais e internacionais), de forma que se possa garantir um maior grau de compatibilidade, complementaridade e comparabilidade entre eles.

Trzeciak (2009) considera que um observatório dessa natureza pode trazer benefícios, por meio do fornecimento de informações e análises relevantes, auxiliando na tomada de decisão, na proposição de novas políticas, na identificação de novos negócios, no compartilhamento de informações, na articulação entre organizações e indivíduos, estimulando a participação efetiva dos principais interessados, na construção de visões a curto, médio e longo prazos, em assuntos que tenham a Ciência, Tecnologia & Inovação (C,T&I) como pontos centrais.

Em se tratando de aspectos referentes aos observatórios de C,T&I, tais como conceitos, benefícios, funções e funcionamento, será apresentada, segundo (TRZECIAK, 2009), uma síntese:

**Quadro 1:** aspectos referentes aos observatórios de C,T&I

<b>Quanto ao conceito</b>	Um observatório fornece informações estratégicas para auxiliarem na identificação de ameaças, oportunidades e tendências, além de oferecerem subsídios à tomada de decisão, de um determinado setor, seja este em nível regional, nacional ou internacional, visando –incrementar a competitividade do setor e o desenvolvimento da economia do país.
<b>Quanto aos benefícios</b>	Conhecer melhor os sistemas de P&D de determinada região, medir e avaliar os esforços de C,T&I, oferecer elementos para a proposição de diagnósticos, identificar as tendências de determinado setor, monitorar novas oportunidades, gerar conhecimento novo, contribuir para a elaboração de indicadores, promover a interação entre universidade-empresa,

	apoiar a formação de redes de especialistas em temas específicos, dentre outros.
<b>Quanto às funções</b>	A principal função é o fornecimento de informações com valor agregado, para auxiliar na tomada de decisão e nas análises estratégicas dos vários atores envolvidos (universidades, governos, empresas etc.). Em suma, verifica-se que um observatório tem como papel essencial transformar dados em informações, oferecendo produtos/serviços (relatórios, análises periódicas, análises prospectivas, dentre outros), para atenderem às demandas específicas do público a que se destina.
<b>Quanto ao funcionamento</b>	Possuem missões específicas, envolvendo um conjunto de atividades variadas, mas que possuem articulação entre si. Os observatórios executam estas atividades de forma direta ou indireta e com diferentes graus de intensidade, independente do modelo de operação e do formato institucional que adotam.

**Fonte:** elaborado e baseado em TRZECIAK, 2009.

Um levantamento realizado por Trzesniak (2014) apontou cerca de trinta observatórios internacionais, operando em países de mais variados estágios de desenvolvimento, como:

França, Canadá, Alemanha, Suíça, Portugal, Espanha, Finlândia, Irlanda, Japão, Coreia do Sul, Colômbia, Cuba, Peru, Costa Rica, Malásia, Singapura, Tailândia, Líbano, Síria, Tunísia, África do Sul, na maioria dos casos, com financiamento governamental. Porém, mais do que isso, há observatórios criados e mantidos por consórcios multinacionais, a exemplos a Comunidade Europeia (operam, sob a referida responsabilidade, o ERAWACH e o ETEPS, sucessores da primeira iniciativa desse continente, e o ESTO) e as Nações Unidas, que planejam o GO-SPIN (baseado no protótipo SPIN), um grande observatório mundial que operaria em parceria com o ERAWACH. Outros destaques são o ESCWA, cobrindo a região do Oriente Médio, o AOSTI, da Comunidade Africana e o RICyT, que cobre a América Latina. O foco principal, comum a todos, é a Pesquisa e Desenvolvimento em Ciência, Tecnologia e Inovação.

No Brasil, apesar de existirem importantes manifestações dos setores públicos e privados, com relação à C,T&I, são poucas as experiências relacionadas a observatórios, principalmente nessa área (TRZECIAK, 2009; SILVA et al. 2013). Merece destaque o Observatório de Tecnologia e Inovação (OTI), que teve início em 2002, por intermédio do IPT (Instituto de Pesquisas Tecnológicas do Estado de São Paulo) e da Secretaria de Ciência, Tecnologia, Desenvolvimento Econômico e Turismo do Estado de São Paulo. O Observatório da Inovação e Competitividade, sediado no Instituto de Estudos Avançados (IEA) da Universidade de São

Paulo, fora lançado em outubro de 2007. E o Centro de Gestão e Estudos Estratégicos (CGEE), organização social ligada ao MCT, lançada em 2001.

### **2.1.1 Formas de atuação dos observatórios**

Verifica-se que não há um modelo único de observatório. Alguns deles surgiram para atenderem a uma determinada demanda existente; outros, como uma oportunidade de oferta de produtos e serviços de informação estratégicos para determinado setor; e outros, ainda, com base em observatórios já existentes. Todavia, diferenciam-se em termos de objetivos, serviços, produtos, missão, estrutura organizacional e funcional e fontes de apoio financeiro.

Portanto, os observatórios assumem diferentes concepções, finalidades, atividades e desenhos conceituais. Desde as administrações públicas, nacionais, regionais e locais; desde instituições acadêmicas, sindicatos, empresas, fundações e entidades não lucrativas, vem sendo aprimoradas e sustentadas a criação de observatórios.

Os observatórios podem ser temáticos, pois estudam, registram e analisam a situação e a evolução de determinada área temática. Existem observatórios relacionados com temas multivariados: ciência e tecnologia, violência de gênero, imigração, racismo e xenofobia, saúde, sustentabilidade, globalização, responsabilidade social corporativa, inovação, terceiro setor etc.

Para Trzesniak e Santos (2014), independente de sua natureza, os observatórios, em geral, estabelecem um foco, ou seja, definem limites temáticos e de âmbito para sua atuação. O âmbito pode ser uma instituição, um bairro, uma cidade, um continente ou uma comunidade. O tema pode ser uma área do conhecimento, restrita, como história ou física, ou ampla, como Ciências Sociais Aplicadas, ou, então, um ramo de atividade humana, como o setor hoteleiro ou o industrial. A escolha é uma questão de demanda, interesse, vocação e competência. Segundo os respectivos autores, há quatro formas de atuação dos observatórios:

**Quadro 2:** Formas de atuação dos observatórios

<b>Básico ou de primeira ordem</b>	Identifica fontes de dados já coletados, que sejam relevantes para o seu foco e, a partir destes, gera e armazena séries históricas correspondentemente descritivas, oferecendo-as a pesquisadores, formadores de opinião e formuladores de políticas em geral.
<b>Integrador ou de segunda ordem</b>	Agrega informações colhidas de observatórios básicos, de modo que seja ampliada, essencialmente, a região (geográfica) abrangida. Por exemplo, um Observatório Nacional, que se alimenta de Observatórios Estaduais.
<b>De ordem zero</b>	Pode haver situações em que um observatório trabalhe como provedor de dados primários, indo buscá-los diretamente em campo, mas, obviamente,

	dentro de um padrão adequado à sua utilização, para sua atividade fim. Ou seja, a coleta é feita visando à utilização para os propósitos do observatório. Não deve ser confundido com provedores de dados que os colhem para finalidades próprias, não pensando em observatórios. Com esses dados, conforme mencionado, os observatórios têm a função de recuperá-los, avaliá-los, retratá-los e, eventualmente, aproveitá-los, da melhor maneira.
<b>De estudos e prospecção</b>	Como categoria adicional, esses observatórios escolhem e focalizam um tipo específico de problema e tratam de descobrir tudo sobre ele, num determinado prazo. Depois, migram para outro, e assim por diante. Diferente dos anteriores, esses não se preocupam com a construção de séries temporais de indicadores, mas com exaurir o que se é capaz de descobrir de algum contexto ou objeto, ou seja, não são observatórios de terceira ordem. São, sim, de outra natureza e o que têm em comum com os demais é o fato de <i>observarem</i> de modo similar: garimpando informações, a partir de dados existentes. O propósito, porém, difere bastante.

**Fonte:** (TRZESNIAK; SANTOS, 2014)

Vale ressaltar que existe uma diferença significativa entre *básicos* e *integradores*. Os primeiros trabalham com dados provenientes das mais diversas fontes, a maioria cuja criação não foi realizada para suprir um observatório, mas para atender demandas como avaliação, gerenciamento e acompanhamento de processos. Eventualmente, na recuperação de séries históricas, precisará lidar com dados não disponíveis *online* ou, até mesmo, existentes apenas em papel. Os integradores, por sua vez, recebem informações já no “padrão observatório”. A ênfase do seu trabalho é gerar indicadores compostos – que são mais complexos –, efetuar estudos comparativos e analisar evolução. Assim, o principal trabalho na retaguarda difere nos dois casos (TRZESNIAK; SANTOS, 2014).

Os observatórios devem funcionar como “antenas, radares, ou faróis para antecipar ideias, tendências, iluminar caminhos futuros, identificar questões e soluções” (CGEE, 2006, p.23). Foi possível observar que as características dos observatórios variam, de acordo com seu campo de ação, seus alcances e sua cobertura e, também, dependem do fenômeno a observar, assim como das distintas dimensões ou facetas do objeto de observação.

Como já fora salientado, não há um modelo único de observatório. Silva et al. (2013) afirmam que, como fruto de uma sociedade que, cada vez mais, produz e consome informação e conhecimento, os observatórios vêm-se expandindo em nível mundial, tendo, como uma de suas características, a diversidade tipológica. Verificam-se diferenças, em termos de origem, temática abordada, finalidades (objetivos), metodologias, serviços e produtos, estrutura organizacional, âmbito de atuação, dependência administrativa, fontes de financiamento, grau de evolução, entre outros pontos. Nesse sentido, Yu Abraham et al. (2006) enfatizam que:

Todos eles diferem, por exemplo, em termos de suas missões, serviços e produtos, estrutura organizacional e fontes de apoio financeiro. A definição do foco operacional de um observatório depende, claramente, das necessidades dos seus clientes e promotores. Em termos operacionais, a dependência é fortemente vinculada à disponibilidade de recursos (financeiros, materiais e humanos).

A criação de Observatórios de C,T&I deve ser vista como uma oportunidade, para que haja um espaço de discussão e construção de consensos entre os provedores de informação, produtores de indicadores e usuários. Trazendo como missão não apenas manter uma estrutura capaz de produzir indicadores sofisticados e, estatisticamente, confiáveis, mas, sobretudo, fazer com que isso contribua para a reflexão e o debate, de maneira mais ampla, apoiados sobre uma base ampla e variada de atores e co-participantes.

## **2.2 Recuperação semântica da informação**

Nesta seção investiga-se os fundamentos teóricos da busca semântica e a aplicação das tecnologias semânticas como recurso na recuperação precisa e eficaz da informação. Inicia-se com um relato histórico sobre a evolução da tecnologia desde Gutenberg até a Web Semântica.

### ***2.2.1 Evolução da tecnologia: contexto histórico***

A tecnologia da informação tem uma maneira surpreendente de mudar a cultura, muitas vezes, de formas inimagináveis para os inventores. Quando Gutenberg inventou a prensa, no meio do século XV, o seu objetivo principal foi desenvolver um mecanismo para acelerar a impressão de Bíblias. Gutenberg, provavelmente, nunca pensou em sua tecnologia, em termos de difusão geral do conhecimento humano, através de meios impressos. A prensa de Gutenberg impactou, diretamente, na relação que a tecnologia detinha com a sociedade, pela questão religiosa da época. Por meio da imprensa era possível que cada cristão tivesse a palavra em seu lar. Ou seja, a epistemologia da comunicação oral, cujo detentor era um sacerdote, foi alterada (RIBEIRO et al., 2004)<sup>4</sup>. Gutenberg, procurando uma maneira melhor para produzir Bíblias, mudou fundamentalmente o controle do conhecimento na sociedade ocidental. Dentro de alguns anos, qualquer pessoa que possuísse uma prensa para impressão poderia distribuir amplamente os conhecimentos para alguém, disposto a lê-lo.

---

<sup>4</sup>[http://www.uniritter.edu.br/eventos/sepesq/vi\\_sepesq/arquivosPDF/27194/2426/com\\_identificacao/Anexo%20com%20t\\_tulo%2000862.pdf](http://www.uniritter.edu.br/eventos/sepesq/vi_sepesq/arquivosPDF/27194/2426/com_identificacao/Anexo%20com%20t_tulo%2000862.pdf)

Tecnologias herdaram partes das tecnologias que as antecederam, combinando-as. Esse efeito reverte em abruptas as aparências das inovações radicais, anteriormente menos abruptas. Em suma, as tecnologias recentes se formam usando as primitivas como componentes. O conjunto de tecnologias promove outras novas e a tecnologia se cria a partir de si própria (RIBEIRO et al., 2004)<sup>5</sup>.

Com o surgimento da Documentação, introduzida por Paul Otlet e Henri La Fontaine, em meados de 1930, a questão da recuperação da informação começou a receber maior visibilidade. Segundo Mattelart (2002), Otlet e La Fontaine desenvolveram técnicas e ferramentas de organização e tratamento de uma grande massa de conhecimento registrado, evidenciando seu trabalho de documentação. Paul Otlet, quando participou do Movimento Bibliográfico, no final do século XIX, sua intenção era dar à documentação um caráter científico. Diversos instrumentos documentários foram concebidos e construídos por Otlet, durante seu trabalho no Institut International de Bibliographie (IIB). Segundo Santos (2007), seu mérito, sem dúvida, foi o de ter reunido teoria e prática em um trabalho incansável, para a consolidação de novas metodologias para a análise e síntese do conhecimento, visando à sua circulação. A erudição, o interesse pelas ciências em geral e pelas Ciências Sociais em particular e seu entusiasmo por certas vertentes do positivismo e cientificismo, dominantes na época, marcaram sua atuação.

Esses são os principais traços do modelo desenhado por Paul Otlet, para pensar e trabalhar o conhecimento e a informação, prenunciando as formas de tratamento e circulação da informação que vieram a ser adotadas, a partir da segunda metade do século XX.

Problemas informacionais existem, há um longo tempo, e sempre estiveram mais ou menos presentes, mas sua importância real ou percebida mudou e essa mudança foi responsável pelo surgimento da Ciência da Informação. Os avanços científicos e tecnológicos, impulsionados pela Segunda Guerra Mundial, foram fatores decisivos para o surgimento da CI. Apesar da existência da Documentação, que já havia tomado para si a árdua tarefa de organizar a informação científica, o crescimento exponencial do volume organizacional exigia o aperfeiçoamento das técnicas de organizar e recuperar a informação. No novo cenário mundial, além de organizar, era necessário ter um maior controle sobre a informação produzida e, principalmente, recuperá-la de forma mais rápida e eficaz.

---

<sup>5</sup>[http://www.uniritter.edu.br/eventos/sepesq/vi\\_sepesq/arquivosPDF/27194/2426/com\\_identificacao/Anexo%20com%20t\\_tulo%2000862.pdf](http://www.uniritter.edu.br/eventos/sepesq/vi_sepesq/arquivosPDF/27194/2426/com_identificacao/Anexo%20com%20t_tulo%2000862.pdf)

Outro acontecimento histórico, que marca o impulso do desenvolvimento e a própria origem da Ciência da Informação (CI), pode ser identificado no artigo *As we may think*, de VANNENVAR BUSH, respeitado cientista e chefe do esforço científico americano, durante a Segunda Guerra Mundial (BUSH, 1945). Também preocupado com a recuperação da informação, Bush se concentra na ideia de instrumentos de registro e transmissão de informação, que ele considera estarem entre os principais desafios para os cientistas – como ler e entender tantos artigos e relatórios e acessar tantas informações e, ali, selecionar o que é relevante –. O problema era (e, basicamente, ainda é) a tarefa massiva de tornar mais acessível um acervo crescente de conhecimento. Para Bush, se é para ser útil à ciência, deve ser, continuamente, prolongado, armazenado e, acima de tudo, consultado (BUSH, 1945).

Bush propôs uma máquina, chamada MEMEX, que seria capaz de associar ideias, conceitos ou palavras, para a organização da informação, argumentando que esse seria o padrão utilizado pelo cérebro humano para transformar informação em conhecimento. Segundo o autor, os processos repetitivos de pensamento não se encontram presos a questões simplesmente aritméticas ou estatísticas. Efetivamente, cada vez que combinamos e registramos, de acordo com certos processos lógicos estabelecidos, o aspecto criativo do pensamento entra em jogo, unicamente, na seleção dos dados e no processo a ser utilizado, a manipulação posterior é de natureza repetitiva e suscetível de ser realizado por uma máquina. Bush deixa claro que o ser humano não pode sonhar com a duplicação do processo mental artificialmente, mas deve ser capaz de aprender com ele e melhorá-lo, em alguns pequenos detalhes (BUSH, 1945).

Saracevic (1996), ao explorar a evolução e a natureza mutante da CI, observou que cientistas e engenheiros de todo o mundo e os mais importantes governos e agências de financiamento, em muitos países, se preocuparam com o problema da explosão informacional, identificado por Bush. Possivelmente, esses esforços foram responsáveis pelo desenvolvimento da moderna indústria da informação e das concepções que a direcionam. O raciocínio original, que fundamentou tais esforços, era o seguinte: uma vez que a ciência e a tecnologia são críticas para a sociedade (por exemplo, para a economia, a saúde, o comércio e a defesa), é, também, crítico prover os meios para o fornecimento de informações relevantes a indivíduos, grupos e organizações com elas envolvidas, pressupondo que a informação é um dos mais importantes insumos para se atingir e sustentar o desenvolvimento em tais áreas (SARACEVIC, 1996).

O aparecimento dos computadores e sua aplicação em sistemas de recuperação e disponibilização de informações, em larga escala, e a necessidade de idealizar estes sistemas em um funcionamento, otimizado em ambientes digitais, favoreceu para que decisões implícitas, relativas às metodologias de representação da informação, tivessem que ser

evidenciadas. Os computadores passaram de máquinas de processamento numérico para máquinas de processamento de conteúdos. Porém, a maior parte do conteúdo digital armazenado está em formato textual, legível por seres humanos, mas não, ainda, em formato adequado ao processamento semântico por computadores (MARCONDES, 2011).

Ainda no final do século XX, Berners-Lee, com o objetivo de fornecer acesso rápido aos relatórios técnicos online e a outros documentos, criados por laboratórios de física de alta energia do mundo, buscou tornar mais fácil o acesso à literatura distribuída, a partir de uma variedade de centros de pesquisa, espalhados pelo mundo. Nesse processo, Berners-Lee lançou as bases para a World Wide Web. No entanto, só em 1989, Berners-Lee pôde imaginar como a sua proposta de vincular relatórios técnicos via hipertexto, algum dia, poderia mudar fundamentalmente os aspectos essenciais da comunicação humana e interação social (DAVIES; STUDER; WARREN, 2006). Não era sua intenção revolucionar a comunicação da informação para e-commerce, para o raciocínio geográfico, para serviços governamentais ou por qualquer uma das inúmeras aplicações baseadas na *web*. No entanto, a sociedade mudou, de forma irreversível, quando Berners-Lee inventou a *Linguagem de Marcação de Hipertexto* HTML e o Protocolo de Transferência de Hipertexto HTTP (em 1989).

Segundo Trzesniak (2012, p.85), o início dos anos 90 foi a época da experimentação, da descoberta, da aventura no espaço virtual. Por isso, foi, pelo autor, denominada de “web romântica”, nome (licença didática do autor) que contrasta com a fase mais evoluída e profissional, conhecida como a web semântica. Foi nessa época que apareceram os pioneiros das revistas eletrônicas, como a Scientific Electronic Library Online (SciELO) e a Fapesp (Fundação de Amparo à Pesquisa do Estado de São Paulo) (TRZESNIAK, 2012, p.86).

Nesse caso, não se tratava de aventureiros, mas desbravadores e descobridores, sem dúvida: profissionais da Biblioteconomia, com larga experiência em tratamento, disseminação e recuperação da informação, que souberam, com competência, extrair o máximo e o melhor dos recursos da época, sendo, com toda justiça, recompensados, por um enorme sucesso que – cumpre ressaltar – se pereniza, porque também sabem acompanhar as evoluções, que não param de surgir (TRZESNIAK, 2012, p.86).

É certo que, ao longo do tempo, inúmeros pesquisadores vêm trabalhando na concretização de soluções para a comunicação homem-máquina. Nessa perspectiva, muito se tem discutido acerca de inovações e possibilidades das ferramentas de cunho semântico nos sistemas de recuperação da informação, em geral.

### 2.2.2 Web Semântica

A World Wide Web foi concebida por Tim Berners-Lee, em 1989, a partir de um projeto do CRN (Conselho Europeu para a Investigação Nuclear). Atualmente, vem sendo convertido em um gigantesco repositório de informação, em contínuo crescimento, fator que a torna em uma plataforma universal para a produção, disponibilização, armazenamento, compartilhamento e acessibilidade à informação.

Na década 2000, houve uma intensa atividade no desenvolvimento das ideias baseadas na Web Semântica. Esse esforço mundial foi coordenado pelo World Wide Web Consortium (W3C), por meio da seção, W3C Semantic Web Activity. A proposta, delineada por Berners-Lee et al. (2001), acentua a questão de que conteúdos na Web Semântica teriam “significado” para programas e a informação adquire significado, mediante o uso de metadados, para prover uma categorização semântica de seu conteúdo, tal como permitir um raciocínio automatizado sobre a informação.

O termo Web Semântica ficou conhecido através do artigo *The Semantic Web*, publicado na Scientific American em maio de 2001, de autoria de Tim Berners-Lee, James Hendler e Ora Lassila. Caracterizada por se tratar de uma extensão da web de 2000, a Web Semântica é apresentada como uma estrutura que possibilita a compreensão e o gerenciamento dos conteúdos armazenados na web, independentemente da forma que esses se apresentam (texto, som, imagem), a partir da valorização semântica desses conteúdos, advindos de fontes diversas, capazes de processar as informações. A Web Semântica (Berners-Lee et al., 2001), através de programas inteligentes, tem por objetivo realizar tarefas num grau mais sofisticado de inteligência, realizando inferências e deduções. À medida que elementos de um artefato computacional estejam inter-relacionados, como em uma base de dados ou em uma ontologia, um “agente de software” é capaz de realizar “inferências” automáticas, com base nessas relações, seguindo os “links” entre os nós de uma rede semântica.

Há pesquisas e desenvolvimento de programas de computador, comumente chamados de “agentes inteligentes”, que têm a possibilidade de fazerem a colheita (ou harvesting) de informações em outros computadores, agentes e dispositivos eletrônicos, para, assim, tomarem decisões baseadas em heurísticas embutidas (SOUZA, 2005). Agentes são programas que capturam o conteúdo de várias fontes na *web*, processam estas informações e fazem intercâmbio desses resultados com outros programas (DIAS; SANTOS, 2003). Sánchez (2007) define “agente inteligente” como um agente de software situado em um ambiente específico, capaz de mostrar um comportamento autônomo, reativo e proativo sobre este ambiente e preparado para

interagir com outros agentes, com a finalidade de satisfazer determinados objetivos, estabelecidos por um ente (software ou humano), representado pelo agente. Para que esse cenário seja viável, o conhecimento de diferentes domínios deve estar formalizado dentro de uma compreensão comum, padronizada, e representado, também, em formato legível por programas – nas chamadas ontologias –, para que possam ser utilizados por programas “agentes inteligentes”.

A proposta da Web Semântica *introduz* uma nova concepção de Sistemas de Organização do Conhecimento (SOC), diferente de seus predecessores destinados ao uso humano. Agora, os SOCs podem ser usados por “agentes inteligentes” para a busca de informações. Para alcançar tais características, torna-se necessária a explicitação das relações semânticas entre os termos, possibilitando a realização de inferências automáticas, dando origem a uma nova categoria de SOC, que passou a ser denominada “ontologia”. Em grego, o termo onto significa “Ser”, enquanto logia, “estudo” ou “conhecimento”.

Segundo Chauí (2003), ontologia significa o estudo dos entes, das coisas como verdadeiramente são, isto é, do “Ser”. Davies, Studer e Warren, (2006), definem ontologia como uma especificação explícita e formal de uma conceitualização de um domínio de interesse e realça dois pontos importantes: a conceitualização é formal e, portanto, permite o raciocínio por computador; e que uma ontologia é concebida para algum domínio de particular de interesse.

De acordo com Schiessl e Bräscher (2012), para a Ciência da Informação, ontologia, como explicação sistemática da própria existência, se estende ao domínio de sistemas de informação (SI), sendo possível descrever uma ontologia pela definição de um grupo de termos representativos de determinado domínio. A tarefa desse domínio é representar formalmente essa existência. Nesse contexto, ontologia é um modelo que reflete uma visão de mundo. Depreende-se das argumentações de Schiessl e Bräscher (2012) que o objetivo da ontologia é delimitar a quantidade de interpretações e caracterizar os significados básicos de categorias, utilizadas para descrever determinado domínio.

Outro princípio fundamental da Web Semântica é a criação e utilização de metadados semânticos. Os metadados, por um lado, podem descrever um documento ou parte de um documento. Por outro lado, eles podem descrever entidades dentro do documento (por exemplo, uma pessoa ou empresa). Em qualquer caso, o importante é que a semântica dos metadados informe sobre o conteúdo de um documento (por exemplo, seu tema, ou relação com outros documentos) ou sobre uma entidade dentro do documento (DAVIES, STUDER; WARREN, 2006).

Para Rodríguez Perojo e Ronda León (2005), de maneira formal, pode-se dizer que um metadado é um dado que se encarrega de manter um registro sobre o significado, contexto ou propósito de um objeto de informação, podendo descobrir, entender, extrair e administrar esse objeto. Em geral, esses registros são bem mais concisos ~~de~~ que os objetos que descrevem, de forma que podem estabelecer intercâmbio com facilidade.

Usando semântica, os sistemas podem compreender “se” palavras ou frases são equivalentes. Ao procurar por referências à palavra “Jaguar”, no contexto da indústria automobilística, o sistema pode ignorar as referências aos felinos. Usando semântica, pode-se melhorar a forma pela qual a informação é apresentada, na sua forma mais simples, ao invés de uma pesquisa em uma lista linear de resultados, e esses podem ser agrupados por significados. Assim, uma busca por “Jaguar” pode fornecer documentos agrupados, de acordo com o que eles são (carros, animais ou temas diferentes), todos juntos. No entanto, pode-se ir mais longe ~~de~~ que isso, usando da semântica, para mesclar informações de todos os documentos relevantes, e fazendo inferências, a partir do conhecimento existente, para criar novos conhecimentos.

Segundo Marcondes (2011), desde que os filósofos gregos desenvolveram a forma de raciocínio, conhecido como silogismo lógico, que o papel das relações para viabilizar a inferência é reconhecido. O autor faz menção ao conhecido exemplo do silogismo lógico “Todo homem é mortal; Sócrates é homem; portanto Sócrates é mortal”. O poder do raciocínio inferencial se sustenta, essencialmente, na relação estabelecida entre os três termos (“homem”, “Sócrates” e “mortal”) e não no seu conteúdo.

A partir da proposta da Web Semântica, a ideia de representar documentos em meio digital, através da estruturação de seu conteúdo, ganha ênfase, a partir do momento que propõe novas técnicas para a representação da informação e do conhecimento. Esse enfoque pretende enriquecer a estrutura da informação e agregar componentes semânticos, que podem ser processados de forma automática.

#### *2.2.2.1 A estrutura da Web Semântica*

Moreiro-Gonzalez (2011, p. 75) especifica a Web Semântica em camadas:

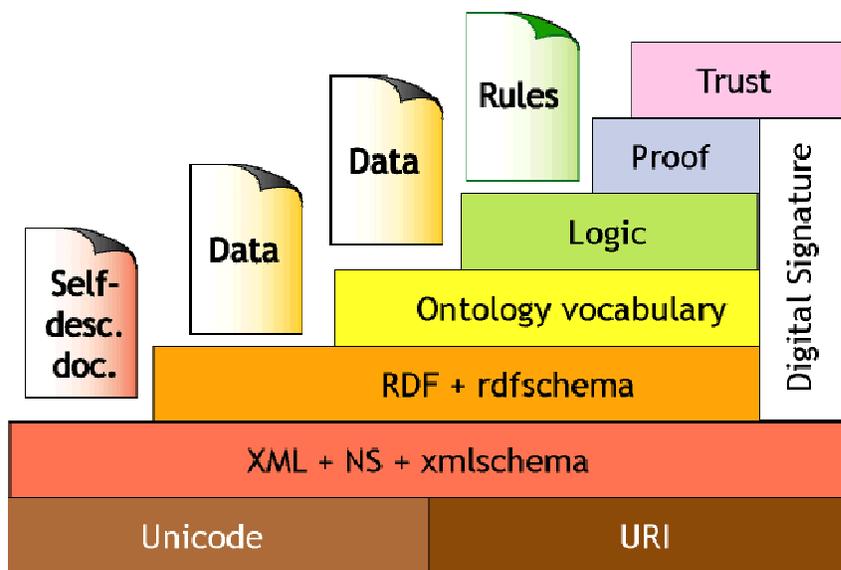
- sintática;
- semântica;
- ontológica;
- lógica;
- de prova e confiança.

O autor destaca que as três primeiras camadas – sintática, semântica e ontológica – dão as bases para poder representar a informação, de maneira semanticamente acessível. As outras duas definem os elementos, que permitem aos agentes de software a autenticação e a comprovação de confiabilidade dos diferentes elementos do modelo (recursos, agentes, inferências obtidas etc). No entanto, para que a camada ontológica possa inferir conhecimento, é necessário contextualizar a informação, em um determinado cenário.

Em geral, uma ontologia contém uma lista finita de termos e relações entre esses termos. Os termos denotam conceitos (classes de objetos) importantes do domínio, servindo como vocabulário comum (sintático e semântico), que favorece a comunicação e a interoperabilidade de recursos. Por isso, a arquitetura da Web Semântica visa estruturar conteúdos (documentos) de forma crescentemente mais complexa, através de mais e de diferenciadas relações (MARCONDES, 2011).

Segundo Berners-Lee, a arquitetura da Web Semântica<sup>6</sup> pode ser apresentada da seguinte forma:

Figura 1: Arquitetura da Web Semântica



Fonte: Tim Berners-Lee. *Semantic Web -XML2000. Architecture*

O W3C desempenha um papel fundamental no desenvolvimento e padronização das tecnologias semânticas e, desde a sua fundação, vem favorecendo a criação de inúmeras

<sup>6</sup> <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide11-0.html>

tecnologias e padrões, que deram origem à denominada arquitetura da Web Semântica (figura 1), na qual é constituída a partir de uma série de camadas, que representam as etapas e as tecnologias necessárias para sua concretização, de forma que cada camada posterior possibilite estender as funcionalidades presentes nas anteriores, favorecendo uma estrutura semanticamente compatível e escalonável (RAMALHO; OUCHI, 2011).

As camadas Unicode e URI determinam um conjunto de caracteres internacionais e fornecem recursos para a identificação de objetos na Web Semântica. A camada XML, denominada de namespace e schema, propõe a integração de definições da Web Semântica com outros padrões baseados em XML. A camada RDF e RDF Schema permite estabelecer afirmações sobre objetos com URIs, tal como definir vocabulários, que podem ser atribuídos por URIs.

A camada Ontologia tanto suporta o desenvolvimento de vocabulários como pode definir relações entre conceitos diferentes. Já a camada Digital Signature identifica as alterações em documentos. A camada da Logic permite a escrita de regras e a camada da Proof executa estas regras e examina, juntamente com a camada da Trust, os mecanismos que possibilitam às aplicações a confiabilidade (ou não) nas provas realizadas.

A estruturação de conteúdos, segundo os padrões da proposta da Web Semântica, de forma resumida, é assim descrita (RODRÍGUEZ PEROJO; RONDA LEÓN, 2005):

- XML: Surgiu em 1996, com o objetivo de possibilitar o intercâmbio de documentos estruturados por meio da *web*, uma vez que permitia o uso do hipertexto. Um dos aspectos mais importantes do XML é que é um conjunto de tecnologias, baseadas em padrões abertos, que formam módulos opcionais e que ampliam suas possibilidades. XML desloca o foco da exibição da info (HTML) para seu conteúdo (XML). Esse é o aspecto mais importante;
- XML Schema é uma linguagem para definir a estrutura dos documentos XML;
- RDF: em 1997, aparece RDF, uma linguagem para representar conhecimento na *WWW*, baseado em XML. RDF não é um formato de arquivo, mas um formalismo de dados, desenhado para apoiar a gestão de dados distribuídos em ambiente *web*, que permite realizar afirmações sobre recursos. Mediante RDF, é possível representar qualquer tipo de propriedade de um recurso. O modelo de dados de RDF é uma forma de sintaxe neutra para representar expressões RDF e se usa para avaliar a equivalência em significado. Essa definição de equivalência permite algumas variações sintáticas em expressões, sem alterar o significado;

- RDF Schema é um conjunto de informações relativas às classes de recursos, que serve para explicar relações hierárquicas, que se estabelecem entre si, ou para esclarecer o caráter obrigatório ou opcional das propriedades e outras restrições, como o número de ocorrências;
- OWL, finalmente, adiciona mais vocabulário, para descrever propriedades e classes. Entre outras melhoras na expressividade se encontram a possibilidade de definir relações entre classes, cardinalidade, igualdade, tipologias de propriedade mais complexas, caracterização de propriedades ou classes enumeradas.

XML fornece informações semânticas como um subproduto para definir a estrutura do documento, porque define uma estrutura em forma de árvore para um documento, de modo que suas folhas contêm a informação. Pode-se observar, então, que a estrutura e a semântica de um documento XML estão entrelaçadas, porque RDF fornece meios para agregar semântica a um documento, sem fazer referência à sua estrutura. Seu objetivo geral é definir um mecanismo para descrever os recursos, de modo a não criar dependência alguma sobre um domínio de aplicação particular, nem definir, a priori, a semântica de algum domínio de aplicação. A definição do mecanismo deve ser neutra, com respeito ao domínio, porém, isso deve se adaptar, para descrever informações sobre qualquer domínio (RODRÍGUEZ PEROJO; RONDA LEÓN, 2005). A partir de descrições precisas de recursos de informação codificada, utilizando RDF, um agente inteligente será capaz de recuperar informações relevantes e precisas, a partir de diferentes repositórios de informações acessíveis na *web*.

Os computadores não têm por que compreenderem a informação. Só precisam de metadados e ontologias para indexarem e classificarem os documentos. Graças à semântica, um software é capaz de processar seu conteúdo, estabelecer relações, combinar e realizar deduções lógicas, para resolver problemas automaticamente. Com a efetiva implantação de um ambiente digital, onde os recursos estão semanticamente especificados, a descoberta de conhecimento nesse ambiente tornar-se-á ou torna-se muito mais facilitada, uma vez que mecanismos automatizados poderão ser utilizados, de forma mais eficiente (VASCONCELOS; CARVALHO, 2004).

### ***2.2.3 A semântica e os sistemas de organização do conhecimento***

A semântica (derivada do grego *sêmainô*, que quer dizer “significar”) é definida, estritamente, *como o estudo do sentido das palavras* (GUIRAUD, 1972). Sentido, na definição do autor, quer dizer *significado* ou *emprego*. Esse campo de estudo envolve níveis diferenciados de tratamento dos problemas do significado, nos quais especificam-se o caráter e o vínculo disciplinar. Para Almeida (2011), significado é um elemento analisável pela estrutura contextual que o circunda. Já González (2011) afirma que pode-se chegar ainda a um conceito mais amplo sobre semântica, que distinga uma semântica referencial (relação com a realidade extralinguística) de uma semântica orientada ao conteúdo, ambas entendidas como relações inerentes ao sistema, que se dão entre o conteúdo dos signos: conteúdo global do discurso, seu tema.

Na intenção de melhorar métodos e técnicas de organização do acesso aos recursos de conhecimento digital no ambiente de rede, a tentativa deve ser feita para substituir sistemas de organização do conhecimento (SOC), semanticamente fracos, por sistemas com uma semântica mais forte. Isso pode ser alcançado, por meio da criação de novos modelos de SOC ou com a modernização e o desenvolvimento de SOC tradicional.

De acordo com Sosinska-Kalata (2014), a criação de SOC semanticamente fortes exige uma identificação precisa de estruturas conceituais de vários domínios do conhecimento, bem como uma análise precisa de contextos epistemológicos e culturais da criação de conhecimento e de aspectos pragmáticos da aplicação do conhecimento.

Sob essa perspectiva, a autora aborda o termo semantização, no qual é conceituado como o processo de aumento do grau de representação semântica de conteúdo disponível no ambiente de rede, em particular na Internet, com a aplicação de metadados que permitem o processamento por máquina dos recursos de conhecimento registrados digitalmente em um nível conceitual. Potencialmente, semantização deve assegurar um acesso mais preciso a recursos de conhecimento digitais, ultrapassando os métodos convencionais de organização de dados, usados para recuperar a informação.

Encontrar soluções para os problemas relacionados com a semantização de métodos e ferramentas de representação do conhecimento é o foco atual de um número considerável de projetos de investigação, dentro do campo da Organização do Conhecimento (OC). A afirmação acima mencionada foi verificada por Sosinska-Kalata (2014), com base na análise de temas de artigos publicados na revista *Knowledge Organization* no período de 2010 a 2013. Foi descoberto pela autora que mais de 60% dos 140 artigos publicados, durante esse período, discutiam amplamente aspectos semânticos, relacionados com a organização do conhecimento.

Trinta por cento, aproximadamente, dos trabalhos analisados pela autora citada eram pesquisas sobre o projeto e experimentação de novos métodos e técnicas para representações do conhecimento semanticamente fortes. Em particular, os mapas de tópicos e ontologias. Também, sobre o reforço do poder da semântica nos Sistemas de Organização do Conhecimento (SOC) tradicionais. Outra tendência importante nas pesquisas sobre novos modelos de SOC é que são requeridos no desenvolvimento da teoria da classificação facetada e novos sistemas de classificação, de forma que permitam a criação de representações multidimensionais do conteúdo do documento.

Pode-se notar, também, um interesse crescente em tecnologias semânticas, seguido pelo desenvolvimento de novos tipos de SOC, mais eficazes que os sistemas tradicionais, ambos ajustados às capacidades tecnológicas de ambiente de informação em rede e à representação multidimensional de conhecimento de vários domínios. Para Sosinska-Kalata (2014), não é de surpreender que as questões de semântica, a análise dos significados e seus métodos de representação ocupam tal posição exposta na pesquisa da OC.

#### ***2.2.4 Semântica e tesauros, taxionomias, ontologias***

Os novos formatos de representação surgem para responderem às necessidades de organização do conhecimento em ambiente digital, diante do crescimento da informação na *web*, e emergem novas perspectivas para o desenvolvimento de classificações, tesauros e ontologias. Essas linguagens colaboram com a descrição dos diferentes recursos de informação, fator que pode melhorar a recuperação da informação na *web*, em termos de efetividade, rapidez e facilidade de acesso.

Diante desse contexto Dziekaniak, Pacheco e Kern, (2011), afirmam que níveis de exigências sugerem que atenções específicas devam ser dadas aos mecanismos e instrumentos de apoio tradicionais, como os tesauros, utilizados para representar e organizar o conhecimento, no âmbito das unidades de informação, agregando-lhes, dentro de uma perspectiva mais moderna e atual, as regras essenciais de desenvolvimento, sob a forma de normas e padrões, comuns para a representação, a transferência, acesso e organização do conhecimento.

Não é o objetivo dessa pesquisa assumir que as tradicionais linguagens documentárias, como o caso dos tesauros, estão obsoletas, mas, sim, que há novos formatos, metodologias e possibilidades de modelagem de conhecimento, a partir do uso dessas linguagens. Até porque toda teoria, na qual sua construção está embasada, é amplamente apta para ser utilizada como contribuição.

Diferentemente das ontologias, que têm origem no contexto da era digital, os tesauros tiveram origem no contexto das bibliotecas tradicionais. Foram instrumentos criados, originalmente, para auxiliarem os bibliotecários na indexação. Posteriormente, principalmente quando inseridos no contexto digital, passaram a ser disponibilizados também aos usuários, para auxiliarem na busca por informações (WEISS, 2014).

Segundo Martins (2013), diante da Web Semântica, a evolução mais significativa dos tesauros é a da estruturação terminológica numa dimensão de representação conceitual, onde as equivalências se criam entre conceitos e não entre palavras.

Ontologias, taxonomias e tesauros, por serem dotados de técnicas de indexação automática, de normas e procedimentos terminológicos e de programas de interoperabilidade linguística, demonstram ser os instrumentos mais adaptáveis aos propósitos informacionais da Web Semântica (DIAS; COSTA, 2011).

De acordo com Rodríguez Perojo e Ronda León (2005), as ontologias agregam valor aos tesauros tradicionais, por meio de uma semântica mais profunda, desde um prisma conceitual, relacional e informático. Para o autor, de fato, uma maior profundidade semântica pode implicar níveis mais detalhados de hierarquia, ricas relações entre classes e conceitos, assim como a capacidade de formular regras de inferência. Contudo, também podem ser encontrados pontos de encontro: a crescente complementaridade entre a linguagem natural e as linguagens controladas e a diminuição das diferenças entre as linguagens documentárias. O autor ainda afirma que as ontologias são instrumentos claramente conectados com os “agentes inteligentes” em seu objetivo de filtrar informação e que, em muitos casos, podem derivar em tesauros.

Portanto, é de se esperar que, apesar da posição bem estabelecida de padrões de classificação tradicionais e tendências de simplificação na contemporânea organização do acesso aos recursos de conhecimento digitais, essas novas soluções serão colocadas em prática e abrirão o caminho para a normalização e semantização de ferramentas de organização do conhecimento (SZOSTAK, 2004 apud SOSINSKA-KALATA, 2014).

### ***2.2.5 Motor de busca semântico***

Um motor de busca semântico tem a capacidade de compreender as relações entre as palavras, frases ou partes do discurso dentro de uma pesquisa, permitindo, assim, entender o significado subjacente da frase inteira. Por exemplo, um motor de busca semântico seria capaz

de distinguir os diferentes significados entre as seguintes frases compostas pelos mesmos "termos":

*Vítimas de crimes juvenis.*

*Vítimas juvenis de crimes.*

No exemplo, as frases são constituídas pelas mesmas palavras, porém em ordem invertida. Em buscadores tradicionais, baseados em algoritmos de classificação, uma vez que as relações entre as partes das frases são desconsideradas, os motores apresentariam resultados idênticos, ou quase idênticos, apesar do sentido das duas frases ter sido completamente diferente.

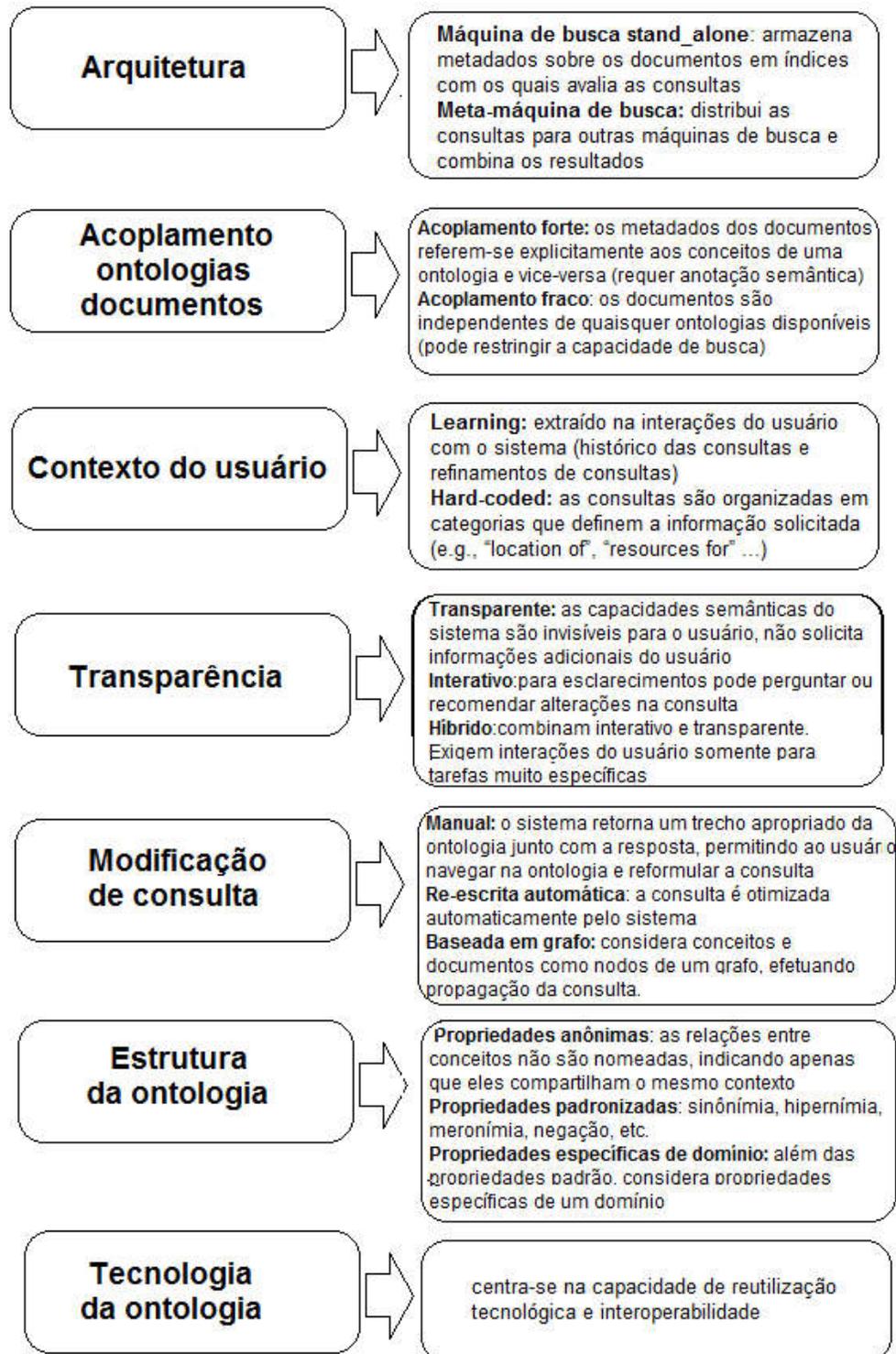
Um motor de busca semântico tenta fazer sentido aos resultados da pesquisa com base no contexto, identificando automaticamente os conceitos que estruturam os textos. Um segundo exemplo seria: ao procurar pelo termo "eleição", um motor de busca semântico pode recuperar documentos que contenham a palavra "voto", "campanha" e "votação", mesmo que a palavra "eleição" não seja encontrada no documento de origem.

Segundo Amanqui (2014), para aumentar a eficiência de ferramentas de recuperação de documentos, ontologias e mecanismos de inferência, são utilizados para explorar o conhecimento do domínio e compartilhar a mesma estrutura de informação entre pessoas e agentes de software (programas), no processo de recuperação de documentos. Essa abordagem é denominada "busca semântica". Mangold (2007) define busca semântica como um processo de recuperação de documentos, que aproveita o conhecimento de domínio do contexto semântico dos termos de consulta, aumentando sua precisão e revocação.

A busca semântica tem mostrado um potencial significativo na função de melhorar o desempenho da recuperação da informação. Comparados aos motores de busca tradicionais, que se concentram na frequência de aparecimento das palavras no texto, os motores de busca semântica são mais propensos a compreenderem os significados, escondidos por meio da adição de tags semânticas em textos, a fim de estruturarem e conceituarem os objetos dentro dos documentos.

Considerando uma série de publicações, que tratam das abordagens sobre a busca semântica, Mangold (2007) identifica e apresenta um esquema de categorização, com a finalidade de classificar diferentes abordagens para pesquisa semântica, em várias dimensões. É válido salientar que existem outros critérios para classificar motores de busca semântica, porém, enfatizamos os critérios selecionados por Mangold (2007), conforme a figura 2:

Figura 2: critérios da busca semântica



Fonte: elaborado de acordo com Mangold (2007)

Mangold (2007) considera que há um grande número de abordagens promissoras para a recuperação de documentos, de acordo com a busca semântica. Neste levantamento, o autor elencou as abordagens que exploram domínio de conhecimento para processar solicitações de pesquisa, apresentando utilizações que compreendem a navegação ontológica, a modificação de consultas – manual e automática – e modelagem de contexto do usuário. O objetivo é maximizar precisão e cobertura do resultado da busca e minimizar a interação do usuário.

### ***2.2.6 As barreiras linguísticas***

Tarefas que envolvem a gestão do conhecimento estão intimamente relacionadas com os processos linguísticos que ocorrem no cérebro e se materializam na forma de conhecimento, representado através da linguagem. Nesse sentido, Lima, Nunes e Vieira (2007) ressaltam que as tecnologias da informação e da língua são, cada vez mais, elementos-chave do ingresso na sociedade da informação e na sociedade do conhecimento, por sua natureza de viabilização, especialmente, na geração, no acesso a conteúdos e na constituição de redes de comunicação. Para as autoras, essa temática perpassa a interatividade e contempla aspectos do multilinguismo e do acesso à informação, que promovem a cidadania, o desenvolvimento tecnológico e a redução do abismo digital.

As revoluções tecnológicas que, no passado, envolveram a linguagem natural (o advento da escrita, a imprensa mecânica etc.), fizeram com que muitas línguas perdessem a sua relevância e algumas fossem extintas, à medida que os seus falantes deixavam de ser beneficiados por esses avanços tecnológicos. Para uma língua prosperar na era digital, é necessário que esteja devidamente equipada, do ponto de vista tecnológico, de forma a poder ser usada para beneficiar a todas as pessoas, serviços e bens que irão ficando disponíveis, através da sociedade da informação (BRANCO et al., 2012). De acordo com os autores, a tecnologia da linguagem é o fator que irá desencadear em uma nova revolução tecnológica para a linguagem natural. Apenas a tecnologia da linguagem, desenvolvida e adaptada especificamente para a uma dada língua, permitirá que essa sobreviva na era digital e que os seus falantes e a sua cultura tenham uma cidadania plena, assegurada na sociedade da informação.

É notório que a informação e o conhecimento são fatores determinantes na criação de riqueza, transformação social e desenvolvimento humano. Com isso, as barreiras geográficas foram transpostas mais facilmente. Entretanto, a barreira linguística é um ponto relevante a ser discutido. Essa barreira imposta pela língua tem se tornado um ponto crítico na transferência de informações e, principalmente, na análise e representação de conteúdos informacionais (HUDON, 1997). Para tanto, Branco et al. (2012) afirmam que muitos idiomas não estão ainda equipados com a tecnologia básica para a análise de texto, nem com os recursos linguísticos,

essenciais para o desenvolvimento dessa tecnologia. Desse modo, é preciso realizar um esforço, em grande escala, para que seja alcançado o objetivo ambicioso de se assegurar tecnologia da linguagem de alta qualidade, para todas as línguas.

Hudon (1997) assinala que o tratamento igualitário das línguas é uma forma de respeito, em se tratando do valor ético, fazendo com que as representações dos conceitos sejam identificadas e aceitas por usuários de diferentes línguas, refletindo, assim, suas crenças. Portanto, infere-se que a língua, além de mediar a transmissão de ideias, também é a expressão cultural de uma nação.

Lima, Nunes e Vieira (2007) ressaltam que a maior densidade de produtos e resultados se concentra na língua inglesa, e estes resultados não são, na área do processamento da língua, diretamente transportáveis para outras línguas. São necessários estudos fundamentais que envolvem a construção de recursos de base, como léxicos, ontologias, grandes coleções de textos ou de amostras de fala processadas e anotadas – os *corpora*, gramáticas e analisadores, que somente se tornam disponíveis, após um trabalho árduo de estudo, seja este estatístico-computacional ou linguístico-cognitivo, do comportamento da linguagem e do seu uso.

No caso do Português, o apoio da tecnologia da linguagem tem melhorado gradualmente, mas é necessário garantir o incremento estratégico do esforço aplicado, para que seja alcançado um patamar de desenvolvimento sustentado. São necessárias medidas imediatas, para que possam ser obtidos progressos importantes para o Português e assegurar a sua posição como língua de comunicação internacional, com projeção global. Há uma boa comunidade de centros de investigação, tanto em Portugal quanto no Brasil, que cooperam ativamente entre si e que têm capacidade instalada, para fazerem avançar a tecnologia da linguagem para a língua portuguesa (BRANCO et al., 2012).

Outro desafio é encontrado, quando os programas precisam resolver situações de ambiguidade, em qualquer nível linguístico (léxico, semântico ou sintático). Entende-se ambiguidade como uma expressão da língua (palavra ou frase) que possui vários significados distintos, podendo, conseqüentemente, ser compreendida de diferentes maneiras por um receptor (BRÄSCHER, 2002). Qualquer idioma utilizado pelos humanos é composto por ambiguidades e nuances de interpretação, que são extremamente dependentes de diversos fatores, além da questão puramente gramatical. A interpretação de frases, muitas vezes, depende do contexto no qual ela foi utilizada, da avaliação de ambiguidades, de regras gramaticais, de questões culturais e de conceitos, muitas vezes, abstratos. Portanto, possuir conhecimento linguístico e domínio das técnicas do processamento sobre a língua portuguesa é ponto fundamental para prosseguir com a construção de ferramentas a serem disponibilizadas, para a busca do conhecimento.

## 2.3 Processamento de Linguagem Natural

O poder de processamento de dados permitiu um aumento na habilidade de guardar, coletar e manipular informações. Com a intenção de capacitar as máquinas para o reconhecimento e interpretação de informações, nos deparamos com uma questão filosófica secular: o que é o conhecimento e como adquiri-lo? A área de Inteligência Artificial (IA) tem devotado esforços para desenvolver sistemas capazes de “copiar” o raciocínio humano, isto é, de adquirir novo conhecimento a partir de bases de conhecimento (SCHIESSL; BRÄSCHER, 2012).

A parte da ciência que estuda e desenvolve recursos para a recuperação da informação está em constante atualização e criação. Atualmente, se discute a Linguística Computacional, que pode ser entendida, segundo Vieira e Lima (2001), como a área do conhecimento que explora as relações entre linguística e informática, tornando possível a construção de sistemas, com capacidade de reconhecer e produzir informação, apresentada em linguagem natural. Envolve várias áreas de pesquisa tradicionais, como: Linguística Teórica e Aplicada, Sintaxe, Semântica, Fonética e Fonologia, Pragmática, Análise do Discurso, entre outros. Isso tudo para tentar processar (compreender e produzir) as línguas naturais, dominar o conhecimento linguístico humano e passá-lo para as máquinas – os agentes inteligentes – que possam localizar, filtrar e devolver as consultas, de acordo com as necessidades de cada usuário (FACHIN, 2010).

Segundo Mendonça (2000), a contribuição da linguística é fornecer subsídios que auxiliem os outros campos do conhecimento na construção de conceitos e parâmetros que definirão a sua própria estrutura científica. Já a contribuição da CI seria a função social via comunicação, estudando o tratamento e a transmissão da informação mediante procedimentos metodológicos de caráter científico. Diante disso, pode-se afirmar que as técnicas de PLN são bem utilizadas na área da Recuperação da Informação, tanto para facilitar a descrição do conteúdo dos documentos como para representar a consulta formulada pelo usuário.

Para que um computador realize algumas tarefas humanas, é primordial que ele entenda e utilize a linguagem natural. Porém, é difícil fazer com que o computador entenda informações contextuais presentes em todas as situações. Isso se deve ao fato de que a compreensão da linguagem natural é difícil, devido à sua complexidade e ambiguidade. Essa compreensão exige não somente o conhecimento linguístico de uma determinada língua, mas, também, o conhecimento de mundo relacionado ao assunto utilizado, ou seja, o domínio.

A tarefa de processar uma linguagem natural permite que os seres humanos se comuniquem com os computadores, da forma mais natural possível, utilizando a linguagem na qual

mais estão habituados. Uma das áreas de pesquisa da Linguística Computacional é a linguística baseada em corpus, que se utiliza do agrupamento de documentos de texto que após se tornarem legíveis para a máquina, por sua vez, podem rapidamente ser pesquisados, a fim de obter informações sobre seu conteúdo. Trabalho este que envolve tanto um processamento morfossintático e semântico, como também pragmático, etapas fortemente apoiadas pela Inteligência Artificial (MARTINS, 2014).

A Inteligência Artificial, uma área de pesquisa da Ciência da Computação, vem sendo utilizada pela CI como fonte de ferramenta, ligada ao desenvolvimento e ao uso de algoritmos, que permitem a utilização mais aprimorada dos recursos computacionais, a fim de ampliar a eficácia dos mecanismos de recuperação da informação. Um dos vários usos da Inteligência Artificial ocorre nos sistemas de recuperação da informação, com o objetivo principal de que, a partir de uma questão formulada pelo usuário, o sistema seja capaz de retornar os resultados compatíveis com a questão apresentada.

Nesse sentido, devido à complexidade de se criar programas computacionais capazes de “compreender” as línguas, a maioria dos pesquisadores passou a focalizar o desenvolvimento de programas que realizam tarefas bastante específicas, como correção ortográfica, análise sintática, tradução entre outras. A área na qual esses programas têm sido desenvolvidos recebeu o nome de Processamento Automático dos Línguas Naturais - PLN (DI FELIPPO, 2004).

PLN visa capacitar um computador para lidar com a língua, permitindo utilização/adaptação da linguagem humana, ou seja, a criação de programas que possam, ainda que de modo rudimentar, emular o conhecimento e o desempenho linguístico dos humanos (DIAS-DASILVA et al., 2007).

Como o texto armazenado em meio digital é o objeto básico tratado pelo PLN, a linguística e a computação se tornaram fundamentais, para que o computador possa se tornar instrumento capaz de identificar minimamente os fenômenos da língua, para criar taxonomias que sejam úteis na interação entre computadores e seres humanos (MARTINS, 2014). Fellipo e Dias-da-Silva (2009) consideram que as pesquisas, derivadas da união dessas duas áreas, têm propiciado não só desenvolvimento de tecnologias ou recursos aplicáveis a várias atividades, mas, também, o próprio desenvolvimento da Linguística e da Ciência da Computação, duas das várias disciplinas matrizes do PLN. Para os autores, o estudo do PLN pode ser concebido como um tipo de “engenharia do conhecimento linguístico”, reflexo do fato de que a delimitação do conhecimento, necessário para a construção de sistemas de PLN, exige a organização e a representação de uma variedade de dados complexos necessários à simulação da competência e do

desempenho linguísticos e se beneficiar da estratégia desenvolvida para esse campo (FELLIPO; DIAS-DA SILVA, 2009).

Projetar um sistema de PLN, ou seja, um sistema que simule parcelas da competência e do desempenho linguístico humano, requer a especificação de vários conhecimentos e habilidades que os falantes (especialistas nesse domínio) possuem. Mais precisamente, Fellipo e Dias-da-Silva (2009) ressaltam que, para simular uma língua natural de modo satisfatório, um sistema de PLN deve conter vários sistemas de conhecimento e realizar uma série de atividades cognitivas. Essas atividades ficam evidentes nas palavras de (DIAS-DA-SILVA et al., 2007):

- (i) possuir um “modelo simples de sua própria mentalidade”; (ii) possuir um “modelo detalhado do domínio específico do discurso”; (iii) possuir um modelo que represente “informações morfológicas, sintáticas, semânticas, (iv) contextuais e do conhecimento de mundo físico”; (v) “compreender o assunto que está em discussão”; (vi) “lembrar, discutir, executar seus planos e ações”; (vii) participar de um diálogo e responder, com ações e frases, às frases digitadas pelo usuário; (viii) solicitar esclarecimentos quando seus programas heurísticos não conseguirem; (ix) compreender uma frase.

Um sistema de PLN é concebido como um tipo de sistema automático de conhecimentos, cujas especialidades, entre outras, incluem: fazer revisões ortográficas de textos, fazer análises sintáticas, traduzir frases ou textos, fazer perguntas e respostas e auxiliar os pesquisadores na própria construção de modelos linguísticos. Para Martins (2014), os sistemas baseados em conhecimento podem ser vistos como conectores semânticos, recebendo informações de diversas origens e sendo capazes de analisá-las, interpretá-las, identificando a sua relevância e estando aptos a direcionar soluções de acordo com interesses dinâmicos.

Segundo Lima, Nunes e Vieira (2007), o enriquecimento do processo do tratamento linguístico da informação pode-se dar pela:

consideração de informação linguística e contextual, em ordem crescente de complexidade, nos níveis morfológico (a formação das palavras da consulta e do texto pode guiar, de alguma forma, a busca), morfossintático (as classes das palavras – nome, verbo, etc. – são indicadores de relevância), sintático (a inter-relação entre os elementos de um texto – nome-adjetivo, sujeito-verbo, verbo-objeto, etc. adiciona informação importante), semântico (em nível lexical – o sentido daquela ocorrência – ou em nível sentencial – o significado de uma sentença, ambos contribuem para o sucesso de uma busca) e pragmático-discursivo.

A linguagem é um processo comunicativo, onde o emissor e receptor processam determinada informação, em função de um conhecimento linguístico e um conhecimento do mundo (pragmático) compartilhado. Para um sistema computacional interpretar uma sentença em lin-

guagem natural, é necessário manter informações morfológicas, sintáticas e semânticas, armazenadas em um dicionário, juntamente com as palavras que o sistema compreende. As etapas do PNL são apresentadas no quadro abaixo:

**Quadro 3:** Etapas do Processamento de Linguagem Natural

<b>Análise morfológica</b>	O analisador morfológico identifica palavras ou expressões isoladas em uma sentença, sendo este processo auxiliado por delimitadores (pontuação e espaços em branco). As palavras identificadas são classificadas, de acordo com seu tipo de uso ou, em linguagem natural, categoria gramatical. Neste contexto, uma instância de uma palavra em uma sentença gramaticalmente válida pode ser substituída por outra do mesmo tipo, configurando uma sentença ainda válida (exemplo: substantivos, pronomes, verbos, etc.). Dentro de um mesmo tipo de palavra, existem grupos de regras que caracterizam o comportamento de um subconjunto de vocábulos da linguagem (exemplo: formação do plural de substantivos terminados em "ão", flexões dos verbos regulares terminados em "ar" etc.). Assim, a morfologia trata as palavras quanto à sua estrutura, forma, flexão e classificação, no que se refere a cada um dos tipos de palavras.
<b>Análise sintática</b>	Através da gramática da linguagem a ser analisada e das informações do analisador morfológico, o analisador sintático procura construir árvores de derivação para cada sentença, mostrando como as palavras estão relacionadas entre si. Durante a construção da árvore de derivação, é verificada a adequação das sequências de palavras às regras de construção impostas pela linguagem, na composição de frases, períodos ou orações. Dentre estas regras, pode-se citar a concordância e a regência nominal e/ou verbal, bem como o posicionamento de termos na frase. Um termo corresponde a um elemento de informação (palavra ou expressão) e é tratado como unidade funcional da oração, participando da estrutura como um de seus constituintes, denominados sintagmas.
<b>Análise semântica</b>	Significados compõem as estruturas criadas pelo analisador sintático. O analisador semântico examina o sentido das estruturas das palavras que foram reagrupadas pelo analisador sintático, uma vez que o analisador morfológico permitiu identificar estas palavras individualmente. Pode-se dizer que a semântica se desdobra em semântica léxica e em semântica gramatical. A semântica léxica busca uma representação conceitual para descrever o sentido, entretanto, para construir esta representação, pode ser feita a decomposição semântica das unidades léxicas (em primitivas ou em traços

	semânticos), ou podem ser utilizadas redes semânticas. Esta última forma de representação foi originada da psicologia e leva em conta a forma com a qual os seres humanos categorizam e memorizam conceitos.
<b>Pragmática</b>	As palavras podem se associar através de dois tipos de relações: paradigmáticas e sintagmáticas. As relações paradigmáticas associam palavras através do significado. As relações sintagmáticas conectam palavras que são frequentemente encontradas no mesmo discurso.

**Fonte:** Oliveira Neto, Tonin e Pietrich (2010)

A análise morfológica e sintática inclui o estudo de como as palavras se associam para formarem estruturas, em nível de sentença. Na morfológica, cada palavra é etiquetada de acordo com a sua categoria gramatical. A semântica está relacionada com o significado, não só de cada palavra, mas também do conjunto decorrente dessas palavras. A compreensão da relação entre as palavras é tão importante quanto a compreensão das próprias palavras.

Enfoques formais para a semântica gramatical tentam descrever o sentido de uma frase mediante a tradução de sua estrutura sintática para uma fórmula lógico-semântica. Como não existe uma correspondência imediata e biunívoca entre sintaxe e semântica, uma mesma estrutura sintática pode dar origem a diferentes representações semânticas (OLIVEIRA NETO; TONIN; PIETRICH, 2010). Os termos são considerados unidades que representam o conhecimento em um domínio específico, com a função de descreverem conceitos para que possam ser transmitidos, são uma parte fundamental da estrutura de texto que representa o conhecimento.

A análise sintática de uma oração em português deve levar em conta os seguintes sintagmas: termos essenciais (sujeito e predicado), termos integrantes (complementos verbal e nominal) e termos acessórios (adjunto adverbial, adjunto adnominal e aposto). A análise do período, por sua vez, deve considerar o tipo de período (simples ou composto), sua composição (por subordinação, por coordenação) e a classificação das orações (absoluta, principal, coordenada ou subordinada).

A semântica e a sintaxe têm papéis importantes na recuperação da informação, na medida em que permitem ao software identificar a estrutura lexical das frases e o significado dos termos que representam o conteúdo do documento. Na busca semântica, a geração de resultados relevantes envolve, por exemplo, compreensão da intencionalidade do pesquisador e o contexto do termo pesquisado.

A análise semântica profunda de documentos possibilita a extração de informação de domínio especializado, em uma área rica em conhecimento, expresso em língua natural (LIMA; NUNES; VIEIRA, 2007). Ao fazerem uso da linguagem natural, as ferramentas de busca precisam de conhecimento sobre o significado das expressões que são tratadas e das relações que se estabelecem entre elas. Essas ferramentas devem, ainda, ser capazes de tratar determinados fenômenos linguísticos que afetam a qualidade da recuperação (BRÄSCHER, 2002).

Segundo Oliveira Neto, Tonin e Pietrich (2010), à medida que se avança no processamento da linguagem natural, é necessário fazer uma interpretação do todo e cessar a análise do significado de suas partes, como ocorre na análise das informações morfológicas (léxicas), sintáticas (regras gramaticais) e semânticas (significados). Em nível pragmático, baseia-se na relação da linguagem com o contexto na qual é utilizada. Em muitos casos, não se pode realizar uma interpretação literal e automatizada dos termos utilizados. Em determinadas circunstâncias, o sentido das palavras que formam uma frase tem que ser interpretado num nível superior, recorrendo ao contexto em que a frase é formulada.

Para Borges (2009), toda língua tem seu próprio recorte e sua própria semântica e essa língua pode ser repleta de regionalismos, metáforas, gírias, linguagem figurada, denotação e conotação. Tudo aquilo que está presente na vida das pessoas possui um nome, que é parte do léxico. A estrutura lexical compreende o conjunto de vocábulos de uma língua e abrange o conhecimento linguístico, partilhado pela sociedade na qual é falada, possuindo valor diferente de língua para língua.

Sendo a língua natural o meio mais usado para a transmissão de conhecimento, é imprescindível deduzir que o domínio da tecnologia para o processamento da língua é também relevante nesse cenário. Porém, a construção dessas tecnologias não foi o foco dessa pesquisa, mas sim buscou-se compreender questões relativas a processos, métodos e recursos necessários à construção dos sistemas de PLN.

A CI está focalizada na investigação de caráter mais reflexivo, desenhando todo um campo teórico-conceitual, em busca de embasamentos e entendimentos teóricos e metodológicos, relativos às aplicações passadas e futuras. A Ciência da Computação, responsável pela maioria dos estudos de PNL, não deixa de lado as reflexões teóricas, mas prioriza a construção dos aparatos informáticos, concentra suas investigações no desenvolvimento e aplicação dos programas de PNL.

Levando em consideração a existência de tecnologias inovadoras, no tocante ao processamento de linguagens – não surgidas, porém, no âmbito da CI –, tais tecnologias precisam ser compreendidas pela área de CI, para não deixarem de atender às prerrogativas para

as quais foram criadas, especificadas por Mendonça (2000): investigar as propriedades e comportamento da informação, as forças governantes dos fluxos e os meios de processar a informação, tendo como objetivo a sua organização, armazenamento, recuperação e disseminação.

## **2.4 Tecnologias de Mineração**

Esta seção trata das tecnologias de mineração, ressaltando a mineração de texto e a extração automática do conhecimento em documentos textuais.

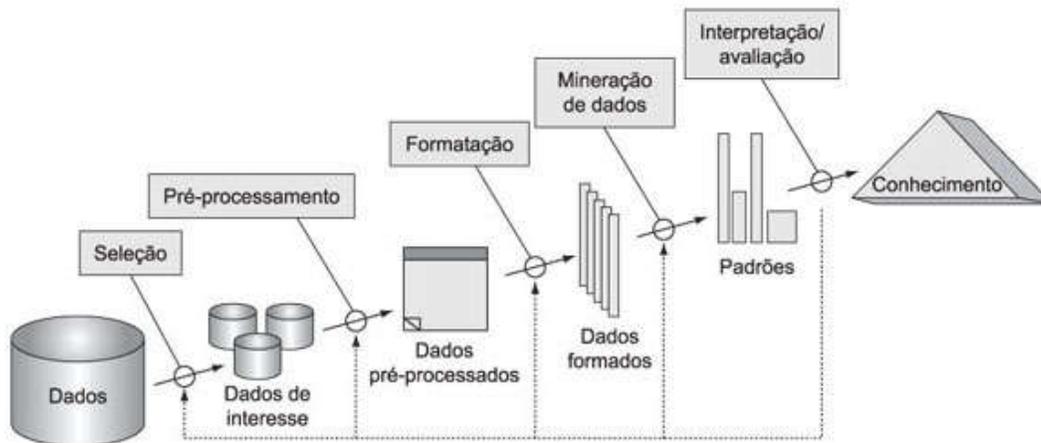
### **2.4.1 Mineração de Dados**

Considerando que, em geral, os acervos de documentos digitais contenham uma grande quantidade de dados, tornou-se necessário o desenvolvimento de processos de análise automática, como a Extração de Conhecimento de Bases de Dados (mineração de dados). Para Rezende (2003), tal processo tem o objetivo de encontrar conhecimento a partir de um conjunto de dados, para ser utilizado em um processo decisório. Segundo a autora, um requisito de suma importância é esse conhecimento descoberto ser compreensível a humanos, além de útil e interessante para os usuários finais do processo, geralmente, tomadores de decisão.

Na década de 1960, os estatísticos utilizavam termos como "Data Fishing" (pesca de dados) ou "Data Dredging" (dados de dragagem), para se referirem ao que eles consideraram uma prática pouco eficaz de análise de dados, sem nenhum tipo de parâmetro. O termo "Mineração de Dados" apareceu por volta de 1990, nas comunidades de banco de dados. Outros termos, tais como *Arqueologia de dados*, *Colheita de Informações*, *Pesquisa de Informações e Extração de Conhecimento*, também eram utilizados. Em 1989, Gregory Piatetsky-Shapiro criou o termo "Knowledge Database Discover – KDD (Descoberta de conhecimento em bancos de dados)", para o primeiro workshop sobre o mesmo tema, e esse termo se tornou mais popular na comunidade científica, especialmente em Inteligência Artificial (PIATETSKY-SHAPIRO; PARKER, 2011).

O KDD, apresentado na figura 3, é utilizado como ferramenta de análise em muitas áreas que produzem grande volume de dados, que podem ser ordenados e processados conjuntamente, e consiste em alguns passos, que vão desde a forma pela qual os dados são acessados até a interpretação e a visualização dos resultados. Os passos podem variar, dependendo da metodologia utilizada, porém, na maioria das vezes, um deles é a mineração de dados.

Figura 3: Processo de KDD



Fonte: Elaborado, de acordo com FAYYAD, PIATESTKY-SHAPIRO E SMYH (1996)

Segundo o dicionário Aurélio, o termo “mineração” significa: *trabalho ou exploração de minas; depuração do minério extraído*. Diante dessa definição, percebe-se que a ideia de utilizar o termo mineração faz sentido, já que o objetivo da mineração de dados é explorar e extrair informações, a partir de um conjunto de dados. A mineração de dados, em princípio, aplica-se a qualquer tipo de informação, sendo as técnicas de mineração diferentes para cada uma delas.

O campo da Mineração de Dados tem grande potencial de uso para a automação, na manipulação de grandes quantidades de dados. O desenvolvimento de algoritmos mais avançados possibilita a ampliação do campo de busca das informações que são relevantes para os gestores de tecnologia.

Na atual sociedade da informação, onde se multiplica a quantidade de dados armazenados, a Mineração de Dados é uma ferramenta fundamental para analisar e exportar, de forma eficaz, informações. Uma vez que o usuário não tenha condições para imaginar todas as possíveis relações e associações, existentes em um grande volume de dados, faz-se necessária a utilização de técnicas de análise dirigidas por computador, que possibilitem a extração automática de novos conhecimentos, a partir de um grande repositório de dados.

#### 2.4.2 Mineração de Texto

Para tal feito, técnicas específicas para mineração de textos têm sido desenvolvidas para processarem uma parte importante da informação disponível, que pode ser encontrada na forma de dados não-estruturados. “A mineração de texto agrega técnicas de visualização de dados, permite, todavia, descobrir conceitos-chave e grupos similares de documentos, sem que haja necessidade prévia de leitura integral dos documentos” (SILVA, 2002, p. 39).

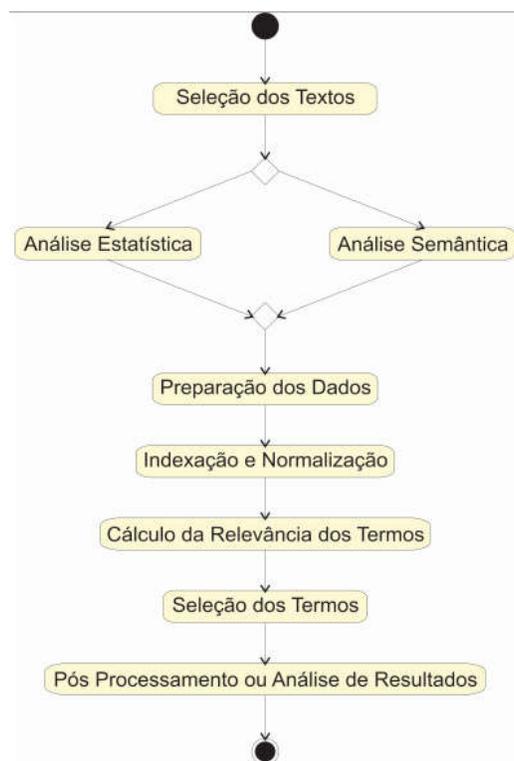
Na mineração de dados, as informações estão pré-organizadas e o processo de mineração precisa apenas tratá-las. Quando se refere à mineração de texto, tais dados se encontram escritos em língua natural, na forma de textos livres, e geram maiores desafios para que se obtenha deles entendimento e extração de conhecimento. Por isso, elas diferem.

A mineração de texto utiliza métodos de busca, baseados em análises gramaticais e léxicas ou, ainda, em técnicas de *clustering* (agrupamento). Além disso, “a mineração de texto agrega técnicas de visualização de dados e, permite, todavia, descobrir conceitos-chave e grupos similares de documentos, sem que haja necessidade prévia de leitura integral dos documentos”. (SILVA, 2002, p. 39).

Essas técnicas permitem mostrar conceitos-chave e relações entre palavras e ideias. Partindo dessa visualização, é possível detalhar dados ou “trilhar” caminhos para outros documentos. Silva (2002), em resumo, afirma que os seguintes pontos abrangem a mineração de texto:

- i. concentra algoritmos inteligentes e, em alguns casos, análise léxica;
- ii. processa documentos eliminando a análise “manual” direta. Categoriza, classifica ou constrói árvores de tópicos e índices de documentos;
- iii. provê identificação automática e indexação de conceitos entre os textos;
- iv. apresenta, por meio de técnicas de visualização, o escopo global dos dados. Permite detalhamento quanto ao grau de relevância;
- v. permite aos usuários fazerem associações, correlacionamentos e âncoras entre os documentos, para posterior análise.

Morais e Ambrósio (2007) definem, de forma geral, as etapas do processo de mineração de textos: seleção de documentos, definição do tipo de abordagem dos dados (análise semântica ou estatística), preparação dos dados, indexação e normalização, cálculo da relevância dos termos, seleção dos termos e pós-processamento (análise de resultados). Estas etapas podem ser visualizadas através do diagrama de atividades, representado na figura 4.

**Figura 4:** Etapas do processo de mineração de texto

**Fonte:** Elaborado, de acordo com (MORAIS; AMBRÓSIO, 2007).

Com relação aos tipos de abordagem dos dados textuais (análise semântica ou estatística), Ebecken, Lopes e Costa (2003) ressaltam que esses tipos de abordagem podem ser utilizados sozinhos ou em conjunto. O quadro 4 apresenta um resumo das características de cada uma das análises.

**Quadro 4:** Tipos de abordagem dos dados

<b>Análise Semântica</b>	<b>Análise estatística</b>
<ul style="list-style-type: none"> <li>- Emprega técnicas que avaliam a sequência dos termos no contexto da frase;</li> <li>- Utiliza fundamentos e técnicas baseadas no processamento de linguagem natural;</li> <li>- Permite a melhoria em qualidade da mineração de texto, quando incrementado de um processamento linguístico mais complexo;</li> </ul>	<ul style="list-style-type: none"> <li>- A importância dos termos é dada, basicamente, pelo número de vezes que eles aparecem nos textos;</li> <li>- Pode ser conduzida, independentemente do idioma;</li> </ul>

**Fonte:** Elaborado, com base em (EBECKEN; LOPES; COSTA, 2003)

Segundo Ebecken, Lopes e Costa (2003), no entendimento da linguagem natural, os seguintes tipos de conhecimentos são relevantes:

- **conhecimento morfológico:** conhecimento da estrutura, da forma e das inflexões das palavras;
- **conhecimento sintático:** conhecimento estrutural das listas de palavras e como as palavras podem ser combinadas para produzirem sentenças;
- **conhecimento semântico:** o que as palavras significam, independentemente do contexto, e como significados mais complexos são formados pela combinação de palavras;
- **conhecimento pragmático:** o conhecimento do uso da língua, em diferentes contextos, e como os significados e a interpretação são afetadas pelo contexto;
- **conhecimento do discurso:** como as sentenças imediatamente precedentes afetam a interpretação da próxima sentença;
- **conhecimento do mundo:** conhecimento geral do domínio ou o mundo com que a comunicação da linguagem natural se relaciona.

#### 2.4.3 Metodologia Tech Mining ou mineração de tecnologia

*Tech Mining* (Mineração de Tecnologia, em tradução literal) é uma expressão adotada por Alan Porter e Scott Cunningham, em sua obra “*Tech mining. Exploiting New Technologies for Competitive Advantage (2005)*”. Consiste em explorar a informação sobre tecnologias emergentes, principalmente, procedente de bases de dados, com a intenção de detectar tendências e oportunidades. Para Marcondes (2011), pode-se afirmar, estendendo o conceito apresentado por Porter e Cunningham, que *Tech Mining* é o próprio processo KDD, porém, utilizado, exclusivamente, para a extração de informações úteis para a gestão de inovação, a partir de bases de dados científico-tecnológicas.

*Tech Mining* é uma abordagem sistemática. Inicia-se com questões-chave relacionadas com a gestão da tecnologia. Em seguida, aplica-se a mineração à informação externa, sobre ciência e tecnologia, e a fontes de informação, sobre o ambiente de negócios. Com isso, geramos indicadores de inovação que podem responder às perguntas mais importantes (PORTER; CUNNINGHAM, 2005). O quadro 5 mostra vários tipos de análises tecnológicas, que podem ser realizadas utilizando técnicas de *Tech Mining*:

**Quadro 5.** Possíveis análises tecnológicas, através do uso de *Tech Mining*

A	Vigilância tecnológica	Cataloga, caracteriza e interpreta as atividades do desenvolvimento tecnológico
B	Inteligência tecnológica competitiva	Encontrar do ambiente externo “Quem está fazendo o que?”
C	Previsão tecnológica	Antecipar os possíveis desenvolvimentos futuros em tecnologias particulares
D	Mapeamento tecnológico	Seguir os passos da evolução, dentro das tecnologias relacionadas e famílias de produtos
E	Evolução tecnológica	Antecipar possíveis mal entendidos, indiretas e intempestivas consequências de uma mudança tecnológica em particular
F	Gestão dos processos tecnológicos	Fornecer informações sobre tecnologias aos que tomam decisões

**Fonte:** Elaborado, com base em Porter e Cunningham (2005).

São consideradas áreas mais significativas, para uma análise de *Tech Mining*:

- inteligência: a informação externa é rastreada e interpretada, para servir a um propósito particular da organização. É vital para A e B e contribui para C a F;
- investigações futuras: antecipam-se desenvolvimentos futuros, com ênfase no desenvolvimento tecnológico. É vital em C, D, E e F;
- com relação aos fatores contextuais socioeconômicos, enfatizando quais são as influências e seu efeito sobre as mudanças na tecnologia. É vital em E e F e importante em B e G;
- análise de oportunidades: interpretação das mudanças tecnológicas relacionadas com as ameaças e oportunidades para a organização. É vital em A, D e F;
- considerações de processo: envolve fornecedores em ações determinadas, neste caso em particular sobre a tecnologia, é especialmente direcionado para E.

Atividades de *Tech Mining* variam consideravelmente e podem ser adaptadas, de acordo com as necessidades da organização. O foco é geralmente na tecnologia, mas de uma forma contextualmente rica. Muitos usuários de *Tech Mining* querem que os aspectos contextuais sejam totalmente integrados com o tecnológico. Por exemplo, eles podem exigir que a informação de custos, projeções de mercado, e planos de transição sejam combinados para informar as decisões sobre escolhas tecnológicas. Para (PORTER; CUNNINGHAM, 2005).

*Tech Mining* é um processo de resolução de problemas, ele se baseia em um conjunto estruturado de fases (etapas) que orientam o usuário através do processo. Porter e Cunningham (2005) se basearam no processo de tomada de decisão de três etapas (quadro 5) de Simon

(1960), no qual tem-se aplicado em diversos domínios, incluindo projetos dos sistemas de informação. Os passos são: (1) inteligência, (2) projeto (que inclui a análise) e (3) escolha.

A fase de *inteligência* envolve o planejamento e a coleta dos dados a serem minerados, revisados, analisados e descritos.

A fase de *projeto e análise* envolve conhecimento derivado, a partir dos dados coletados, para resolver problemas específicos de inovação ou gestão de tecnologia. Muitas vezes, misturam-se análises dedutivas e indutivas. Analistas podem testar hipóteses específicas sobre os dados através de modelos analíticos, capazes de (por exemplo) descreverem atividades de ciência e tecnologia e previrem desenvolvimentos futuros.

A fase de *escolha* envolve indicação de opções via *Tech Mining*, bem como a seleção das melhores oportunidades de inovação para organização. A chave para este esforço é a criação de métricas e medidas de desempenho específicas, para avaliar as opções disponíveis, usando os critérios que são mais importantes para os usuários-alvo.

**Quadro 6** – fases de decisão do processo Tech Mining

Inteligência	1. Identificação de questões
	2. Seleção de fontes de Informação
	3. Procura, refinamento e recuperação de dados
Análise e Projeto	4. Limpeza dos dados
	5. Análises básicas
	6. Análises avançadas
Escolha	7. Representação
	8. Interpretação
	9. Utilização

**Fonte:** elaborado, com base em Porter e Cunningham (2005)

Percebe-se que a principal intenção dos autores é demonstrar que os recursos informacionais em C,T&I podem ser extraídos para gerar inteligência técnica competitiva efetiva, a partir da metodologia *Tech Mining*. É necessário trabalhar com ênfase nas necessidades dos usuários para gerar produtos de inteligência tecnológica bem orientados. A justificativa para perseguir esta premissa é a de que a Gestão de Tecnologia (*Management of Technology* - MOT) pode ser beneficiada por indicadores empíricos, para complementar a perícia (PORTER; CUNNINGHAM, 2005). MOT é a ligação de gestão, ciência e disciplinas de engenharia, para gerir as capacidades tecnológicas de forma que atinja os objetivos organizacionais, assegurando uma orientação para o cliente, ao invés de uma orientação para a tecnologia (HEQUET, 1991).

O campo de estudos de *Tech Mining* é definido sobre duas operações essenciais, ou conceitos importantes: *Exploration* e a *Exploitation*. Alguns autores, com base em uma analogia "evolucionária", chamam essas duas operações de variação e seleção. A primeira, é o ato de variar o conhecimento já existente na busca de novos conhecimentos, significa descobrir o contexto de ideias, inovadores e instituições. Às vezes, o processo envolve variar ideias que têm sido bem-sucedidas no passado. A segunda, é o ato de selecionar e juntar conhecimentos de áreas diferentes, combinando e recombinao para a extração de um novo produto, buscando atender uma necessidade de mercado (CORDEIRO, 2010). Envolve a utilização plena de conhecimento e relacionamentos existentes; bem como inclui estabelecer ligações, anteriormente não existentes. Muitas vezes, envolve a seleção e fomento de alguns componentes do processo de inovação, em detrimento de outros.

Parte da procura do núcleo das pesquisas, e posterior busca de meios para aprimorar o fluxo principal de atividades (por exemplo, o aspecto "quente" do desenvolvimento de determinado produto). Por outro lado, pode-se prosseguir a exploração quando avistamos uma justaposição não-usual de descobertas de pesquisa ou nova aplicação de uma ferramenta. Esses conceitos são importantes, pois mostram o potencial de uso de *Tech Mining*.

### ***2.5 Teoria da Solução Inventiva de Problemas - TRIZ***

TRIZ é uma sigla que, em russo, quer dizer: Teoria da Solução Inventiva de Problemas, criada pelo engenheiro russo G. S. Altshuller, que iniciou seus estudos na década de 1940. Quando Altshuller e seus companheiros estudaram centenas de milhares de patentes, eles verificaram que os tipos de problemas e os tipos de soluções se repetiam entre as diferentes áreas técnicas (CARVALHO, 2007). A princípio, buscava-se desenvolver uma metodologia generalizada o suficiente para a resolução de problemas inventivos, porém, a TRIZ expandiu-se para além desta perspectiva. A TRIZ analisa o micro para compreender o macro, ou seja, a análise de múltiplas patentes proporcionou que os problemas resolvidos por elas pudessem ser generalizados. Assim, a solução de muitos dos problemas, que surgem de novos desenvolvimentos, podem ser sistematizados.

A investigação, usando o método TRIZ, parte da hipótese que existem princípios universais de invenção e que estes são a base para as inovações e o avanço tecnológico. Identificando e codificando os princípios criativos, estes podem ser ensinados. Como

consequência, a metodologia TRIZ pretende aprofundar e dinamizar o processo criativo (LOPEZ; ALMEIDA; MOREIRA, 2004).

Verbitsky (2004) define " Semântica TRIZ " como uma aplicação de software com inteligência suficiente para "entender", em termos de leitura, milhões de documentos de C,T&I. Semântica TRIZ é baseada em indexação semântica, e sua aplicação lança uma nova abordagem para o processo de inovação. Verbitsky nota que, utilizando a relação entre elementos para analisar os termos de uma patente ou de um artigo científico, uma série de relações entre frases substantivas e seus verbos (ações) pode ser extraída. Isto permite que se relacione conceito com uso ou função e, assim, ajudar a compreender melhor a evolução do sistema inventivo ou tecnologia em estudo.

Os Princípios Inventivos (PIs) são a base central de um dos métodos de solução de problemas da teoria TRIZ – o Método dos Princípios Inventivos e Matriz de Contradições. Estas contradições são continuamente recorrentes em diferentes setores, e essas mesmas contradições têm sido frequentemente resolvidas, usando abordagens de repetição. As contradições aparecem quando, tentando melhorar uma característica ou parâmetro de um sistema, faz-se com que outra característica ou parâmetro do sistema piore (LOPEZ; ALMEIDA; MOREIRA, 2004). Também foi descoberto que os padrões de evolução do sistema técnico são repetidos em diferentes setores, assim, especificamente, os sistemas estão desenvolvidos nas direções de aumento da idealidade, harmonização, controlabilidade e grau de liberdade. E, finalmente, tornou-se amplamente reconhecido que as melhores soluções inovadoras parecem adotar efeitos científicos de diversas áreas da ciência e engenharia.

Em ciência, é sempre um desafio derivar uma lei geral a partir de observações empíricas específicas. Para Verbitsky (2004), aqueles que empreendem esse desafio devem ser capazes de responder a algumas perguntas sobre a validade da extrapolação. Para ferramentas baseadas em TRIZ, essas perguntas são bastante óbvias, por exemplo:

- a) Ao realizar uma análise de patentes e eventualmente analisar todas as patentes existentes, irá a matriz contradição mudar? Se sim, então como é que vai mudar?
- b) São descobertas tendências de evolução da tecnologia estatisticamente estáveis para um número de patentes analisadas? Será que elas cobrem todas as tendências existentes no mundo atual da tecnologia?
- c) Como atravessar o abismo de recomendação geral para uma ideia inovadora específica?

Para responder a essas perguntas, são necessárias soluções na forma de softwares capazes de lerem milhões de documentos e serem "inteligentes" o suficiente para compreendê-los. A

solução que providencia tais meios é baseada na tecnologia de indexação semântica, e sua aplicação na prática de resolução de problemas constitui uma nova abordagem para o processo de inovação, denominada Semântica TRIZ™ (VERBITSKY, 2004).

## 2.6 Tecnologia de busca semântica – *Goldfire*

*Goldfire* é uma plataforma de inteligência que integra métodos eficazes de mineração de dados, acesso semântico ao conhecimento e a abordagem criativa TRIZ. Comercializado pela *Invention Machine Corporation*, *Goldfire* inclui um poderoso motor de busca semântico. Oferece aos engenheiros, cientistas e pesquisadores a descoberta de conhecimentos e habilidades analíticas da tendência de inovação, bem como o acesso às patentes e seu conteúdo científico. Disponibiliza acesso às fontes mais relevantes do mundo globalizado, levando o usuário a formular os problemas de forma pertinente e adequada.

Criada em 1992, a *Invention Machine* focou exclusivamente no desenvolvimento e implementação de *softwares* de inovação concebidos para ajudar os fabricantes na identificação, validação e classificação de ideias.<sup>7</sup> Em 2012, a *Invention Machine* foi comprada pela IHS Inc. e, desde então, essa aquisição tem permitido a empresa IHS à oportunidade de se expandir ainda mais, por meio da criação de novos produtos de qualidade, aplicabilidade de normas e gestão do conhecimento.<sup>8</sup>

Capurso e Magri (2009) definem *Goldfire* como um software inovador que combina fluxo de trabalho com capacidades comprovadas de inovação e de colaboração com o acesso ao conhecimento coletivo exterior.

Algumas características sobre o *Goldfire*, de acordo com Capurso e Magri (2009), são:

- A interface de consulta (perguntas) em linguagem natural e síntese intuitiva automática de documentos permite que engenheiros e cientistas encontrem soluções às suas necessidades funcionais;
- Cada frase de um documento é indexada com base no seu significado funcional - o assunto, ação e objeto - a fim de facilitar o acesso rápido e a compreensão do seu conteúdo técnico;

---

<sup>7</sup> [http://everything.explained.today/Invention\\_Machine/](http://everything.explained.today/Invention_Machine/)

<sup>8</sup> <http://press.ihs.com/press-release/corporate-financial/ihs-acquires-invention-machine-and-releases-updated-fy2012-guidanc>

- O sistema possui uma série de tecnologias semânticas que entendem o significado da pergunta ou da sentença, dentro do contexto da busca. Seleciona o problema, extrai os conceitos apropriados e gera certos pressupostos / ideias;
- É capaz de analisar corretamente o conteúdo do texto, incluindo informações e links para as patentes, *websites* e documentos técnicos.
- A Mereologia (ciência que estuda a relação entre as partes em um todo) é a base dos processos de qualidade voltados para simplificação do sistema (produto, máquina, sistema, processo, etc.), permitindo a identificação de relações funcionais entre as diferentes partes/entidades constituintes do conjunto.
- É multilíngue (inglês, francês, alemão e japonês), trabalha com informação no formato sujeito/predicado/complemento, contextualizando a frase de forma adequada e retornando uma resposta coerente com os temas abordados.

O *software* oferece uma gama de ferramentas analíticas e metodológicas para suporte de resolução de problemas. Baseando-se em Lněnička (2012), o quadro 7 apresenta as funções mais eficazes desse *software*.

**Quadro 7:** Algumas funções do *software Goldfire*

<b>Busca de informações sobre um determinado tópico</b>	Visão geral das patentes, as tendências e os agentes; potenciais causas ou consequências de uma ação ou evento; entendimento dos diferentes elementos da técnica, incluindo seus relacionamentos.
<b>Busca sobre determinada tecnologia</b>	Explora oportunidades (habilidades, potenciais) ou a área de mercado em que a tecnologia funciona; revela as vantagens e desvantagens da tecnologia, o impacto do projeto; avalia oportunidades de concorrentes.
<b>Evolução de um produto</b>	Ajuda a criar o mapa em termos de desenvolvimento de produtos, a entender as tendências atuais e onde essas tendências estão indo; se concentra no produto, a partir de um ponto de vista estratégico, e ajuda a criar uma vantagem competitiva útil ou reviver um produto que já perdeu a sua atratividade.
<b>Projetando um novo produto</b>	Ajuda a atender as exigências ou oportunidades do mercado atual; gera um conjunto de possíveis estratégias para enfrentar as oportunidades e recomendações específicas para a entrada ideal do produto no mercado.
<b>Solução para defeitos do produto</b>	Através de relatos sobre problemas de produtos e das soluções na tentativa de encontrar uma maneira mais eficiente, identifica a raiz do problema, gera um conjunto de recomendações para soluções corretivas.
<b>Diagnóstico das causas da falha do produto</b>	Ajuda a determinar exatamente o que causa falha do sistema técnico, gerando um conjunto de causas.
<b>Melhora do produto para atingir o objetivo específico</b>	Ajuda a aumentar o lucro de um produto, responder a um novo produto competitivo, aumentar funcionalidade (idealidade) do produto ou encontrar outros requisitos regulamentar ou legislativo.

<b>Resolução de conflitos</b>	Esta ferramenta pode ser utilizada quando é necessário melhorar uma certa característica (traço ou parâmetro) do sistema, sem prejudicar outras características. O resultado é uma série de possíveis estratégias para resolver contradições.
<b>Previsão analítica de possíveis distúrbios</b>	Esta ferramenta irá ajudar a prever o potencial de falha do produto ou processo, e desenvolver e avaliar estratégias para melhorar a segurança e confiabilidade de um determinado produto, a fim de aumentar a satisfação do cliente e reduzir os custos associados com queixas e período de garantia para o produto.
<b>Quebrando patentes</b>	Com esta ferramenta é possível traçar as soluções possíveis de patentes, que podem ser livres para serem usadas ou evitar taxas de licença.
<b>Encontrando um novo mercado para uma tecnologia já existente</b>	Ajuda a identificar a capacidade técnica da tecnologia e a oportunidade de ser usada em um novo mercado. O resultado é uma lista de potenciais oportunidades de mercado, usuários, parceiros e concorrentes.
<b>Análise das causas profundas</b>	Usada para analisar as causas de eventos adversos. Identifica o desempenho inferior do produto ou do processo, gerando uma árvore de causas e consequências de eventos adversos, problema ou desempenho abaixo do ideal e possíveis estratégias para aliviar as causas.
<b>Análise de Produtos</b>	Ajuda a modelar e analisar a função de um determinado produto, apresentando um modelo das funções do produto, incluindo a lista de problemas detectados e matriz de ligações entre os componentes do produto.
<b>Análise do processo</b>	Ajuda a modelar e analisar a funcionalidade de um determinado processo, onde é possível entender todas as operações do processo ou modelo, para captar a sua sequência de relação funcional.

Fonte: elaborado de acordo com Lněnička (2012)

Em março de 2014, foi concedida uma nova patente para a tecnologia de perguntas-respostas por trás do software IHS *Goldfire*, a Patente No. 8666730, denominada como *Sistema de Pergunta-Resposta e Método Baseado em Marcação Semântica de Documentos de Texto*, para questões formuladas pelo usuário em linguagem natural. A patente abrange o conceito semântico "lentes", que IHS *Goldfire* usa para acelerar a pesquisa. As lentes se correlacionam com o sistema de conhecimento humano, organizando e apresentando respostas para as perguntas dos usuários, permitindo explorar rapidamente definições de conceito, aplicações, vantagens, desvantagens e muito mais.

Diferente das tecnologias de perguntas e respostas atuais, que funcionam como motores de meta-busca para procurar fragmentos de texto por palavra-chave e, em seguida, tentar extrair respostas semelhantes ao fragmento de texto, a abordagem IHS *Goldfire* é inteiramente original. Diferentes tipos de documentos podem ser analisados por um processador semântico para extrair relações como sujeito-ação-objeto, causa-e-efeito e dezenas de outros (TODHUNTER; SOVPEL; PASTANOHAU, 2014).

Simultaneamente, o sistema pode executar marcação semântica em termos de tipos de perguntas (que modela o comportamento humano) para as chamadas palavras-alvo (do inglês *target words*), usando classificador predefinido de tipos de perguntas e componentes de possíveis respostas (para documentos de texto).

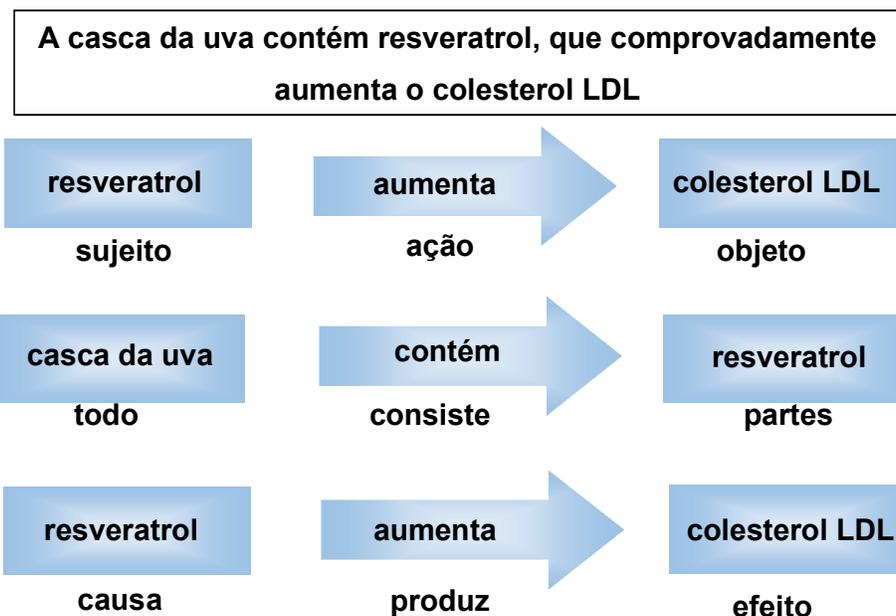
*Goldfire* processa semanticamente um pedido do usuário em linguagem natural para identificar e armazenar estruturas linguísticas sujeito-ação-objeto (SAO). A análise semântica pode incluir: reconhecer um ou mais fatos sob a forma de um ou mais conjuntos expandidos de (SAO) nos documentos de texto e pergunta do usuário. Reconhecendo um ou mais conjuntos expandidos de (SAO) e relações causa-efeito nos documentos de texto e pergunta do usuário, pode-se incluir o acesso a uma base de conhecimentos linguísticos, tendo um banco de dados de padrões que definem componentes (SÃO) e de causa-efeito.

Na prática, aplicando essa análise para a seguinte frase *Dissociação eletrolítica pode ser utilizada com sucesso para medir a umidade do ar?* O computador irá determinar que o sujeito é a dissociação eletrolítica, a ação é a medida e o objeto é a umidade do ar. Esses itens contêm informações sobre quais perguntas podem ser feitas, e que respostas podem ser dadas. Por exemplo, se alguém faz a pergunta: *Como posso medir a umidade?* A ação é *medir* e o objeto é *umidade*. O que é desconhecido para o usuário é o sujeito (o que mede a umidade?).

Ao fazer a análise semântica de todos os documentos disponíveis, incluindo a estrutura sintático-semântica acima, serão identificados todos os assuntos, ações e objetos, relacionados a esta pergunta. Serão fornecidas as combinações de sujeito-ação-objeto, que tenham a combinação ação e objeto como a da pergunta, devolvendo uma resposta exata a correspondente pergunta (por exemplo, a dissociação eletrolítica - medir – umidade do ar).

O método é baseado em um processador semântico que realiza preformatação de documentos de texto, análise linguística de documentos de texto/consultas de usuário (por exemplo, análise léxica, morfossintática, sintática e análise semântica); bem como a sua rotulagem semântica (do inglês *automatic semantic labeling*), usando termos (marcadores) de tipos de conhecimentos básicos (como objetos/classes de objetos; fatos e as regras que refletem regularidades do mundo/domínio de conhecimento, sob a forma de relações de causa-efeito), seus componentes e atributos. Extrai significados subjacentes, para que o usuário obtenha de volta respostas relevantes de forma precisa, como pode ser visto na figura 5.

Figura 5: tecnologia semântica do *Goldfire*



Fonte: elaborado pela autora

O sistema analisa o texto solicitado e, automaticamente, marca cada palavra com um código, indicando o tipo de cada uma delas. Uma vez que todas as palavras do pedido tiverem sido marcadas, o sistema executa uma análise semântica que, em um exemplo, inclui determinar e armazenar os grupos de verbo na primeira frase do pedido, em seguida, determina e armazena os grupos de substantivo dentro da frase solicitada. Este processo é repetido para todas as frases no pedido.

O quadro 8 apresenta algumas categorias e subcategorias que o *software* utiliza para fazer a rotulagem semântica.

**Quadro 8:** categorias e subcategorias do *software Goldfire*

General Facts	Parts and functions	Parameters	Causes and effects	People&Roles
Definitions	Made up of	Speed	Causes	People&Roles
More specific	Part of	Distance	Effects	Persons (object of sentence)
Concepts	Objects acted upon	Height	Preventions	Persons (subject of sentence)
Advantages	Function (action + object)	Depth		Achievements
Disadvantages	Action subjects	Length		Appraised as
Applications	Interactions	Mass		Awards
Methods		Age		Education
Conditions		Size		Occupations
Locations		Thickness		Skills
Materials		Width		Employed by
		Frequency		Achivers
		Color		Award winners
		Duration		Person educated at
		Shape		Professionals
		Number		Specialists
		Amount		Employees
		Values of a parametr		
		Parametr		

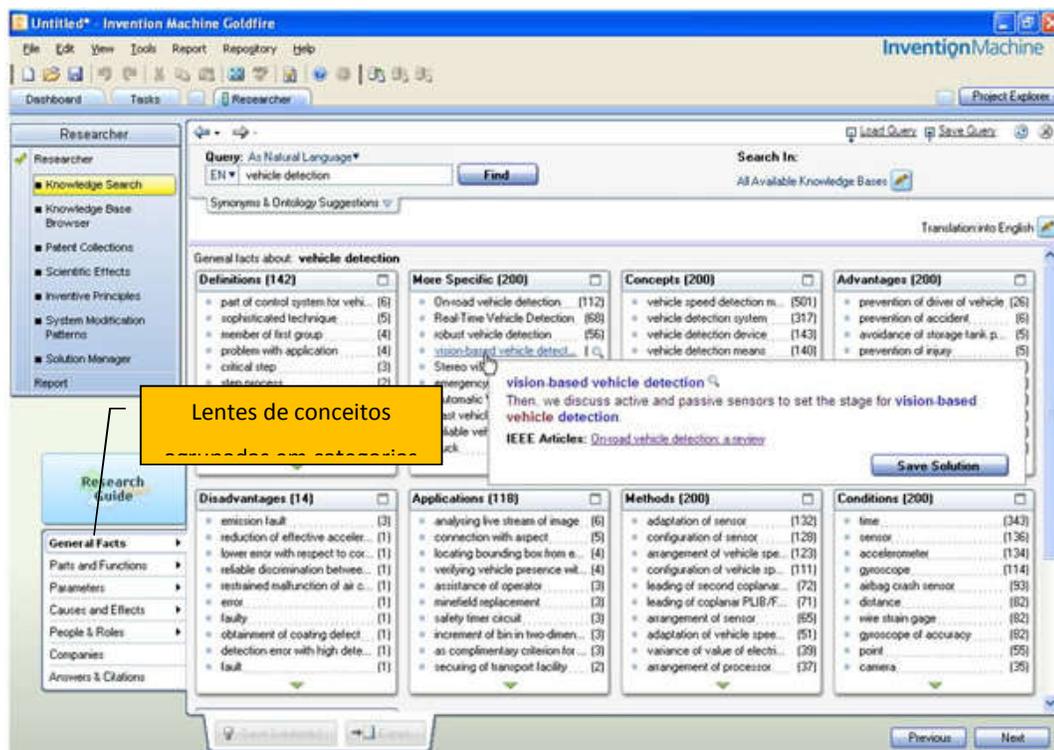
**Fonte:** *software Goldfire*

O sistema executa a análise linguística de uma pergunta recebida e realiza rotulagem semântica em termos dos principais tipos de conhecimento e de seus componentes. O sistema faz isso usando modelos de rotulagem texto, que correspondem ao comportamento humano. Neste caso, um procedimento de escolha utiliza os tipos de rótulos semânticos mencionados e permite uma resposta exata para a pergunta solicitada (TODHUNTER; SOVPEL; PASTANO-HAU, 2014).

A partir disso, o usuário pode obter respostas geradas semanticamente, compostas de citações diretas ou resumos das fontes de conteúdo relacionados com a sua pesquisa. Citações incluem links para os documentos finais. As lentes de conceitos (figura 6) fornecem *insights* poderosos em definições, vantagens, desvantagens, aplicações, materiais, causas e efeitos, peças, funções, interações, e muito mais.

Focado no significado semântico da consulta do usuário, *Goldfire* sugere respostas novas e relevantes de fontes documentais e fornece resoluções além de seu âmbito tradicional.

Figura 6: tela do software Goldfire



Fonte: tela do software Goldfire

## 2.7 Revisão sistemática da literatura científica sobre o *Software Goldfire*

A revisão sistemática da literatura apresentada nesta seção, apesar de estar inserida no referencial teórico deste trabalho dissertativo, configura-se como um dos resultados desta pesquisa.

Os procedimentos metodológicos para a realização da revisão sistemática são descritos na seção 3 desta Dissertação (o qual visa descrever a metodologia do presente estudo). Tendo em vista que a presente pesquisa tem como objetivo analisar o uso do *software Goldfire* ao ser mencionado na produção científica da comunidade mundial, os resultados apresentados nesta

seção foram alcançados adentrando-se no conteúdo das publicações selecionadas, ou seja, a partir da análise dos 29 trabalhos (artigos, anais de congresso, dissertações, teses), os quais permitiram acesso ao texto integral. Os documentos foram analisados observando-se a intensão do autor ao mencionar o *software Goldfire* e a relação do *software* com os objetivos e a metodologia do trabalho. Dentre outras coisas, foi verificado se os trabalhos têm apresentado experimentos práticos, ou se são de cunho teórico.

A seguir são apresentadas as discussões que emergiram durante a análise do conteúdo das 29 publicações, das quais foi possível ter acesso ao texto completo. Tendo em vista que esta análise permitiu adentrar no conteúdo das publicações, optou-se por chamá-la de *síntese das publicações analisadas*.

### ***2.7.1 Síntese das publicações analisadas***

#### **2.7.1.1 Impacto**

Dois trabalhos merecem destaque, primeiro o artigo *CONG, He; TONG, Loh Han. Grouping of TRIZ Inventive Principles to facilitate automatic patent classification. Expert Systems with Applications, v. 34, n. 1, p. 788-795, 2008*, que teve um impacto de 88 citações no Google Acadêmico. Essa informação foi obtida através do programa “*Publish or Perish*”, que também auxilia na sistematização dos dados de citação. Ao realizar uma pesquisa no programa, o usuário recupera uma lista das referências pertinentes à expressão de busca e gera estatísticas de produtividade e de impacto de autores e periódicos.

O artigo, acima citado, propõe um sistema especialista baseado em texto, que permite classificar patentes de acordo com princípios inventivos TRIZ. Cong e Tong (2008) afirmam que a classificação automática de patentes facilita a procura de documentos de patentes anteriores e, para este efeito, é necessário um sistema de classificação de patente baseado na TRIZ, o que gerou grande interesse por parte de pesquisadores nessa metodologia. Usuários da TRIZ precisam pesquisar patentes com base nas soluções TRIZ, princípios inventivos e contradições abordadas nas patentes, o que é diferente da pesquisa tradicional com base nos campos de aplicação das invenções.

Princípios combinados com Contradição irão fornecer uma visão mais ampla para os usuários da TRIZ, ajudando-os a encontrar inspiração possível, a partir de um campo que pode ser totalmente diferente do deles. Ainda sobre o impacto das publicações analisadas, em segundo lugar vem o artigo *SÁNCHEZ, Jenny Marcela; MEDINA, Javier E.; LEÓN, Andrés Mauricio*.

*Publicación internacional de patentes por organizaciones inventores de origen colombiano. Cuadernos de Economía*, v. 26, n. 47, p. 247-270, 2007, com 12 citações no Google Acadêmico. Os demais trabalhos obtiveram menos de 10 citações.

### 2.7.1.2 Abordagem teórica

As publicações enquadradas nesta categoria tratam de estudos desenvolvidos com o intuito de descrever ou aprofundar a compreensão sobre o *software Goldfire*. Foram selecionados neste grupo dois trabalhos que dedicaram pelo menos um capítulo ou seção para abordar as principais vantagens do *software* quanto à busca semântica e as aplicações metodológicas para suporte de resolução de problemas por meio da TRIZ.

	ANO	AUTOR	TÍTULO
1	2009	Davide Capurso; Eugenio Magri.	Modelli di rappresentazione della conoscenza a supporto del benchmarking tecnologico.
2	2012	Luboš Lněnička	Tvorba inovačních zadání a jejich řešení

#### 1- Capurso e Magri (2009)

A proposta da dissertação foi discutir o tema *avaliação comparativa de tecnologia*, que tem como objetivo o monitoramento das atividades inventivas de concorrentes, resultando em economia de tempo, recursos e investimentos para ser utilizado no desenvolvimento de novos produtos ou processos. No capítulo intitulado “Invention Machine Goldfire Versão 4.0”, os autores se referem ao *Goldfire* como um *software* inovador que combina fluxo de trabalho com capacidades comprovadas de inovação e de colaboração com o acesso ao conhecimento coletivo exterior. Funcionando como um motor de busca semântica, o *Goldfire* oferece aos engenheiros, cientistas ou pesquisadores a descoberta de conhecimentos e habilidades analíticas da tendência de inovação, por meio de uma pesquisa de precisão, bem como um acesso sem precedentes à patente e seu conteúdo científico.

A partir de um exemplo realizado com patentes, o trabalho procurou demonstrar a diferença entre uma busca booleana e uma busca semântica. Propôs um exemplo onde foi realizada uma

pesquisa booleana pelo termo “vibração”, os resultados não foram satisfatórios, devido ao reconhecimento incorreto da ligação entre a ação e o princípio físico. O termo “vibração” foi relacionado com várias outras questões, quando deveria ser relacionado apenas ao princípio físico. Essa questão pode ser resolvida pelo uso do motor de busca semântica *Goldfire* que, realizando pesquisas de patentes em linguagem natural, pode realmente ser capaz de retornar patentes em que o princípio físico e a ação estejam relacionados na busca.

## 2- Lněnička (2012)

No segundo trabalho, também uma dissertação de mestrado, o autor trata sobre uma metodologia para a criação de projetos e soluções inovadoras. Mais especificamente trata do uso da metodologia TRIZ na criação de soluções inovadoras em tarefas inventivas. A segunda seção do trabalho é dedicada ao *software Goldfire Innovator*. Lněnička (2012) ressalta que o *software* centra-se em todo o ciclo de vida dos produtos diante das exigências do mercado atual, tendo como principais vantagens a utilização de uma base de dados de patentes, o conhecimento recuperado através da busca semântica e os princípios definidos de desenvolvimento de sistemas técnicos (princípios heurísticos, modelos e soluções).

O trabalho analisado também ressaltou um novo recurso do *Goldfire*, que seria a possibilidade do *software* de se conectar a redes sociais e *feeds* RSS (*Really Simple Syndication*), e desenhar, a partir de comentários e sugestões, a melhoria de produtos existentes. Para finalizar, foi realizado um estudo de caso sobre o desenvolvimento da energia eólica com o apoio do *software Goldfire Innovator* e os benefícios da metodologia TRIZ, onde, durante o processo da pesquisa, foi observado como pesquisadores podem trabalhar na resolução de problemas usando TRIZ.

### 2.7.1.3 Aplicado em estudo de caso

O objetivo desta categoria foi revelar os experimentos práticos para os quais o *Goldfire* foi utilizado, enfatizando como os pesquisadores têm usado o *software* em termos de ferramenta para resolver problemas, ou seja, em quais aplicações este tipo de tecnologia foi utilizada. Foram selecionados, para esta categoria, 11 trabalhos.

	<b>ANO</b>	<b>AUTOR</b>	<b>TÍTULO</b>
1	2004	Siegfried Luger	Lighting helmet for Formula 1
2	2007	Jenny Marcela; Sánchez, Javier E. Medina; Andrés Mauricio León.	Publicación internacional de patentes por organizaciones inventores de origen colombiano
3	2008	Nan Si; Pengfei Yu; Lin Zhang.	Semantic SOA based integrated development environment for complex information systems
4	2008	Juan Gonzalo Castellanos Méndez; Juan Carlos Rodríguez Cárdenas; Claudia Liliana Luengas Camelo	Informe de vigilancia tecnológica em mejoramiento genético florestal
5	2008	Oscar Fernando Castellanos D., Luz Marina Torres P., Sandra Lorena Fonseca R., Victor Mauricio Montañez F.	Tendencias en investigación, desarrollo Tecnológico y mercado en la cadena productiva de cacao-chocolate.
6	2008	Jenny Marcela Sánchez-Torres,	Apoyo en la definición de políticas públicas en ciencia, tecnología e innovación a través de vigilancia tecnológica e inteligencia competitiva
7	2013	Yi Zhang; Xiao Zhou; Alan L. Porter; Jose M. Vicente Gomila; An Yan.	Triple Helix innovation in China's dye-sensitized solar cell industry: hybrid methods with semantic TRIZ and technology roadmapping
8	2014	Yi Zhang; Xiao Zhou; Alan L. Porter; Jose M. Vicente Gomila.	How to combine term clumping and technology roadmapping for newly emerging science & technology competitive intelligence: The semantic TRIZ tool and case study
9	2014	Pavel šic	Dimenzování proudovodné dráhy elektrických přístrojů
10	2014	Bohuslav Busov; Milada Bartlova	Improvement of active hinge of the car bonnet
11	2014	José Vicente Gomila; Fernando Palop Marro; Angela Machado Rocha; Maria Fernanda Ariza; Maria Clara Melro.	O acesso transversal ao conhecimento no processo da pesquisa médica: diagnósticos e biomarcadores no câncer da próstata

1- Siegfried Luger (2004)

Trabalho apresentado na 4ª edição da TRIZ *Future Conference*, explica o desenvolvimento de um novo conceito para capacete de carros de corrida com base em Metodologia TRIZ.

O *software Goldfire Innovator* foi utilizado para modelar o sistema; as soluções padrões foram agrupadas em função das matérias-alvo do problema, definiu-se "Aumentar a eficiência" como a categoria de ordenação. Além disso, o *software* sublinha as tendências evolucionárias para esta solução: controlando a ação diretamente no objeto - ação por meio de acionamento de mecanismo. O conceito de capacete com iluminação leva a cooperação com os principais capacetes fabricados para a Fórmula 1. Este conceito, baseado em TRIZ, direciona para as patentes europeias e proporciona novos impulsos de inovação na indústria de capacete.

Segundo o autor deste trabalho, a vantagem de usar a metodologia TRIZ pode estar em expandir todos os pensamentos obstruídos, que vêm de experiências existentes. Abrir a mente para formulação de contradições e conceito de pesquisa baseado em princípios inventivos foi muito importante. Analisar o sistema hierarquicamente, centrando-se na zona de operação, nos leva a uma nova visão do problema. Com TRIZ encontramos novas maneiras essenciais para criar as necessidades dos clientes e satisfazê-las, garantindo a inovação a longo prazo e a liderança de mercado (SIEGFRIED LUGER, 2004).

## 2- Sanchez, Medina e Leon (2007)

Os autores apresentaram uma análise do registro de patentes por parte de instituições ou inventores colombianos nas bases de dados das Oficinas de Patentes dos Estados Unidos e Europa, entre 1968 e 2007.

O *software GoldFire* foi utilizado na obtenção dos dados para medir a atividade de patentabilidade. Com os dados obtidos foram realizadas as seguintes análises: comparação da atividade de patentabilidade; principais áreas do conhecimento de patentes colombianas em oficinas internacionais; evolução das áreas do conhecimento de patentes colombianas em oficinas internacionais de patentes; inventores com maior índice de patentabilidade; patrocinadores com maior índice de patentabilidade, patentabilidade registrada por universidades colombianas em oficinas internacionais; patentes que são usadas como referências na atribuição das patentes colombianas.

## 3- N Si, P Yu, L Zhang (2008)

Os autores implementaram um protótipo de sistema para CISR (do inglês *Command Control Communications Computers Intelligence Surveillance and Reconnaissance*). Na concepção

dos principais recursos do sistema protótipo foi mencionado o uso do *software Goldfire*, como plataforma de inovação conceitual. Entretanto, não foi relatado no artigo mais informações sobre o uso do *software*.

#### 4- Méndez, Cárdenas e Camelo (2008)

O objetivo do trabalho foi obter informações de fontes acadêmicas e científicas, válidas e reconhecidas mundialmente, a respeito do tema “o melhoramento de sementes florestais” como fator de aumento da produtividade e rentabilidade das plantações comerciais. Para busca e análise de patentes foi utilizado o *software Goldfire Researcher*, com o objetivo de conhecer as invenções, os inventores, as empresas patenteadoras e a dinâmica do tema nas patentes.

#### 5- Castellanos D. e Torres P. (2008)

Trata de um estudo de Vigilância Tecnológica e Comercial, que permitiu estabelecer o estado atual da investigação e o desenvolvimento tecnológico a nível mundial, bem como determinar as características mais importantes do mercado de produtos finais na cadeia produtiva do cacau.

O *Goldfire* foi utilizado como ferramenta especializada para análise e processamento. O estudo foi realizado a fim de contribuir para a identificação dos fatores críticos que contribuem para a construção da Agenda Prospectiva de Investigação, por meio da exploração de tendências mundiais e das capacidades nacionais. Para isso, foi elaborado um conjunto de perguntas de investigação que guiaram o estudo. O desenvolvimento tecnológico expresso nas patentes revelou uma alta concentração de solicitação e concessão de patentes que abordam a composição de produtos e alimentos derivados do cacau, fundamentalmente os produtos de chocolate.

Também, com uma importante participação, se encontram os desenvolvimentos tecnológicos em termos de processos e maquinaria para a produção de chocolate, tal é o caso das patentes relacionadas com processos de resfriamento na fabricação de chocolate, que é um dos principais problemas enfrentados pelos pequenos produtores colombianos, sendo possível, por meio da revisão dessas investigações e da contextualização das condições de processo, estabelecer estratégias de melhoramento.

#### 6- Sánchez-Torres (2008)

O trabalho teve como objetivo definir as agendas temáticas de investigação por meio de exercícios de Vigilância Tecnológica (VT) e Inteligência Competitiva (IC) em quatro cadeias produtivas agroindustriais na Colômbia. Para conhecer as tecnologias, os inventores e a dinâmica

das empresas foi feita uma busca por patentes em bases de dados internacionais com o *software Goldfire*, inclusive bases japonesas.

7- Zhang et al. (2013)

Mais uma abordagem da semântica TRIZ foi utilizada para avaliar os padrões de inovação por associação de temas da *Newly Emerging Science & Technology* (NEST), usando frases substantivas para abordar temas e objetos; e ações via verbos. A análise foi demonstrada por meio de um estudo de caso para células solares sensibilizadas por corante. Baseado nos métodos da semântica TRIZ com o *GoldFire Innovator*, foram analisados os itens e as sentenças relacionadas por meio do rastreamento das atividades anuais de patentes. Tais informações, indicadas pelos padrões de problema e solução, oferecem um rico potencial de inteligência técnica competitiva para rastrear interesses globais ou individuais de uma organização ao longo do tempo.

8- Zhang et al. (2014)

Mesmo autor do trabalho anterior, propôs um método bibliométrico para associar resultados de “*term clumping*” com TRM, usando ferramenta da semântica TRIZ (nesse caso, o *software Goldfire*). Abordou a semântica TRIZ como uma ferramenta útil para representar resultados “*term clumping*” e aplicá-los a Tecnologia *Roadmapping* (TRM), com a ajuda da semântica “*Problem & Solution*” (P&S).

Com o objetivo de construir uma ponte entre os resultados de “*term clumping*” e abordagens TRM, de forma efetiva, este artigo enfatizou estudos do conceito e metodologia se baseando nos subprocessos da TRIZ semântica. Além disso, envolveu as estruturas “Sujeito - Ação - Objeto (SAO)” para representar os principais conceitos e conhecimentos dos inventores que definem as relações explícitas entre os componentes. O *Goldfire* foi utilizado para processar o conteúdo dos registros. Após recuperar os verbos relacionados com o componente do sistema, combinou tais verbos com a estrutura SAO.

9- Pavel (2014)

Centra-se na questão de projetar dispositivos de comutação elétricos, especialmente dimensionamento dos circuitos de corrente de transporte, de acordo com os efeitos das forças eletrodinâmicas. Por meio do *software Goldfire Innovator* foi recuperada soluções técnicas progressivas no estado da arte das patentes.

## 10- Busov e Bartlova (2014)

Apresenta uma solução inovadora para uma dobradiça ativa de capô do carro. A dobradiça ativa pode diminuir os efeitos mais comuns e graves de lesões na cabeça para pedestres, especialmente nas travessias de pedestres nas cidades. O artigo apresenta a utilização parcial da metodologia TRIZ e *Goldfire Innovator* para melhorar a dobradiça ativa do capô do carro. Utilizando resultados de contradição em várias recomendações inspiradoras (princípios heurísticos) vantagem e desvantagem: - Força / fatores prejudiciais.

## 11- Gomila et al (2014)

Por meio de programas de tecnologia de mineração como *VantagePoint* e *Goldfire*, o Instituto Nacional de Oncologia da Colômbia identificou os sistemas de detecção e monitoramento mais promissores assim como os grupos de pesquisadores e suas parcerias participantes relacionados com a o diagnóstico e prognóstico da pesquisa sobre o câncer de próstata. Os dados obtidos na coleção completa da *Pubmed*, em patentes internacionais, artigos da *ScienceDirect*, bem como todos os sites de *research.uk.org*, foram sintática e semanticamente processados a fim de obter métodos, tendências e trazer respostas às buscas críticas com o *software Goldfire*. O trabalho demonstrou que, por meio de ferramentas de apoio a inovação, é possível acessar conhecimento novo, de forma transversal, e identificar tendências com evidências objetivas, tudo isso com forte embasamento nas atividades das pesquisas mundiais das ciências médicas e da saúde.

## 2.7.1.4 Mencionado no texto

Este grupo apresenta uma breve visão da intensão dos autores em mencionar o *software Goldfire* ao longo do estudo, bem como enfatiza as dificuldades e limitações do *software*. Os 10 trabalhos selecionados nesta categoria abordam o *software Goldfire* com diversas acepções.

	<b>ANO</b>	<b>AUTOR</b>	<b>TÍTULO</b>
1	2004	<i>Valeri Souchkov</i>	Innovative enterprise infrastructure
2	2004	Peter Chuksin	Selecting of key problems and solution search area in forecasting

3	2005	Rafael Nunes Armada	Um ambiente de apoio a externalização e mapeamento do conhecimento em equipes distribuídas de trabalho
4	2008	Moisés Mir Mauri; Martí Casadesús Fa.	UNE 166002:2006: estandarizar y sistematizar la i+d+i la norma y la importancia de las TIC en su implementación
5	2011	Hans Bakker; Leonid Chechurin; Wessel Wits.	Integrating TRIZ function modeling in CAD Software
6	2011	Giovanni Arturo López Isaza; Martha Judith Correa Vallejo.	Fuentes de información e inteligência organizacional en investigación. El caso de la Universidad Tecnológica de Pereira
7	2011	Umakant Mishra	50 Usable Software on Innovation, Concept Mapping and Idea Management
8	2012	H.M. Bakker	Improving software assistance for design engineers by integrating mathematical and function modeling
9	2012	SUROSO, Ir Arif Imam	Mata Kuliah Sistem Informasi Manajemen Dosen.
10	2013	Katherine Fu; Joel Chan; Christian Schunn; Jonathan Cagan.	Expert representation of design repository space: A comparison to and validation of algorithmic output
11	2013	Katherine Fu; Jeremy Murphy; Maria Yang; Kevin Otto; Dan Jensen; Kristin Wood.	Investigating the effect of functionality level of Analogical Stimulation on Design Outcomes
12	2015	Katherine Fu; Jeremy Murphy; Maria Yang; Kevin Otto; Dan Jensen; Kristin Wood.	Design-by-analogy: experimental evaluation of a functional analogy search methodology for concept generation improvement

### 1- Souchkov (2004)

Discute a importância da infraestrutura baseada em inovação sistemática, para melhorar a gestão da inovação e o aumento da produtividade inovadora. Enfatiza, também, que a eficiência da inovação depende da nossa capacidade de produzir ideias de sucesso, que podem ser

realizadas sob a forma de produtos e tecnologias viáveis. O autor cita o *software Goldfire Innovator* como um tipo de sistema de Tecnologia da Informação (TI) importante na resolução de problemas e na geração de soluções inovadoras.

#### 2- Chuksin (2004)

Menciona o *software Goldfire Innovator* como um motor de busca relacionado à semântica TRIZ, que fornece excelentes possibilidades de informação para pesquisa por meio da internet.

#### 3- Armada (2005)

Em sua tese de doutorado, o autor apresenta, no capítulo 3, algumas tecnologias que apoiam a Gestão do Conhecimento e Aprendizagem Organizacional. Ressalta que, embora essas tecnologias não sejam capazes de gerar conhecimento por si só, quando usadas num ambiente propício, são capazes de agilizar e promover o processo de difusão e aquisição do conhecimento. O trabalho menciona que o *software Goldfire* é uma ferramenta capaz de trazer um processo estruturado ao uso inventivo do conhecimento na busca de soluções para problemas complexos. “Ele apoia o processo de inovação, guiando profissionais durante os estágios do projeto conceitual de um novo produto ou no processo de evolução contínua de produtos já existentes. Estimula a utilização de conhecimentos vindos de diferentes disciplinas e facilita o compartilhamento de conhecimento entre indústrias, através da representação abstrata de conceitos, melhorando a solução de problemas com o uso de metodologias de desenvolvimento de produtos comprovadamente eficazes” (ARMADA, 2005, p.56).

#### 4- Mauri e Casadesús (2008)

Analisam a importância das Tecnologias de Informação e Comunicação (TIC) no processo de inovação, sistematizando algumas atividades chave: vigilância tecnológica, criatividade, gestão do conhecimento e previsão tecnológica. O *Goldfire* foi mencionado como um tipo de ferramenta computacional que permite analisar tecnologias emergentes para a realização de estudos de previsão tecnológica. Como exemplo de tal tecnologia, foi gerado um gráfico- “Evolução de Patentes (previsão tecnológica) com o *software Goldfire*”.

#### 5- Bakker, Chechurin e Wits (2011)

Mencionam que *softwares* que fornecem função CAI (*Computer Aided Invenção*) - inovação auxiliada por computador - como parte de um fluxo de trabalho maior, como *Goldfire Innovator*, são geralmente muito caros. Isso os torna atraente, principalmente para as grandes empresas. Em segundo lugar, todas as suites CAI disponíveis são autônomos. Eles não oferecem integração com outros *softwares*.

#### 6- Isaza e Vallejo (2011)

Tiveram como objetivo conhecer as estratégias de busca de informação relevantes, empregadas por grupos de pesquisa registrados na Vice-reitoria de Investigações, Inovação e Extensão da Universidade Tecnológica de Pereira, Colombia. Os resultados foram obtidos a partir do processamento da informação proveniente de 69 entrevistados. Um dado interessante levantado pelos autores foi em relação às ferramentas e motores de busca (*software* ou ferramentas online utilizadas pelos investigadores), foi constatado que 62% utilizam ferramentas diferentes das oferecidas pela biblioteca da universidade. Dos pesquisadores que utilizam as ferramentas disponíveis pela universidade, (57%) utilizam *Science Direct College*, (55%) utilizam *Proquest*, (43%) utilizam *Google* e (12%) utilizam o *Goldfire Researcher*.

#### 9- Isaza e Vallejo (2011)

Observaram que, em geral, existe um desconhecimento ou uma subutilização das ferramentas ideais para obter informação com valor agregado na atividade investigativa. Em relação ao pouco uso do *software Goldfire*, os autores ressaltam que os pesquisadores estão perdendo a oportunidade de aproveitar uma ferramenta adquirida pela universidade por um alto custo, por desconhecerem o valor das patentes como ferramentas estratégicas para a aprendizagem tecnológica, e outros usos promissores mencionados neste artigo. Enfatizam, ainda, que o *software* permite acesso a mais de 30 milhões de documentos de patentes e outras fontes adicionais de conhecimento.

#### 10- Umakant Mishra (2011)

Ressalta que um *software* pode não ajudar diretamente a inovar, mas pode, definitivamente, ajudar na prática de qualquer processo específico de inovação, seja de geração e gestão de ideias, mapeamento de conceito ou análise de problema. O autor faz um levantamento de 50

*softwares* de inovação e mapeamento de conceito, o *software Goldfire* foi citado entre os 50 listados.

11- Bakker (2012)

Os *softwares* baseados em TRIZ podem ser divididos em dois grupos: o *software* que ajuda os engenheiros a aplicar um princípio desenvolvido usando TRIZ, e o *software* que orienta no processo de aplicação da TRIZ. O *software Goldfire* foi enquadrado pelo autor como o *software* que orienta no processo de aplicação da TRIZ, usando gestão estratégica, gestão de ideia e gestão de patentes.

12- Barata (2012)

Dedica um capítulo ao estudo dos tipos de sistemas informáticos existentes para o processo de inovação. O autor cita o *Goldfire Innovator* como exemplo de um *software* da área de “Sistemas de Inovação Auxiliada por Computador (*Computerized aided Innovation - CAI*)”. CAI corresponde às ferramentas de suporte ao utilizador na criação de modelos de resolução de problemas na inovação (BARATA, 2012, p.25).

13- Katherine Fu, Chan, Schunn e Cagan (2013)

Enfatizam quem existem muitas tentativas de automatizar, ajudar ou agilizar a busca da base de dados de patentes dos EUA, incluindo buscas simples de palavras-chave, bem como teorias como TRIZ e suas ferramentas, a exemplo do *software Goldfire*.

14- Katherine Fu, Murphy, Yang, Otto, Jensen e Wood (2013)

Citam o *software Goldfire* como uma ferramenta de suporte computacional da teoria TRIZ.

15- Katherine Fu, Murphy, Yang, Jensen e Wood (2015)

Apresentam os resultados do experimento que testa um novo método para a extração de analogia em fontes de dados, tais como bancos de dados de patentes para auxiliar pesquisadores na busca e identificação sistemática de analogias. Menciona a TRIZ como uma teoria que apre-

senta regras heurísticas ou princípios para auxiliar os pesquisadores a superar impasses no raciocínio funcional, com base em estudos iniciais de patentes em termos de contradições. Segundo os autores, *Goldfire* é uma ferramenta de exploração de patente, centra-se no objetivo da compreensão de paisagens competitivas, facilita a criação-por-analogia, por meio da transferência de domínio cruzado, baseado em função do conhecimento.

### **2.7.2 Considerações sobre as análises**

As vantagens e benefícios que podem ser conquistados com o uso do *software Goldfire* foram descritos e estabelecidos no desenvolvimento deste estudo, como também foram elencados os experimentos ou estudos de caso que enfatizaram questões sobre metodologias que se concentraram nos processos da TRIZ. Além disso, a estrutura *Sujeito - Ação - Objeto (SAO)* está envolvida e representa os principais conceitos e conhecimentos dos inventores que definem as relações explícitas entre os componentes. Na concepção dos principais recursos dos sistemas protótipos, foi mencionado o uso do *software* como plataforma de inovação conceitual. Como afirma Kiatake (2004), modelos de inovação conceitual são aqueles que buscam estruturar os problemas, não por meio de uma forma hierárquica, mas pelas relações entre as ideias, com vistas a possibilitar análises para posterior tomada de decisão.

Um outro ponto interessante diz respeito ao desconhecimento ou a subutilização do *software Goldfire*. Alguns autores observaram que muitos pesquisadores estão perdendo a oportunidade de aproveitar ferramentas ideais para obter informação com valor agregado na atividade investigativa, talvez por questões de custo ou por desconhecerem o valor das patentes como ferramentas estratégicas para a aprendizagem tecnológica e outros usos promissores, mencionados nas publicações analisadas.

Após a realização deste estudo, fica mais evidente as vantagens do *software*, o qual opera como uma ferramenta de apoio a geração de novos conceitos, com base no conhecimento científico e técnico existente em bases de dados de patentes e publicações científicas. O uso do *Goldfire* se baseia na utilização de técnicas avançadas de ferramentas semânticas, que são capazes de extrair de documentos um significado técnico. Utiliza cerca de 40 diferentes critérios semânticos. Em grande medida, estes critérios são baseados na teoria da inovação e resolução de problemas TRIZ.

A partir da análise dos trabalhos, foi possível observar que, em um mercado cada vez mais competitivo, é necessário ter acesso a dados precisos e atualizados para fazer escolhas consistentes e bem informadas, possibilitando tomar as melhores decisões. Nesse contexto, se

insere o *software Goldfire*, pois, por meio dele, é possível acessar conhecimento novo e identificar tendências com evidências objetivas, tudo isso com forte embasamento nas atividades das pesquisas mundiais, baseadas no método TRIZ para a resolução de problemas inventivos.

### 3 METODOLOGIA

Nesta seção será apresentado o método que caracteriza a pesquisa, bem como o percurso metodológico utilizado para sua realização.

#### 3.1 Tipo de pesquisa

De acordo com os tipos de pesquisa propostos por Gerhardt e Silveira (2009), quanto à abordagem, é adequado classificar a presente pesquisa como qualitativa, já que busca explicar o porquê de algo, exprimindo o que convém ser feito, mas não quantifica os valores e as trocas simbólicas, nem se submete à prova de fatos, pois os dados analisados são não-métricos (suscitados e de interação) e se valem de diferentes abordagens. É aplicada, quanto à natureza, porque esse tipo de pesquisa objetiva gerar conhecimentos para aplicação prática, dirigidos à solução de problemas específicos. Quanto aos objetivos, é empírico-descritiva, pois o estudo pretende descrever os fatos e fenômenos de determinada realidade. Em relação aos procedimentos, a pesquisa pode ser classificada como experimental, já que tem o objetivo de comprovar experimentalmente hipóteses, através da experiência em compreensão computadorizada de linguagem natural.

#### 3.2 Etapas da Pesquisa

##### 3.2.1 Revisão sistemática da literatura científica sobre o software Goldfire

A princípio, uma pesquisa foi realizada no Portal de Periódicos da CAPES, nas bases *Web of Science* e *Scopus*, onde foram utilizados os termos “*Goldfire*” e “*Goldfire software*” no campo busca por assunto. Porém, esse tipo de busca é realizada apenas nos títulos, resumos e palavras-chave, o que resultou em apenas 6 publicações. A mesma busca foi feita nas bases *Redalyc* e *Scielo*, retornando também 6 publicações.

Para uma busca mais abrangente envolvendo toda a literatura acadêmica disponível na web, foi utilizado o programa denominado *Publish or Perish*, um programa simples e gratuito, desenvolvido por Anne-Wil Harzing (2007)<sup>9</sup>, que funciona como uma interface de busca do

---

<sup>9</sup> <http://www.harzing.com/>

*Google Acadêmico*. Ao realizar a pesquisa no programa, foi recuperada uma lista com 53 referências pertinentes à expressão de busca “*Goldfire software*” e “*Software Goldfire*”. A seleção dos trabalhos foi realizada no mês de setembro de 2015. Não se optou por fazer um recorte temporal, já que o objetivo seria analisar toda a publicação científica sobre o *software Goldfire*.

O objetivo em utilizar o meta-buscador do *Google Acadêmico* foi devido ao fato desse tipo de busca considerar o texto integral de cada publicação, os resultados consideraram o termo pesquisado em qualquer lugar do texto completo. Jacsó (2005) ressalta que a grande vantagem da recuperação por meio dessa ferramenta consiste no fato de que a busca se dá em índices criados a partir do texto completo do documento primário.

Alguns dos 53 resultados apresentados pelo *Google acadêmico* eram de conteúdos pagos e não estavam acessíveis para *download*. Em se tratando de trabalhos pagos, o *Google Acadêmico* indexa os trabalhos publicados, porém o usuário acessa gratuitamente apenas sumários e resumos. Nesse caso, foi feita uma segunda busca no Portal de Periódicos da CAPES, dessa vez a busca foi realizada pelo título de cada trabalho não disponibilizado gratuitamente pelo *Google Acadêmico*. Assim, foi possível desfrutar dos recursos do *Google Acadêmico* como meta-buscador e, ainda, acessar o texto completo das publicações.

Após verificar a disponibilidade de todos os trabalhos para leitura integral, tanto os disponibilizados pelo *Google Acadêmico* como os disponibilizados pelo Portal de Periódicos da CAPES, foi realizado o tratamento de eliminação dos trabalhos redundantes, ficando, assim, 29 trabalhos, pois, mesmo no Portal de Periódicos da CAPES, algumas publicações não estavam disponíveis.

### **3.2.2 Extração de relacionamentos semânticos do tipo causa e efeito**

#### **3.2.2.1 Escolha do corpus**

O experimento foi realizado por meio da leitura dos resumos de um conjunto de artigos científicos relacionados à agronomia, todos escritos na língua portuguesa. A coleta dos artigos científicos foi realizada em outubro de 2015, por meio da *home page* da *Scientific Electronic Library Online – SciELO*. A seleção dos artigos se deu utilizando-se do seguinte critério de busca, na opção Assunto: Ciências Agrárias > Títulos correntes > Pesquisa agropecuária Brasileira. Nessas opções, foram feitas duas buscas, uma inserindo a palavra *fertilizante* e outra inserindo a palavra *aminoácidos*. Foram selecionados 30 artigos para o termo *fertilizante* e 30

para o termo *aminoácidos*, publicados nos últimos anos, até o ano de 2014, totalizando 60 artigos.

A importância de delimitar o assunto dos textos em uma área específica – no caso, a área de Ciências Agrárias – foi devido à necessidade de escopo e contextualização. Não há, entretanto, restrições de aplicabilidade da metodologia para documentos textuais oriundos de outras áreas do conhecimento.

### 3.2.2.2 Transferência do formato de arquivo para texto simples

As ferramentas tecnológicas utilizadas nesta metodologia necessitavam de documentos submetidos em formato de arquivos de texto simples. Como os documentos coletados na *web* se encontravam em formato PDF, esses documentos precisaram ser convertidos para o formato texto simples. O processo foi feito manualmente, o resumo em português de cada artigo foi copiado e transferido para arquivos em formato TXT (bloco de notas).

### 3.2.2.3 Análise morfossintática

Para a análise morfossintática dos 60 resumos, utilizou-se o *software Palavras*, uma ferramenta de processamento morfossintático de textos em português, desenvolvida por Bick (1996) em sua tese de doutorado na *Southern University of Denmark*, e que faz parte de um conjunto de ferramentas multilíngues chamado VISL<sup>10</sup> (*Virtual Interactive Syntax Learning*).

Sua função básica é identificar as classes gramaticais e os elementos sintáticos e semânticos que compõem uma sentença ou texto. O princípio de análise da ferramenta é a gramática de restrições (*constraint grammar* – CG), que faz a análise do texto morfológicamente (lexemas), de grupos de palavras e da composição da oração. Com isso, o programa obtém uma análise nos níveis ortográfico, semântico e sintático. Após a aplicação da identificação do léxico, o programa elimina as ambiguidades encontradas em cada palavra, por meio da aplicação de um conjunto de regras na sentença, identificando e eliminando possibilidades de formas sintáticas inexistentes (MAIA, 2008).

O *software* disponibiliza uma interface (figura 7) na internet, que permite uma diversidade de opções de consulta em várias formas de visualização, onde um usuário submete sentenças ou textos completos e recebe de volta os textos marcados.

---

<sup>10</sup> Disponível no endereço da Internet: <http://visl.sdu.dk/visl/>

Figura 7: interface do software *PALAVRAS*

The screenshot shows the web interface of the PALAVRAS software. The browser address bar displays `beta.visl.sdu.dk/visl/pt/parsing/automatic/parse.php`. The page title is "Portuguese -> Automatic parse -> Flat structure" under the heading "Visual Interactive Syntax Learning". The SYDDANSK UNIVERSITET logo is visible in the top left. A navigation menu includes "Skip | Games | Quizzes | Tools | Sentence Analysis | Corpus" and a search box. A "Log in / register or be a guest" link is in the top right.

The main content area is titled "Flat structure". It features a text input field containing the sentence: "O emprego do fertilizante de liberação lenta aumentou o diâmetro de caule". Below the input are "Go!" and "Reset" buttons. The "Parser:" dropdown is set to "Full morphosyntactic parse" and the "Visualization:" dropdown is set to "Default".

The output shows a flat structure parse of the sentence, with words and their morphosyntactic information:
 

- o [o] <artd> DET M S @>N
- emprego [emprego] <pos-soc> <ac> N M S @SUBJ>
- de [de] <sam-> PRP @N<
- o [o] <-sam> <artd> DET M S @>N
- fertilizante [fertilizante] <cm-chem> N M S @P<
- de [de] PRP @N<
- liberação [liberação] <act> N F S @P<
- lenta [lento] ADJ F S @N<
- aumentou [aumentar] <vt> <fmc> V PS 3S IND VFIN @FMV
- o [o] <artd> DET M S @>N
- diâmetro [diâmetro] N M S @<ACC
- de [de] PRP @N<
- caule [caule] N M S @P<
- .

A sidebar on the left contains navigation links:
 

- Portuguese VISL**: Overview, Credits, Info
- Sentence Analysis**: Pre-analyzed, Machine Analysis, Flat structure, Tree structure, Dependency, links, Complex interface, Upload interface, Remote interface
- Edutainment**: Games, Quizzes
- Corpora**: SDU corpus search, Floresta sintá(c)tica
- Dictionaries**

Fonte: site do software

A figura 7 mostra o resultado do *parsing* de um trecho de um dos resumos do corpus analisado, com o texto “O emprego do fertilizante de liberação lenta aumentou o diâmetro de caule”, submetido ao processamento morfossintático do processador *PALAVRAS*. É possível ver a análise realizada por meio da figura 8.

**Figura 8:** Resultado de um texto submetido ao processador PALAVRAS

```

o [o] <artd> DET M S @>N
emprego [emprego] <pos-soc> <ac> N M S @SUBJ>
de [de] <sam-> PRP @N<
o [o] <-sam> <artd> DET M S @>N
fertilizante [fertilizante] <cm-chem> N M S @P<
de [de] PRP @N<
liberação [liberação] <act> N F S @P<
lenta [lento] ADJ F S @N<
 aumentou [aumentar] <vt> <fmc> V PS 3S IND VFIN @FMV
o [o] <artd> DET M S @>N
diâmetro [diâmetro] N M S @<ACC
de [de] PRP @N<
caule [caule] N M S @P<

```

**Fonte:** elaborado pelo autor

Observa-se, na figura 8, que em cada linha do arquivo de saída aparecem a forma do lexema, tal qual ocorre no texto submetido. Em seguida, a forma canônica do lexema e, por fim, a classificação morfossintática. No exemplo ilustrado na figura 7, temos para o lexema “emprego” a forma canônica “emprego” e as classificações N (substantivo), M (masculino), S (singular). O analisador morfológico identifica palavras ou expressões isoladas em uma sentença, sendo este processo auxiliado por delimitadores (pontuação e espaços em branco). Os significados de todos os símbolos da classe gramatical das palavras encontram-se no Apêndice A desta Dissertação. Os dados foram retirados do site do programa. Para consultar o conjunto de símbolos completo do VISL, pode-se visitar a página do programa na Internet<sup>11</sup>.

A próxima etapa foi transferir os dados gerados pelo programa para o Excel. Para tanto foi realizado o seguinte procedimento: copiando o trecho da página HTML, contendo a análise morfossintática e, com o uso de um programa que automatiza a substituição das marcas HTML de fonte por “;”, foi possível construir, para cada resumo, um arquivo no formato .csv, que pode ser carregado no *Excel*, a fim de ser melhor visualizado e manipulado. O arquivo gerado para a mesma frase mencionada anteriormente é apresentado na figura 9.

<sup>11</sup> <http://visl.sdu.dk/visl/pt/info/symbolset-manual.html>

**Figura 9:** arquivo no formato .csv

```

o;o;<artd>;DET; M S ;@>N
emprego;emprego;<pos-soc> <ac>;N; M S ;@SUBJ>
de;de;<sam->;PRP; ;@N<
o;o;<-sam> <artd>;DET; M S ;@>N
fertilizante;fertilizante;<cm-chem>;N; M S ;@P<
de;de;;PRP; ;@N<
liberação;liberação;<act>;N; F S ;@P<
lenta;lento;;ADJ; F S ;@N<
aumentou;aumentar;<vt> <fmc>;V; PS 3S IND VFIN ;@FMV
o;o;<artd>;DET; M S ;@>N
diâmetro;diâmetro;;N; M S ;@<ACC
de;de;;PRP; ;@N<
caule;caule;;N; M S ;@P<

```

**Fonte:** elaborado pelo autor

A tabela 1 mostra o arquivo gerado carregado no *Excel*, onde cada coluna corresponde a um tipo de marcação que o *PALAVRAS* atribui às palavras da frase, e que são exibidas no cabeçalho da tabela, para fins de entendimento da estrutura do arquivo.

**Tabela 1:** marcação com o *software* gerada no *Excel*

<b>PALAVRA</b>	<b>LEMA</b>	<b>MAR- CA- ÇÃO SE- MÂN- TICA</b>	<b>CLASSE GRAMATICAL</b>	<b>MARCA- ÇÃO MORFO- LÓGICA</b>	<b>MARCAÇÃO SINTÁTICA</b>
o	o	<artd>	DET	M S	@>N
emprego	emprego	<pos- soc> <ac>	N	M S	@SUBJ>
de	de	<sam->	PRP		@N<
o	o	<-sam> <artd>	DET	M S	@>N
fertilizante	fertilizante	<cm- chem>	N	M S	@P<
de	de		PRP		@N<
liberação	liberação	<act>	N	F S	@P<
lenta	lento		ADJ	F S	@N<
aumentou	aumentar	<vt> <fmc>	V	PS 3S IND VFIN	@FMV

o	o	<artd>	DET	M S	@>N
diâmetro	diâmetro		N	M S	@<ACC
de	de		PRP		@N<
caule	caule		N	M S	@P<

Fonte: elaborado pelo autor

Lema é a forma canônica da palavra. A lematização é o ato de representar as palavras através do infinitivo dos verbos e masculino singular dos substantivos e adjetivos.

#### 3.2.2.4 *Análise Semântica- processo manual*

Para essa avaliação, tomamos o corpus construído com 60 resumos de artigos científicos da área de Ciências Agrárias. Os 60 resumos foram analisados individualmente, localizando e extraíndo as relações de causa e efeito presentes em sentenças distintas de cada resumo. No *Word*, utilizando a ferramenta *cor da fonte*, foram marcados da cor azul a expressão de causa, da cor vermelha o verbo que faz a ligação entre a causa e o efeito, e da cor verde a expressão que designa o efeito, conforme figura 10. Para facilitar o trabalho de análise, todas as sentenças marcadas foram transferidas para o *Excel*, onde foram separadas por colunas (causa, verbo e efeito). Em seguida foram listados todos os verbos (na forma canônica) de acordo com sua frequência de ocorrência.

Figura 10: marcação da relação causa e efeito

Disponibilidade de nutrientes e crescimento de porta-enxertos de citros fertilizados com fertilizantes convencionais e de liberação lenta.

Os fertilizantes de liberação lenta são usados para reduzir a frequência das adubações por disponibilizar gradualmente os nutrientes às plantas. Entretanto, a liberação de nutrientes deve coincidir com a demanda nutricional das plantas. O objetivo deste trabalho foi avaliar a disponibilidade de N, P e K e o crescimento de dois porta-enxertos de citros quando fertilizados com fertilizantes de liberação lenta e fertilizantes convencionais. Foram avaliados dois manejos de adubação em substrato comercial (com fertilizantes de liberação lenta e somente com fertilizantes convencionais), dois porta-enxertos (limoeiro 'Cravo' e tangerineira 'Sunki') e cinco épocas de avaliação (30; 60; 90; 120 e 180 dias após o replantio das mudas), com cinco repetições e duas plantas por parcela. O emprego do fertilizante de liberação lenta aumentou o diâmetro de caule, número de folhas, teores de P nas folhas e disponibilidade de P e K no substrato, quando comparado ao emprego da fertilização convencional.

Fonte: elaborado pelo autor

### ***3.2.2.5 Análise Semântica- processo automático***

Para extrair as sentenças de causa e efeito automaticamente foram utilizados os seguintes procedimentos: com os dados transferidos do *software PALAVRAS* para a planilha do *Excel*, foi possível realizar uma programação para localizar sentenças de causa e efeito automaticamente. A rotina para realizar esta tarefa, criada pelo Prof. Piotr Trzesniak em planilha do excel contendo a saída formatada do Parser *PALAVRAS* (Lema, Classe Gramatical e Marcação Morfológica) será apresentada na próxima seção. O objetivo foi comparar as sentenças identificadas diretamente pelo pesquisador e as sentenças reconstruídas automaticamente a partir de um conjunto de células programadas no *Excel*.

## 4 RESULTADOS E DISCUSSÃO

Nesta seção, busca-se consolidar o aprendizado decorrente da aplicação da metodologia ao *corpus* analisado.

### 4.1 Relacionamentos semânticos de causa e efeito

A análise semântica dos verbos se divide entre a semântica léxica, entendida como a análise das relações de significado entre unidades do componente léxico, e a semântica relacional, entendida como a análise das relações de significado entre os argumentos lógicos de um predicado e sua correlação com as relações sintáticas de dependência. Scheicher (2013) diz que um relacionamento semântico é todo tipo de associação que existe entre termos de um texto no nível de seus significados. Segundo o autor, dentre os diversos tipos de relacionamentos semânticos, os mais frequentemente considerados são os apresentados na tabela 2.

**Tabela 2:** Relacionamentos semânticos mais comuns

	Relacionamento Semântico	Sentença exemplo	Relacionamento extraído
1.	hiponímia/hiperonímia: <i>is-a</i> (subclasse, superclasse)	Maçã é uma fruta	<i>is-a</i> (maçã, fruta)
2.	meronímia/holonímia: <i>part-of</i> (todo, parte)	Parafuso é uma parte de uma máquina	<i>part-of</i> (máquina, parafuso)
3.	causa e efeito: <i>effect-of</i> (ação/estado, consequência)	Gripe causa febre	<i>effect-of</i> (gripe, febre)

Fonte: elaborado de acordo com Scheicher (2013)

Geralmente, um relacionamento é representado por um verbo ou uma expressão verbal e um ou mais sintagmas nominais, que são associados ao verbo por uma dada relação que imprime sentido à mensagem. Para Garcia (2004) os verbos causais são aqueles que expressam uma relação de causa-e-efeito entre dois elementos do domínio do verbo, essa relação é vista tipicamente como um acontecimento. Os verbos causais pressupõem um elemento no caso objetivo, o que faz com que eles admitam a voz passiva, são exemplos deles: causar, provocar, acarretar, ocasionar; originar; resultar em; derivar, provir, proceder, advir, resultar de; suscitar.

É evidente que não pode haver efeito sem causa. Assim, pode-se afirmar que todo verbo de causa pressupõe um elemento ou um fato responsável pelo evento referido por ele. No Exemplo

*Fertilizantes fornecem micronutrientes para o solo*, o verbo comanda uma função do tipo fornecer (fertilizantes, micronutrientes), que faz uma relação de um argumento agente com um argumento tema (o resultado), ao mesmo tempo que expressa um tipo de ação determinada. Do ponto de vista semântico, a formação de verbo de causa pode ser explicada usando o processo da fusão. A fusão, segundo Ortiz (2013), refere-se à condensação linguística de elementos em uma única forma e se manifesta de maneira individual e independente. A fusão incide sobre as consequências conceituais dessa união, já que as consequências morfossintáticas são a perda de independência dos elementos unidos e a transição de categoria substantivo para a categoria verbo. Para Maculan e Lima (2014), os fatores semânticos são de grande importância para a compreensão deste processo, uma vez que auxiliarão na determinação dos seus papéis temáticos de forma mais rigorosa.

Os relacionamentos que compõem a estrutura conceitual de uma busca inteligente podem ser mais claros e inteligíveis para humanos e máquinas quando explicitamente representados. De forma geral, essa representação é realizada através de verbos ou expressões verbais. Por essa razão, a Linguística forneceu importantes aportes teóricos. Considerando que o texto científico é a manifestação linguística de um grupo que espera ver-se como a ele pertencente, o que se constatou foi que a especialização do verbo em um domínio refletirá, não apenas nos aspectos semânticos, mas também, muitas vezes nos aspectos sintáticos. Cabe também ressaltar que esse tipo de amostra pode sofrer alterações e se estabelecer de maneiras diferentes, conforme as áreas estudadas e até mesmo conforme os gêneros textuais, isso sem contar as particularidades inerentes à línguas portuguesa.

#### 4.2 Verbos localizados nas sentenças de causa e efeito

Por meio da leitura dos 60 resumos, foram identificadas 54 sentenças que, de alguma forma, continham verbos que expressavam causalidade. Em seguida, foram listados todos os verbos de acordo com sua frequência de ocorrência (ver tabela 3). A referida tabela consiste de um conjunto de possíveis verbos de sentido causal, cujo sentido foi verificado em seu uso nos textos. Alguns desses verbos estão explicitamente se referindo a causalidade, porém a maioria deles são ambíguos, expressando uma relação de causa e efeito somente em um contexto particular.

**Tabela 3:** lista dos verbos encontrados nas sentenças marcadas

Verbos localizados nas sentenças			
8	aumentar	1	incrementar

4	apresentar	1	induzir
5	ser	1	Influenciar
4	proporcionar	1	mostrar
4	reduzir	1	possibilitar
3	Melhorar	1	produzir
3	Poder	1	promover
2	Afetar	1	propiciar
2	alterar	1	provoca
2	Causar	1	recomendar
2	diminuir	1	relacionar
2	limitar	1	aproximar
2	permitir	1	depende
1	elaborar	1	verificar
1	fornecer		

Fonte: elaborado pelo autor

Vale ressaltar que nas 5 vezes em que o verbo “Ir” apareceu nas sentenças ele não estava sozinho, estava acompanhado de complementos: *foi menor; foi observada; foi superior foi viável; foram obtidos*. Observa-se que dentro da área do conhecimento das Ciências Agrárias no que se trata do assunto aminoácidos e fertilizantes o verbo “aumentar” apareceu mais vezes na amostra estudada, como pode ser observado nos exemplos abaixo:

- *O amadurecimento do trato digestório, aumenta o aproveitamento proteico em aves que consomem milho e gérmen integral de milho.*
- *A utilização de fertilizantes nitrogenados de eficiência aumentada aumentou o suprimento de N para o milho aumentando ainda a produtividade de grãos.*
- *A utilização de doses crescentes de fertilizantes na cultura anterior (soja) aumentou a produção de biomassa seca, concentração e acúmulo de nutrientes nas plantas de milho.*

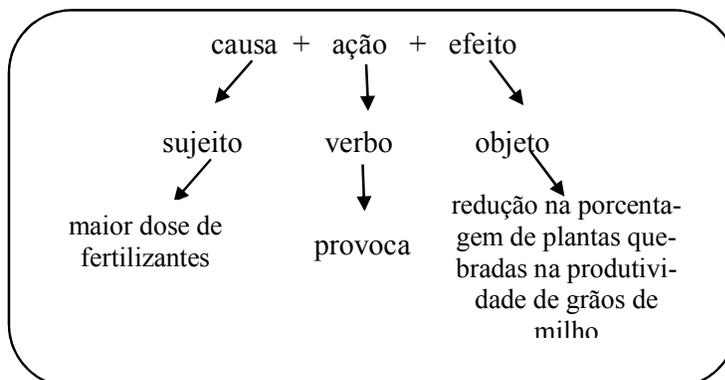
Assim, pode-se sugerir ao especialista do domínio, que pode existir uma relação de causa e efeito entre fertilizantes e produtividade de grãos de milho e de soja, entre o amadurecimento do trato digestório e o aproveitamento proteico. A partir do pressuposto, o especialista pode realizar testes que comprovem ou rejeitem tais afirmações.

### 5.3 Procedimentos para extração automática de sentenças de causa e efeito

A principal tarefa de um programa que recupera sentenças de causa e efeito é buscar os verbos que constituem a parte mais importante da sentença de causa e efeito. Uma vez que se encontra verbos na oração, se busca o sujeito e o objeto correspondente a cada verbo.

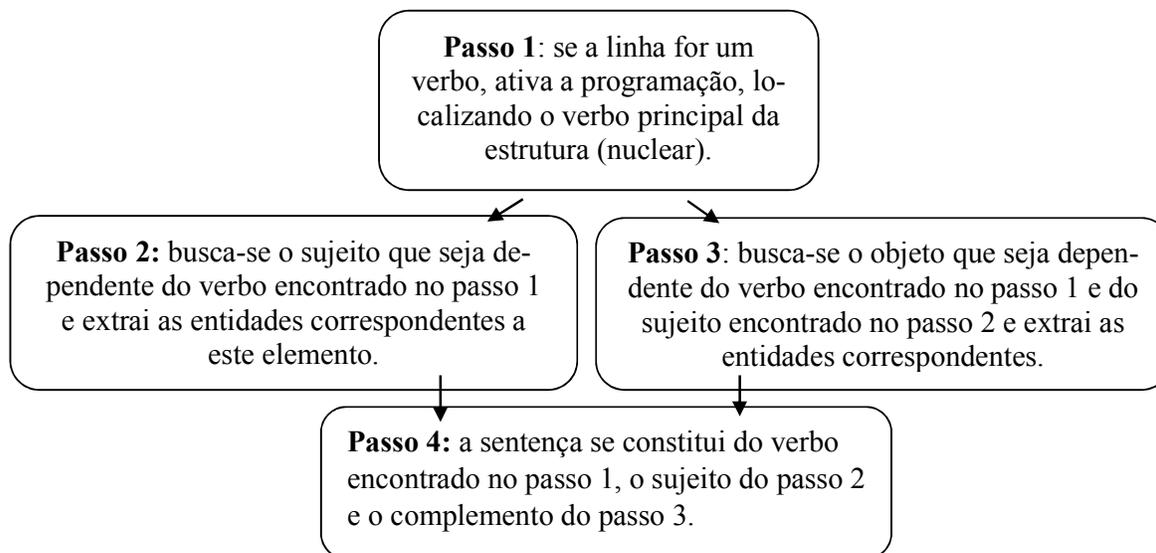
Exemplo de uma sentença de causa e efeito:

*Maior dose de fertilizantes provoca redução na porcentagem de plantas quebradas na produtividade de grãos de milho.*



De posse da lista dos verbos que exprimem relação de causa e efeito localizados na análise manual, a rotina para realizar a extração das sentenças de causa e efeito, criada pelo Prof. Piotr Trzesniak em planilha do excel contendo a saída formatada do Parser *PALAVRAS*, está ilustrada na figura 11 conforme os passos a seguir:

**Figura 11:** procedimentos das células programadas no *Excel*



**Fonte:** elaborado pelo autor

A descrição da rotina é dividida em três grupamentos:

## GRUPAMENTO SUJEITO + GRUPAMENTO AÇÃO + GRUPAMENTO OBJETO

*1- Grupamento de ação*

Para localizar uma ação de causa e efeito seis tipos de estrutura são possíveis:

- a) Verbo simples (nuclear). Exemplo: causar, provocar
- b) Verbos VIR, ESTAR, SER ou TER (VEST) + verbo nuclear (ex. Com a redução dos níveis de proteína na dieta, foi observada redução linear dos coeficientes de digestibilidade aparente da energia bruta).
- c) VEST + adverbio + verbo nuclear
- d) Verbos PODER ou DEVER + verbo nuclear
- e) Verbos PODER ou DEVER + SER + verbo nuclear (ex. FAI e a QA podem ser utilizados nas rações de aves em substituição ao milho)
- f) Verbos PODER ou DEVER+ adverbio+ SER+ verbo nuclear
- g) Verbos PODER ou DEVER+SER+ adverbio + verbo nuclear

*2- Grupamento do sujeito*

Para localizar o sujeito se parte do princípio que todo substantivo é potencialmente um sujeito. As seguintes atividades são realizadas neste passo:

- a) Localizar o substantivo, como todo substantivo é potencialmente um sujeito são destacados todos os substantivos;
- b) Partindo de cada substantivo antes do verbo (elementos pré-verbais), se avança no texto agrupando as palavras até:
  - (i) Encontrar um sinal de pontuação (vírgula, ponto final, interrogação, exclamação, ponto e vírgula), neste caso a busca é interrompida, e o sujeito foi encontrado contendo as palavras agrupadas;
  - (ii) Ao encontrar um verbo, forma-se o sujeito. Complementando com no máximo 15 palavras após o primeiro substantivo até chegar ao verbo.

*3- Grupamento objeto*

O princípio deste passo é que o objeto é o que vem após o verbo nuclear. As seguintes atividades localizam o objeto da frase:

- a) Localiza e reúne tudo que vem após o verbo nuclear até:
  - (i) Encontrar um ponto sinal de pontuação (vírgula, ponto final, interrogação, exclamação, ponto e vírgula);
  - (ii) Encontrar outro verbo
  - (iii) Completar um total de 8 palavras

#### 4- Sentença montada

Após os passos 1, 2 e 3 as sentenças são montadas. Se o núcleo verbal do agrupamento de ação for de causa-efeito, então o programa seleciona a sentença como de causa e efeito.

### 5.4 Comparação entre as sentenças identificadas manualmente e automaticamente

Os resultados, apresentados nesta seção, permitiram estabelecer algumas considerações sobre as extrações manual e automática em um *corpus* anteriormente processado de forma manual. Apresenta-se considerações de ordem qualitativa e quantitativa para tecer uma comparação entre os processos manual e automático de extração de sentenças de causa e efeito da área de Ciências Agrárias. Esses documentos foram previamente analisados de forma manual e as sentenças de causa e efeito foram extraídas.

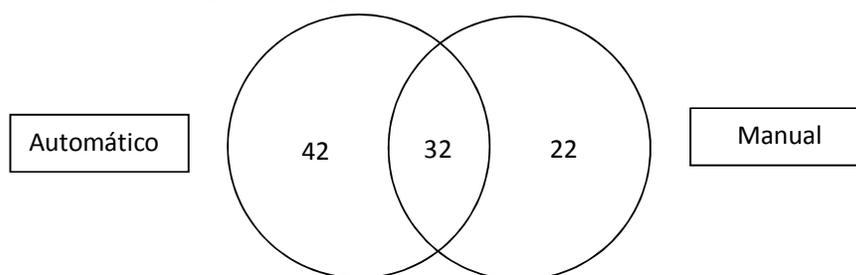
**Quadro 9:** quantidade de sentenças recuperadas

	Extração manual	Extração automática
Total de sentenças de causa e efeito identificadas	54	74

Fonte: elaborado pelo autor

Como pode ser observado no quadro 9, a extração de sentenças de causa e efeito, por meio do processo automático, retornou um número maior de sentenças. A figura 12 ilustra através do diagrama a relação dos dois resultados (processo manual e processo automático):

Figura 12: diagrama dos resultados quantitativos



Fonte: elaborado pelo autor

Em relação ao conteúdo das sentenças, observou-se que dentro das 74 sentenças identificadas pelo processo automático, 42 não foram identificadas pelo processo manual. Dentre essas 42 sentenças, 8 são sentenças que incluem o advérbio de negação “não”. A princípio, na busca manual, não foi levado em consideração esse tipo de sentença, por considerar que se está reportando ao “não acontecimento do efeito esperado”, porém, ao se fazer a análise dessas sentenças recuperadas pelo processo automático, foi possível perceber que as sentenças são de causa e efeito, independente se estão acompanhadas do advérbio de negação ligado ao verbo. Alguns exemplos de sentenças de causa-efeito com advérbio de negação são:

- *O aminoácido não alterou as variáveis analisadas.*
- *Aminoácidos essenciais não melhorou com a utilização de a Pbc como variável independente.*
- *A redução do teor de proteína bruta da ração não afetou o ganho de peso.*
- *O gesso não proporcionou alterações na produção de biomassa seca do milheto.*
- *A produtividade das culturas não foi alterada pela adição de qualquer dos fosfatos no primeiro ciclo de rotação.*

Em relação as outras sentenças que não foram encontradas no processo manual, é possível afirmar que houve uma desatenção humana do autor dessa Dissertação no momento da busca e identificação das sentenças de causa e efeito. Todas as sentenças localizadas pelo processo automático e manual encontram-se no Apêndice B desta Dissertação.

Ainda em relação ao conteúdo das sentenças, dentro das 54 sentenças localizadas no processo manual, 22 não foram encontradas pelo processo automático. Em relação às sentenças encontradas no processo manual que não foram encontradas no processo automático, possivelmente houve alguma falha que poderá ser revista em um esforço futuro de continuidade desta

pesquisa. Entretanto, podemos identificar alguns problemas específicos do processo, que podem ter influenciado na não localização das sentenças:

- Falhas do processador *PALAVRAS*, na identificação errônea de palavras ou sinais especiais de formatação (ex. números seguidos por um ponto, números romanos, títulos de seções do texto sem pontuação final, abreviaturas, sinais gráficos como \$, &, etc.);
- Sujeitos e objetos muito longos nas sentenças (mais de 8 palavras);
- A vírgula, às vezes, separa sentenças e, às vezes, separa elementos de uma sentença.

Vale ressaltar, também, que um olhar mais atento e minucioso permitiu verificar que a identificação manual oferece tratamento melhor para a exploração das sentenças de causa e efeito presentes nas estruturas dos resumos. Porém, mesmo considerando os problemas apontados e a eficácia qualitativa, se compararmos as performances, levando em conta a velocidade relativa dos processos de extração e o grande percentual de sentenças extraídas corretamente pelo programa, consideramos que o pressuposto apresentado na introdução (estabelecer uma rotina computacional capaz de ler, interpretar e marcar textos científicos em português, de modo a possibilitar sua inclusão em buscas semânticas inteligentes) se verificou.

Como pode ser visto, a possibilidade de usar técnicas automáticas acelera o processo de criação e extração de relações de causa e efeito. A principal contribuição desta dissertação consiste do próprio método proposto para extração de relacionamentos do tipo causa e efeito, tanto nos textos relacionados com a agronomia quanto em textos relacionados com outros assuntos. Foi gerado um modelo inicial de busca pelos relacionamentos de causa e efeito, que demonstra a grande utilidade deste artefato na geração de novas hipóteses de pesquisa sobre o assunto fertilizantes e aminoácidos por pesquisadores especialistas das Ciências Agrárias.

## 5 CONSIDERAÇÕES FINAIS

O enorme volume de informações disponíveis (e em constante crescimento) torna difícil não apenas recuperá-las, mas, também, gerenciar a recuperação e seus resultados de modo inteligente. Por essa razão, se faz necessário buscar respaldo teórico e prático no intuito de conhecer os recursos eficazes para organizar e dar acesso ao conhecimento. Nesse sentido, o Observatório Temático e Laboratório – Ensino, Tecnologia, Ciência e Informação (OtletCI) funciona como facilitador do processo de desenvolvimento de competências, métodos e técnicas para a realização de monitoramento e elaboração de diagnósticos. É no âmbito do OtletCI que essa pesquisa se desenvolveu, com o objetivo de avançar na questão de como extrair informação relevante e de como representá-la para fins de recuperação semântica da informação, em particular no caso de textos de publicações científicas em português.

A necessidade de explicitação sobre as técnicas para aumentar a eficiência dos mecanismos de busca, por meio da utilização de linguagens que permitam definir dados e regras para o raciocínio sobre esses dados, motivou a proposta de estabelecer uma rotina computacional capaz de ler, interpretar e marcar textos científicos em português, de modo a possibilitar sua inclusão em buscas semânticas inteligentes.

A presente pesquisa partiu do estudo sobre a busca semântica para ampliar a compreensão da natureza da mesma e para colaborar metodologicamente (na fase experimental) para o desenvolvimento de mecanismos de busca de informação científica e tecnológica, cuja temática é de interesse para a CI. Tradicionalmente, esta área de conhecimento se ocupa de técnicas de automação da informação documentária, porém, a complexidade associada a linguagem natural exige uma atenção especial quando se trata de recuperar a informação textual que satisfaça a necessidade de informação de um usuário, como verificado na literatura. É nesse cenário que a CI tem se aproximado dos estudos linguísticos, em especial os estudos semânticos, como consequência da necessidade de analisar o conteúdo de textos para deles se extrair elementos que permitam sua eficaz recuperação.

Para tanto, investigaram-se os fundamentos teóricos e a aplicação da tecnologia da busca semântica como recurso na recuperação precisa e eficaz da informação; com embasamento na epistemologia da Ciência da Informação; com aportes da linguística computacional, Processamento de Linguagem Natural e das tecnologias oriundas da Web Semântica e mineração de texto. Uma forma de conhecer uma aplicação da busca semântica foi por meio do software Goldfire. Por meio da análise da produção científica sobre o software foi possível conhecer suas vantagens e benefícios, observou-se ainda os experimentos práticos para os quais o Goldfire foi utilizado, enfatizando como os pesquisadores têm usado o software em termos de ferramenta para resolver problemas, ou seja, em quais aplicações este tipo de tecnologia foi utilizada.

Portanto, mais que identificar e discorrer o que vem sendo estudado pelos pesquisadores sobre o software Goldfire, o presente estudo evidenciou a tecnologia de indexação e busca semântica, ressaltando o poder do software de entender sentidos, de identificar relações de causa-efeito, de comparar e fornecer vantagens e desvantagens das várias soluções de um problema específico, de localizar os experts num dado tema, enfim, de ir além dos inputs, de fato, acrescentando inteligência semântica às palavras ingressadas na consulta.

Das contribuições da revisão bibliográfica desta pesquisa, podemos citar as que diretamente identificamos: a) a linguagem natural precisa ser representada de tal maneira que possa ser posteriormente manipulada e extraída por processos automatizados, o que exige que a mesma seja convertida em alguma estrutura lógica. b) esse tipo de organização permite que ferramentas de busca por informação sejam capazes de fazer a seleção e a filtragem dessa informação, baseadas na semântica dos termos de busca e dos itens pesquisados. c) a busca semântica permite que o conhecimento seja organizado em universos conceituais, de acordo com seu significado. d) a extração e representação computacional de relações de causa e efeito, pelo maior grau de informação semântica embutida, podem vir a se tornar mais eficaz do que as palavras-chave usualmente extraídas e utilizadas como descritores em outros processos automatizados de representação de documentos.

Continuando com os resultados deste trabalho, retornar-se-á às questões que encadearam a pesquisa. Inicialmente, as questões que instigaram a pesquisa indagavam sobre a possibilidade de enumerar e identificar vários tipos de relações semânticas no corpus analisado, evidenciando fatos que estão expressos nos textos. Considerando-se que existem vários tipos de relações semânticas, podemos inferir a impossibilidade de estudá-las a priori, devido ao curto período de tempo. Mesmo com esta limitação, neste trabalho, chegou-se à seguinte sistematização: trabalhando diretamente com resumos da área de Ciências Agrárias, localizam-se sentenças que envolvam relações de causa-efeito. Por outro lado, empregando recursos computacionais de identificação morfológica e sintática, decompõem-se e se recompõem os textos, igualmente destacando-se sentenças que se presume atendam as mesmas condições (relações de causa-efeito).

O estudo apresentou, por meio de considerações de ordem qualitativa e quantitativa, uma comparação entre os processos manual e automático de extração de sentenças de causa e efeito. Para essa avaliação, tomamos o corpus construído com 60 resumos de artigos científicos da área de Ciências Agrárias. Esses documentos foram previamente analisados de forma manual e as sentenças de causa e efeito foram extraídas.

Embora se tenha constituído a partir de algumas contribuições, o presente trabalho pode ser considerado útil, na medida em que abre caminho para aperfeiçoamento constante de metodologias de extração de relações semânticas de causa e efeito. A principal contribuição destacou-se com o próprio método proposto para extração de relacionamentos do tipo causa e efeito, tanto nos textos relacionados com a agronomia quanto em textos relacionados com outros assuntos. Foi gerado um modelo inicial de busca pelos relacionamentos de causa e efeito, que demonstra a grande utilidade deste artefato na geração de novas hipóteses de pesquisa sobre o assunto "fertilizantes e aminoácidos" por pesquisadores especialistas das Ciências Agrárias.

A possibilidade de usar técnicas automáticas acelera o processo de criação e extração de relações de causa e efeito e pode ser usada como alternativa ao processo custoso de identificação manual de informações semânticas. Dessa forma, busca-se superar o gargalo existente devido a grande demanda por dados semânticos, e a escassez de tal conhecimento e de mão de obra qualificada e disponível para gerá-lo em tempo hábil.

Afinal, conclui-se que mais importante que propor uma estrutura de relações de causa e efeito para a construção de sistemas de busca, o que podemos apontar como o resultado mais expressivo da presente pesquisa é o estabelecimento preliminar de rotinas para a versão automatizada. Espera-se, com esta pesquisa então, viabilizar a identificação de indícios de novas descobertas científicas no âmbito do Observatório Temático e Laboratório – Ensino, Tecnologia, Ciência e Informação (OtletCI).

Para trabalhos futuros, nesse contexto, podemos indicar: a) verificação da aplicação das relações semânticas extraídas de artigos científicos de outras áreas do conhecimento. b) verificação da aplicação de outros tipos de relações semânticas extraídas de artigos científicos como: agente-paciente, vantagem-desvantagem, etc. c) aprofundar a pesquisa em semântica e linguística, de forma a explicitar sua inserção em Ciência da Informação, em especial para o desenvolvimento sistemas de busca semântico. d) resolver as fontes de divergências do programa utilizado no processo automático. e) desenvolver um sistema que realize a extração de relações semânticas automaticamente, a partir de textos escritos em língua portuguesa.

## REFERÊNCIAS

ALMEIDA, C. C. de. **Elementos de Linguística e Semiologia na Organização da Informação**. 1.ed. São Paulo: Cultura Acadêmica, 2011.

ARMADA, R. N. **Um ambiente de apoio a externalização e mapeamento do conhecimento em equipes distribuídas de trabalho**. 2005. 187f. Dissertação (Mestrado em Informática) – Programa de Pós-Graduação em Informática, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2005. Disponível em: <[http://www.nce.ufrj.br/ginape/publicacoes/dissertacoes/d\\_2005/d\\_2005\\_rafael\\_nunes\\_armada.pdf](http://www.nce.ufrj.br/ginape/publicacoes/dissertacoes/d_2005/d_2005_rafael_nunes_armada.pdf)>. Acesso em: 05 set. 2016.

BAKKER, H.M., CHECHURIN, L.S., WITS, W.W. Integrating TRIZ function modeling in CAD software. In: TRIZ FEST CONFERENCE, 2011, Russia. **Anais...** Russia: University of Twenty, 2011. Disponível em: <[http://doc.utwente.nl/93359/1/paper\\_HB\\_LCH\\_WW\\_final.pdf](http://doc.utwente.nl/93359/1/paper_HB_LCH_WW_final.pdf)>. Acesso em: 05 set. 2016.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic *web*. **Scientific American**, San Francisco, EUA, p. 28-37, 2001.

BORGES, G. S. B. **Indexação automática de documentos textuais: critérios essenciais**. 2009. 111 f. Dissertação (Mestrado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2009.

BRANCO, A. et al. **A Língua Portuguesa na era digital / The Portuguese Language in the Digital Age**. White Paper Series, Berlin: Springer-Verlag, 2012. 85p.

BRÄSCHER, M. A Ambigüidade na Recuperação da Informação. **DataGramaZero: revista de Ciência da Informação**, Rio de Janeiro, v. 3, n. 1, p. 1-27, fev. 2002. Disponível em: <[http://repositorio.unb.br/bitstream/10482/924/2/ARTIGO\\_AmbiguidadeRecuperacaoInformacao.pdf](http://repositorio.unb.br/bitstream/10482/924/2/ARTIGO_AmbiguidadeRecuperacaoInformacao.pdf)>. Acesso em: 08 dez. 2015.

BUSH, V. As We May Think. **The Atlantic**, v.176, n.1, p.101-108, jul., 1945. Disponível em: <<http://www.theatlantic.com/doc/194507/bush>>. Acesso em: 30 jun. 2015.

BUSOV, B.; BARTLOVA, M. Improvement of active hinge of the car bonnet. In: MECHANICS-MECHATRONIKA (ME)- INTERNATIONAL CONFERENCE, 16, 2014, Brno. **Anais...** Brno: IEEE, 2014. p. 575-581. Disponível em: <[http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=7018323&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs\\_all.jsp%3Farnumber%3D7018323](http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=7018323&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs_all.jsp%3Farnumber%3D7018323)>. Acesso em: 05 nov. 2015.

CAMILO, C. O.; DA SILVA, J. C. **Um estudo sobre a interação entre Mineração de Dados e Ontologias**. Goiás: Instituto de Informática da Universidade Federal de Goiás, 2009. 15p. (Relatório técnico, RT-INF 002-09). Disponível em: <[http://www.portal.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF\\_002-09.pdf](http://www.portal.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_002-09.pdf)>. Acesso em: 30 jul. 2015.

CANONGIA, C. et al. Integração entre inteligência competitiva, gestão do conhecimento e visão de futuro: reflexão sobre um sistema de prospecção tecnológica e do conhecimento para o setor de ciência, tecnologia e inovação brasileiro. In: WORKSHOPBRASILEIRO DE INTELIGÊNCIA COMPETITIVA DE GESTÃO DO CONHECIMENTO, 3., 2002, São Paulo. **Anais...** São Paulo: IBICT, 2002.

CAPURSO, D.; MAGRI, E. **MODELLI DI RAPPRESENTAZIONE DELLA CONOSCENZA A SUPPORTO DEL BENCHMARKING TECNOLOGICO**. 2009. 162f. Dissertação (Mestrado em Engenharia Mecânica) - Facoltà di Ingegneria, Università Degli Studi Di Bergamo, Bergamo, 2009. Disponível em: <[http://www.ideam.it/upload/proposte/tesi/2011.05.23-10.05.24\\_tesi\\_6.pdf](http://www.ideam.it/upload/proposte/tesi/2011.05.23-10.05.24_tesi_6.pdf)>. Acesso em: 03 set. 2015.

CARVALHO, M. A. **Metodologia IDEATRIZ para a Ideação de Novos Produtos**. 2007. 232 f. Tese (Doutorado em Engenharia da Produção) – Programa de Pós-Graduação em Engenharia da Produção, Universidade Federal de Santa Catarina, Florianópolis, 2007. Disponível em: <<http://www.tede.ufsc.br/teses/PEPS5195-T.pdf>>. Acesso em: 10 jul. 2015.

CASTELLANOS, O. F. et al. Tendencias en investigación, desarrollo Tecnológico y mercado en la cadena productiva de cacao-chocolate. In: CASTELLANOS, O.; L. TORRES; S. FONSECA; MONTAÑES. V. **Estudios de vigilancia tecnológica aplicados a cadenas productivas del sector agropecuario colombiano: agendas prospectivas**. Bogotá: Giro Editores Ltda, 2008. P. 29-65. Disponível em: <<https://issuu.com/biogestion/docs/cacao-chocolate/151>>. Acesso em: 05 set. 2015.

CHAUÍ, M. Convite à filosofia. 13. ed. São Paulo: Ed. Ática, 2003. 424 p.

CHUKSIN, P. Selecting of key problems and solution search area in forecasting. In: TRIZ FUTURE CONFERENCE, 16., 2004, Florence. **Anais...** Florence: Firenze University Press, 2004. p. 247-252. Disponível em: <<http://www.torrossa.it/resources/an/2251149#page=259>>. Acesso em: 05 set. 2016.

CORDEIRO, P. V. M. **Avaliação de conhecimentos explicitados em patentes para levantamento de indícios de possíveis parcerias empresariais**. 2010. 136 f. Dissertação (Mestrado em Tecnologia) - Programa de Pós-graduação em Tecnologia, Universidade Tecnológica Federal do Paraná, Curitiba, 2010.

DAVIES, J.; STUDER, R; WARREN, P. **Semantic Web technologies: trends and research in ontology-based systems**. England: John Wiley & Sons, 2006. 327p.

DI FELIPPO, A. **Delimitação e alinhamento de conceitos lexicalizados no inglês norte-americano e no português brasileiro**. 2008. 237 f. Tese (Doutorado em Letras) - Faculdade de Ciências e Letras de Araraquara, Universidade Estadual Paulista, Araraquara, 2008. Disponível em: <<http://hdl.handle.net/11449/103583>>. Acesso em: 30 jun. 2015.

DIAS-DA-SILVA, B. C. **A face tecnológica dos estudos da linguagem: o processamento automático das línguas naturais**. 1996. 272 f. Tese (Doutorado em Letras) - Faculdade de Ciências e Letras, Universidade Estadual Paulista, Araraquara, 1996.

DIAS-DA-SILVA, B. C. O estudo linguístico-computacional da linguagem. **Letras de Hoje**, Porto Alegre, v. 41, n. 2, p. 103-138, 2006.

DIAS-DA-SILVA, B. C. et al. **Introdução ao Processamento das Línguas Naturais e algumas aplicações**. São Paulo, 2007, 121p. (Série de Relatórios do Núcleo Interinstitucional de Linguística Computacional NILC-TR-07-10). Disponível em: <<http://www.lettras.etc.br/ebraic/NILCTR0710-DiasDaSilvaEtAl.pdf>>. Acesso em: 03 nov. 2015.

DIAS, T. D.; SANTOS, N. Web Semântica: Conceitos Básicos e Tecnologias Associadas. **Cadernos do IME-Série Informática**, v. 14, p.80-92, 2013. Disponível em: <<http://www.e-publicacoes.uerj.br/ojs/index.php/cadinf/article/viewFile/6619/4734>>. Acesso em: 30 nov. 2015.

DIAS, E. A. V.; COSTA, H. G. Mapeamento da produção científica no escopo da Ontologia. **Sistemas & Gestão**, v. 6, n. 4, p. 481-507, 2012. Disponível em: <<http://www.revis-tasg.uff.br/index.php/sg/article/download/V6N4A6/V6N4A6>>. Acesso em: 30 jun. 2015.

DI FELIPPO, A.; DIAS-DA-SILVA, B. C. O processamento automático de línguas naturais enquanto engenharia do conhecimento linguístico. **Calidoscópico**, v. 7, n. 3, p. 183- 191, set./dez. 2009.

DZIEKANIAK, G.; PACHECO, R.; KERN, V. M. Revisitando a organização do conhecimento através dos tesouros, folksonomias e ontologias: aportes da engenharia do conhecimento. In: SILVA, F. C. C. da; SALES, R. de. (Org.). **Cenários da organização do conhecimento: linguagens documentárias em cena**. Brasília: Thesaurus, 2011, p. 201-237.

EBECKEN, N. F. F.; LOPES, M. C. S.; COSTA, M. C. A. Mineração de Textos. In: REZENDE, S. O. **Sistemas Inteligentes: fundamentos e aplicações**. Barueri, SP: Manole, 2003. P. 307-333.

ENJUTO, N. Razón de ser de los Observatorios. In: **OBSERVATORIO DEL VOLUNTARIADO**. Observando Observatorios: ¿Nuevos agentes en el Tercer Sector? Madrid: Plataforma del Voluntariado de España, 2010. p. 10-17. Disponível em: <<http://www.plataformavoluntariado.org/descargaRecursos>>. Acesso em: 12 jul. 2015.

FACHIN, G. R. B. Recuperação inteligente da informação e ontologias: um levantamento na área da Ciência da Informação. **BIBLOS**, Rio Grande, v. 23, n. 1, p. 259-283, mar. 2010. Disponível em: <<http://www.seer.furg.br/biblos/article/view/1282/576>>. Acesso em: 17 jan. 2016.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI magazine**, v. 17, n. 3, p. 37, 1996. Disponível em:

<<http://www.aaai.org/ojs/index.php/aimagazine/article/viewArticle/1230>>. Acesso em: 12 jul. 2015.

FU, K. Design-by-analogy: experimental evaluation of a functional analogy search methodology for concept generation improvement. **Research in Engineering Design**, v. 26, n. 1, p. 77-95, 2015. Disponível em: <[http://scholar.google.com/scholar?cites=2563217979092148580&as\\_sdt=2005&sci-odt=0,5&hl=en&num=20](http://scholar.google.com/scholar?cites=2563217979092148580&as_sdt=2005&sci-odt=0,5&hl=en&num=20)>. Acesso em: 05 set. 2015.

FU, K. et al. Investigating the Effect of Functionality Level of Analogical Stimulation on Design Outcomes. In: KOREA-JAPAN DESIGN ENGINEERING WORKSHOPS (DEWS), 13, 2013, Japan. **Anais...** Japan: JSME DSD, 2013. p. 28-30. Disponível em: <[https://www.researchgate.net/profile/Kristin\\_Wood/publication/270183800\\_Investigating\\_the\\_Effect\\_of\\_Functionality\\_Level\\_of\\_Analogical\\_Stimulation\\_on\\_Design\\_Outcomes/links/54a258470cf267bdb902cf3d.pdf](https://www.researchgate.net/profile/Kristin_Wood/publication/270183800_Investigating_the_Effect_of_Functionality_Level_of_Analogical_Stimulation_on_Design_Outcomes/links/54a258470cf267bdb902cf3d.pdf)>. Acesso em: 05 set. 2015.

GARCIA, A. UMA TIPOLOGIA SEMÂNTICA DO VERBO NO PORTUGUÊS. **Revista SOLETRAS**, Rio de Janeiro, n. 8, p. 52-70, Dez. 2012. Disponível em: <<http://www.e-publicacoes.uerj.br/index.php/soletras/article/view/4527>>. Acesso em: 12 Fev. 2016..

GARCÍA, F. S. **Sistema Basado en Tecnologías del Conocimiento para Entornos de Servicios Web Semánticos**. 2008. 419f. Tese (Doutorado em Engenharia da Informação e das Comunicações)- Departamento de Ingeniería de la Información y las Comunicaciones, Universidad de Murcia, España, 2008. Disponível em: <<https://digitum.um.es/xmlui/handle/10201/154>>. Acesso em: 25 abr. 2015.

GERHARDT, T. E.; SILVEIRA, D. T. **Métodos de pesquisa**. Porto Alegre: Editora da UFRGS, 2009.

GOMILA, J. F. O acesso transversal ao conhecimento no processo da pesquisa médica: diagnósticos e biomarcadores no câncer da próstata. In: COLÓQUIO INTERNACIONAL - A MEDICINA NA ERA DA INFORMAÇÃO – MEDINFOR, 3., 2014, Bahia. **Anais...** Bahia: UFBA, 2014. Disponível em: <[https://www.researchgate.net/publication/267151749\\_O\\_acesso\\_transversal\\_ao\\_conhecimento\\_no\\_processo\\_da\\_pesquisa\\_medica\\_diagnosticos\\_e\\_biomarcadores\\_no\\_cancer\\_da\\_prostata\\_The\\_cross\\_access\\_to\\_knowledge\\_in\\_the\\_case\\_of\\_medical\\_research\\_diagnostic\\_and\\_bio](https://www.researchgate.net/publication/267151749_O_acesso_transversal_ao_conhecimento_no_processo_da_pesquisa_medica_diagnosticos_e_biomarcadores_no_cancer_da_prostata_The_cross_access_to_knowledge_in_the_case_of_medical_research_diagnostic_and_bio)>. Acesso em: 05 set. 2015.

GUSMÃO, M. R. P. D. Estruturas e dispositivos nacionais de produção e difusão de indicadores de C,T&I: deficiências e possíveis avanços. In: CONFERÊNCIA NACIONAL DE CIÊNCIA, TECNOLOGIA E INOVAÇÃO PARA O DESENVOLVIMENTO, 3., 2005, Brasília. **Anais...** Brasília: MCT, 2005. p.1-25.

GUSMÃO, M. R. P. D; RAMOS, M. Y. Concentração regional da C&T no Brasil: perfil da liderança paulista no cenário nacional. **São Paulo em Perspectiva**, São Paulo, v. 20, n. 3, p. 120-141, jul./set. 2006.

HEARST, M. **What Is Text Mining?** Berkeley: UC Berkeley, 2003. Disponível em: <<http://people.ischool.berkeley.edu/~hearst/text-mining.html>>. Acesso em: 13 maio. 2015.

HEQUET, M. Management of technology. **Training**, Minneapolis, v. 28, n. 4, p.61, abr. 1991. Disponível em <<http://go.galegroup.com/ps/i.do?id=GALE%7CA10656136&v=2.1&u=capes&it=r&p=AONE&sw=w&asid=3d26a28ed3f6128db5d7ceefae46751f>>. Acesso em: 13 maio. 2015.

HUDON, M. Multilingual thesaurus construction: integrating the views of different cultures in one gateway to knowledge and concepts. **Knowledge Organization**, Würzburg, v. 24, n. 2, p. 84-91, 1997.

IGNÁCIO, S. E. O processo da derivação frasal nas frases dinâmicas do português escrito contemporâneo do Brasil. **ALFA: Revista de Linguística**, São Paulo, v. 38, p. 155-163 1994. Disponível em: <<http://hdl.handle.net/11449/107715>>. Acesso em: 12 out. 2015.

KEENAN, M. A prospecção tecnológica no Reino Unido. **Parcerias Estratégicas**, Brasília, v. 11, n. 23, p. 127-152, 2006.

LADEIRA, A. P. **Processamento de linguagem natural [manuscrito]**: caracterização da produção científica dos pesquisadores brasileiros. 2010. 259 f. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2010.

LIMA, V. L. S.; NUNES, M. das G. V.; VIEIRA, R. Desafios do processamento de línguas naturais. In: SEMISH-SEMINÁRIO INTEGRADO DE SOFTWARE E HARDWARE, 34., 2007, Rio de Janeiro. **Anais...** Rio de Janeiro: SBC, 2007. p. 2202-2216.

LNĚNIČKA, L. **Tvorba inovačních zadání a jejich řešení**. 2012. 58f. Dissertação (Mestrado em Engenharia Elétrica) - Fakulta elektrotechniky a komunikačních technologií, Vysoké učení technické v Brně, Brně, 2012.

LÓPEZ ISAZA, G. A.; CORREA VALLEJO, M. J. Sources of organizational information and intelligence in research: The case of Universidad Tecnológica de Pereira. **Cuadernos de Administración**, Bogotá, v. 24, n. 42, p. 231-252, 2011. Disponível em: <[http://www.scielo.org.co/scielo.php?pid=S0120-35922011000100011&script=sci\\_arttext&tlng=en](http://www.scielo.org.co/scielo.php?pid=S0120-35922011000100011&script=sci_arttext&tlng=en)>. Acesso em: 05 set. 2015.

LOPEZ, J.; ALMEIDA, R. L. de; ARAUJO-MOREIRA, F. M. TRIZ: criatividade como uma

ciência exata? **Revista Brasileira de Ensino de Física**, São Paulo, v. 27, n. 2, p. 205-209, 2005. Disponível em: <<http://www.scielo.br/pdf/rbef/v27n2/a04v27n2.pdf>>. Acesso em: 25 Jan. 2016.

LUGER, S. Lighting Helmet for Formula 1. In: TRIZ FUTURE CONFERENCE, 16., 2004, Florence. **Anais...** Florence: Firenze University Press, 2004. p. 29-38. Disponível em: <<http://www.triz-austria.com/files/LUGER%20RESEARCH%20-%20Lighting%20Helmet%20for%20Formula%201%20-%20Siegfried%20Luger%20-%20ETRIA%202004.pdf>>. Acesso em: 04 set. 2015.

MACHADO, A. P. et al. Mineração de texto em Redes Sociais aplicada à Educação a Distância. **Colabor@** - A Revista Digital da CVA-RICESU, v. 6, n. 23, 2010. Disponível em: <<http://pead.ucpel.tche.br/revistas/index.php/colabora/article/viewFile/132/115>>. Acesso em: 12 jul. 2015.

MACULAN, B. C.; LIMA, G. A. B. O. Relacionamentos em tesauros: o valor semântico dos verbos. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 19, n. 4, p. 182-201, 2014.

MANGOLD, C. A survey and classification of semantic search approaches. **International Journal of Metadata, Semantics and Ontologies**, v. 2, n. 1, p. 23-34, 2007. Disponível em: <[ftp://inf.informatik.uni-stuttgart.de/pub/library/ncstrl.ustuttgart\\_fi/ART-2007-09/ART-2007-09.pdf](ftp://inf.informatik.uni-stuttgart.de/pub/library/ncstrl.ustuttgart_fi/ART-2007-09/ART-2007-09.pdf)>. Acesso em: 12 nov. 2015.

MARCONDES, C. H. O papel das relações semânticas na Organização e Representação do Conhecimento em ambientes digitais. In: SILVA, F. C. C. da; SALES, R. de. (Org.). **Cenários da organização do conhecimento: linguagens documentárias em cena**. Brasília: Thesaurus, 2011, p. 129-168.

MARTINS, A. L. **O uso do sintagma nominal na recuperação de documentos**: proposta de um mecanismo automático para a classificação temática de textos digitais. 2014. 192 f. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2014.

MATTELART, A. **História da sociedade da informação**. São Paulo: Loyola, 2002. 200p.

MÉNDEZ, J. G. C.; CÁRDENAS, J. C. R.; CAMELO, C. L. L. Informe de vigilancia tecnológica em mejoramiento genético forestal. In: CASTELLANOS, O.; L. TORRES; S. FONSECA; MONTAÑES. V. **Estudios de vigilancia tecnológica aplicados a cadenas productivas del sector agropecuario colombiano: agendas prospectivas**. Bogotá: Giro Editores Ltda, 2008. p. 151- 202. Disponível em: <<https://issuu.com/biogestion/docs/cacao-chocolate/151>>. Acesso em: 05 set. 2015.

MISHRA, U. 50 Useful Software on Innovation, Concept Mapping and Idea Management. **TRIZsite Journal**, P. 1-8, 2011. Disponível em: <[http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1976625](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1976625)>. Acesso em: 05 set. 2015.

MIR-MAURI, M., CASADESUS-FA, M. UNE 166002: 2006: ESTANDARIZAR Y SISTEMATIZAR LA I+ D+ I LA NORMA Y LA IMPORTANCIA DE LAS TIC EN SU IMPLEMENTACIÓN. **DYNA-Ingeniería e Industria**, v. 83. n.6, p. 325-331, 2008. Disponível em: <<http://www.revistadyna.com/Documentos/pdfs/%5C200806sep%5C1479DYNAIN-DEX.pdf>>. Acesso em: 05 set. 2015.

MORAIS, E. A. M.; AMBRÓSIO, A. P. L. Mineração de textos. Goiás: Instituto de Informática da Universidade Federal de Goiás, 2007. 30p. (Relatório Técnico, RT- INF 005-07). Disponível em: <[http://www.portal.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF\\_005-07.pdf](http://www.portal.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_005-07.pdf)>. Acesso em: 12 jun. 2015.

MOREIRO GONZÁLEZ, J. A. **Linguagens documentárias e vocabulários semânticos para a web**: elementos conceituais. Salvador: EDUFA, 2011.

OLIVEIRA, F. A. D. **Processamento de linguagem natural: princípios básicos e a implementação de um analisador sintático de sentenças da língua portuguesa**. Porto Alegre: Universidade Federal do Rio Grande do Sul, 2009.

OLIVEIRA NETO, J. M.; TONIN, S. D.; PIETRICH, S. S. Processamento de linguagem natural e suas aplicações computacionais. In: ESCOLA REGIONAL DE INFORMÁTICA, 2., 2010, Manaus. **Anais...** Manaus: INPA, 2010. p. 1-10. Disponível em: <<https://www.inpa.gov.br/erin2010/Artigo/Artigo9.pdf>>. Acesso em: 05 out. 2015.

ORTEGA, C.; DEL VALLE, R. S. S. Nuevos retos de los observatorios culturales. **Boletín GC: Gestión Cultural**, n. 19, p.1-15, 2010. Disponível em: <<http://www.gestioncultural.org/boletin/files/bgc19-CortegaRSansalvador.pdf>>. Acesso em: 15 out. 2015.

ORTIZ, R. M. L. Los verbos denominales de cambio de estado experiencial: consideraciones sintáctico-semánticas. **Revista internacional de lingüística iberoamericana**, n. 22, p. 29-46, 2013. Disponível em: <<https://dialnet.unirioja.es/servlet/articulo?codigo=4610920>>. Acesso em: 23. Nov. 2015.

PIATETSKY-SHAPIRO, G.; PARKER, G. Lesson: Data Mining, and Knowledge Discovery: An Introduction. **Introduction to Data Mining**, 2011. Disponível em: <[http://www.kdnuggets.com/data\\_mining\\_course/xIntro-to-data-minin-notes.html](http://www.kdnuggets.com/data_mining_course/xIntro-to-data-minin-notes.html)>. Acesso em: 15 maio. 2015.

PORTER, A. et al. **La inteligencia competitiva: factor clave para la toma de decisiones estratégicas en las organizaciones**. España: Editorial Comunidad de Madrid, 2007.

PORTER, A. L. Tech mining for future oriented technology analysis. **Text Mining: A. U. M. Project. Futures Research Methods. V. 2**, p. 1-21, 2009. Disponível em: <<https://www.thevanagepoint.com/resources/articles/FRM%20chapter-TextMining-2009mar3.pdf>> Acesso em: 05 abr. 2015.

PORTER, A. L.; DETAMPEL, M. J. Technology Opportunities Analysis, Technological Forecasting and Social Change. **Elsevier**, v.49, n. 3, p. 237-255, 1995. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_nlinks&ref=000172&pid=S0100-1965199900030001000002&lng=en](http://www.scielo.br/scielo.php?script=sci_nlinks&ref=000172&pid=S0100-1965199900030001000002&lng=en)>. Acesso em: 8 jun. 2015.

PORTER, A.; CUNNINGHAM, S. W. **Tech Mining: Exploiting New technologies for Competitive Advantage**. New Jersey: John Wiley & Sons, 2005.

RAMALHO, R. A. S.; OUCHI, M. T. Tecnologias Semânticas: Novas Perspectivas para a Representação de Recursos Informacionais. **Informação & Informação**, v. 16, n. 3, p. 60-75, 2011. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/9829/10643>>. Acesso em: 12 jul. 2015.

RENEAR, A. H.; PALMER, C L. Strategic reading, ontologies, and the future of scientific publishing. **Science**, Nova Iorque, v. 325, n. 5942, p. 828-832, 2009.

REZENDE, S. O.; MARCACINI, R. M.; MOURA, M. F. O uso da Mineração de Textos para Extração e Organização Não Supervisionada de Conhecimento. **Revista de Sistemas de Informação da FSMA**, Rio de Janeiro, n. 7, p. 7-21, 2011. Disponível em: <[http://www.fsma.edu.br/si/edicao7/FSMA\\_SI\\_2011\\_1\\_Principal\\_3.pdf](http://www.fsma.edu.br/si/edicao7/FSMA_SI_2011_1_Principal_3.pdf)>. Acesso em: 10 maio. 2015.

RODRÍGUEZ PEROJO, K.; RONDA LEÓN, R. Web semántica: un nuevo enfoque para la organización y recuperación de información en el web. **Acimed**, Ciudad de La Habana, v. 13, n. 6, p. 1-56, 2005.

SÁNCHEZ-TORRES, J. M. Apoyo en la definición de políticas públicas en ciencia, tecnología e innovación através de vigilancia tecnológica e inteligencia competitiva. In: CASTELLANOS, O.; L. TORRES; S. FONSECA; MONTAÑES. V. **Estudios de vigilancia tecnológica aplicados a cadenas productivas del sector agropecuario colombiano: agendas prospectivas**. Bogotá: Giro Editores Ltda, 2008. p. 13-28. Disponível em: <<https://issuu.com/biogestion/docs/cacao-chocolate/151>>. Acesso em: 05 set. 2015.

SANCHEZ, J. M.; MEDINA, J. E.; LEON, A. M. Publicación internacional de patentes por organizaciones inventores de origen colombiano. **Cuadernos de Economía**, Bogotá, v. 26, n. 47, p. 1-24, Dec. 2007. Disponível em: <[http://www.scielo.org.co/scielo.php?pid=S0121-47722007000200010&script=sci\\_arttext&tlng=en](http://www.scielo.org.co/scielo.php?pid=S0121-47722007000200010&script=sci_arttext&tlng=en)>. Acesso em: 05 set. 2016.

SANTOS, R. N. M. dos. **Formulação de diretrizes para concepção e implantação de um observatório na UFPE** (Projeto de pesquisa). Recife: Departamento de Ciência da Informação/UFPE, 2013.

SCHIESSL, M.; BRÄSCHER, M. Do texto às ontologias: uma perspectiva para a ciência da informação. **Ciência da Informação**, Brasília, v. 40, n. 2, p. 301-311, 2012. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/1947/1427>>. Acesso em: 12 jul. 2015.

SI, N.; YU, P.; ZHANG, L. Semantic SOA based integrated development environment for complex information systems. In: SYSTEM SIMULATION AND SCIENTIFIC COMPUTING. ASIA SIMULATION INTERNATIONAL CONFERENCE, 7., 2008, Beijing. **Anais...** Beijing: IEEE, 2008. p. 1521-1526. Disponível em: <[http://scholar.google.com/scholar?cites=12429780024365329850&as\\_sdt=2005&scioldt=0,5&hl=en&num=20](http://scholar.google.com/scholar?cites=12429780024365329850&as_sdt=2005&scioldt=0,5&hl=en&num=20)>. Acesso em: 05 set. 2015.

ŠIC, P. **Dimenzování proudovodné dráhy elektrických přístrojů**. 2014. 59f. Dissertação (Mestrado em Engenharia Elétrica) - Fakulta elektrotechniky a komunikačních technologií, Vysoké učení technické v Brně, Brně, 2014. Disponível em: <<https://dspace.vutbr.cz/handle/11012/33591>>. Acesso em: 05 out. 2015.

SILVA et al. OBSERVATÓRIOS DE INFORMAÇÃO E CONHECIMENTO: DISCUTINDO BASES CONCEITUAIS E PERSPECTIVAS DE EFETIVIDADE. In: CONGRESSO NACIONAL DE EXCELÊNCIA EM GESTÃO, 9, 2013, Rio de Janeiro. **Anais...** Rio de Janeiro: CNEG, 2013, p. 1-22. Disponível em: <<http://www.inovarse.org/node/964>>. Acesso em: 13 jun. 2015.

SILVA, E. M. **Descoberta de conhecimento com o uso de Text Mining**: Cruzando o abismo de moore. 2002. 174 f. Dissertação (Mestrado em Gestão do Conhecimento e da Tecnologia da Informação)- Programa de Pós-Graduação *Stricto Sensu* em Informática, Universidade Católica de Brasília, Brasília, 2002. Disponível em: <[http://edilms.eti.br/uploads/file/publicacoes/dissertacao\\_edilberto\\_textmining.pdf](http://edilms.eti.br/uploads/file/publicacoes/dissertacao_edilberto_textmining.pdf)> Acesso em: 02 maio. 2015.

SILVA, F. C. C.; SALES, R. de. **Cenários da Organização do Conhecimento**: linguagens documentárias em cena. Brasília: Thesaurus, 2011. 296p.

SILVA, F.; SILVA, R.; Abordagens da representação da informação hoje. In: SILVA, F. C. C. da; SALES, R. de. (Org.). **Cenários da organização do conhecimento**: linguagens documentárias em cena. Brasília: Thesaurus, 2011. p. 45-70.

DA SILVA, Augusto Soares. Verbos y construcciones causativas analíticas en portugués y en español. **Estudios de lingüística**, n. 1, p. 581-598, 2004. Disponível em: <<http://rua.ua.es/dspace/handle/10045/9799>> Acesso em: 12 fev 2016t.

SIQUEIRA, I. C. P. Mecanismos de busca na *web*: passado, presente e futuro. **Ponto de Acesso**, Salvador, v. 7, n. 2, p. 47-67, 2013. Disponível em: <<http://www.portalseer.ufba.br/index.php/revistaici/article/view/6355/6136>>. Acesso em: 12 jul. 2015.

SOSISKA-KALATA, B. Semantization and standardization – cooperative or conflicting trends in knowledge organization? **Knowledge organization**, Würzburg, v.14, n. 2, p. 580, 2014.

SOUCHKOV, V. Innovative Enterprise Infrastructure. In: TRIZ FUTURE CONFERENCE, 16., 2004, Florence. **Anais...** Florence: Firenze University Press, 2004. p. 227-236. Disponível em: <<http://www.triz-austria.com/files/LUGER%20RESEARCH%20-%20Lighting%20Helmet%20for%20Formula%201%20-%20Siegfried%20Luger%20-%20ETRIA%202004.pdf>>. Acesso em: 04 set. 2015.

SOUZA, E. N. P. **Classificação de Relações Semânticas Abertas Baseada em Similaridade de Estruturas Gramaticais na Língua Portuguesa**. 2014. 89f. Dissertação (Mestrado em Ciência da Informação) - Programa de Pós-Graduação em Ciência da Informação, Universidade Federal da Bahia, Salvador, 2014.

SOUZA, R. R.; ALVARENGA, Lídia. A *Web Semântica* e suas contribuições para a ciência da informação. **Ciência da Informação**, Brasília, v. 33, n. 1, p. 132-141, 2004. Disponível em: <<http://revista.ibict.br/cienciadainformacao/index.php/ciinf/article/view/50/49>>. Acesso em: 12 jul. 2015.

SOUZA, R. R. Sistemas de Recuperação de Informação e Mecanismos de Busca na web: panorama atual e tendências. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.11, n. 2, p. 161-173, maio/ago. 2006.

SI, N.; YU, P.; ZHANG, L. Semantic SOA based integrated development environment for complex information systems. In: SYSTEM SIMULATION AND SCIENTIFIC COMPUTING. ASIA SIMULATION INTERNATIONAL CONFERENCE, 7., 2008, Beijing. **Anais...** Beijing: IEEE, 2008. p. 1521-1526. Disponível em: <[http://scholar.google.com/scholar?cites=12429780024365329850&as\\_sdt=2005&sciodt=0,5&hl=en&num=20](http://scholar.google.com/scholar?cites=12429780024365329850&as_sdt=2005&sciodt=0,5&hl=en&num=20)>. Acesso em: 05 set. 2015.

TODHUNTER, J; SOVPEL, I.; PASTANOHAU, D. **Question-answering system and method based on semantic labeling of text documents and user questions**. U.S. Patent n. 8,666,730, 4 mar. 2014.

TRZECIAK, D.S. **Modelo de observatório para arranjos produtivos locais**. 2009. 236 f. Tese (Doutorado em Engenharia de Produção) – Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, 2009.

TRZESNIAK, P. A questão do livre acesso aos artigos publicados em periódicos científicos. **Em Aberto**, Brasília, v. 25, n. 87, p. 77-112, jan./jun. 2012. Disponível em: <<http://emaberto.inep.gov.br/index.php/emaberto/article/view/2364/2326>>. Acesso em: 02 jul. 2015.

TRZESNIAK, P. OtletCI, um sistema informacional de alto desempenho (projeto de pesquisa). Departamento de Ciência da Informação, Universidade Federal de Pernambuco (2015).

TRZESNIAK, P.; SANTOS, R. N. M. dos. Um observatório para diagnóstico e fomento da pesquisa na América Latina e Caribe (projeto de pesquisa) Departamento de Ciência da Informação, Universidade Federal de Pernambuco (2014).

VASCONCELOS, L. M. R.; CARVALHO, C. L. Aplicação de Regras de Associação para Mineração de Dados na *Web*. Goiás: Instituto de Informática da Universidade Federal de Goiás, 2004. p. 11-14. (Relatório técnico, RT-INF 004-04). Disponível em: <[http://www.portal.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF\\_004-04.pdf](http://www.portal.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_004-04.pdf)>. Acesso em: 12 jul. 2015.

VERBITSKY, M. Semantic TRIZ. **The TRIZ journal**, p. 1-14, fev. 2004. Disponível em: <<http://www.triz-journal.com/archives/2004/>>. Acesso em: 15 jun. 2015.

VERGARA, S. C. **Métodos de pesquisa em administração**. São Paulo: Atlas, 2007.

VIEIRA, R.; LIMA, V. L. S. **Linguística computacional: princípios e aplicações**. In: CONGRESSO DA SBC - I JORNADA DE ATUALIZAÇÃO EM INTELIGÊNCIA ARTIFICIAL, 21., 2001, Fortaleza. **Anais...** Fortaleza: SBC, 2001. p. 47-86.

WEISS, L.C. **Relações semânticas em tesouros: um estudo da abordagem Pragmática**. 2014. 166 f. Dissertação (Mestrado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação, Universidade Federal de Santa Catarina, Florianópolis, 2014. Disponível em: <<https://repositorio.ufsc.br/xmlui/bitstream/handle/123456789/128984/331626.pdf?sequence=1&isAllowed=y>>. Acesso em: 12 jul. 2015.

YANG, Y. et al. Text mining and visualization tools—impressions of emerging capabilities. **World Patent Information**, Princeton, NJ, v. 30, n. 4, p. 280-293, 2008. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0172219008000094>>. Acesso em: 14 maio. 2015.

YU, ABRAHAM et al. Observatório de tecnologia e inovação no IPT: sua evolução e aprendizado. **Parcerias Estratégicas**, Brasília, n. 23, p. 95-126, Dez. 2006.

ZHANG, Y. et al. How to combine term clumping and technology roadmapping for newly emerging science & technology competitive intelligence: "problem & solution" pattern based

semantic TRIZ tool and case study. **Scientometrics**, Netherlands, v. 101, n.2, p. 1375-1389, 2014. Disponível em: <<http://link.springer.com/article/10.1007/s11192-014-1262-2#page-1>>. Acesso em: 05 set. 2015.

ZHANG, Y. et al. Triple Helix innovation in China's dye-sensitized solar cell industry: hybrid methods with semantic TRIZ and technology roadmapping. **Scientometrics**, Netherlands, v. 99, n. 1, p. 55-75, 2014. Disponível em: <<http://link.springer.com/article/10.1007/s11192-013-1090-9>>. Acesso em: 05 set. 2015.

**APÊNDICE A - CLASSE DE PALAVRAS (FORMA MORFOLÓGICA).....113**

Símbolo		Categoria
N		substantivo
PROP		nome próprio
ADJ		adjetivo
V	VFIN	verbo finito
	INF	infinitivo
	PCP	particípio
	GER	gerúndio
DET <artd>		artigo
PRON	PERS	pronome pessoal
	DET	pronome determinativo
	SPEC	pronome independente
ADV		advérbio
NUM		numeral
PRP		preposição
IN		interjeição
conj	KS	conjunção subordinativa
	KC	conjunção coordenativa
pu (unused)		pontuação

**APÊNDICE B – SENTENÇAS DE CAUSA E EFEITO RECUPERADAS PELO PROCESSO AUTOMÁTICO E PELO PROCESSO MANUAL.....114**

**Verbos considerados com característicos de causa/efeito para esta recuperação:**

acarretar	acrescentar	advir	alterar	aumentar
causar	diminuir	fornecer	incrementar	induzir
limitar	mediar	melhorar	ocasionar	originar
permitir	possibilitar	prejudicar	produzir	promover
propiciar	proporcionar	provir	provocar	reduzir
resultar	sofrer			

Processo automático		Processo manual	
RES01	154		
RES02	177		
RES02	241	Res02	A utilização de fertilizantes nitrogenados de eficiência aumentada não apenas aumentou mas também melhorou o suprimento de N para o milho aumentando ainda a produtividade de grãos
RES02	249	Res02	A utilização de fertilizantes nitrogenados de eficiência aumentada não apenas aumentou mas também melhorou o suprimento de N para o milho aumentando ainda a produtividade de grãos
RES03	124		
RES03	171	Res03	Medidas de TGA/DTA possibilitaram a confirmação das composições e dos teores de íons de nitrato intercalados.

RES05	25	plantas de cobertura podem <i>proporcionar</i> maior ciclagem de nutrientes .		
RES05	225	o calcário <i>proporcionou</i> incrementos em a produção de biomassa seca e em o acúmulo de nutrientes ( n		
RES05	260	o gesso não <i>proporcionou</i> alterações em a produção de biomassa seca de o milho .		
RES05	286	a utilização de doses crescentes de fertilizantes em a cultura anterior ( soja ) <i>aumentou</i> a produção de biomassa seca , concentração e acúmulo de nutrientes em as plantas de	Res05	A utilização de doses crescentes de fertilizantes na cultura anterior (soja) aumentou a produção de biomassa seca, concentração e acúmulo de nutrientes nas plantas de milho
RES05	334	o efeito residual de o fertilizante aplicado a a cultura de verão , <i>proporcionou</i> ciclagem de nutrientes que	Res05	O emprego do milho como planta de cobertura, na entressafra, utilizando o efeito residual do fertilizante aplicado à cultura de verão, proporcionou ciclagem de nutrientes que pode ser utilizada pela cultura em sucessão
RES06	44	frequentemente empregada em a citricultura com o intuito de <i>reduzir</i> os custos de as aplicações .		
			Res06	A aplicação dos fertilizantes foliares cloreto de zinco, cloreto de manganês, ureia e a mistura de fosfite de potássio + ureia + cloreto de zinco não afetaram a ação dos acaricidas cyhexatin, propargite e acrinathrin sobre o controle de B. phoenicis
RES06	284	e de manganês com o sulfato de magnésio e a adição de fosfite ALT xxxito de potássio <i>diminuíram</i> a eficiência de os acaricidas propargite ALT xxxite e acrinathrin ALT xxx , não		
			RES7	A interação entre (salinidade x manejo) e a salinidade isolada influenciou significativamente os componentes estudados, exceto para fitomassa seca do caule e raiz na interação.

RES08	251	os fertilizantes minerais com p <b>proporcionaram</b> maior produção de matéria seca de a parte aérea ( MSPA ) e de raízes		
RES08	305	por areia foi a que <b>proporcionou</b> os maiores valores de MSPA e de MSRA .		
RES09	27	os fertilizantes de liberação lenta são usados para <b>reduzir</b> a frequência ALT frequência de as adubações por		
RES09	172	o emprego de o fertilizante de liberação lenta <b>aumentou</b> o diâmetro de caule , número de folhas , teores de p em as folhas	Res09	O emprego do fertilizante de liberação lenta aumentou o diâmetro de caule, número de folhas, teores de P nas folhas e disponibilidade de P e K no substrato
RES10	27	em os solos , <b>fornecem</b> micronutrientes como impurezas residuais de a sua fabricação , os quais	Res10	Fertilizantes e corretivos, ao serem aplicados nos solos, fornecem micronutrientes como impurezas residuais da sua fabricação
RES11	40	a a entrada , a concentração observada e a alta de preços de os fertilizantes <b>induziram</b> a a desconfiança de possibilidade de exercício de o poder de mercado por parte de as empresas	Res11	As barreiras à entrada, a concentração observada e a alta de preços dos fertilizantes induziram à desconfiança de possibilidade de exercício do poder de mercado por parte das empresas representativas da indústria brasileira de fertilizantes NPK
RES12	221	maior comprimento de brotações e mudas mais altas as formulações 1 e 2 <b>proporcionaram</b> mudas maiores a os 120 dias , para as variedades estudadas as variedades Ascolano 315		
RES13	340	significativamente de a testemunha , todos os produtos <b>proporcionaram</b> incremento em o teor de sólidos solúveis .		
			RES13	Maiores quantidades de frutos por planta foram obtidas com fosfito de potássio, seguido pelo gesso agrícola e CPAC-GE

RES14	220	com relação a as doses de adubação de pré-plantio que <b>propiciaram</b> maior número e massa de frutos por planta ,		
RES15	170	à exceção de a cevada , a produtividade de as culturas não foi <b>alterada</b> por a adição de qualquer de os fosfatos em o primeiro ciclo de rotação ,		
RES15	241	dm-3 ALT xxx de p , a aplicação de p em a semeadura de a aveia branca <b>incrementou</b> em 11 % a produtividade de as culturas de verão e em 20 % a		
RES15	292	em este ciclo , o SFT <b>proporcionou</b> maior produtividade do que os FNR em as culturas de inverno ( aveia branca , trigo		
			RES16	O tratamento composto por Biofert® acrescido de polpa de banana apresentou os melhores resultados para o desenvolvimento in vitro do híbrido
RES17	17	a aplicação de fertilizantes organominerais ALT xxxais ALT xxxs tem <b>permitido</b> respostas positivas em diversas olerícolas ALT xxxícolas .	Res17	A aplicação de fertilizantes organominerais tem permitido respostas positivas em diversas olerícolas
			RES18	A fertilização com nitrogênio (N) foi viável economicamente, principalmente em doses de, aproximadamente, 200 kg de N/ha/ano, quando houve fornecimento de concentrados e 400 kg de N/ha/ano, quando não ocorreu essa suplementação
RES19	168	se que a dosagem de 40 mL L-1 de Biofertilizante <b>propiciou</b> melhores resultados , com a utilização de AIB .	Res19	A dosagem de 40 mL L-1 de Biofertilizante propiciou melhores resultados, com a utilização de AIB.
RES20	206	o etefon ALT xxx <b>proporcionou</b> melhor perfilhamento , mas a resposta		

RES20	241	a aplicação de Stimulate® e fertilizantes líquidos não <b>proporcionou</b> efeitos em a qualidade de a cana-de-açúcar .	Res20	A aplicação de Stimulate® e fertilizantes líquidos não proporcionou efeitos na qualidade da cana-de-açúcar.
RES22	33	intensamente em a agricultura , pois a baixa disponibilidade de p frequentemente <b>limita</b> o rendimento de as culturas em as condições brasileiras .	Res20	A baixa disponibilidade de P frequentemente limita o rendimento das culturas nas condições brasileiras
RES22	200	a aplicação de fosfato de Gafsa <b>proporcionou</b> as maiores concentrações de pb em a parte aérea de o milho em o primeiro		
			RES22	Houve aumento da produtividade de colmos e de açúcar, independente do genótipo, com o emprego do biorregulador Stimulate®, com ou sem complementação de fertilizante líquido
			RES23	o FERTCALC-Abacaxi constitui-se numa alternativa mais versátil para recomendar doses de fertilizantes e corretivos para o abacaxizeiro
RES24	175	a ureia ( 5 g L-1 ) <b>proporcionou</b> efeitos benéficos somente em a avaliação a os 28 dias após a aplicação .		
RES24	200	em laboratório , o aumento de a concentração de glifosato <b>promoveu</b> gradativa acidificação de a calda de pulverização , com estabilização de o pH de a		
RES24	223	, com estabilização de o pH de a solução em 4,5. o sulfato de amônio <b>causou</b> pequena acidificação de a calda herbicida , enquanto a ureia não	Res24	O sulfato de amônio causou pequena acidificação da calda herbicida, enquanto a ureia não alterou o pH
RES24	235	pequena acidificação de a calda herbicida , enquanto a ureia não <b>alterou</b> o pH .	Res24	O sulfato de amônio causou pequena acidificação da calda herbicida, enquanto a ureia não alterou o pH
RES24	248	o uso de o pulverizador pressurizado por CO2 pouco <b>alterou</b> o pH de a calda de glifosato .		

			Res24	A maior eficácia do glifosato após a adição de fertilizantes nitrogenados à calda está pouco relacionada com alterações no pH da solução
			Res 25	A uniformidade transversal da distribuição de fertilizantes e corretivos a lanço tem grande dependência da qualidade e da condição física do produto a ser aplicado
RES26	260	a adoção de a maior dose de fertilizantes não <b>proporciona</b> incrementos significativos em a produtividade de grãos de milho , porém		
RES26	272	incrementos significativos em a produtividade de grãos de milho , porém <b>provoca</b> redução em a porcentagem de plantasacamadas ALT xxxadas e	Res26	Maior dose de fertilizantes não proporciona incrementos significativos na produtividade de grãos de milho, porém provoca redução na porcentagem de plantas camadas e quebradas.
RES26	288	a densidade de 55000 plantas ha-1 ALT xxx <b>proporciona</b> a maior produtividade de grãos .		
			Res28	A formatação de arranjos do tipo bundling pode ser motivada por razões de eficiência, sendo a economia em custos de transação uma dessas razões.
			Res28	A aversão ao risco e a confiança motivam o uso do arranjo troca.
RES29	169	que a adição de os fertilizantes a o solo <b>causou</b> uma pequena elevação de os teores de k , u e th e	Res29	A adição dos fertilizantes ao solo causou uma pequena elevação dos teores de K, U e Th e diminuiu cerca de dez vezes os valores de resistividade elétrica
RES29	183	uma pequena elevação de os teores de k , u e th e <b>diminuiu</b> cerca de dez vezes os valores de resistividade elétrica .	Res29	A adição dos fertilizantes ao solo causou uma pequena elevação dos teores de K, U e Th e diminuiu cerca de dez vezes os valores de resistividade elétrica
RES30	134	os fertilizantes fosfatados com maiores teores de cd <b>aumentaram</b> a quantidade de esse metal acumulado em as plantas .	Res30	Os fertilizantes fosfatados com maiores teores de Cd aumentaram a quantidade desse metal acumulado nas plantas

			Res31	O perfil de aminoácidos obtido com a mistura de caseína e gelatina à proporção de 8:1 se aproxima do requerido para o crescimento de juvenis de <i>Piaractus brachypomus</i> .
			Res32	A suplementação com microminerais complexados a aminoácidos não influencia no desempenho das matrizes, variáveis de incubação e desempenho das progênes.
RES33	144	ração diário e a conversão alimentar , e o pH de o estômago também não <i>sofreu</i> influência , mas houve diminuição de o pH de o duodeno .		
RES33	194	, enquanto o peso de o pâncreas <i>diminuiu</i> com a redução de a proteína bruta de a dieta .	Res33	O peso do pâncreas diminuiu com a redução da proteína bruta da dieta
RES34	200	o nitrogênio <i>incrementou</i> linearmente as variáveis altura de plantas , altura de inserção de espiga , número de	Res34	O nitrogênio incrementou linearmente as variáveis altura de plantas, altura de inserção de espiga, número de grãos por espiga e número de grãos por fileira
			Res35	As variações na composição química das farinhas de penas e sangue podem estar relacionadas à falta de padronização no processamento a que são submetidas
RES36	274	o peso vivo e o ganho de peso , mas <i>alteram</i> o consumo de ração , a conversão alimentar e o custo de alimento .	Res36	O tipo sexual do macho (castrado ou inteiro) e a suplementação aminoácida não afetam o peso vivo e o ganho de peso, mas alteram o consumo de ração, a conversão alimentar e o custo de alimento.
RES37	248	Ronozyme Np ( m ) por tonelada de ração ) , o nível de fósforo disponível <i>reduziu</i> em 0.15 e o de cálcio em 0.3 ponto percentual em relação a o de a ração		
RES37	368	é possível <i>reduzir</i> em até três pontos percentuais o nível proteico ALT protéico de as rações , em 0.15 o	Res37	É possível reduzir em até três pontos percentuais o nível proteico das rações, em 0,15 o de fósforo disponível e em 0,30 ponto percentual o de cálcio quando as rações são suplementadas com fitase e aminoácidos.

			Res37	A excreção de nitrogênio e fósforo foi menor quando as aves receberam ração com níveis reduzidos desses nutrientes.
RES38	359	se que , com o amadurecimento de o trato digestório ALT xxxo , <b>aumenta</b> o aproveitamento proteico ALT protéico em aves que	Res38	O amadurecimento do trato digestório, aumenta o aproveitamento proteico em aves que consomem milho e gérmen integral de milho
RES39	238	de peso e eficiência de retenção de proteína em o músculo de os alevinos de tambatinga ALT xxxa <b>aumentaram</b> linearmente com a redução de proteína bruta em a dieta , enquanto o teor de	Res39	Os teores de proteína bruta, a proporção de proteína no ganho de peso e eficiência de retenção de proteína no músculo dos alevinos de tambatinga aumentaram linearmente com a redução de proteína bruta na dieta, enquanto o teor de extrato etéreo e sua proporção no ganho de peso diminuíram.
RES39	263	dieta , enquanto o teor de extrato etéreo e sua proporção em o ganho de peso <b>diminuíram</b> .	Res39	Os teores de proteína bruta, a proporção de proteína no ganho de peso e eficiência de retenção de proteína no músculo dos alevinos de tambatinga aumentaram linearmente com a redução de proteína bruta na dieta, enquanto o teor de extrato etéreo e sua proporção no ganho de peso diminuíram.
RES39	294	em dietas para alimentação de os alevinos de tambatinga ALT xxxa , pois não <b>prejudica</b> seu desempenho .		
RES41	306	é possível <b>reduzir</b> , de 26.74 para 24.53 % , a proteína digestível em dietas para tilápias-donilo em		
			Res41	Com a redução dos níveis de proteína na dieta, foi observada redução linear dos coeficientes de digestibilidade aparente da energia bruta, proteína bruta, extrato etéreo e fósforo das dietas.
RES42	146	a adição de fitase a as dietas <b>aumentou</b> em 2 % a dia de a arginina ALT xxxina , em 14 % a de o		

RES42	213	o fósforo fítico e as fibras , em as dietas , <b>reduzem</b> a dia de o cálcio , de o fósforo e de os aminoácidos essenciais .	Res42	O fósforo fítico e as fibras, nas dietas, reduzem a DLa do cálcio, do fósforo e dos aminoácidos essenciais
RES42	240	o uso de fitase e xilanase , em as dietas , <b>melhora</b> o aproveitamento de cálcio , fósforo e alguns aminoácidos .	Res42	O uso de fitase e xilanase, nas dietas, melhora o aproveitamento de cálcio, fósforo e alguns aminoácidos
RES42	262	no entanto , o excesso de cálcio e fósforo em as dietas <b>reduz</b> a ação de a fitase sobre a digestibilidade ileal de os nutrientes .		
			Res42	As concentrações de fósforo fítico e as frações fibra em detergente neutro, fibra em detergente ácido e lignina em detergente ácido, nas dietas, apresentaram correlações baixas e negativas com a DLa do cálcio, fósforo e aminoácidos
RES43	178	a variação de proteína corporal , a gordura e a variação de gordura corporal nem <b>alteraram</b> o intervalo desmame-estro ou a produção de leite de as porcas .		
			Res43	As relações dos aminoácidos metionina + cistina, treonina, triptofano e valina digestíveis com a lisina digestível propostas com base na proteína ideal atendem às exigências de porcas em lactação para melhores desempenhos produtivo e reprodutivo.
			Res44	Os valores obtidos dos coeficientes de digestibilidade verdadeira de aminoácidos essenciais e não-essenciais dos alimentos estudados permitem elaborar rações mais eficientes para aves.
RES46	272	a redução de os teores de aminoácidos em as dietas <b>proporciona</b> aumento de o marmoreio , independentemente de o genótipo .	Res46	A redução dos teores de aminoácidos nas dietas proporciona aumento do marmoreio, independentemente do genótipo.

RES47	109	se uma sub ALT xxx amostra aleatória dentro de cada uma de as categorias anteriores , que <b>resultou</b> em uma sub ALT xxx amostra de 15 milhos , os quais		
RES47	284	a predição de os aminoácidos essenciais não <b>melhorou</b> com a utilização de a PBC como variável independente .		
RES49	147	efeito quadrático de os níveis de PB sobre a conversão alimentar , que <b>melhorou</b> até o nível estimado de 15.7 % de proteína bruta .		
RES50	358	o aquecimento <b>reduziu</b> as concentrações de todos os aminoácidos em as 3 polpas .		
RES50	377	o meio fortemente alcalino ( pH 12 ) <b>produziu</b> a maior degradação de aminoácidos .	Res50	O meio fortemente alcalino (pH 12) produziu a maior degradação de aminoácidos.
RES51	22	a utilização de aminoácidos digestíveis e de a proteína ideal auxiliou a <b>reduzir</b> as excreções de nitrogênio para o ambiente e o custo de a ração formulada ,	Res51	A utilização de aminoácidos digestíveis e da proteína ideal auxiliou a reduzir as excreções de nitrogênio para o ambiente e o custo da ração formulada, sem afetar negativamente o desempenho dos frangos de corte.
RES52	165	o aumento de a temperatura de secagem <b>ocasionou</b> redução linear de a digestibilidade de metionina + cistina e isoleucina e efeito quadrático sobre		
RES52	276	a temperatura de secagem e o tempo de armazenamento de os grãos de milho <b>reduzem</b> a digestibilidade de a maioria de os aminoácidos .	Res52	A temperatura de secagem e o tempo de armazenamento dos grãos de milho reduzem a digestibilidade da maioria dos aminoácidos.
			Res53	FAI e a QA podem ser utilizados nas rações de aves em substituição ao milho, uma vez que tiveram níveis maiores de proteína bruta e aminoácidos digestíveis.

RES56	283	), assim como os critérios de formulação de ração ( aminoácidos totais ou digestíveis ) <b>proporcionam</b> desempenho semelhante em aves de reposição .		
RES57	312	de proteína bruta de a ração para alevinos de tilápia- de o nilo pode ser <b>reduzido</b> em quatro pontos percentuais ( de 32 para 28 % ) , pois essa redução não		
RES57	327	em quatro pontos percentuais ( de 32 para 28 % ) , pois essa redução não <b>prejudica</b> o desempenho de os animais , desde que as rações		
			Res57	Os peixes alimentados com rações com 27% de PB apresentaram pior conversão alimentar, menor eficiência de utilização da lisina para ganho, menor consumo de nitrogênio e umidade corporal, maior eficiência protéica para ganho e mais alto teor de gordura corporal
			Res58	Os coeficientes de digestibilidade ileal verdadeira de metionina, treonina, triptofano e serina determinados pelas perdas endógenas utilizando-se a dieta CHE foram superiores àqueles determinados utilizando a dieta DIP
RES59	291	a matéria seca ( MS ) e de a matéria orgânica ( me o ) <b>aumentou</b> em o experimento 2 , em virtude de a menor inclusão de farelo de soja .		
RES59	346	o desempenho , a ração de ADig <b>proporcionou</b> maior digestibilidade de a me o e tendência a a maior digestibilidade de a MS		
RES59	371	as rações com 17.5 % PB e ADig <b>promoveram</b> balanço mais positivo e maior retenção relativa de proteína em comparação a as de digestibilidade padrão	Res59	As rações com 17,5% PB e ADig promoveram balanço mais positivo e maior retenção relativa de proteína em comparação às de digestibilidade padrão

RES60	82	a fim de <i>provocar</i> redução de o estoque de glicogênio muscular e , por conseguinte ,		
-------	----	--	--	--