

Rudger Nowasky do Nascimento Taxweiler

**UM MODELO PARA A EXTRAÇÃO DE PERFIL DE
ESPECIALISTA APLICADO ÀS FERRAMENTAS DE
EXPERTISE LOCATION E APOIO À GESTÃO DO
CONHECIMENTO**

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

Orientador: Prof. Denilson Sell, Dr.
Coorientador: Prof. Paulo Maurício Selig, Dr.

Florianópolis
2016

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

Taxweiler, Ruder Nowasky do Nascimento

Um modelo para a extração de perfil de especialista aplicado às ferramentas de Expertise Location e apoio à Gestão do Conhecimento / Ruder Nowasky do Nascimento Taxweiler; orientador, Denilson Sell; coorientador, Paulo Maurício Selig - Florianópolis, SC, 2016.

171 p.

Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro Tecnológico. Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento.

Inclui referências

1. Engenharia e Gestão do Conhecimento. 2. Expertise Location. 3. Extração de perfis de especialistas. 4. Processamento de Linguagem Natural. 5. DBPedia. I. Sell, Denilson. II. Selig, Paulo Maurício. III. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento. IV. Título.

Rudger Nowasky do Nascimento Taxweiler

**UM MODELO PARA A EXTRAÇÃO DE PERFIL DE
ESPECIALISTA APLICADO ÀS FERRAMENTAS DE
EXPERTISE LOCATION E APOIO À GESTÃO DO
CONHECIMENTO**

Esta Dissertação foi julgada adequada para a obtenção do Título de Mestre em Engenharia e Gestão do Conhecimento e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento.

Florianópolis, 16 de março de 2016.

Prof. Roberto Carlos dos Santos Pacheco, Dr.
Coordenador do Curso

Banca Examinadora:

Prof. Denilson Sell, Dr.
Orientador
Universidade Federal de Santa Catarina

Prof. Paulo Maurício Selig, Dr.
Coorientador
Universidade Federal de Santa Catarina

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

Prof. Alexandre Leopoldo Gonçalves, Dr.
Universidade Federal de Santa Catarina

Prof. Jordan Paulesky Juliani, Dr.
Universidade do Estado de Santa Catarina

Este trabalho é dedicado aos meus professores e mestres, aos meus amigos e colegas, aos meus pais e irmãs, e, em especial, à minha querida esposa.

Uma boa educação não é a que enche sua cabeça de fatos, mas sim a que estimula sua curiosidade, porque, daí em diante, você continua aprendendo pelo resto da sua vida.

(Neil deGrasse Tyson, 2014)

RESUMO

As ferramentas de *Expertise Location* podem ser utilizadas amplamente na Gestão do Conhecimento para apoiar a identificação e o compartilhamento do conhecimento. Porém, manter os dados dos colaboradores de uma organização atualizados nessas ferramentas pode ser desafiador. Muitas vezes, os colaboradores precisam preencher as mesmas informações em diversos sistemas. Como uma abordagem alternativa para simplificar esse processo de atualização dos dados, este trabalho propõe um modelo para a extração automática de perfis de especialistas a partir de seus documentos não estruturados. Assim, realizou-se uma pesquisa aplicada e exploratória com base em uma revisão integrativa da literatura, a qual resultou na identificação das abordagens atuais para a extração de perfil de especialista que permitisse a construção desse modelo. A partir dessas abordagens, foram elaborados um modelo conceitual e um protótipo baseados em Processamento de Linguagem Natural para a tarefa de extração de informações de perfil de especialistas que possam fornecer insumos para a identificação de seus conhecimentos e de suas áreas de interesse. A implementação do protótipo resultou também em uma ferramenta de código aberto. Tal ferramenta é disponibilizada em um site público, em conjunto com o seu código-fonte, e gera uma página de perfil com o uso de componentes de *tag cloud* e *timeline*. Com o intuito de verificar a viabilidade do modelo proposto, a partir de documentos de voluntários, foram executados testes comparando os perfis gerados pela ferramenta com os perfis presentes na rede social LinkedIn. Os resultados dos testes demonstraram que o modelo proposto pode representar uma alternativa viável para a geração de perfis de especialistas de forma automática com o objetivo de apoiar as ferramentas de *Expertise Location* em uma organização. Conseqüentemente, a adoção desse modelo pode reduzir a necessidade de atualizações constantes dos perfis de especialistas de forma manual.

Palavras-chave: Expertise Location. Extração de perfis de especialistas. Processamento de Linguagem Natural. DBPedia.

ABSTRACT

The Expertise Location Tools can be widely used in Knowledge Management in order to support the identification and sharing of the knowledge. However, to keep the data of the employees of an organization updated in those tools can be challenging. From time to time, employees need to fill out the same data in different systems. As an alternative approach to simplify this process of updating the data, this paper proposes a model for the automatic extraction of expert profiles from their own non-structural documents. Thus, an applied and exploratory research based on an integrative literature review was carried out, resulting in the identification of the current approaches to the extraction of an expert profile that could allow the construction of this model. From these approaches were elaborated a conceptual model and a prototype based on Natural Language Processing for the task of extraction of information from expert profiles that could provide inputs to the identification of their expertise and their areas of interest. The prototype implementation has also resulted in an open source tool. This tool is available on a public website together with its source code and it generates a profile page using the tag cloud and timeline components. In order to verify the feasibility of the proposed model, tests from documents of volunteers were performed comparing the profiles generated by the tool with those profiles on LinkedIn social network. The test results demonstrated that the proposed model can represent a viable alternative to the generation of automatically expert profiles in order to support Expertise Location tools in an organization. Consequently, the adoption of this model can reduce the need for constant updates of the expert profiles.

Keywords: Expertise Location. Expert Profiles Extraction. Natural Language Processing. DBPedia.

LISTA DE FIGURAS

Figura 1 – Tipos de repositórios de conhecimento.....	27
Figura 2 – Conhecimento tácito: característica e forma de armazenamento	28
Figura 3 – Conhecimento explícito: característica e forma de armazenamento	29
Figura 4 – Interdisciplinaridade da Gestão do Conhecimento	31
Figura 5 – Resumo das etapas e dos resultados da revisão integrativa da literatura	42
Figura 6 – Arquitetura lógica do modelo de extração de entidades e geração de ontologia de Ceci	54
Figura 7 – Modelo de mineração de textos voltado às aplicações de EC e GC de Gonçalves.....	55
Figura 8 – Modelo de <i>Temporal Knowledge Discovery in Texts</i> de Bovo	56
Figura 9 – Procedimentos metodológicos para desenvolvimento da pesquisa.....	58
Figura 10 - Tipos de revisão da literatura.....	59
Figura 11 – Etapas da revisão integrativa	65
Figura 12 – <i>Tag cloud</i> sobre palavras-chave e resumos dos estudos da revisão integrativa	79
Figura 13 – Visão conceitual do modelo para extração de perfil de especialista	85
Figura 14 – Modelo proposto – etapa de extração do conteúdo.....	88
Figura 15 – Exemplo de <i>tag cloud</i> como resultado da etapa de extração do perfil.....	93
Figura 16 – Exemplo de <i>timeline</i> como resultado da etapa de extração do perfil	93
Figura 17 – Arquitetura de <i>pipelines</i> do DKPro e seus respectivos componentes.....	98
Figura 18 – Arquitetura de <i>pipelines</i> do protótipo do modelo	98

Figura 19 – Resultados da etapa de extração do conteúdo – anotações de metadados do documento.....	101
Figura 20 – Resultados da etapa de extração do conteúdo – anotações de metadados complementares	102
Figura 21 – Resultados da etapa de extração do conteúdo – anotações de sentenças	103
Figura 22 – Resultados da etapa de extração do conteúdo – anotações de tokens	104
Figura 23 – Resultados da etapa de extração do conteúdo – anotações de stopwords	105
Figura 24 – Estrutura do índice Lucene gerado a partir da SKOS da DBPedia.....	108
Figura 25 – Resultados da etapa de reconhecimento de entidades – anotações de conceitos	111
Figura 26 – Resultados da etapa de reconhecimento de entidades – anotações de datas.....	112
Figura 27 – Site que disponibiliza a ferramenta de resultado do protótipo	120
Figura 28 – Ferramenta Simple Desktop Application em uso	121
Figura 29 – Exemplo de perfil com representação em <i>tag cloud</i> gerado pelo protótipo	122
Figura 30 – Exemplo de perfil com representação em <i>timeline</i> gerado pelo protótipo	122
Figura 31 – Exemplo de perfil com representação em relatório gerado pelo protótipo.....	123
Figura 32 – Demonstração de viabilidade com a geração de perfil de especialista #1	124
Figura 33 – Demonstração de viabilidade com a geração de perfil de especialista #2	124
Figura 34 – Demonstração de viabilidade com a geração de perfil de especialista #3	125

LISTA DE QUADROS

Quadro 1 – Definições de conhecimento	25
Quadro 2 – Métodos e ferramentas de apoio aos processos da GC.....	34
Quadro 3 – Diferenças entre revisão sistemática e revisão narrativa	60
Quadro 4 - Comparação dos tipos de revisão de literatura.....	63
Quadro 5 – Detalhamento das etapas da revisão integrativa.....	65
Quadro 6 – Termos correlatos das palavras-chave da revisão integrativa da literatura.....	70
Quadro 7 – Total de estudos encontrados na base Scopus a partir das palavras-chave de pesquisa	72
Quadro 8 – Estudos encontrados na base Scopus a partir da revisão integrativa.....	73
Quadro 9 – Estudos selecionados na revisão integrativa.....	80
Quadro 10 – Matriz de síntese da revisão integrativa com as categorias dos estudos	82
Quadro 11 – Exemplo de anotações de conceitos e datas da etapa de reconhecimento de entidades	89
Quadro 12 – Exemplo de armazenamento de anotações da etapa de extração e armazenamento	90
Quadro 13 – Exemplo de mapa do resultado da etapa de associação de entidades.....	91
Quadro 14 – Lista de tarefas e componentes disponíveis no DKPro	96
Quadro 15 – N-gramas gerados a partir de uma sentença e as respectivas consultas realizadas.....	109
Quadro 16 – Campos do índice Lucene gerados pela etapa de extração e armazenamento de entidade	114
Quadro 17 – Perfil do LinkedIn e Perfil extraído pela ferramenta de protótipo dos especialistas.....	127

LISTA DE ABREVIATURAS E SIGLAS

APQC – American Productivity & Quality Center

EC – Engenharia do Conhecimento

EGC – Engenharia e Gestão do Conhecimento

EL – Expertise Location

ELS – Expertise Location Systems

GC – Gestão do Conhecimento

IDC – International Data Corporation

NER – Named Entity Recognition

PPEGC – Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento

SBC – Sistema Baseado em Conhecimento

SGC – Sistema de Gestão do Conhecimento

UFSC – Universidade Federal de Santa Catarina

SUMÁRIO

1 INTRODUÇÃO.....	17
1.1 CONTEXTUALIZAÇÃO E PROBLEMA DE PESQUISA	17
1.2 OBJETIVOS	19
1.2.1 Objetivo geral.....	19
1.2.2 Objetivos específicos.....	19
1.3 DELIMITAÇÃO DO ESCOPO DA PESQUISA	20
1.4 INTERDISCIPLINARIDADE E ADERÊNCIA AO OBJETO DE PESQUISA DO PROGRAMA	21
1.5 ESTRUTURA DO TRABALHO.....	22
2 FUNDAMENTAÇÃO TEÓRICA	24
2.1 CONCEITUALIZAÇÃO	24
2.1.1 Conhecimento	24
2.1.1.1 <i>Conhecimento tácito</i>	27
2.1.1.2 <i>Conhecimento explícito.....</i>	28
2.1.1.3 <i>Dados, informação e conhecimento.....</i>	29
2.1.2 Gestão do Conhecimento.....	30
2.1.3 Engenharia do Conhecimento	34
2.1.4 Expertise Location	36
2.1.5 Considerações sobre a conceitualização	37
2.1.5.1 <i>Uso dos termos Expertise e Expertise Location.....</i>	39
2.1.5.2 <i>Expertise Location e a área de Recuperação de Informação..</i>	39
2.2 TRABALHOS RELACIONADOS AO TEMA DE PESQUISA..	40
2.2.1 Expertise Location no contexto da GC	43
2.2.2 Métodos conceituais e tecnológicos para a extração de perfis de especialistas	48
2.2.3 Considerações sobre os trabalhos relacionados ao tema de pesquisa	52
2.2.3.1 <i>Modelos de reconhecimento de entidades, correlação e análise temporal.....</i>	53
3 PROCEDIMENTOS METODOLÓGICOS.....	57

3.1	REVISÃO DA LITERATURA	59
3.2	REVISÃO INTEGRATIVA DA LITERATURA	64
3.3	EXECUÇÃO DO MÉTODO DE REVISÃO INTEGRATIVA ...	68
3.3.1	Identificação do tema e seleção da questão de pesquisa	68
3.3.2	Definição dos critérios de inclusão e exclusão de estudos	72
3.3.3	Identificação dos estudos pré-selecionados e selecionados ...	78
3.3.4	Categorização dos estudos selecionados	81
4	APRESENTAÇÃO DOS RESULTADOS	83
4.1	MODELO DE EXTRAÇÃO DE PERFIL DE ESPECIALISTA PROPOSTO	83
4.1.1	Extração do conteúdo.....	87
4.1.2	Reconhecimento de entidades	88
4.1.3	Extração e armazenamento de entidades	89
4.1.4	Associação de entidades	90
4.1.5	Extração do perfil de especialista.....	91
4.2	PROTÓTIPO DA IMPLEMENTAÇÃO DO MODELO PROPOSTO	93
4.2.1	Arquitetura do protótipo	94
4.2.2	Implementação da etapa de extração do conteúdo.....	99
4.2.3	Implementação da etapa de reconhecimento de entidades .	106
4.2.4	Implementação da etapa de extração e armazenamento de entidade	113
4.2.5	Implementação da etapa de associação de entidades	116
4.2.6	Implementação da etapa de extração do perfil de especialista	118
4.2.7	Ferramenta e resultados gerados a partir do protótipo.....	119
4.3	DEMONSTRAÇÃO DE VIABILIDADE	123
4.3.1	Comparação dos perfis extraídos com os perfis publicados no LinkedIn.....	126
5	CONCLUSÕES, CONTRIBUIÇÕES E TRABALHOS FUTUROS	130
5.1	CONTRIBUIÇÕES	132
5.2	TRABALHOS FUTUROS	133

REFERÊNCIAS	135
APÊNDICE A – PERFIL COMPLETO DO ESPECIALISTA #1	146
APÊNDICE B – PERFIL COMPLETO DO ESPECIALISTA #2	159
APÊNDICE C – PERFIL COMPLETO DO ESPECIALISTA #3	162
APÊNDICE D – PERFIL EXTRAÍDO VS LINKEDIN ESPECIALISTA #1	169
APÊNDICE E – PERFIL EXTRAÍDO VS LINKEDIN ESPECIALISTA #2	170
APÊNDICE F – PERFIL EXTRAÍDO VS LINKEDIN ESPECIALISTA #3	171

1 INTRODUÇÃO

Este capítulo tem por objetivo apresentar o propósito do trabalho, tendo sido estruturado em cinco seções. A primeira seção contextualiza o leitor no que tange ao problema de pesquisa abordado. A segunda seção apresenta os objetivos – geral e específicos – que norteiam os estudos que culminaram na redação deste trabalho. A terceira seção detalha o escopo da pesquisa. A quarta seção traz a visão de aderência ao Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento (PPEGC) da Universidade Federal de Santa Catarina (UFSC). E, por fim, a quinta seção apresenta os demais capítulos que exploram este trabalho de acordo com o problema identificado, os objetivos e o seu escopo.

1.1 CONTEXTUALIZAÇÃO E PROBLEMA DE PESQUISA

Segundo Schreiber et al. (1999), a cada dia o conhecimento torna-se mais importante na economia, e pesquisas apontaram que os executivos o consideram como um dos fatores mais importantes para o sucesso de suas organizações. Em indústrias de manufatura, por exemplo, o conhecimento é relacionado a cerca de 65% a 75% do valor total dos produtos (SCHREIBER et al., 1999).

Dada a importância do conhecimento e da sua visão como um recurso-chave, a Gestão do Conhecimento (GC) torna-se uma atividade crucial nas organizações modernas (WIIG, 1997). O seu objetivo está no planejamento, na organização, na motivação e no gerenciamento de pessoas, de processos e de sistemas em organizações para assegurar que os seus ativos de conhecimento sejam empregados de forma eficaz (BASSI, 1999; WEGGEMAN, 1997; WIIG, 1997 apud STEIL, 2007). Esses ativos de conhecimento incluem o conhecimento na forma de documentos impressos, conhecimentos armazenados em repositórios eletrônicos, o conhecimento dos funcionários sobre a melhor maneira de fazer o seu trabalho, o conhecimento que é realizado por equipes e o conhecimento que está incorporado em produtos, em processos e em relacionamentos da organização (SUN, Z; HAO, G. HSM, 2006).

Segundo Young (2010), *os Expertise Location Systems (ELS)* – ou Sistemas para Localização de Especialistas – têm sido uma das principais ferramentas para a Gestão do Conhecimento, pois possuem o potencial de auxiliar as seguintes etapas: (a) identificação de conhecimento; (b) criação de conhecimento; (c) armazenamento de conhecimento; (d) compartilhamento de conhecimento; e (e) aplicação de conhecimento em um programa de GC. Recentemente, a APQC

(2015), em seu relatório anual sobre tendências e prioridades em programas de GC, destacou a *Expertise Location* como uma das três principais iniciativas a serem implementadas em 2015 na área de Gestão de Conhecimento, demonstrando a importância da EL no contexto da GC na atualidade.

De acordo com os estudos de Balog et al. (2012), a construção de ELS ocorreu a partir da década de 1960, e uma série de sistemas foram desenvolvidos para suportar a busca por especialistas em determinados assuntos. Os ELS abordam problemas típicos que grandes organizações enfrentaram: muitas vezes, não é possível encontrar o conhecimento adequado que existe em algum lugar na organização. Esse problema pode induzir a organização a “reinventar a roda” para resolver um mesmo problema já solucionado anteriormente só porque não se sabe quem fez o que ou quem possui determinado conhecimento (BALOG et al., 2012).

Nesse contexto, a EL envolve ferramentas que permitem compartilhar o conhecimento existente, conectando pessoas que precisam de determinados conhecimentos àquelas que possuem esses conhecimentos (YOUNG, 2010). Às vezes, o sistema também auxilia o processo de formação de novas equipes de projetos, agregando os conhecimentos necessários. Essa ferramenta pode ser materializada em uma simples página amarela eletrônica ou em sistemas mais sofisticados para buscar automaticamente *expertise*, com vistas a apoiar a localização e a conexão de pessoas que necessitam de um conhecimento a pessoas que detêm esse conhecimento (BALOG et al., 2012; MAYBURY, 2006; YOUNG, 2010).

Embora seja uma ferramenta importante no papel da GC, muitos sistemas de EL ainda dependem do preenchimento das informações dos colaboradores de forma manual. Vários estudos sugerem que sistemas de EL devem incluir a integração de tecnologias de Inteligência Artificial, *Text Mining*, *Web Mining*, Processamento de Linguagem Natural, entre outras, para melhorar o processo de atualização dos perfis dos colaboradores com base em repositórios eletrônicos de documentos e informações provenientes da internet (BALOG et al., 2012; BALOG; RIJKE, 2007; BECERRA-FERNANDEZ, 2006; MARWICK, 2001; SOHAIL; AFZAL; AHMAD, 2013; YAO; TANG; LI, 2007). Assim, os sistemas de EL possibilitariam uma menor dependência do preenchimento de informações de cada colaborador e permitiriam que os dados dos seus perfis fossem mais facilmente atualizados.

Esses estudos também apontam que o problema com o uso de perfis explícitos, isto é, preenchidos pelos próprios colaboradores, é a

necessidade de atualização constante das informações dos colaboradores, o que demanda mais tempo para essas atividades de atualização. Desse modo, é preferível reunir informações automaticamente, se possível a partir de fontes existentes na organização, para extrair informações relevantes sobre os especialistas da organização (BECERRA-FERNANDEZ, 2006).

De modo a corroborar com essa visão, os estudos realizados pela IDC – International Data Corporation (2010) – descrevem que as fontes de informações não estruturadas (e.g., documentos em linguagem natural) representarão aproximadamente 90% de todo o conteúdo gerado nesta década (2010 a 2020). Por meio de técnicas da Engenharia do Conhecimento, essas informações, quando trabalhadas com a metodologia e com ferramentas adequadas, podem explicitar ativos de conhecimento, tais como redes de relacionamento de especialistas, competências e interesses dos especialistas da organização, informações sobre projetos de uma organização, entre outros (GONÇALVES, 2006). Esses ativos do conhecimento podem, desse modo, auxiliar no processo de criação e atualização dos perfis de especialistas de uma organização.

Dessa forma, com base nos referenciais expostos, esta dissertação aborda a seguinte pergunta de pesquisa: como extrair perfis de especialistas de forma automática a partir de documentos não estruturados para melhorar as ferramentas de *Expertise Location* utilizadas em iniciativas de Gestão de Conhecimento?

1.2 OBJETIVOS

1.2.1 Objetivo geral

Propor um modelo para a extração do perfil de especialista de forma automática a partir de documentos não estruturados para apoiar o desenvolvimento de soluções de *Expertise Location*.

1.2.2 Objetivos específicos

- Identificar as abordagens atuais que possibilitem a extração de perfis de especialistas a partir de documentos não estruturados no contexto de *Expertise Location*.
- Definir um modelo conceitual, de acordo com as abordagens existentes, que suporte a extração automática de perfis de especialistas e identifique os elementos necessários para a implementação do modelo.

- Desenvolver um protótipo funcional para verificar a viabilidade do modelo proposto utilizando documentos não estruturados para extração do perfil de especialista.

1.3 DELIMITAÇÃO DO ESCOPO DA PESQUISA

Conforme apresentado anteriormente, o principal resultado deste trabalho é a proposição de um modelo para a extração automática de perfis de especialistas a partir de documentos não estruturados para apoio à implantação ou atualização de ferramentas de *Expertise Location* que contribuam para iniciativas de Gestão do Conhecimento.

Portanto, o foco do modelo proposto está na extração e na criação de perfis de especialistas a partir de documentos. Não faz parte desta pesquisa a especificação de etapas para a construção de um sistema de *Expertise Location* em sua plenitude. Por exemplo, este modelo não prevê como se fará a busca de especialistas de uma organização, mas se atém quanto à criação do perfil que será apresentado como resultado de buscas nessas ferramentas.

Este modelo também não tem por objetivo quantificar ou assegurar as competências, os conhecimentos e os interesses que um especialista possui. Dessa forma, esta pesquisa limita-se à extração de insumos que possam indicar conhecimentos e interesses dos especialistas a partir da exploração dos seus documentos criados ao longo dos anos.

Outro limite desta pesquisa são as questões que trabalham a comparação dos resultados gerados entre dois ou mais especialistas. O objetivo desta proposta não contempla identificar se um especialista possui mais conhecimento ou interesses em determinados assuntos do que outros especialistas.

Para fins de verificação da viabilidade do modelo proposto, são utilizados documentos de especialistas produzidos ao longo de suas carreiras profissionais. Os testes empregam a base de conhecimento da DBPedia¹ nos idiomas português e inglês, e os documentos dos tipos DOC, DOCX, PPT, PPTX, PDF e TXT.

¹ DBPedia é um projeto cujo objetivo é extrair conteúdo estruturado das informações da Wikipédia. Disponível em: <<http://wiki.dbpedia.org>>.

1.4 INTERDISCIPLINARIDADE E ADERÊNCIA AO OBJETO DE PESQUISA DO PROGRAMA

A área de EC do PPEGC tem por objetivo “a formação de pesquisadores e profissionais responsáveis pela codificação de conhecimento organizacional”. Fornece “uma metodologia científica para analisar e engenhar conhecimento” em sinergia com as áreas de Gestão e Mídia do Conhecimento e “inclui a pesquisa e o desenvolvimento de técnicas e ferramentas para a formalização, codificação e gestão do conhecimento” (EGC, 2015a). Assim, a aderência desta proposta ao PPEGC é observada pela aplicação de métodos e de técnicas ligadas à EC para extração de perfis de especialistas, de modo a explicitar seus conhecimentos a partir de fontes de informação não estruturadas.

Há uma contribuição para a área de Gestão do Conhecimento pelo fato de esta dissertação buscar o aprimoramento de ferramentas *Expertise Location* que são amplamente utilizadas em iniciativas de GC e, também, por abordar o conhecimento “como produto, processo e resultado de interações sociais e tecnológicas entre agentes humanos e tecnológicos” (EGC, 2015b). Para realizar essa contribuição, utilizam-se construções teóricas interdisciplinares relacionadas à administração, à engenharia de software e à inteligência artificial para o desenvolvimento do modelo proposto. Essas construções teóricas são fundamentadas em estudos já realizados no programa na área de Engenharia do Conhecimento, conforme a lista de trabalhos listados a seguir, principalmente na área de recuperação de informação e arquitetura de sistemas.

- BOVO, A. B. **Um modelo de descoberta de conhecimento inerente à evolução temporal dos relacionamentos entre elementos textuais**. 2011. Tese (Doutorado em Engenharia de Produção) – Universidade Federal de Santa Catarina, Florianópolis, 2011.
- CECI, F. **Um modelo semi-automático para a construção e manutenção de ontologias a partir de bases de documentos não estruturados**. 2010. 131 f. Dissertação (Mestrado em Engenharia de Gestão do Conhecimento) – Universidade Federal de Santa Catarina, Florianópolis, 2010.
- BEPLER, F. D. **Um modelo para recuperação e busca de informação baseado em ontologia e no círculo hermenêutico**.

2008. Tese (Doutorado em Engenharia e Gestão do Conhecimento) – Universidade Federal de Santa Catarina, Florianópolis, 2008.

- **SELL, D. Uma arquitetura para business intelligence baseada em tecnologias semânticas para suporte a aplicações analíticas.** 2006. Tese (Doutorado em Engenharia de Produção²) – Universidade Federal de Santa Catarina, Florianópolis, 2006.
- **GONÇALVES, A. L. Um modelo de descoberta de conhecimento baseado na correlação de elementos textuais e expansão vetorial aplicado à engenharia e gestão do conhecimento.** 2006. 196 f. Tese (Doutorado em Engenharia de Produção³) – Universidade Federal de Santa Catarina, Florianópolis, 2006.

1.5 ESTRUTURA DO TRABALHO

Este trabalho é composto por cinco capítulos, considerando-se esta introdução, sendo os demais descritos a seguir.

- Capítulo 2. Fundamentação teórica: neste capítulo, apresenta-se o conteúdo acerca do tema do problema da pesquisa, trazendo-se uma visão dos conceitos de Gestão do Conhecimento, Engenharia do Conhecimento e *Expertise Location* para a conceitualização do leitor e a fundamentação dos conceitos utilizados neste trabalho. Apresenta-se também o conteúdo resultante da revisão sistemática de literatura que gera insumos para a revisão dos trabalhos relacionados às temáticas abordadas nesta pesquisa.
- Capítulo 3. Procedimentos metodológicos: este capítulo tem por objetivo apresentar os procedimentos realizados para materializar os objetivos desta pesquisa. Assim, contextualiza o leitor no que tange à revisão da literatura e aos seus resultados,

² O Dr. Denilson Sell, orientador desta dissertação, defendeu a sua tese no Programa de Engenharia de Produção, mas a contribuição do seu trabalho está relacionada às áreas de Engenharia e Gestão do Conhecimento.

³ O Dr. Alexandre Gonçalves, professor da disciplina Métodos e Técnicas de Engenharia do Conhecimento, ministrada no PPEGC, defendeu a sua tese no Programa de Engenharia de Produção, mas a contribuição do seu trabalho também está relacionada às áreas de Engenharia e Gestão do Conhecimento.

bem como detalha os procedimentos realizados para a construção do modelo proposto.

- Capítulo 4. Resultados: apresenta-se o modelo proposto com base na revisão integrativa da literatura e o protótipo baseado nesse modelo, demonstrando os resultados alcançados.
- Capítulo 5. Conclusões, contribuições e trabalhos futuros: neste capítulo, são destacadas as conclusões sobre o trabalho realizado e as suas contribuições, bem como são delineados os trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta os conceitos dos temas abordados neste trabalho que fomentam a construção do modelo para a extração de perfis de especialistas com base em documentos não estruturados. Assim, a fundamentação teórica é constituída pela conceitualização dos temas Conhecimento, Gestão do Conhecimento, Engenharia do Conhecimento e *Expertise Location*, e pela apresentação do conteúdo da revisão sistemática da leitura com base no objetivo do trabalho.

Na primeira seção, apresenta-se a conceitualização embasada em uma revisão narrativa da leitura, a partir dos conteúdos adotados pelo Programa de Engenharia e Gestão do Conhecimento da Universidade Federal de Santa Catarina, com o intuito de situar e relacionar os temas quanto ao trabalho proposto. Por fim, apresenta-se o conteúdo do resultado da combinação desses conceitos na revisão sistemática integrativa da literatura, que fomenta a identificação das abordagens atuais, as quais possibilitam a extração de perfis de especialistas a partir de documentos não estruturados.

2.1 CONCEITUALIZAÇÃO

A conceitualização da fundamentação teórica apresenta os conteúdos de referência para a elaboração da pesquisa descrita neste trabalho. Para realizar tal pesquisa, é importante fornecer as definições de cada um dos temas abordados na dissertação. Essas definições contribuem para o entendimento do contexto trabalhado e também para a visão dos autores sobre os temas relacionados. Ademais, esta seção apresenta os conceitos utilizados para a elaboração da revisão sistemática da literatura com o objetivo de identificar os trabalhos que abordam os temas de interesse desta pesquisa.

2.1.1 Conhecimento

Nessas últimas décadas, tem-se observado um interesse cada vez maior em utilizar o conhecimento como um recurso organizacional, tornando-se um importante ativo nas organizações (ALAVI; LEIDNER, 2001; GRANT, 1996; SCHREIBER et al., 1999; SPENDER, 1996, TEECE; PISANO; SHUEN, 1997; VENZIN; KROGH; ROOS, 1998). Apesar de sua crescente importância, o conhecimento é um conceito bastante amplo e abstrato que tem sido debatido de forma epistemológica desde a era clássica grega (ALAVI; LEIDNER, 2001; SCHREIBER et al., 1999). Em Teeteto, um diálogo sobre a natureza do conhecimento escrito em aproximadamente 369 A.C., Sócrates definiu o

conhecimento como “uma crença verdadeira justificada”, mas essa definição vem se tornando inadequada (ANAND; SINGH, 2011).

De modo a ter uma definição mais objetiva sobre conhecimento, Anand e Singh (2011), em sua revisão da literatura de entendimento da Gestão do Conhecimento, buscaram referências em autores que trabalharam com Gestão do Conhecimento e também trouxeram definições sobre Conhecimento (e.g. BECKMAN, 1997; DAVENPORT; PRUSAK, 1998; GRANT, 1996; NONAKA; TAKEYCHI, 1995; WIIG, 1993, entre outros). Essas definições são apresentadas no Quadro 1 a seguir.

Quadro 1 – Definições de conhecimento

Definição	Referência
O conhecimento é um fator de produção.	Nonaka e Takeuchi (1995)
O conhecimento reside na cabeça do indivíduo [...] conhecimento é aquilo que é conhecido.	Grant (1996)
O conhecimento consiste em verdades e crenças, perspectivas e conceitos, julgamentos e expectativas, metodologias e know-how.	Wiig (1993)
O conhecimento é informação no contexto juntamente com uma compreensão de como usá-lo.	Davenport e Prusak (1998)
O conhecimento é informação combinada com experiência, contexto, interpretação e reflexão.	Davenport e Long (1998)
O conhecimento é o raciocínio sobre informações para orientar ativamente a execução da tarefa, a resolução de problemas e a tomada de decisão a fim de realizar, aprender e ensinar.	Beckman (1999)
O conhecimento é definido como a compreensão dos efeitos das variáveis de entrada na saída.	Bohn (1994)

O conhecimento é como uma percepção nova ou modificada ou um entendimento preditivo.	Kock e McQueen (1998)
O conhecimento é o conjunto de ideias, experiências e procedimentos que são considerados corretos e verdadeiros, e que, portanto, define os pensamentos, os comportamentos e a comunicação das pessoas.	Van der Spek e Spijkervet (1997)
O conhecimento é crença pessoal justificada que aumenta a capacidade de um indivíduo tomar medidas eficazes.	Alavi e Leidner (1999)
O conhecimento refere-se ao estoque de informação, habilidades, experiência, crenças e memórias de um indivíduo.	Alexander, Schallert e Hare (1991)
O conhecimento tem origem na mente de um indivíduo (o estado mental de ter ideias, fatos, conceitos, dados e técnicas, registrados na memória) e baseia-se na informação que é transformada e enriquecida pela experiência pessoal, pelas crenças e pelos valores com a decisão e a ação que ganham significado. O conhecimento de um indivíduo pode ser diferente do conhecimento adquirido por outra pessoa que recebe a mesma informação.	Bender e Fish (2000)

Fonte: adaptado de Anand e Singh (2011, p. 928)

A partir dessas definições, Anand e Singh (2011) observam que o termo ‘conhecimento’ está associado a três diferentes significados: (1) o estado de saber; (2) a capacidade de ação; e (3) o lugar onde esse conhecimento está codificado. O estado do saber tem relação com a familiaridade de um indivíduo com fatos, métodos, técnicas e princípios. Já a capacidade de ação se refere ao entendimento e à compreensão de fatos, métodos, técnicas e princípios na execução de uma ação (i.e., *know-how*). E, por último, o lugar onde esse conhecimento está codificado tem relação com a maneira em que o conhecimento é articulado e capturado em forma de livros, manuais, código de computador, fórmulas, entre outros.

Anand e Singh (2011) destacam ainda que o conhecimento é gerado na mente humana e pode ser obtido a partir de fontes externas. Uma vez gerado, o conhecimento pode ser armazenado em um

repositório para ser acessado e compartilhado por outros indivíduos ou grupos dentro de uma organização. Dessa forma, existem vários meios de armazenamento em que o conhecimento pode residir, como na mente humana, na organização, nos documentos e em computadores, porém com diferentes graus de acessibilidade, conforme é apresentado na Figura 1 a seguir.

Figura 1 – Tipos de repositórios de conhecimento



Fonte: adaptado de Anand e Singh (2011, p. 929)

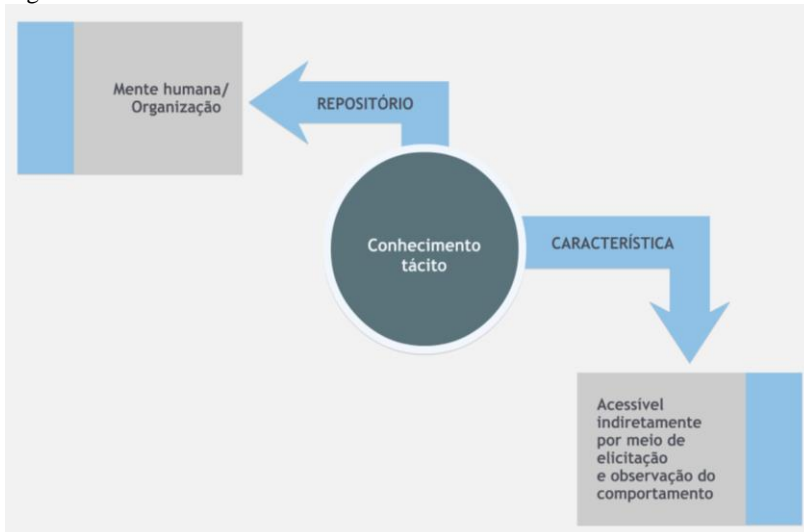
Esse grau de acessibilidade, também conhecido pela dimensão ‘acessibilidade do conhecimento’ (ANAND; SINGH, 2011), tem sido discutido por diversos autores ao longo dos anos (COOK; BROWN 1999; DAVENPORT; PRUSAK, 1998; HANSEN, 1999; NONAKA; TAKEUCHI, 1995; TEECE, 1998; WINTER, 1987). Os autores Nonaka e Takeuchi (1995), a partir dos estudos de Polanyi (1966), dividiram a acessibilidade em duas categorias: tácito e explícito.

2.1.1.1 *Conhecimento tácito*

Uma das definições mais difundidas sobre conhecimento tácito encontrada na literatura foi cunhada por Polanyi (1966), que tenta traduzir o entendimento sobre o conhecimento tácito na frase “nós sabemos mais do que podemos expressar” (ANAND; SINGH, 2011;

BROWN; DUGUID, 1998). Basicamente o conhecimento tácito é incomunicável e intuitivo; é o conhecimento não escrito, de difícil transferência, perceptível em habilidades e na experiência (NONAKA, 1994). O conhecimento tácito é orientado à ação e à prática (i.e., *know-how*), adquirido pela experiência pessoal, raramente expresso verbalmente e muitas vezes se assemelha à intuição (SMITH, 2001). A Figura 2 resume as características do conhecimento tácito e a sua forma de armazenamento.

Figura 2 – Conhecimento tácito: característica e forma de armazenamento



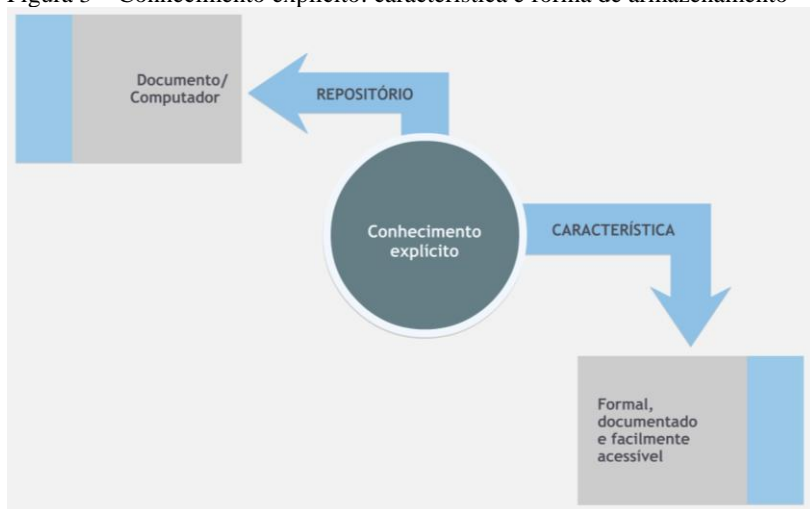
Fonte: adaptado de Anand e Singh (2011, p. 930)

2.1.1.2 Conhecimento explícito

Diferentemente do conhecimento tácito, o conhecimento explícito é aquele que pode ser expresso na forma de palavras, números, textos, especificações e manuais, e também é facilmente comunicável e compartilhável (NONAKA; TAKEUCHI, 1995). Para Smith (2011), uma importante parte desse conhecimento explícito (i.e., *knowwhat*) pode ser classificada como conteúdo técnico ou conteúdo acadêmico, descritos em linguagem formal. Esse conhecimento pode ser compartilhado por meio de documentos impressos, meios eletrônicos e outros meios formais. Smith (2011) destaca ainda que o conhecimento explícito, quando é técnico, exige um nível de conhecimento acadêmico

ou entendimento de que é adquirido através da educação formal ou de estudo estruturado. Esse tipo de conhecimento também pode ser acessado rapidamente quando se encontra em um sistema de informação. A Figura 3 resume as características do conhecimento tácito e a sua forma de armazenamento.

Figura 3 – Conhecimento explícito: característica e forma de armazenamento



Fonte: adaptado de Anand e Singh (2011, p. 930)

2.1.1.3 Dados, informação e conhecimento

O que é conhecimento e qual a sua diferença entre dados e informação é uma dúvida frequente de pessoas da Gestão e Engenharia do Conhecimento (SCHREIBER et al., 1999). Dados, informação e conhecimento são termos utilizados comumente como sinônimos, mas possuem significados diferentes. Com base nas definições encontradas na literatura referentes aos trabalhos de Ackoff (1989), Argyris(1993), Aune (1970), Awad e Ghazi (2004), Desouza (2005), Gallup et al. (2002), Goldman (1991), King (1993), Nonaka e Takeuchi (1995), Thierauf e Hoctor (2006), os autores Anand e Singh (2011) concluíram que dados podem ser considerados como uma representação “crua” e não processada da realidade; informação pode ser considerada como dados que foram processados em alguns aspectos significativos; e conhecimento é considerado como a informação que foi processada em alguns aspectos significativos. Schreiber et al. (1999) também

diferenciaram esses conceitos, conforme exibido a seguir, e apresentaram os seus respectivos exemplos.

- **Dados.** Os dados são os sinais não interpretados que atingem os nossos sentidos (e.g., visão). Um exemplo é uma luz vermelha, verde ou amarela em um cruzamento. Computadores são repletos de dados: sinais que consistem em números, caracteres e outros símbolos que são mecanicamente manipulados em grandes quantidades.
- **Informação.** A informação é considerada como os dados providos de significado. Para um motorista de carro, uma luz vermelha de um semáforo não é apenas um sinal de algum objeto colorido, é uma indicação de “pare”. Ficticiamente, um ser extraterrestre que chegar a Terra provavelmente não atribuirá o mesmo significado à luz vermelha. O dado é o mesmo, mas a informação será diferente.
- **Conhecimento.** Conhecimento é o conjunto de dados e informações que as pessoas utilizam para algum uso prático em ação, para executar tarefas e para criar novas informações. O conhecimento acrescenta dois aspectos diferentes: (a) um senso de propósito; e (b) uma capacidade geradora, porque uma das principais funções do conhecimento é produzir novas informações, tanto que é considerado como um novo “fator de produção”.

2.1.2 Gestão do Conhecimento

Conforme exposto na seção anterior, o conhecimento tornou-se um importante ativo nas organizações. Um dos motivos para o conhecimento ter uma relevância cada vez maior, de modo a ser considerado um dos fatores mais importantes na atual economia (SCHREIBER et al., 1999; WIIG, 1997), é porque tornou-se uma arma estratégica que tem o potencial de aumentar os lucros de uma organização de maneira sustentável (CHOI; LEE, 2002) e torná-la competitiva (ALAVI; LEIDNER, 2001). Essas vantagens podem ser obtidas quando a organização busca, de forma eficiente e eficaz, criar, localizar, capturar e compartilhar o conhecimento e as experiências da

sua organização para aplicá-los na resolução de problemas e na exploração de oportunidades (ZACK, 1999). É nesse contexto que surge a Gestão do Conhecimento para garantir que os processos e as tecnologias façam fluir o conhecimento na organização.

A crescente literatura de Gestão do Conhecimento oferece diversas definições, conceitos e formas de aplicar a GC na prática (ALAVI; LEIDNER, 2001; ANAND; SINGH, 2011; CHOI; LEE, 2002; SMITH, 2001). No que tange às definições, é possível observar uma classificação de Gestão de Conhecimento nas dimensões de objetivos, demandas, processos, tecnologia de informação, estratégia, práticas, natureza holística e capital intelectual (ANAND; SINGH, 2011). Essas diferentes classificações representam uma visão geral do desafio que são as iniciativas de Gestão de Conhecimento nas organizações, uma vez que a GC incorpora uma série de disciplinas necessárias para ser implantada em uma organização. A Figura 4 ilustra as disciplinas ligadas à GC.

Figura 4 – Interdisciplinaridade da Gestão do Conhecimento



Fonte: adaptado de Dalkir (2005, p. 7)

Contudo, a GC pode ser definida de forma geral como a coordenação sistemática de pessoas, tecnologias, processos e estrutura organizacional, com o objetivo de agregar valor por meio da reutilização do conhecimento em uma organização (DALKIR, 2005). Essa coordenação é obtida na criação, no compartilhamento e na aplicação do conhecimento, bem como na aplicação de lições aprendidas e de melhores práticas para uma aprendizagem organizacional contínua.

Logo, essa reutilização do conhecimento cria novas capacidades organizacionais que permitem melhorar o seu desempenho e atender melhor os seus clientes.

A coordenação sistemática de pessoas, tecnologias processos e estrutura organizacional refere-se ao processo da GC. Vários modelos são observados na literatura: Bukowitz e Williams (2003), Daghfous e Kah (2006), Lee, Lee e Kang (2005), McElory (1999), Meyer e Zack (1996), Wiig (1993) e Wong e Aspinwall (2004). De forma geral, esses modelos consideram quatro processos: criação do conhecimento; armazenamento e recuperação do conhecimento; transferência do conhecimento; e aplicação do conhecimento (ANAND; SINGH, 2011; ALAVI; LEIDNER, 2001). A seguir, cada um desses processos é detalhado com base no *Framework* descrito por Alavi e Leider (2001).

- **Criação do conhecimento.** A criação de conhecimento organizacional envolve o desenvolvimento de novo conteúdo ou a substituição de conteúdo existente a partir dos conhecimentos tácito e explícito detidos pela organização. Isso ocorre por meio de processos sociais e de colaboração, bem como através de processos cognitivos do indivíduo (por exemplo, a reflexão) em que o conhecimento é criado, compartilhado, amplificado, ampliado e justificado em ambientes organizacionais. Esse processo de criação do conhecimento abarca uma interação contínua entre as dimensões tácita e explícita de conhecimento em um fluxo espiral, enquanto o conhecimento se move entre os níveis individual, grupal e organizacional.
- **Armazenamento e recuperação do conhecimento.** O armazenamento, a organização e a recuperação do conhecimento organizacional, também conhecido como memória organizacional, consistem no gerenciamento do conhecimento que reside em documentos, informações estruturadas armazenadas em bases de dados eletrônicas, conhecimento humano codificado armazenado em sistemas especialistas, procedimentos e processos organizacionais documentados e em conhecimento tácito adquirido por indivíduos e redes de indivíduos. Esse conhecimento de eventos e experiências do passado pode ser utilizado em atividades no presente.

- **Transferência do conhecimento.** A transferência de conhecimento ocorre em vários níveis: entre indivíduos, de indivíduos para fontes explícitas, de indivíduos para grupos, entre grupos, entre indivíduos de um mesmo grupo e de grupos para a organização. É um processo que envolve o compartilhamento de conhecimentos entre todos dentro da organização, tanto de forma tácita quanto de forma explícita. Um elemento importante na transferência de conhecimento é o canal de transferência, que pode ser formal ou informal, pessoal ou impessoal. Mecanismos informais, tais como reuniões não programadas, seminários informais ou conversas em cafés, podem ser eficazes na promoção da socialização do conhecimento, mas também podem impedir uma divulgação mais ampla desses conhecimentos. Mecanismos de transferência formais, tais como sessões de formação e treinamentos, podem garantir uma maior distribuição do conhecimento, mas podem, por outro lado, inibir a criatividade.
- **Aplicação do conhecimento.** A aplicação do conhecimento refere-se ao processo de aplicação e uso do conhecimento na organização para obtenção de vantagens competitivas. Para que a aplicação do conhecimento ocorra, são necessários três mecanismos: normas diretivas, rotinas organizacionais e equipes de trabalho. As normas diretivas compõem um conjunto específico de regras, normas, procedimentos e instruções desenvolvidos por meio da conversão do conhecimento tácito de especialistas em conhecimento explícito. As rotinas organizacionais estão relacionadas à criação de padrões, protocolos de interação e especificações de processo que permitam aos colaboradores da organização aplicar e integrar os seus conhecimentos. Já as equipes de trabalho dizem respeito à criação de grupos com conhecimento específicos que possam trabalhar em conjunto para a resolução de problemas.

Por fim, para apoiar os processos de conhecimento, (i.e., criação, armazenamento e recuperação, transferência e aplicação do conhecimento), há uma série de técnicas e de ferramentas que contribuem para a implantação da Gestão do Conhecimento em uma organização. Young (2010) apresenta uma lista de métodos e

ferramentas classificados como tecnológicos (i.e., utilizam TI) e não tecnológicos (i.e., não utilizam TI) que suportam esses processos, conforme demonstrado no Quadro 2 a seguir.

Quadro 2 – Métodos e ferramentas de apoio aos processos da GC

Métodos e ferramentas não tecnológicos	Métodos e ferramentas tecnológicos
<ul style="list-style-type: none"> • Brainstorming • Captura de ideias e aprendizagem • Assistência em pares • Revisão de aprendizagem • Revisão pós-ação • Storytelling • Estrutura colaborativa física • Café do conhecimento • Comunidades de prática • Taxonomia 	<ul style="list-style-type: none"> • Expertise Locator • Sistema de gestão de documentos • Bases de conhecimento • Blogs • Redes sociais • VOIP (Voice and Voice-over-Internet) • Buscas avançadas • Clusters para criação de conhecimento • Estrutura colaborativa virtual

Fonte: adaptado de YOUNG (2010, p. 9)

2.1.3 Engenharia do Conhecimento

A Engenharia do Conhecimento surgiu a partir da Inteligência Artificial como uma subárea dedicada à concepção, ao desenvolvimento e à implementação de Sistemas Baseados em Conhecimentos (SBC). Na Engenharia do Conhecimento Clássica os especialistas extraíam o conhecimento e o transferiam para uma base de conhecimento computacional (STUDER; BENJAMINS; FANSEL, 1998). Mais recentemente, a nova Engenharia do Conhecimento teve a sua ênfase no paradigma de modelagem de conhecimento (SCHREIBER et al., 1999).

Segundo os autores Studer, Benjamins e Fensel (1998), as primeiras tentativas da Engenharia do Conhecimento foram com a construção de sistemas denominados *Problem-Solving Method* (PSM), que resultou na concepção de métodos genéricos para a resolução de problemas, mas perdeu especificidade e aplicabilidade. Essa primeira

geração de Sistemas Baseados em Conhecimento era fundamentada em *shells* e em linguagens de representação de conhecimento simbólico e dedutivo, o que acabou culminando em problemas de escala, dificuldade em modelagem e aplicação em problemas complexos.

Na segunda era da Engenharia do Conhecimento, segundo Studer, Benjamins e Fensel (1998), surgiram os métodos estruturados chamados de KADS, em 1985, em função da necessidade de criação de metodologias e de técnicas mais formais para auxiliar o desenvolvimento de SBC. Essas metodologias e técnicas retornaram à visão inicial da EC, que rogava generalidade, separando a modelagem do domínio da tarefa. Mais tarde, em 1995, surgiu o CommonKADs, uma metodologia da EC mais madura. O CommonKADs tornou-se uma valiosa fonte de princípios e de técnicas de modelagem do conhecimento, oferecendo um conjunto de etapas para o desenvolvimento de Sistemas Baseados em Conhecimento (SCHREIBER et al., 1999).

Partindo de um mesmo objeto – o conhecimento como ativo intangível –, a Engenharia do Conhecimento e a Gestão do Conhecimento diferenciam-se pela forma com que se posicionam nas organizações. Para Kendal e Creen (2006), o termo “gestão” da Gestão do Conhecimento relaciona-se ao exercício executivo e à direção administrativa e de supervisão, ao passo que o termo “engenharia” está relacionado ao ato de se construir, inventar ou planejar. Enquanto a GC estabelece uma visão estratégica para posicionamento do conhecimento organizacional, a EC concebe e desenvolve mecanismos inteligentes (modelos formais e tecnologias) para que a tecnologia da informação esteja alinhada a essa visão. Em termos de produtos, a GC propõe Sistemas de Gestão do Conhecimento (SGC), e a EC sugere os Sistemas Baseados em Conhecimento (SBC).

A Engenharia do Conhecimento atual, segundo Schreiber (1999), oferece vários benefícios: permite identificar as oportunidades e os gargalos na forma como as organizações criam, distribuem e aplicam os seus conhecimentos, de modo a fornecer ferramentas para a gestão do conhecimento organizacional; fornece métodos para obter uma compreensão completa das estruturas e dos processos na criação, distribuição e aplicação do conhecimento, possibilitando uma melhor integração da tecnologia da informação e a utilização do conhecimento; e, por último, permite a construção de sistemas de conhecimento mais fáceis de usar, bem estruturados arquiteturalmente e mais simples de se manter.

2.1.4 *Expertise Location*

Frequentemente, as pessoas consultam outras para ajudá-las na busca do caminho correto no intuito de resolver um determinado problema, seja no trabalho, em ambientes acadêmicos ou até mesmo em seus lares. Inclusive, pode existir um grande volume de informações disponíveis acerca do problema, mas contar com uma pessoa que possui conhecimento sobre o tema (i.e., um especialista) torna a sua resolução mais rápida (BALOG et al., 2012).

Nesse contexto, as ferramentas de *Expertise Location* (i.e., *Expertise Location Systems* ou *Expertise Locator*) contribuem na identificação e na localização de especialistas que possuem conhecimento e experiência necessários às outras pessoas e têm ganhado destaque na Gestão do Conhecimento (APQC, 2015; BALOG et al., 2012; MARWICK, 2001; MCDONALD; ACKERMAN, 1998). Principalmente, em grandes empresas, em que saber quem são as pessoas que possuem experiências, habilidades e conhecimento é uma ação dificultada pela quantidade de pessoas (YOUNG, 2010).

De acordo com Balog et al. (2012), as evoluções nos sistemas de *Expertise Location* nas últimas décadas destacaram o surgimento de duas categorias: *Expert Finding Systems* (i.e., Sistemas de Busca de Especialistas) e *Expert Profiling Systems* (i.e., Sistemas de Perfis de Especialistas). Enquanto o objetivo dos Sistemas de Busca de Especialista é justamente encontrar um especialista em um determinado tópico, o objetivo dos Sistemas de Perfis de Especialistas é apresentar quais são os tópicos em que uma pessoa é especialista.

Balog et al. (2012) também afirmam que a área que aborda a elaboração desses sistemas de *Expertise Location* é referenciada como *Expertise Retrieval* (ou recuperação de *expertise*) e é tradicionalmente considerada como um tema de investigação na área de *Information Retrieval* (Recuperação de Informação ou RI). Basicamente, um sistema de recuperação de *expertise* consiste nos componentes descritos a seguir.

- **Aquisição de dados.** A aquisição de dados tem por objetivo a coleta de evidências para a geração de perfis profissionais a partir de fontes de dados disponíveis em uma organização. Essa etapa envolve descobrir as fontes de dados disponíveis, sejam elas estruturadas ou não para compor os dados que serão processados.

- **Pré-processamento e indexação.** O pré-processamento de dados e a indexação são componentes comuns na construção de sistemas de recuperação de documentos. Entretanto, há um desafio adicional para a recuperação de *expertise* na identificação de ocorrências de colaboradores. Para tal, algumas técnicas precisam ser aplicadas nos documentos não estruturados: remoção de *stopwords* de documentos para aumentar a precisão de buscas; *stemming* para converter palavras flexionadas com os mesmos significados semânticos; e *named entity recognition* (NER) para identificar citações de nomes de pessoas em diferentes documentos.
- **Modelo e recuperação.** Esse componente de modelo e recuperação consiste na definição da abordagem de associação entre os elementos obtidos a partir das fontes de dados pré-processados e indexados, que frequentemente relacionam pessoas, documentos e tópicos (e.g., entidade nomeada). Alguns modelos encontrados na literatura são: probabilísticos, que representam a estimativa de associação entre tópicos e pessoas; discriminativos, que determinam de forma binária a relação entre pessoas e tópicos; baseados em votos, que permitem a classificação de documentos, tópicos e pessoas; baseados em grafos, que também determinam as associações entre documentos, tópicos e pessoas; e, por fim, modelos que incluem distribuição de variáveis latentes.
- **Forma de interação.** A forma de interação, também mencionada como design de interação, determina como buscar e apresentar os dados de um sistema de recuperação de *expertise* de modo acessível aos usuários. Pontos importantes, tais como dados de apresentação de resultados (e.g., foto e identificação), *ranking*, documentos relacionados, tópicos relacionados, vocabulários controlados (i.e., taxonomias) e *tags* de identificação de *expertise*, precisam ser trabalhados nesse componente.

2.1.5 Considerações sobre a conceitualização

Esta seção de conceitualização é iniciada pela apresentação do conceito de conhecimento. Observa-se que, apesar de ser objeto de estudo desde a Grécia antiga, não há uma definição de consenso na literatura. As referências definem o conhecimento com base naquilo que uma pessoa sabe, na sua capacidade de utilizar a informação no contexto

correto, nas suas experiências, no saber fazer e na sua importância como um novo fator de produção. A origem do conhecimento é dada na mente das pessoas, e ele pode ser codificado em diversos meios, tais como documentos, manuais, livros, códigos de computadores (e suas respectivas regras de negócios). Além disso, o conhecimento possui classificação quanto à sua acessibilidade, sendo dividido em tácito e explícito. Tácito quando é de difícil transferência, que ocorre por observação; e explícito é de fácil transferência, podendo ser externalizado em documentos ou verbalizado. Observa-se ainda a utilização de dados, informação e conhecimento com o mesmo sinônimo, porém representam conceitos diferentes.

Para que o conhecimento seja aplicado de forma eficiente e eficaz com o intuito de agregar valor às organizações, é apresentada a Gestão do Conhecimento. Conceitualmente, a GC é descrita como coordenação das pessoas, de tecnologias e de processos de estrutura organizacional para uma melhor utilização e reutilização de conhecimento em uma organização, de modo a torná-la mais competitiva. Para que isso ocorra, precisa-se utilizar métodos e técnicas de GC nos processos de criação, armazenamento e recuperação, transferência e aplicação do conhecimento. Por isso, várias disciplinas são empregadas na GC, desde a ciência organizacional até técnicas de escrita e tecnologias de banco de dados.

Quando empregadas em sistemas para a GC, essas várias disciplinas geram Sistemas de Gestão do Conhecimento. Por outro lado, quando é necessária a criação de sistemas que utilizam o conhecimento, isto é, Sistemas Baseados em Conhecimento, atribui-se a responsabilidade à disciplina de Engenharia do Conhecimento. A EC possui ferramentas e métodos formais para esse objetivo, como, por exemplo, o CommonKADS, que possibilita a modelagem do conhecimento necessário para o desenvolvimento de um SBC (e.g., tarefas intensivas em conhecimento, tais como classificação).

Chega-se, então, às ferramentas de *Expertise Location*, que se tornaram grandes aliadas da Gestão do Conhecimento e são objeto de interesse desta pesquisa. Classificadas em Sistemas de Busca de Especialistas e Sistemas de Perfis de Especialistas, possuem foco diferente, porém estão diretamente relacionadas. A primeira é responsável por gerenciar a busca de pessoas relacionadas a um determinado tópico, e a segunda encarrega-se de montar e apresentar o perfil de uma pessoa com a sua expertise. A área que aborda essas ferramentas é conhecida como recuperação de *expertise* e relaciona-se com a área de Recuperação de Informação. Ambos os tipos de

ferramentas, de busca de especialistas e de perfis de especialistas, são objetos desta pesquisa. Da mesma forma, os produtos da Gestão do Conhecimento e da Engenharia do Conhecimento que estão ligados à *Expertise Location* também são objetos desta pesquisa.

Por fim, dado o contexto apresentado, duas ponderações são necessárias quanto às ferramentas de *Expertise Location*. A primeira ponderação tem relação com o uso dos termos *Expertise* e *Expertise Location*, e a segunda refere-se à exploração da área de Recuperação de Informação. Essas ponderações são detalhadas nas próximas subseções.

2.1.5.1 *Uso dos termos Expertise e Expertise Location*

De origem francesa, segundo o dicionário online Priberam⁴, o significado de *expertise* apresentado no site Significados⁵ consiste no “conjunto de habilidades e conhecimentos de uma pessoa, de um sistema ou tecnologia” e representa “experiência, especialização e perícia”. O site Significados detalha ainda que a *expertise* “é o conhecimento adquirido com base no estudo de um assunto e a capacidade de aplicar tal conhecimento, resultando em experiência, prática e distinção naquele campo de atuação” e está relacionada “com as habilidades e competências para executar algo”, sinônimo de *know-how*. Já o tradutor do Google⁶ define *expertise* como “habilidade ou conhecimento em um campo de estudo particular” e traz como sinônimos os termos “competência”, “proficiência” e “especialização”.

Por apresentar vários significados e sinônimos e não possuir uma definição em língua portuguesa (i.e., os dicionários Aurélio⁷ e Michaelis⁸ não têm definição), decidiu-se não traduzir o termo neste trabalho para que não se perca semântica e por entender que se aproxima das definições de conhecimento, seja ele tácito ou explícito. Também se optou por não traduzir *Expertise Location*, que basicamente consiste na localização de *expertise*.

2.1.5.2 *Expertise Location e a área de Recuperação de Informação*

Segundo Balog et al. (2012), a área que aborda as ferramentas de *Expertise Location* é conhecida como recuperação de *expertise* e está ligada à área de Recuperação de Informação. Essa área, por sua vez,

⁴ Disponível em: <<http://priberam.pt/dlpo/expertise>>.

⁵ Disponível em: <<http://www.significados.com.br/expertise>>.

⁶ Disponível em: <<https://translate.google.com.br/#auto/pt/expertise>>.

⁷ Disponível em: <<http://dicionariodoaurelio.com>>.

⁸ Disponível em: <<http://michaelis.uol.com.br/moderno/portugues>>.

conforme Salton (1968 apud BEPLER, 2008), tem foco na estrutura, na análise, na organização, no armazenamento e na busca de informação. A RI deve possibilitar a localização de documentos com base em argumentos de busca de usuários (GONÇALVES, 2006) e oferecer mecanismos para solucionar desafios com relação à natureza de informação não estruturada, documentos descritos em linguagem natural e ampla variedade de assuntos presentes em documentos (MITRA; CHAUDHURI, 2000 apud GONÇALVES, 2005).

Gonçalves (2005) e Bepler (2008) destacam que novas abordagens estão sendo desenvolvidas na área de RI para melhorar o contexto semântico nos documentos. Na fase de identificação de documentos, por exemplo, dicionários controlados, taxonomias, tesouros e extração de entidades estão sendo utilizados. Na fase de representação de documentos, observa-se também o emprego de ontologias e dicionários léxicos para aumentar a precisão na recuperação de documentos. Os autores destacam que esses elementos podem ser representados por meio de modelos, tais como Modelo de Espaço Vetorial (VSM), Modelo de Vetores de Contexto (CVM), Modelo de Indexação Semântica Latente (LSI) e modelos que utilizam processamento de linguagem natural.

Apesar de a área de RI fornecer uma série de métodos e de técnicas para trabalhar documentos que podem indicar tópicos de *expertise* de especialistas, o objetivo desta seção não é esgotar as possibilidades descritas na literatura com relação à área. A aplicação dessas técnicas e os resultados obtidos no contexto da extração de perfil de especialista para apoio às ferramentas de *Expertise Location* e, conseqüentemente, as iniciativas de GC são enfatizados na próxima seção, que busca apresentar os trabalhos que relacionam os temas observados na conceitualização da literatura (i.e., GC, EC e EL) e fomentam a identificação das abordagens que exploram a geração de perfis de especialistas.

2.2 TRABALHOS RELACIONADOS AO TEMA DE PESQUISA

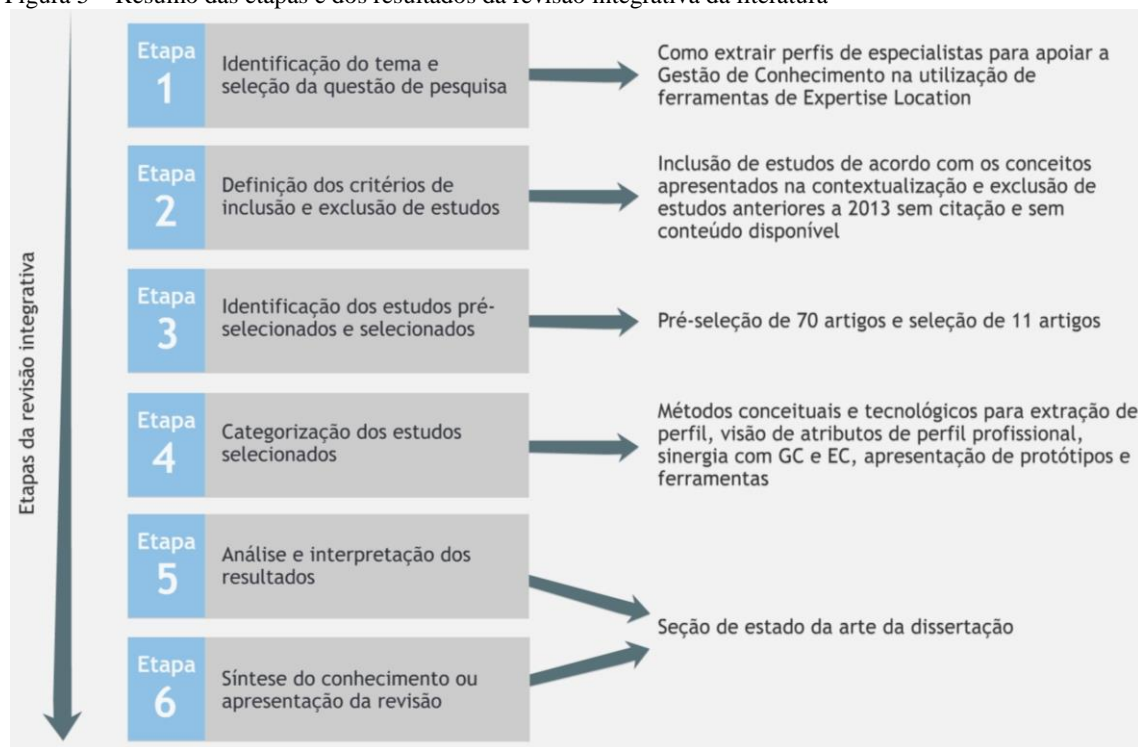
No âmbito da pesquisa bibliográfica, a revisão dos trabalhos sobre um tema de pesquisa possibilita a visualização do conteúdo publicado sobre esse tema, de modo que se possa identificar as lacunas existentes e onde se encontram entraves teóricos ou metodológicos (SILVA, 2005). Esse conteúdo permite, ainda, obter as informações atuais sobre um determinado tema de pesquisa, identificar publicações e confrontar opiniões.

Para determinar os trabalhos relacionados ao tema de pesquisa desta dissertação e identificar os métodos e as técnicas que possibilitam a extração de perfis de especialistas no contexto da Gestão do Conhecimento e ferramentas de *Expertise Location*, optou-se por utilizar o método de revisão integrativa da literatura (seção 3). Esse método, além de permitir a identificação dos estudos realizados sobre um determinado tema, estabelece um protocolo que possibilita a inclusão de estudos empíricos nessa revisão.

Na realização da revisão integrativa (BOTELHO; CUNHA; MACEDO, 2011), a primeira etapa corresponde à identificação do tema e à seleção da questão de pesquisa. Essa é a etapa que guia todas as etapas subsequentes. Além da pergunta de pesquisa bem definida, também há a necessidade da especificação das palavras-chave utilizadas na busca em base de dados. Os passos seguintes definem os critérios de exclusão e inclusão, a identificação dos estudos pré-selecionados e selecionados, a categorização dos estudos e a análise e interpretação dos resultados.

A Figura 5 a seguir apresenta as etapas da revisão integrativa executadas nesta pesquisa e os seus respectivos resultados de forma resumida para cada etapa. A descrição completa da execução do método e os seus resultados são detalhados no capítulo 3.3 deste trabalho.

Figura 5 – Resumo das etapas e dos resultados da revisão integrativa da literatura



Fonte: elaborado pelo autor

No que tange à análise e à interpretação dos resultados e, conseqüentemente, à síntese do conhecimento encontrado nessa revisão integrativa da literatura, os seus achados são mostrados nas próximas subseções. Essas subseções detalham os conteúdos com base na matriz de síntese que consta na seção 3.3.4. O foco desse detalhamento está, primeiramente, na apresentação do apoio da *Expertise Location* na GC e, posteriormente, dos métodos conceituais e tecnológicos para a construção de sistemas de EL. As demais categorias da matriz serviram de base para a construção do modelo proposto detalhado na seção 4.

2.2.1 *Expertise Location* no contexto da GC

A análise e a interpretação dos estudos quanto ao apoio da *Expertise Location* na Gestão do Conhecimento buscam apresentar como esses temas estão relacionados na visão dos autores que tiveram os seus trabalhos selecionados. Com base nesse conteúdo, é possível observar como iniciativas anteriores, de 2006 a 2013, conforme os anos de publicação dos estudos selecionados e categorizados, buscaram tratar essas temáticas.

De forma cronológica, esse conteúdo começa a partir do estudo de 2006 realizado por Becerra-Fernandez (2006). Em seu estudo, este autor analisa a utilização de Sistemas de *Expertise Location* (*Expertise Location Systems* - ELS) e de Gestão do Conhecimento (GC) com base em um cenário de montagem de equipe para um projeto de um ônibus espacial para a NASA. Becerra-Fernandez (2006) classifica o ELS como um sistema que contribui para a GC, isto é, como um Sistema de Gestão do Conhecimento (SGC). Para a autora, os SGC seguem a classificação listada a seguir, na qual se encontra também a classificação dos ELS.

- Sistemas de Captura de Conhecimento (*Knowledge Capture Systems*): preservam e formalizam o conhecimento de especialistas para que possam ser compartilhados com outros especialistas. Esses sistemas formalizam conhecimento em modelos como mapas conceituais, que permitem que outras pessoas compreendam o domínio formalizado.
- Sistemas de Aplicação de Conhecimento (*Knowledge Application Systems*): auxiliam na resolução de problemas. Organizações com expressivo capital intelectual necessitam da identificação e da captura de conhecimento para reutilizar na resolução de novos problemas, bem como em problemas recorrentes já identificados pela organização.

- Sistemas de Descoberta de Conhecimento (*Knowledge Discovery Systems*): criam novos conhecimentos por meio da implementação de algoritmos inteligentes, tais como mineração de dados, e também por meio da inferência das relações entre os dados.
- Sistemas de Compartilhamento de Conhecimento (*Knowledge Sharing Systems*): organizam e distribuem o conhecimento. São repositórios de conhecimento que constituem e representam o maior número de tipos de Sistema de Gestão do Conhecimento. Vários tipos de repositórios de conhecimento são descritos para apoiar a captura e a reutilização de experiência em diferentes contextos. Além de experiências corporativas, existem sistemas de lições aprendidas, bancos de dados de relatórios de incidentes, sistemas de alerta e melhores práticas que também são descritos como sistemas de compartilhamento de conhecimentos. Sistemas de *Expertise Location* (ELS), também conhecidos como páginas amarelas ou sistemas de localização de pessoas (EL), são um tipo especial de sistema de compartilhamento de conhecimento e permitem a localização de especialistas, ou seja, aqueles que detêm o conhecimento e o compartilhamento de conhecimento quando há a interação entre os especialistas.

Essa classificação possibilita o entendimento da aplicação dos sistemas de *Expertise Location* no contexto da GC. Porém, a maior contribuição abordada no estudo de Becerra-Fernandez (2006) é a proposição, de forma conceitual, de tecnologias de aprendizagem de máquina para ajudar no desenvolvimento de ELS, melhorando o processo de atualização das informações do perfil de especialista. Portanto, esse estudo encontra-se alinhado à proposta e ao contexto do presente trabalho.

Seguindo a ordem cronológica dos estudos selecionados, o segundo estudo que apresenta a correção dos temas de EL e GC é o de Balog e Rijke (2007). Os autores destacam que os conhecimentos mais valiosos de uma organização estão nas mentes de seus funcionários. Assim, as empresas devem combinar informação digital de seus repositórios com o conhecimento da experiência dos funcionários. O estudo de Balog e Rijke (2007) foi utilizado como referência no estudo publicado por Balog et al. (2012), apresentado na seção 2.1.4.

Para Balog e Rijke (2007), a localização de especialistas aborda a tarefa de encontrar a pessoa certa com as competências e com os conhecimentos adequados. Por exemplo: “Quem são os especialistas sobre o tema X?”. O contexto e as evidências são necessários para ajudar os usuários de ELS a encontrar e decidir com quem entrar em contato, de modo a comprovar a experiência na área desejada. Por exemplo, dado um especialista cujo nome é resultado de uma busca em um ELS, quais são as suas áreas de especialização? Com quem ele trabalha? Quais são os seus dados de contato?

A principal tarefa apresentada no trabalho de Balog e Rijke (2007) é a de determinar o perfil de um especialista, ou seja, uma descrição concisa das áreas em que uma pessoa é especialista, além de uma descrição de seu ambiente de colaboração. Sua principal contribuição é a introdução de geração de perfis de especialistas como uma tarefa específica anterior à busca por especialista, juntamente com algoritmos para abordar essa tarefa. Balog e Rijke (2007), assim como apresentado nos estudos de Balog et al. (2012), separam a atividade de *Expertise Location* em duas tarefas distintas: (a) construção de perfil de especialista e (b) localização de especialistas. Essa mesma distinção é adotada neste trabalho de pesquisa para, primeiramente, construir um perfil de especialista a partir de documentos não estruturados e, em seguida, criar um ambiente de busca desses perfis.

Balog e Rijke (2007) ainda situam o seu trabalho quanto à Gestão do Conhecimento ao esclarecerem que essas tarefas são previamente tratadas por sistemas de páginas amarelas ligadas à GC. Porém, os autores também discutem que a maioria das ferramentas depende de pessoas para o preenchimento de seus perfis, o que abre espaço para a aplicação de tecnologias inteligentes que podem melhorar o processo de atualização de perfis, citando, inclusive, o trabalho de Becerra-Fernandez (2000).

Yang e Huh (2008), de forma aderente aos estudos de Becerra-Fernandez (2006), abordam a importância da Gestão do Conhecimento e de Sistemas de Gestão do Conhecimento para os ambientes organizacionais. Para os autores, os funcionários podem se utilizar das ferramentas da GC para a recuperação, o armazenamento, o compartilhamento e a busca de informações críticas para a tomada de decisões em seus negócios. Além de citar a importância da GC, Yang e Huh (2008) explicam os tipos de conhecimentos, tácitos e explícitos, conceituados por Nonaka e Takeuchi (1995) e que são apresentados na seção 2.1.1 deste trabalho. Nesse contexto, os autores também fazem a classificação do ELS como um Sistema de Gestão de Conhecimento.

Yang e Huh (2008) explicam que os sistemas de *Expertise Location*, até 2008, eram representados, em sua maioria, por sistemas que necessitavam do preenchimento manual dos perfis das pessoas da organização. Isso representava, inclusive, sistemas comerciais como, por exemplo, sistemas de EL da Microsoft. Embora essa abordagem explícita de preenchimento do perfil facilitasse o desenvolvimento das ferramentas de EL, o método de preenchimento manual das informações requeria constantes custos administrativos para atualização dos perfis. Além dessa desvantagem com o custo de preenchimento dos perfis, havia, ainda, a dificuldade de refletir novos conceitos e terminologias utilizados pelas pessoas da organização nesses sistemas; a dificuldade para medir o nível de conhecimento com base numa tarefa que envolvia a classificação das pessoas por essas pessoas ou por um administrador do sistema; e, por fim, a falta de experiência de uma pessoa para preencher e entender a classificação em todos os campos solicitados no sistema.

Os problemas apresentados por Yang e Huh (2008) foram abordados de maneira que o seu estudo pudesse contribuir, primeiramente, na definição da estrutura do perfil do especialista e, na sequência, na geração automática desse perfil. Para uma primeira contribuição, os autores fizeram uso de três informações que representavam a quantidade de atividade, relevância e utilização do conhecimento como base do perfil do especialista. Para a geração do perfil, segunda contribuição, utilizaram material armazenado nos SGC de uma organização, como documentos e artigos das pessoas, e aplicaram técnicas de categorização de textos nesses artefatos para a extração do conhecimento.

Seguindo a cronologia dos estudos, agora com base no estudo de Li, Liu e Li (2011), que seguem a mesma linha de Yang e Huh (2008), verificou-se, mais uma vez, a declaração explícita da GC. A diferença dos demais estudos está na ligação com a *Expertise Location*. Os autores explicam a importância da GC para utilizar melhor o conhecimento na organização, mas abordam a *Expertise Location* na recomendação de especialistas. Essa recomendação, por sua vez, trabalha a construção dos perfis das pessoas a partir de técnicas de categorização de textos que analisam documentos armazenados em Sistemas de Gestão do Conhecimento.

Li, Liu e Li (2011) explicam em seu estudo que o conhecimento é um recurso estratégico essencial para uma organização e também o categorizam em conhecimento tácito e conhecimento explícito com base em Nonaka e Takeuchi (1995). Li, Liu e Li (2011) discutem a

importância dos Sistemas de Gestão do Conhecimento, em especial a importância dos sistemas de *Expertise Location* para gerenciar esse conhecimento tácito, muitas vezes utilizando-se de ferramentas como páginas amarelas. Nesse ponto de vista, o interessado em determinado conhecimento pode pesquisar nesses sistemas e encontrar especialistas que possuem o conhecimento necessário para ajudá-los.

Contudo, a abordagem de Li, Liu e Li (2011) está relacionada com a utilização de sistemas de recomendação, ao contrário dos demais trabalhos apresentados até esse momento, que trazem abordagens para melhorar os ELS na criação de perfis automáticos e no aprimoramento das buscas nesses sistemas. Essa recomendação de especialistas é pautada em duas prerrogativas: (a) o conhecimento necessário para realizar uma busca de especialistas é incerto porque pode envolver várias áreas de conhecimento; e (b) o usuário pode se perder ou se confundir na utilização da taxonomia dos sistemas devido à quantidade crescente de conhecimentos e de categorias nessa taxonomia. Assim, o trabalho de Li, Liu e Li (2011) ajuda ao descrever soluções que podem auxiliar na utilização de taxonomias nos sistemas de EL.

Por fim, Sohail, Afzal e Ahmad (2013) discutem a importância dos sistemas de EL, mas sem detalhar os conceitos da Gestão do Conhecimento. Para os autores, provenientes da área da medicina, a busca por especialistas é um problema imperativo e vem ganhando popularidade a partir do final da década de 1990, quando a Microsoft, a Hewlett-Packard e a NASA desenvolveram sistemas para localizar especialistas que pudessem compor equipes na realização de novos projetos. Essas afirmações são pautadas nos estudos de Davenport (1997) e de Becerra-Fernandez (2000), autores que, por sua vez, abordam a Gestão do Conhecimento em seus estudos.

Sohail, Afzal e Ahmad (2013) destacam que sistemas de EL são decisivos nos cenários de resolução de problemas, preenchimento de vagas de trabalho, aumento de produtividade e busca por colaboradores que trabalham na mesma área de atuação. A revisão da literatura de seu trabalho mostra, por exemplo, que várias áreas como engenharia de software, pesquisa científica, medicina e administração empresarial tiveram casos de sucesso na utilização de ferramentas de EL, principalmente ligados ao aumento de produtividade. A contribuição do estudo realizado pelos autores está na proposta de identificação dos especialistas da área médica a partir da coleta de dados do Google Scholar e PubMed. Para a criação dos perfis dos especialistas, com base na coleta e processamento desses dados, os autores também utilizaram uma taxonomia própria para classificar esses perfis.

2.2.2 Métodos conceituais e tecnológicos para a extração de perfis de especialistas

Um dos resultados que fomentam a base desta pesquisa no que tange à revisão integrativa da leitura, conforme os objetivos deste trabalho, é a exploração de métodos conceituais e tecnológicos para a extração de perfis de especialistas no contexto das ferramentas de EL. Essa exploração permite encontrar as técnicas utilizadas pelos autores em diferentes campos da ciência, de forma interdisciplinar, para traçar o panorama dos trabalhos relacionados aos temas de pesquisa. Assim, de forma similar à seção 2.2.1, esta seção aborda os estudos resultantes de forma cronológica, isto é, de 2006 a 2013. Esse período compreende os estudos resultantes da busca na revisão integrativa da literatura.

Essa análise dos métodos conceituais e tecnológicos é iniciada pelo estudo de Becerra-Fernandez (2006). Apesar de o seu estudo trazer uma comparação de sistemas de *Expertise Location* até 2006 e sugerir a aplicação de técnicas para melhorá-los, principalmente quanto ao Expert Seeker da NASA, projeto no qual a autora esteve envolvida, as ideias apresentadas nesse estudo ainda são exploradas na atualidade. Sua abordagem traz a utilização de informações de pessoas divulgadas na internet. Becerra-Fernandez (2006) lembra que o conteúdo da internet não é estruturado e, portanto, técnicas usualmente utilizadas para a extração de informação como *data mining* não podem ser aplicadas diretamente. Seu estudo, então, apresenta a utilização de inteligência artificial para o Processamento de Linguagem Natural (PLN) e classifica essa extração de conhecimento como *Web Mining*.

O estudo de Becerra-Fernandez (2006) descreve que o uso de *Web Mining* pode mitigar alguns dos problemas inerentes ao preenchimento de dados de forma tendenciosa pelas próprias pessoas para manter os seus dados em sistemas de EL. Em seu estudo, Becerra-Fernandez (2006) relata também que a *Web Mining* pode mitigar a necessidade de desenvolver taxonomias manualmente para expressar os conhecimentos de uma organização.

Para demonstrar como utilizar a *Web Mining*, Becerra-Fernandez (2006) apresenta um processo do uso dessa técnica em um sistema de EL. Em primeiro lugar, todos os dados relevantes em HTML são transferidos para um diretório local para processamento. A segunda etapa identifica todas as instâncias de nomes de colaboradores de forma programática examinando cada arquivo HTML no diretório local. Os nomes são extraídos a partir de um dos bancos de dados de informações pessoais que faz a comparação no HTML. Os nomes no banco de dados

dos colaboradores são organizados em uma estrutura de mapa, o qual é composto por todos os nomes de funcionários, em todas as formas possíveis que podem aparecer. A terceira etapa envolve a identificação de palavras-chave dentro do conteúdo HTML. Isso é feito utilizando-se um cálculo de frequência. Primeiro, o texto é dividido em palavras individuais. Qualquer sequência de caracteres alfabéticos é reconhecida como uma palavra, enquanto pontuação, números e caracteres de espaço em branco são ignorados. Em seguida, a lista resultante de palavras é processada para determinar se uma palavra foi utilizada em uma lista de palavras irrelevantes (e.g., *stopwords*). Essa lista resultante é então processada com um algoritmo para remover o sufixo de uma palavra (i.e., *stemming*). Tal processo é realizado para agrupar palavras que podem ser escritas de forma diferente, mas que possuem o mesmo significado semântico. A quarta e última etapa envolve o cálculo da frequência de cada termo. A técnica TF-IDF (*Term Frequency – Inverse Document Frequency*) é utilizada durante o processo de seleção de palavras-chave para determinar quais são os termos de índice. A *expertise* é atribuída quando um colaborador aparece de modo recorrente em muitos documentos juntamente com uma determinada palavra-chave. Essa recorrência indica que a pessoa possui algum conhecimento (i.e., *expertise*) em relação a essa palavra-chave.

O segundo estudo, de forma cronológica, que contribui para a análise dos modelos conceituais e tecnológicos da revisão da literatura é o de Balog e Rijke (2007). A contribuição desse trabalho com esta revisão se dá pela apresentação de um método para descobrir e identificar possíveis áreas de conhecimento de uma determinada pessoa. Além desse método, os autores também propõem outro método para medir a competência das pessoas nessas áreas.

O método de Balog e Rijke (2007) para identificar as áreas de conhecimento de uma pessoa tem por objetivo descobrir a *expertise* dessa pessoa com base na exploração de documentos não estruturados. Dessa forma, Balog e Rijke (2007) esclarecem que as áreas de conhecimento são obtidas por meio de processamento linguístico. As áreas de conhecimento extraídas através das técnicas linguísticas são associadas com as pessoas identificadas nesses documentos. Isto é, para identificar as pessoas, também é realizado um algoritmo de PLN que encontra as entidades relacionadas a elas (essa técnica também é conhecida como *Named Entity Recognition – NER*). Para atribuir as áreas de conhecimento às pessoas, os autores utilizam-se das mesmas técnicas apresentadas por Becerra-Fernandez (2007), ou seja, extraem os termos mais significativos de cada documento. Essa extração é feita com

base em *tokenization* e remoção de *stopwords*, porém sem *stemming*, e aplicando-se a técnica TF-IDF.

O terceiro estudo, dos autores Yao, Tang e Li (2007), tem por objetivo a separação no tratamento das informações para gerar os perfis profissionais de forma automática. Ou seja, os autores propõem um método para processar diferentes áreas de um perfil de especialista de forma segmentada, aplicando-se algoritmos especializados para cada unidade de informação classificada por eles. Assim, o método proposto pelos autores emprega três etapas: definição de página relevante, pré-processamento dos dados e *tagging*. O resultado final desse método é um arquivo baseado na ontologia Friend Of A Friend (FOAF) que descreve pessoas, suas atividades e suas relações com outras pessoas e objetos. Basicamente, trata-se de um *parser* de informações que percorre páginas da internet e classifica o conteúdo dessas páginas com base em *tags* esperadas. Os estudos desses autores apontam a utilização de *Support Vector Machines* (SVM) para a classificação das informações dos conteúdos encontrados.

O quarto trabalho, dos autores Yang e Huh (2008), aborda a identificação do especialista usando uma técnica de categorização de texto. Seu processo de categorização é dividido em duas etapas, por meio da abordagem *Vector Space Model* (VSM): 1) preparação e treinamento de documentos; e 2) criação de um vetor para cada área de assunto identificada no documento. Nessas etapas, os autores também utilizam a remoção de *stopwords* e *stemming*. Para o processo de criação de perfil de especialista, são utilizados os documentos registrados nas bases de conhecimento da organização e associados aos especialistas. Após a criação de vetores de cada área de conhecimento, os vetores são atribuídos aos especialistas de acordo com a autoria desses documentos. Isto é, todos os documentos anexados às bases de conhecimento da organização são atribuídos aos especialistas que os inseriram nessas bases. O processo completo envolve os seguintes passos: registro de documento na base de conhecimento da organização; criação do vetor do documento; cálculo de aderência dos documentos aos assuntos identificados na organização; e atualização dos perfis de especialistas na base de dados da organização.

O quinto trabalho, de Yang et al. (2008), propõe a construção de perfil de especialista de forma automática utilizando as informações disponíveis sobre as suas publicações. Para classificar os seus conhecimentos, são utilizados os domínios da Wikipédia. Basicamente, o trabalho desses autores relata a construção de uma taxonomia com base nas informações da Wikipédia e a amarração desses conceitos ao

especialista por meio dos conceitos encontrados em suas publicações. O destaque de sua abordagem é observado no método de extração dos termos das publicações dos especialistas. Sua abordagem utiliza o método C-value/NC-value, uma combinação de linguística e estatística para a identificação de termos compostos por várias palavras (e.g., inteligência artificial, redes neurais, entre outros).

O sexto trabalho, de Li, Liu e Li (2011), aborda a melhoria dos dois processos de um sistema de *Expertise Location*: a construção do perfil do especialista e a busca dos especialistas. Essa melhoria é proposta com a adoção do método linguístico baseado em lógica difusa (i.e., lógica *fuzzy*) para a extração de áreas de conhecimento de especialistas a partir de documentos. Esse método se chama modelo linguístico 2-tupla (2-Tuple Linguistic Model), proposto por Herrera e Martínez (2000).

O trabalho de Krusheva e Tsiporkova (2012), sétimo trabalho descrito, introduz uma visão conceitual para comparar perfis de especialista. Apesar de não focar na construção do perfil de especialista, esse trabalho recomenda a exploração de conteúdos da internet por meio de *Web Mining* para a captura dos dados dos especialistas. Alguns sites que contêm conteúdo de especialistas são citados, tais como LinkedIn, DBLP, Microsoft Academic Search e Google Scholar Citation. As autoras sugerem a utilização de PLN para a captura de palavras-chave que possam descrever as expertises dos especialistas.

O oitavo trabalho, de Liu et al. (2012), aborda a construção de perfis de especialistas utilizando dados disponíveis em sistemas de perguntas e respostas online, como, por exemplo, o Yahoo! Answers. Seu método combina as informações de conhecimento dos especialistas com as reputações de suas respostas e análise de links entre temas e especialista para determinar se uma pessoa possui uma determinada *expertise*. Nesse cenário, os dados históricos de respostas dos usuários são obtidos e analisados utilizando técnicas de RI e representam os assuntos de conhecimento dos especialistas. Sua abordagem emprega as técnicas de VSM e TF-IDF para processar o conteúdo de pares dos sistemas de pergunta e resposta.

O nono e último trabalho da revisão integrativa, dos autores Sohail, Afzal e Ahma (2013), aborda a construção do perfil de especialista na área médica utilizando as informações de publicações contidas nos sites Google Scholar e PubMed. O perfil de especialista é construído a partir de três informações disponíveis: número de publicações, número de citações recebidas e rede de coautoria. Na etapa de criação do perfil, os autores utilizam um *crawler* que captura o

conteúdo dos especialistas previamente selecionados e processa as informações disponíveis, armazenando-as em uma base de dados. Essas informações são, então, trabalhadas especificamente em uma página de informação do especialista, isto é, o perfil do especialista é criado. Na última etapa, o conteúdo do perfil do especialista é associado a uma taxonomia da MeSH (Medical Subject Headings).

2.2.3 Considerações sobre os trabalhos relacionados ao tema de pesquisa

A seção sobre os trabalhos relacionados ao tema de pesquisa, resultado da revisão integrativa da literatura sobre os temas de perfil de especialista, Gestão do Conhecimento, Engenharia do Conhecimento e *Expertise Location*, apresenta um panorama sobre dois aspectos importantes para esta pesquisa: a contribuição das ferramentas de *Expertise Location* para a Gestão do Conhecimento e os métodos conceituais e tecnológicos para a extração de perfis de especialistas a partir de documentos.

Sobre o primeiro aspecto, observa-se um conjunto de estudos que correlacionam esses temas e nos quais se destacam os trabalhos realizados pelos seguintes autores: Becerra-Fernandez (2006), da Universidade Internacional da Flórida (Florida International University – FIU), EUA; Balog e Rijke (2007), da Universidade de Stavanger, na Noruega, e da Universidade de Amsterdam, na Holanda, respectivamente; Yang e Huh (2008), da Universidade de Keimyung (Keimyung University – KMU), Coreia do Sul; e Li, Liu e Li (2011), da Universidade do Petróleo, na China. Nessa revisão, destaca-se que os estudos selecionados na busca de trabalhos com os temas de EL, GC e EC, no contexto de perfis de especialistas, abordam as ferramentas de *Expertise Location* como Sistemas de Gestão do Conhecimento.

Com relação ao segundo aspecto, métodos conceituais e tecnológicos para a extração de perfis de especialistas a partir de documentos, observou-se um diversificado conjunto de soluções para extrair informações e conhecimentos de documentos com o objetivo de contribuir com a *Expertise Location*, seja pela extração ou pela geração de perfil de especialista, isto é, pela contribuição na busca e recomendação de especialistas. É importante ressaltar que, apesar de descrever os métodos conceituais e tecnológicos, esta seção abordou também processos, técnicas, algoritmos e tecnologias que pudessem contribuir com o tema e fornecer subsídios para a proposta de modelo de extração de perfis de especialistas deste trabalho.

Observa-se que trabalhos analisados nesta seção, que apresentam métodos conceituais e tecnológicos, também descrevem o uso de técnicas provenientes das áreas de Recuperação da Informação e Inteligência Artificial, como pode ser observado na seção 2.1.4, que trata sobre a conceitualização dos temas no que tange ao processo de recuperação de *expertise*. Destacam-se nesses trabalhos a utilização de Processamento de Linguagem Natural e suas técnicas de tokenização, normalização, remoção de *stop-words*, *lemmatization*, *stemming*, *Part-Of-Speech Tagging* e *Named Entity Recognition*; *Web-mining*; *Support Vector Machines (SVM)*; *2-Tuple Linguistic Model*; *C-value/NC-value*; e *TF-IDF*.

Conclui-se que esses trabalhos, apesar de fornecerem visões sobre diferentes propostas conceituais e tecnológicas para a exploração do tema de *Expertise Location*, não apresentam soluções para uma questão importante quando se trata de *expertise*: a dimensão tempo. Saber se um especialista explorou ou se ainda explora um determinado tema pode ser um fator crucial em um momento de contato, por exemplo. Crescentes demonstrações de exploração de um tema, quando observado pela dimensão tempo, também podem indicar uma maior familiaridade de um especialista nesse tema. Assim, a seção a seguir apresenta modelos que podem contribuir com essa revisão.

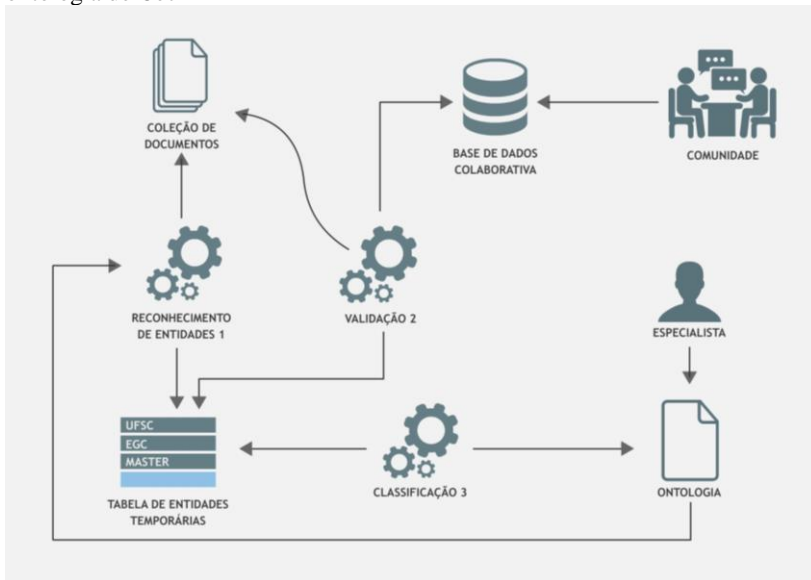
2.2.3.1 Modelos de reconhecimento de entidades, correlação e análise temporal

Com o objetivo de combinar perspectivas que possam contribuir com novas visões sobre os trabalhos selecionados acerca dos métodos conceituais e tecnológicos para a extração do perfil de especialistas a partir de documentos, inclui-se a essa discussão os modelos de Gonçalves (2006), Ceci (2010) e Bovo (2011), atualmente professor e doutores, respectivamente, pelo PPGEHC da UFSC na área de Engenharia do Conhecimento. O trabalho de Ceci (2010) é incorporado pelo seu modelo para a extração de entidades, importante passo na identificação dos termos de *expertise* contidos nos documentos ligados aos especialistas. O modelo de Gonçalves (2006) é incorporado porque aborda a correlação e a associação de entidades para extração do conhecimento. E, por fim, o modelo de Bovo (2011) é incorporado pela contribuição com o processo de análise temporal, elemento de destaque não explorado nos trabalhos selecionados da revisão integrativa.

O modelo ilustrado na Figura 6, proposto por Ceci (2010), possui três etapas: (1) reconhecimento de entidades da coleção de documentos por meio de clusterização; (2) validação das entidades recuperadas com

base nos termos da base colaborativa da Wikipédia; e (3) classificação das entidades e adição das instâncias encontradas na ontologia de domínio da organização por meio de utilização de dicionário léxico, expressões regulares e método estatístico de identificação de instâncias.

Figura 6 – Arquitetura lógica do modelo de extração de entidades e geração de ontologia de Ceci

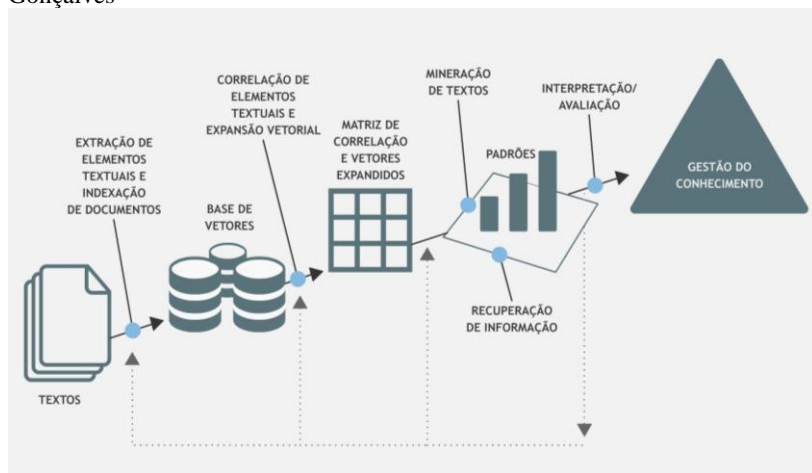


Fonte: adaptado de Ceci (2010, p. 66)

O modelo de Gonçalves (2006), apresentado na Figura 7, aborda a extração de elementos textuais como entidades, conceitos e termos, e os relaciona por meio de mapeamento de espaço vetorial. Seu modelo descreve cinco fases da extração de elementos textuais até a visualização dessas informações para apoio à Gestão do Conhecimento. A primeira fase é a extração de elementos textuais por meio do NER utilizando bases de conhecimento e análises de padrões léxicos. Após essa extração de elementos, é realizada a fase de correção dos elementos textuais, que ocorre por meio de análise de vetores. Na fase seguinte de expansão vetorial, é suportada a matriz de correção, que destaca os elementos mais relevantes dos vetores. Esses vetores expandidos são, então, agrupados para indicar os seus possíveis relacionados. Por fim, o modelo de Gonçalves (2006) apresenta a fase de visualização dos padrões produzidos, que resultam em mapas de conhecimento e são

apresentados de forma gráfica para que ocorram *insights* em seus observadores.

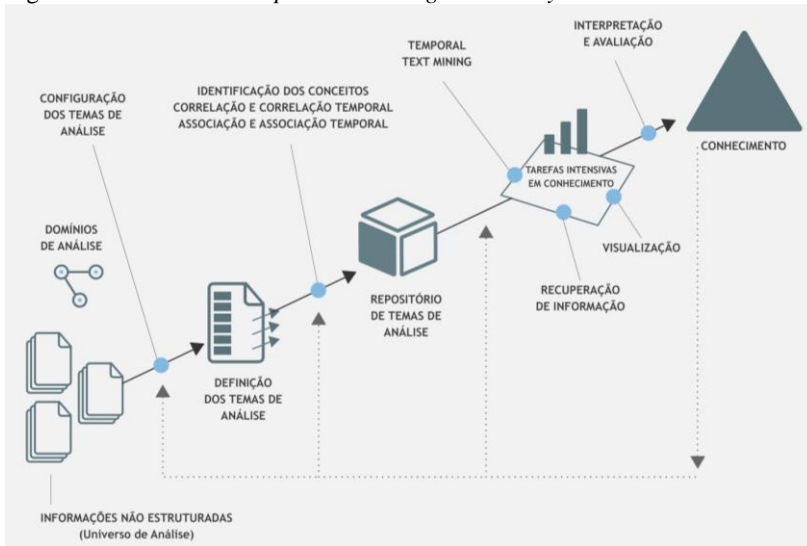
Figura 7 – Modelo de mineração de textos voltado às aplicações de EC e GC de Gonçalves



Fonte: adaptado de Gonçalves (2006, p. 73)

O terceiro modelo, o de Bovo (2011), apresentado na Figura 8 e similar ao modelo de Gonçalves (2006) no que tange às etapas de pré-processamento de dados, mineração de textos e pós-processamento de dados, propõe a inclusão da dimensão tempo no processo de descoberta de conhecimento em textos. O modelo de Bovo (2011) é composto por seis etapas. A primeira etapa é a configuração de temas de análise, que corresponde à modelagem do conhecimento envolvido no processo de extração de conhecimento e inclusão de bases de conhecimento. A segunda etapa, de identificação de ocorrência dos conceitos, é responsável pela identificação e marcação temporal desses conceitos com base no domínio de análise. A terceira etapa, de correlação e correlação temporal, determina a força de correção entre os conceitos identificados em relação ao domínio de análise. A quarta etapa, de associação e associação temporal, determina a força de associação entre dois conceitos. A penúltima etapa, de repositório de temas de análise, representa um modelo de cinco dimensões que permite navegar no conteúdo gerado. A última etapa, de tarefas intensivas em conhecimento, prevê a utilização do repositório de temas de análise para a descoberta de conhecimento.

Figura 8 – Modelo de *Temporal Knowledge Discovery in Texts* de Bovo



Fonte: adaptado de Bovo (2011, p. 67)

A partir dos modelos de Bovo (2011), de Ceci (2010) e de Gonçalves (2006), que apresentam processos de correlação temporal, identificação de conceitos com base na Wikipédia e em dicionários léxicos para a explicitação de conhecimento, e com base nos demais trabalhos identificados na revisão integrativa da literatura, que abordam principalmente o Processamento de Linguagem Natural (PNL) como solução para se identificar *expertise* em documentos de especialistas, apresenta-se o modelo proposto, resultado desta pesquisa. Esse modelo é definido de acordo com as abordagens existentes para a extração de perfis de especialistas a partir de documentos não estruturados e pode ser observado na seção 4.

3 PROCEDIMENTOS METODOLÓGICOS

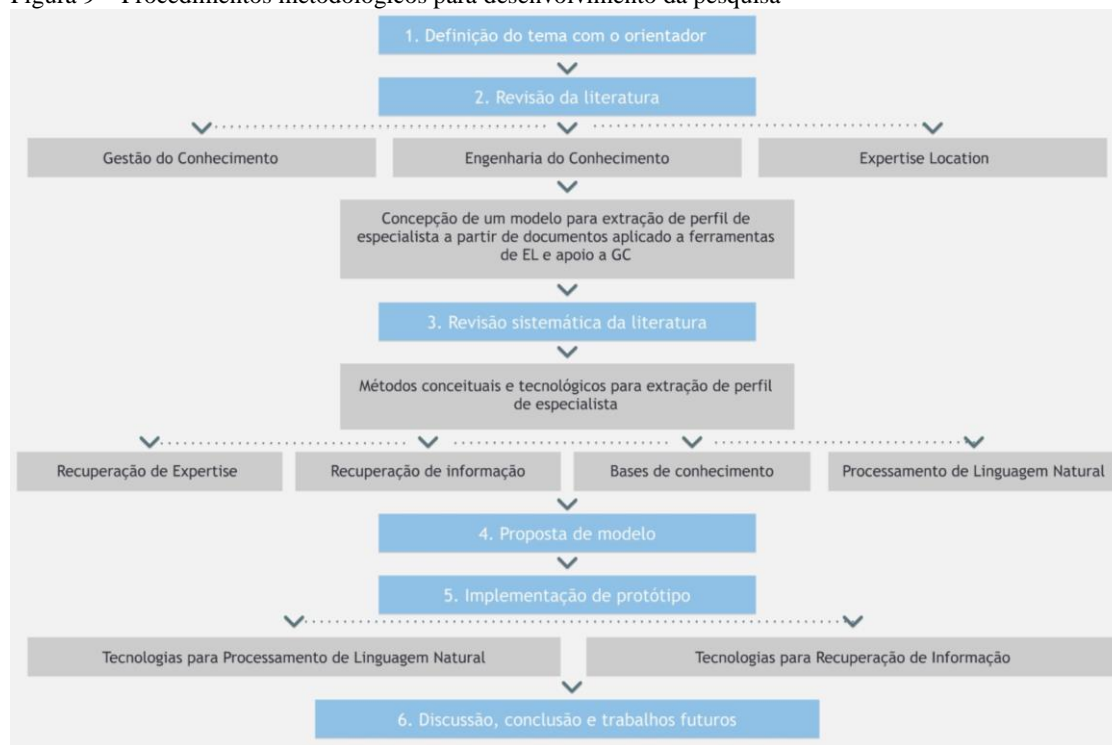
A ciência, da forma que a conhecemos hoje, é fruto da necessidade humana de desvendar e entender o funcionamento da natureza. Várias civilizações como a chinesa, a grega, a indiana e a europeia contribuíram, por meio de realizações científicas e tecnológicas, para o nascimento da chamada “ciência moderna” (KNELLER, 1980). Um dos pilares dessa ciência moderna é o conhecimento racional, obtido por meio da realização de procedimentos metódicos e metodológicos que permitem que todo conhecimento novo possa ser verificado, experimentado, dando origem ao chamado “conhecimento científico” (APOLINÁRIO, 2012).

Esse conhecimento científico pode ser definido como “um conhecimento concreto, real (vem dos fatos), organizado e sistematizado, obtido por meio de um processo bem definido (método científico) e que pode ser replicado (outros pesquisadores [...])” (APOLINÁRIO, 2012, p. 12). Naturalmente, o conhecimento científico é influenciado pela objetividade e pela subjetividade dos pesquisadores, sendo construído com base em visões de mundo do homem. A sua visão gera verdades da sua própria realidade e da sua concepção filosófica (CRESWELL, 2010; MORGAN, 1980).

Essa conceitualização da ciência e do conhecimento científico é importante para justificar a explicação dos procedimentos metodológicos adotados nesta pesquisa, de modo que permita a investigação desse tema por outros pesquisadores. Destaca-se que a visão de mundo dos autores desta pesquisa se enquadra no paradigma funcionalista de Morgan (1980), e a modalidade é caracterizada como pesquisa tecnológica, com o objetivo de projetar artefatos tecnológicos à luz do conhecimento científico (FREITAS et al., 2012).

Dessa forma, a natureza desta pesquisa se enquadra em pesquisa aplicada e exploratória por prover uma aplicação prática e dirigida à solução de problemas específicos (SILVA; MENEZES, 2005). Para embasamento e exploração dos trabalhos relacionados a esta pesquisa, adotou-se a revisão sistemática da literatura, que é detalhada na seção 3.3. A Figura 9 apresenta os procedimentos metodológicos para alcançar os objetivos propostos.

Figura 9 – Procedimentos metodológicos para desenvolvimento da pesquisa



Fonte: elaborado pelo autor

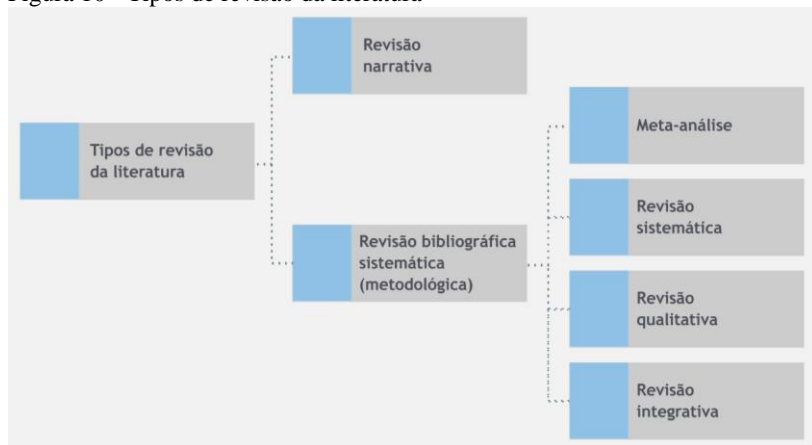
3.1 REVISÃO DA LITERATURA

A revisão da literatura é uma etapa vital na elaboração de projetos acadêmicos. Uma revisão eficaz cria uma base sólida para o avanço do conhecimento, uma vez que facilita o desenvolvimento de novas teorias e possibilita a descoberta de áreas que necessitam de investigação (WEBSTER; WATSON, 2002). A revisão da literatura envolve localizar, analisar, sintetizar e interpretar uma investigação prévia, com base em revistas científicas, livros, anais de congressos, entre outros documentos relacionados à área de estudo (BENTO, 2012). Neste trabalho, a revisão da literatura é fundamental para que se atinja o objetivo específico de identificação das abordagens que possibilitam a extração de perfis de especialistas.

Existem diferentes formas de se realizar uma revisão da literatura: as que tomam por base técnicas como a revisão bibliográfica tradicional e as que se utilizam de mecanismos e metodologias empregados por acadêmicos e pesquisadores para descrever o estado da arte de um tema (BOTELHO; CUNHA; MACEDO, 2011). Essas revisões são classificadas em narrativas e sistemáticas, respectivamente. As revisões sistemáticas são subclassificadas em meta-análise, revisão sistemática, revisão qualitativa e revisão integrativa, sendo diferenciadas essencialmente pelo seu rigor metodológico.

A Figura 10 apresenta a visão hierárquica dos tipos de revisão de literatura classificados por Botelho, Cunha e Macedo (2011).

Figura 10 - Tipos de revisão da literatura



Fonte: adaptado de Botelho, Cunha e Macedo (2011, p. 125)

Por sua vez, a revisão narrativa, também chamada de tradicional, não utiliza critérios explícitos e sistemáticos para a busca e análise da literatura quando comparada à revisão sistemática (CORDEIRO et al., 2007). A busca pelos estudos não precisa esgotar as fontes de informações e também não aplica estratégias de busca sofisticadas e exaustivas. A seleção dos estudos e a interpretação das informações podem estar sujeitas à subjetividade e à análise crítica pessoal do pesquisador (BERNARDO; NOBRE; JANETE, 2004).

Por outro lado, a revisão sistemática é uma revisão planejada para responder a uma pergunta específica claramente formulada. Nesse tipo de revisão, utilizam-se métodos explícitos e sistemáticos para identificar, selecionar e avaliar criticamente os estudos prévios realizados, o que conduz a uma síntese dos resultados de múltiplos estudos primários (CASTRO, 2006 apud BOTELHO; CORDEIRO et al., 2007; CUNHA; MACEDO, 2011).

Para Cook, Mulrow e Raynes (1997), a qualidade de uma revisão da literatura e, conseqüentemente, o seu valor científico têm relação com a aplicação de métodos para minimizar erros. Essa é a característica fundamental que distingue as revisões narrativas tradicionais e as revisões sistemáticas. Se uma revisão é elaborada com base em métodos da revisão sistemática, por exemplo, é mais provável que os resultados apresentem conclusões imparciais. A síntese das diferenças entre uma revisão narrativa e uma revisão sistemática é apresentada no Quadro 3 a seguir.

Quadro 3 – Diferenças entre revisão sistemática e revisão narrativa

Características	Revisão narrativa	Revisão sistemática
Quanto às perguntas	Ampla, frequentemente não especificada	Específica, baseada em critérios uniformemente aplicados
Quanto à seleção de fontes	Subjetiva, muitas vezes não especificada	Fontes específicas e explícitas
Quanto à avaliação	Subjetiva	Crítica e rigorosa

Quanto à síntese	Frequentemente qualitativa	Resumo quantitativo, quando se inclui uma síntese estatística é considerada uma meta-análise
Quanto às inferências	Às vezes é baseada em evidências	Geralmente baseada em evidências

Fonte: adaptado de Cook, Mulrow e Raynes (1997, p. 378)

Apesar da convergência entre os subtipos de revisões sistemáticas, pode-se observar especificidades em seus propósitos, suas amostragens e seus processos de análise. A seguir, são apresentadas as definições gerais de cada subtipo de revisão sistemática. Essas definições são comparadas no Quadro 4, explorando-se também o proposto, o escopo, a amostra e o forma de análise dos tipos de revisão de literatura apresentados e resumidos por Whitemore e Knafl (2005) e Botelho, Cunha e Macedo (2011).

- **Meta-análise:** combina estudos primários, aplicando-se fórmulas estatísticas para melhorar a objetividade e a validade dos resultados de pesquisas progressas (GLASS, 1976). Assim, a meta-análise consiste em colocar diferentes estudos juntos em um mesmo banco de dados e empregar metodologias analíticas e estatísticas para explicar a variância dos resultados utilizando fatores comuns aos estudos (ROSCOE; JENKINS, 2005). Uma de suas vantagens é elevar a objetividade das revisões de literatura, minimizando possíveis vieses e aumentando a quantidade de estudos analisados.
- **Revisão sistemática:** combina evidências de múltiplos estudos a respeito de um problema específico e formulado para avaliá-los e sintetizá-los em um tópico específico. Uma revisão sistemática requer uma pergunta bem especificada, métodos explícitos e uma pesquisa abrangente de estudos primários relevantes (COUNSELL, 1997; GREENHALGH, 1997).
- **Revisão qualitativa:** tem por objetivo sintetizar os estudos primários qualitativos. Whitemore e Knafl (2005) destacam

que esse tipo de revisão de literatura foi aprimorado na última década com a combinação de vários métodos, como, por exemplo, metassínteses, metaestudos, *grounded theory* e metaetnografia. Os autores também observam que sintetizar as evidências sobre os múltiplos aspectos que incorporam a revisão qualitativa é um procedimento complexo, porém possui um grande potencial para a criação de novos estudos.

- Revisão integrativa: é o tipo mais amplo dos métodos de avaliação de pesquisa que permitem a inclusão simultânea de investigação experimental e não experimental, a fim de melhor compreender um fenômeno de interesse do pesquisador (WHITTEMORE, 2005). O termo “integrativa”, segundo Botelho, Cunha e Macedo (2011), tem origem na integração de opiniões, conceitos ou ideias provenientes das pesquisas utilizadas no método. Assim, revisões integrativas também podem combinar dados teóricos e empíricos, além de possibilitar a aplicação em diferentes propósitos como definição de conceitos, revisão de teorias, avaliação de evidências e análise de questões metodológicas de um tema específico (BROOME, 1993 apud WHITTEMORE; KNAFL, 2005).

Quadro 4 - Comparação dos tipos de revisão de literatura

Tipo de revisão	Definição	Propósito	Escopo	Amostra	Análise
Meta-análise	Um sumário de pesquisas passadas, que usa técnicas de estatísticas para transformar descobertas de estudos com hipóteses idênticas ou relativas em uma medida comum e que calcula o efeito total, a magnitude do efeito e os efeitos de subamostras	Estimar o efeito de intervenções ou de relacionamentos	Limitado	Pesquisa quantitativa de metodologia similar	Estatística
Revisão sistemática	Um sumário de pesquisas passadas que usa um objetivo e uma abordagem rigorosa de estudos com hipóteses idênticas ou relativas	Resumir evidências relativas a um problema específico	Limitada	Pesquisa quantitativa de metodologia similar	Narrativa ou estatística
Revisão qualitativa	Um sumário de pesquisas passadas que combina as descobertas de múltiplos estudos qualitativos	Informar pesquisas ou práticas pelo resumo de processos ou experiências	Limitada ou ampla	Pesquisa qualitativa	Narrativa
Revisão integrativa	Um sumário da literatura, num conceito específico ou numa área de conteúdo, em que a pesquisa é resumida, analisada, e as conclusões totais são extraídas	Revisar métodos, teorias e/ou estudos empíricos sobre um tópico particular	Limitada ou ampla	Pesquisa quantitativa ou qualitativa; literatura teórica; literatura metodológica	Narrativa

Fonte: adaptado de Botelho, Cunha e Macedo (2011, p. 128)

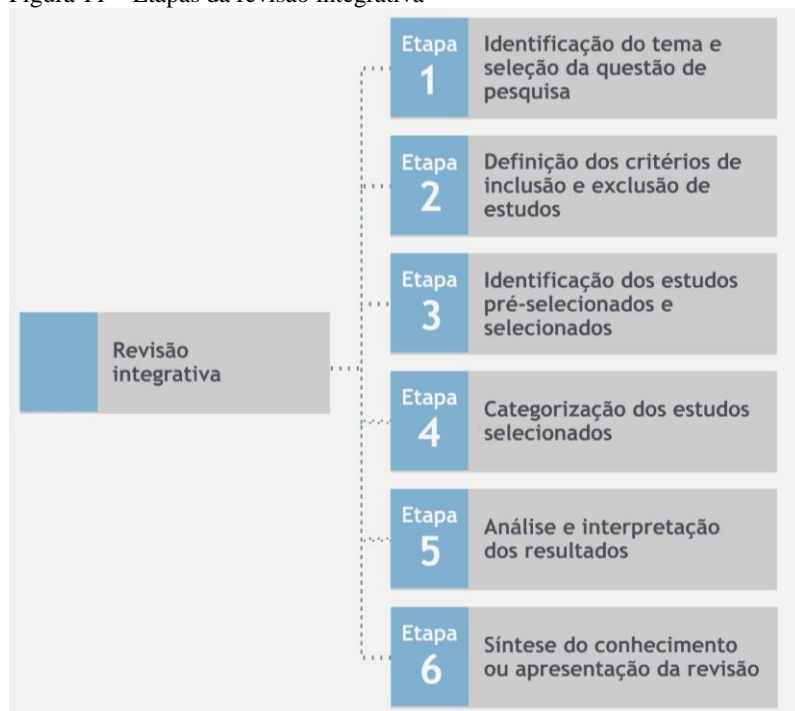
Conforme exposto, a revisão da literatura utilizada neste trabalho tem por finalidade identificar as abordagens atuais que possibilitem a extração de perfis de especialistas a partir de documentos para que se defina um método. Por se tratar de uma revisão de métodos, teorias e estudos empíricos numa área específica de aplicação, optou-se por se utilizar o tipo de revisão integrativa no desenvolvimento deste trabalho, em especial na elaboração do modelo que suporte a extração de perfis. O método da revisão integrativa, que guia a revisão da literatura, é apresentado na seção a seguir.

3.2 REVISÃO INTEGRATIVA DA LITERATURA

A revisão integrativa é um método que permite o resumo de um conceito da literatura, seja teórica ou empírica, de modo a fornecer uma compreensão ampla a respeito do tema de interesse do pesquisador (WHITTEMORE, 2005). Essas análises de pesquisas relevantes dão suporte à tomada de decisão e possibilitam a síntese do estado do conhecimento de um determinado assunto, apontando também lacunas do conhecimento que precisam ser preenchidas com a realização de novos estudos (MENDES; SILVEIRA; GALVÃO, 2008).

Para elaborar uma revisão integrativa relevante, Botelho, Cunha e Macedo (2011) e Mendes, Silveira e Galvão (2008) afirmam que é necessário que as etapas a serem seguidas estejam claramente descritas e percorram 6 etapas distintas. Essas etapas são descritas na Figura 11 e detalhadas no Quadro 5.

Figura 11 – Etapas da revisão integrativa



Fonte: adaptado de Botelho, Cunha e Macedo (2011, p. 129)

Quadro 5 – Detalhamento das etapas da revisão integrativa

Etapa	Descrição
1 – Identificação do tema e seleção da questão de pesquisa	Inicia-se com a definição de um problema e a formulação de uma pergunta de pesquisa. A pergunta de pesquisa deve ser clara e específica. Com a pergunta de pesquisa definida, o próximo passo é a definição dos descritores de palavras-chave, da estratégia de busca, bem como dos bancos de dados a serem utilizados. Com a pergunta de pesquisa, definem-se os descritores e constrói-se a estratégia de busca.

Etapa	Descrição
2 – Definição dos critérios de inclusão e exclusão de estudos	<p>Executa-se a busca nas bases de dados para identificação dos estudos que serão incluídos na revisão.</p> <p>Esta etapa depende muito dos resultados encontrados ou delineados na etapa anterior. Frequentemente, a seleção de estudos inicia-se de forma mais ampla e afunila-se à medida que o pesquisador retorna à sua questão inicial.</p> <p>Os critérios de inclusão e exclusão devem ser identificados no estudo, sendo claros e objetivos, mas podem sofrer reorganização durante o processo de busca dos estudos e também durante a elaboração da revisão integrativa.</p>
3 – Identificação dos estudos pré-selecionados e selecionados	<p>Realiza-se a leitura criteriosa dos títulos, dos resumos e das palavras-chave de todas as publicações completas localizadas pela estratégia de busca para posteriormente se verificar a sua adequação aos critérios de inclusão do estudo.</p> <p>Nos casos em que o título, o resumo e as palavras-chave não sejam suficientes para definir a sua seleção, busca-se a publicação do artigo na íntegra.</p> <p>Com a conclusão desse procedimento, elaborase uma tabela com os estudos pré-selecionados para a revisão integrativa.</p>

Etapa	Descrição
4 – Categorização dos estudos selecionados	<p>Sumarizam-se e documentam-se as informações extraídas dos estudos científicos encontrados nas fases anteriores.</p> <p>Um dos instrumentos utilizados para se extraírem as informações dos estudos selecionados é a matriz de síntese usada como ferramenta de extração e organização de dados de revisão da literatura em várias disciplinas. Esse uso deve-se à sua capacidade para resumir aspectos complexos do conhecimento.</p> <p>A matriz serve de ferramenta para interpretar e construir a redação da revisão integrativa para os pesquisadores.</p> <p>O processo de construção da matriz depende da criatividade pessoal do pesquisador. Não há matriz de síntese correta, apenas matrizes funcionais ou não. Dessa forma, a construção da matriz depende da interpretação do pesquisador e da maneira como ele organiza os seus dados.</p>
5 – Análise e interpretação dos resultados	<p>Discute-se sobre os textos analisados na revisão integrativa. O pesquisador, guiado pelos achados, realiza a interpretação dos dados e, com isso, é capaz de levantar as lacunas de conhecimento existentes e sugerir pautas para futuras pesquisas.</p> <p>O pesquisador, fundamentado nos resultados da avaliação crítica dos estudos incluídos, realiza a comparação com o conhecimento teórico, a identificação de conclusões e implicações resultantes da revisão integrativa.</p>
6 – Síntese do conhecimento ou apresentação da revisão	<p>Elabora-se um documento que contemple a descrição de todas as fases percorridas pelo pesquisador, de forma criteriosa, e que apresente os principais resultados obtidos.</p> <p>A revisão deve possibilitar a replicação do estudo e contemplar informações que possibilitem que os leitores avaliem a pertinência dos procedimentos empregados na elaboração da revisão.</p>

Fonte: adaptado de Botelho, Cunha e Macedo (2011, p. 129)

Em sua execução, a revisão integrativa possibilita a síntese e a análise do conhecimento científico já produzido sobre um tema investigado. Também facilita a obtenção de informações que permitem aos leitores avaliarem a pertinência dos procedimentos empregados numa etapa de revisão da literatura (BOTELHO; CUNHA; MACEDO, 2011).

Botelho, Cunha e Macedo (2011, p. 133) afirmam que a revisão integrativa da literatura “permite ao pesquisador aproximar-se da problemática que deseja apreciar, traçando um panorama sobre a sua produção científica, de forma que possa conhecer a evolução do tema ao longo do tempo e, com isso, visualizar possíveis oportunidades de pesquisa”.

3.3 EXECUÇÃO DO MÉTODO DE REVISÃO INTEGRATIVA

A primeira etapa para se realizar uma revisão integrativa corresponde à identificação do tema e à seleção da questão de pesquisa. Essa é a etapa mais importante da revisão integrativa e a que guia todas as etapas subsequentes. Além da pergunta de pesquisa bem definida, também há a necessidade da especificação das palavras-chave utilizadas na busca em base de dados. Os passos seguintes definem os critérios de exclusão e inclusão, de identificação dos estudos pré-selecionados e selecionados, de categorização dos estudos e de análise e interpretação dos resultados. O último passo da revisão é a síntese do conhecimento ou a apresentação da revisão, demonstrada na seção 2.2.

3.3.1 Identificação do tema e seleção da questão de pesquisa

Este trabalho busca explorar soluções para o problema recorrente de atualização de informações de perfis profissionais de pessoas em ferramentas de *Expertise Location* que apoiam a Gestão do Conhecimento em organizações. Vários trabalhos (BALOG et al., 2012; BECERRA-FERNANDEZ, 2006; MARWICK, 2001) apontam que as organizações que utilizam ferramentas de *Expertise Location* possuem uma grande dependência de seus colaboradores no que tange ao preenchimento das informações nos perfis profissionais, o que pode acarretar num desuso dessas ferramentas. Esse desuso pode ser originado por várias razões, desde a desmotivação pela morosidade na atualização dos dados em um sistema de informação ou até por falta de políticas que incentivem os colaboradores a preencherem os seus dados e que demonstrem a real utilização dessas informações pela organização ou pelos próprios colaboradores. Contudo, existe espaço para que esse preenchimento do perfil de especialista possa ocorrer de forma

automática, o que permitiria uma simplificação no processo de atualização desses conteúdos e uma possível maximização do uso de ferramentas de *Expertise Location* (BALOG et al., 2012; BECERRA-FERNANDEZ, 2006; MARWICK, 2001).

A partir desse contexto como tema de pesquisa, a pergunta que sintetiza esse problema e guia o desenvolvimento deste trabalho, apresentada no primeiro capítulo, é: “*como extrair perfis de especialistas a partir de documentos para apoiar a Gestão de Conhecimento na utilização de ferramentas de Expertise Location?*”.

O tema e a pergunta de pesquisa, no contexto de ferramentas de *Expertise Location* e associado com a Gestão do Conhecimento e com a Engenharia do Conhecimento, foram convertidos em descritores de palavras-chave para a busca de documentos na base de dados da Scopus⁹. A base Scopus foi selecionada devido à sua grande relevância acadêmica, por conter um expressivo número de *journals*, por ser multidisciplinar, por possuir um vocabulário controlado e por possuir um conteúdo mais abrangente do que a base de dados Web of Science CHADEGANI et al., 2013; KUMAR, 2013; UNIVERSITY OF WASHINGTON, 2015).

Para a conversão dos descritores em palavras-chave, foi realizada a tradução dos termos de referência para a língua inglesa, uma vez que se utilizou a base Scopus também de língua inglesa. Além disso, foi utilizada a ferramenta Planejador de Palavras-chave do Google¹⁰, que possibilita a identificação de termos correlatos e o acesso às estatísticas sobre as palavras-chave buscadas. Apesar de o uso da ferramenta do Planejador ser voltado para a realização de campanhas no AdWords¹¹ (i.e., realização de publicidade), essa ferramenta permite uma ampla visão sobre termos similares, além de possibilitar a visão de seu impacto nas buscas dos usuários que usam a internet, o que comporta uma definição mais rica e abrangente dos descritores em palavras-chave (e seus termos) que foram utilizados na busca da base de dados da Scopus.

O Quadro 6 a seguir, apresenta a lista dessas palavras-chave, conforme o tema, a questão de pesquisa, as traduções em inglês e o

⁹ Scopus: um banco de dados de resumos e citações de artigos para jornais/revistas acadêmicos. Disponível em <<http://www.scopus.com>>.

¹⁰ Planejador de palavras-chave do Google: Planeje suas campanhas da Rede de Pesquisa e saiba o que seus clientes procuram. Disponível em: <<https://adwords.google.com/KeywordPlanner>>.

¹¹ Google AdWords: Publicidade online. Disponível em: <<https://www.google.com.br/adwords>>.

resultado da expansão das palavras-chave com a utilização da ferramenta denominada Planejador de Palavras-chave do Google.

Quadro 6 – Termos correlatos das palavras-chave da revisão integrativa da literatura

Palavra-chave	Palavra-chave em inglês	Termos correlatos
Perfil de especialista	Expert profile	<ul style="list-style-type: none"> • Expert Data • Expert Identification • Expert Profile • Personal Data • Personal Profile • Professional Data • Professional Resume • Resume Builder • Resume Building • Resume Writer
Localização de expertise ¹²	Expertise Location	<ul style="list-style-type: none"> • Expert Finder • Expert Finding • Expert Location • Expert Locator • Expert Seeker • Expert Seeking • Expertise Finder • Expertise Finding • Expertise Location • Expertise Location System • Expertise Locator • Expertise Search • Expertise Searching • Expertise Software

¹² Foram encontradas duas possibilidades de tradução para o *Expertise Location*: Localização de expertise ou Locação de perícias. Contudo, optou-se por empregar o termo em inglês na redação deste trabalho por não serem encontradas fortes referências de trabalhos em português com essas traduções.

Palavra-chave	Palavra-chave em inglês	Termos correlatos
Gestão do conhecimento	Knowledge Management	<ul style="list-style-type: none"> • Knowledge Base System • Knowledge Based Systems • Knowledge Management Information Systems • Knowledge Management System • Knowledge Systems
Engenharia do Conhecimento	Knowledge Engineering	<ul style="list-style-type: none"> • Data and Knowledge Engineering • Knowledge and Data Engineering • Knowledge Base Applications • Knowledge Base System • Knowledge Based Engineering • Knowledge Based Systems • Knowledge Management Application • Knowledge Search Engine

Fonte: elaborado pelo autor

De posse dos termos correlatos das palavras-chave da revisão da literatura, utilizou-se a seguinte estratégia de combinação das palavras-chave na busca de documentos da base Scopus:

- busca pelo grupo de palavras-chave e por seus termos correlatos de “Perfil de especialista”, “*Expertise Location*”, “Gestão do Conhecimento” e “Engenharia do Conhecimento”;
- busca pelo grupo de palavras-chave e seus termos correlatos de “Perfil de especialista” e “*Expertise Location*”;
- busca pelo grupo de palavras-chave e seus termos correlatos de “Perfil de especialista” e “Gestão do Conhecimento”;

- busca pelo grupo de palavras-chave e seus termos correlatos de “Perfil de especialista” e “Engenharia do Conhecimento”; e
- combinação da lista de resultados e remoção de redundâncias.

Observa-se que a prioridade em relação às palavras-chave e aos termos correlatos buscados é dada para o perfil de especialista, uma vez que se está interessando em buscar estratégias para a sua extração nos contextos de *Expertise Location*, Gestão do Conhecimento e Engenharia do Conhecimento. Conforme apresentado no Capítulo 1 deste trabalho, há uma ligação direta com o tema de *Expertise Location* para apoio à Gestão do Conhecimento, da mesma forma que também há uma ligação direta com o tema de Engenharia do Conhecimento para apoio à Gestão do Conhecimento.

Ainda na estratégia de busca, concluiu-se que os melhores opções para realização dessa busca por documentos seriam por meio dos campos de título, de palavras-chave e de resumo da base de dados Scopus.

3.3.2 Definição dos critérios de inclusão e exclusão de estudos

Seguindo o método para revisão integrativa, segunda etapa de definição dos critérios de inclusão e exclusão de estudos, o primeiro procedimento é realizar a busca nas bases de dados selecionadas. Essa busca foi realizada conforme a estratégia apresentada na seção anterior, e seu resultado encontra-se no Quadro 7 a seguir.

Quadro 7 – Total de estudos encontrados na base Scopus a partir das palavras-chave de pesquisa

Grupos de termos de busca	Estudos encontrados na Scopus
“Perfil de especialista” e “ <i>Expertise Location</i> ” e “Gestão do Conhecimento” e “Engenharia do Conhecimento”	2
“Perfil de especialista” e “ <i>Expertise Location</i> ”	20
“Perfil de especialista” e “Gestão do Conhecimento”	52
“Perfil de especialista” e “Engenharia do Conhecimento”	49
Total de documentos sem redundância	70

Fonte: elaborado pelo autor

A partir do resultado da lista de estudos encontrados, definiu-se a utilização do critério de exclusão de estudos não citados por outros trabalhos, ou seja, foram eliminados estudos que não apresentam citação na base da Scopus. Nos casos de estudos mais recentes encontrados na Scopus, de 2013 a 2015, entendeu-se que esse critério poderia excluir estudos com grande potencial, mas sem citações por serem muito recentes. Logo, nesses casos, foram mantidos os estudos mesmo sem qualquer citação. Assim, o resultado consolidado dessa etapa foi de 44 estudos apresentados no Quadro 8 a seguir.

Quadro 8 – Estudos encontrados na base Scopus a partir da revisão integrativa

#	Autores	Título do trabalho	Ano de publicação	Citações
1	Balog K., De Rijke M.	Determining expert profiles (with an application to expert finding)	2007	42
2	Pedrycz W.	Fuzzy clustering with a knowledge-based guidance	2004	40
3	Becerra-Fernandez I.	Searching for experts on the Web: A review of contemporary expertise locator systems	2006	38
4	Larizza C., Moglia A., Stefanelli M.	M-HTP: A system for monitoring heart transplant patients	1992	36
5	Seebregts C.J., Zwarenstein M., Mathews C., Fairall L., Flisher A.J., Seebregts C., Mukoma W., Klepp K.-I.	Handheld computers for survey and trial data collection in resource-poor settings: Development and evaluation of PDACT, a Palm™ Pilot interviewing system	2009	32
6	Wang M.-H., Lee C.-S., Hsieh K.-L., Hsu C.-Y., Acampora G., Chang C.-C.	Ontology-based multi-agents for intelligent healthcare applications	2010	30

#	Autores	Título do trabalho	Ano de publicação	Citações
7	Li M., Liu L., Li C.-B.	An approach to expert recommendation based on fuzzy linguistic method and fuzzy text classification in knowledge management systems	2011	28
8	Macdonald C., Ounis L.	Expertise drift and query expansion in expert search	2007	15
9	Yang K.-W., Huh S.-Y.	Automatic expert identification using a text categorization technique in knowledge management systems	2008	14
10	Sanchez D., Castella-Roca J., Viejo A.	Knowledge-based scheme to create privacy-preserving but semantically-related queries for web search engines	2013	13
11	Liu D.-R., Chen Y.-H., Kao W.-C., Wang H.-W.	Integrating expert profile, reputation and link analysis for expert finding in question-answering websites	2013	8
12	Yao L., Tang J., Li J.	A unified approach to researcher profiling	2007	7
13	Blanco Y., Pazos J.J., Gil A., Ramos M., Fernandez A., Diaz R.P., Lopez M., Barragans B.	AVATAR: An approach based on semantic reasoning to recommend personalized TV programs	2005	7
14	Karimzadehgan M., White R.W., Richardson M.	Enhancing expert finding using organizational hierarchies	2009	6

#	Autores	Título do trabalho	Ano de publicação	Citações
15	Yang K.-H., Chen C.-Y., Lee H.-M., Ho J.-M.	EFS: Expert finding system based on wikipedia link pattern analysis	2008	6
16	Niu B., Li Q., Zhu X., Cao G., Li H.	Achieving k-anonymity in privacy-aware location-based services	2014	5
17	Niu X., McCalla G., Vassileva J.	Purpose-based expert finding in a portfolio management system	2004	5
18	Afzal M.T., Maurer H.	Expertise recommender system for scientific community	2011	4
19	Chang Y.-H., Yang C.-L.	A high-efficiency knowledge management system based on habitual domains and intelligent agents	2008	4
20	Espinosa R., Garcia-Saiz D., Zorrilla M., Zubcoff J.J., Mazon J.-N.	Development of a knowledge base for enabling non-expert users to apply data mining algorithms	2013	3
21	Ciuciu I., Zhao G., Mulle J., Von Stackelberg S., Vasquez C., Haberecht T., Meersman R., Bohm K.	Semantic support for security-annotated business process models	2011	3
22	Squicciarini A., Mont M.C., Bhargav-Spantzel A., Bertino E.	Automatic compliance of privacy policies in federated digital identity management	2008	3

#	Autores	Título do trabalho	Ano de publicação	Citações
23	Zhang W., Thurow K., Stoll R.	A Knowledge-based telemonitoring platform for application in remote healthcare	2014	2
24	Naudet Y., Latour T., Vidou G., Djagloul Y.	Towards a novel approach for High-Stake Decision Support System based on Community Contributed Knowledge Base	2010	2
25	Weir W.C.S., Sisson Jr. R.D., Bose S.	Rule-based software tool to specify the manufacturing process parameters for a contoured EB-PVD coating	2000	2
26	El-Sawy A.A., Abdalla H.S.	Hybrid approach for machining process optimization using multiple experts data	1999	2
27	Adair Kristin L., Levis Alan P., Hruska Susan I.	Expert network development environment for automating machine fault diagnosis	1996	2
28	Boeva V., Krusheva M., Tshiporkova E.	Measuring expertise similarity in expert networks	2012	1
29	Allard T., Nguyen B., Pucheral P.	Safe realization of the Generalization privacy mechanism	2011	1
30	Li M.	An approach to expert finding based on multi-granularity two-tuple linguistic information	2011	1

#	Autores	Título do trabalho	Ano de publicação	Citações
31	Anthopoulos L., Gerogiannis V., Fitsilis P., Kameas A.	Training in profiling, negotiation and crisis management: Using an immersive and adaptive environment	2010	1
32	Haruechaiyasak C., Kongthon A., Thaiprayoon S.	Building a Thailand researcher network based on a bibliographic database	2009	1
33	Li M., Liu L., Wang J., Huang Z.	Approach to expert recommendation with multiple knowledge areas based on fuzzy text categorization	2009	1
34	Mahdi H., Attia S.S.	MASCE: A multi-agent system for collaborative e-learning	2008	1
35	Espinosa R., Garcia-Saiz D., Zorrilla M., Zubcoff J.J., Mazon J.-N.	Enabling non-expert users to apply data mining for bridging the big data divide	2015	0
36	Ur Rehman M.H., Liew C.S., Wah T.Y., Shuja J., Daghighi B.	Mining personal data using smartphones and wearable devices: A survey	2015	0
37	Wang M.-H., Kurozumi K., Kawaguchi M., Lee C.-S., Tsuji H., Tsumoto S.	Healthy diet assessment mechanism based on fuzzy markup language for Japanese food	2014	0
38	Boeva V., Boneva L., Tsiporkova E.	Semantic-aware expert partitioning	2014	0
39	Sebillo M., Tucci M., Tortora G., Vitiello G., Ginige A., De Giovanni P.	Combining personal diaries with territorial intelligence to empower diabetic patients	2014	0

#	Autores	Título do trabalho	Ano de publicação	Citações
40	Hussain S., Kang B.H., Lee S.	A Wearable Device-Based Personalized Big Data Analysis Model	2014	0
41	Di Valentin C., Emrich A., Werth D., Loos P.	Context-sensitive and individualized support of employees in business processes: Conceptual design of a semantic-based recommender system	2014	0
42	Baquero A., Schiffman A.M., Shrager J.	Blend me in: Privacy-preserving input generalization for personalized online services	2013	0
43	Hristoskova A., Tsiporkova E., Tourwe T., Buelens S., Putman M., De Turck F.	A graph-based disambiguation approach for construction of an expert repository from public online sources	2013	0
44	Sohail A., Afzal M.T., Ahmad J.	Discovery of medical experts	2013	0

Fonte: elaborado pelo autor

3.3.3 Identificação dos estudos pré-selecionados e selecionados

A combinação das palavras-chave e dos termos correlatos de Perfil de especialista, *Expertise Location*, Gestão do Conhecimento e Engenharia do Conhecimento já se demonstrava aderente, com base nos títulos e nas palavras-chave encontrados nos estudos na busca da base de dados Scopus, ao objeto de pesquisa deste trabalho para responder à pergunta de como gerar perfis profissionais de forma automática.

No passo de definição dos critérios de inclusão e exclusão de estudos, priorizou-se a seleção dos estudos que já haviam sido citados em trabalhos publicados na base Scopus, o que poderia indicar uma maior relevância desses estudos. Nesse momento, realizou-se uma macroanálise dos resultados obtidos com o uso da técnica de geração de

internet dos estudos resultantes da etapa anterior. Nesse processo, 7 dos 44 estudos resultantes do Quadro 8, ou 16% dos estudos, não foram encontrados integralmente. São os estudos do Quadro 8 de números 30, 32, 33, 35, 38, 40 e 43. Na sua maioria, esses estudos representavam capítulos de livros, aos quais os autores não conseguiram acesso. Além de ser um conjunto pequeno de estudos, esses trabalhos possuíam pouca ou nenhuma citação, indicando mínima relevância.

A leitura cuidadosa e criteriosa dos 37 estudos resultou na seleção de um total de 11 estudos que poderiam contribuir diretamente com os objetos deste trabalho, principalmente ao ajudar a responder à pergunta de como gerar perfis profissionais de forma automática e de se estabelecer um modelo para essa geração. Esses estudos são listados no Quadro 9 e concluem essa etapa da revisão integrativa. Optou-se por manter o mesmo número do estudo para que o leitor possa comparar, quando necessário, os resultados obtidos nessa etapa a partir do Quadro 8.

Quadro 9 – Estudos selecionados na revisão integrativa

#	Autores	Título do trabalho	Ano de publicação
1	Balog K., De Rijke M.	Determining expert profiles (with an application to expert finding)	2007
3	Becerra-Fernandez I.	Searching for experts on the Web: A review of contemporary expertise locator systems	2006
7	Li M., Liu L., Li C.-B.	An approach to expert recommendation based on fuzzy linguistic method and fuzzy text classification in knowledge management systems	2011
8	Macdonald C., Ounis L.	Expertise drift and query expansion in expert search	2007
9	Yang K.-W., Huh S.-Y.	Automatic expert identification using a text categorization technique in knowledge management systems	2008
11	Liu D.-R., Chen Y.-H., Kao W.-C., Wang H.-W.	Integrating expert profile, reputation and link analysis for expert finding in question-answering websites	2013

#	Autores	Título do trabalho	Ano de publicação
12	Yao L., Tang J., Li J.	A unified approach to researcher profiling	2007
15	Yang K.-H., Chen C.-Y., Lee H.-M., Ho J.-M.	EFS: Expert finding system based on wikipedia link pattern analysis	2008
18	Afzal M.T., Maurer H.	Expertise recommender system for scientific community	2011
28	Boeva V., Krusheva M., Tsiporkova E.	Measuring expertise similarity in expert networks	2012
44	Sohail A., Afzal M.T., Ahmad J.	Discovery of medical experts	2013

Fonte: elaborado pelo autor

3.3.4 Categorização dos estudos selecionados

A leitura dos estudos identificados na seleção dos estudos pré-selecionados e selecionados da revisão integrativa contribuiu de maneira significativa para a categorização desses estudos e possibilitou a definição da matriz de síntese. Essa matriz, de acordo com o método de revisão integrativa adotado neste trabalho, é recomendada para a extração e para a organização dos dados dos estudos selecionados.

É importante ressaltar que o resultado da etapa de categorização dos estudos selecionados a partir do método da revisão integrativa da literatura gera uma síntese dos estudos selecionados com o apoio de uma matriz de síntese, conforme explicam os autores Botelho, Cunha e Macedo (2011). Essa matriz é elaborada a partir dos conhecimentos e do senso crítico dos autores da revisão, e tem por objetivo proteger os pesquisadores de erros durante a análise dos estudos selecionados, servindo de ferramenta para interpretação e construção do texto (BOTELHO; CUNHA; MACEDO, 2011).

Dessa forma, o Quadro 10 a seguir relaciona as categorias, os objetivos das categorias e os estudos aderentes que resultaram na matriz de síntese desta etapa da revisão integrativa, conforme exemplo apresentado por Botelho, Cunha e Macedo (2011, p. 131). A análise e a classificação dos estudos são descritas na seção seguinte, a qual discute os resultados aderentes a essas categorias.

Quadro 10 – Matriz de síntese da revisão integrativa com as categorias dos estudos

Categoria	Objetivo da categoria	Estudos aderentes
Conceitos sobre <i>Expertise Location</i> , Gestão do Conhecimento ou Engenharia do Conhecimento	Identificar estudos que apoiem a <i>Expertise Location</i> , e a Gestão do Conhecimento ou a Engenharia do Conhecimento e explicitamente cite seus conceitos	#1, #3, #7, #8, #9, #44
Exploração de documentos não estruturados para utilização no perfil de especialista	Identificar estudos que trabalhem com a exploração de dados e demais informações não estruturadas para a criação de um perfil de especialista	#1, #3, #7, #9, #11, #12, #15
Métodos conceituais e tecnológicos para criação do perfil de especialista	Identificar estudos que apresentem métodos que possibilitem a geração de perfil de especialista e que discutam essas estratégias	#1, #3, #7, #8, #9, #11, #12, #15, #28, #44
Definições de atributos para constituição do perfil de especialista	Identificar estudos que contribuam para a especificação de atributos e de informações que compoñham uma definição de perfil de especialista	#7, #8, #9, #12, #15, #18, #28, #44
Utilização de ontologias ou taxonomias para classificação do conhecimento	Identificar estudos que contribuam para a codificação do conhecimento e se utilizem de ontologias ou taxonomias para essa codificação	#3, #12, #15
Apresentação de protótipos para validação das propostas	Identificar estudos que apresentem protótipos de validação de conceitos para definição de atributos, ou de métodos, ou ainda de tecnologias para a geração de perfil de especialista	#9, #12, #15, #18

Fonte: elaborado pelo autor

4 APRESENTAÇÃO DOS RESULTADOS

Este capítulo tem o intuito de apresentar os resultados gerados com base no objetivo geral desta pesquisa, que é um modelo para a extração do perfil de especialista de forma automática a partir de documentos não estruturados para apoiar o desenvolvimento de soluções de *Expertise Location*. Esse objetivo geral é detalhado em seus objetivos específicos, que visam (a) identificar as abordagens atuais que possibilitem essa extração, (b) definir um modelo conceitual que suporte a extração automática de perfis de especialistas e identifique os elementos necessários para a implementação do modelo e (c) desenvolver um protótipo funcional para se verificar a viabilidade do modelo proposto. Os resultados que contemplam o objetivo específico (a) são detalhados na seção 2.2. Os resultados referentes aos objetivos (b) e (c) podem ser observados, respectivamente, nas seções 4.1 e 4.2 deste capítulo. A seguir, é apresentado o modelo conceitual de extração de perfil de especialista, o qual fornece uma visão geral que guia a implementação do protótipo.

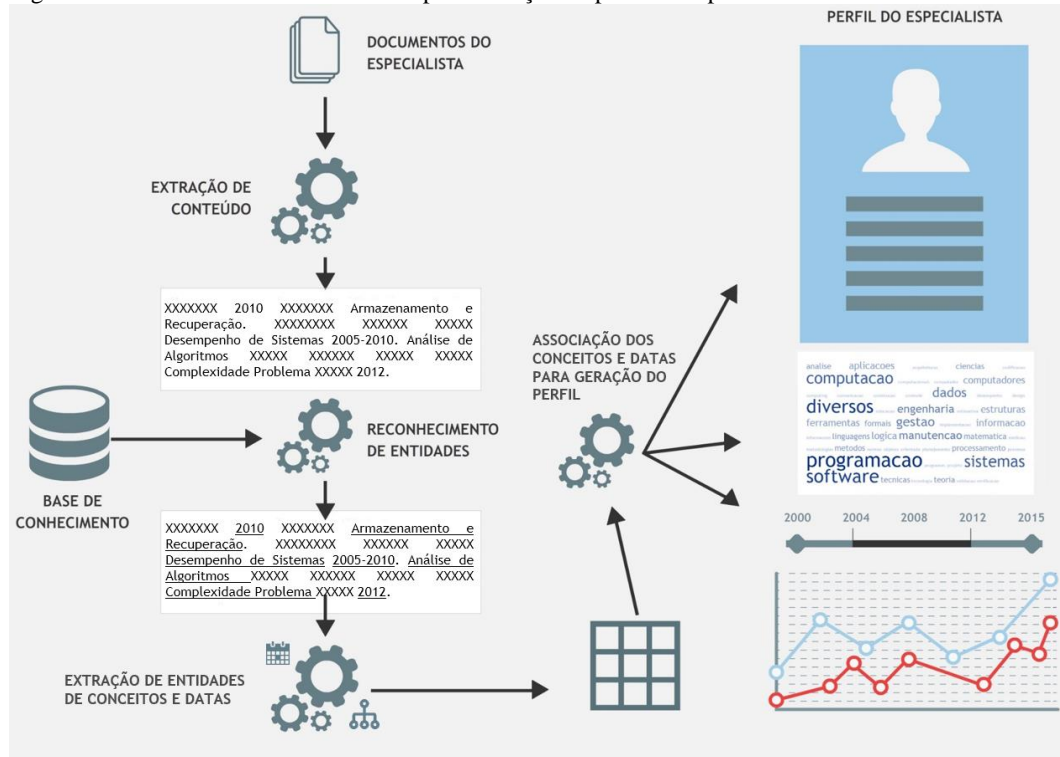
4.1 MODELO DE EXTRAÇÃO DE PERFIL DE ESPECIALISTA PROPOSTO

A proposta de modelo de extração de perfil de especialista foi concebida com base nas abordagens atuais observadas na literatura (seção 2.2). Essas abordagens, conceituais e tecnológicas, que são compostas por processos, técnicas, algoritmos e tecnologias, forneceram subsídios para a definição do modelo proposto. Além disso, as referidas abordagens, principalmente pautadas nos trabalhos realizados por Balog e Rijke (2007), Becerra-Fernandez (2006), Li, Liu e Li (2011) e Yang e Huh (2008) e complementadas pelos trabalhos de Bovo (2011), Ceci (2010) e Gonçalves (2006) permitiram a criação do modelo apresentado nesta seção.

De forma resumida, este modelo proposto espera como entrada os diversos documentos produzidos por um especialista e, com base no conteúdo identificado nesses documentos, cria uma série de instrumentos visuais que exploram as informações neles contidas. A Figura 13 apresenta a visão conceitual do modelo proposto para a extração de perfil de especialista. Esse modelo é composto por cinco etapas: (1) extração do conteúdo, (2) reconhecimento de entidades, (3) extração e armazenamento de entidades, (4) associação de entidades; e (5) extração do perfil de especialista.

Inicialmente, os propósitos e os resultados esperados de cada uma dessas etapas são apresentados de forma resumida e, posteriormente, detalha-se o conteúdo de cada etapa. É importante observar que esse modelo trata as questões conceituais do processamento para a extração do perfil dos especialistas, isto é, trata as etapas como processos com entradas e saídas esperadas. O “como” cada uma dessas etapas faz o seu processamento é detalhado na seção 4.2 de implementação do protótipo. Essa seção 4.2 também considera os resultados da revisão integrativa da literatura (seção 2.2) para a implementação. Dessa forma, o modelo conceitual proposto prevê a criação de componentes separados para cada uma das etapas, de modo a possibilitar a comparação de diferentes tipos de implementação para análise dos resultados a partir de novos estudos futuros.

Figura 13 – Visão conceitual do modelo para extração de perfil de especialista



Fonte: elaborado pelo autor

- **Extração do conteúdo.** Esta etapa tem por objetivo a extração do conteúdo dos documentos em seus diversos formatos (e.g., DOC, DOCX, PPT, PPTX, XLS, XLSX) e que requerem diferentes estratégias para essa tarefa. Espera-se, como resultado, a extração das informações de metainformação do documento, data de criação e data de modificação, além do texto do conteúdo dos documentos. Esses resultados são encaminhados para a etapa de reconhecimento de entidades.
- **Reconhecimento de entidades.** A etapa de reconhecimento de entidades ou reconhecimento de entidades nomeadas (NER) é fundamental para que se identifiquem os conceitos, as pessoas e as datas registradas nos documentos. Como resultado dessa etapa, espera-se um conjunto de entidades nomeadas, classificadas por tipo (i.e., conceitos e datas), com a informação da posição de sua ocorrência no texto do documento.
- **Extração e armazenamento de entidades.** Após a validação das entidades descobertas, esta etapa de extração e armazenamento de entidades de conceitos e datas visa à extração e ao armazenamento dessas informações em tabelas para serem utilizadas no processo de associação. Ao final dessa etapa, registram-se as entidades de conceitos e as entidades de datas em forma estruturada para que seja possível a realização de processos de combinação desses dados na etapa seguinte.
- **Associação de entidades.** Com as informações registradas de forma estruturada dos conceitos e das datas encontrados no processo de extração das entidades, combinadas com as informações dos metadados e com as sentenças dos documentos, esta etapa tem por objetivo associar essas informações para a geração do perfil do especialista. Todas as datas e os conceitos reconhecidos e extraídos são associados por documento para fomentar a etapa de extração do perfil de especialista apresentada a seguir.
- **Extração do perfil de especialista.** A última etapa do modelo proposto tem por objetivo a extração e a geração do perfil de especialista com base nos conceitos e nas datas extraídos e associados. Essa extração deve indicar, também, as respectivas frequências de cada uma dessas entidades, conceitos e datas

para a geração de nuvem de termos e *timeline* de perfil dos conceitos com mais ocorrência nos documentos. Esse processo deve considerar todas as entidades e as datas descobertas nos vários documentos associados e fornecidos pelos especialistas, de modo a agrupar esses dados por ano na apresentação da *timeline*.

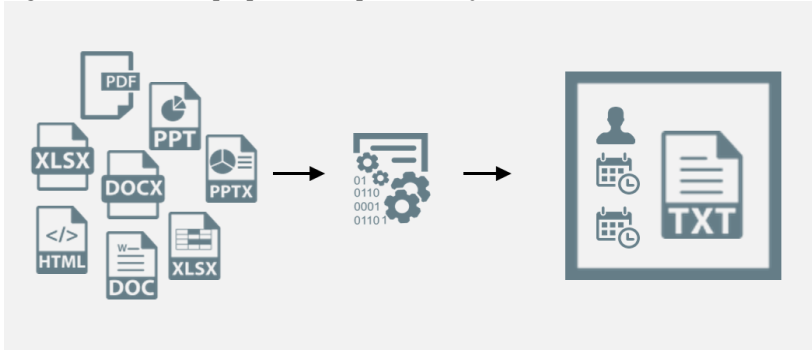
Como observado, cada uma dessas etapas tem uma função específica e depende do resultado de etapas anteriores. Por isso, o modelo proposto considera a modularidade dessas etapas. Assim, essa modularidade deve permitir a criação de novas implementações (e.g., extrator de conteúdo, extrator de entidades) com o passar do tempo para que seja possível experimentar diferentes resultados a partir de novas implementações, porém, respeitando-se as interfaces e os resultados esperados. A seguir, apresenta-se o detalhamento de cada uma dessas etapas deste modelo conceitual.

4.1.1 Extração do conteúdo

A primeira etapa do modelo proposto consiste em extrair o conteúdo dos documentos produzidos pelos especialistas, sejam eles pessoais ou corporativos, que podem ser utilizados no processo de extração do perfil de especialista. Esses documentos existem em uma série de formatos e são reconhecidos por suas extensões, como, por exemplo, DOC, DOCX, PPT, PPTX, XLS, XLSX. Esses são apenas alguns exemplos de formatos de documentos que existem aos milhares (MATTMANN; ZITTING, 2012). Cada formato requer, muitas vezes, um programa específico para a sua leitura.

Entretanto, para que o conteúdo dos documentos em seus diferentes formatos possa ser processado por uma implementação do modelo proposto, é importante que suas informações sejam reconhecidas e que esse conteúdo seja extraído. Isso inclui o conteúdo escrito em linguagem natural e suas metainformações, que são utilizadas nas demais etapas do modelo proposto. São elas: data de criação, data de modificação e conteúdo do documento sem formatação (i.e., texto) e a sua segmentação em sentenças (e.g., frases). A Figura 14 ilustra o processo desta etapa e o resultado esperado. Esse conteúdo será utilizado como entrada na etapa seguinte de reconhecimento de entidades.

Figura 14 – Modelo proposto – etapa de extração do conteúdo



Fonte: elaborado pelo autor

4.1.2 Reconhecimento de entidades

A etapa de reconhecimento de entidades (NER) consiste em identificar e anotar as entidades que existem em documentos. Essas entidades são categorizadas como pessoas, organizações, conceitos, localizações, expressões de tempos, entre outros (BALOG; RIJKE, 2007; CECI, 2010; NEGRI; MAGNINI, 2004). Diferentes abordagens são identificadas na literatura e podem utilizar estratégias de modelos estatísticos (CECI, 2010; GONCALVES, 2006), de modelos linguísticos (BALOG et al., 2012; BALOG; RIJKE, 2007) ou modelos que se utilizam de técnicas de *Machine Learning* (BALOG et al., 2012).

O resultado esperado desta etapa são as anotações no conteúdo do documento sobre conceitos e datas e a identificação da posição do documento em que essas anotações foram feitas, independentemente da estratégia utilizada. O Quadro 11 ilustra como ficariam as anotações no conteúdo do documento quando identificadas as entidades dos tipos conceitos e datas. Esse exemplo teve como base um resumo retirado de um currículo de um pesquisador que integra o EGC e serve apenas para ilustrar o exemplo de anotação. O resultado dessa etapa é utilizado como entrada na etapa seguinte de extração e armazenamento das entidades de conceitos e datas.

Quadro 11 – Exemplo de anotações de conceitos e datas da etapa de reconhecimento de entidades

Possui bacharelado em Ciência da Computação pela Universidade do Vale do Itajaí (1997), mestrado (2001) e doutorado (2006) em Engenharia de Produção pela Universidade Federal de Santa Catarina, com estágio de doutoramento na The Open University. Atualmente é professor no Departamento de Administração Pública da Universidade do Estado de Santa Catarina e no Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento na Universidade Federal de Santa Catarina. Atua também como diretor e pesquisador no Instituto Stela. Atuou como pesquisador e coordenou diversos projetos de P&D com organizações públicas (como a Plataforma Lattes e a Plataforma Aquarius com o MCTI, Portal SINAES com o MEC, DCVISA com a ANVISA e SIBEA com o MMA), organizações privadas (como Busca Semântica com Embraer e Plataforma de Gestão da Ética e da Integridade com o Itaú/Unibanco) e terceiro setor (como a Plataforma para Gestão do Absenteísmo com o SESI/BA e Plataforma da Gestão do Conhecimento do SENAI/CE-FIEC). Tem experiência na área de Ciência da Computação, atuando principalmente nos seguintes temas: engenharia do conhecimento, gestão do conhecimento, data warehouse, business intelligence, ontologias, semantic web, planejamento de sistemas de informações e governo eletrônico.

Fonte: elaborado pelo autor

4.1.3 Extração e armazenamento de entidades

A etapa de extração e armazenamento de entidades tem por objetivo a coleta e a persistência das entidades identificadas. Essa persistência pode ocorrer, por exemplo, em tabelas de dados e deve permitir a realização de consultas posteriormente. Esse armazenamento possibilita a realização da etapa de análise para que ocorra a associação entre os conceitos e as datas na etapa descrita na próxima seção.

Os tipos de entidades nomeadas, objeto dessa etapa de extração e armazenamento, são as entidades do tipo conceitos e datas, conforme a classificação da anotação ocorrida na etapa de reconhecimento de entidades deste modelo. Além disso, as informações de posição das entidades, em relação ao texto e às sentenças do texto, também devem ser coletadas para gerar insumos à próxima etapa do modelo, denominada associação dos conceitos. O Quadro 12 a ilustra como seria o armazenamento dessas entidades com os seus respectivos atributos. Esse exemplo simula uma simples tabela de dados com esse conteúdo.

Quadro 12 – Exemplo de armazenamento de anotações da etapa de extração e armazenamento

Entidade	Tipo da entidade	Posição inicial (#)	Posição final (#)	Sentença (#)
bacharelado	Conceito	7	18	1
Ciência da Computação	Conceito	21	43	1
Universidade do Vale do Itajaí	Conceito/ organização	49	80	1
1997	Data	81	85	1
mestrado	Conceito	88	96	1
2001	Data	98	102	1
doutorado	Conceito	106	115	1
2006	Data	117	121	1
Engenharia de Produção	Conceito	126	148	1
Universidade Federal de Santa Catarina	Conceito/ organização	154	192	1
The Open University	Conceito/ organização	225	244	1
professor	Conceito	259	268	2
Administração Pública	Conceito	288	310	2

Fonte: elaborado pelo autor

4.1.4 Associação de entidades

A etapa de associação de entidades tem o seu foco na criação de relações entre as entidades extraídas e armazenadas dos tipos conceitos

e datas. Essas relações devem ser geradas quando as entidades estão presentes em um mesmo documento. A partir dessas relações, gera-se um mapa das informações e das associações por especialista, documento, data de modificação do documento e respectivas entidades coletadas. Com esses mapas, é possível gerar insumos para a extração do perfil de especialista descrito na etapa seguinte desse modelo. O resultado é ilustrado no Quadro 13 a seguir e exemplifica um mapa com o conteúdo extraído de três documentos e que agrupa os conceitos e as datas por frequência.

Quadro 13 – Exemplo de mapa do resultado da etapa de associação de entidades

Documento	Data do documento	Entidade	Frequência
Documento 1	01/03/2015	Plataforma Lattes	4
Documento 1	01/03/2015	Sistemas de informação	3
Documento 2	02/07/2014	Engenharia do Conhecimento	9
Documento 2	02/07/2014	Processamento de Linguagem Natural	4
Documento 2	02/07/2014	Gestão do Conhecimento	10
Documento 3	27/03/2013	Arquitetura de sistemas	3
Documento 3	27/03/2013	2012	6
Documento 3	27/03/2013	Governo eletrônico	5
Documento 3	27/03/2013	Ontologia	7

Fonte: elaborado pelo autor

4.1.5 Extração do perfil de especialista

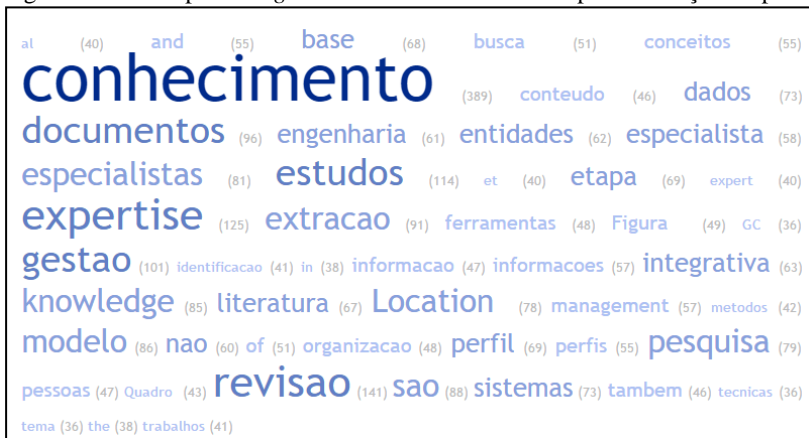
A partir das informações obtidas na etapa de associação das entidades, que gera mapas com as entidades de conceitos, entidades de datas e informações dos documentos, esta última etapa do modelo tem por objetivo a criação de instrumentos visuais que permitam rapidamente identificar as entidades às quais o especialista está

vinculado e que estão registradas nos seus documentos. Essas entidades podem indicar os conhecimentos tácito e explícito que o especialista adquiriu ao longo dos anos bem como os seus interesses (BALOG; RIJKE, 2007; BECERRA-FERNANDEZ, 2006; KRUSHEVA; TSIPORKOVA, 2012; LI; LIU; LI, 2011; LIU et al.; 2013; YANG; HUH, 2008).

É importante observar que a qualidade do resultado do conteúdo dos instrumentos visuais está diretamente relacionada aos documentos processados por uma implementação deste modelo. Isto é, se os documentos não forem realmente do especialista ou se não representarem de fato os seus interesses (i.e., documentos baixados, mas nunca lidos), poderá existir uma distorção no conteúdo gerado. Desse modo, é importante selecionar de forma adequada o conteúdo que será processado.

Espera-se que o resultado desta etapa gere instrumentos visuais como nuvem de termos e *timeline* a partir dos conceitos mais frequentes presentes nos documentos do especialista. A nuvem de termos (i.e., *tag cloud*) possibilita, por exemplo, a visualização dos conceitos mais frequentes nos documentos do especialista de forma global. Já a *timeline* também possibilita a representação dos conceitos mais frequentes nos documentos do especialista, só que organizado por datas. Assim, a *timeline* permite a observação das evoluções e das tendências de conceitos aos quais um especialista está associado. Novos instrumentos que possibilitem uma visão prática do conteúdo gerado poderão ser agregados como melhorias dessa etapa do modelo. Os resultados esperados nesse momento são ilustrados na Figura 15 e na Figura 16 a seguir. O primeiro exemplo foi criado a partir do texto deste documento (i.e., conteúdo desta dissertação), enquanto o segundo exemplo foi elaborado com dados aleatórios.

Figura 15 – Exemplo de *tag cloud* como resultado da etapa de extração do perfil



Fonte: elaborado pelo autor com o uso da ferramenta TagCrowd, disponível em: <<http://tagcrowd.com>>.

Figura 16 – Exemplo de *timeline* como resultado da etapa de extração do perfil

2013	2014	2015	2016
Gestão do conhecimento (15)	Engenharia do conhecimento(20)	Expertise Location (28)	Processamento de linguagem natural(28)
Administração (10)	Gestão do conhecimento (19)	Engenharia do conhecimento (15)	Engenharia do conhecimento (20)
Conhecimento (5)	Administração (5)	Gestão do conhecimento (6)	Inteligência artificial (10)

Fonte: elaborado pelo autor

4.2 PROTÓTIPO DA IMPLEMENTAÇÃO DO MODELO PROPOSTO

O modelo proposto nesta pesquisa foi concebido a partir das abordagens identificadas na literatura que possibilitam a extração e a criação automática do perfil. Este modelo, apresentado na seção 4.1, especifica etapas para o processamento de documentos de especialistas com vistas à extração de conceitos e à apresentação dessas informações em instrumentos visuais como *tag cloud* e *timeline*.

As definições das etapas do modelo proposto descrevem as entradas e saídas esperadas por cada uma delas, bem como o seu propósito. Porém, fica sob responsabilidade da implementação a forma de como fazer para processar e gerar essas entradas e saídas. Nesse sentido, o modelo proposto está em nível conceitual e permite a sua replicação com o emprego de outras técnicas não abordadas neste trabalho, quando necessário.

Para fins de análise do modelo proposto e para ilustrar uma estratégia de implementação, foi estabelecido um protótipo baseado nas abordagens de Processamento de Linguagem Natural (PLN) observadas nos trabalhos de Becerra-Fernandez (2006) e de Balog e Rijke (2007). O PLN envolve, por exemplo, técnicas de tokenização, de normalização, de remoção de *stopwords*, *lemmatization*, *stemming*, *Part-Of-Speech Tagging* e *Named Entity Recognition*. Também foram utilizadas a associação temporal observada no trabalho de Bovo (2011) e a Wikipédia como dicionário de dados para o reconhecimento e a validação dos conceitos encontrados nos documentos na etapa de Reconhecimento de Entidades (NER), conforme observado nos trabalhos de Ceci (2010) e de Yang et al. (2007).

A partir desse contexto, as próximas seções apresentam os passos para a construção do protótipo e seus resultados em termos de aplicação. Dessa forma, detalham-se a arquitetura do protótipo, os componentes PLN adotados, a implementação de cada uma das etapas do modelo proposto e, por último, os resultados do protótipo materializado em uma aplicação que está disponível como projeto de código aberto na internet em <http://rudgertaxweiler.github.io/expert-profiling>.

4.2.1 Arquitetura do protótipo

A arquitetura é uma descrição das estruturas de sistemas de software que apresenta, de forma global, as decisões de projetos em termos de abstrações adotadas para a construção de um sistema em relação às suas camadas, responsabilidades e processos (DOMINGOS, 2004; SELL, 2005). Assim, a arquitetura deste protótipo teve como base o modelo conceitual proposto na seção 4.1 e a utilização das bibliotecas de PLN existentes na linguagem de programação Java. Essa linguagem foi escolhida por ser de domínio do autor deste trabalho e também por possuir uma grande variedade de bibliotecas existentes que suportam as tarefas de PLN.

Até que fosse definida a arquitetura do protótipo, foram executados testes em diferentes bibliotecas de PLN para identificar qual a melhor tecnologia para a implementação das etapas descritas no

modelo. Esses testes foram realizados porque as tecnologias de base para implementação podem interferir na arquitetura elaborada, uma vez que a tecnologia também define os elementos arquiteturais necessários. Nesse contexto, esses testes ocorreram com as ferramentas Stanford CoreNLP¹³, Apache OpenNLP¹⁴ e Apache UIMA¹⁵, principalmente no tocante à transformação de documentos em sentenças e à criação de anotações de NER.

Esses testes, porém, não foram satisfatórios. Trabalhar com PLN requer, muitas vezes, um conjunto misto de transformações que necessitam de diferentes componentes de ferramentas distintas que possam ser combinadas para um melhor resultado (e.g., o tokenizador do Stanford CoreNLP e o NER do Apache OpenNLP). Os *pipelines* do Apache UIMA, isto é, o processamento segmentado de diferentes anotadores guiados por uma configuração, também não ofereciam praticidade de configuração por requererem a configuração em arquivos XMLs. Consequentemente, os arquivos de configuração não dispunham de compilador para indicar erros de configuração, sendo necessário rodar as ferramentas para identificar possíveis erros nessas configurações.

A partir de uma busca por estudos sobre as dificuldades encontradas nesses testes preliminares, encontrou-se o trabalho de Castilho e Gurevych (2014), da Universidade Tecnológica de Darmstadt, na Alemanha. Esse trabalho relata o framework DKPro¹⁶, que aglutina uma coleção das mais diversas bibliotecas de PLN disponíveis atualmente, incluindo as testadas nessa fase preliminar (i.e., Stanford CoreNLP e Apache OpenNLP). O DKPro também utiliza o Apache UIMA, mas disponibiliza uma biblioteca chamada Apache UIMA Fit que possibilita a construção de *pipelines* de anotadores de PLN sem a necessidade de configuração de arquivos XML, o que torna as configurações dos anotadores mais simples. A questão do emprego de diferentes bibliotecas em conjunto também foi trabalhada pelo DKPro

¹³ Stanford CoreNLP fornece um conjunto de ferramentas de análise de linguagem natural. Disponível em: <<http://stanfordnlp.github.io/CoreNLP>>.

¹⁴ Apache OpenNLP fornece um conjunto de ferramentas baseadas em aprendizagem de máquina para o processamento de texto de linguagem natural. Disponível em: <<https://opennlp.apache.org>>.

¹⁵ Apache UIMA fornece frameworks, ferramentas e anotadores para facilitar a análise de conteúdos não estruturados, tais como texto, áudio e vídeo. Disponível em: <<https://uima.apache.org>>.

¹⁶ DKPro fornece um conjunto de componentes adaptados ao Apache UIMA para que possam ser usados alternadamente *pipelines*.

com a criação de *wrappers* das bibliotecas para que possam ser utilizadas em um mesmo processo de *pipeline*. A título de demonstração da variedade de bibliotecas presentes do DKPro, o Quadro 14 a seguir apresenta a lista de componentes disponíveis, em quais tarefas esses componentes atuam e as linguagens suportadas.

Quadro 14 – Lista de tarefas e componentes disponíveis no DKPro

Tarefa	Componentes (#)	Linguagens
Identificação de linguagem	2	Português, inglês, espanhol, alemão, francês e mais 64 linguagens
Tokenização e detector de sentenças	5	Português, inglês, espanhol, alemão, francês e mais 24 linguagens
Lemmatization	7	Inglês e alemão
Stemming	1	Português, inglês, espanhol, alemão, francês e mais 10 linguagens
Part-of-speech tagging	9	Português, inglês, espanhol, alemão, francês e mais 13 linguagens
Análise morfológica	2	Inglês, alemão, francês, italiano e mais uma língua
Reconhecimento de entidades nomeadas	2	Inglês, alemão, espanhol e holandês
Chunking	1	Inglês
Parser de circunscrição	3	Inglês, alemão, francês, chinês e mais uma língua
Parser de dependência	5	Inglês, alemão, francês e mais 8 línguas

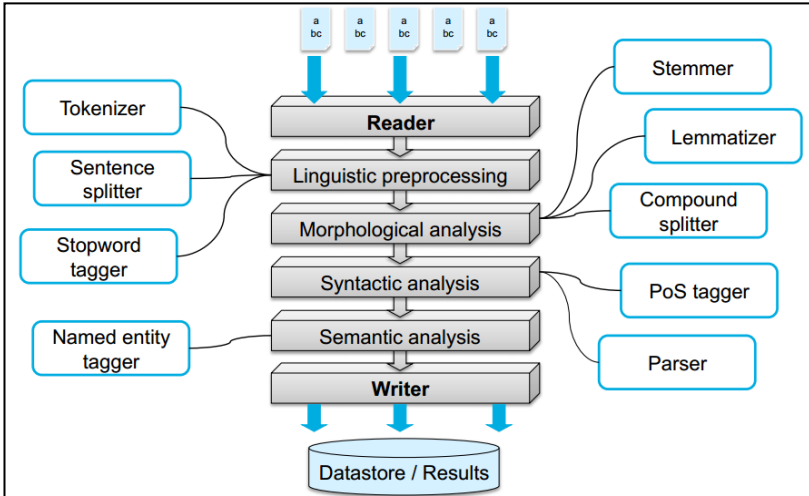
Tarefa	Componentes (#)	Linguagens
Análise de correferência	1	Inglês
Rotulagem semântica	1	Inglês
Verificação e gramática verificação ortográfica	3	Português, inglês, espanhol, alemão, francês e mais 24 linguagens

Fonte: adaptado de Castilho e Gurevych (2014, p. 7)

Por essas razões, definiu-se, então, o DKPro como o framework-base para a implementação do protótipo do modelo proposto. O DKPro possibilita uma maior extensibilidade e configurabilidade na criação de anotadores de forma desacoplada.

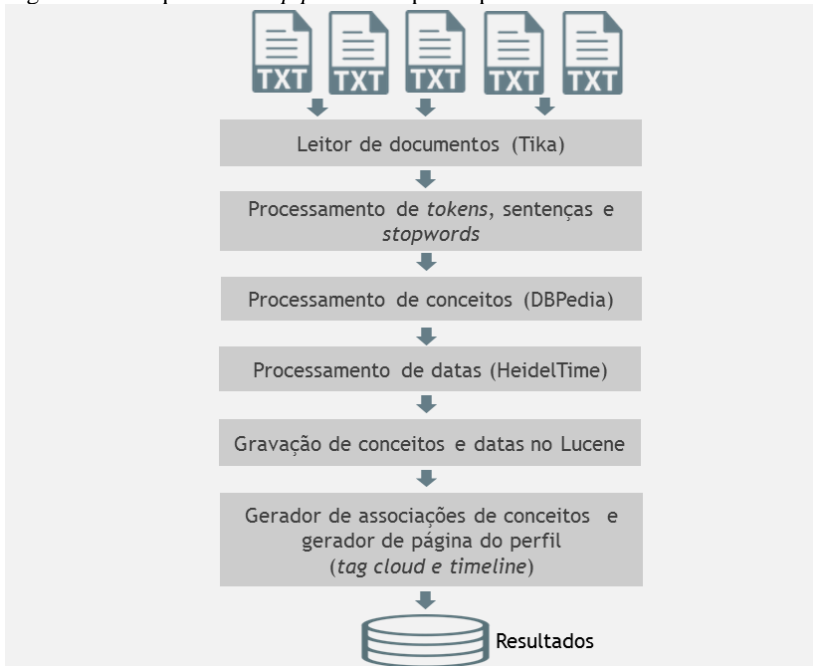
O modelo utilizado pelo DKPro também orientou a arquitetura do protótipo por definir uma arquitetura de *pipelines* baseada no Apache UIMA. Portanto, a arquitetura do protótipo também possui um *pipeline* e é composta por: um leitor de documentos, um anotador de *tokens*, um anotador de sentenças, um anotador de *stopwords*, um anotador de entidades de conceitos, um anotador de entidades de datas, um escritor de anotações e um gerador de arquivos. Para fins de comparação, apresentam-se a Figura 17 e a Figura 18, que demonstram a arquitetura do DKPro e a arquitetura adotada na implementação do protótipo, respectivamente. Os componentes de *pipelines* do modelo proposto são descritos nas próximas seções.

Figura 17 – Arquitetura de *pipelines* do DKPro e seus respectivos componentes



Fonte: Castilho (2013)

Figura 18 – Arquitetura de *pipelines* do protótipo do modelo



Fonte: elaborado pelo autor

4.2.2 Implementação da etapa de extração do conteúdo

A implementação do protótipo quanto à etapa de extração de conteúdo corresponde à implementação de dois elementos da Figura 18 da arquitetura do protótipo: (a) “Leitor de documentos (Tika)” e (b) “Processamento de *tokens*, sentenças e *stopwords*”. Esta etapa de extração de conteúdo é responsável pela leitura do conteúdo e das metainformações dos documentos de um especialista. Após a leitura dessas informações, as sentenças, os *tokens* (e.g., sequência de caracteres que possam formar uma palavra) e as respectivas *stopwords* presentes no conteúdo do documento são anotados.

No que tange à implementação, apesar de o DKPro oferecer uma quantidade grande de componentes, não foi observado um componente de leitura de documentos que conseguisse ler os diferentes formatos de documentos como DOC, DOCX, XLS, XLS, PPT e PPTX. São disponibilizados apenas leitores de textos planos e PDFs. A partir dessa constatação, implementou-se um novo leitor de documentos utilizando a biblioteca Apache Tika¹⁷, que atende aos requisitos da etapa de extração do conteúdo do modelo proposto. O Apache Tika consegue identificar qual tipo de documento é fornecido pelo especialista, além de extrair apropriadamente o conteúdo textual e o conteúdo das metainformações. Essa implementação conta ainda com opções de normalização de texto, que possibilitam a extração do conteúdo dos documentos em caixa-baixa e sem acentos. Essas opções são parametrizáveis.

Após a criação do leitor de documentos com o Apache Tika, utilizou-se o componente do DKPro chamado “OpenNlpSegmenter”, *wrapper* do Apache OpenNLP, para a transformação dos textos em sentenças e *tokens* de acordo com a pontuação encontrada no texto. Essa implementação considera, por exemplo, até mesmo números que contenham pontuação, identificando corretamente os limites de uma sentença.

Por fim, implementou-se também o anotador de *stopwords*, que identifica e anota as *stopwords* nas línguas portuguesa e inglesa declaradas a partir de um dicionário (i.e., lista simples de palavras). Essa implementação foi necessária porque o componente-padrão fornecido pelo DKPro apenas excluiu as palavras-chave e não as anotava.

Em resumo, implementou-se o TikaReader para leitura de documentos (Figura 19 e Figura 20). Também se utilizou o

¹⁷ Apache Tika – uma biblioteca para leitura de textos e metadados de diferentes tipos de documentos. Disponível em: <<https://tika.apache.org>>.

OpenNlpSegmenter para as anotações de sentenças e *tokens* (Figura 21 e Figura 22). E, por fim, implementou-se o componente StopWordAnnotator para anotar as *stopwords* a partir de dicionários (Figura 23). As figuras apresentadas a seguir demonstram os resultados e as anotações obtidas com cada um dos componentes. Essas figuras foram capturadas com a utilização da ferramenta Annotation Viewer, do pacote Apache UIMA.

Figura 19 – Resultados da etapa de extração do conteúdo – anotações de metadados do documento

The screenshot displays the Apache UIMA Annotation Viewer interface. The main window shows the text of a document with several segments highlighted in red, indicating metadata annotations. The text is: "possui bacharelado em ciencias da computacao pela universidade do vale do itajai (1997), mestrado (2001) e doutorado (2006) em engenharia de producao pela universidade federal de santa catarina, com estagio de doutoramento na the open university. atualmente e professor no departamento de administracao publica da universidade do estado de santa catarina e no programa de pos-graduacao em engenharia e gestao do conhecimento na universidade federal de santa catarina. atua tambem como diretor e pesquisador no instituto stela. atuou como pesquisador e coordenou diversos projetos de p&d com organizacoes publicas (como a plataforma lattes e a plataforma aquarius com o mcti, portal sinaes com o mec, dcvisa com a anvisa e sibeia com o mma), organizacoes privadas (como busca semantica com embraer e plataforma de gestao da etica e da integridade com o itau/unibanco) e terceiro setor (como a plataforma para gestao do absenteismo com o sesi/ba e plataforma da gestao do conhecimento do senai/ce-fiec). tem experiencia na area de ciencia da computacao, atuando principalmente nos seguintes temas: engenharia do conhecimento, gestao do conhecimento, data warehouse, business intelligence, ontologias, semantic web, planejamento de sistemas de informacoes e governo eletronico."

At the bottom left, the "Annotation Types" section shows a checked box for "DocumentMetaData". On the right side, a panel titled "Click In Text to See Annotation Detail" shows a tree view of the annotations. The selected annotation is "DocumentMetaData", which has the following properties:

- begin = 0
- end = 1274
- language = pt
- documentTitle = Example3.txt
- documentId = Example3.txt
- documentUri = file:/C:/Users/Rudger/git/expert-profiling/c
- collectionId = file:/C:/Users/Rudger/git/expert-profiling/co
- documentBaseUri = file:/C:/Users/Rudger/git/expert-profi
- isLastSegment = false

Fonte: elaborado pelo autor com o uso da ferramenta Apache UIMA Annotation Viewer

Figura 20 – Resultados da etapa de extração do conteúdo – anotações de metadados complementares

The screenshot displays the Apache UIMA Annotation Viewer interface. The main window shows the document text with highlighted annotations. The right-hand pane, titled "Click In Text to See Annotation Detail", provides a hierarchical view of the selected annotations, showing their metadata.

Document Text:

possui bacharelado em ciencias da computacao pela universidade do vale do itajai (1997), mestrado (2001) e doutorado (2006) em engenharia de producao pela universidade federal de santa catarina, com estagio de doutoramento na the open university. atualmente e professor no departamento de administracao publica da universidade do estado de santa catarina e no programa de pos-graduacao em engenharia e gestao do conhecimento na universidade federal de santa catarina. atua tambem como diretor e pesquisador no instituto stela. atuou como pesquisador e coordenou diversos projetos de p&sd com organizacoes publicas (como a plataforma lattes e a plataforma aquarius com o mcti, portal sinaes com o mec, dcvisa com a anvisa e sibeia com o mma), organizacoes privadas (como busca semantica com embraer e plataforma de gestao da etica e da integridade com o itau/unibanco) e terceiro setor (como a plataforma para gestao do absenteismo com o sesi/ba e plataforma da gestao do conhecimento do senai/ce-fiec). tem experiencia na area de ciencia da computacao, atuando principalmente nos seguintes temas: engenharia do conhecimento, gestao do conhecimento, data warehouse, business intelligence, ontologias, semantic web, planejamento de sistemas de informacoes e governo eletronico.

Annotation Detail:

- Annotations
 - MetadataStringField
 - MetadataStringField ("possui bacharelado em ciencias da comp
 - begin = 0
 - end = 1274
 - key = Content-Type
 - value = text/plain; charset=UTF-8
 - MetadataStringField ("possui bacharelado em ciencias da comp
 - begin = 0
 - end = 1274
 - key = Content-Encoding
 - value = UTF-8
 - MetadataStringField ("possui bacharelado em ciencias da comp
 - begin = 0
 - end = 1274
 - key = X-Parsed-By
 - value = org.apache.tika.parser.DefaultParser

Annotation Types:

- MetadataStringField

Fonte: elaborado pelo autor com o uso da ferramenta Apache UIMA Annotation Viewer

Figura 21 – Resultados da etapa de extração do conteúdo – anotações de sentenças

The screenshot displays the Apache UIMA Annotation Viewer interface. The main window title is "Annotation Results for Example3.txt.xml in C:\Users\Rudger\AppData\Local\Temp\expert-profiling\document_annotation_indexanalysis\result".

The main text area contains a paragraph of text with several sentences highlighted in blue. The highlighted text is: "possui bacharelado em ciencias da computacao pela universidade do vale do itajai (1997), mestrado (2001) e doutorado (2006) em engenharia de producao pela universidade federal de santa catarina, com estagio de doutoramento na the open university, atualmente e professor no departamento de administracao publica da universidade do estado de santa catarina e no programa de pos-graduacao em engenharia e gestao do conhecimento na universidade federal de santa catarina. atua tambem como diretor e pesquisador no instituto stela. atuou como pesquisador e coordenou diversos projetos de p&d com organizacoes publicas (como a plataforma lattes e a plataforma aquarius com o mcti, portal sinaes com o mec, dcvisa com a anvisa e sibeia com o mma), organizacoes privadas (como busca semantica com embraer e plataforma de gestao da etica e da integridade com o itau/unibanco) e terceiro setor (como a plataforma para gestao do absenteismo com o sesi/ba e plataforma da gestao do conhecimento do senai/ce-fiec). tem experiencia na area de ciencia da computacao, atuando principalmente nos seguintes temas: engenharia do conhecimento, gestao do conhecimento, data warehouse, business intelligence, ontologias, semantic web, planejamento de sistemas de informacoes e governo eletronico."

At the bottom left, there is a section titled "Annotation Types" with a checkbox labeled "Sentence" that is checked.

On the right side, there is a panel titled "Click In Text to See Annotation Detail". It shows a tree view of annotations:

- Annotations
 - Sentence
 - Sentence ("possui bacharelado em ciencias da computacao pel...")
 - begin = 0
 - end = 246

Fonte: elaborado pelo autor com o uso da ferramenta Apache UIMA Annotation Viewer

Figura 22 – Resultados da etapa de extração do conteúdo – anotações de tokens

The screenshot displays the Apache UIMA Annotation Viewer interface. The main window title is "Annotation Results for Example3.txt.xml in C:\Users\Rudger\AppData\Local\Temp\expert-profiling\document_annotation_indexanalysis\result". The main text area contains a paragraph of text with several words highlighted in pink, indicating token annotations. The highlighted words are: "possui", "bacharelado", "em", "ciencias", "da", "computacao", "pela", "universidade", "do", "vale", "do", "itajai", "1997", "mestrado", "2001", "e", "doutorado", "2006", "em", "engenharia", "de", "producao", "pela", "universidade", "federal", "de", "santa", "catarina", "com", "estagio", "de", "doutoramento", "na", "the", "open", "university", "atualmente", "e", "professor", "no", "departamento", "de", "administracao", "publica", "da", "universidade", "do", "estado", "de", "santa", "catarina", "e", "no", "programa", "de", "pos-graduacao", "em", "engenharia", "e", "gestao", "do", "conhecimento", "na", "universidade", "federal", "de", "santa", "catarina", "atua", "tambem", "como", "diretor", "e", "pesquisador", "no", "instituto", "stela", "atuou", "como", "pesquisador", "e", "coordenou", "diversos", "projetos", "de", "p&d", "com", "organizacaoes", "publicas", "como", "a", "plataforma", "lattes", "e", "a", "plataforma", "aquarius", "com", "o", "mcti", "portal", "sinaes", "com", "o", "mec", "divisa", "com", "a", "anvisa", "e", "sibea", "com", "o", "mma", "organizacaoes", "privadas", "como", "busca", "semantica", "com", "embraer", "e", "plataforma", "de", "gestao", "da", "etica", "e", "da", "integridade", "com", "o", "itau/unibanco", "e", "terceiro", "setor", "como", "a", "plataforma", "para", "gestao", "do", "absenteismo", "com", "o", "sesi/ba", "e", "plataforma", "da", "gestao", "do", "conhecimento", "do", "senai/ce-fiec", "tem", "experiencia", "na", "area", "de", "ciencia", "da", "computacao", "atuando", "principalmente", "nos", "seguintes", "temas", "engenharia", "do", "conhecimento", "gestao", "do", "conhecimento", "data", "warehouse", "business", "intelligence", "ontologias", "semantic", "web", "planejamento", "de", "sistemas", "de", "informacoes", "e", "governo", "eletronico".

At the bottom left, the "Annotation Types" section shows a checkbox for "Token" which is checked. On the right side, there is a panel titled "Click In Text to See Annotation Detail" showing a tree view of annotations. The tree structure is as follows:

- Annotations
 - Token
 - Token ("possui")
 - begin = 0
 - end = 6

Fonte: elaborado pelo autor com o uso da ferramenta Apache UIMA Annotation Viewer

Figura 23 – Resultados da etapa de extração do conteúdo – anotações de stopwords

The screenshot displays the Apache UIMA Annotation Viewer interface. The main window shows a text document with several words highlighted in green, indicating they are stopwords. The text is: "possi bacharelado em ciencias da computacao pela universidade do vale do itajai (1997), mestrado (2001) e doutorado (2006) em engenharia de producao pela universidade federal de santa catarina, com estagio de doutoramento na the open university. atualmente e professor no departamento de administracao publica da universidade do estado de santa catarina e no programa de pos-graduacao em engenharia e gestao do conhecimento na universidade federal de santa catarina. atua tambem como diretor e pesquisador no instituto stela. atuou como pesquisador e coordenou diversos projetos de p&d com organizacoes publicas (como a plataforma lattes e a plataforma aquarius com o mcti, portal sinaes com o mec, dcvisa com a anvisa e sibeia com o mma), organizacoes privadas (como busca semantica com embraer e plataforma de gestao da etica e da integridade com o itau/unibanco) e terceiro setor (como a plataforma para gestao do absenteismo com o sesi/ba e plataforma da gestao do conhecimento do senai/ce-fiec). tem experiencia na area de ciencia da computacao, atuando principalmente nos seguintes temas: engenharia do conhecimento, gestao do conhecimento, data warehouse, business intelligence, ontologias, semantic web, planejamento de sistemas de informacoes e governo eletronico."

The right-hand pane, titled "Click In Text to See Annotation Detail", shows a tree view of the annotations. It includes a "StopWord" annotation with a "begin" value of 19 and an "end" value of 21. The "Annotation Types" section at the bottom left shows "StopWord" selected with a checked checkbox.

Fonte: elaborado pelo autor com o uso da ferramenta Apache UIMA Annotation Viewer

4.2.3 Implementação da etapa de reconhecimento de entidades

A etapa de reconhecimento de entidades na implementação do protótipo engloba os elementos (a) “Processamento de conceitos (DBPedia)” e (b) “Processamento de datas (HeidelTime)”, que são apresentados na Figura 18 da arquitetura do protótipo. É nesta etapa que o processamento de entidades nomeadas (NER) é realizado, e são esperadas, como resultado, as anotações de entidades do tipo conceito e do tipo data, conforme aponta o modelo conceitual (4.1.2).

Diferentes abordagens podem ser adotadas na implementação do reconhecimento de entidades, conforme prevê o modelo conceitual. Os componentes presentes no DKPro que possuem o recurso de NER são `OpenNlpNamedEntityRecognizer` e `StanfordNamedEntityRecognizer`, da `ApacheNLP` e da `Stanford CoreNLP`, respectivamente.

Enquanto o componente do `ApacheNLP` possui modelos apenas para reconhecimento de entidades do tipo pessoa, o componente do `Stanford CoreNLP` não possui modelos para reconhecimento de entidades em língua portuguesa. Por esse motivo, foi necessário implementar um novo anotador de entidades específico para o desenvolvimento do protótipo, que efetua as anotações de entidade do tipo conceitos.

Em relação ao processo de anotações do tipo data, identificou-se na literatura um componente chamado `HeidelTime` (STROTGEN; GERTZ, 2010) que anota o conteúdo de datas em 13 línguas diferentes, incluindo português e inglês. Essa anotação de datas é realizada por meio de reconhecimento de expressões regulares específicas para cada língua. Para utilizar esse componente, bastou implementar um adaptador de anotações de sentenças e *tokens* para incorporá-lo ao uso do DKPro. Originalmente, o `HeidelTime` está disponível apenas para o `Apache UIMA`, e não para o DKPro. Essa etapa de implementação é chamada de *wrapper*, e o componente resultante dessa adaptação foi chamado de `HeidelTimeWrapper`.

No tocante à implementação do anotador de entidades do tipo conceito, tentou-se utilizar o componente do DKPro, chamado `DictionaryAnnotator`, antes de se implementar um novo anotador. Esse anotador varre as sentenças de um documento e faz anotações conforme os termos identificados em um dicionário de dados (i.e., uma lista de conjunto de termos). Internamente, o componente `DictionaryAnnotator` utiliza o algoritmo baseado em n-grama, que combina termos de uma sentença para validá-los no dicionário de dados carregado.

Um n-grama é uma sequência contínua de n itens de uma dada sequência de texto. Esses itens podem ser fonemas, sílabas, letras ou palavras (BRODER et al., 1997). Esse modelo de n-grama é amplamente utilizado nas áreas de probabilidade, teoria da comunicação, linguística computacional (processamento estatístico da linguagem natural), entre outras.

Contudo, o problema observado com o componente DictionaryAnnotator é que os dados contidos em um arquivo de dicionário são carregados em memória após a sua leitura, o que demanda uma grande quantidade de memória disponível quando existem dicionários mais robustos e completos. Após essa constatação, optou-se por criar uma nova implementação, denominada DbpediaAnnotator, a qual possui três características fundamentais: (1) a utilização do Apache Lucene¹⁸; (2) a utilização do dicionário de conceitos da DBPedia¹⁹; e (3) a utilização do modelo de conhecimento no formato SKOS²⁰.

O Apache Lucene foi utilizado no DictionaryAnnotator para sanar o problema de memória na utilização de grandes dicionários. A DBPedia foi empregada na obtenção dos conceitos para anotação dos documentos, conforme observado nos trabalhos de Ceci (2010) e de Yang et al. (2007). Por fim, o modelo SKOS foi utilizado por fornecer uma estrutura taxonômica dos conceitos, o que possibilita, por exemplo, uma desambiguação de conceitos identificados nos textos em uma possível evolução dessa implementação do protótipo. É importante destacar que a DBPedia disponibiliza as suas estruturas também no modelo SKOS, e nessa implementação utilizou-se o conteúdo de categorias da Wikipédia que a DBPedia torna disponível.

A implementação do DbpediaAnnotator contou, ainda, com o recurso de download automático do conteúdo da DBPedia a partir de páginas da Web (e.g., <http://downloads.dbpedia.org/2015-10/core-118n/pt/skos_categories_pt.ttl.bz2>). Isso ocorre após a inicialização do

¹⁸ Apache Lucene é um software de busca e uma API de indexação de documentos de alta performance. Disponível em: <<https://lucene.apache.org/core>>.

¹⁹ DBPedia é um projeto cujo objetivo é extrair conteúdo estruturado das informações da Wikipédia. Disponível em: <<http://wiki.dbpedia.org>>.

²⁰ SKOS – Simple Knowledge Organization System – é uma área de trabalho de desenvolvimento de especificações e normas para apoiar o uso de sistemas de organização do conhecimento, como enciclopédia, esquemas de classificação, listas de cabeçalhos de assuntos e taxonomias dentro da estrutura da Web Semântica.

anotador, quando há uma verificação se o conteúdo da DBPedia está presente no diretório da aplicação. Caso não esteja, o conteúdo é transferido automaticamente, conforme as URLs configuradas no anotador, e um índice Lucene é criado com a estrutura apresentada na Figura 24. A título de exemplo, as categorias da DBPedia nas línguas portuguesa e inglesa possuem juntas 1,5 milhão de conceitos. Com essa quantidade de termos, ficaria inviável utilizar o componente do DKPro DictionaryAnnotator.

Figura 24 – Estrutura do índice Lucene gerado a partir da SKOS da DBPedia

The screenshot shows the Luke - Lucene Index Toolbox interface. The index name is `C:\Users\Rudger\AppData\Local\Temp\expert-profiling\dbpedia_index_971b2fd41faeda1f1e721ae39b08c436`. Key statistics include: Number of fields: 5, Number of documents: 1542934, Number of terms: 5222464, and Has deletions?: No / Yes. The index format is Lucene ver.: 5.1.0. The interface also shows a table of available fields and term counts per field, and a table of top ranking terms.

Name	Term count	%	Decoder
broader	585.766	11,22 %	string utf8
pref	1.531.503	29,33 %	string utf8
pref_norm	1.528.206	29,28 %	string utf8
related	33.975	0,65 %	string utf8
uri	1.542.934	29,54 %	string utf8

Rank	Freq	Field	Text
1	329	related	http://dbpedia.org/resource/Category/Wikipedians
2	162	related	http://dbpedia.org/resource/Category/People
3	176	related	http://dbpedia.org/resource/Category/Wikipedians_by_locali
4	151	related	http://dbpedia.org/resource/Category/Buildings_and_structu
5	143	related	http://dbpedia.org/resource/Category/People_by_nationality
6	120	related	http://dbpedia.org/resource/Category/American_architects
7	101	related	http://dbpedia.org/resource/Category/Wikipedians_by_inter
8	97	related	http://dbpedia.org/resource/Category/Actors_by_medium
9	85	related	http://dbpedia.org/resource/Category/English_words
10	84	related	http://dbpedia.org/resource/Category/Filipino_people
11	71	related	http://dbpedia.org/resource/Category/Provinces_of_the_Phil
12	66	related	http://dbpedia.org/resource/Category/Wikipedians_in_the_L
13	63	related	http://dbpedia.org/resource/Category/Wikipedians_by_inter
14	60	related	http://dbpedia.org/resource/Category/Wikipedians_intereste
15	58	related	http://dbpedia.org/resource/Category/Writers_by_genre
16	57	related	http://dbpedia.org/resource/Category/Wikipedians_by_alma

Fonte: elaborado pelo autor com o uso da ferramenta Lucene Luke

No que tange à implementação do componente DbpediaAnnotator, foi utilizado o algoritmo de n-grama observado no componente DictionaryAnnotator do DKPro. A diferença na implementação é que o tamanho do n-grama do DictionaryAnnotator era definido conforme o tamanho de *tokens* de uma sentença que estava sendo processada pelo anotador. No caso da implementação do DbpediaAnnotator, esse tamanho do n-grama é configurado com o tamanho máximo de *tokens* existentes em todos os conceitos contidos na DBPedia. Por exemplo, no caso da combinação das taxonomias da DBPedia em inglês e em português, o número máximo de *tokens* é de 24 (i.e., existe um conceito com 24 tokens na DBPedia). Isso significa que o número máximo de combinações por sentença de *tokens* para a

verificação de conceitos no índice Lucene da DBPedia é de 24 possibilidades no máximo.

Para exemplificar como são gerados os n-grams, o Quadro 15 a seguir ilustra a criação de n-gramas a partir do texto “*Exemplo de n-grama de por sentença*”. Ademais, apresenta-se a forma de consulta que é realizada para cada um dos n-gramas gerados que são decompostos até que seja encontrado um conceito válido no índice Lucene da DBPedia.

Quadro 15 – N-gramas gerados a partir de uma sentença e as respectivas consultas realizadas

N-grama gerado	Consulta realizada no índice
1. exemplo de n-grama de por sentença	1. exemplo de n-grama de por sentença 2. exemplo de n-grama de por 3. exemplo de n-grama de 4. exemplo de n-gram 5. exemplo de 6. exemplo
2. de n-grama de por sentença	1. de n-grama de por sentença 2. de n-grama de por 3. de n-grama de 4. de n-gram 5. de
3. n-grama de por sentença	1. n-grama de por sentença 2. n-grama de por 3. n-grama de 4. n-gram
4. de por sentença	1. de por sentença 2. de por 3. de

N-grama gerado	Consulta realizada no índice
5. por sentenca	1. por sentenca 2. por
6. sentenca	sentenca

Fonte: elaborado pelo autor

Quanto às consultas realizadas no índice Lucene, é importante destacar que, em função das anotações de *stopwords* geradas na etapa anterior do protótipo denominado extração do conteúdo, os conceitos que são encontrados no índice Lucene da base da DBpedia e que foram anotados como *stopwords* não são considerados válidos como resultado do processo.

Outra informação relevante referente à implementação do DbpediaAnnotator é que a busca dos termos no índice Lucene é realizada no campo normalizado. Isto é, no momento da indexação, cria-se um campo de busca com os textos dos conceitos normalizados. Essa normalização inclui apenas o tratamento de texto em caixa-baixa e a remoção de acentos. Não foram consideradas questões de *stemming* devido à utilização de uma base muito grande de conceitos, o que poderia resultar em anotações demasiadas no documento. Contudo, recomenda-se realizar esse experimento em trabalhos futuros.

Em resumo, foram criados os componentes DbpediaAnnotator e HeidelTimeWrapper para a implementação da etapa de reconhecimento de entidades do protótipo. A Figura 25 apresenta um exemplo das anotações de conceitos, enquanto a Figura 26 apresenta um exemplo das anotações de datas. Essas figuras foram capturadas com a utilização da ferramenta Annotation Viewer do pacote Apache UIMA.

Figura 25 – Resultados da etapa de reconhecimento de entidades – anotações de conceitos

Annotation Results for Example3.txt.xmi in C:\Users\Rudger\AppData\Local\Temp\expert-profiling\analysis\result

possui bacharelado em **ciencia da computacao** pela **universidade do vale do itajai (1997)**, **mestrado (2001)** e **doutorado (2006)** em **engenharia de producao** pela **universidade federal de santa catarina**, com estagio de doutoramento na the open university. atualmente e professor no departamento de administracao publica da **universidade do estado de santa catarina** e no programa de pos-graduacao em **engenharia e gestao do conhecimento** na **universidade federal de santa catarina**. atua tambem como diretor e pesquisador no instituto stela. atuou como pesquisador e coordenou diversos **projetos de p&d** com **organizacoes publicas** (como a **plataforma lattes** e a **plataforma aquarius** com o mcti, portal **sinaes** com o mec, **dcvisa** com a **anvisa** e **sibea** com o **mma**), **organizacoes privadas** (como **busca semantica** com **embraer** e **plataforma de gestao da etica** e da **integridade** com o **itau/unibanco**) e terceiro setor (como a **plataforma para gestao do absenteismo** com o **sesi/ba** e **plataforma da gestao do conhecimento do senai/ce-fiec**). tem experiencia na **area de ciencia da computacao**, atuando principalmente nos seguintes temas: **engenharia do conhecimento**, **gestao do conhecimento**, **data warehouse**, **business intelligence**, **ontologias**, **semantic web**, **planejamento** de sistemas de informacoes e **governo eletronico**.

Annotation Types

DbpediaCategory

Click In Text to See Annotation Detail

- Annotations
 - DbpediaCategory
 - DbpediaCategory ("ciencia da computacao")
 - begin = 22
 - end = 43
 - uri = <http://pt.dbpedia.org/resource/Cat>

Fonte: elaborado pelo autor com o uso da ferramenta Apache UIMA Annotation Viewer

Figura 26 – Resultados da etapa de reconhecimento de entidades – anotações de datas

Annotation Results for Example3.txt.xmi in C:\Users\Rudger\AppData\Local\Temp\expert-profiling\analysis\result

possui bacharelado em ciencia da computacao pela universidade do vale do itajai (1997), mestrado (2001) e doutorado (2006) em engenharia de producao pela universidade federal de santa catarina, com estagio de doutoramento na the open university. atualmente e professor no departamento de administracao publica da universidade do estado de santa catarina e no programa de pos-graduacao em engenharia e gestao do conhecimento na universidade federal de santa catarina. atua tambem como diretor e pesquisador no instituto stela. atuou como pesquisador e coordenou diversos projetos de p&d com organizacoes publicas (como a plataforma lattes e a plataforma aquarius com o mcti, portal sinaes com o mec, dcvisa com a anvisa e sibeia com o mma), organizacoes privadas (como busca semantica com embraer e plataforma de gestao da etica e da integridade com o itau/unibanco) e terceiro setor (como a plataforma para gestao do absenteismo com o sesi/ba e plataforma da gestao do conhecimento do senai/ce-fiec). tem experiencia na area de ciencia da computacao, atuando principalmente nos seguintes temas: engenharia do conhecimento, gestao do conhecimento, data warehouse, business intelligence, ontologias, semantic web, planejamento de sistemas de informacoes e governo eletronico.

Annotation Types

Timex3

Click In Text to See Annotation Detail

- Annotations
 - Timex3
 - Timex3 ("1997")
 - begin = 81
 - end = 85
 - filename = null
 - sentId = 0
 - firstTokId = 0
 - allTokIds = BEGIN<-->0
 - timexId = t1
 - timexInstance = 0
 - timexType = DATE
 - timexValue = 1997
 - foundByRule = date_r11a-explicit
 - timexQuant =
 - timexFreq =

Fonte: elaborado pelo autor com o uso da ferramenta Apache UIMA Annotation Viewer

4.2.4 Implementação da etapa de extração e armazenamento de entidade

A implementação da etapa de extração e armazenamento de entidade do protótipo corresponde à implementação do elemento “Gravação de conceitos e datas no Lucene”, apresentado na Figura 18 da arquitetura do protótipo. O resultado esperado para essa etapa é, basicamente, a persistência das anotações realizadas nos documentos dos especialistas. Essa persistência deve possibilitar as consultas dessas informações posteriormente.

Para essa implementação, inicialmente, pensou-se em utilizar um banco de dados relacional portátil, como, por exemplo, o H2²¹. Porém, ao se analisarem os requisitos de persistência dessas anotações, observou-se que não havia a necessidade de garantia de consistência entre os dados, procedimento que diminui o desempenho no momento da gravação, e também não havia a necessidade de utilização de forma compartilhada da base de dados.

Partindo-se dessas observações, optou-se, então, por utilizar a persistência das anotações em um índice Lucene, o que torna a leitura das informações mais veloz pelo fato de o Lucene ser otimizado para a realização de consultas. Além disso, verificou-se que o Lucene também permite a realização de consulta aos seus dados de modo semelhante aos bancos relacionais (MCCANDLESS, 2012). Essa funcionalidade permite, por exemplo, efetuar consultas de documentos-pais que possuam itens-filhos aninhados. Esse é o caso das anotações nesta etapa, pois é necessário gravar as informações de metadados dos documentos e também todo o conjunto de anotações que foram realizadas para esses documentos.

Com essas premissas, implementou-se o componente chamado Annotations2Lucene. Esse componente captura as anotações realizadas nas etapas anteriores (i.e., entidades de conceitos e entidades de datas) e as persiste em um índice que possui os campos listados no Quadro 16 a seguir.

²¹ H2 é um sistema de gerenciamento de banco de dados relacional escrito em Java e pode ser integrado em aplicativos Java ou executar no modo cliente-servidor.

Quadro 16 – Campos do índice Lucene gerados pela etapa de extração e armazenamento de entidade

Campo do índice	Descrição
Type	Tipo do documento no índice Lucene (documento ou anotação). Dependendo do tipo, o resultado da consulta retorna as informações do documento do especialista ou as informações de uma anotação do documento.
doc_id	Identificador do documento do especialista. É gerado de forma sequencial e controlado pelo protótipo.
doc_title	Título do documento do especialista
doc_uri	Localização do documento do especialista no momento da execução da aplicação do protótipo.
doc_language	Linguagem do documento. A implementação atual considera apenas as línguas portuguesa e inglesa.
doc_creation_time	Data de criação do documento do especialista. Essa data pode sofrer alterações quando o documento é baixado pela internet.
doc_last_modification_time	Data da última modificação do documento do especialista.
doc_last_access_time	Data do último acesso ao documento do especialista.

Campo do índice	Descrição
doc_metadata_json	Conteúdo das metainformações do documento do especialista no formato Json para consulta.
doc_md5	Código de criptografia do conteúdo do documento do especialista. Esse código não pode se repetir para o mesmo especialista duas vezes. Garante que o especialista insira apenas uma vez o mesmo documento no processo anotação.
doc_content	Conteúdo binário do documento do especialista para futuras implementações.
doc_owner_name	Nome do especialista dono do documento processado.
doc_owner_code	Código de referência do especialista para ser utilizado nas consultas para extração do perfil.
doc_annotation_type	Tipo da anotação do documento quando o DOC do Lucene for específico para o tipo anotação.
doc_annotation_id	Identificador da anotação. É gerado de forma sequencial e controlado pelo protótipo.
doc_annotation_document_id	Código de referência do documento que serviu de base para a anotação gerada.
doc_annotation_value	Valor do texto correspondente ao alcance da anotação.

Campo do índice	Descrição
doc_annotation_sentence_seq	Identificador sequencial da sentença do documento onde a anotação ocorreu.
doc_annotation_sentence_begin	Posição de início da sentença do documento.
doc_annotation_sentence_end	Posição de fim da sentença do documento.
doc_annotation_timex3_value	Valor específico para a anotação do tipo data.
doc_annotation_timex3_type	Código do tipo da anotação gerada pelo HeidelTime. Pode ser data ou timestamp.
doc_annotation_dbpedia_category_uri	Valor da URI da categoria da DBPedia.

Fonte: elaborado pelo autor

4.2.5 Implementação da etapa de associação de entidades

A penúltima etapa de implementação do modelo proposto tem por objetivo a leitura e a associação das entidades identificadas no conteúdo dos documentos do especialista, conforme é apresentado no modelo conceitual. Esta etapa corresponde ao elemento “Gerador de associações de conceitos e gerador de página do perfil (*tag cloud* e *timeline*)” apresentado na Figura 18 da arquitetura do protótipo. Entretanto, a etapa de associação de entidades corresponde apenas à parte do “Gerador de associações de conceitos”. Já a parte “gerador de página do perfil” é apresentada na seção 4.2.6. Isso se deve ao fato de o componente criado ser responsável pelas duas etapas.

Esse componente de associação de entidades foi denominado Lucene2ProfilePage. Sua função é ler as informações das anotações extraídas e armazenadas no índice Lucene e criar as relações entre as entidades para a extração e a geração do perfil de especialista. Contudo, é importante mencionar que o Lucene2ProfilePage é executado apenas uma vez ao final do processamento de todos os documentos de um especialista. Esse procedimento acontece de forma diferente nos demais

componentes implementados nas etapas anteriores. Nos componentes anteriores, por exemplo, o acionamento do processamento de cada etapa é executado para cada documento. Por isso, foi necessário implementar a persistência de todo o conteúdo processado referente às anotações dos documentos do especialista no componente Annotations2Lucene (seção 4.2.4).

Dois algoritmos são esperados nesta etapa. O primeiro algoritmo deve combinar as anotações de conceitos com as datas dos documentos para fomentar as informações da geração do perfil de especialista. O segundo algoritmo visa relacionar as anotações de datas e de conceitos presentes em uma mesma sentença por proximidade. Para o protótipo, foi considerada a implementação do primeiro algoritmo, que combina as informações das anotações com as datas da última modificação dos documentos. Essa data da última modificação foi escolhida por representar a última edição feita pelo especialista em seu documento. O segundo algoritmo deverá ser implementado em trabalhos futuros.

Desse modo, quanto à associação das entidades de conceitos e às datas de modificação dos documentos, a implementação do componente Lucene2ProfilePage gera dois conjuntos de dados: (1) o conjunto de dados para o componente de *tag cloud* e (2) o conjunto de dados para o componente de *timeline*. O procedimento para a geração de cada um desses conjuntos de dados é explicado a seguir.

- **Conjunto de dados para *tag cloud*.** O conjunto de dados para o componente de *tag cloud* é obtido por meio da leitura de todas as anotações dos documentos de um especialista. Após a leitura dessas anotações, os conceitos são agrupados pela URI, informação proveniente da DBPedia, e uma contabilização é gerada para cada um desses itens agrupados. Nesse ponto, obtêm-se todos os conceitos representados pela URI e a sua frequência total observada nos documentos do especialista. Por fim, o último procedimento executado é a limitação em até 100 conceitos ordenados pela quantidade de forma decrescente, os quais serão repassados para o componente visual de *tag cloud*.
- **Conjunto de dados para *timeline*.** O conjunto de dados para o componente de *timeline* é obtido de forma semelhante à do conjunto de dados para o componente de *tag cloud*. A diferença observada é que há dois agrupamentos para a geração dos dados. O primeiro agrupamento é gerado pelos anos registrados

nas dadas de modificação dos documentos. Assim, obtém-se uma lista de todas as anotações dos documentos dos especialistas e realiza-se o agrupamento por ano. A partir desse resultado, é realizado o segundo agrupamento, agora pelos conceitos que estão agrupados pelos anos. Isto é, para cada lista de conceitos agrupados por anos é realizado um novo agrupamento pela URI, e uma contabilização é gerada para cada um desses itens agrupados. Com esse resultado, apenas se ordena o agrupamento de itens gerados pela URI e pela quantidade de ocorrências de forma decrescente e limita-se em até 100 conceitos por ano que serão repassados no componente visual de *timeline*.

Após essa associação de entidades com as informações de datas dos documentos e os respectivos agrupamentos por totais e anos, o componente *Lucene2ProfilePage* envia esses dados a um *template* do Apache Velocity²², responsável pela etapa de extração do perfil de especialista, de modo a apresentá-lo visualmente. Esse processo é descrito na etapa a seguir.

4.2.6 Implementação da etapa de extração do perfil de especialista

A última etapa da implementação do modelo proposto visa à extração e à apresentação do perfil de especialista em instrumentos visuais de *tag cloud* e de *timeline*, conforme orientações do modelo conceitual proposto. Assim como a implementação da etapa de associação de entidades, esta etapa corresponde ao elemento “Gerador de associações de conceitos e gerador de página do perfil (*tag cloud* e *timeline*)” apresentado na Figura 18 da arquitetura do protótipo. Porém, a sua responsabilidade corresponde à parte de “gerador de página do perfil”.

Na implementação da etapa de associação de entidades, foram gerados mapas com as informações das entidades de conceitos encontradas nos documentos de um especialista. Esses mapas representam dois conjuntos de dados que já se encontram montados para o consumo nesta etapa de extração do perfil de especialista. Esses conjuntos representam os dados para a *tag cloud* e para a *timeline*.

²² Apache Velocity é uma biblioteca para o processamento de *templates* e que permite a geração de relatórios nos formatos de HTML, XML, RTF e outros. Disponível em: <<http://velocity.apache.org>>.

Nesse contexto, o componente responsável pela implementação desta etapa é o Lucene2ProfilePage, o mesmo utilizado na etapa de associação de entidades. Porém, a parte específica desta etapa se dá pela apresentação do conteúdo em componentes visuais. Portanto, conforme mencionado anteriormente, um *template* do Apache Velocity é utilizado para a criação de uma página em HTML com as entidades extraídas dos documentos para, enfim, a criação do perfil do especialista.

O *template* responsável pela montagem dos elementos visuais em HTML, que utiliza os dados resultantes das associações de entidades, foi construído com base em dois componentes: (1) jQCloud²³, para a criação de *tag cloud* e (2) Timeline Chap Link²⁴, para a criação de *timeline*. Além disso, foi criada uma visualização adicional chamada de “relatório”, que apresenta as mesmas informações do conteúdo de *timeline*, tendo como diferença o formato vertical de apresentação do conteúdo. Os exemplos visuais da utilização desses componentes e os respectivos resultados da etapa de extração do perfil podem ser observados na seção seguinte, que apresenta os resultados finais da construção do protótipo.

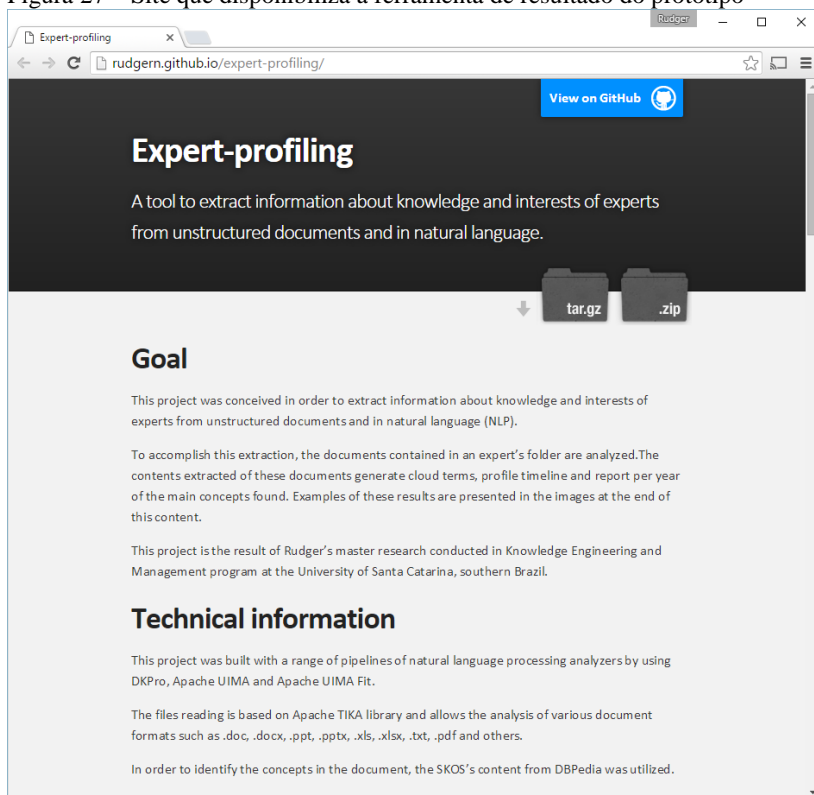
4.2.7 Ferramenta e resultados gerados a partir do protótipo

A implementação do protótipo que contempla as etapas do modelo conceitual proposto nesta pesquisa com base nas abordagens observadas na literatura resultou na criação de uma ferramenta de código aberto disponível no site <http://rudgertaxweiler.github.io/expert-profiling>. Esse site, além de explicar o objetivo, as informações técnicas e os resultados desta pesquisa, disponibiliza também todo o código-fonte gerado. Dessa forma, possibilita a divulgação da ferramenta gerada e também permite a contribuição de outras pessoas interessadas no tema desta pesquisa. A Figura 27 a seguir apresenta uma imagem do site mencionado.

²³ jQCloud é um componente baseado em JQuery para a criação de word cloud. Disponível em: <<https://github.com/lucaong/jQCloud>>.

²⁴ Timeline Chap Link é um componente visual para representação interativa de eventos no tempo. Disponível em: <<https://almende.github.io/chap-links-library/timeline.html>>.

Figura 27 – Site que disponibiliza a ferramenta de resultado do protótipo

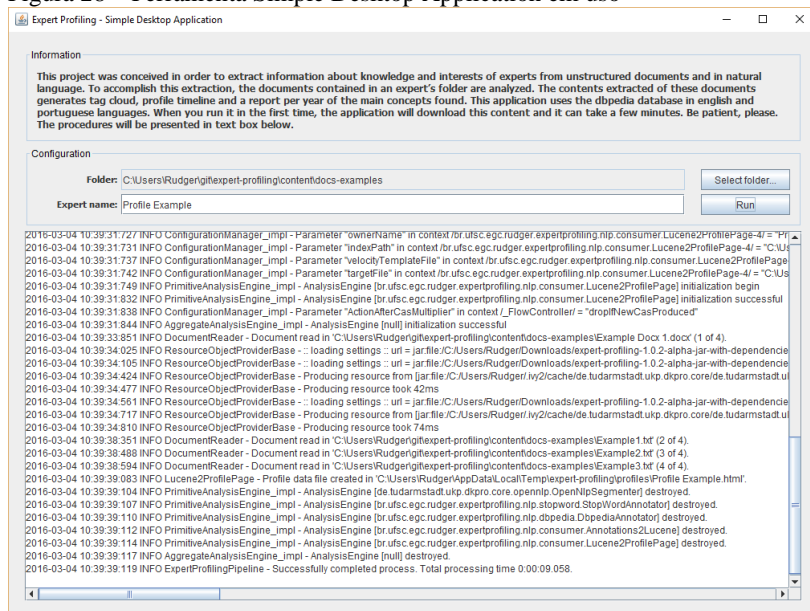


Fonte: elaborado pelo autor

Essa implementação possibilitou, ainda, a criação de uma ferramenta visual simples para a utilização de qualquer especialista. Essa ferramenta, denominada “Simple Desktop Application”, permite a configuração e a execução das etapas mencionadas no modelo proposto para gerar páginas que representem possíveis conhecimentos e interesses de um especialista a partir de seus documentos. Basicamente, é necessário configurar apenas o diretório no qual os documentos serão obtidos e o nome do especialista que terá a sua página gerada. A Figura 28 mostra um exemplo de utilização da ferramenta “Simple Desktop Application” com os seus respectivos resultados apresentados na

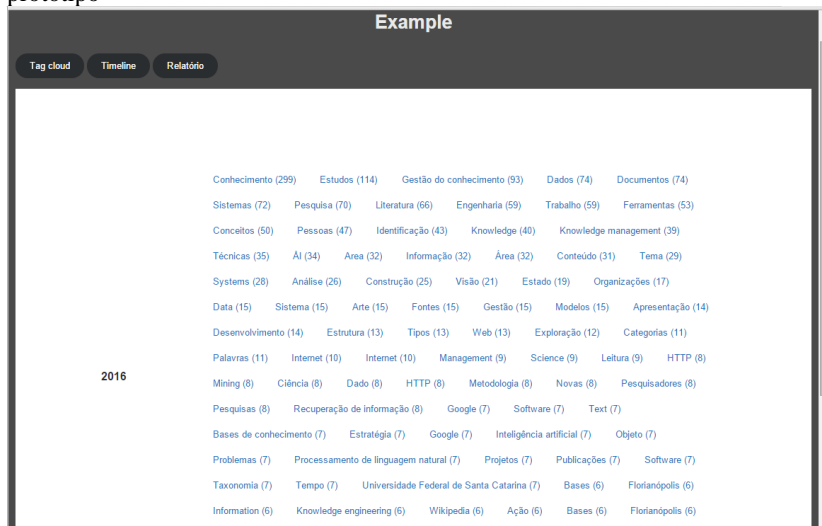
Figura 29, com a *tag cloud*, Figura 30, com a *timeline*, e Figura 31, com o *relatório*. O conteúdo deste documento fomentou a geração do perfil de exemplo ilustrado nessas imagens.

Figura 28 – Ferramenta Simple Desktop Application em uso



Fonte: elaborado pelo autor

Figura 31 – Exemplo de perfil com representação em relatório gerado pelo protótipo



Fonte: elaborado pelo autor

4.3 DEMONSTRAÇÃO DE VIABILIDADE

Com base no objetivo desta pesquisa, que é propor um modelo para a extração de perfis de especialistas a partir de documentos não estruturados, foi desenvolvido um protótipo funcional que implementa as etapas do modelo conceitual proposto. Ao final da implementação desse protótipo, chegou-se, então, a uma ferramenta que pode ser utilizada por qualquer especialista que deseja analisar seus documentos para extrair insumos de seus conhecimentos e interesses ao longo dos anos. Além disso, o protótipo da implementação também pode ser utilizado como um componente de negócio para integração em sistemas de *Expertise Location*.

Essa implementação corresponde ao último objetivo específico deste trabalho, que é desenvolver um protótipo funcional para verificar a viabilidade do modelo proposto utilizando documentos para a extração do perfil de especialista. Assim, foi realizada a geração de três perfis de especialistas, que se colocaram à disposição para executar a ferramenta a partir de seus documentos. Os perfis extraídos desses especialistas permitem observar a viabilidade do modelo proposto. Para fins de demonstração, foram selecionados os conteúdos de *tag cloud* dos resultados da extração do perfil apresentados na Figura 32 (especialista #1), Figura 33 (especialista #2) e Figura 34 (especialista #3). Os

4.3.1 Comparação dos perfis extraídos com os perfis publicados no LinkedIn

Com base nos resultados dos perfis extraídos dos três especialistas que participaram do teste com a ferramenta de protótipo, realizou-se uma comparação desses perfis no que tange ao conteúdo de “Competências e recomendações” e “conhecimentos” da ferramenta LinkedIn. Para a comparação, utilizou-se a ferramenta Document Similarity²⁵, que realiza a comparação entre dois documentos via algoritmo de Cosseno, o qual pode ser observado nos trabalhos de Nascimento et al. (2014), Bovo (2010), Ceci (2010) e Gonçalves (2006). O conteúdo utilizado para comparação entre os perfis com a ferramenta Document Similarity é apresentado no Quadro 17.

Antes de se compararem os resultados dos perfis extraídos com a ferramenta em relação aos perfis do LinkedIn, foram removidas as quantidades de vezes em que cada conceito ocorre nos documentos do especialista para o perfil extraído bem como as quantidades de vezes em que o conhecimento foi recomendado no LinkedIn. Os textos comparados e os resultados de cada comparação de perfil de especialista estão disponíveis em forma completa nos apêndices D, E e F deste trabalho.

Como resultado, obtiveram-se os índices 0,6569 para o perfil do especialista #1, 0,7796 para o perfil do especialista #2 e, por fim, 0,6858 para o perfil do especialista #3. Esses resultados geraram uma média de **0,7074** de similaridade entre os perfis originados com os perfis mantidos no LinkedIn.

Esse índice demonstra que a ferramenta tem potencial para extrair de forma automática os perfis de especialistas a partir de seus documentos. As implicações desse resultado são discutidas mais amplamente no capítulo de conclusões.

²⁵ Document Similarity é uma ferramenta baseada em PLN e Machine Learning que compara a semelhança semântica de dois documentos de texto com o algoritmo de similaridade de Cosseno.

Quadro 17 – Perfil do LinkedIn e Perfil extraído pela ferramenta de protótipo dos especialistas

#	LinkedIn	Perfil extraído (<i>tag cloud</i>)
1	<p>Business Intelligence (57), Project Management (37), Knowledge Management (35), Management (21), Strategic Planning (18), Business Development (16), Product Development (15), Strategy (14), Team Leadership (13), Business Analysis (13), Innovation Management(8), Entrepreneurship (8), Business Strategy (7), E-government (7), Scrum (7), Software Development (6), SQL (5), Software Project (5), Product Management (5), Project Planning (4), Business Process (4), MS Project (4), Java (4), ITIL (4), Requirements Analysis (3), Semantic Technologies (3), Team Management (3), UML (3), Databases (3), Business Planning (2), Data Analysis (2), Governance (2), Systems Analysis (2), Software Engineering (2), Portuguese (2), ERP (2), Management Consulting (2), Agile Project Management (1), PMBOK (1), PMP (1), Agile Methodologies (1), Knowledge Engineering (1), Semantics (1), Semantic Web, Artificial Intelligence, PMI, Data Mining</p>	<p>Conhecimento (57130), Dados (48939), Plataforma (43924), Sistema (39979), Desenvolvimento (35756), Gestão (30816), Pesquisa (30757), Serviços (24976), Informação (23442), Sistemas (23119), Inovação (22761), Trabalho (22443), Projetos (21640), Análise (21122), Data (21002), Área (19370), Valor (16531), Empresas (15160), Tecnologia (15063), Knowledge (14041), Nacional (13938), Fontes (13680), Avaliação (13510), Plano (13506), Ambiente (13130), Produtos (12652), SENAI (12383), Nome (12302), HTTP (12243), Qualidade (11986), Arquitetura (11971), Componentes (11290), Educação (11126), Capital (11033), Software (10818), Software (10818), Engenharia (10809), Gestão do conhecimento (10670), Governo (10602), Apresentação (10384), Técnicas (10313), Information (10054), Planejamento (9975), Ações (9905), Identificação (9785), Documentos (9573), Rede (9470), Pessoas (9314), Brasil (9228), Value (9050), Instituição (8969), Tempo (8900), Estrutura (8691), Ontologia (8479), Visão (8231), Av (8116), Estratégia (8111), Registro (7879), Ação (7647), Comunicação (7268), Segurança (7182), Dias (7119), Ferramentas (7106), Fortaleza (6986), Gerenciamento (6943), Florianópolis (6910), Organizações (6858), Web (6854), Equipamentos (6794), Conteúdo (6695), Ál (6679), Instituições (6641), Metodologia (6595), Anos (6536), Construção (6472), Objeto (6446), Estado (6441), Ciência (6290), Mercado (6196), Atores (5960), Art (5845), Oracle (5845), Example (5842), Evolução (5774), Redes (5660), Saúde (5550), Conceitos (5487), Clientes (5480), Santos (5466), Management (5463), Sistemas de informação (5425), Novas (5397), Referência (5216)</p>

#	LinkedIn	Perfil extraído (<i>tag cloud</i>)
2	<p>Business Intelligence (31), Java (24), Scrum (20), Agile Methodologies (16), Knowledge Engineering (14), OLAP (11), Big Data (10), Software Engineering (9), Solr (1), PostgreSQL (13), Lucene (2), ElasticSearch (1), Linked Data (5), Eclipse (6), JUnit (5), ERP (4), Databases (5), JBoss Application Server (4), SQL(17), Tomcat (3), JPA (3), RUP (2), Hibernate (8), Universal Information (1), Certified Scrum Master(5), Spring (1), Software Development (2)</p>	<p>Data (10971), Information (4671), Dados (4222), Ontology (3098), XML (2901), Example (2769), Business (2090), Knowledge (2052), Dimension (1979), Systems (1833), Database (1786), Text (1692), Time (1679), Documents (1577), Analysis (1551), OLAP (1320), Análise (1300), Sources (1280), Framework (1273), Web (1228), HTTP (1214), SQL (1178), Architecture (1120), Value (1097), Tables (1050), Language (1027), Research (1024), Concepts (993), Design (955), Tools (923), Property (909), Structure (909), Management (882), Data warehouse (882), Single (866), Java (847), Tubarão (830), Hierarchy (813), Conteúdo (809), Business intelligence (808), Business intelligence (808), Environment (806), Models (767), Definition (760), Informação (748), Sales (747), Technology (737), Fontes (737), Modeling (724), Sistema (711), Rules (698), Conhecimento (677), Development (672), Databases (671), Processes (666), Software (663), Software (663), API (658), Area (640), Tempo (636), Tool (610), Big data (610), Tool (610), Ontologia (599), Technologies (577), Ál (570), Arquitetura (568), Metadata (564), Filter (557), Paper (557), Semantics (548), Driver (535), Objects (531), Main (525), Pesquisa (505), Knowledge management (504), Components (504), Conceitos (486), Topics (486), Ferramentas (478), Products (477), Documentos (475), Standards (472), Means (469), Size (459), Techniques (457), Propriedade (457), Web services (453), Valor (447), Semantic Web (447), Brasil (443), Form (441)</p>

#	LinkedIn	Perfil extraído (<i>tag cloud</i>)
3	<p>Scrum (28), Java (20), UML (17), PostgreSQL (12), JavaScript (8), Java Enterprise Edition (8), Agile Project Management (7), XML (6), Spring (6), RUP (4), HTML (4), JSP (4), CMMI (3), JPA (3), Gestão ágil de projetos (3), Gerência de projeto (2), MS Project (2), Arquitetura de software (2), Spring Framework (2), Engenharia do conhecimento (1), e-Governance (1), MPS.Br (1), AspectJ (1), Oracle SQL (1), Microsoft Project (1), Análise de requisitos, iReport, Jasper Reports, Enterprise Architecture, BPMN, Análise de processos, XSLT</p>	<p>Knowledge (27135), Conhecimento (24080), Tipos (15246), B5 (14009), Ciências (12975), Trabalho (11534), Plano (10928), Research (10776), Sistema (10490), Dados (10313), Pesquisa (10282), Medicina (9628), Area (9459), Área (9459), Information (9440), Data (8288), Ensino (8047), Sistemas (7420), Avaliação (7359), Desenvolvimento (7220), AI (7202), Análise (7125), Segurança (6420), Gestão (5956), Analysis (5801), Saúde (5434), Informação (5425), Conceitos (5359), Knowledge management (5278), Técnicas (5233), Management (5198), Educação (5126), Gestão do conhecimento (5042), Example (4946), Ambiente (4840), Inovação (4808), Design (4785), Design (4785), Qualidade (4641), Tecnologia (4625), Ferramentas (4515), Software (4430), Software (4430), Systems (4400), Equipamentos (4389), Java (4353), HTTP (4214), HTTP (4214), Ontologia (4182), Máquinas (4098), Development (4084), Engenharia (4078), Materiais (3997), Ciência (3907), Time (3887), Documents (3867), People (3856), March (3833), Comunicação (3786), Estudos (3785), Serviços (3764), Tools (3712), Science (3691), Web (3690), Learning (3670), Technology (3609), Wednesday (3573), Projetos (3565), Rede (3536), Documentos (3491), Organizations (3400), Value (3329), Text (3302), Texto (3263), Protected (3231), Pessoas (3181), Framework (3147), Psicologia (3141), Estrutura (3087), Administração (3078), Organizações (3077), Tempo (3075), Planejamento (3074), Empresas (3025), História (3015), Produtos (3007), Construção (3000), Identificação (2802), Direito (2801), Models (2795), Valor (2748), Biotecnologia (2737), Business (2705), Redes (2659), Linguística (2654), Linguística (2654), Química (2648), Rio de Janeiro (2607), Rio de Janeiro (2607), Structure (2598)</p>

Fonte: elaborado pelo autor

5 CONCLUSÕES, CONTRIBUIÇÕES E TRABALHOS FUTUROS

Conforme exposto nas primeiras seções deste trabalho, o conhecimento é um importante recurso para o ganho de competitividade nas organizações e, por esse motivo, tem se observado uma preocupação cada dia maior para torná-lo fluido nos ambientes organizacionais. É nesse contexto que entra a Gestão do Conhecimento, que possibilita o gerenciamento desse valioso recurso cada vez mais necessário para as organizações. Diante desse cenário, uma importante ferramenta para a Gestão do Conhecimento, que apoia o gerenciamento dos ativos de conhecimento, são os sistemas de *Expertise Location*. Tal ferramenta auxilia a identificação e o compartilhamento de conhecimento, seja ele tácito ou explícito, por meio da integração de pessoas que precisam de determinados conhecimentos com aquelas que os possuem.

Evidentemente, a construção das ferramentas de *Expertise Location* é observada há muitas décadas na literatura, conforme apresentado nas seções 1.1, 2.1.4 e 2.2 deste trabalho, antes mesmo das iniciativas da Gestão do Conhecimento. Contudo, tais ferramentas estão ganhando cada vez mais recursos para serem melhoradas. Um desses recursos, que tem se destacado nesta última década, é a utilização de tecnologias das áreas de Recuperação de Informação e de Inteligência Artificial para a extração e criação automática de perfil de especialistas a partir dos documentos elaborados por esses especialistas (seções 1.1, 2.1.4 e 2.2).

Nesse contexto, este trabalho teve por objetivo propor um modelo para a extração do perfil de especialista de forma automática, a partir de documentos para apoiar a Gestão do Conhecimento na aplicação de ferramentas de *Expertise Location* em organizações. Um dos objetivos específicos deste trabalho foi justamente identificar as atuais abordagens que possibilitam a extração de perfis de especialistas a partir de documentos não estruturados no contexto de *Expertise Location*. Tal objetivo foi alcançado por meio da realização de uma revisão sistemática da literatura (seções 3.1, 3.2 e 3.3), que resultou em 70 trabalhos para análise. Desses 70 trabalhos, foram selecionados 11, que contribuíram para a identificação dessas abordagens atuais, conforme foi apresentado na seção 2.2.

De posse da visão das abordagens atuais que possibilitaram atingir o objetivo desta pesquisa, foi sugerido um modelo conceitual que combinou as várias propostas dos trabalhos observados na literatura. Inclusive, combinaram-se abordagens realizadas no âmbito do Programa

de Engenharia e Gestão do Conhecimento da UFSC, desenvolvidas pelos autores Bovo (2011) e Ceci (2010). Dessa maneira, foi atingido também o segundo objetivo específico deste trabalho, que era definir um modelo conceitual, de acordo com as abordagens existentes, que suportasse a extração automática de perfis de especialistas e identificasse os elementos necessários para que o modelo pudesse ser implementado. Esse modelo foi apresentado na seção 4.1 e teve como proposta a realização de cinco etapas para o processamento de documentos de um especialista até que fossem extraídas as referências aos possíveis conhecimentos (i.e., tácito ou explícito) e os interesses que esse especialista possui.

Finalmente, com a proposição do modelo conceitual para a extração de perfil de especialista a partir de documentos, implementou-se um protótipo funcional que materializava o modelo proposto neste trabalho. Tal protótipo gerou ainda uma ferramenta denominada Expert Profiling, que foi disponibilizada no site <http://rudgertaxweiler.github.io/expert-profiling> para download de qualquer especialista que tenha interesse em processar os seus documentos para extrair insumos de modo a gerar um perfil de especialista automaticamente. Essa ferramenta foi liberada juntamente com o seu código-fonte, que é aberto à comunidade. Ademais, a sua disponibilização abre a possibilidade, por exemplo, de ganho de novas contribuições provenientes de pesquisadores que estão trabalhando com o mesmo tema de pesquisa abordado neste trabalho.

Com essa implementação do protótipo, atingiu-se, então, o último objetivo específico desta pesquisa, que foi desenvolver um protótipo funcional para verificar a viabilidade do modelo proposto utilizando-se documentos para a extração do perfil de especialista.

Adicionalmente, este trabalho foi além do objetivo específico e apresentou exemplos para a demonstração de viabilidade do modelo proposto, extraindo perfis de forma automática de três especialistas voluntários. Esses perfis foram comparados aos perfis públicos dos especialistas mantidos na rede social LinkedIn. Dando sequência a esse procedimento, ao se realizar uma verificação de similaridade com o algoritmo de cosseno dos perfis extraídos pelo protótipo em relação ao perfil mantido na rede social, chegou-se ao índice de **0,7074** de similaridade. Esse resultado demonstrou que o modelo conceitual e o protótipo funcional podem representar uma robusta opção para a geração automática de perfis de especialista a partir de seus documentos.

Contudo, ainda são necessários mais testes para que seja possível indicar com mais propriedade o modelo proposto como uma das

soluções definitivas para que se consiga a geração de perfis de especialistas. Essas ponderações são observadas mais adiante como indicações de trabalhos futuros. A seguir, são apresentadas as contribuições provenientes desta pesquisa.

5.1 CONTRIBUIÇÕES

Ao se atingir o objetivo proposto deste trabalho e, conseqüentemente, os objetivos específicos, entende-se que esta pesquisa resultou em três contribuições principais vinculadas a esses objetivos específicos. Tais contribuições são apresentadas a seguir.

A primeira contribuição se dá pela revisão integrativa da literatura no que tange aos temas de perfil de especialista, de localização de expertise, de gestão do conhecimento e de engenharia do conhecimento. Tal revisão proporciona uma visão geral de como essas temáticas estão sendo tratadas de forma interdependente, como também é importante ressaltar que os resultados dessa revisão sistemática culminaram na proposta de modelo para a extração de perfis de especialistas a partir de documentos não estruturados. De forma complementar, é apresentado um guia de como realizar a revisão integrativa e a diferença do método da revisão integrativa com relação ao método tradicional de revisão. Esse é um importante recurso para se explorar nos trabalhos quando, por exemplo, inicia-se uma pesquisa de doutorado.

A segunda contribuição que esta pesquisa gerou foi a criação do modelo conceitual para a extração de perfis de especialista a partir de documentos (e.g., documentos nos formatos DOC, DOCX, PDF, TXT, PPT, PPTX, entre outros). Como já foi mencionado, o modelo conceitual proposto é fruto da combinação de diferentes abordagens identificadas na literatura para se tratarem as temáticas envolvidas nesta pesquisa. Para futuras evoluções ou propostas de novas abordagens, este trabalho poderá ser utilizado como base de comparação. Além disso, a pesquisa abordou a evolução dos conceitos relacionados a um especialista com o cruzamento da dimensão tempo, proposta de Bovo (2010). Nenhum dos trabalhos de EL identificados na revisão sistemática abordava tal dimensão.

A terceira contribuição observada foi a implementação do modelo proposto em forma de ferramenta que está disponibilizada como código aberto para a comunidade. Essa contribuição possui três papéis importantes: a contribuição em termos comparativos para uma nova implementação do modelo proposto, com base em PLN; a utilização do componente gerado em uma possível integração com ferramentas de

Expertise Location existente; e, por último, a ferramenta desktop disponibilizada para que qualquer especialista extraia o seu perfil a partir de seus documentos.

5.2 TRABALHOS FUTUROS

Ao longo do desenvolvimento desta pesquisa, foram identificadas possibilidades de melhorias futuras para cada um dos objetivos específicos mencionados, assim como foram reveladas oportunidades de pesquisas que podem ser aplicadas para se avançar no conhecimento produzido com a pesquisa proposta. A seguir, são apresentadas algumas sugestões.

Quanto à revisão integrativa da literatura para identificar as abordagens que possibilitaram a extração de perfil de especialista, observou-se que duas áreas podem ser mais trabalhadas no levantamento bibliográfico: a área de recuperação de informação e a área de extração de informação. Em uma próxima pesquisa, sugere-se incluir esses temas para encontrar novas abordagens possíveis para se trabalhar com os documentos de especialistas no intuito de se produzirem insumos para a extração de conhecimentos e interesses a partir de seus documentos.

Outra sugestão de trabalho futuro, principalmente para a realização de um doutorado, é a identificação do estado da arte quanto ao tema *Expertise Location*. Isto é, sugere-se pesquisar desde os primeiros sistemas de EL até os sistemas mais atuais que não estão vinculados às áreas de pesquisa deste trabalho, com o propósito de traçar um panorama da área. O trabalho de Balog et al. (2012) pode guiar essa visão do estado da arte realizada até 2012.

Quanto ao modelo conceitual proposto, identificou-se na literatura resultante da revisão integrativa a possibilidade de utilização de Web Mining, da técnica de TF-IDF e da ontologia FOAF, os quais podem ser explorados numa nova proposta de modelo, já que não foram trabalhados nesta pesquisa.

Quanto à implementação do protótipo, nesta pesquisa foi utilizada a base SKOS da DBPedia nas línguas portuguesa e inglesa. Ocorre que essas duas bases juntas possuem mais de 1,5 milhão de conceitos. Isso pode gerar uma quantidade excessiva de anotações que não são relevantes para os especialistas. Por isso, sugere-se a criação de uma visão final de perfil por nodos da SKOS da DBPedia. Por exemplo, uma visão que apresente somente as anotações extraídas do perfil de um especialista e que sejam filhas de um nodo específico da DBPedia como [http://dbpedia.org/page/Category:Subfields by academic discipline](http://dbpedia.org/page/Category:Subfields_by_academic_discipline). Essa visão poderia trazer apenas o conteúdo referente aos campos de

estudos acadêmicos, demonstrando somente os conceitos que estão ligados aos campos científicos presentes nos documentos do especialista.

Ainda no tocante à implementação do protótipo, sugere-se a exploração das técnicas de PLN como *streaming* e *lemmatization* para a melhoria da busca de conceitos da SKOS da DBPedia no índice Lucene. Contudo, a inclusão definitiva dessas técnicas precisa ser observada e comparada por meio do uso de estatísticas da área de RI, como, por exemplo, *precision e recall*.

Outra opção para a melhoria do protótipo é a utilização de implementações de NER que pudessem anotar pessoas e organizações a partir dos documentos do especialista. Além disso, como resultado complementar, a etapa de extração do perfil de especialista e a implementação dos instrumentos visuais poderiam utilizar um componente de visualização de redes para que se identifique com quais pessoas e organizações o especialista vem se relacionando ao longo dos anos.

Sugere-se, ainda, a análise de uma nova implementação de algoritmo de correlação entre datas e conceitos presentes no documento, conforme apresentado na seção 4.2.5 de associação de entidades. Essa implementação pode trabalhar também os elementos de ligação dos conceitos e das datas ao longo dos anos e, dessa forma, possibilitar uma nova visão dos conceitos tratados pelo especialista.

Ao se falar de correlação, abre-se espaço também para uma implementação de protótipo quanto à visualização das ligações latentes entre os conceitos que estão num mesmo documento ou até mesmo numa mesma sentença do documento. Essa implementação pode seguir as linhas de pesquisas destacadas nos trabalhos de Gonçalves (2006) e de Bovo (2011).

Por fim, sugere-se a continuidade na verificação do modelo proposto quanto à sua aplicabilidade em ferramentas de EL. Para isso, pode-se trabalhar mais a comparação de diferentes perfis de especialistas gerados com perfis preenchidos manualmente. Essa comparação com um maior número de especialistas pode trazer uma métrica mais confiável no que tange à similaridade entre os perfis gerados e os perfis fornecidos. Tal métrica pode, então, servir de base para se comparar a implementação de novos protótipos e melhorias a partir do modelo proposto.

REFERÊNCIAS

ACKOFF, R. L. From data to wisdom. **Journal of Applied Systems Analysis**, v. 16, p. 3-9, 1989.

ADOLFO, L. B. **Uma ontologia de apoio à classificação de processos judiciais**. 2013. Dissertação (Mestrado em Engenharia e Gestão do Conhecimento) – Universidade Federal de Santa Catarina, Florianópolis, 2013.

ALAVI, M.; LEIDNER, D. E. Knowledge management and knowledge management systems: conceptual foundations and research issues. **MIS Quarterly**, Minnesota, v. 25, n. 1, p. 107-136, 2001.

_____. Knowledge Management systems: emerging views and practices from the field. In: HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES, 32., 1999, Hawaii. **Proceedings...** Hawaii, 1999.

ALEXANDER, P. A.; SCHALLERT, D. L.; HARE, V. C. Coming to terms: how researchers in learning and literacy talk about knowledge. **Review of Educational Research**, v. 61, n. 3, p. 315-343, 1991.

ANAND, A.; SINGH, M. D. Understanding knowledge management: a literature review. **International Journal of Engineering Science and Technology**, v. 3, n. 2, p. 926-939, Feb. 2011.

APOLINÁRIO, F. **Metodologia da ciência: filosofia e prática da Pesquisa**. São Paulo: Cenage Learning, 2012. p. 3-14.

APQC - American Productivity & Quality Center. **2015 Knowledge Management Priorities**. Disponível em: <https://www.apqc.org/knowledge-base/download/340305/K05834_2015_KM_Priorities_ExecSum.pdf>. Acesso em: 2 abr. 2015.

ARGYRIS, C. **On organizational learning**. Cambridge, MA: Blackwell, 1993.

AUNE, B. **Rationalism, empiricism, and pragmatism: an introduction**. New York: Random House, 1970.

AWAD, M. A.; GHAZIRI. H. M. **Knowledge management**. Upper Saddle River, NJ: Pearson Education International, 2004.

BALOG K.; RIJKE M. **Determining expert profiles** (with an application to expert finding). 2007.

BALOG, K. et al. Expertise Retrieval. **Foundations and Trends in Information Retrieval**, [S.L], v. 6, n. 2-3, p. 127-256, 2012.

BECERRA-FERNANDEZ, I. Searching for experts on the web: a review of contemporary expertise locator systems. **ACM Transactions on Internet Technology**, New York, v. 6, n. 4, p. 333-355, 2006.

_____. The role of artificial intelligence technologies in the implementation of People-Finder knowledge management systems. **Knowledge-Based Systems**, Miami, v. 13, n. 5, p. 315-320, 2000.

BECKMAN, T. The current state of knowledge management. In: LIEBOWITZ, J. (Ed.). **Knowledge Management Handbook**. Boca Raton: CRC Press, 1999.

BENDER, S.; FISH, A. The transfer of knowledge and the retention of expertise: the continuing need for global assignments. **Journal of Knowledge Management**, v. 4, n. 2, p. 125-37, 2000.

BENTO, A. Como fazer uma revisão da literatura: considerações teóricas e práticas. **Revista JA (Associação Acadêmica da Universidade da Madeira)**, n. 65, p. 42-44, ano VII.

BEPPLER, F. D. **Um modelo para recuperação e busca de informação baseado em ontologia e no círculo hermenêutico**. 2008. Tese (Doutorado em Engenharia e Gestão do Conhecimento) – Universidade Federal de Santa Catarina, Florianópolis, 2008.

BERNARDO, W. M.; NOBRE, M. R. C.; JATENE, F. B. A prática clínica baseada em evidências: parte II - buscando as evidências em fontes de informação. **Revista da Associação Médica Brasileira**, São Paulo, v. 50, n. 1, p. 104-8, 2004.

BOEVA, V.; KRUSHEVA, M.; Tsiporkova, E. **Measuring expertise similarity in expert networks**. 2012.

BOHN, R. E. Measuring and managing technological knowledge. **Sloan Management Review**, v. 26, n. 1, p. 61-73, '1994.

BOTELHO, L. L. R.; CUNHA, C. C. A.; MACEDO, M. O método da revisão integrativa nos estudos organizacionais. **Gestão e Sociedade**, Belo Horizonte, v. 5, n. 11, p. 121-136, 2011.

BOVO, A. B. **Um modelo de descoberta de conhecimento inerente à evolução temporal dos relacionamentos entre elementos textuais**. 2011. Tese (Doutorado em Engenharia de Produção) – Universidade Federal de Santa Catarina, Florianópolis, 2011.

BRODER, A. Z. et al. Syntactic clustering of the web. **Computer Networks and ISDN Systems**, v. 29, n. 8, p. 1157-1166, 1997.

BROWN, J. S.; DUGUID, P. Organizing knowledge. **California Management Review**, California, v. 40, n. 3, p. 90-111, 1998.

BUKOWITZ, W.; WILLIAMS, R. **The knowledge management field book**. London: Financial Times/Prentice Hall, 2000.

CASTILHO, E. UIMA Tutorial – uimaFIT & DKPro Core. **3rd UIMA@GSCL Workshop, GSCL 2013, Darmstadt**. Disponível em: <<http://www.informatica.uniroma2.it/upload/2015/IA2/GSCL2013UIMATutorialUKP.pdf>>. Acesso em: 1. fev. 2016.

CASTILHO, E.; GUREVYCH, I. A broad-coverage collection of portable NLP components for building shareable analysis pipelines. In: WORKSHOP ON OPEN INFRASTRUCTURES AND ANALYSIS FRAMEWORKS FOR HLT, 2014, Dublin. **Proceedings...** Dublin, 2014, p. 1-11.

CECI, F. **Um modelo semi-automático para a construção e manutenção de ontologias a partir de bases de documentos não estruturados**. 2010. 131 f. Dissertação (Mestrado em Engenharia de Gestão do Conhecimento) – Universidade Federal de Santa Catarina, Florianópolis, 2010.

CHADEGANI et al. A Comparison between two main academic literature collections: Web of Science and Scopus Databases. **Asian Social Science**, v. 9, n. 5, p. 18-26, abr. 2013.

CHEN, H. **Knowledge management systems: a text mining perspective.** Tucson: The University of Arizona Press, 2001.

CHOI, B.; LEE, H. Knowledge management strategy and its link to knowledge creation process. **Expert Systems with Applications**, v. 23, p. 173-187, 2002.

COOK, D. J.; MULROW, C. D.; RAYNES, R. B. Systematic reviews: synthesis of best evidence for clinical decisions. **Annals of Internal Medicine**, Philadelphia, v. 126, n. 5, p. 376-380, Mar. 1997.

COOK, S. N. D.; BROWN, J. S. Bridging epistemologies: the generative dance between organizational knowledge and organizational knowing. **Organization Science**, v. 10, n. 4, p. 381-400, 1999.

CORDEIRO, A. M. et al. Revisão sistemática: uma revisão narrativa. **Revista do Colégio Brasileiro de Cirurgiões**, Rio de Janeiro, v. 34, n. 5, p. 428-431, nov./dez., 2007.

COUNSELL, C. Formulating questions and locating primary studies for inclusion in systematic reviews. **Annals of internal medicine**, v. 127, n. 5, p. 380-387, 1997.

CRESWELL, J. W. **Projeto de pesquisa: métodos qualitativo, quantitativo e misto.** 3. ed. Porto Alegre: Artmed, 2010. p. 25-47.

DAGHFOUS, A.; KAH, M. M. O. Knowledge management implementation in SMEs: a framework and a case illustration. **Journal of Information and Knowledge Management**, v. 5, n. 2, p. 107-115, 2006.

DALKIR, K. **Knowledge management in theory and practice.** Burlington: Elsevier Butterworth-Heinemann, 2005.

DAVENPORT, T. H. Ten principles of knowledge management and four case studies. **Knowledge and process Management**, v. 4, n. 3, p. 187-208, 1997.

DAVENPORT, T. H.; PRUSAK, L. **Working knowledge: how organizations manage what they know.** Boston: Harvard Business School Press, 1998.

DESOUZA, K. C. **New frontiers of knowledge management**. New York: Palgrave Macmillan, 2005.

DOMINGOS, M. **Uma arquitetura de referência para sistemas de informação e portais de serviços de governo eletrônico**. 2004. Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal de Santa Catarina, Florianópolis, 2004.

EGC – Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento. **Áreas de Concentração**. UFSC, Florianópolis. Disponível em: <<http://www.egc.ufsc.br/index.php/pt/egc/pos-graduacao/programa/areas-de-concentracao>>. Acesso em: 26 jul. 2015a.

EGC – Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento. **O programa**. UFSC, Florianópolis. Disponível em: <<http://www.egc.ufsc.br/index.php/egc/pos-graduacao/programa>>. Acesso em: 26 jul. 2015b.

FREITAS, V. et al. A pesquisa científica e tecnológica. **Revista de Educação Técnica e Tecnológica em Ciências Agrícolas**, ISSN 2177-8086. 2012.

GALLUP, S. D.; DATTERO, R.; HICKS, R. C. Knowledge management systems: an architecture for active and passive knowledge. **Information Resource Management Journal**, v. 15, n. 1, p. 22-7, 2002.

GLASS, G. V. Primary, secondary, and meta-analysis of research. **American Educational Researcher Association**, Denver, v. 5, n. 10, p. 3-8, 1976.

GOLDMAN, A. Discrimination and perceptual knowledge. **The Journal of Philosophy**, n. 73, p. 771-791, 1976.

GONÇALVES, A. L. **Um modelo de descoberta de conhecimento baseado na correlação de elementos textuais e expansão vetorial aplicado à engenharia e gestão do conhecimento**. 2006. 196 f. Tese (Doutorado em Engenharia de Produção) – Universidade Federal de Santa Catarina, Florianópolis, 2006.

GRANT, R. M. Toward a knowledge-based theory of the firm. **Strategic Management Journal**, Washington-DC, v. 17, p. 109-122, 1996.

GREENHALGH, T. How to read a paper: papers that summarize other papers (systematic reviews and meta-analyses). **BMJ Clinical Research**, Londres, v. 315, n. 7109, p. 672-675, 1997.

GRUBER, T. A translation approach to portable ontology specifications. **Knowledge Acquisition**, Califórnia, v. 5, n. 2, p. 199-220, 1993.

GUARINO, N. **Formal ontology and information systems**. FOIS'98. Amsterdam: IOS Press, 1998.

HANSEN, M. T. The search-transfer problem: the role of weak ties in sharing knowledge across organization subunits. **Administrative Science Quarterly**, v. 44, p. 82-111, 1999.

HERRERA, F.; MARTÍNEZ, L. A 2-tuple fuzzy linguistic representation model for computing with words. **IEEE Transactions on Fuzzy Systems**, v. 8, n. 6, p. 746-752, 2000.

IDC - International Data Corporation. **Extracting Value from Chaos**. Boston: Emc Corporation, 2010. 12 p.

KENDAL, S.; CREEN, M. **An Introduction to Knowledge Engineering**, Springer, 2006.

KING, J. Editorial notes. **Information Systems Research**, v. 4, n. 4, p. 291-8, 1993.

KNELLER, G. F. **A ciência como atividade humana**. Rio de Janeiro: Zahar, 1980. p. 11-35.

KOCK, N.; MCQUEEN, R. Knowledge and information communication in organizations: an analysis of core, support and improvement process. **Knowledge and Process Management**, v. 5, n. 1, p. 29-40. 1998

KUMAR, K. **Comparative analysis of search features of Scopus and Web of Science**. Department of Library & Information Science Sri Venkateswara University, 2013.

LEE, K. C.; LEE, S.; KANG, I.W. Knowledge management: measuring knowledge management performance. **Information and Management**, v. 42, n. 3, p. 469-82, 2005.

LI, M.; LIU, L.; LI, C.-B. **An approach to expert recommendation based on fuzzy linguistic method and fuzzy text classification in knowledge management systems**. 2011.

LIU, D.-R. et al. **Integrating expert profile, reputation and link analysis for expert finding in question-answering websites**. 2013.

LOH, S. **Abordagem baseada em conceitos para descoberta de conhecimentos em textos**. 2011. Tese (Doutorado em Ciência da Computação) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2001.

MCCANDLESS, M. **Searching relational content with Lucene's BlockJoinQuery**. 2012. Disponível em: <<http://blog.mikemccandless.com/2012/01/searching-relational-content-with.html>>. Acesso em: 1. fev. 2016.

MARTINEZ, A. et al. Las categorías o facetas fundamentales: una metodología para el diseño de taxonomías corporativas de sitios Web argentinos. **Ciência da Informação**, Brasília, v. 33, n. 2, p.106-111, maio/ago. 2004.

MARWICK, A. Knowledge management technology. **IBM Systems Journal**, v. 40, n. 4, p. 814-830, 2001.

MAYBURY, M. **Expert Finding Systems**. Bedford, Massachusetts: Center for Integrated Intelligence Systems, 2006.

MAYER, M.; ZACK, M. The design and implementation of information Products. **Sloan Management Review**, v. 37, n. 3, p. 45-59, 1996.

MATTMANN, C. A.; ZITTING, J. L. **Tika in action**. Shelter Island: Manning Publications Co., 2012.

MCELROY, M. The knowledge life cycle. In: ICM CONFERENCE ON KNOWLEDGE MANAGEMENT, 1999, Miami, Flórida, 1999.

MCDONALD, D. W.; ACKERMAN, M. S. Just talk to me: a field study of expertise location. In: ACM CONFERENCE ON COMPUTER SUPPORTED COOPERATIVE WORK, 1998, Seattle, Washington, USA. **Proceedings...** Seattle, Washington, USA, 1998. p. 14-18.

MENDES, K. D. S.; SILVA, R. C. C. P; GALVÃO, C. M. Revisão integrativa: método de pesquisa para a incorporação de evidências na saúde e na enfermagem. **Texto e Contexto Enfermagem**, Florianópolis, v. 17, n. 4, p. 758, out. 2008.

MORGAN, G. Paradigms, Metaphors, and Puzzle Solving in Organization Theory. **Administrative Science Quarterly**, v. 25, p. 605-622, 1980.

NAIR, P.; PRAKASH, K. **Knowledge management: facilitator's guide**. Tokyo: Asion Productivity Organization (APO), 2009.

NASCIMENTO, R. N. do et al. Classificação semântica de informações não estruturadas sobre colaboradores. In: CONGRESSO INTERNACIONAL DO CONHECIMENTO E INOVAÇÃO, 4., 2014, Loja, Equador. **Proceedings...** Loja, Equador, 2014.

NONAKA, I. A dynamic theory of organizational knowledge creation. **Organization Science**, v. 5, n. 1, p. 14-37, 1994.

NONAKA, I; TAKEUCHI, H. **The knowledge-creating company: how Japanese companies create the dynamics of innovation**. Oxford: Oxford University Press, 1995.

NORTH, K.; RIVAS, R. **Gestión del conocimiento. Una guía práctica hacia la empresa inteligente**. Libros en Red. 2008.

POLANYI, M. **The tacit dimension**. New York: Doubleday, 1966.

RAO, M. **Knowledge management tools and techniques: practitioners and experts evaluate KM Solutions**. Burlington: Elsevier, Butterworth-Heinemann, 2005.

RAUTENBERG, S.; TODESCO, J. L.; STEIL, A. V. Ontologias de domínio no mapeamento de instrumentos da gestão do conhecimento e de agentes computacionais da engenharia do conhecimento: o estado da arte. **Perspectivas em Ciência da Informação**, v. 15, n. 2, p. 163-182, maio/ago. 2010.

ROSCOE, D. D.; JENKINS, S. A Meta-Analysis of Campaign Contributions' Impact on Roll Call Voting. **Social Science Quarterly**, v. 86, n. 1, p. 52-68, 2005.

SCHREIBER, G. et al. **Knowledge engineering and management: the CommonKADS methodology**. Massachusetts: MIT Press, 1999.

SELL, D. **Uma arquitetura para business intelligence baseada em tecnologias semânticas para suporte a aplicações analíticas**. 2006. Tese (Doutorado em Engenharia de Produção) – Universidade Federal de Santa Catarina, Florianópolis, 2006.

SILVA, E. L. da; MENEZES, E. M. **Metodologia da pesquisa e elaboração de dissertação**. 4. ed. Florianópolis: UFSC, 2005.

SMITH, E. A. The role of tacit and explicit knowledge in the workplace. **Journal of Knowledge Management**, v. 5, n. 4, p. 311-321, 2001.

SOHAIL, A.; AFZAL, M. T.; AHMAD J. **Discovery of medical experts**. 2013.

SPENDER, J. C. Making knowledge the basis of a dynamic theory of the firm. **Strategic Management Journal**, v. 17, p. 45-62, 1996.

STEIL, A. V. **Estado da arte das definições de gestão do conhecimento e seus subsistemas**. Technical Report. Florianópolis: Instituto Stela, 2007.

STROTGEN, J.; GERTZ; M. HeidelTime: High quality rule-based extraction and normalization of temporal expressions. In: INTERNATIONAL WORKSHOP ON SEMANTIC EVALUATION, 5., 2010, Suécia. **Proceedings...** Suécia, 2010. p. 321-324.

STUDER, R.; BENJAMINS, R; FANSEL, D. Knowledge Engineering: Principles and methods. **Data & Knowledge Engineering**, v. 25, n. 1, p. 161-197, 1998.

SUN, Z.; HAO, G. HSM: a hierarchical spiral model for knowledge management. In: INTERNATIONAL CONFERENCE ON INFORMATION MANAGEMENT AND BUSINESS (IMB2006), 2., Sydney. **Proceedings...** Sydney, 2006. p. 542-555.

TEECE, D. J. Capturing value from knowledge assets: the new economy, markets for know-how, and intangible assets. **California Management Review**, California, v. 40, n. 3, p. 55-79, 1998.

TEECE, D. J.; PISANO, G.; SHUEN, A. Dynamic capabilities and strategic management. **Strategic Management Journal**, v. 18, n. 7, p. 509-533, 1997.

THIERAUF, R.; HOCTOR, J. **Optimal knowledge management**. Hershey, PA: Idea Group, 2006.

UNIVERSITY OF WASHINGTON. **Scopus and Web of Science Comparison Chart**. 2015. Disponível em: <<http://guides.lib.uw.edu/c.php?g=99232&p=642081>>. Acesso em: 5 jun. 2015.

URSI, E. S. **Prevenção de lesões de pele no perioperatório**: revisão integrativa da literatura. 2005. 130 f. Dissertação (Mestrado em Enfermagem) – Escola de Enfermagem, Universidade de São Paulo, Ribeirão Preto, 2005.

VAN DER SPEK, R.; SPIJKERVET, A. Knowledge management: dealing intelligently with knowledge. In: LIEBOWITZ, W. (Ed.). **Knowledge management and its integrative elements**. Boca Raton: CRC Press, 1997.

VENZIN, M.; KROGH, G.; ROOS, J. Future Research into Knowledge Management. In: KROGH, G.; ROOS, J.; KLEINE, D. **Knowing in Firms**. Londres: Sage Publication, 1998.

W3C - World Wide Web Consortium. **SKOS Simple Knowledge Organization System Reference**, 2009.

WEBSTER, J.; WATSON, J. T. Analyzing the past to prepare for the future: writing a literature review. **MIS Quarterly & The Society for Information Management**, v. 26, n. 2, p. 13-23, 2002.

WHITTEMORE, Robin; KNAFL, Kathleen. The integrative review: updated methodology. **Journal of Advanced Nursing**, v. 52, n. 5, p. 546-553, 2005.

WIIG, K. M. Integrating Intellectual Capital and Knowledge Management. **Long Range Planning**, v. 30, n. 4, p. 399-405, 1997.

WIIG, K. M. **Knowledge management foundations: thinking about thinking – how people and organizations create, represent and use knowledge**. Arlington: Schema Press, 1993.

WINTER, S. **Knowledge and competence as strategic assets**. Cambridge: Ballinger Publishing, 1987.

WONG, K. Y.; ASPINWALL, E. Characterizing knowledge management in the small business environment. **Journal of Knowledge Management**, v. 8, n. 3, p. 44-61, 2004.

YANG, K.-H et al. **EFS: expert finding system based on wikipedia link pattern analysis**. 2008.

YANG, K.-W.; HUH, S.-Y. **Automatic expert identification using a text categorization technique in knowledge management systems**. 2008.

YAO, L.; TANG, J.; LI, J. **A unified approach to researcher profiling**. 2007.

YOUNG, R. **Knowledge Management Tools and Techniques Manual**. Japan: Hirakawacho, Chiyoda-ku: Asian Productivity Organization (APO), 2010.

ZACK, M. H. Managing codified knowledge. **MIT Sloan Management Review**, jul. 1999.

APÊNDICE A – PERFIL COMPLETO DO ESPECIALISTA #1



2015	<p>Dados (1280) Conhecimento (1275) Trabalho (1045) Gestão (928) Plataforma (922) Sistema (879) Serviços (838) Análise (761) Desenvolvimento (592) Pesquisa (539) Valor (461) Referência (459) Sistemas (455) Informação (409) Plano (407) Registro (403) Qualidade (402) Area (400) Área (400) Taxonomia (395) Days (382) Documentos (372) Estrutura (369) Identificação (358) Fontes (349) Arquitetura (323) Produtos (319) Objeto (318) Apresentação (317) Gestão do conhecimento (303) Pessoas (300) Currículos (294) Gerenciamento (291) Nome (279) Ações (258) Conteúdo (255) Projetos (251) Comunicação (245) Planejamento (244) Visão (242) Estado (231) Problemas (226) Governo (225) Servidores (218) Data (217) Avaliação (215) Tempo (214) Ambiente (213) Dias (213) Estratégia (212) Filtros (211) Software (203) Software (203) Ferramentas (190) Instituição (188) Organizações (185) Engenharia (177) Padrões (175) Conceitos (174) Técnicas (164) Gestão de processos (162) Meses (162) Titulação (162) Acompanhamento (161) Inovação (161) Metodologia (160) Evolução (159) Estudos (154) Clientes (151) Infraestrutura (137) Reunião (137) Tipos (136) Web (136) Ação (134) Nacional (130) Círculos (129) Diagramas (126) Eventos (126) Publicação (123) Anos (121) Construção (121) Negócios (105) Notação (104) Testes (101) SILVA (101) Modelos (98) Normas (97) Telefone (95) Vi (94) Sociedade (93) Atores (92) Gestores (92) Cultura (91) Rede (90) Segurança (85) Tecnologia (85) Custos (84) Ontologia (81) Papel (81) Brasil (77)</p>
------	---

2014	<p>Conhecimento (7534) Plataforma (6137) Gestão (4746) Sistema (4472)Desenvolvimento (4406) Pesquisa (4133) Inovação (4062) Dados (3986)Gestão do conhecimento (3282) Florianópolis (3267) Florianópolis (3267)Serviços (3082) Análise (3053) Trabalho (2942) Projetos (2636)Santos (2539) Santos (2539) Sistemas (2415) Empresas (2278)Informação (2172) Valor (2016) Engenharia (2008) Governo (1919)Roberto Carlos (1899) Brasil (1842) Tecnologia (1814) HTTP (1766)HTTP (1766) Software (1689) Software (1689) Area (1607) Área (1607)Data (1591) Nacional (1588) Ambiente (1518) Organizações (1512)Estado (1481) Plano (1477) Rede (1434) Montenegro (1433)Montenegro (1433) Pessoas (1364) Estratégia (1338) Identificação (1275)Ações (1248) Avaliação (1239) Planejamento (1224) Instituição (1191)Produtos (1188) Gerenciamento (1130) Visão (1098) Arquitetura (1069)Comunicação (1019) Knowledge (951) Sistemas de informação (933)Ciência (916) Currículos (904) Sociedade (901)Engenharia de produção (896) Documentos (882) Registro (878)Qualidade (867) Santa Catarina (860) Santa Catarina (860) Eventos (851)Educação (845) Nome (836) Guimarães (822) Guimarães (822)Taxonomia (813) Dias (812) Referência (811) Metodologia (808)Capital (790) Management (790) Capital (790) Web (778) Direitos (772)Técnicas (770) Estrutura (769) Rio de Janeiro (762) Rio de Janeiro (762)Apresentação (750) SILVA (748) Information (738) Fontes (726)Tempo (720) Ação (719) Segurança (711) Construção (708)Martins (679) Problemas (676) Anos (669) Objeto (666)Instituições (650) Mercado (638) Marcos (636) Aprendizagem (633)Componentes (627) Novas (620)</p>
2013	<p>SENAI (10965) Av (7800) Trabalho (7280) Fortaleza (6716)Fortaleza (6716) Conhecimento (5521) Educação (4839) Técnicas (4663)Sistema (4344) Plataforma (4294) Dados (4173) Equipamentos (3915)Desenvolvimento (3689) Gestão (3590) Segurança (3508) Area (3354)Área (3354) Qualidade (2926) Tecnologia (2914) Sistemas (2878)Almeida (2854) Ibiapina (2781) Avaliação (2753) Nome (2615)Ceará (2577) Ceará (2577) Pesquisa (2572) Normas (2545)Ambiente (2453) Análise (2431) Máquinas (2269) Barra (2191)Serviços (2188) Plano (2181) Empresas (2146) Saúde (2133)Projetos (2047) Materiais (2034) Indústria (2000) Dias (1986)Ferramentas (1976) Rodrigues (1907) Rodrigues (1907) Tipos (1885)Informação (1881) Produtos (1851) Maracanaú (1790)Distrito Industrial (1785) Planejamento (1608) Inovação (1546)Aprendizagem (1457) Rua (1445) Registro (1385) Data (1348)Documentos (1310) Componentes (1301) Identificação (1293)Nacional (1245) HTTP (1217) HTTP (1217) Pessoas (1211)Conceitos (1208) Evolução (1146) Comunicação (1114) Media (1113)Engenharia (1104) Tempo</p>

	(1095) Rede (1075) Valor (1069)Visão (1050) Custos (1031) Ensino (1020) Anos (997) Ações (997)Apresentação (995) Mercado (950) Gestão do conhecimento (904)Ação (892) Fontes (868) Mecânica (864) Construção (857)Conteúdo (850) Padrões (829) Ambientes (787) Brasil (781)Problemas (775) Estrutura (770) Situações (757) Software (753)Software (753) Princípios (745) Novas (725) Natureza (718)Alimentos (707) Ontologia (704) Metodologia (694) Administração (691)Círculos (691) Gerenciamento (673) João Pessoa (660)
2012	Conhecimento (6657) Plataforma (5975) Knowledge (4894) Sistema (4651)Desenvolvimento (4248) Dados (3556) Gestão (3537) Pesquisa (3260)Serviços (3208) Inovação (2677) Data (2663) Sistemas (2444)Projetos (2187) Trabalho (2121) Information (2098) Capital (2026)Capital (2026) Análise (2016) Empresas (1950) Valor (1873)Informação (1846) Nacional (1731) Tecnologia (1631) Area (1499)Área (1499) Plano (1417) Turismo (1367) Produtos (1341)Avaliação (1333) Identificação (1272) Software (1255) Software (1255)HTTP (1225) HTTP (1225) Rede (1181) Brasil (1169)Documentos (1153) Governo (1134) Ambiente (1111) Planejamento (1082)Knowledge management (1048) Indústria (1010) Example (1008)SENAI (1007) Management (993) Engenharia (990) Ál (984)Florianópolis (980) Florianópolis (980) Ações (975) People (967)Qualidade (956) Pessoas (929) Dias (927) Santos (897) Santos (897)Gestão do conhecimento (892) Systems (878) Research (867) Web (859)Apresentação (858) Tempo (846) Sistemas de informação (844)Ciência (822) Registro (815) Arquitetura (812) Mercado (812)Fontes (798) Choreography (796) Processes (787) Comunicação (771)Time (762) Objeto (744) Design (743) Design (743) Clientes (734)Referência (723) Educação (718) Estratégia (716) Componentes (713)Instituições (706) Documents (703) Notation (701) Business (696)Instituição (696) Business process (693) Anos (689) Value (688)Analysis (681) Administração (668) Visão (668) Models (663)Roberto Carlos (647) Nome (646) Direitos (640) Estrutura (636)Ação (624) Redes (624) Saúde (624) Events (616)

2011	<p>Dados (19120) Conhecimento (14974) Plataforma (8964)Desenvolvimento (7390) Projetos (7374) Data (7247) Pesquisa (7212)Sistema (7112) Valor (6359) Capital (6000) Capital (6000)Informação (5366) Serviços (5245) Gestão (5150) Ontologia (5045)Área (4758) Área (4758) Oracle (4713) Oracle (4713) Fontes (4639)Arquitetura (4562) HTTP (4245) HTTP (4245) Análise (4168)Trabalho (4117) Example (4043) Information (4033) Plano (4007)Knowledge (3972) Empresas (3860) Nome (3801) Fundos (3766)Inovação (3764) Value (3758) Governo (3743) Qualidade (3698)Estrutura (3590) Apresentação (3560) Nacional (3485) Ontology (3177)Sistemas (3021) Documentos (2729) Tecnologia (2729) Tempo (2729)Ações (2670) Brasil (2582) Estratégia (2580) Identificação (2542)Gerenciamento (2531) Pessoas (2392) Visão (2376) Engenharia (2332)Produtos (2240) Planejamento (2173) Management (2146) Avaliação (2113)Evolução (2099) Ação (1988) Receita (1984) Development (1976)Software (1976) Software (1976) Conceitos (1974) Art (1909) Ál (1908)Componentes (1894) Anos (1847) Text (1828) Organizações (1730)Filtros (1685) Natureza (1636) Geometry (1622) Days (1601)Rede (1594) Ciência (1592) Research (1568) Comunicação (1561)Metodologia (1544) Tema (1539) Ambiente (1538) Ferramentas (1535)Gestão do conhecimento (1507) Registro (1493) Estado (1487)Innovation (1476) Técnicas (1441) Construção (1419) Tabelas (1419)Instituições (1381) Time (1380) Administração (1374) Padrões (1369)Crédito (1330) Política (1316) Web (1315) Sociedade (1310)Custos (1305) Instituição (1296) SQL (1286) SQL (1286)</p>
-------------	---

2010	<p> Conhecimento (11558) Inovação (6020) Gestão (5986)Desenvolvimento (5937) Pesquisa (5435) Plataforma (4725)Informação (4571) Dados (4497) Projetos (3125) Sistema (3109)Sistemas (2942) Knowledge (2907) Serviços (2833) Engenharia (2623)Tecnologia (2480) Análise (2468) Fontes (2454) Nacional (2393)Gestão do conhecimento (2324) Ação (2289) Arquitetura (2201) Area (2111)Área (2111) Plano (2099) Data (2052) Information (1975)Produtos (1905) Trabalho (1821) Ações (1729) Empresas (1704)Planejamento (1604) Ciência (1578) Knowledge management (1556)Componentes (1545) HTTP (1519) HTTP (1519) Research (1431)AI (1421) Ontology (1362) Identificação (1346) Visão (1312)Estratégia (1266) Valor (1256) Ontologia (1243) Brasil (1233)Apresentação (1145) Organizações (1132) Avaliação (1108) Atores (1095)Embraer (1049) Software (1049) EMBRAER (1049) Embraer (1049)Software (1049) Governo (1040) Documentos (1017) Estrutura (1014)Ambiente (997) Acompanhamento (989) Innovation (988) Rede (986)Pessoas (981) Capital (973) Capital (973) Redes (965)Comunicação (956) Qualidade (930) Técnicas (916) Systems (913)Tempo (908) Web (887) Ferramentas (869) Sistemas de informação (796)Management (794) Pesquisadores (789) Construção (788)Gestão estratégica (788) Development (758) Semântica (758)Instituição (757) Anos (749) Conceitos (746) Instituições (745)Gerenciamento (744) Conteúdo (727) Business (717) Modelos (709)Pesquisas (703) Novas (700) Padrões (680) Estado (671) Política (668)Sociedade (665) Metodologia (650) Estudos (623) Tema (615)Technology (578) Referência (573) Bases (565) Bases (565) </p>
-------------	---

2009	<p>Plataforma (1976) Pesquisa (1731) Conhecimento (1611) Serviços (1195) Art (1100) Desenvolvimento (1036) Dados (1024) Gestão (1012) Análise (947) Projetos (891) Inovação (886) Sistema (858) Informação (807) Sistemas (768) Data (756) Instituição (726) Componentes (650) Empresas (612) Pessoas (581) Área (555) Área (555) Fontes (514) Valor (497) Ensino (496) Software (484) Software (484) Tecnologia (470) Produtos (466) Capital (444) Capital (444) Avaliação (442) Documentos (433) Mercado (417) Apresentação (405) Knowledge (396) Negócios (396) HTTP (379) HTTP (379) Plano (373) Trabalho (352) Nome (341) Web (324) Rede (322) Estratégia (310) Tempo (295) Conteúdo (293) Ambiente (289) Identificação (281) Dias (279) Anos (278) Inteligência competitiva (277) Relações (277) Clientes (275) Redes (266) Visão (266) Florianópolis (261) Florianópolis (261) Currículos (251) Planejamento (250) Gestão do conhecimento (249) Innovation (242) Nacional (241) Meses (241) Pesquisadores (239) Mapas (238) Professores (233) Normas (225) Índices (224) Administração (223) Instituições (222) Universidade do Contestado (219) Estrutura (212) Evolução (208) Financiamento (208) Brasil (206) Engenharia (203) Objeto (199) Ontologia (195) Registro (188) Direitos (188) Ações (185) Time (180) Qualidade (172) Técnicas (171) Bases (169) Marketing (169) Bases (169) Marketing (169) Arquitetura (167) Filtros (164) Acompanhamento (161) Research (160) Estudos (159) Construção (156) Órgãos (156) Educação (155) Titulação (155) Softwares (152) Estado (151) Governo (151)</p>
2008	<p>Conhecimento (1002) Desenvolvimento (871) Dados (702) Sistema (688) Plataforma (669) Serviços (659) Pesquisa (616) Gestão (542) Análise (531) Informação (514) Software (499) Software (499) Biofuels (441) Área (438) Sistemas (438) Área (438) Mercado (409) Valor (398) Inovação (393) Componentes (360) Empresas (356) Projetos (356) Data (351) Tecnologia (336) Petrobras (320) Petrobras (320) Petrobrás (320) Documentos (301) Apresentação (294) Negócios (292) Gestão do conhecimento (286) Conteúdo (274) Planejamento (273) Fuel (267) Florianópolis (260) Florianópolis (260) Information (259) Produtos (255) Inteligência (247) Engenharia (240) Clientes (200) Tempo (198) Fontes (195) Nome (195) Pesquisas (194) Segurança (193) Ambiente (189) Qualidade (187) Ações (181) Estado (180) Trabalho (178) Ferramentas (176) Anos (175) Técnicas (164) Redes (162) Avaliação (159) Energia (159) Objeto (154) Dias (150) Filtros (148) Nacional (147) Novas (147) Santa Catarina (144) Santa Catarina (144) Justiça (143) Retailers (139) Taxonomia (139) Brasil (135) Instituição (134) Acompanhamento (133) Meses (133) Business intelligence (132) Business intelligence (132) Estudos (130) Plano (128) Comunicação (126) Arquitetura</p>

	<p>(124)Indexação (123) Construção (120) Rede (116) Visão (115)Biodiesel (114) Organizações (113) Oil (112) Identificação (112)Pessoas (111) Knowledge (109) Markets (108) Web (107)Direitos (105) Tema (104) Mapas (104) Demand (103) Business (102)Gerenciamento (102) Estrutura (101) Evolução (101) Sociedade (100)HTTP (97) HTTP (97)</p>
2007	<p>Conhecimento (3127) Desenvolvimento (2847) Plataforma (2829)Sistema (2804) Serviços (2436) Dados (2421) Sistemas (2190)Área (2125) Área (2125) Pesquisa (2074) Gestão (1954)Informação (1879) Software (1675) Software (1675) Inovação (1514)Data (1498) Produtos (1342) Componentes (1302) Análise (1187)Tecnologia (1160) Trabalho (1159) Fontes (1069) Ambiente (1051)Projetos (1039) Abril (894) Instituição (862) Atores (810) Plano (796)Empresas (770) Justiça (765) Segurança (747) Pessoas (695)Nacional (685) Clientes (684) Apresentação (679) Instituições (668)Arquitetura (662) Construção (650) Comunicação (627) Saúde (625)Reunião (616) Qualidade (614) Nome (613) Crédito (609)Metodologia (609) Ações (603) Conteúdo (592) Governo (583)Planejamento (564) Tempo (561) Registro (552) Web (538)Mercado (533) Redes (518) Comunidade (511) Ferramentas (496)Engenharia (487) Ál (483) Objeto (471) Administração (470)Novas (463) Agências (454) Gestão do conhecimento (450) HTTP (449)HTTP (449) Análise complexa (448) Business intelligence (445)Business intelligence (445) Rede (438) Montenegro (424) Montenegro (424)Meses (401) Santos (397) Santos (397) Valor (394) Dias (385)Documentos (385) Negócios (385) Avaliação (379) Brasil (379)Comunidades (372) Técnicas (371) Bases (362) Bases (362)Marketing (361) Marketing (361) Pesquisadores (345) Marcos (343)Semântica (342) Estado (336) Anos (333) Padrões (328) Canais (309)Interoperabilidade (307) Visão (306) Gerência de serviços (303) Mapas (296)Bank (295) Engenharia de produção (295) Gerenciamento (289)</p>

2006	<p>Sistema (2280) Dados (1806) Desenvolvimento (1487) Conhecimento (1325) Plataforma (1222) Serviços (1104) Informação (1101) Nome (1070) Sistemas (1048) Avaliação (1037) Inovação (994) Gestão (986) Instituição (971) Componentes (914) Ambiente (876) Pesquisa (856) Análise (806) Registro (717) Empresas (709) Estratégia (703) Produtos (676) Área (641) Área (641) Valor (636) Data (631) Trabalho (597) Projetos (587) Clientes (571) Nacional (570) Educação ambiental (566) Ações (504) Instituições (462) Estrutura (454) Conteúdo (448) Arquitetura (434) Web (418) Educação (417) Governo (410) Metodologia (408) Software (399) Software (399) Tempo (397) Construção (390) Organizações (388) Atores (382) Apresentação (381) Fontes (381) Sistemas de informação (360) Administração (353) Ensino (352) HTTP (351) HTTP (351) Qualidade (348) Norton (338) Visão (332) Comunicação (331) Tecnologia (320) Ál (315) XML (312) Comunidade (312) XML (312) Internet (309) Internet (309) Plano (297) Mercado (296) Professores (296) Redes (291) Currículos (288) Documentos (285) Novas (283) Custos (271) Referência (270) UNESCO (264) Unesco (264) Ação (264) UNESCO (264) Arquivos (262) Identificação (262) Relações (259) Value (253) Brasil (250) Estado (249) Sociedade (246) Gestão do conhecimento (240) Planejamento (236) Fundamental (232) Pessoas (231) Anos (217) Objeto (216) Gestores (215) Educadores (213) Ferramentas (212) Aprendizagem (211) Testes (203) Dias (201) Rede (200) Ensino superior (196) Negócios (195) Bases (194) Bases (194)</p>
2005	<p>Sistema (6485) Plataforma (5900) Dados (3411) Sistemas (3341) Value (3206) Educação (2891) Informação (2576) Avaliação (2419) Desenvolvimento (2198) Ambiente (2009) Data (1864) Conhecimento (1862) Gestão (1692) Componentes (1683) Análise (1634) Fontes (1589) Serviços (1553) Objeto (1519) Pesquisa (1462) Nacional (1447) Instituição (1397) Protected (1298) Instituições (1206) Filters (1106) Área (1105) Área (1105) Valor (1051) Arquitetura (981) Currículos (969) Nome (957) Interoperabilidade (949) Tecnologia (940) Sistemas de informação (931) Web (930) Conteúdo (920) Produtos (849) Dias (764) Governo (761) Registro (757) Atores (736) XML (733) Espaço (733) XML (733) Ensino superior (718) Bases (713) Bases (713) Erro (707) Projetos (700) Inovação (694) Software (658) Software (658) Pies (618) HTTP (614) HTTP (614) Apresentação (606) Documentos (587) Construção (523) Texto (517) Filter (515) Ambientes (511) Author (504) Identificação (495) Pesquisadores (482) Arquivos (477) Empresas (477) Metodologia (472) Trabalho (470) Meses (452) Ensino médio (443) Índices (441) Acompanhamento (437) Pessoas (430) Data</p>

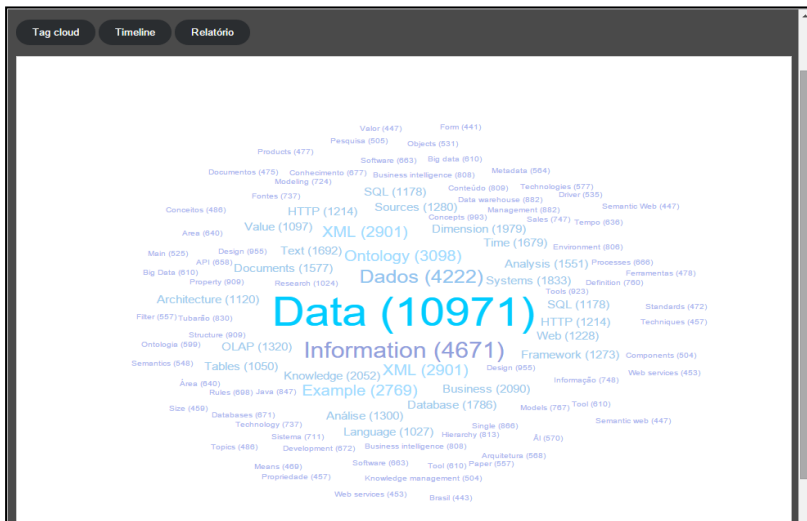
	warehouse (421) Tempo (404) Ensino (393)Brasil (392) Estado (389) Dinâmica (388) Sociedade (380) Visão (376)Ciência (374) Comissões (373) Comunidade (373) Indexação (357)Ações (356) Referência (350) Estudos (333) Padrões (325) Filtros (314)Novas (305) Engenharia (303) Rede (295) Saúde (295)Divulgação (279) Qualidade (279) Anos (278) Instituições de ensino (267)Técnicas (267) Ferramentas (265) Publicação (264)
2004	Água (551) Ambiente (456) Solo (435) Sistema (422)Recursos naturais (322) Vocabulário (307) Estrutura (215) Sistemas (179)Nome (165) Conhecimento (159) Desenvolvimento (151) Tempo (149)Ação (146) Minerais (145) Area (140) Área (140) Plantas (137)Terra (123) Solos (122) Temperatura (119) Pedologia (116)Rochas (109) Família (105) Erosão (102) Natureza (102) Massa (101)Energia (100) Matéria (100) Partículas (97) Plano (95) MÃR (92)Cor (89) Ferro (89) Plataforma (89) Valor (85) Dados (84)Estado (83) Organismos (82) Tipos (82) Luz (78) Química (77)Volume (76) Animais (76) Botânica (76) Volume (76) Areia (75)Vigilância (75) Gestão (74) Vegetais (71) Ponto (70) Ondas (69)Regiões (69) Bases (68) Bases (68) Vida (67) Geologia (66)Atmosfera (63) Mente (63) Objeto (63) Produtos (63) Aves (62)Aves (62) Pressão (62) Corrente (61) Vegetação (61) Cristais (60)Rio de Janeiro (59) Rio de Janeiro (59) Serviços (59) Media (58)Currículos (58) Materiais (57) Bactérias (56) Oxigênio (55)Componentes (53) Anos (52) Evolução (52) Pesquisa (52)Nacional (51) Ál (51) Biologia (51) Insetos (51) Dias (50)Qualidade (50) Clima (49) Rede (49) Análise (48) Espaço (48)Falha (48) Registro (46) Carbono (46) Ciência (46) Células (46)Animal (45) Brasil (45) Alumínio (44) Movimento (44) Palinologia (43)Correntes (42) Praia (41)

2003	<p>Dados (139) Mineração de dados (53) Sistema (52) Rede (46) Gestão (44) Desenvolvimento (39) Nome (39) Data warehouse (35) Pesquisa (33) Apresentação (32) Conhecimento (32) Técnicas (32) Redes (30) Sistemas (30) Acompanhamento (29) Análise (28) Informação (28) Empresas (26) Instituição (26) Clientes (25) Estratégia (23) Energia (22) Data warehousing (21) Metodologia (21) Produtos (21) Qualidade (21) Visão (20) Area (19) Data (19) Área (19) Florianópolis (18) Florianópolis (18) Trabalho (16) Tempo (15) Ações (14) Ferramentas (14) Plataforma (14) Serviços (14) Santos (13) Dado (13) Planejamento (13) Roberto Carlos (13) Santos (13) Fax (12) Norton (12) Algoritmos (12) Padrões (12) Problemas (12) Avaliação (11) Comportamento (11) Fontes (11) Infraestrutura (11) Plano (10) Santa Catarina (10) Tomar (10) Ambiente (10) Direito (10) Estado (10) Mercado (10) Negócios (10) Santa Catarina (10) Tomar (10) Custos (9) Evolução (9) Gerenciamento (9) Organizações (9) Projetos (9) HTTP (8) Vi (8) Arquitetura (8) Construção (8) Energia elétrica (8) Ensino (8) Equipamentos (8) Funções (8) Gestão estratégica (8) HTTP (8) Identificação (8) Tabelas (8) Valor (8) Data mining (7) Futuro (7) Medida (7) Mineração (7) Natureza (7) Relações (7) Sistemas operacionais (7) Críticos (6) Físicas (6) Inovação (6) Instituições de ensino superior (6) Novas (6) Pessoas (6) Universidade Federal de Santa Catarina (6) Vetores (6) IEEE (5) Brasil (5) Cidade (5) Engenharia (5) Ensino superior (5)</p>
-------------	---

2002	<p>Dados (1791) Rede (1060) Sistema (1034) Análise (619)Equipamentos (447) MÄR (437) Santa Catarina (411) Santa Catarina (411)Sistemas (369) Area (366) Área (366) Energia elétrica (361)Serviços (359) Desenvolvimento (352) Mineração de dados (349)Falha (347) Ambiente (346) Data (335) Valor (322) Energia (316)Planejamento (305) Vegetação (296) Qualidade (290) Conhecimento (283)HTTP (273) HTTP (273) Simo (265) Animal (264) Art (264)Data warehouse (263) Redes (263) Técnicas (260) Gestão (253)Oracle (219) Oracle (219) Acompanhamento (194) Pesquisa (194)Clientes (193) Tempo (183) Consumo (182) Informação (180)Padrões (178) Apresentação (168) Mercado (160) Cálculo (156)Projetos (156) Componentes (152) Estudos (148) Índices (146)Genesis (145) Dias (135) Metodologia (134) Custos (129)Ferramentas (121) Firewall (119) Construção (118) Trabalho (118)Avaliação (115) Identificação (115) Io (115) Distância (114)Algoritmos (112) Comportamento (112) Empresas (111) Materiais (110)Dado (109) Running (108) Nome (108) Tabelas (108) Chá (106)Gerenciamento (106) Machines (105) Problemas (105) Components (103)Database (98) Medição (98) Meses (96) Data warehousing (94)Ações (94) Estado (92) Estrutura (90) Software (89) Software (89)HTML (87) PATH (87) HTML (87) Information (86) Media (85)Documentos (84) Produtos (83) Engenharia (81) Iluminação (81)BASIC (80) Engenharia de produção (78) Cabos (77) Comunicação (77)Florianópolis (76) Florianópolis (76) Evolução (75) Contratos (74)</p>
-------------	---

2001	<p>Dados (949) Sistema (789) Sistemas (601) Pesquisa (588) Rede (584)Desenvolvimento (513) Energia (494) Análise (425) Técnicas (413)Data (407) Metodologia (393) Redes (383) Power (365) Gestão (322)Systems (287) Famílias (287) Consumo (285) Equipamentos (273)Projetos (261) Area (252) Área (252) Qualidade (246)Energia elétrica (237) Curvas (227) Conhecimento (210)Mineração de dados (208) Serviços (203) Nome (201) Data warehouse (197)Padrões (195) Trabalho (193) Engenharia de produção (186) Tempo (186)Planejamento (184) Models (180) Time (177) Demand (175)Apresentação (174) Algoritmos (171) Nacional (169) Avaliação (161)Acompanhamento (160) IEEE (153) IEEE (153) Error (151)Clustering (150) Energy (150) Pesquisas (147) Plataforma (144)Martins (143) Analysis (141) Custos (141) Electricity (140)Estudos (136) Empresas (135) Series (131) Séries (131)Identificação (130) Training (127) Clientes (116) Day (114)Inteligência artificial (114) Tecnologia (113) Ál (112) Materiais (107)Valor (106) Comportamento (105) Redes neurais (105) Research (103)Rio de Janeiro (103) Construção (103) Linear (103) Rio de Janeiro (103)Brasil (102) Rules (98) Neural networks (96) Value (96) Cálculo (96)Florianópolis (94) Florianópolis (94) Electric power (90) Meses (89)Pesquisadores (89) Universidade Federal de Santa Catarina (88)Temperature (87) Mercado (86) Data warehousing (85) Techniques (83)Engenharia (83) Instituição (83) Management (82) Media (81) Anos (80)Sistemas operacionais (79) Arquitetura (78) Learning (77) Prediction (77)Variables (76) Paper (75) Software (74)</p>
-------------	--

APÊNDICE B – PERFIL COMPLETO DO ESPECIALISTA #2



2014	<p>Data (2003) Big Data (610) Big data (610) Example (542) API (509)Property (414) Hadoop (390) Value (330) Documents (316)Information (301) Time (299) Analysis (277) SQL (266) SQL (266)Patterns (247) Java (241) Database (237) Single (221) Apache (217)Memory (215) Tools (211) Architecture (193) HTTP (193) HTTP (193)Filter (167) NoSQL (148) Índices (147) Indexing (134) Arquitetura (134)Rules (133) Optimization (130) Size (129) Systems (129)Analytics (125) Text (122) Replicas (118) Running (118) Sources (116)Garbage (109) Garbage (109) Means (107) Setting (106) Works (101)Structure (97) Objects (96) Routing (92) Plataforma (91)Databases (90) Space (84) Behavior (81) Management (81) Tool (80)Tool (80) Security (78) Business (77) Policy (76) Power (68)Dados (68) Infrastructure (67) Form (65) Streaming (64) Streaming (64)Products (63) Contents (62) Driver (62) Columns (61) Books (60)Software (60) Solutions (60) Software (60) Filters (59) IBM (58)IBM (58) Administration (56) Thread (56) Language (54) Library (54)Phrase (54) Oracle (53) Oracle (53) Open Source (52) Open source (52)Serviços (52) Resources (51) Building (50) Examples (50) Map (50)Volume (50) Volume (50) Framework (49) Design (48) Family (48)JSON (48) Microsoft (48) Design (48) Microsoft (48) Components (47)Environment (47) World (47) Help (46)</p>
2010	<p>Data (8150) Information (3927) Ontology (2977) Example (1970)Knowledge (1904) Dimension (1786) Business (1759) Systems</p>

	<p>(1580)Text (1477) Database (1474) XML (1391) XML (1391) Time (1326)Documents (1148) OLAP (1113) Sources (1031) Analysis (976)Research (922) Architecture (899) Tables (895) SQL (881) SQL (881)Design (867) Design (867) Concepts (855) Language (787)Management (758) Structure (753) Environment (749) Sales (739)Web (728) Models (719) Modeling (700) Technology (686)Definition (676) Hierarchy (669) Framework (632) Value (609)Business intelligence (601) Business intelligence (601) Processes (593)Tools (578) Data warehouse (575) Single (573) Databases (570)Development (566) HTTP (563) HTTP (563) Technologies (527)Tool (492) Tool (492) Software (484) Software (484) Rules (472)Semantics (469) Paper (461) Knowledge management (451) Ål (445)Semantic Web (433) Semantic web (433) Topics (430) Techniques (428)Standards (421) Metadata (420) Components (415) Main (412)Products (410) Interoperability (400) Dados (380) Business process (371)Objects (366) Form (360) Records (342) Means (341) Area (323)Área (323) IBM (316) IBM (316) ACM (305) BASIC (305)Word (292) Future (287) Future (287) Cells (283) Size (283)Words (281) Names (276) Focus (276) Help (269) Categories (256)Reasoning (250) Best (247) Columns (245) Examples (244)Building (242) Property (241) Solutions (235) Handling (234)Reference (231) World (231)</p>
2008	<p>Data (110) Business (67) Ontology (66) Sources (42) Information (40)Rules (37) Concepts (36) Tools (36) Analysis (28) Knowledge (24)Framework (18) OLAP (18) Brazil (17) Florianópolis (16) Semantics (16)Florianópolis (16) Strategy (15) Business intelligence (14) Tool (14)Business intelligence (14) Tool (14) Dimension (13) Solutions (12)Filter (12) Database (11) Inference (11) Students (11) Documents (10)Tables (10) XML (10) Driver (10) Example (10) XML (10)Organizations (9) Time (9) Dados (9)Universidade Federal de Santa Catarina (9) Collections (8) Definitions (8)Drivers (8) UNIT (8) Components (6) Hierarchy (6) Main (6)Reasoning (6) Semantic Web (6) Semantic web (6) Terminology (6)Area (5) Cubes (5) Definition (5) Indexing (5) Person (5)Presentation (5) Projects (5) Systems (5) University (5) Web (5)Área (5) Algorithms (4) Architecture (4) Dice (4) Exploration (4)HTTP (4) Knowledge management (4) Map (4) MÄR (4) Nacional (4)Paper (4) Property (4) Research (4) Student (4) Technologies (4)Theme (4) Works (4) Arquitetura (4) Data warehouse (4) HTTP (4)Informação (4) Ontologia (4) Abstraction (3) Alumni (3) Education (3)Filters (3) Illustration (3) Jena (3) Language (3) Logic (3)Management (3) Maps (3) Metadata (3) Reference (3) Tests (3)Volume (3) Análise (3) Brasil (3) Jena (3) Metodologia (3)Projetos (3) Semântica (3)</p>
2006	<p>Dados (3765) XML (1484) XML (1484) Análise (1169) Tubarão (830)Conteúdo (767) Data (705) Fontes (664) Sistema (626)Informação (625) Conhecimento (589) Framework (574) Tempo (565)Ontologia (507) Pesquisa (478) Documentos (461) HTTP (451)HTTP (451) Web</p>

	<p>(450) Propriedade (444) Brasil (437) Driver (428) Ferramentas (427) Conceitos (425) Valor (399) Information (398) Drivers (382) Arquitetura (382) Semântica (381) Desenvolvimento (369) Java (359) HTML (302) HTML (302) Área (289) Área (289) Apresentação (286) Style (283) Texto (279) Data warehouse (271) Analysis (270) Ciências (270) Pesquisadores (269) Nome (262) Property (250) Web services (237) Example (237) Tipos (237) Web services (237) Sistemas (235) Março (234) Visão (224) Filtros (223) Serviços (220) Estrutura (211) Propriedades (206) Componentes (201) Business (187) UNIT (187) OLAP (184) Color (182) Language (181) Filter (181) Dimension (177) Lógica (177) Empresas (176) Projetos (175) Tabelas (171) Linguagem (170) Internet (168) Coleções (168) Internet (168) Plásticos (164) Tecnologia (164) Média (163) Engenharia (151) Business intelligence (149) Business intelligence (149) Pesquisas (143) Identificação (142) Value (141) Sistemas de informação (139) Dado (131) Exploração (130) Hierarchy (129) Tema (123) Organizações (123) UML (123) Construção (121) Dimensão (120) Software (119) Software (119) Systems (118) Ál (118) Novas (116) Objetos (114) Metadados (111) CIA (109) CIA (109) Medida (109) Indexação (106)</p>
2005	<p>Java (93) C++ (29) C++ (29) Objects (17) Value (16) Food (15) Compiler (13) Inheritance (11) Map (11) Reference (11) Arrays (10) Length (10) Example (10) Protected (9) Time (9) Primitive types (8) PATH (7) Error (6) Memory (6) Single (6) Size (6) Garbage (5) Information (5) Strategy (5) Unicode (5) Variables (5) Garbage (5) Unicode (5) Data types (4) Design (4) Design patterns (4) Examples (4) Graphics (4) Hierarchy (4) Root (4) Works (4) Design (4) Thread (4) Arguments (3) Arithmetic (3) Art (3) Chess (3) Composition (3) Containers (3) Data (3) Definition (3) Development (3) Drawing (3) Feet (3) HTTP (3) Libraries (3) Power (3) Sun (3) Wind (3) Cartoon (3) Game (3) HTTP (3) Web (3) Zero (3) Abstraction (2) Bath (2) Behavior (2) Compilers (2) Definitions (2) Environment (2) Help (2) Interfaces (2) LEN (2) Language (2) Learning (2) Library (2) Names (2) Pizza (2) Polymorphism (2) Pool (2) Resources (2) Syntax (2) Utility (2) BASIC (2) Pizza (2) Rock (2) Os (1) 1990s (1) Algorithms (1) Architecture (1) Associative arrays (1) Bar-Ilan University (1) Books (1) Brass (1) Building (1) CORBA (1) Cheating (1) Collections (1) Commerce (1) Contents (1) Corn (1) Counts (1) Database (1) Debugging (1) Duplicate (1)</p>

	(182)
2015	<p>Conhecimento (6538) Knowledge (4638) Protected (3083) Java (2950) Pesquisa (2940) Ál (2858) Dados (2737) Data (2549) Information (2485) Análise (1959) Trabalho (1895)Gestão do conhecimento (1884) Analysis (1753) Example (1565) Ontologia (1531) Documents (1452) Research (1321) Informação (1306) Sistema (1203) Area (1183) Área (1183) Conceitos (1103)Estudos (1098) Knowledge management (1097) Text (1095) HTTP (1057) HTTP (1057) Management (1036) Models (1026) Desenvolvimento (1024) Engenharia (1017) Inovação (1001) Value (1000)Organizações (968) Sistemas (929) Documentos (917) Framework (898) Systems (842) Ontology (805) Gestão (793) Time (791) Web (790) Literatura (774) Language (715) Artigos (685)Estrutura (660) Learning (653) Empresas (652) Evaluation (645) Técnicas (631) Structure (628) Pessoas (624) Valor (613) Pesquisas (604) Arquitetura (603) Avaliação (588) Tempo (585)Texto (580) Ferramentas (546) Modelos (544) Medida (542) XML (538) XML (538) Indexing (532) Software (527) Software (527) Literature (517) Media (506) Fontes (502) Erro (490)Development (489) Topics (485) Paper (475) Identificação (474) Ambiente (473) Serviços (471) Conteúdo (470) Produtos (465) Components (455) Business (451) Best (450) Tipos (449)Construção (448) Relações (448) Map (439) Form (437) Main (433) Processes (433) People (431) Sustentabilidade (428) Words (424) Tema (408) Design (399) Design (399) Privacy (394)ACM (390) Novas (382) Rules (380) Semântica (380) Technology (379)</p>
2014	<p>Tipos (13203) Plano (9779) Ensino (6412) Segurança (5441) Trabalho (4360) Máquinas (3638) Equipamentos (2821) Qualidade (2528) Materiais (2416) Mecânica (2315) Ambiente (2031)Sistemas (1985) Gestão (1977) Design patterns (1974) Sistema (1950) Ferramentas (1942) Técnicas (1929) Comunicação (1652) Conceitos (1604) Eletricidade (1545) Flags (1415) Tecnologia (1375)Rio de Janeiro (1348) Rio de Janeiro (1348) Apresentação (1322) Android (1263) Desenvolvimento (1193) Análise (1124) Saúde (1104) Produtos (1095) Redes (1047) Planejamento (1046)Eletrônica (1009) Normas (949) Documents (938) Projetos (929) Length (886) Soldagem (882) Dados (871) Conhecimento (849) Componentes (832) Programação (829) Rio Grande do Sul (823)Rio Grande do Sul (823) Serviços (808) Medição (777) Resíduos (776) Data (770) Contents (769) Aço (761) Media (759) Construção (752) Text (747) Books (736) Rede (736)Identificação (717) Informática (704) Princípios (688) Copyright (661) Gerenciamento (656) Rights (655) Reproduction (654) Higiene (640) Prevenção (626) Desenho (620) Leitura (618)Metrologia (614) Minas Gerais (594) Minas Gerais (594) Web (593) Pressão (591) Area (590) Área (590) Motores (587) Desenho técnico (580) Indexing (574) Industriais (567) Aprendizagem (548)Campinas (544) Campinas (544) Time (543) Corrente (541) ISO</p>

	(540) ISO (540) Eletrodo (539) Eléctrodo (539) Temperatura (530) Administração (527) Example (524) Padrões (519)Calçados (503) Information (499) Acidentes (498) Avaliação (494) Funções (494) Usinagem (492) Estrutura (491) Veículos (483) Pesquisa (477) Tempo (472)
2013	Knowledge (15499) B5 (13981) Ciências (12352) Medicina (9316) Research (6965) Conhecimento (6239) Area (5175) Área (5175) Avaliação (4567) Educação (4038) Information (4001)Saúde (3775) March (3688) Wednesday (3571) Pesquisa (3404) Design (3375) Design (3375) Psicologia (2964) Development (2763) Ciência (2743) Management (2714) Dados (2714)Science (2666) Biotecnologia (2665) Knowledge management (2591) Trabalho (2585) História (2562) Linguística (2542) Lingüística (2542) Direito (2487) Biodiversidade (2472) Technology (2390)Tools (2373) Data (2296) Sociologia (2284) Systems (2279) Learning (2268) Ciências agrárias (2253) Sistema (2212) People (2123) Química (2098) Administração (2089) Engineering (1921)SENAI (1911) Desenvolvimento (1842) Analysis (1816) Farmácia (1814) Ál (1807) Educação Física (1762) Educação física (1762) Física (1690) Odontologia (1669) Time (1666) Ciências sociais (1593)Example (1582) Organizations (1574) Tecnologia (1567) Geografia (1547) Técnicas (1504) Filosofia (1471) Business (1466) Gestão (1466) Medicina veterinária (1466) Ciência da computação (1426)Astronomia (1421) Fortaleza (1399) Fortaleza (1399) Texto (1396) Materiais (1375) Estudos (1358) Análise (1338) Economia (1328) Matemática (1312) Sistemas (1308) Enfermagem (1285)Processes (1274) Planejamento (1244) Solutions (1236) Structure (1229) Software (1221) Software (1221) Ensino (1216) Probabilidade (1215) Zootecnia (1175) Engenharia (1159) Framework (1158)Models (1146) Web (1127) Informação (1104) Alimentos (1099) HTTP (1098) HTTP (1098) Antropologia (1084) Ciência política (1072) Gestão do conhecimento (1067) Value (1054) Ambiente (1046)Arqueologia (1019) Communication (1015) Documents (1009)
2012	Conhecimento (6025) Sistema (3775) Inovação (2546) Dados (2394) Desenvolvimento (2116) Sistemas (2046) Pesquisa (2022) Rede (1816) Análise (1574) Informação (1566) Trabalho (1546)Ontologia (1431) Conceitos (1365) Serviços (1352) Area (1285) Área (1285) Knowledge (1265) Empresas (1260) Software (1223) Software (1223) Avaliação (1192) Ál (1057) Gestão (998)Tecnologia (993) Ambiente (992) Data (988) Gestão do conhecimento (977) Redes (971) Documentos (965) Treinamento (939) Aprendizagem (928) Pessoas (916) Engenharia (901) Valor (867)Objetos (859) Estrutura (858) Tempo (839) Brasil (794) Técnicas (776) Visão (776) Java (761) Projetos (760) Organizações (753) Mundo (738) Mercado (720) HTTP (719) HTTP (719)Estado (715) Governo (707) Nacional (690) Construção (653) Identificação (653) Metodologia (653) Ferramentas (643) Modelos (642) Produtos (637) Instituição

	(624) Tipos (616) Anos (610)Componentes (593) Ações (582) Relações (581) Comunicação (578) Conteúdo (569) Information (565) Nome (555) Ciência (547) Plataforma (530) Texto (521) Orientação (518) Objeto (512)Educação (498) Qualidade (497) Novas (495) Linguagem (493) Padrões (493) Estudos (465) Text (460) Instituições (450) Palavras (445) Espaço (438) Sociedade (427) Vida (417)Planejamento (406) Apresentação (404) Problemas (403) Referência (403) Web (401) Florianópolis (396) Florianópolis (396) Energia (386) Papel (386) Fontes (383) Plano (373) Equipamentos (370)Natureza (362) Value (357) Pesquisas (350) Internet (348) Internet (348)
2011	Research (1303) Conhecimento (1041) Area (610) Área (610) Innovation (608) Propostas (570) People (514) Knowledge (510) Sistema (445) Organizations (403) Data (391) Equipamentos (390)Information (341) Management (341) Example (340) BASIC (327) Estado (325) Development (321) Library (319) Technology (318) Gestão (314) Inovação (313) Pesquisa (311)Gestão do conhecimento (300) Serviços (296) Desenvolvimento (281) Informação (254) Avaliação (253) Sinalização (243) Conflict (242) Análise (238) Time (237) Mídia (236) Mídia (236)Science (235) Nacional (218) Scientists (218) Organizações (211) Sistemas (211) Strategy (201) Person (196) Princípios (196) Projetos (188) Individuals (184) Design (176) Comunicação (176)Cálculo (176) Design (176) Communication (174) Tempo (171) Technology transfer (170) Leadership (169) University (168) Nome (168) Behavior (164) Pessoas com deficiência (164) Estrutura (161)Investment (160) Turismo (159) Eventos (158) Tecnologia (155) Dados (151) Sociedade (150) Projects (149) Ál (147) Industry (146) Empresas (146) Titulação (143) Environment (141)Engenharia (141) Março (141) Ações (140) Resources (138) Julho (138) Estacionamento (135) Instituição (134) Engineers (132) Engineering (130) Atores (130) Government (129) Lazer (129)Novembro (128) Maio (127) Telefone (127) Researchers (126) Fevereiro (125) Setembro (124) Structure (123) Junho (123) HTTP (121) Dezembro (121) Governo (121) HTTP (121)Outubro (121) Abril (120) Janeiro (120) Business (119) Capital (119) Systems (119) Training (119)
2010	Conhecimento (270) Organizações (69) Sistemas (66) Mídia (57) Mídia (57) Knowledge (51) Informação (48) Humanos (33) Gestão do conhecimento (30) Engenharia (27) Tecnologia (27)Sistema (21) Conceitos (21) Economia (21) Comunicação (18) Globalização (18) População (18) Serviços (18) Valor (18) Information (15) Media (15) Empresas (15) Capital (13) Capital (13)Bens (12) Bibliotecas (12) Desenvolvimento (12) Países (12) Pessoas (12) Trabalho (12) Development (9) Dados (9) Diferença (9) Metodologias (9) Qualidade (9) Riqueza (9) Texto (9)Identificação (7) Action (6) Barreiras (6) Communication (6) Digital media (6) Languages (6) Sources (6) Tics (6) Transportation (6) Ambiente (6) Aprendizagem

	<p>(6) Atividades econômicas (6) Barreiras (6) Ciências (6) Competição (6) Construção (6) Cultura (6) Físicos (6) Lógica (6) Matéria (6) Mente (6) Natureza (6) Nomes (6) Novas (6) Ontologia (6) Ordenação (6) Papel (6) Projetos (6) Publicações (6) Redes (6) Sociedade (6) Softwares (6) Tecnologia da informação (6) Títulos (6) Vida (6) Física (4) Algorithms (3) Area (3) Bases (3) Boas (3) Building (3) Components (3) Concepts (3) Expert systems (3) Future (3) Gera (3) Goal (3) Humans (3) Knowledge engineering (3) Knowledge markets (3) Learning (3) Libraries (3) Models (3) Organizations (3) Past (3) Procedural knowledge (3) Processes (3) Renting (3) Research (3) Software (3) Switzerland (3) TV (3) Technology (3)</p>
2009	<p>Dados (15) Projetos (12) Tecnologia (10) Trabalho (9) XML (8) XML (8) Valor (7) Estado (6) Information (5) Data (4) HTTP (4) Internet (4) Knowledge (4) Knowledge management (4) Sistema (4) Systems (4) Arquitetura (4) Dado (4) HTTP (4) Internet (4) Sociedade (4) Agriculture (3) Brasília (3) Energy (3) Florianópolis (3) Nacional (3) Sources (3) Web services (3) Brasil (3) Brasília (3) Comunicação (3) Diferença (3) Estratégia (3) Florianópolis (3) Gestão (3) Reunião (3) Web (3) Web services (3) Area (2) Capital (2) Germany (2) Industry (2) Processes (2) Solutions (2) Acompanhamento (2) Apresentação (2) Capital (2) Desenvolvimento (2) Example (2) Exercício (2) Gestores (2) Motivação (2) Padrões (2) Pessoas (2) Secretaria da Receita Federal (2) Sistemas informatizados (2) Área (2) Aid (1) Alfred North Whitehead (1) Analysis (1) Best (1) Business (1) Concrete (1) Conservation (1) Countries (1) Development (1) Education (1) Email (1) Engineering (1) Environment (1) Farmers (1) Forestry (1) Fur (1) GIS (1) GPS (1) Geo (1) Gera (1) German (1) Goal (1) HTML (1) Information systems (1) Infrastructure (1) Innovation (1) Interfaces (1) Interpretation (1) Kaiserslautern (1) Knowledge sharing (1) Leaders (1) Mario (1) Media (1) Months (1) Nature (1) Nutrition (1) October (1) Partnerships (1) Pliers (1) Pool (1) Programs (1) Public sector (1) Science (1)</p>
2008	<p>Paths (48) PATH (46) Paper (22) Names (20) Design (19) Information (19) Design (19) Analysis (15) Modeling (13) ER (12) Triad (11) Authors (10) BASIC (9) Single (7) Categorization (6) Quality (6) Accept (6) Standards (5) Database (4) Form (4) IEEE (4) Research (4) Semantics (4) Structure (4) Example (4) IEEE (4) Concepts (3) Data (3) Language (3) Lead (3) Main (3) Models (3) Observation (3) Presentation (3) Rating (3) University (3) Area (2) Brazil (2) Business (2) Examples (2) Identification (2) Internet (2) Migration (2) Nature (2) Relational model (2) States (2) Technology (2) Trials (2) ACM (2) Internet (2) Área (2) Action (1) Architecture (1) Banks (1) BoA (1) Building (1) Capital (1) Communications (1) Database management systems (1) Databases (1) December (1) Definition (1) Departments (1) Dorset (1) Duplicate (1) Environment (1) Florianópolis (1) Foundations (1) Help (1) Information systems (1) Knowledge (1) Lent (1) Medina (1) National Institute of Standards</p>

	and Technology (1) Objects (1) Paris (1) Perception (1) Publishing (1) Quotas (1) Rules (1)Sound (1) Student (1) Syntax (1) Systems (1) Time (1) USA (1) Value (1) Works (1) Abusos (1) Bancos de dados (1) BoA (1) Brasil (1) Cabeção (1) Capital (1) Dorset (1)Exercício (1) Florianópolis (1) Framework (1) Física (1) House (1)
2007	Sistema (483) Dados (428) Medição (242) Turismo (224) Multimedia (218) Qualidade (183) Modelos (167) Conhecimento (164) Software (154) Software (154) Trabalho (144) Area (142)Video (142) Vídeo (142) Área (142) Medida (136) Ferramentas (131) Semantic Web (129) Semantic web (129) Paper (127) Nacional (122) Data (121) HTTP (121) HTTP (121) Conteúdo (117)Desenvolvimento (111) Web (110) Ontology (106) Produtos (106) Serviços (105) Information (103) Nome (103) Ál (100) Análise (98) Example (97) Media (91) Ontologia (91) XML (83)Linguagem (83) XML (83) Knowledge (82) Informação (78) Sistemas (77) Metadata (76) Tool (73) Conceitos (73) Tool (73) Analysis (69) Concepts (69) Multimídia (67) Instituição (66)Registro (65) Internet (64) Internet (64) Projetos (63) Gestão (60) Organizações (59) Language (58) Equipamentos (57) Comunicação (56) Syntax (53) Research (52) IEEE (51) IEEE (51)Tools (50) Web semântica (49) Ambiente (48) Time (47) Author (46) Systems (46) Dado (46) Documentos (45) Metadados (45) Problemas (45) Páginas (45) Estrutura (44) Database (43)Images (43) Pessoas (43) Tipos (43) Framework (42) Padrões (42) Técnicas (42) Image (41) Gerenciamento (41) Relações (41) Lógica (40) Semântica (40) Tempo (40) Structure (38)Arquivos (37) Engenharia de software (37) Pesquisa (37) Objetos (36) Construção (35) Semantics (34) Avaliação (34) Componentes (34) Texto (34) Recursos (33)
2006	Knowledge (485) Informação (302) Conhecimento (270) Pesquisa (260) XML (242) XML (242) Trabalho (235) Desenvolvimento (212) Tools (203) Information (190) Knowledge management (185)Software (185) Software (185) Metrics (154) Governo (135) HTTP (134) HTTP (134) Plataforma (127) Metodologia (118) Unidades de informação (108) Objeto (106) Objetos (105) Construção (102)Area (96) Área (96) Dados (95) Sistemas (94) Identificação (92) Management (91) Value (88) Arquitetura (84) Web (80) Data (77) Linguagem (77) Comunidade (73) Tool (71) Crianças (71)Nome (71) Tool (71) Comunicação (69) Pensamento (69) Análise (68) Example (67) Gestão do conhecimento (67) Texto (66) Research (59) Aprendizagem (57) Informática (57) Tecnologia (57)UML (57) Educação (56) Sistema (55) People (54) Ál (54) Softwares (54) Systems (51) Gestão (51) Pessoas (51) Ontology (50) Fontes (50) Business (49) Technology (48) Time (48)Conceitos (48) Qualidade (48) Referência (48) Documents (47) Measurement (47) Artigos (47) Ciência (47) Anos (46) Conteúdo (45) Modelos (44) Ações (42) Ação (40) Deficiência (40)Engenharia (40) Apresentação (39) Internet (38) Projects (38) Internet (38) Jogos (38) Projetos (38) Tipos (37) Organizations (36) Ambiente (36) Help (33) Processes

	(33) Form (32)Examples (31) Atenção (31) Sources (30) Novas (30) Relações (30) Serviços (30) Learning (29) Educação especial (29) Mundo (29) Professores (29) Tempo (29)
2005	Software (17) Software (17) Erro (12) Qualidade (12) Java (10) Desenvolvimento (6) Produtos (5) Tempo (5) Sistema (4) Framework (4) Linguagem (4) Programação orientada a aspectos (4)Software quality (3) Tool (3) Compilação (3) Nome (3) Ponto (3) Tecnologia (3) Tool (3) Trabalho (3) Aspect-oriented programming (2) Help (2) Information (2) Language (2) Maintenance (2)Metrics (2) XML (2) Apresentação (2) Comportamento (2) Construções (2) Cálculo (2) Desenvolvimento de software (2) Mecanismos (2) Novas (2) Programação (2) Referência (2) Tratados (2)XML (2) Action (1) Area (1) Barra (1) Boston (1) Data (1) Development (1) Education (1) Engineering (1) Error (1) Gera (1) Greenwich (1) Hand (1) Indianapolis (1) Joining (1)Logging (1) Models (1) Paper (1) Pizza (1) Products (1) Quality (1) Quantity (1) Rio de Janeiro (1) Source code (1) Time (1) Vi (1) Ambiente (1) Ambientes (1) Arquivos (1)Avaliação (1) Boston (1) Complexidade (1) Conceitos (1) Dado (1) Dados (1) Diários (1) Escrita (1) Estrutura (1) Futuro (1) Gera (1) Gravação (1) Guarda (1) Indianápolis (1)Meses (1) Objetos (1) Orientação (1) Pesquisa (1) Pizza (1) Programação orientada a objetos (1) Projetos (1) Qualidade de software (1) Rio de Janeiro (1) Softwares (1) Tipos (1) Técnicas (1)Área (1)

APÊNDICE D – PERFIL EXTRAÍDO VS LINKEDIN ESPECIALISTA #1

<i>Original Text One</i>	<i>Original Text Two</i>
business intelligence project management knowledge management management strategic planning business development product development strategy team leadership business analysis innovation management, entrepreneurship business strategy e- government scrum software development sql software project product management project planning business process ms project java itil requirements analysis semantic technologies team management uml databases business planning data analysis governance systems analysis software engineering portuguese erp management consulting agile project management pmbok pmp agile methodologies knowledge engineering semantics semantic web artificial intelligence pmi data mining	conhecimento dados plataforma sistema desenvolvimento gestão pesquisa serviços informação sistemas inovação trabalho projetos análise data área valor empresas tecnologia knowledge nacional fontes avaliação plano ambiente produtos senai nome http qualidade arquitetura componentes educação capital software software engenharia gestão do conhecimento governo apresentação técnicas information planejamento ações identificação documentos rede pessoas brasil value instituição tempo estrutura ontologia visão av estratégia registro ação comunicação segurança dias ferramentas fortaleza gerenciamento florianópolis organizações web equipamentos conteúdo ãl instituições metodologia anos construção objeto estado ciência mercado atores art oracle example evolução redes saúde conceitos clientes santos management sistemas de informação novas referência

Similarity: 0.6569 (cosine similarity from 0 to 1, 0 means absolute different, 1 means absolute same)

Fonte: <http://documentsimilarity.com/document-similarity-demo>

APÊNDICE E – PERFIL EXTRAÍDO VS LINKEDIN ESPECIALISTA #2

<i>Original Text One</i>	<i>Original Text Two</i>
business intelligence java scrum agile methodologies knowledge engineering olap big data software engineering solr postgresql lucene elasticsearch linked data eclipse junit erp databases jboss application server sql tomcat jpa rup hibernate universal information certified scrum master spring software development	data information dados ontology xml example business knowledge dimension systems database text time documents analysis olap análise sources framework web http sql architecture value tables language research concepts design tools property structure management data warehouse single java tubarão hierarchy conteúdo business intelligence business intelligence environment models definition informação sales technology fontes modeling sistema rules conhecimento development databases processes software software api area tempo tool big data tool ontologia technologies ãl arquitetura metadata filter paper semantics driver objects main pesquisa knowledge management components conceitos topics ferramentas products documentos standards means size techniques propriedade web services valor semantic web brasil form

Similarity: 0.7796 (*cosine similarirty from 0 to 1, 0 means absolute different, 1 means absolute same*)

Fonte: <http://documentsimilarity.com/document-similarity-demo>

APÊNDICE F – PERFIL EXTRAÍDO VS LINKEDIN ESPECIALISTA #3

<i>Original Text One</i>	<i>Original Text Two</i>
<p>scrum java uml postgresql javascript java enterprise edition agile project management xml spring rup html jsp cmmi jpa gestão ágil de projetos gerência de projeto ms project arquitetura de software spring framework engenharia do conhecimento e- governance mps.br aspectj oracle sql microsoft project análise de requisitos, ireport, jasper reports, enterprise architecture, bpmn, análise de processos, xslt</p>	<p>knowledge conhecimento tipos b ciências trabalho plano research sistema dados pesquisa medicina area área information data ensino sistemas avaliação desenvolvimento ál análise segurança gestão analysis saúde informação conceitos knowledge management técnicas management educação gestão do conhecimento example ambiente inovação design design qualidade tecnologia ferramentas software software systems equipamentos java http http ontologia máquinas development engenharia materiais ciência time documents people march comunicação estudos serviços tools science web learning technology wednesday projetos rede documentos organizations value text texto protected pessoas framework psicologia estrutura administração organizações tempo planejamento empresas história produtos construção identificação direito models valor biotecnologia business redes linguística linguística química rio de janeiro rio de janeiro structure</p>

Similarity: 0.6858 (cosine similarity from 0 to 1, 0 means absolute different, 1 means absolute same)

Fonte: <http://documentsimilarity.com/document-similarity-demo>