**Rudger Nowasky do Nascimento**

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

Dissertação submetido(a) 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. Dr. Denilson Sell

Coorientador: Prof. Dr. Paulo M. Selig

Florianópolis

2015

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. Em busca de alternativas que possam simplificar esse processo de atualização dos dados, este trabalho propõe um modelo para a extração de perfis de especialistas a partir de documentos presentes na organização ou que possam ser buscados na internet. Para a construção desse modelo, realizou-se uma revisão integrativa da literatura que resultou na identificação das abordagens atuais para extração de perfil de especialista. A partir dessas abordagens, elaborou-se um modelo composto por um método conceitual e um protótipo tecnológico para a tarefa de extração de informações de perfil de especialistas com o objetivo de apoiar as ferramentas de *Expertise Location*. Observa-se que a utilização dos documentos não estruturados de uma organização pode ajudar na extração de perfis de especialistas, possibilitando a utilização de menos tempo para atualização de seus dados em ferramentas de *Expertise Location*.

Palavras-chave: *Expertise Location*. Gestão do Conhecimento. Engenharia do Conhecimento. Recuperação de Expertise. Extração de perfis de especialistas.

LISTA DE FIGURAS

[Figura 1 - Tipos de revisão da literatura. 15](#_Toc442219737)

[Figura 2 - Etapas da revisão integrativa. 20](#_Toc442219738)

[Figura 3 – Tipos de repositórios de conhecimento. 26](#_Toc442219739)

[Figura 4 – Conhecimento tácito: característica e forma de armazenamento. 27](#_Toc442219740)

[Figura 5 – Conhecimento explícito: característica e forma de armazenamento. 28](#_Toc442219741)

[Figura 6 – Pirâmide do conhecimento (dado, informação e conhecimento). 30](#_Toc442219742)

[Figura 7 – Escada do conhecimento (dado, informação e conhecimento). 30](#_Toc442219743)

[Figura 8 – Interdisciplinaridade da Gestão do Conhecimento. 32](#_Toc442219744)

[Figura 9 – Resumo das etapas e dos resultados da revisão integrativa da literatura. 42](#_Toc442219745)

[Figura 10 – Arquitetura lógica do modelo de extração de entidades e geração de ontologia de Ceci. 53](#_Toc442219746)

[Figura 11 – Modelo de mineração de textos voltados a aplicações de EC e GC de Gonçalves. 54](#_Toc442219747)

[Figura 12 – Modelo de *Temporal Knowledge Discovery in Texts* de Bolvo. 55](#_Toc442219748)

[Figura 13 – Procedimentos metodológicos para desenvolvimento da pesquisa. 57](#_Toc442219749)

[Figura 14 – *Tag cloud* sobre as palavras-chaves e resumos dos estudos da revisão integrativa. 66](#_Toc442219750)

[Figura 15 – Visão conceitual do modelo para extração de perfil de especialista. 72](#_Toc442219751)

LISTA DE QUADROS

[Quadro 1 - Diferenças entre revisão sistemática e revisão narrativa 16](#_Toc442219752)

[Quadro 2 - Comparação dos tipos de revisão de literatura 18](#_Toc442219753)

[Quadro 3 – Detalhamento das etapas da revisão integrativa 21](#_Toc442219754)

[Quadro 4 – Definições de conhecimento. 24](#_Toc442219755)

[Quadro 5 – Métodos e ferramentas de apoio aos processos da GC. 34](#_Toc442219756)

[Quadro 6 – Termos correlatos das palavras-chaves da revisão integrativa da literatura. 59](#_Toc442219757)

[Quadro 7 – Total de estudos encontrados na base Scopus a partir das palavras-chaves de pesquisa. 61](#_Toc442219758)

[Quadro 8 – Estudos encontrados na base Scopus a partir da revisão integrativa. 62](#_Toc442219759)

[Quadro 9 – Estudos selecionados na revisão integrativa. 67](#_Toc442219760)

[Quadro 10 – Matriz de síntese da revisão integrativa com as categorias dos estudos. 69](#_Toc442219761)

LISTA DE ABREVIATURAS E SIGLAS

APO – Asian Productivity Organization

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 Baseados em Conhecimento

SGC – Sistemas de Gestão do Conhecimento

UFSC - Universidade Federal de Santa Catarina

SUMÁRIO

1 INTRODUÇÃO 8

1.1 CONTEXTUALIZAÇÃO E PROBLEMA DE PESQUISA 8

1.2 OBJETIVOS 10

1.2.1 Objetivo geral 10

1.2.2 Objetivos específicos 10

1.3 DELIMITAÇÃO DO ESCOPO DA PESQUISA 11

1.4 INTERDISCIPLINARIDADE E ADERÊNCIA AO OBJETO DE PESQUISA DO PROGRAMA 11

1.5 ESTRUTURA DO TRABALHO 12

2 FUNDAMENTAÇÃO TEÓRICA 14

2.1 REVISÃO DA LITERATURA 14

2.1.1 Revisão integrativa da literatura 19

2.2 CONTEXTUALIZAÇÃO 23

2.2.1 Conhecimento 23

2.2.1.1 Conhecimento tácito 26

2.2.1.2 Conhecimento explícito 27

2.2.1.3 Dados, informação, conhecimento 28

2.2.2 Gestão do Conhecimento 31

2.2.3 Engenharia do Conhecimento 35

*2.2.4* *Expertise Location* 36

2.2.5 Discussão e conclusão da contextualização 38

2.2.5.1 Uso dos termos Expertise e Expertise Location. 39

2.2.5.2 Expertise Location e a área de Recuperação de Informação 40

2.3 ESTADO DA ARTE 41

2.3.1 *Expertise Location* no contexto da GC 42

2.3.2 Métodos conceituais e tecnológicos para a extração de perfis de especialistas 47

2.3.3 Discussão e conclusão do estado da arte 51

2.3.3.1 Modelos de reconhecimento de entidades, correlação e análise temporal 52

3 PROCEDIMENTOS METODOLÓGICOS 56

3.1 Execução do método de revisão integrativa 57

3.1.1 Identificação do tema e seleção da questão de pesquisa 57

3.1.2 Definição dos critérios de inclusão e exclusão de estudos 61

3.1.3 Identificação dos estudos pré-selecionados e selecionados 66

3.1.4 Categorização dos estudos selecionados 68

4 APRESENTAÇÃO DOS RESULTADOS 70

4.1 Modelo de extração de perfil de especialista proposto 70

4.1.1 Extração do conteúdo 74

4.1.2 Reconhecimento de entidades 74

4.1.3 Extração de entidades de conceitos e datas 74

4.1.4 Associação dos conceitos e datas para geração do perfil 74

4.1.5 Visualização do perfil do especialista 74

4.2 Protótipo baseamento no modelo proposto 74

4.3 Demonstração de viabilidade 74

5 CONCLUSÕES E TRABALHOS FUTUROS 75

5.1 CONSIDERAÇÕES FINAIS 75

5.2 RECOMENDAÇÕES PARA TRABALHOS FUTUROS 75

6 REFERÊNCIAS 76

# INTRODUÇÃO

Este capítulo tem por objetivo apresentar o propósito deste trabalho e foi estruturado em cinco seções. A primeira seção contextualiza o leitor no que tange ao problema de pesquisa abordado. A segunda seção apresenta os objetos – 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 desenvolvem este trabalho de acordo com o problema identificado, os objetivos e o seu escopo.

## 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 consideram o conhecimento 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.

Dada a importância do conhecimento e a 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 (WEGGEMAN, 1997; WIIG, 1997; BASSI, 1999 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, 2005).

Segundo a APO - Asian Productivity Organization (2010), *os Expertise Location Systems* (ELS) – ou Sistemas para Localização de Especialistas - tem sido uma das principais ferramentas para a Gestão do Conhecimento, pois possuem o potencial de auxiliar as etapas de (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 Conhecimentos, demonstrando a sua importância 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 ELSs abordam problemas típicos que grandes organizações enfrentaram: muitas vezes não é possível encontrar o conhecimento adequado que existe em algum lugar na organização ou na cabeça de alguém. Esse problema pode induzir a organização a “reinventar a roda” para resolver um mesmo problema já resolvido anteriormente, só porque não se sabe quem fez o quê ou quem possui determinado conhecimento.

A EL envolve ferramentas que permitem compartilhar o conhecimento existente, conectando pessoas que precisam de determinados conhecimentos às pessoas que possuem esses conhecimentos (APO, 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 sistemas mais sofisticados para buscar automaticamente *expertise* que apoiam o processo de encontrar e de conectar pessoas que necessitam de um conhecimento a pessoas que detêm esse conhecimento (BALOG et al., 2012; APO, 2010; MAYBURY, 2006).

Embora seja uma ferramenta importante no papel da GC, muitos sistemas de EL ainda dependem do preenchimento das informações dos colaboradores de forma manual, e vários estudos (SOHAIL; AFZAL; AHMAD, 2013, BALOG et al., 2012; BALOG; RIJKE, 2007; YAO; TANG; LI, 2007, BECERRA-FERNANDEZ, 2006; MARWICK, 2001) 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. Assim, os sistemas de EL possibilitariam uma menor dependência do preenchimento de informações de cada colaborador e permitiriam que os dados dos perfis fossem mais facilmente atualizados. Esses estudos também apontam que problema com uso de perfis explícitos, isto é, preenchidos pelos próprios colaboradores, é a atualização constante das informações. Desse modo, é preferível reunir informações automaticamente, se possível a partir de fontes existentes.

De modo a corroborar com essa visão, os estudos realizados pela IDC - International Data Corporation (2010) descrevem que as fontes de informações não estruturadas representarão aproximadamente 90% de todo o conteúdo gerado nesta década (2010 a 2020). Por meio de técnicas da Engenheira do Conhecimento, essas informações, quando trabalhadas com a metodologia e ferramentas adequadas, suscitam em conhecimento e em ativos do conhecimento conforme exemplos citados por Gonçalves (2006): redes de relacionamento, competências e interesses de colaboradores. Esses ativos do conhecimento podem, desse modo, auxiliar no processo de criação e atualização dos perfis de especialistas de uma organização.

Portanto, com base nos referenciais expostos, esta dissertação visa abordar a seguinte pergunta de pesquisa: como extrair perfis de especialistas a partir de documentos para apoiar a Gestão de Conhecimento na utilização de ferramentas de *Expertise Location*?

## OBJETIVOS

### Objetivo geral

Propor um modelo para a extração do perfil de especialista a partir de fontes de dados não estruturadas para apoiar a Gestão do Conhecimento na aplicação de ferramentas de *Expertise Location* em organizações.

### Objetivos específicos

* Identificar as abordagens atuais que possibilitem a extração de perfis de especialistas com base em fontes de dados não estruturadas no contexto de *Expertise Location*.
* Definir um modelo, de acordo com as abordagens existentes, para a extração de perfis de especialistas com base em fontes de dados não estruturadas.
* Desenvolver um protótipo funcional para verificar a viabilidade do modelo proposto utilizando informações não estruturadas (e.g., relatórios técnicos) para extração do perfil de especialista.

## DELIMITAÇÃO DO ESCOPO DA PESQUISA

Conforme apresentado anteriormente, o principal resultado deste trabalho é a proposição de um modelo para a extração 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.

Nesse trabalho, são identificados métodos técnicos e conceituais que permitem essa extração de perfis de especialistas. Entretanto, esta pesquisa não busca preencher lacunas existentes na literatura, mas, sim, utilizar e combinar métodos existentes para propor um modelo que apresente uma possibilidade de utilização de documentos não estruturados para a extração de perfil de especialista.

Para fins de verificação da viabilidade do modelo proposto, são utilizados documentos de especialistas no contexto de currículos e atuações profissionais. O protótipo proposto focou na exploração de dados relacionados à atuação profissional sobre documentos produzidos em atividades de consultoria e seus resultados são específicos para esse domínio.

## 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 “incluem 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 técnicas ligadas à EC para extração de perfis de especialistas de modo a explicitar seus conhecimentos a partir de fontes de dados não estruturadas (e.g. documentos em linguagem natural).

Há, também, uma contribuição para a área de Gestão do Conhecimento pelo fato desta dissertação buscar o aprimoramento de ferramentas *Expertise Location* que são amplamente utilizadas em iniciativas de GC e abordar o conhecimento “como produto, processo e resultado de interações sociais e tecnológicas entre agentes humanos e tecnológicos” (EGC, 2015b). Para a materialização dessa contribuição, utiliza-se construções teóricas interdisciplinares relacionadas principalmente 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, Alessandro Botelho. Um modelo de descoberta de conhecimento inerente à evolução temporal dos relacionamentos entre elementos textuais. Tese de doutorado em Engenharia e Gestão do Conhecimento, 2011.
* CECI, Flávio. Um Modelo semi-automático para a construção e manutenção de ontologias a partir de bases de documentos não estruturados. Dissertação de mestrado em Engenharia de Gestão do Conhecimento, 2010.
* BEPPLER, Fabiano Duarte. Um modelo para recuperação e busca de informação  
  baseado em ontologia e no círculo hermenêutico, 2008. Tese de doutorado em Engenharia e Gestão do Conhecimento, 2008.
* SELL, Denilson. Uma arquitetura para *business intelligence* baseada em tecnologias semânticas para suporte a aplicações analíticas. Tese de doutorado em Engenharia de Produção[[1]](#footnote-1), 2006.
* GONÇALVES, Alexandre Leopoldo. Um modelo de descoberta de conhecimento baseado na correlação de elementos textuais e expansão vetorial aplicado à engenharia e gestão do conhecimento. Tese de doutorado em Engenharia de Produção[[2]](#footnote-2), 2006.

## ESTRUTURA DO TRABALHO

Este trabalho é composto por cinco capítulos, sendo os demais descritos a seguir.

* Capítulo 2. Fundamentação teórica: neste capítulo se apresenta 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 contextualização do leitor e fundamentação dos conceitos utilizados neste trabalho. Apresenta-se também, o conteúdo resultante da revisão sistemática de literatura que gerar insumos para a visão do estado da arte das temáticas abordadas neste trabalho.
* Capítulo 3. Métodos de pesquisa: este capítulo tem por objetivo apresentar os procedimentos realizados para materializar os objetivos desata pesquisa. Assim, contextualiza o leitor no que tange à revisão da literatura e seus resultados, bem como detalha os procedimentos realizados para construção do modelo proposto.
* Capítulo 4. Resultados: apresenta-se o modelo proposto com base da revisão integrativa da literatura e o protótipo baseado nesse modelo, demonstrando os resultados alcançados.
* Capítulo 5. Conclusões e Trabalhos Futuros: neste capítulo são destacadas as conclusões sobre o trabalho realizado bem como são delineados os trabalhos futuros.

# FUNDAMENTAÇÃO TEÓRICA

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

Na primeira seção, são apresentados, brevemente, os conceitos sobre os tipos de revisão de leitura realizados nesta pesquisa. Posteriormente, apresenta-se a contextualização 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. Por fim, apresenta-se o conteúdo do resultado da combinação desses conceitos na revisão sistemática integrativa da literatura.

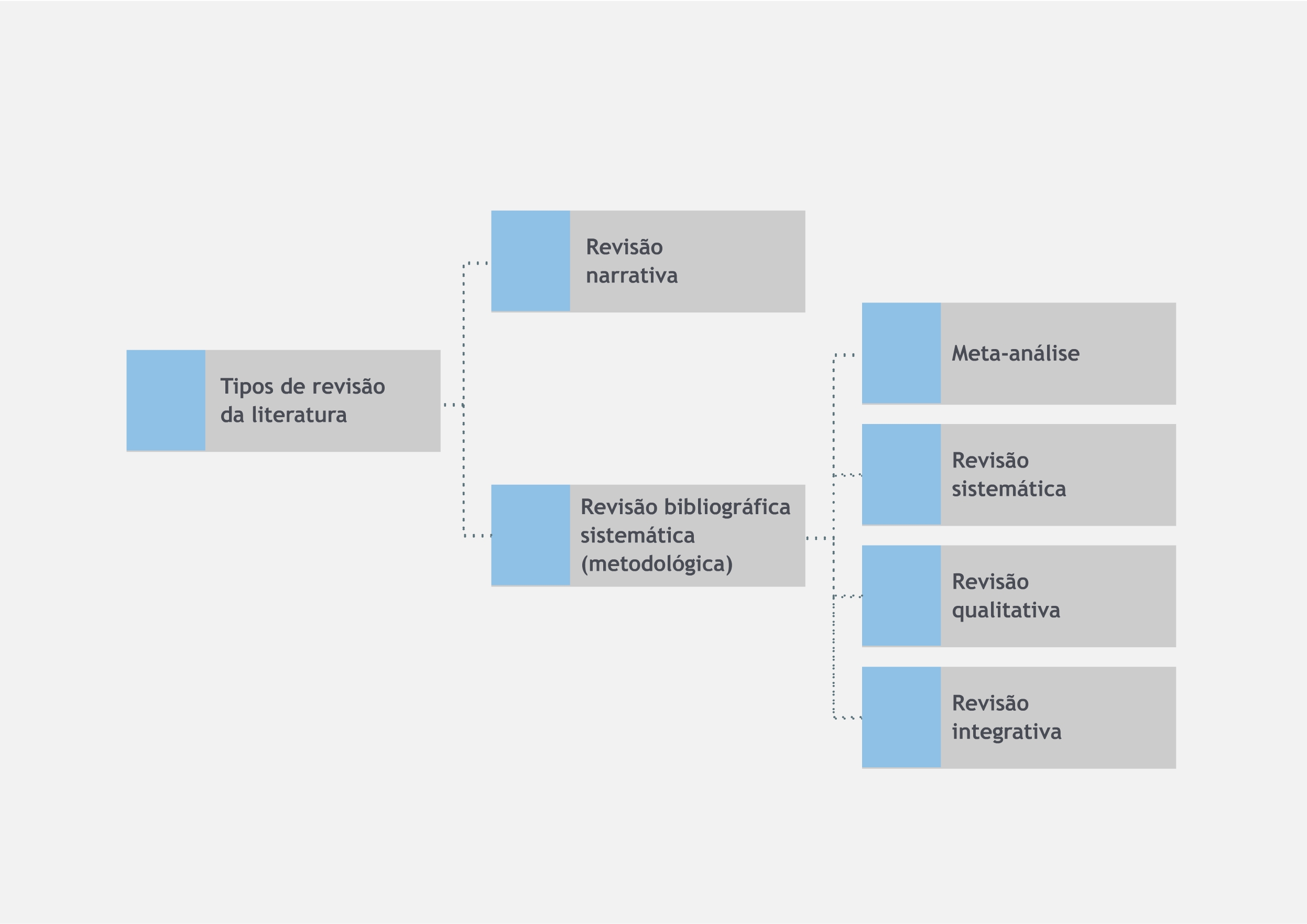
## REVISÃO DA LITERATURA

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 cientificas, 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 atinjam os objetivos específicos 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 utilizados por acadêmicos e pesquisadores para descrever o estado da arte de um tema (BOTELHO; CUNHA; MACEDO, 2011). Essas revisões são classificadas em revisões narrativas e revisões sistemáticas, respectivamente. As revisões sistemáticas, por sua vez, são subclassificadas em meta-análise, revisão sistemática, revisão qualitativa e revisão integrativa, e são diferenciadas essencialmente pelo seu rigor metodológico.

A Figura 1 apresenta a visão hierárquica dos tipos de revisão de literatura classificados por Botelho, Cunha e Macedo (2011).

Figura - Tipos de revisão da literatura.



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

A revisão narrativa, também chamada de revisão tradicional, não utiliza critérios explícitos e sistemáticos para a busca e análise da literatura quando se comparada à revisão sistemática (CORDEIRO et al., 2012). 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, utiliza-se métodos explícitos e sistemáticos para identificar, selecionar e avaliar criticamente os estudos prévios realizados conduzindo a uma síntese dos resultados de múltiplos estudos primários (CORDEIRO et al., 2012; CASTRO, 2006 apud BOTELHO; CUNHA; MACEDO, 2011).

Para Cook, Mulrow e Raynes (1997), a qualidade de uma revisão da literatura e, consequentemente, 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 são apresentadas no Quadro 1 a seguir.

Quadro - 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 baseadas 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, amostragens e processos de análise. A seguir, são apresentadas as definições gerais de cada subtipo de revisão sistemática. Essas definições são comparadas no Quadro 2, explorando-se também o proposto, o escopo, a amostra e o tipo 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 pesquisa pregressas (GLASS, 1976). Assim, a meta-análise consiste em colocar diferentes estudos juntos em um mesmo banco de dados e utilizar 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, *grouded theory* e a 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 (WHITEMORE, 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 possibilitarem a aplicação em diferentes propósitos como definir conceitos, rever teorias, avaliar evidências e analisar questões metodológicas de um tema específico (BROOME, 1993 apud WHITEMORE; KNAFL, 2005).

Quadro - Comparação dos tipos de revisão de literatura

| **Tipo de revisão** | **Definição** | **Propósito** | **Escopo** | **Amostra** | **Análise** |
| --- | --- | --- | --- | --- | --- |
| **Méta-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 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 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).

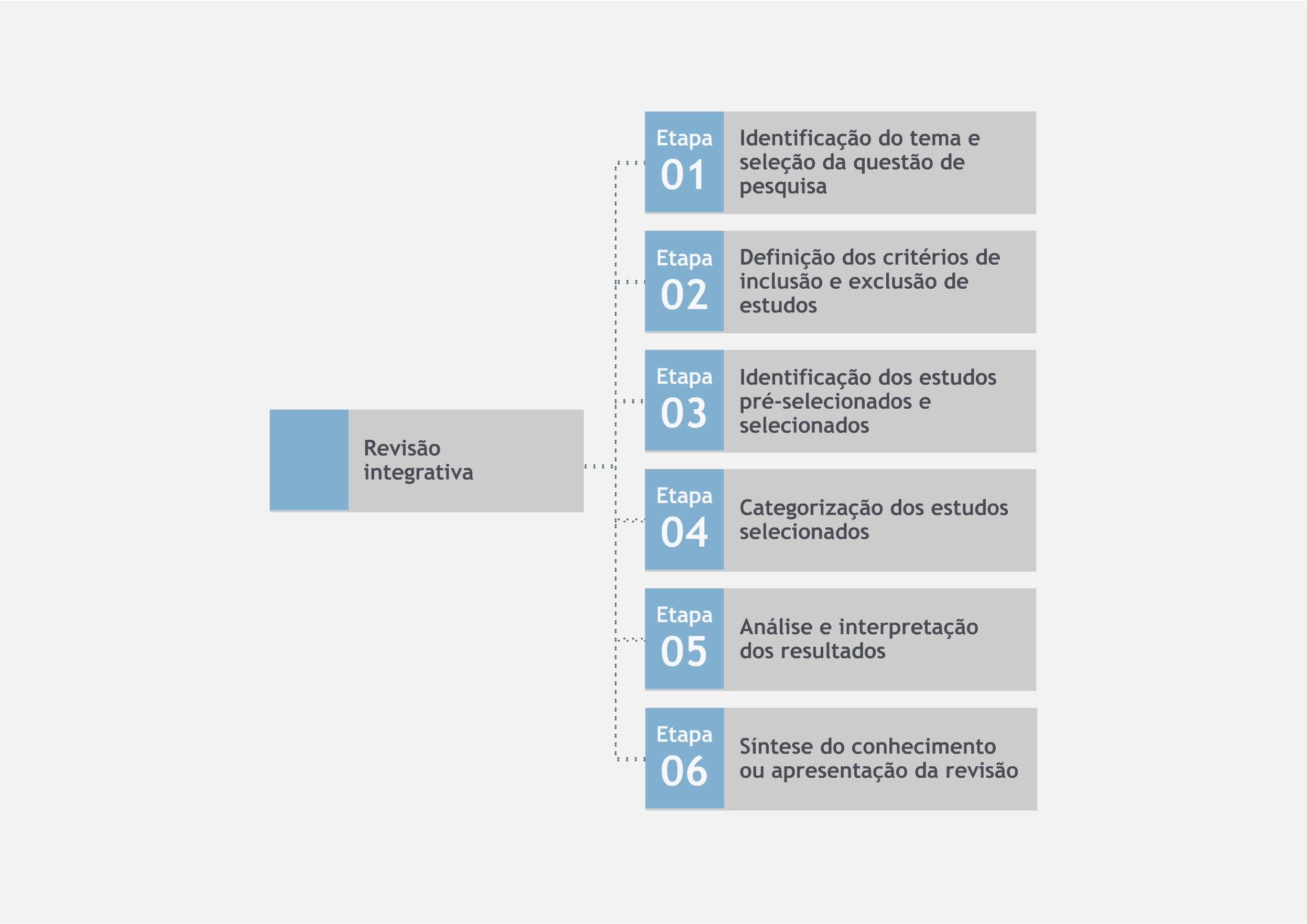
Conforme exposto, a revisão da literatura utilizada neste trabalho tem por finalidade identificar as abordagens atuais que possibilitem a extração de perfis de especialistas a partir de documentos para que se defina um método. Por se tratar de uma revisão de métodos, teorias e estudos empíricos numa 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.

### 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 (WHITEMORE, 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, Mendes, Silveira e Galvão (2008) e Botelho, Cunha e Macedo (2011), afirmam que é necessário que as etapas a serem seguidas estejam claramente descritas e percorram 6 etapas distintas descritas na Figura 2Figura 3 e detalhadas no Quadro 3.

Figura - Etapas da revisão integrativa.

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

Quadro – 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 utilizado. Com a pergunta de pesquisa define-se os descritores e constrói-se a estratégia de busca. |
| **2 – Definição dos critérios de inclusão e exclusão de estudos** | Executa-se a busca nas bases de dados, para identificação dos estudos que serão incluídos na revisão.  Essa 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 na medida em que o pesquisador retorna à sua questão inicial.  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 durante a elaboração da revisão integrativa. |
| **3 – Identificação dos estudos pré-selecionados e selecionados** | Realiza-se a leitura criteriosa dos títulos, resumos e palavras-chave de todas as publicações completas localizadas pela estratégia de busca, para posteriormente verificar sua adequação aos critérios de inclusão do estudo.  Nos casos em que o título, o resumo e as palavras-chave não sejam suficientes para definir sua seleção, busca-se a publicação do artigo na íntegra.  Com a conclusão desse procedimento, elabora-se uma tabela com os estudos pré-selecionados para a revisão integrativa. |
| **4 – Categorização dos estudos selecionados** | Sumarizar-se e documenta-se as informações extraídas dos estudos científicos encontrados nas fases anteriores.  Um dos instrumentos utilizados para extrair as informações dos estudos selecionados é a matriz de síntese utilizada como ferramenta de extração e organização de dados de revisão da literatura em várias disciplinas, devido à sua capacidade para resumir aspectos complexos do conhecimento.  A matriz serve de ferramenta de interpretação e construção da redação da revisão integrativa para os pesquisadores.  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 seus dados. |
| **5 – Análise e interpretação dos resultados** | 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.  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. |
| **6 – Síntese do conhecimento ou apresentação da revisão** | Elabora-se um documento que contempla a descrição de todas as fases percorridas pelo pesquisador, de forma criteriosa, e que apresente os principais resultados obtidos.  A revisão deve possibilitar a replicação do estudo. Deve contemplar informações que possibilitem que os leitores avaliem a pertinência dos procedimentos empregados na elaboração da revisão. |

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

Em sua execução, a revisão integrativa possibilita a síntese e 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 a que possa conhecer a evolução do tema ao longo do tempo e, com isso, visualizar possíveis oportunidades de pesquisa”.

## CONTEXTUALIZAÇÃO

A contextualização 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 com a visão dos autores sobre os temas relacionados. Ademais, essa seção situa a pesquisa para a elaboração da revisão sistemática da literatura com o objetivo de observar o estado da arte que abordam os temas de interesse.

### 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; SCHREIBER et al., 1999; VENZIN; KROGH; ROOS, 1998; TEECE; PISANO; SHUEN, 1997; SPENDER; GRANT, 1996). Apesar de sua crescente importância, o conhecimento é um conceito bastante amplo e abstrato que tem sido debatido de forma epistemológica desde era clássica grega (ALAVI; LEIDNER, 2001; SCHREIBER et al., 1999). Sócrates em Teeteto, um diálogo sobre a natureza do conhecimento escrito em aproximadamente 369 A.C., 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 seu trabalho de revisão da literatura de entendimento da Gestão do Conhecimento, buscaram referências em autores que trabalharam com Gestão do Conhecimento desde sua afirmação no campo científico e também trouxeram definições sobre Conhecimento (e.g. NONAKA; TAKEYCHI, 1995; GRANT, 1996; WIIG, 1993; DAVENPORT; PRUSAK, 1998; BECKMAN, 1997; entre outros). Essas definições são apresentadas no Quadro 4 a seguir.

Quadro – Definições de conhecimento.

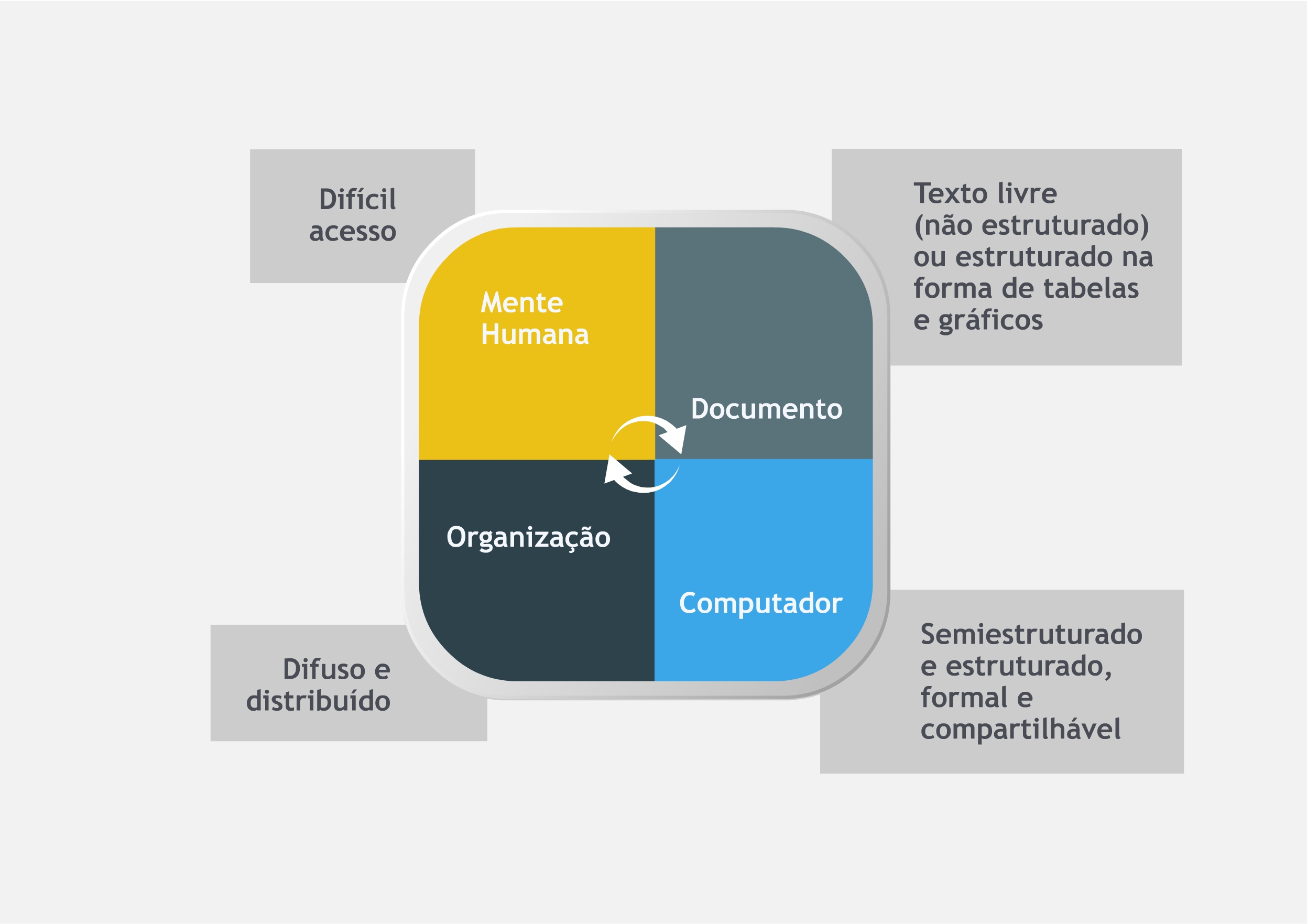
|  |  |
| --- | --- |
| Definição | Referência |
| O conhecimento é um fator de produção. | Nonaka e Takeuchi (1995) |
| Conhecimento reside na cabeça do indivíduo de [...] conhecimento é aquilo que é conhecido. | Grant (1996) |
| Conhecimento consiste em verdades e crenças, perspectivas e conceitos, julgamentos e expectativas, metodologias e *know-how*. | Wiig (1993) |
| Conhecimento é informação no contexto juntamente com uma compreensão de como usá-lo. | Davenport e Prusak (1998) |
| Conhecimento é informação combinada com a experiência, contexto, interpretação e reflexão. | Davenport e Long (1998) |
| O conhecimento é o raciocínio sobre informações para orientar ativamente a execução da tarefa, resolução de problemas e tomada de decisão a fim de realizar, aprender e ensinar. | Beckman (1997) |
| Conhecimento é definido como a compreensão dos efeitos das variáveis de entrada na saída. | Bohn (1994) |
| Conhecimento é como uma percepção nova ou modificada ou um entendimento preditivo. | Kock e Queen (1998) |
| O conhecimento é o conjunto de ideias, experiências e procedimentos que são considerados corretos e verdadeiros, e que, portanto, guia os pensamentos, 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) |
| Conhecimento refere-se a ao estoque de informação, habilidades, experiência, crenças e memórias de um indivíduo. | Alexander, Schallert e Hare (1991) |
| 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, crenças e valores com a decisão e a ação que ganham significado. Conhecimento de um indivíduo pode ser diferente do conhecimento adquirido por uma 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: o estado de saber; capacidade de ação; e 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 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 aborda o como 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 no qual 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 3 a seguir.

Figura – Tipos de repositórios de conhecimento.



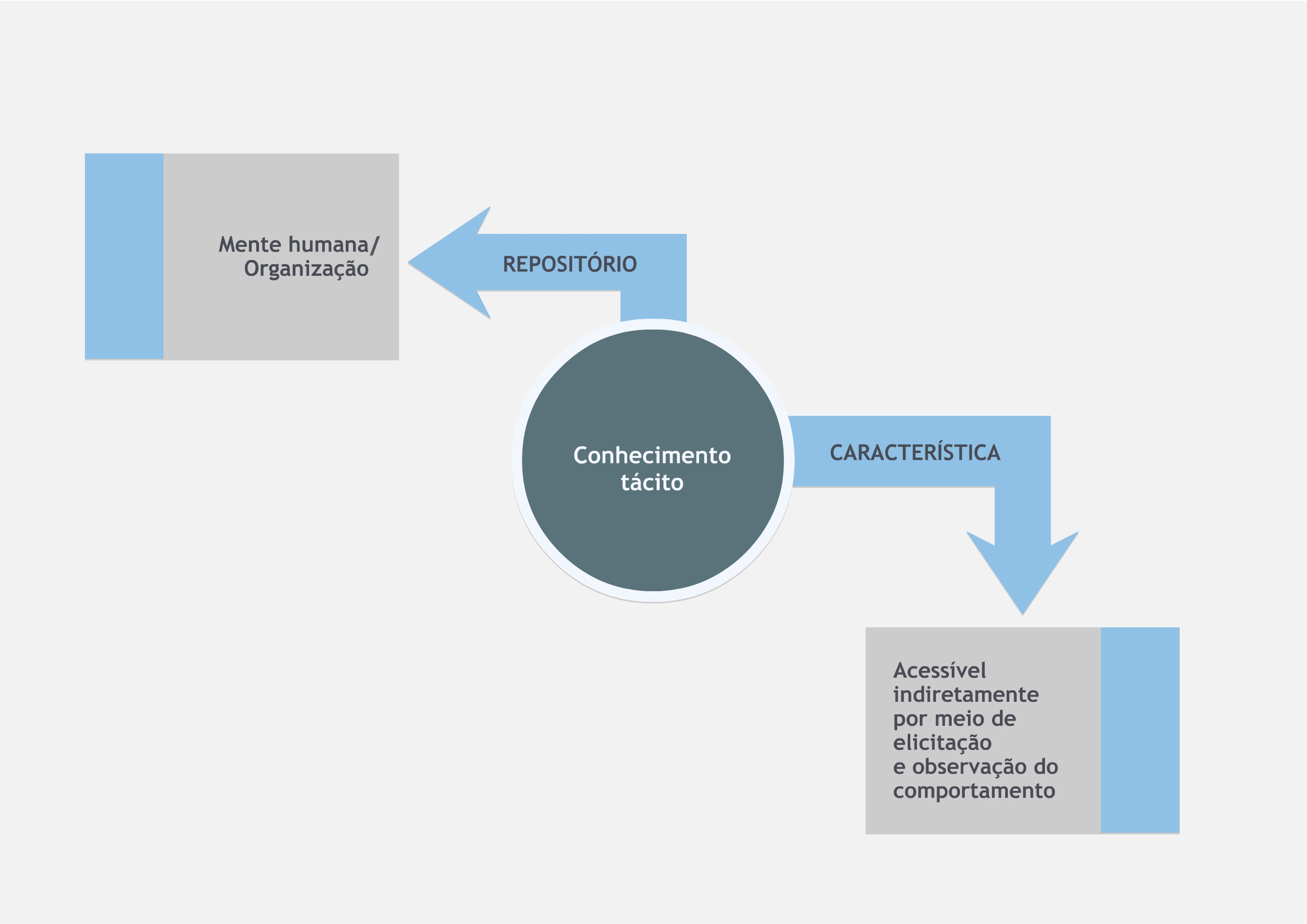
Fonte: Adaptado de Anand e Singh (2011, p. 929).

Esse grau de acessibilidade, também conhecido pela dimensão acessibilidade do conhecimento (ANAND; SINGH, 2011), tem sido discutido por diversos autores ao longo dos anos (e.g. NONAKA; TAKEUCHI, 1995; WINTER, 1987; TEECE, 1998; HANSEN, 1999; DAVENPORT; PRUSAK, 1998; COOK; BROWN 1999). Os autores Nonaka e Takeuchi (1995), a partir dos estudos de Polanyi (1967), dividiram a acessibilidade em duas categorias: tácito e explícito.

#### Conhecimento tácito

Uma das definições mais difundidas sobre conhecimento tácito encontrada na literatura foi cunhada por Polanyi (1966) no qual 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 a ação e a prática (i.e. *know how*), adquirido pela experiência pessoal, raramente expresso verbalmente e muitas vezes se assemelha a intuição (SMITH, 2001). A Figura 4 resume as características do conhecimento tácito e sua forma de armazenamento.

Figura – Conhecimento tácito: característica e forma de armazenamento.

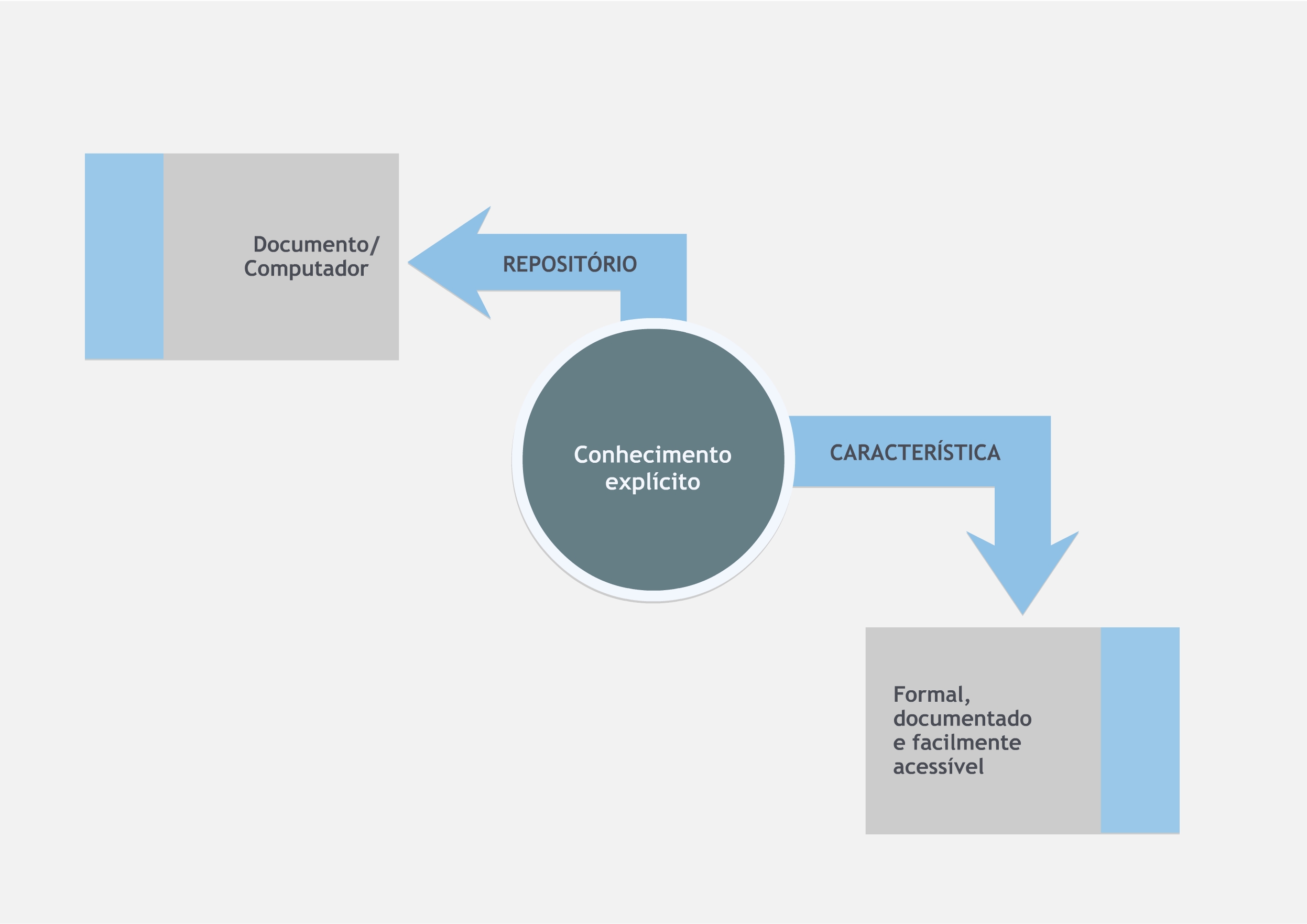


Fonte: Adaptado de Anand e Singh (2011, p. 930).

#### Conhecimento explícito

Diferentemente do conhecimento tácito, o conhecimento explícito é aquele que pode ser expressado na forma de palavras, números, textos, especificações, manuais e também é facilmente comunicável e compartilhável (NONAKA; TAKEUCHI, 1995). Para Smith (2011), uma importante parte desse conhecimento explícito (i.e. *know what*) pode ser classificado como conteúdo técnico ou conteúdo acadêmico, que são 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 estudo estruturado. Esse tipo de conhecimento também pode ser acessado rapidamente quando se encontra em um sistema de informação. A Figura 5 resume as características do conhecimento tácito e sua forma de armazenamento.

Figura – Conhecimento explícito: característica e forma de armazenamento.

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

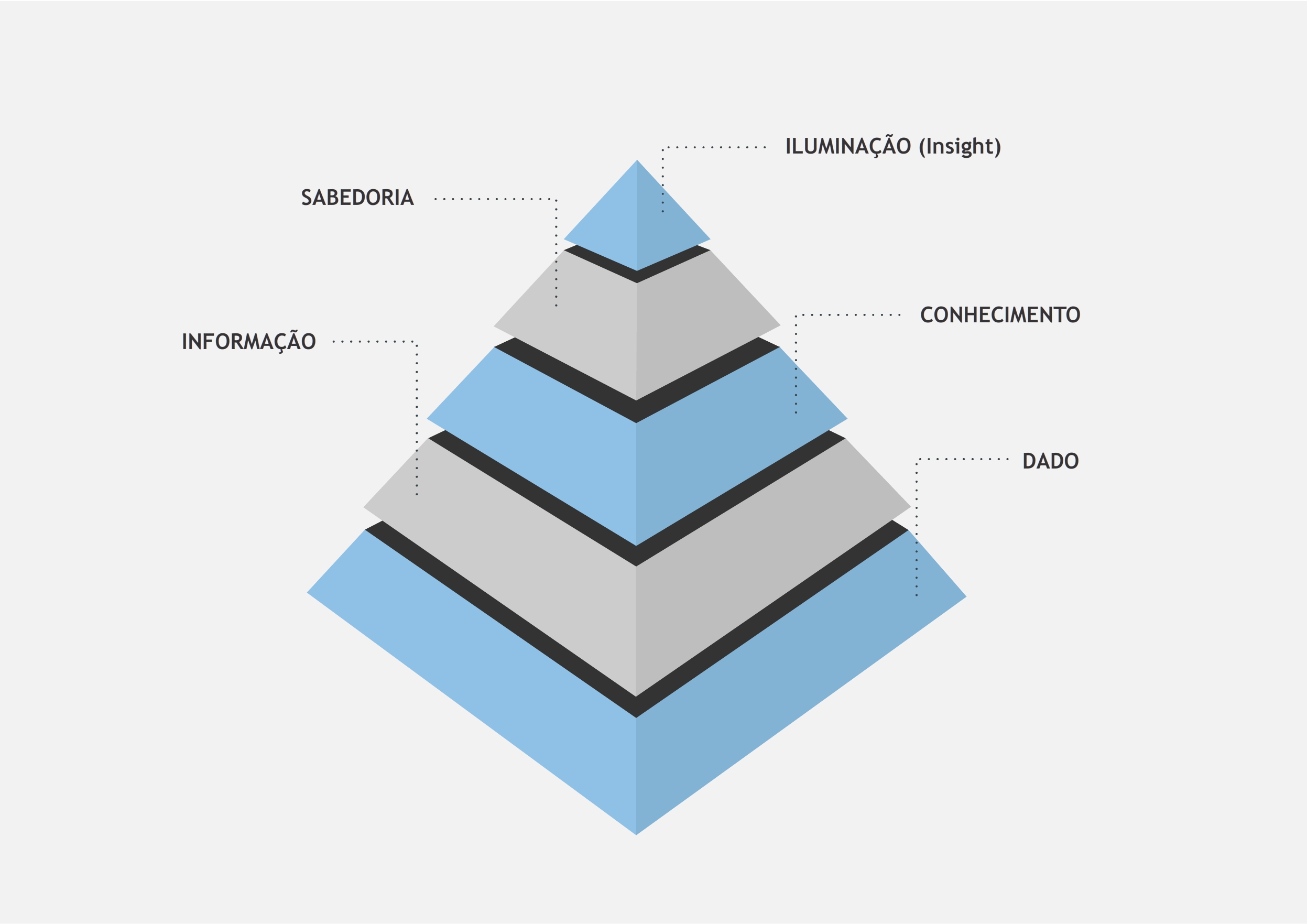
#### Dados, informação, conhecimento

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 Aune (1970), Ackoff (1989), Goldman (1991), King (1993), Argyris( 1993), Nonaka e Takeuchi (1995) , Gallup et al. (2002), Awad e Ghazi ( 2004), Desouza (2005), 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 a informação que foi processada em alguns aspectos significativos. Schreiber et al. (1999) também diferenciaram esses conceitos, conforme é exibido a seguir, e apresentaram 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 amarelo 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 à 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 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”.

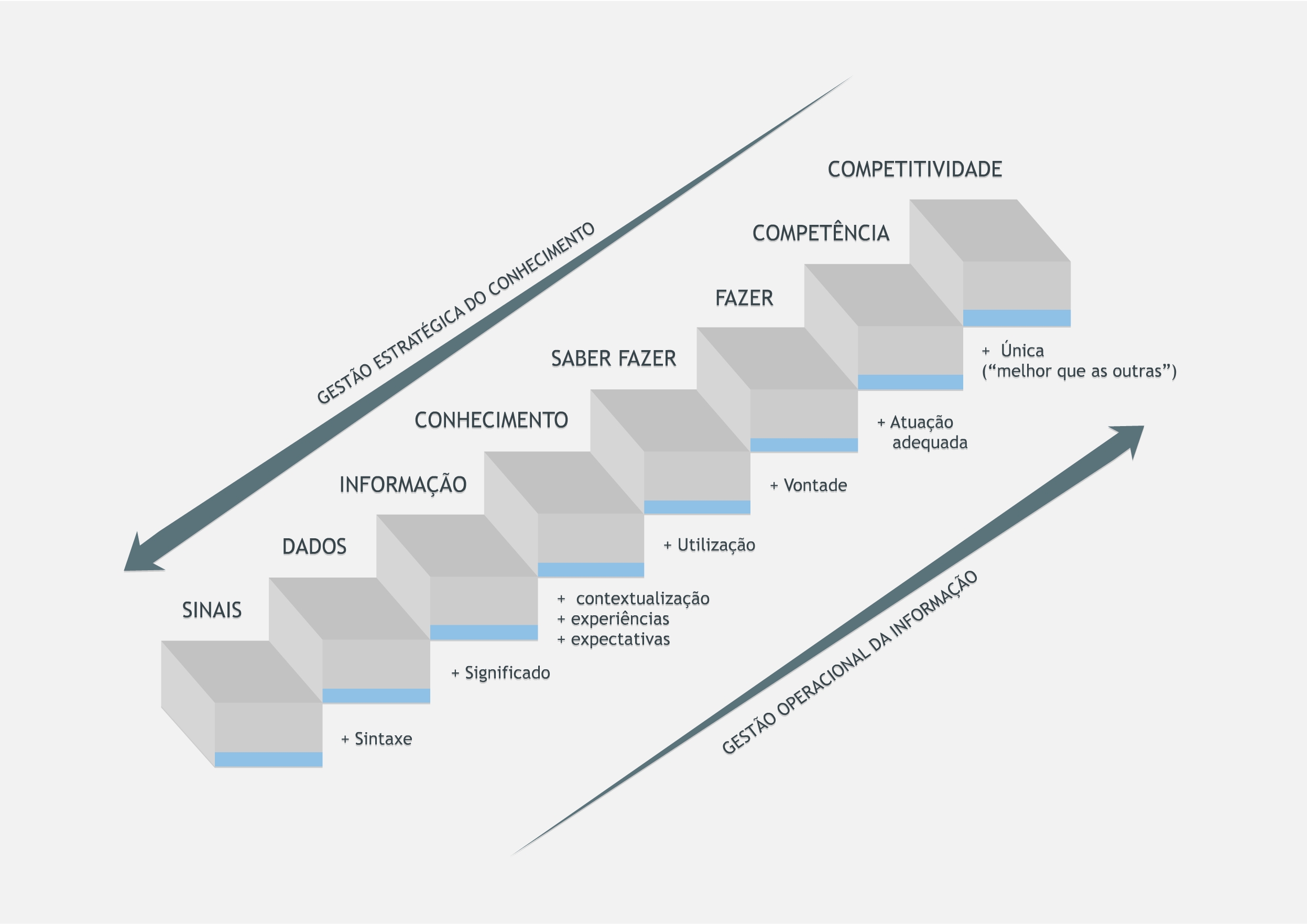
Além de apresentar os conceitos e as diferenças entre dado, informação e conhecimento, há trabalhos como de Anand e Singh (2011) e de North e Rivas (2008) que apresentam a pirâmide do conhecimento (Figura 6) e a escada do conhecimento (Figura 7), respectivamente. Esses trabalhos demonstram a hierarquia entre esses conceitos, de modo a expandi-los em sabedoria e iluminação (ANAND; SINGH, 2011); e saber fazer, fazer, competência e competitividade (NORTH; RIVAS, 2008).

Figura – Pirâmide do conhecimento (dado, informação e conhecimento).



Fonte: Adaptado de Anand e Singh (2011, p. 932).

Figura – Escada do conhecimento (dado, informação e conhecimento).



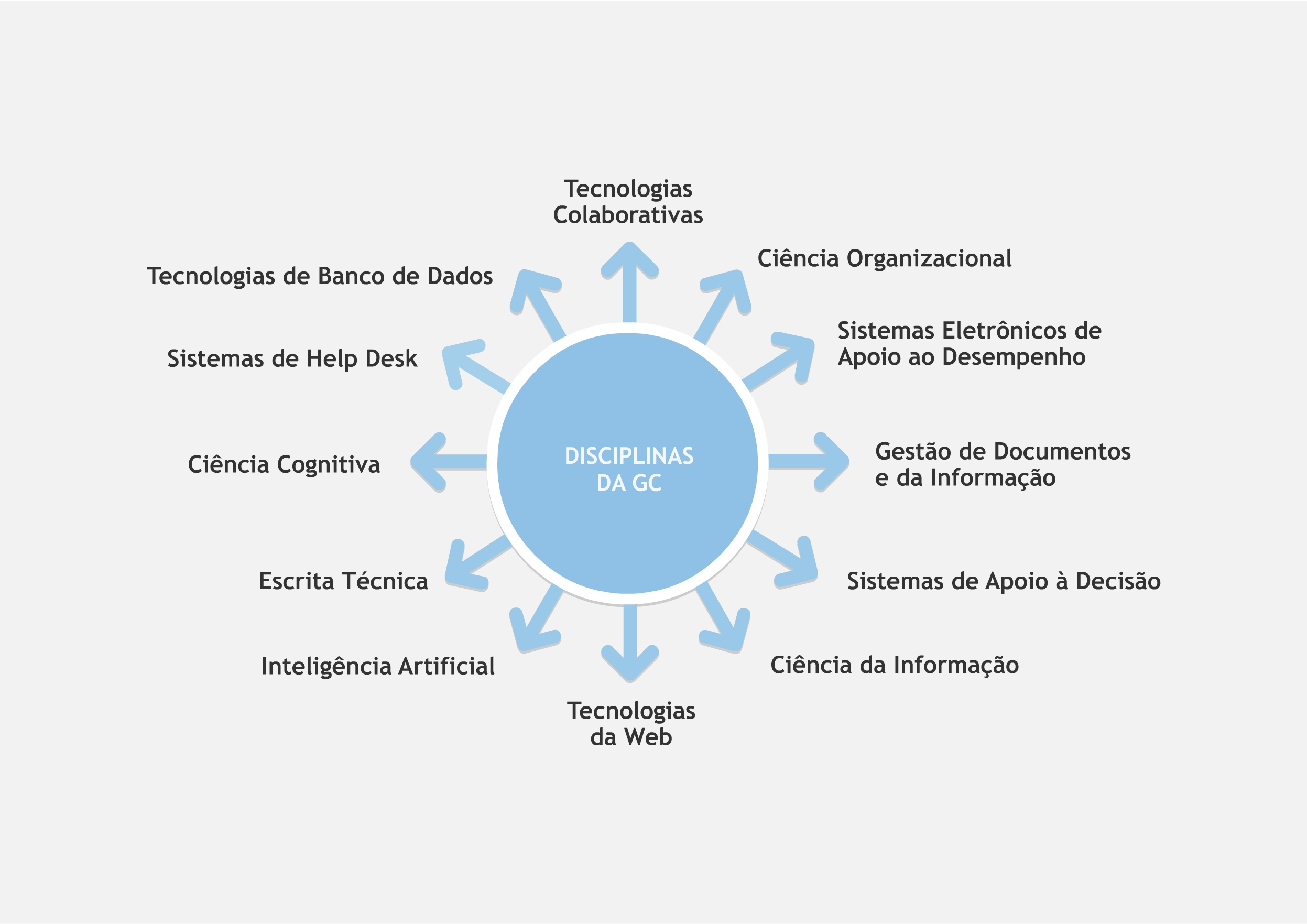
Fonte: Adaptado de North e Rivas (2008, p. 42).

### 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 o conhecimento 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 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 aplicá-la na prática (ANAND; SINGH, 2011; CHOI; LEE, 2002; ALAVI; LEIDNER, 2001; SMITH, 2001). No que tange as 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 que são necessárias para a sua implantação em uma organização. A Figura 8 ilustra as disciplinas ligadas à GC.

Figura – 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, compartilhamento e 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 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: Wiig (1993), Meyer e Zack (1996), Mc Elory (1999), Bukowitz e Williams (2003), Wong e Aspinwall (2004), Lee et.al. (2005) e Dagnfous e Kah (2006). 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, são detalhados cada um desses processos 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 os processos cognitivos do indivíduo (por exemplo, a reflexão), no qual o conhecimento é criado, compartilhado, amplificado, ampliado e justificado em ambientes organizacionais. Este processo 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, organização e recuperação do conhecimento organizacional, também conhecido como memória organizacional, consiste no gerenciamento do conhecimento que reside em documentos, informações estruturadas armazenadas em bases de dados eletrônicas, em 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 a 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 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 desenvolvidas por meio da conversão do conhecimento tácito de especialistas em conhecimento explícito. Enquanto que 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. Por fim, as equipes de trabalho dizem respeito à criação de grupos com conhecimento específicos que possam trabalhar em conjunto para 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 ferramentas que que contribuem na implantação da Gestão do Conhecimento em uma organização. A APO (2010) apresenta uma lista de métodos e ferramentas classificados tecnológicos (i.e. utilizam TI) e não tecnológicos (i.e. não utilizam TI) que suportam esses processos, conforme demonstrado Quadro 5 no a seguir.

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

|  |  |
| --- | --- |
| Métodos e ferramentas não tecnológicos | Métodos e ferramentas tecnológicos |
| * *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 | * *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 APO (2010, p. 9).

### Engenharia do Conhecimento

A Engenharia do Conhecimento teve o seu surgimento a partir da Inteligência Artificial como uma subárea dedicada à concepção, desenvolvimento e implementação de Sistemas Baseados em Conhecimentos (SBC). Nesse período, definido como Engenharia do Conhecimento Clássica, os especialistas extraíam o conhecimento e transferiam para uma base de conhecimento computacional (STUDER; BENJAMINS; FANSEL, 1998). Mais recentemente, a nova Engenharia do Conhecimento teve 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 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 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 se tornou uma valiosa fonte de princípios e técnicas de modelagem do conhecimento, a qual oferece um conjunto de etapas para o desenvolvimento de Sistemas Baseados em Conhecimento (SCHREIBER et al., 2002).

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, enquanto que o termo "engenharia" está relacionada ao ato de 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 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 gestão do conhecimento organizacional; fornece métodos para obter uma compreensão completa das estruturas e processos na criação, distribuição e aplicação do conhecimento, possibilitando uma melhor integração da tecnologia da informação e a utilização do conhecimento; e, por último, possibilita a construção de sistemas de conhecimento mais fáceis de usar, bem estruturados arquiteturalmente e mais simples de se manter.

### *Expertise Location*

Frequentemente, pessoas consultam outras pessoas para ajudá-las na busca do caminho correto para 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 localização de especialistas que possui conhecimento e experiência necessários às outras pessoas e têm ganhado destaque na Gestão do Conhecimento (APQC, 2015; BALOG et al., 2012; APO, 2010; MARWICK, 2001; MCDONALD; ACKERMAN, 1998). Principalmente, em grandes empresas, no qual saber quem são as pessoas que possuem experiências, habilidades e conhecimento é dificultada pela quantidade de pessoas (APO, 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 (e.g. conhecimento) 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, como: remoção de *stopwords* de documentos para aumentar a precisão de buscas; *stemming* para converter palavras flexionadas com os mesmos significados semântico; 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 e 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 mencionado como design de interação, determina a forma de buscar e apresentar os dados de um sistema de recuperação de *expertise* de forma acessível aos usuários. Pontos importantes 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.

### Discussão e conclusão da contextualização

Esta contextualização 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 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. Sua origem é dada na mente das pessoas e pode ser codificado em diversos meios como documentos, manuais, livros, códigos de computares (e.g. e suas respectivas regras de negócios). Além disso, o conhecimento possui classificação quanto à sua acessibilidade, sendo classificado em tácito e explícito. Tácito quando é de difícil transferência, que ocorre por observação; e explícito é de fácil transferência que pode ser externalizado em documentos ou verbalizado. Observa-se ainda, a utilização de dados, informação e conhecimento como o mesmo sinônimo, porém que representam conceitos diferentes.

Para que o conhecimento seja aplicado de forma eficiente e eficaz com o intuito de de valor nas 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 a melhor utilização e reutilização de conhecimento em uma organização de modo a torná-la mais competitiva. Para que isso ocorra, precisa-se aplicar 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 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 como classificação).

Chega-se, então, às ferramentas de *Expertise Location*, que se tornaram grandes aliadas à 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 responsável por 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 tem relação 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 está relacionada à exploração da área de Recuperação de Informação. Essas ponderações são detalhadas nas próximas subseções.

#### Uso dos termos Expertise e Expertise Location.

De origem francesa, segundo o dicionário online Priberam[[3]](#footnote-3), o significado de *expertise* apresentado no site Significados[[4]](#footnote-4) 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 competência para executar algo”, sinônimo de *know-how.* Já o tradutor do Google[[5]](#footnote-5), define expertise como “habilidade ou conhecimento em um campo de estudo particular”, e traz como sinônimos os termos “competência, proficiência e especialização”.

Por trazer vários significados e sinônimos e não possuir uma definição em língua portuguesa (i.e. os dicionários Aurélio[[6]](#footnote-6) e Michaelis[[7]](#footnote-7) não possuem definição), decidiu-se não traduzir o termo neste trabalho para que não se perca semântica e por entender que o termo se aproxima das definições de conhecimento, seja ele tácito ou explícito. Também se optou por não traduzir os termos *Expertise* *Location*, que, basicamente, consiste na localização de *expertise*.

#### Expertise Location e a área de Recuperação de Informação

Segundo os atores 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, segundo Beppler (2008 *apud* SALTON, 1968), tem foco na estrutura, análise, organização, armazenamento e 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 estrutura, documentos descritos em linguagens naturais e ampla variedade de assuntos presentes em documentos (GONÇALVES, 2005 apud MITRA; CHAUDHURI, 2000).

Gonçalves (2005) e Beppler (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, tesauros e extração de entidades estão sendo utilizados. Na fase de representação de documentos, observa-se também a utilização 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 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 dá área de RI fornecer uma série de métodos e técnicas para trabalhar documentos que podem indicar tópicos de *expertise* de especialistas, o objetivo dessa 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, consequentemente, as iniciativas de GC, são observados na próxima seção que busca apresentar o estado da arte com relação aos temas observados na contextualização da literatura (i.e. GC, EC e EL).

## ESTADO DA ARTE

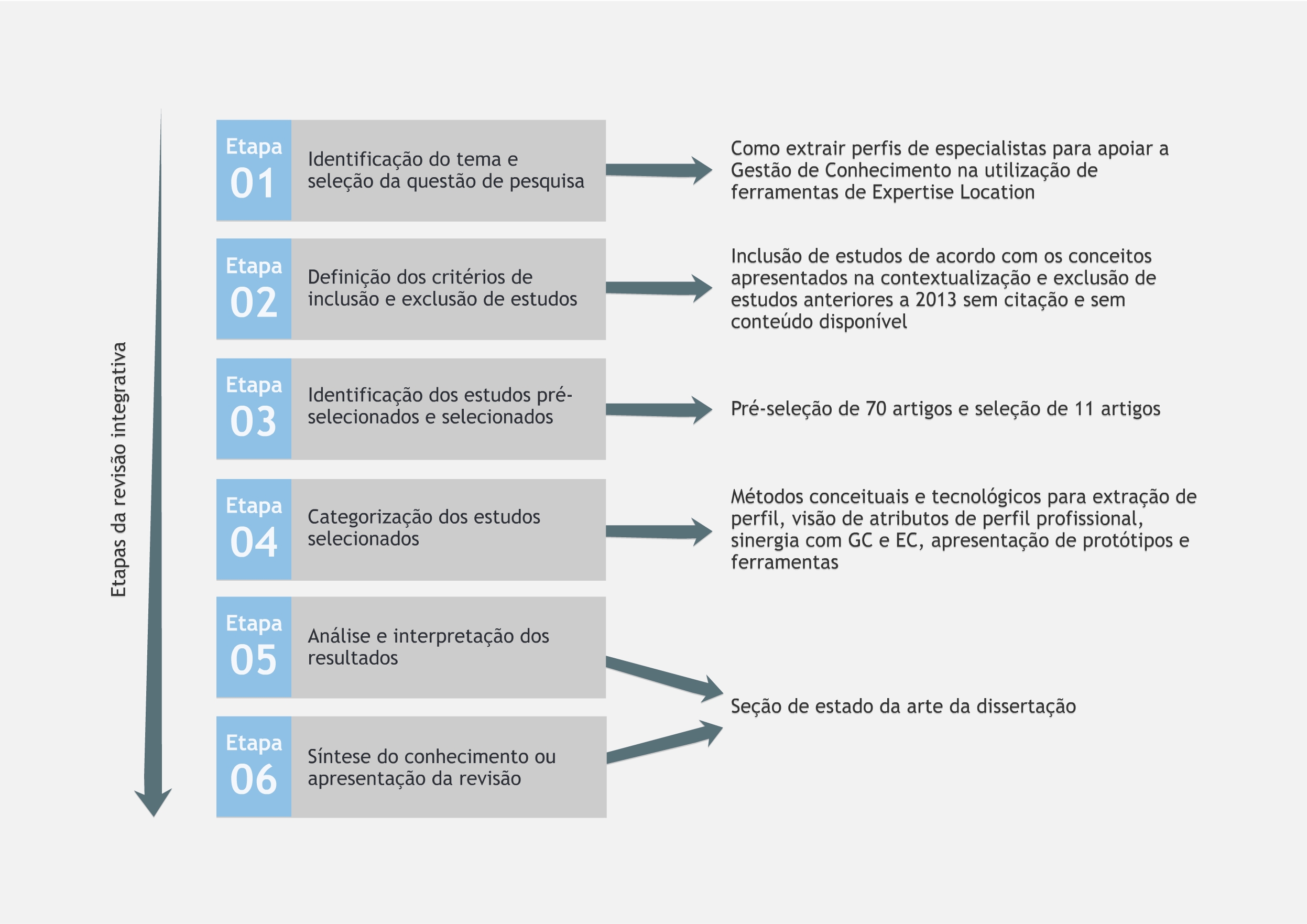
No âmbito da revisão da literatura, o estado da arte busca apresentar o conteúdo já publicado sobre um determinado tema, de modo que se possa identificar as lacunas existentes e onde se encontram entraves teóricos ou metodológicos (SILVA, 2005). Esse conteúdo possibilita, ainda, obter as informações atuais sobre um determinado tema de pesquisa, identificar publicações e confrontar opiniões.

Para determinar o estado da arte sobre métodos e técnicas que possibilitem 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. Esse método, além de possibilitar 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, a primeira etapa é a identificação do tema e a 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-chaves 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 9, a seguir, apresenta as etapas da revisão integrativa executadas nessa pesquisa e os seus respectivos resultados de forma resumida para cada etapa. A descrição completa da execução do método e seus resultados são detalhados no capítulo 3.1 deste trabalho.

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



Fonte: elaboração própria.

Com relação à análise e interpretação dos resultados e, consequentemente, a síntese do conhecimento encontrado nessa revisão integrativa da literatura, os seus achados são apresentados nas próximas subseções. Essas subseções detalham os conteúdos com base na matriz de síntese apresentada no capítulo 3.1.4. O foco desse detalhamento está, primeiramente, na apresentação do apoio da *Expertise Location* na GC e, posteriormente, na apresentação 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.

### *Expertise Location* no contexto da GC

A análise e a interpretação dos estudos quanto ao apoio da *Expertise Location* na Gestão do Conhecimento busca apresentar como esses temas estão relacionados na visão dos autores que tiveram 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. Em seu estudo, Becerra-Fernandez analisa a utilização de Sistemas de *Expertise Location* (*Expetise 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 classifica o ELS como um sistema que contribui para a GC, isto é, como um Sistema de Gestão do Conhecimento (SGC). Para Becerra-Fernandez, 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 reutilização na resolução de novos problemas, bem como 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 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 que localizam especialistas, ou seja, aqueles que detêm o conhecimento.

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 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 profissional. Portanto, esse estudo encontra-se alinhado em todas as propostas e contexto deste 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 de 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. Esse estudo de Balog e Rijke de 2007 foi utilizado como referência no estudo publicado por Balog et al. de 2012 apresentado na seção 2.2.4.

Para Balog e Rijke, 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 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 é a de determinar o perfil de um especialista, ou seja, uma descrição concisa das áreas 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, assim como é apresentado nos estudos de Balog et al. (2012), separam a atividade de *Expertise Location* em duas tarefas distintas: (a) construção de perfil profissional e (b) localização de especialistas. Essa mesma distinção é adotada nesse trabalho para, primeiramente, construir um perfil profissional a partir de documentos não estruturados e, em seguida, criar um ambiente de busca desses perfis.

Balog e Rijke ainda situam seu trabalho quanto à Gestão do Conhecimento ao esclarecerem que essas tarefas são previamente tratadas por sistemas de páginas amarelas ligadas à GC. Porém, também discutem que a maioria das ferramentas dependem 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, em 2008, de forma aderente aos estudos de Becerra-Fernandez, aborda 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, armazenamento, compartilhamento e busca de informações críticas para tomada de decisões em seus negócios. Além de citar a importância da GC, Yang e Huh explicam os tipos de conhecimentos, tácitos e explícitos, conceituados por Nonaka e Takeuchi (1995) e que são apresentados na seção 2.2.1 deste trabalho. Nesse contexto, os autores também fazem a classificação do ELS como um Sistemas de Gestão de Conhecimento.

Yang e Huh explicam que os sistemas *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 tivesse facilidade para desenvolvimento relacionada à implementação das ferramentas, o método manual requeria constantes custos administrativos para atualização dos perfis. Além dessa desvantagem com o custo de preenchimento dos perfis, existia, ainda, a dificuldade de refletir novos conceitos e terminologias utilizadas 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 foram abordados de maneira que seu estudo pudesse contribuir, primeiramente, na definição da estrutura do perfil do especialista e, na sequência, na geração automática desse perfil. Para primeira contribuição, os autores utilizaram 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 documentos armazenados 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 de 2011, que seguem na mesma linha de Yang e Huh, verificou-se, mais uma vez, a declaração explicíta 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 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 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 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, esse estudo ajuda ao descrever soluções que podem auxiliar na utilização de taxonomias nos sistemas de EL.

Por fim, Sohail, Afzal e Ahmad, em 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 sistemas de EL da Microsoft, Hewlett-Packard e NASA desenvolveram sistemas para localizar especialistas para 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 destacam que sistemas de EL são decisivos nos cenários de resolução de problemas, preenchimento de vaga de trabalho, aumento de produtividade e busca por colaboradores que trabalham na mesma área de atuação. A revisão da literatura de seu trabalho apresenta, 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.

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

Um dos resultados mais importantes para esta pesquisa com relação à revisão integrativa da leitura 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 do estado da arte sobre a temática. Assim, de forma similar à seção 2.3.1, 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 de 2006. Apesar de seu estudo trazer uma comparação de sistemas de *Expertise Location* até 2006 e sugerir 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 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 descreve que o uso de *Web Mining* pode mitigar alguns dos problemas inerentes ao preenchimento tendencioso de dados pelas próprias pessoas para manter seus dados em sistemas de EL e bem como pode mitigar a necessidade de desenvolver taxonomias manualmente para expressar os conhecimentos de uma organização. 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. Todos os nomes no banco de dados dos colaboradores são organizados em uma estrutura de dados mapa. O mapa é composto por todos os nomes de funcionários, armazenando esses nomes 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 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. A lista resultante de palavras é processada para determinar se uma palavra foi utiliza em uma lista de palavras irrelevantes (e.g. *stopwords*). A lista resultante é então processada com um algoritmo para remover o sufixo de uma palavra (i.e. *stemming*). Esse processo é feito 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 forma 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 de Balog e Rijke, de 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 um método para medir a competência das pessoas nessas áreas.

O método de Balog e Rijke para identificar as áreas de conhecimento de uma pessoa,tem por objetivo descobrir a *expertise* de uma pessoa com base na exploração de documentos não estruturados. Dessa forma, Balog e Rijke esclarecem que as áreas de conhecimento são obtidas por meio de processamento linguístico. As áreas de conhecimento extraídas por meio das técnicas linguísticas são associadas com as pessoas identificadas nesses documentos. Isto é, para se identificar as pessoas também é realizado um algoritmo de PLN que encontra as entidades relacionadas às pessoas (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. 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 a técnica TF-IDF.

O terceiro estudo, de 2007, dos autores de Yao, Tang e Li, 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 profissional de forma segmentada, aplicando-se algoritmos especializados para cada unidade de informação classificadas 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 seu conteúdo com base em *tags* esperadas. Seus estudos apontam a utilização de *Support Vector Machines* (SVM) para a classificação das informações dos conteúdos encontrados.

O quarto trabalho, também de 2007, dos autores Yang e Huh, aborda a identificação do especialista usando uma técnica de categorização de texto. Seu processo de categorização é dividido em duas etapas utilizando a abordagem *Vector Space Model* (VSM): 1) para preparação e treinamento de documentos; 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 passos de: 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 a atualização dos perfis de especialistas na base de dados da organização.

O quinto trabalho de Yang et al., de 2008, propõe a construção de perfil de especialista de forma automática utilizando as informações disponíveis sobre suas publicações. Para classificar seus conhecimentos, são utilizados os domínios da Wikipedia. Basicamente, seu trabalho relata a construção de uma taxonomia com base nas informações da Wikipedia 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).

O sexto trabalho de Li, Liu e Li, de 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 lingüístico 2-tupla (2-Tuple Linguistic Model) proposto por Herrera e Martínez (2000).

O trabalho de Krusheva e Tsiporkova de 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, como: LinkedIn, DBLP, Microsoft Academic Search e Google Scholar Citation. As autoras sugerem a utilização de PLN para a captura de palavras-chaves que possam descrever as expertises dos especialistas.

O oitavo trabalho, também de 2012, de Liu et al., 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 utiliza 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, de 2013, dos autores Sohail, Afzal e Ahma, trabalham 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).

### Discussão e conclusão do estado da arte

A seção do estado da arte, 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 a esta pesquisa: a contribuição das ferramentas de *Expertise Location* à Gestão do Conhecimento; e os métodos conceituais e tecnológicos para a extração de perfis de especialistas a partir de documentos.

Sobre o primeiro aspecto, a contribuição das ferramentas de *Expertise Location* à Gestão do Conhecimento, observa-se um conjunto de trabalhos que correlacionam esses temas e onde se destacam os trabalhos realizados por Becerra-Fernandez (2006) da Universidade Internacional da Florida dos 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 na Coréia do Sul; 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 e geração de perfil de especialista, ou seja pela contribuição na busca e recomendação de especialistas. É importante ressaltar que, apesar de descrever como métodos conceituais e tecnológicos, essa 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 da área de Recuperação da Informação e da Inteligência Artificial como pode ser observado na seção 2.2.4 de contextualização dos temas no que tange ao processo de recuperação de *expertise*. Destacam-se 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.

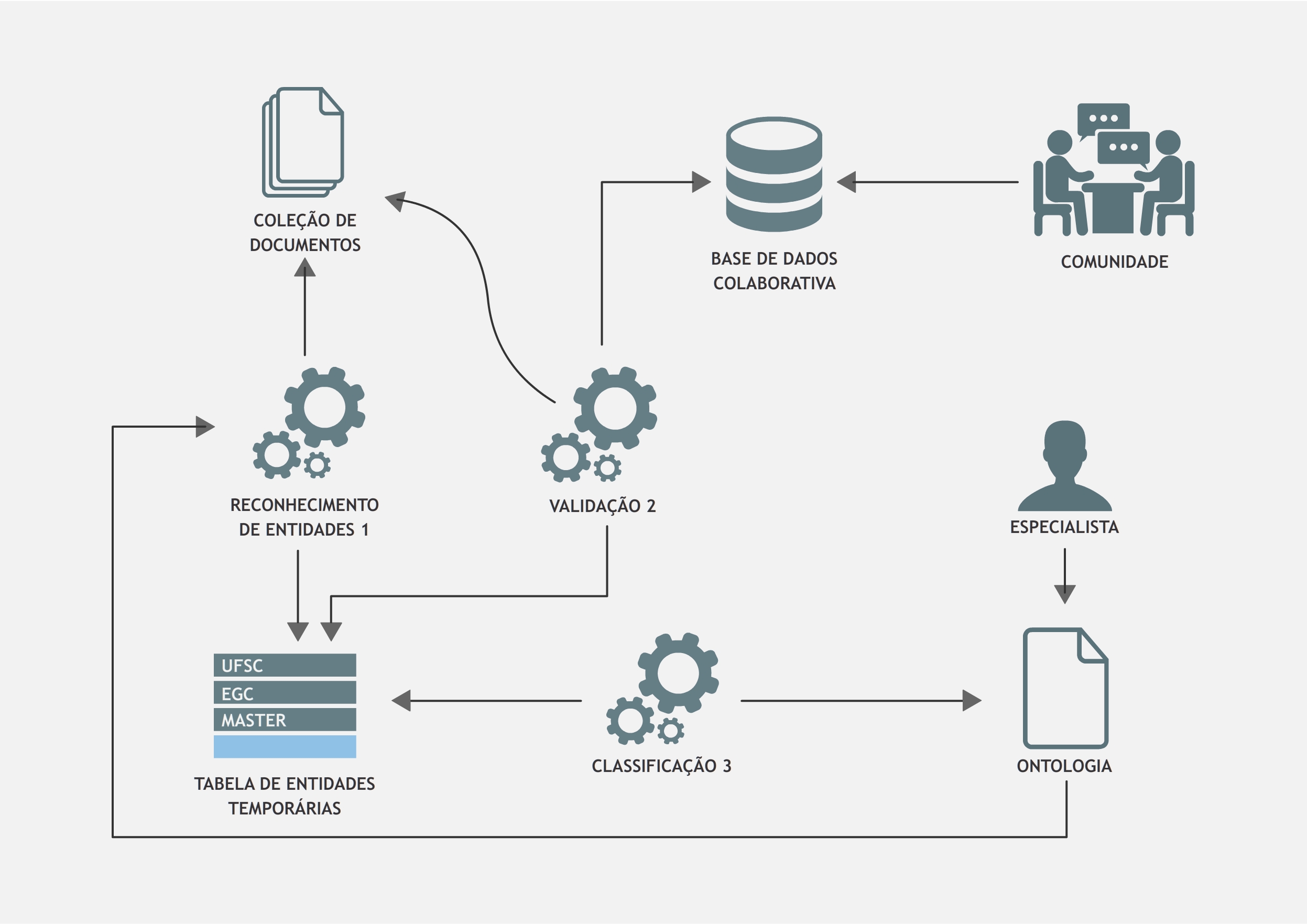
Conclui-se que esses trabalhos, apesar de fornecerem visões sobre diferentes propostas conceituais e tecnológicas para a exploração do tema de *Expertise Location*, não apresentam soluções para uma questão importante quando se trata de *expertise*: a dimensão tempo. Saber se um especialista explorou ou 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.

#### Modelos de reconhecimento de entidades, correlação e análise temporal

Com o objetivo de combinar novas 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 PPGEGC da UFSC na área de Engenharia do Conhecimento. O trabalho de Ceci (2010) com um modelo para extração de entidades, importante passo na identificação dos termos de *expertise* contidos nos documentos ligados aos especialistas. O modelo de Gonçalves (2006), que abordou a correlação e a associação de entidades para extração do conhecimento. E, por fim, o modelo de Bovo (2011), com o processo de análise temporal, elemento de destaque não explorado nos trabalhos selecionados da revisão integrativa.

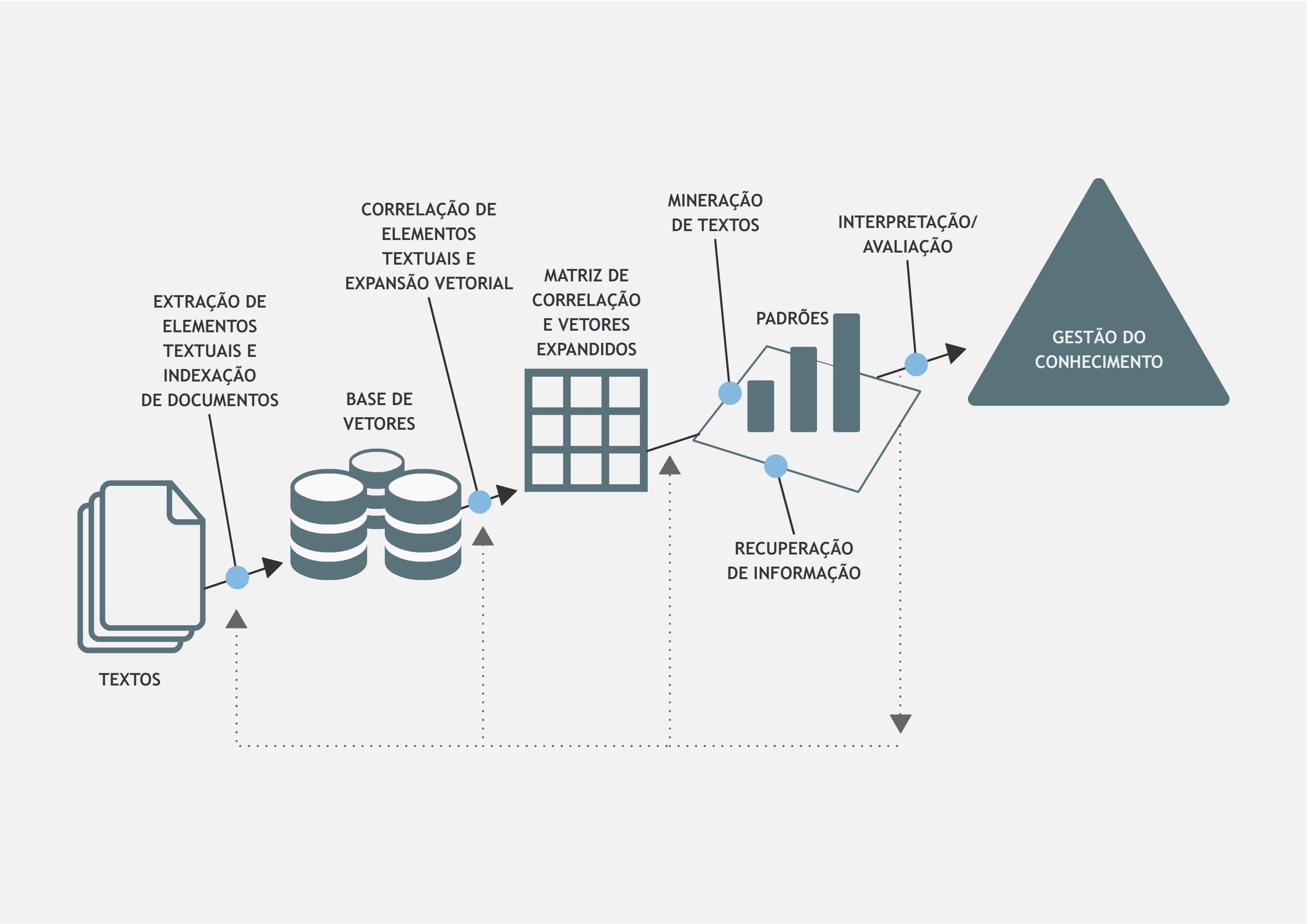
O modelo ilustrado na Figura 10, proposto por Ceci (2010), possui três etapas: reconhecimento de entidades da coleção de documentos por meio de clusterização; validação das entidades recuperadas com base nos termos da base colaborativa da Wikipédia; e classificação das entidades e adição das instâncias encontradas na ontologia de domínio da organização por meio de utilização de dicionário léxico, expressões regulares e método estatístico de identificação de instâncias.

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

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

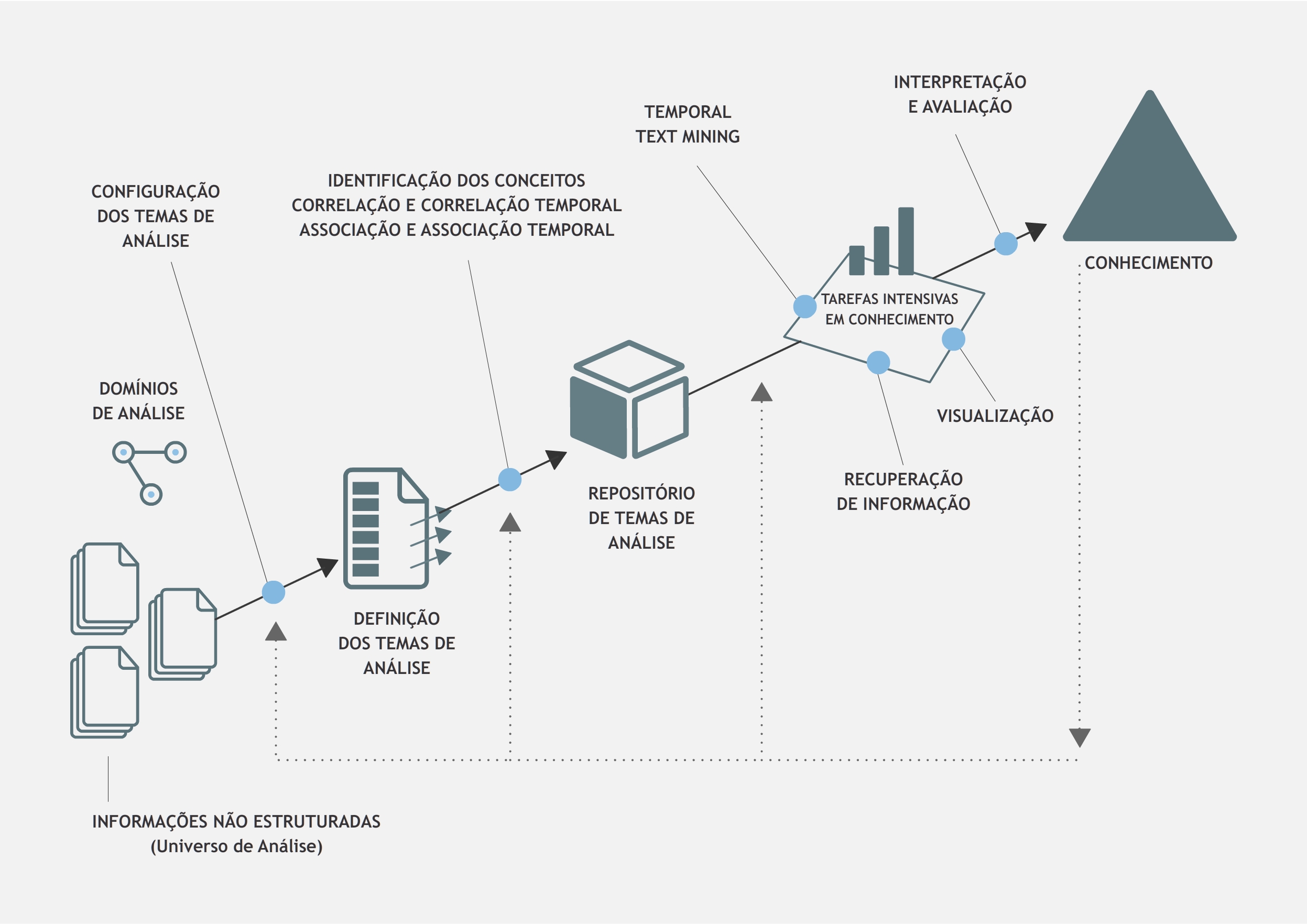
O modelo de Gonçalves (2006), apresentado na Figura 11, aborda a extração de elementos textuais como entidades, conceitos e termos, e os relaciona por meio de mapeamento de espaço vetorial. Seu modelo descreve cinco fases da extração de elementos textuais até a visualização dessas informações para apoio 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 apresenta a fase de visualização dos padrões produzidos, que resultam em mapas de conhecimento e são apresentados de forma gráfica para que ocorra *insights* em seus observadores.

Figura – Modelo de mineração de textos voltados a aplicações de EC e GC de Gonçalves.

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

O terceiro modelo, de Bovo (2011), apresentado na Figura 12 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. Seu modelo 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, 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, por fim, de tarefas intensivas em conhecimento, prevê a utilização do repositório de temas de análise para a descoberta de conhecimento.

Figura – Modelo de *Temporal Knowledge Discovery in Texts* de Bolvo.



Fonte: Adaptado de Bolvo (2011, p. 67).

# PROCEDIMENTOS METODOLÓGICOS

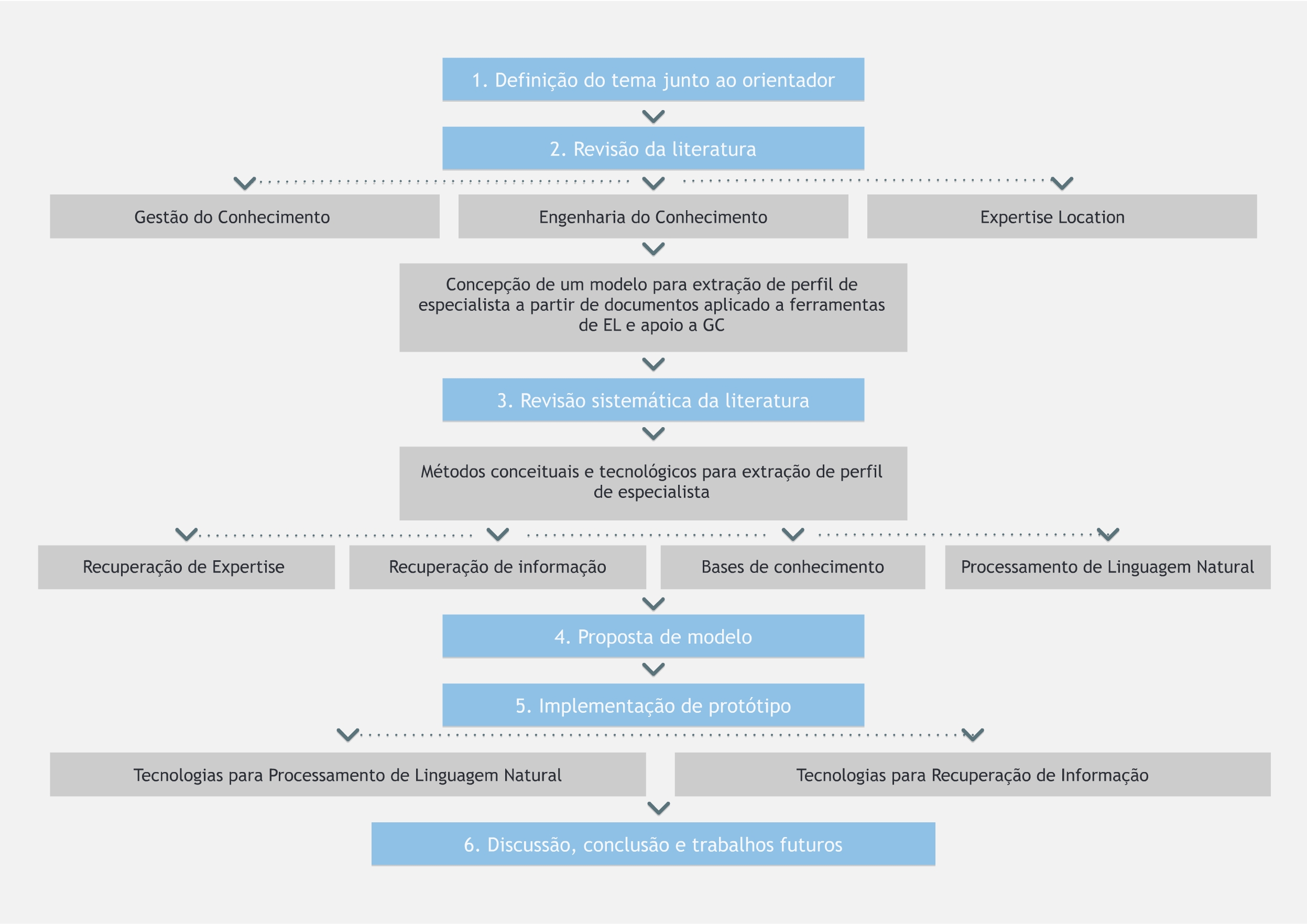
A ciência da forma que conhecemos hoje é fruto da necessidade humana de desvendar e entender o funcionamento da natureza. Várias civilizações como a chinesa, a grega, a indiana e a europeia contribuíram, por meio de realizações científicas e tecnológicas, para o nascimento da chamada “ciência moderna” (KNELLER, 1980). Um dos pilares dessa ciência moderna é o conhecimento racional, obtido por meio da realização de procedimentos metódicos e metodológicos no qual permitem que todo conhecimento novo possa ser verificado, experimentado, dando origem ao chamado “conhecimento cientifico” (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 subjetividade dos pesquisadores e é construído com base em visões de mundo do homem, sendo que asua visão gera verdades da sua própria realidade e da sua concepção filosófica (MORGAN, 1980; CRESWELL, 2010).

Essa contextualização 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. Assim, é necessário apresentar os procedimentos adotados para a concepção e a implementação, em forma de protótipo, do método de extração de perfil de especialista a partir de documentos.

Contudo, antes de se apresentar esses procedimentos, 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, de modo a projetar artefatos tecnológicos à luz do conhecimento cientifico (FREITAS et al, 2012). Dessa forma, a natureza desta esta pesquisa se enquadra em aplicada por prover uma aplicação prática e dirigida à solução de problemas específicos (SILVA; MENEZES, 2005). Para embasamento e exploração do estado da arte, adotou-se a revisão sistemática da literatura que é detalhada na seção 3.1. Mais informações sobre o procedimento da revisão sistemática da literatura podem ser observadas na seção 2.1. A Figura 13 apresenta os procedimentos metodológicos para se alcançar os objetivos desta pesquisa.

Figura – Procedimentos metodológicos para desenvolvimento da pesquisa.



Fonte: Elaboração própria.

## Execução do método de revisão integrativa

A primeira etapa para a realização de uma revisão integrativa é a identificação do tema e a seleção da questão de pesquisa. Essa é a etapa mais importante da revisão integrativa e é a etapa que guia todas as etapas subsequentes. Além da pergunta de pesquisa bem definida, também há a necessidade da especificação das palavras-chaves 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 2.3.

### 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 (MARWICK, 2001; BECERRA-FERNANDEZ, 2006; BALOG et al, 2012) 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 vários motivos, desde a desmotivação pela morosidade na atualização dos dados em um sistema de informação ou até por uma falta de políticas que incentivem os colaboradores a preencherem seus dados e demonstrem a real utilização dessas informações pela organização ou pelos próprios colaboradores. Contudo, existe espaço para que esse preenchimento do perfil profissional possa ocorrer de forma semiautomática, o que permitiria uma simplificação no processo de atualização desses conteúdos e uma possível maximização do uso de ferramentas de *Expertise Location* (MARWICK, 2001; BECERRA-FERNANDEZ, 2006; BALOG et al, 2012).

A partir desse contexto como tema de pesquisa, a pergunta que sintetiza esse problema e guia o desenvolvimento desse 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?”.*

Esse tema e essa pergunta de pesquisa, no contexto de ferramentas de *Expertise Location,* associado com a Gestão do Conhecimento e com a Engenharia do Conhecimento, foram convertidos em descritores de palavras-chaves para busca de documentos na base de dados da Scopus[[8]](#footnote-8). A base Scopus foi selecionada devido à sua grande relevância acadêmica, por conter um expressivo número de *jornals*, por ser multidisciplinar, por possuir um vocabulário controlado e por possuir um conteúdo mais abrangente do que a base de dados Web of Science (UNIVERSITY OF WASHINGTON, 2015; CHADEGANI et al, 2013; KUMAR, 2013).

Para a conversão dos descritores em palavras-chaves, 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-chaves do Google[[9]](#footnote-9), que possibilita a identificação de termos correlatos e o acesso às estatísticas sobre as palavras-chaves buscadas. Apesar do uso da ferramenta do Planejador de Palavras-chaves do Google ser voltado para a realização de campanhas no AdWords[[10]](#footnote-10) (i.e. realização de publicidade), essa ferramenta permite uma ampla visão sobre termos similares, além de permitir a visão de seu impacto nas buscas dos usuários na utilização da internet, o que comporta uma definição mais rica e abrangente dos descritores em palavras-chaves (e seus termos) que foram utilizados na busca da base de dados da Scopus.

O Quadro 6 a seguir, apresenta a lista dessas palavras-chaves, conforme o tema, a questão de pesquisa, as traduções em inglês e o resultado da expansão das palavras-chaves com a utilização da ferramenta Planejador de Palavras-chaves do Google.

Quadro – Termos correlatos das palavras-chaves da revisão integrativa da literatura.

| **Palavra-chave** | **Palavra-chave em inglês** | **Termos correlatos** |
| --- | --- | --- |
| **Perfil profissional** | **Professional profile** | * Expert Data * Expert Identification * Expert Profile * Personal Data * Personal Profile * Professional Data * Professional Resume * Resume Builder * Resume Building * Resume Writer |
| **Localização de expertise[[11]](#footnote-11)** | **Expertise Location** | * 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 |
| **Gestão do conhecimento** | **Knowledge Management** | * Knowledge Base System * Knowledge Based Systems * Knowledge Management Information Systems * Knowledge Managment System * Knowledge Systems |
| **Engenharia do Conhecimento** | **Knowledge Engineering** | * 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: elaboração própria.

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

* busca pelo grupo de palavras-chaves e seus termos correlatos de “Perfil profissional” e “*Expertise Location*” e “Gestão do Conhecimento” e “Engenharia do Conhecimento”;
* busca pelo grupo de palavras-chaves e seus termos correlatos de “Perfil profissional” e “*Expertise Location*”;
* busca pelo grupo de palavras-chaves e seus termos correlatos de “Perfil profissional” e “Gestão do Conhecimento”;
* busca pelo grupo de palavras-chaves e seus termos correlatos de “Perfil profissional” e “Engenharia do Conhecimento”;
* combinação da lista de resultados e remoção de redundâncias.

Observa-se que a prioridade em relação às palavras-chaves e termos correlatos buscados é dada para o perfil profissional, 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 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 campos para realização dessa busca por documentos seriam por meio dos campos de título, de palavras-chaves e de resumo da base de dados da Scopus.

### 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 – Total de estudos encontrados na base Scopus a partir das palavras-chaves de pesquisa.

|  |  |
| --- | --- |
| **Grupos de termos de busca** | **Estudos encontrados na Scopus** |
| “Perfil profissional” e “*Expertise Location*” e “Gestão do Conhecimento” e “Engenharia do Conhecimento” | **2** |
| “Perfil profissional” e “*Expertise Location*” | **20** |
| “Perfil profissional” e “Gestão do Conhecimento” | **52** |
| “Perfil profissional” e “Engenharia do Conhecimento” | **49** |
| Total de documentos sem redundância | **70** |

Fonte: elaboração própria.

A partir do resultado da lista de estudos encontrados, definiu-se a utilização do critério de exclusão de estudos não citados por outros trabalhos. Isto é, excluir 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 – 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 |
| 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 |
| 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 |
| 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., Djaghloul 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., Tsiporkova 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 |
| 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 |
| 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 |

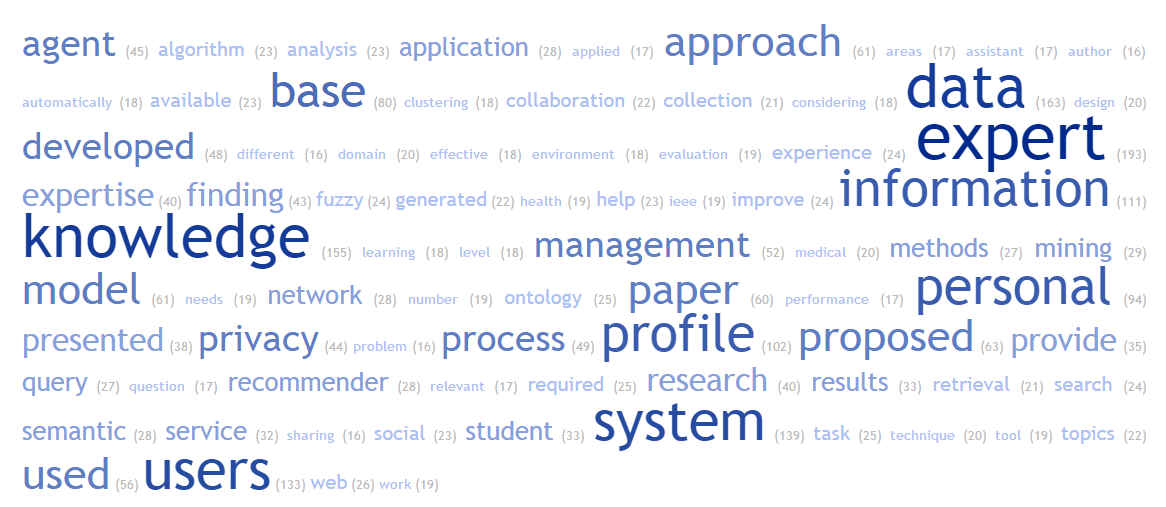
Fonte: elaboração própria.

### Identificação dos estudos pré-selecionados e selecionados

A combinação das palavras-chaves e termos correlatos de Perfil Profissional*, Expertise Location*, Gestão do Conhecimento e Engenharia do Conhecimento já se demonstrava aderente, com base nos títulos e palavras-chaves encontradas nos estudos na busca da base de dados Scopus, ao objeto de pesquisa desse trabalho para responder a pergunta de como gerar perfis profissionais de forma semiautomática.

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 *tag cloud*. Essa técnica de *tag cloud* possibilitou a representação visual da importância das palavras encontradas em um documento textual com base na frequência das ocorrências. A Figura 14, a seguir, traz a visão dessa representação com base nas palavras-chaves e resumos encontrados nos estudos contidos no Quadro 8 da seção anterior.

Figura – *Tag cloud* sobre as palavras-chaves e resumos dos estudos da revisão integrativa.

 Fonte: elaboração própria com uso da ferramenta TagCrowd disponível em < http://tagcrowd.com/ >.

Pôde-se inferir, então, com base na experiência dos autores, que as palavras *data*, *expert*, *information*, *personal* e *profile* poderiam indicar ligação com a palavra-chave de pesquisa Perfil Profissional; as palavras *expertise e finding* com *Expertise Location*; as palavras *knowledge, management, system,* e *approach* com Gestão do Conhecimento; e, por fim, as palavras-chaves *semantic, mining, fuzzy, retieval, search, task* e *ontology* com Engenharia do Conhecimento. Essa ligação indicaria uma notável relevância dos estudos encontrados na Scopus, ou, ao menos, parte desses estudos, com o objetivo da revisão sistemática a partir do tema e da pergunta de pesquisa.

A segunda tarefa realizada nessa etapa, agora de acordo com a identificação dos estudos pré-selecionados e selecionados à luz do método de revisão integrativa, foi realizar a leitura criteriosa dos títulos, resumos, palavras-chave e conteúdo completo, quando necessário, a partir dos arquivos originais disponíveis. Cabe lembrar que a base de dados Scopus não mantém todos os estudos indexados disponíveis para download em seu site. Assim, foi realizada a busca na internet dos estudos resultantes da etapa anterior. Nesse processo, 7 estudos 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, dos quais os autores não conseguiram acesso. Além de ser um conjunto pequeno de estudos, esses trabalhos possuíam pouca ou nenhuma citação, indicando pouca 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 desse trabalho, principalmente ao ajudar a responder a pergunta de como gerar perfis profissionais de forma semiautomática 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 – 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 |
| 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: elaboração própria.

### 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. Essa leitura, possibilitou a identificação de categorias que para a definição da matriz de síntese que é recomendada para a extração e organização dos dados dos estudos selecionados, de acordo com o método de revisão integrativa adotado neste trabalho.

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 de síntese é elaborada a partir dos conhecimentos e do senso crítico dos autores da revisão, tem por objetivo proteger os pesquisadores de erros durante a análise dos estudos selecionados e serve de ferramenta de interpretação e construção da redação da revisão integrativa para os pesquisadores (BOTELHO; CUNHA; MACEDO, 2011).

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

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

| Categoria | Objetivo da categoria | Estudos aderentes |
| --- | --- | --- |
| Conceitos sobre *Expertise Location,* Gestão do Conhecimento ou Engenharia do Conhecimento | Identificar estudos que apoiem a *Expertise Location,* e Gestão do Conhecimento ou Engenharia do Conhecimento explicitamente citem seus conceitos | #1, #3, #7, #8, #9, #44 |
| Exploração de documentos não estruturados para utilização no perfil profissional | Identificar estudos que trabalham com a exploração de dados e demais informações não estruturadas para a criação de um perfil profissional. | #1, #3, #7, #9, #11, #12, #15 |
| Métodos conceituais e tecnológicos para criação do perfil profissional | Identificar estudos que apresentem métodos dos quais possibilitem a geração de perfil profissional e que discutam essas estratégias | #1, #3, #7, #8, #9, #11, #12, #15, #28, #44 |
| Definições de atributos para constituição do perfil profissional | Identificar estudos que contribuem para a especificação de atributos e de informações que compõe uma definição de perfil profissional | #7, #8, #9, #12, #15, #18, #28, #44 |
| Utilização de ontologias ou taxonomias para classificação do conhecimento | Identificar estudos que contribuem 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 de tecnologias para a geração de perfil profissional | #9, #12, #15, #18 |

Fonte: elaboração própria.

# APRESENTAÇÃO DOS RESULTADOS

Este capítulo tem o intuito de apresentar os resultados gerados com base no objetivo geral desta pesquisa, que é propor um modelo de extração de perfis de especialistas a partir de documentos não estruturados, isto é, em linguagem natural, com vistas a apoiar a Gestão do Conhecimento no âmbito de ferramentas para a *Expertise Location*. Por sua vez, os objetivos específicos visam identificar as abordagens atuais que possibilitem essa extração, definir um modelo de extração de perfis de especialistas e desenvolver um protótipo funcional para se verificar a viabilidade do modelo proposto.

A identificação das abordagens atuais que possibilitem essa extração identificadas na literatura são descritas na seção 2.3. Esse foi o primeiro resultado, gerado para cumprir um dos objetivos da pesquisa, que relata os métodos conceituais e tecnológicos observados na literatura e combina os temas de Gestão do Conhecimento, Engenharia do Conhecimento, *Expertise Location* e extração de perfil de especialista. O conteúdo elaborado com a identificação dessas abordagens foi realizado por meio de uma revisão sistemática e metodológica denominada revisão integrativa, cujo método é apresentado na seção 2.1.1 e a sua execução, nesta pesquisa, na seção 3.1.

Os modelos observados na literatura, conceituais e tecnológicos, que abordam processos, técnicas, algoritmos e tecnologias, e que fornecem subsídios para a definição do modelo de extração de perfil de especialista, são apresentados neste capítulo e materializados em forma de software no protótipo funcional. Esses modelos, principalmente pautados nos trabalhos realizados por Becerra-Fernandez (2006), Balog e Rijke (2007), Yang e Huh (2008) e Li, Liu e Li (2011) e complementados pelos trabalhos de Gonçalves (2006), Ceci (2010) e Bovo (2011), permitem a criação de uma nova abordagem para se tratar da extração de *expertise* de documentos, e vai além na proposição de uma nova abordagem que combina a dimensão tempo nessa extração. Esse modelo, resultado desta pesquisa, é apresentado na próxima seção.

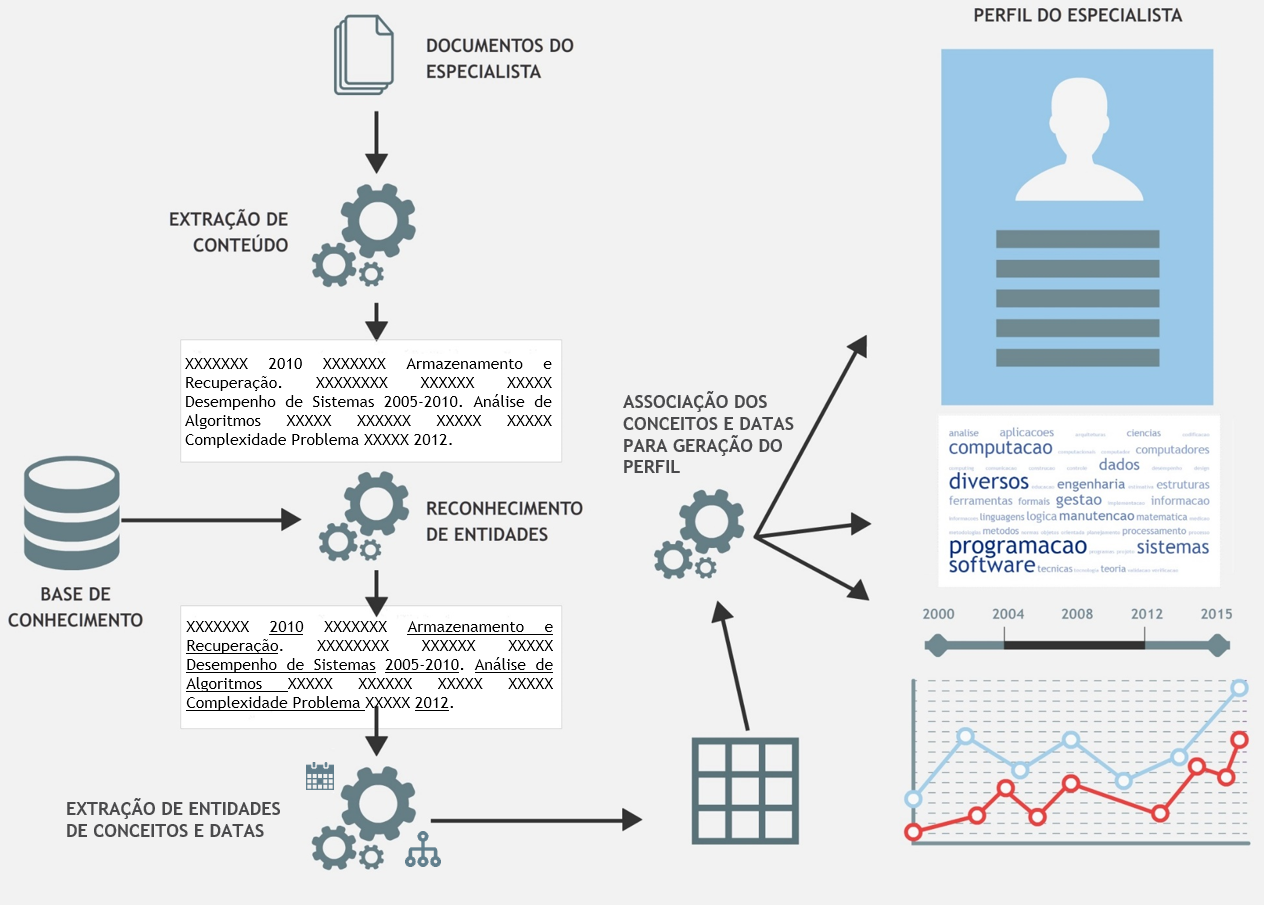
## Modelo de extração de perfil de especialista proposto

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 fluído nos ambientes organizacionais. Com a sua crescente importância, torna-se necessário gerenciar esse valioso recurso, e é nesse contexto que entra a Gestão do Conhecimento. Uma importante ferramenta na Gestão do Conhecimento, que apoia o gerenciamento dos ativos de conhecimento, são os sistemas de *Expertise Location.* Essas ferramentas auxiliam a identificação e o compartilhamento de conhecimento, seja o conhecimento tácito ou o conhecimento explícito, por meio da integração de pessoas que precisam de determinados conhecimentos com as pessoas que possuem esses conhecimentos.

Evidentemente, a construção das ferramentas de *Expertise Location* é observada há muitas décadas na literatura, conforme apresentado nas seções 1 e 2.2.4 deste trabalho, antes mesmo das iniciativas da Gestão do Conhecimento. Contudo, essas ferramentas estão ganhando cada vez mais recursos para melhorá-las. Um desses recursos, que tem ganhado bastante destaque 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 especialista a partir dos documentos elaborados por esses especialistas.

As diferentes de técnicas, métodos, processos e tecnologias identificadas na literatura, por meio da revisão integrativa, são detalhas na seção 2.3.2. Observa-se que o Processamento de Linguagem Natural representa um papel importante nas soluções detalhadas nos trabalhos encontrados de Becerra-Fernandez (2006) e Balog e Rijke (2007) resultantes da revisão integrativa. Por isso, torna-se, também, um elemento essencial no modelo proposto e apresentado na Figura 15 a seguir. O PLN envolve as técnicas de tokenização, normalização, remoção de *stop-words*, *lemmatization*, *stemming, Part-Of-Speech Tagging* e *Named Entity Recognition* que também são adotados nesse modelo. Outras técnicas de reconhecimento de entidades dos trabalhos de Gonçalves (2006) e Ceci (2010), correlação de Gonçalves (2006) e Bovo (2011), e associação temporal de Bovo (2011) são consideradas nesse modelo. A seguir se ilustra o modelo e se detalha cada uma de suas etapas para a extração do perfil de especialista.

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



Fonte: elaboração própria.

* **Extração do conteúdo**. Essa 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 enumeras bibliotecas de software para essa tarefa. Os conteúdos de meta-informação como autores, data de criação e de modificação são extraídos para serem utilizados posteriormente na etapa de associação de conceitos para geração do perfil. Já o conteúdo dos documentos, isto é, o conteúdo presente internamente nos documentos, são extraídos e encaminhados para a etapa de reconhecimento de entidades.
* **Reconhecimento de entidades.** A etapa de reconhecimento de entidades (NER) é fundamental para que se identifique os conceitos, pessoas e dadas registrados nos documentos. Esse processo de NER deve ser realizado com a utilização de bibliotecas de software que permitem o Processamento de Linguagem Natural para a identificação de sentenças e para a identificação entidades candidatas. Após a identificação das entidades candidatas, deve ser realizada a validação com os elementos registrados na base de conhecimento. Essa base de conhecimento, ou bases de conhecimento, devem manter os dicionários léxicos, taxonomias ou tesauros de pessoas e conceitos para a validação das entidades candidatas encontras no processo de reconhecimento. Atualmente, o formato de taxonomias em SKOS (Simple Knowledge Organization System) tem sido bastante utilizado. Um exemplo de taxonomia SKOS é a estrutura de conceitos mantidas pela DBPedia, no qual toda a base colaborativa de artigos presentes na Wikipédia é mantida e disponibilizada para a comunidade. Assim, recomenda-se a utilização da SKOS da DBPedia para a utilização quando não se tem o conhecimento da organização mapeamento nessas bases de conhecimento, e recomenda-se a criação de uma SKOS para mapear o conhecimento envolvido nas tarefas realizadas pelos especialistas da organização.
* **Extração de entidades de conceitos e datas.** Após a validação das entidades descobertas de conceitos, essa etapa de extração de entidades de conceitos e datas visa a extração de entidades do tipo data (e.g. 2014) ou períodos (e.g. 2005-2012 ou 2005 a 2012) e o armazenamento dessas entidades de conceitos e datas em tabelas para serem utilizadas no processo de associação dos conceitos e datas. Ao final dessa etapa, registra-se as informações dos meta-dados do documento, das sentenças presentes nos documentos, das entidades de conceitos e das entidades de datas em forma estruturada para a realização de processos de combinação desses dados.
* **Associação dos conceitos e datas para geração do perfil.** Com as informações registradas de forma estrutura dos conceitos e datas encontrados no processo de extração das entidades, mais as informações dos meta-dados e as sentenças dos documentos, essa etapa tem por objetivo associar esses insumos para a geração do perfil do especialista. Os conceitos encontrados são associados, então, às datas ou aos períodos encontrados no documento com base na janela da sentença. Assim, são associadas todas as datas e conceitos que estiverem diretamente na mesma sentença com peso máximo. Já os conceitos que não forem associados a datas e a períodos na mesma sentença, serão associados a datas presentes em sentenças mais próximas que não estiverem com entidades de datas e conceitos associados. Entretanto, a distância do conceito em relação à data ou ao período mais próximo terá o seu peso dividido pelo número de sentenças percorridos.
* **Visualização do perfil do especialista.** A última etapa do modelo proposto tem por objetivo a apresentação das entidades de conceitos e datas no formato gráfico ao usuário, de modo que possibilite a identificação desses conceitos com as suas respectivas datas e respectivas frequências. Esse processo deve considerar todas as entidades e datas descobertas nos vários documentos associados e fornecidos pelos especialistas. Dois formatos são sugeridos para a apresentação desses conteúdos: gráficos de *tag cloud* com a possibilidade de navegação dos anos; e gráficos de linha com os conceitos e suas frequências por ano.

Com base nessa apresentação das etapas, apresenta-se o detalhamento de cada uma delas nas seções a seguir.

### Extração do conteúdo

### Reconhecimento de entidades

### Extração de entidades de conceitos e datas

### Associação dos conceitos e datas para geração do perfil

### Visualização do perfil do especialista

## Protótipo baseamento no modelo proposto

## Demonstração de viabilidade

# CONCLUSÕES E TRABALHOS FUTUROS

## CONSIDERAÇÕES FINAIS

## RECOMENDAÇÕES PARA TRABALHOS FUTUROS

# REFERÊNCIAS

ADOLFO, L. B. Uma Ontologia de Apoio a Classificação de Processos Judiciais. Dissertação (Mestrado em Engenharia de Gestão do Conhecimento), 2013.

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

ALAVI, M., LEIDNER, D. E. Knowledge Management systems: Emerging views and practices from the field. In International Conference on System Sciences, 239. 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, 61( 3 ), 315–343. 1991.

ANAND, A; SINGH, M. D. Understanding Knowledge Management: A Literature Review", International Journal of Engineering Science and Technology, Vol. 3 No. 2, pp. 926- 939. 2011.

APO - Asian Productivity Organization. Knowledge Management Tools and Techniques Manual. Japan: Hirakawacho, Chiyoda-ku, 2010.

APPOLINÁRIO, 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 02 abr. 2015.

BALOG, K, et al. Expertise Retrieval. Foundations and Trends in Information Retrieval, p. 127-256. 2012.

BECERRA-FERNANDEZ, I. Searching for experts on the web: a review of contemporary expertise locator systems. ACM Trans. on Internet Technology, vol. 6, no. 4, pp. 333-355. 2006.

BECERRA-FERNANDEZ, I. The role of artificial intelligence technologies in the implementation of PeopleFinder knowledge management systems. Knowledge-Based Systems, 13(5):315–320, 2000.

BECKMAN, T. The current state of Knowledge Management. In Liebowitz, J. (Eds), Knowledge Management Handbook, CRC Press, Boca Raton, FL. 1999.

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

BENTO, A. Como fazer uma revisão da literatura: Considerações teóricas epráticas. Revista JA (Associação Académica da Universidade da Madeira), nº 65, ano VII (pp. 42-44). ISSN: 1647-8975. 2012.

BERNARDO, W. M.; NOBRE, M. R. C.; JATENE, Fábio Biscegli. A prática clínica baseada em evidências: parte II-buscando as evidências em fontes de informação. Rev Assoc Med Bras, v. 50, n. 1, p. 104-8, 2004.

BOTELHO, L. L. R.; CUNHA, C. C. A.; MACEDO, M. O método da revisão integrativa nosestudos organizacionais. Gestão e Sociedade, 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. Tese (Doutorado em Engenharia de Produção), 2011.

BROWN, J.S.; DUGUID, P. Organizing Knowledge, California Management Review vol. 40, no.3. 1998.

BUKOWITZ, W.; WILLIAMS, R. The Knowledge Management field book. London: Financial Times/Prentice Hall. 2000.

CECI, F. Um Modelo semi-automático para a construção e manutenção de ontologias a partir de bases de documentos não estruturados. 131 p. Dissertação (Mestrado em Engenharia de Gestão do Conhecimento), 2010.

CHADEGANI et al. A Comparison between Two Main Academic Literature Collections: Web of Science and Scopus Databases. Asian Social Science. ISSN 1911-2017. 2013.

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

CHOI, B.; LEE, H. Knowledge management strategy and its link to knowledge creation process. Expert Systems Appl. 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.

CORDEIRO, A. M. et al. Revisão sistemática: uma revisão narrativa. Revista do Colégio Brasileiro de Corurgiões, 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.

DAGHFOUS, A.; KAH, M. M. O. Knowledge Management implementation in SMEs: a framework and a case illustration. Journal of Information and Knowledge Management,5(2), 107-15. 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. Harvard Business School Press, Boston, MA. 1998.

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.

GLASS, G. V. Primary, secondary, and meta-analysis of research.Educational researcher, p. 3-8, 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. 196 p. Tese (Doutorado em Engenharia de Produção) ênfase em Inteligência Aplicada - Programa de Pós-Graduação em Engenharia de Produção, UFSC, Florianópolis, 2006.

GRANT, R. M. Toward a knowledge‐based theory of the firm. Strategic management journal, v. 17, n. S2, p. 109-122, 1996.

GREENHALGH, T. How to read a paper: Papers that summarise other papers (systematic reviews and meta-analyses). Bmj, v. 315, n. 7109, p. 672-675, 1997.

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

GRUBER, T. A translation approach to portable ontology specification. Knowledge Acquisition, 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.

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.

KNELLER, G. F. A ciência como atividade humana. Rio de Janeiro: Zahr, 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, 5(1), 29-40. 1998.

KUMAR, K. Comparative Analysis of Search Features of Scopus and Web of Science. Department of Uhrary& Information Science Sri Venkateswara University. 2013

LEE, K. C. et al. Knowledge Management: measuring Knowledge Management performance. Information and Management, 42(3), 469-82. 2005.

LOH, S. Abordagem Baseada em Conceitos para Descoberta de Conhecimentos em Textos. Tese (Doutorado em Ciência da Computação). Programa de Pós-graduação em Computação. UFGRS, Porto Alegre, 2001.

MARTINEZ, A.et al. Las categorías o facetas fundamentales: uma 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 40 (4) , p.814-830. 2001.

MAYBURY, M. Expert Finding Systems. MITRE - Center for Integrated Intelligence Systems. Bedford, Massachusetts. 2006.

MAYER, M.; ZACK, M. The design and implementation of information Products. Sloan Management Review, 37 (3), 45 - 59. 1996.

MC ELROY; M. The knowledge life cycle. Presented at the ICM conference on Knowledge Management. Miami, Florida. 1999.

MENDES, K. D. S. et al. Revisão integrativa: método de pesquisa para a incorporação de evidências na saúde e na enfermagem. Texto and Contexto Enfermagem, v. 17, n. 4, p. 758, 2008.

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

NAIR, P.; PRAKASH, K. Knowledge Management: Facilitator’s Guide. APO: Tokyo, 2009.

NASCIMENTO, R. N. do; SELL, Denilson; TODESCO, José L; BEPPLER, Fabiano; PACHECO, Roberto. Classificação semântica de informações não estruturadas sobre colaboradores. IV Congresso Internacional do Conhecimento e Inovação. 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 university press, 1995

NORTH, K.; RIVAS, R. Gestión del conocimiento. Una guía práctica hacia la empresa inteligente. p. 364. 2008.

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

RAO, M. Knowledge Management Tools and Techniques. 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, mai./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.

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, Vol. 5 Iss: 4, pp.311 – 321. 2001.

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.

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

SUN, Z.; HAO, G. HSM: a hierarchical spiral model for knowledge management. In: Proceedings of the 2nd International Conference on Information Management and Business (IMB2006), Sydney, 13-16 February, 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, v. 40, n. 3, p. 55–79, 1998.

TEECE, D. J.; PISANO, G.; SHUEN, A. Dynamic capabilities and strategic management. Strategic Management Journal, 1997.

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 05 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, CRC Press, Boca Raton, FL. 1997.

VENZIN, M.; KROGH, G; ROOS, J. Future Research into Knowledge Manegement, In Knowing in Firms, 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, pp.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, vol. 30, no. 4, 1997.

WIIG, K. M. Knowledge Management Foundations: Thinking about Thinking – How People and Organizations Create, Represent and use Knowledge, Schema Press, Arlington, TX. 1993.

WONG, K.Y.; ASPINWALL, E. Characterizing Knowledge Management in the small business environment. Journal of Knowledge Management, 8(3), 44-61. 2004.

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

1. O Dr. Denilson Sell, orientador desse trabalho, defendeu sua tese no programa de Engenha de Produção, mas a contribuição de sua tese está relacionada às áreas de Engenharia e Gestão do Conhecimento. [↑](#footnote-ref-1)
2. O Dr. Alexandre Golçalves, professor da disciplina de Métodos e Técnicas de Engenharia do Conhecimento no PPEGC, defendeu sua tese no programa de Engenha de Produção, mas a contribuição de sua tese também está relacionada às áreas de Engenharia e Gestão do Conhecimento. [↑](#footnote-ref-2)
3. http://priberam.pt/dlpo/expertise [↑](#footnote-ref-3)
4. http://www.significados.com.br/expertise/ [↑](#footnote-ref-4)
5. https://translate.google.com.br/#auto/pt/expertise [↑](#footnote-ref-5)
6. http://dicionariodoaurelio.com/ [↑](#footnote-ref-6)
7. http://michaelis.uol.com.br/moderno/portugues [↑](#footnote-ref-7)
8. Scopus: um banco de dados de resumos e citações de artigos para jornais/revistas acadêmicos. Disponível em <http://www.scopus.com/>. [↑](#footnote-ref-8)
9. Planejador de palavras-chaves do Google: Planeje suas campanhas da Rede de Pesquisa e saiba o que seus clientes procuram. Disponível em < https://adwords.google.com/KeywordPlanner >. [↑](#footnote-ref-9)
10. Google AdWords: Publicidade online. Disponível em < https://www.google.com.br/adwords/ >. [↑](#footnote-ref-10)
11. Foram encontradas duas possibilidades de tradução para o *Expertise Location*: Localização de expertise ou Locação de perícias. Contudo, optou-se por utilizar o termo em inglês na redação deste trabalho por não se encontrar fortes referências de trabalhos em português com a utilização dessas traduções. [↑](#footnote-ref-11)