



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CAMPUS FLORIANÓPOLIS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA E GESTÃO DO
CONHECIMENTO

MARINA CARRADORE SÉRGIO

**MODELO DE AVALIAÇÃO DE POTENCIAIS IDEIAS ALINHADAS
AO CONTEXTO ORGANIZACIONAL**

FLORIANÓPOLIS

2020

Marina Carradore Sérgio

**MODELO DE AVALIAÇÃO DE POTENCIAIS IDEIAS ALINHADAS AO
CONTEXTO ORGANIZACIONAL**

Tese submetida ao Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento da Universidade Federal de Santa Catarina para a obtenção do título de Doutor em Engenharia e Gestão do Conhecimento. Área de concentração: Engenharia do Conhecimento. Linha de pesquisa: Teoria e Prática em Engenharia do Conhecimento.

Orientador: Prof. Dr. Alexandre Leopoldo Gonçalves.

Coorientador: Prof. Dr. João Artur de Souza.

Florianópolis

2020

Ficha de identificação da obra

Sérgio, Marina Carradore
MODELO DE AVALIAÇÃO DE POTENCIAIS IDEIAS ALINHADAS AO
CONTEXTO ORGANIZACIONAL / Marina Carradore Sérgio ;
orientador, Alexandre Leopoldo Gonçalves, coorientador,
João Artur de Souza, 2020.
242 p.

Tese (doutorado) - Universidade Federal de Santa
Catarina, Centro Tecnológico, Programa de Pós-Graduação em
Engenharia e Gestão do Conhecimento, Florianópolis, 2020.

Inclui referências.

1. Engenharia e Gestão do Conhecimento. 2. Gestão de
Ideias, Representação de Conhecimento, Mineração de Texto.
I. Leopoldo Gonçalves, Alexandre . II. Souza, João Artur
de. III. Universidade Federal de Santa Catarina. Programa
de Pós-Graduação em Engenharia e Gestão do Conhecimento. IV.
Título.]

Marina Carradore Sérgio

Modelo de avaliação de potenciais ideias alinhadas ao contexto organizacional

O presente trabalho em nível de doutorado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof. Denilson Sell, Dr.

Universidade Federal de Santa Catarina

Prof. Patricia de Sá Freire, Dra.

Universidade Federal de Santa Catarina

Prof. Carlos Roberto de Rolt, Dr.

Universidade do Estado de Santa Catarina

Prof. Paulo Henrique de Souza Bermejo, Dr.

Universidade de Brasília

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de doutor em Engenharia e Gestão do Conhecimento.

Prof. Dr. Roberto Carlos dos Santos Pacheco
Coordenador do Programa

Prof. Dr. Alexandre Leopoldo Gonçalves
Orientador

Florianópolis, 22 de abril de 2020.

Aos meus pais João e Maristela, e familiares pelo incentivo, confiança e compreensão.

Ao meu companheiro Gustavo, que sempre esteve comigo no decorrer desta caminhada, me dando estímulo, carinho e atenção. Sem você essa jornada seria difícil.

AGRADECIMENTOS

Para o desenvolvimento desta tese foi necessário muito empenho e dedicação. Contudo, o trabalho não seria possível sem a participação direta ou indireta de algumas pessoas, às quais gostaria de expressar minha sincera gratidão.

Primeiramente a Deus, pelo dom da vida e pela força que me concedeu nesta jornada para enfrentar as adversidades, me proporcionando ânimo e fé para nunca desistir dos meus objetivos.

Aos meus pais, João e Maristela, pelo carinho e fonte de acolhimento, por confiarem em mim em todas as circunstâncias, por serem exemplo de integridade e perseverança.

À minha família, que sempre esteve presente em minha vida, apesar da distância.

Ao meu companheiro Gustavo por todo estímulo, amor e compreensão. Também pela força e fé que depositou em mim, me fazendo crer que seria possível alcançar os meus anseios, tornando-me mais feliz e motivada. Sem ele, essa jornada seria difícil.

Ao meu orientador, professor Dr. Alexandre Leopoldo Gonçalves, meu profundo sentimento de gratidão por ter concedido a oportunidade de conviver e expandir os meus conhecimentos. Agradeço pelos momentos de incentivo, confiança, respeito, paciência, motivação, assim como pelas incansáveis e esclarecedoras conversas e, principalmente, por todo o conhecimento compartilhado. Com certeza foram momentos de muito aprendizado. Também agradeço ao professor Dr. João Artur de Souza pela atenção e coorientação neste trabalho.

Aos professores do Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento que aceitaram compor a banca examinadora, Dr. Denilson Sell e Dra. Patricia de Sá Freire, bem como ao Dr. Carlos Roberto de Rolt da Universidade do Estado de Santa Catarina e ao Dr. Paulo Henrique de Souza Bermejo da Universidade de Brasília. Agradeço também ao professor Dr. José Leomar Todesco que participou da defesa de qualificação. É uma grande honra tê-los como avaliadores deste trabalho.

A todos os professores que tive durante a minha trajetória acadêmica, os quais me ensinaram e também me inspiraram a seguir por este caminho.

Ao Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, pela oportunidade em participar desse prestigiado curso. E, à Universidade Federal de Santa Catarina, aos funcionários e professores.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de estudos durante o doutorado.

Agradeço às demais pessoas que participaram direta ou indiretamente do desenvolvimento deste trabalho. Aos amigos, o meu mais profundo e sincero sentimento de agradecimento.

“A mente que se abre a uma nova ideia, jamais volta ao seu tamanho original”.
(Albert Einstein)

RESUMO

O volume de dados textuais gerados em plataformas de gestão de ideias tem tornado a avaliação destas informações, realizada de forma manual, demorada e dispendiosa. Embora a literatura relate a importância da utilização de ferramentas de mineração de dados e texto para a avaliação de ideias, poucos trabalhos utilizam tais mecanismos para apoiar os gestores na tomada de decisão. O objetivo desta pesquisa é avaliar o potencial de implementação de ideias alinhadas ao contexto organizacional, propondo um modelo com base nestas ferramentas. Para a determinação da relevância de cada ideia foram utilizadas fontes de informação externas visando auxiliar na geração de índices, entre eles: o índice de aderência ao contexto organizacional, o índice de atualidade, o índice de originalidade e o índice de tecnologia. Estes quatro índices compõem ao final o índice de potencialidade da ideia. A representação do domínio ocorreu através de uma ontologia, validada por um conjunto de especialistas. Para a etapa de avaliação do modelo proposto foi desenvolvido um protótipo aplicado em dois cenários de estudo, bem como foram realizados testes com modelos supervisionados e elaborado um questionário avaliado por especialistas. O primeiro cenário de estudo foi composto por ideias coletadas do programa Sinapse da Inovação[®] e informações extraídas do edital de fomento para compor o contexto organizacional. Neste cenário, na etapa de mineração de texto o modelo proposto produziu uma assertividade de aproximadamente 70% na seleção de ideias implementadas e obteve 80% de assertividade na indicação de grupos de potenciais ideias quando analisados os principais agrupamentos. O segundo cenário de estudo envolveu ideias que constam na coletânea #MosquitoNão promovida pelo Ministério da Educação. Neste cenário, o objetivo da análise foi identificar as características da formação dos agrupamentos e, a partir disso, analisar os índices gerados pelo protótipo independente do contexto organizacional, enfatizando a generalidade e aplicabilidade do modelo. Com os resultados obtidos pode-se afirmar que o modelo proposto permite avaliar o potencial de implementação das ideias alinhadas às temáticas estratégicas relacionadas ao contexto organizacional. Isso ocorre em aspectos como a redução do esforço durante a avaliação de ideias, bem como a promoção de indícios sobre o potencial de implementação das mesmas, com o intuito de suportar os especialistas na avaliação e seleção de quais ideias serão colocadas em prática, otimizando os recursos organizacionais.

Palavras-chave: Gestão de Ideias. Representação de Conhecimento. Mineração de Texto.

ABSTRACT

The volume of textual data generated in idea management platforms has made the evaluation of this information, carried out manually, time-consuming and expensive. Although the literature reports the importance of using data and text mining tools for the evaluation of ideas, few studies use such mechanisms to support managers in decision making. The objective of this research is to evaluate the potential for implementing ideas aligned to the organizational context, proposing a model based on these tools. To determine the relevance of each idea, external information sources were used to assist in the generation of indexes, including: the adherence index to the organizational context, the currentness index, the originality index and the technology index. These four indices make up the idea's potentiality index at the end. The representation of the domain occurred through an ontology, validated by a group of specialists. For the evaluation stage of the proposed model, a prototype was developed applied in two study scenarios, as well as tests were carried out with supervised models and a questionnaire was evaluated by experts. The first study scenario was composed of ideas collected from the Sinapse da Inovação[®] program and information extracted from the promotion notice in order to compose the organizational context. In this scenario, during the text mining stage, the proposed model produced approximately 70% assertiveness in the selection of implemented ideas and obtained 80% assertiveness in the indication of groups of potential ideas when the main groupings were analyzed. The second study scenario involved ideas contained in the collection #MosquitoNo promoted by the Ministry of Education. In this scenario, the objective of the analysis was to identify the characteristics of the formation of clusters and, from there, to analyze the indices generated by the prototype regardless of the organizational context, emphasizing the generality and applicability of the model. With the results obtained, it can be said that the proposed model allows assessing the potential for implementing ideas aligned with strategic themes related to the organizational context. This occurs in aspects such as the reduction of effort during the evaluation of ideas, as well as the promotion of indications about the potential of their implementation. It intends to support experts in the evaluation and selection of which ideas will be put into practice, optimizing the organizational resources.

Keywords: Idea Management. Knowledge Representation. Text Mining.

LISTA DE FIGURAS

Figura 1 – Integração das áreas de concentração com o tema de pesquisa	38
Figura 2 – Benefícios da aplicação do processo de Gestão de Ideias	48
Figura 3 – Círculo de inovação: inovação apoiada por computador	50
Figura 4 – Arquitetura de sistema conceitual de um sistema de gerenciamento de ideias	51
Figura 5 - Modelo de inovação aberta.....	54
Figura 6 - Papéis da multidão.....	57
Figura 7 - Matriz termo x documento	70
Figura 8 - Matriz decomposta	70
Figura 9 – Representação da matriz decomposta	71
Figura 10 – Identificação de rótulos (<i>labels</i>).....	71
Figura 11 - DSRM.....	92
Figura 12 - Etapas da pesquisa	94
Figura 13 - Desenvolvimento da tese	95
Figura 14 – Formato de exposição das ideias no site Sinapse da Inovação®	97
Figura 15 - Método de desenvolvimento de ontologia.....	105
Figura 16 – Classes da ontologia.....	107
Figura 17 – Detalhamento das classes presentes na ontologia.....	110
Figura 18 – Pergunta de competência 1	113
Figura 19 – Pergunta de competência 2	114
Figura 20 – Pergunta de competência 3	115
Figura 21 – Pergunta de competência 4	116
Figura 22 – Pergunta de competência 5	116
Figura 23 – Pergunta de competência 6	117
Figura 24 – Pergunta de competência 7	118
Figura 25 – Pergunta de competência 8	118
Figura 26 – Pergunta de competência 9	119
Figura 27 – Pergunta de competência 10	120
Figura 28 – Pergunta de competência 11	120
Figura 29 – Pergunta de competência 12	121
Figura 30 – Pergunta de competência 13	121
Figura 31 – Pergunta de competência 14	122

Figura 32 – Pergunta de competência 15	123
Figura 33 – Modelo de avaliação de potenciais ideias	135
Figura 34 – População da base de conhecimento e seleção de ideias	137
Figura 35 – Pré-processamento de ideias	138
Figura 36 – Transformação de ideias	140
Figura 37 – Mineração de texto.....	142
Figura 38 – Fluxograma do algoritmo de agrupamento	143
Figura 39 – Nuvem de termos do vetor de potenciais áreas de investimento	150
Figura 40 – Nuvem de termos do vetor de tecnologia	150
Figura 41 – Visualização dos agrupamentos através de <i>treemap</i> para o primeiro cenário de análise	152
Figura 42 – <i>Treemap</i> para o <i>cluster</i> 18.....	153
Figura 43 – Nuvem de termos do <i>cluster</i> 18	154
Figura 44 – <i>Treemap</i> para o <i>cluster</i> 6.....	155
Figura 45 – Nuvem de termos do <i>cluster</i> 6	156
Figura 46 – <i>Treemap</i> para o <i>cluster</i> 4.....	156
Figura 47 – Nuvem de termos do <i>cluster</i> 4	158
Figura 48 – <i>Treemap</i> para o <i>cluster</i> 1.....	158
Figura 49 – Nuvem de termos do <i>cluster</i> 1	159
Figura 50 – <i>Treemap</i> para o <i>cluster</i> 15.....	160
Figura 51 – Nuvem de termos do <i>cluster</i> 15	161
Figura 52 – <i>Treemap</i> para o <i>cluster</i> 32.....	161
Figura 53 – Nuvem de termos do <i>cluster</i> 32	162
Figura 54 – <i>Treemap</i> para o <i>cluster</i> 10.....	163
Figura 55 – Nuvem de termos do <i>cluster</i> 10	164
Figura 56 – <i>Treemap</i> para o <i>cluster</i> 14.....	164
Figura 57 – Nuvem de termos do <i>cluster</i> 14	165
Figura 58 – <i>Treemap</i> para o <i>cluster</i> 30.....	166
Figura 59 – Nuvem de termos do <i>cluster</i> 30	167
Figura 60 – <i>Treemap</i> para o <i>cluster</i> 39.....	167
Figura 61 – Nuvem de termos do <i>cluster</i> 39	168
Figura 62 - Formação da rede com peso das arestas igual a 0,4	173
Figura 63 - Formação da rede com peso das arestas igual a 0,5	175
Figura 64 - Formação da rede com peso das arestas igual a 0,6	176

Figura 65 - Formação da rede com peso das arestas igual a 0,7	177
Figura 66 - Formação da rede com peso das arestas igual a 0,8	178
Figura 67 – Visualização dos agrupamentos através de <i>treemap</i> para o segundo cenário de análise	183
Figura 68 – <i>Treemap</i> para o <i>cluster</i> 17.....	184
Figura 69 – Nuvem de termos do <i>cluster</i> 17	185
Figura 70 – <i>Treemap</i> para o <i>cluster</i> 0.....	186
Figura 71 – Nuvem de termos do <i>cluster</i> 0	187
Figura 72 – <i>Treemap</i> para o <i>cluster</i> 7.....	188
Figura 73 – Nuvem de termos do <i>cluster</i> 7	189
Figura 74 – <i>Treemap</i> para o <i>cluster</i> 3.....	189
Figura 75 – Nuvem de termos do <i>cluster</i> 3	190
Figura 76 – <i>Treemap</i> para o <i>cluster</i> 6.....	191
Figura 77 – Nuvem de termos do <i>cluster</i> 6	192
Figura 78 - Matriz de confusão	207

LISTA DE QUADROS

Quadro 1 - Tecnologias e ferramentas utilizadas no domínio de Gestão de Ideias....	33
Quadro 2 - Comparação entre inovação fechada e inovação aberta.....	53
Quadro 3 - Elementos da ontologia.....	64
Quadro 4 - Tipos de ontologias	65
Quadro 5 – Pesquisa científica e pesquisa tecnológica	89
Quadro 6 – Métodos de avaliação da DSRM.....	93
Quadro 7 – Síntese da abordagem metodológica	102
Quadro 8 – Síntese da abordagem utilizada nas etapas da pesquisa	103
Quadro 9 – Perguntas de competência definidas.....	106
Quadro 10 - Descrição das classes da ontologia	107
Quadro 11 - Descrição das classes da ontologia e suas propriedades de dados.....	108
Quadro 12 - Retorno da pergunta de competência 6.....	117
Quadro 13 - Retorno da pergunta de competência 9.....	119
Quadro 14 - Retorno da pergunta de competência 11	120
Quadro 15 - Retorno da pergunta de competência 14.....	122
Quadro 16 - Atividades de impacto.....	149
Quadro 17 - Atividades de impacto.....	180

LISTA DE TABELAS

Tabela 1 - Retorno da pergunta de competência 1	113
Tabela 2 - Retorno da pergunta de competência 2	114
Tabela 3 - Retorno da pergunta de competência 3	115
Tabela 4 - Retorno da pergunta de competência 4	116
Tabela 5 - Retorno da pergunta de competência 5	117
Tabela 6 - Retorno da pergunta de competência 7	118
Tabela 7 - Retorno da pergunta de competência 8	118
Tabela 8 - Retorno da pergunta de competência 10	120
Tabela 9 - Retorno da pergunta de competência 12	121
Tabela 10 - Retorno da pergunta de competência 13	122
Tabela 11 - Retorno da pergunta de competência 15	123
Tabela 12 – Exemplo de cálculo do índice de aderência ao contexto organizacional	127
Tabela 13 – Exemplo de cálculo do índice de atualidade	128
Tabela 14 – Exemplo de cálculo do índice de tecnologia	129
Tabela 15 – Exemplo de cálculo do índice de originalidade.....	131
Tabela 16 - Vetor final da etapa de pré-processamento	139
Tabela 17 - Matriz obtida na etapa de pré-processamento	140
Tabela 18 - Distribuição das ideias coletadas por categoria	146
Tabela 19 - Distribuição das ideias coletadas por grau de maturidade	147
Tabela 20 – Ideias que compõem o <i>cluster</i> 18	153
Tabela 21 – Ideias que compõem o <i>cluster</i> 6	155
Tabela 22 – Ideias que compõem o <i>cluster</i> 4	157
Tabela 23 – Ideias que compõem o <i>cluster</i> 1	159
Tabela 24 – Ideias que compõem o <i>cluster</i> 15	160
Tabela 25 – Ideias que compõem o <i>cluster</i> 32	162
Tabela 26 – Ideias que compõem o <i>cluster</i> 10	163
Tabela 27 – Ideias que compõem o <i>cluster</i> 14	165
Tabela 28 – Ideias que compõem o <i>cluster</i> 30	166
Tabela 29 – Ideias que compõem o <i>cluster</i> 39	168
Tabela 30 – Sumarização dos agrupamentos	169
Tabela 31 – Ideias ordenadas pelo índice de potencialidade.....	170

Tabela 32 - Termos pesquisados no cenário com peso 0,4	174
Tabela 33 - Termos pesquisados no cenário com peso 0,5	175
Tabela 34 - Termos pesquisados no cenário com peso 0,6	176
Tabela 35 - Termos pesquisados no cenário com 0,7.....	178
Tabela 36 - Termos pesquisados no cenário com 0,8.....	179
Tabela 37 - Síntese dos resultados	179
Tabela 38 – Sumarização dos resultados por termo-chave	181
Tabela 39 – Ideias que compõem o <i>cluster</i> 17	184
Tabela 40 – Ideias que compõem o <i>cluster</i> 0	186
Tabela 41 – Ideias que compõem o <i>cluster</i> 7	188
Tabela 42 – Ideias que compõem o <i>cluster</i> 3	190
Tabela 43 – Ideias que compõem o <i>cluster</i> 6	191
Tabela 44 – Análise do coeficiente de Kappa	208
Tabela 45 – Matriz de confusão do algoritmo <i>Naive Bayes</i>	209
Tabela 46 – Matriz de confusão do algoritmo <i>Support Vector Machine</i>	209

LISTA DE GRÁFICOS

Gráfico 1 - Formação dos agrupamentos	151
Gráfico 2 - Porcentagem de acertos	180
Gráfico 3 - Formação dos agrupamentos no segundo cenário	182
Gráfico 4 – Dimensão 1 - Pergunta 1	193
Gráfico 5 – Dimensão 1 - Pergunta 2	194
Gráfico 6 – Dimensão 1 - Pergunta 3	195
Gráfico 7 – Dimensão 2 - Pergunta 1	196
Gráfico 8 – Dimensão 2 - Pergunta 2	196
Gráfico 9 – Dimensão 2 – Pergunta 3	197
Gráfico 10 – Dimensão 2 - Pergunta 4	197
Gráfico 11 – Dimensão 3 - Pergunta de competência 1	198
Gráfico 12 – Dimensão 3 – Pergunta de competência 2	199
Gráfico 13 – Dimensão 3 - Pergunta de competência 3	199
Gráfico 14 – Dimensão 3 - Pergunta de competência 4	200
Gráfico 15 – Dimensão 3 - Pergunta de competência 5	200
Gráfico 16 – Dimensão 3 - Pergunta de competência 6	201
Gráfico 17 – Dimensão 3 - Pergunta de competência 7	201
Gráfico 18 – Dimensão 3 - Pergunta de competência 8	202
Gráfico 19 – Dimensão 3 - Pergunta de competência 9	202
Gráfico 20 – Dimensão 3 - Pergunta de competência 10	203
Gráfico 21 – Dimensão 3 - Pergunta de competência 11	203
Gráfico 22 – Dimensão 3 – Pergunta de competência 12	204
Gráfico 23 – Dimensão 3 – Pergunta de competência 13	204
Gráfico 24 – Dimensão 3 - Pergunta de competência 14	205
Gráfico 25 – Dimensão 3 - Pergunta de competência 15	206

LISTA DE ABREVIATURAS E SIGLAS

ABNT – Associação Brasileira de Normas Técnicas
ARS – Análise de Redes Sociais
DSRM - *Design Science Research Methodology*
EGC – Engenharia e Gestão do Conhecimento
EUA – Estados Unidos da América
GC – Gestão do Conhecimento
GI – Gestão de Ideias
IMS – *Idea Management System*
LDA – *Latent Dirichlet Allocation*
LSA – *Latent Semantic Analysis*
LSI – *Latente Semantic Indexing*
MI – Mineração de Ideias
MT – Mineração de Texto
OI - *Open Innovation*
OWL – *Ontology Web Language*
PLN – Processamento de Linguagem Natural
PSM - *Problem-Solving Method*
RDF - *Resource Description Framework*
RI – Recuperação de Informação
SPARQL - *Protocol and RDF Query Language*
SVD – *Singular Value Decomposition*
SVM – *Support Vector Machine*
SWRL – *Semantic Web Rule Language*
TF-IDF - *Term Frequency–Inverse Document Frequency*
TI – Tecnologias da Informação
TIC – Tecnologias da Informação e Comunicação
UFSC – Universidade Federal de Santa Catarina
USPTO® - *United States Patent and Trademark Office*
VSM - *Vector Space Model*
WS – *Web Semântica*

SUMÁRIO

1	INTRODUÇÃO	23
1.1	DEFINIÇÃO DO PROBLEMA	24
1.2	PERGUNTA DE PESQUISA	27
1.3	OBJETIVOS	28
1.3.1	Objetivo Geral	28
1.3.2	Objetivos Específicos	28
1.4	JUSTIFICATIVA E RELEVÂNCIA DO TEMA	28
1.5	ORIGINALIDADE DA PESQUISA	32
1.5.1	Contribuições	35
1.6	ESCOPO DO TRABALHO	35
1.7	ADERÊNCIA AO PPGE GC	36
1.7.1	Identidade	36
1.7.2	Contexto Estrutural do EGC	38
1.7.3	Referências Factuais	39
1.8	ESTRUTURA DO TRABALHO	42
2	FUNDAMENTAÇÃO TEÓRICA	44
2.1	CONTEXTUALIZAÇÃO	44
2.2	DEFINIÇÃO DE IDEIA	45
2.3	GESTÃO DE IDEIAS	46
2.4	SISTEMAS DE GESTÃO DE IDEIAS	48
2.5	INOVAÇÃO ABERTA	51
2.5.1	Crowdsourcing e Gestão de Ideias	55
2.6	CRITÉRIOS DE SELEÇÃO DE IDEIAS	59
2.7	ONTOLOGIA	63
2.7.1	Ontologias e Gestão de Ideias	66
2.8	MINERAÇÃO DE DADOS E TEXTO	68

2.8.1	Técnicas de Análise Matricial	69
2.8.1.1	<i>Singular Value Decomposition (SVD)</i>	69
2.8.1.2	<i>Latent Dirichlet Allocation (LDA)</i>	71
2.8.1.3	<i>Latent Semantic Indexing (LSI)</i>	72
2.8.2	Análise de Agrupamento.....	73
2.8.2.1	<i>Representação e Normalização de Documentos</i>	74
2.8.2.2	<i>Vector Space Model (VSM)</i>	75
2.8.2.3	<i>Normalização utilizando tf-idf</i>	76
2.8.2.4	<i>Algoritmos de Análise de Agrupamentos</i>	77
2.8.2.4.1	K-means.....	78
2.8.2.4.2	Kohonen	79
2.8.2.4.3	Lingo	80
2.8.2.5	<i>Métricas de Avaliação de Agrupamentos</i>	81
2.8.3	Análise de Redes Sociais	82
2.9	TRABALHOS CORRELATOS.....	84
2.10	CONSIDERAÇÕES FINAIS	86
3	METODOLOGIA DE PESQUISA.....	87
3.1	PESQUISA CIENTÍFICA E PESQUISA TECNOLÓGICA.....	88
3.2	<i>DESIGN SCIENCE RESEARCH METHODOLOGY (DSRM)</i>	90
3.3	DESENVOLVIMENTO DA PESQUISA.....	93
3.3.1	Definição do problema e solução.....	95
3.3.2	Definição dos objetivos para solução	96
3.3.3	Projeto e desenvolvimento	96
3.3.3.1	<i>Coleta dos dados e preenchimento da base de conhecimento</i>	96
3.3.3.2	<i>Pré-processamento dos dados</i>	98
3.3.3.3	<i>Transformação dos dados</i>	99
3.3.3.4	<i>Agrupamento dos dados</i>	99

3.3.4	Demonstração	99
3.3.5	Avaliação	101
3.3.6	Comunicação	102
3.4	SÍNTESE DA METODOLOGIA DE PESQUISA	102
4	REPRESENTAÇÃO DA BASE DE CONHECIMENTO	104
4.1	METODOLOGIA ADOTADA.....	104
4.1.1	Escopo da ontologia	105
4.1.2	Reuso da ontologia	106
4.1.3	Definição dos conceitos e classes	106
4.1.4	Definição das propriedades e restrições	107
4.2	LEVANTAMENTO DE INFORMAÇÕES PARA FORMALIZAÇÃO DA ONTOLOGIA.....	112
4.3	CONSIDERAÇÕES FINAIS	124
5	DEFINIÇÃO DOS ÍNDICES	125
5.1	ANÁLISE DE FONTES EXTERNAS.....	125
5.2	ÍNDICE DE ADERÊNCIA AO CONTEXTO ORGANIZACIONAL.....	126
5.3	ÍNDICE DE ATUALIDADE	127
5.4	ÍNDICE DE TECNOLOGIA	129
5.5	ÍNDICE DE ORIGINALIDADE	130
5.6	ÍNDICE DE POTENCIALIDADE DE IMPLEMENTAÇÃO DA IDEIA.....	133
5.7	CONSIDERAÇÕES FINAIS	134
6	MODELO PROPOSTO	135
6.1	APRESENTAÇÃO DO MODELO.....	135
6.1.1	População da Base de Conhecimento e Seleção de Ideias	136
6.1.2	Pré-processamento	137
6.1.3	Transformação	139
6.1.4	Mineração de Texto	141
6.1.5	Avaliação das Ideias	144

6.2	CONSIDERAÇÕES FINAIS	145
7	ANÁLISE E APRESENTAÇÃO DOS RESULTADOS	146
7.1	APRESENTAÇÃO DO CENÁRIO	146
7.1.1	População da Base de Conhecimento e Seleção das Ideias.....	147
7.1.2	Pré-processamento e Transformação	148
7.1.3	Transformação.....	148
7.1.4	Mineração de Texto.....	148
7.2	ANÁLISE DOS RESULTADOS OBTIDOS	149
7.2.1	Primeiro cenário de análise	149
7.2.1.1	<i>Análise da formação da rede de palavras do primeiro cenário.....</i>	<i>171</i>
7.2.2	Segundo cenário de análise.....	181
7.3	AVALIAÇÃO DA ONTOLOGIA E MODELO PROPOSTOS.....	192
7.3.1	Análise das Respostas às Perguntas da Dimensão 1.....	193
7.3.2	Análise das Respostas às Perguntas da Dimensão 2.....	195
7.3.3	Análise das Respostas às perguntas da Dimensão 3.....	198
7.4	COMPARAÇÃO COM MODELOS PREDITIVOS SUPERVISIONADOS	206
7.5	DISCUSSÃO DOS RESULTADOS	209
7.6	CONSIDERAÇÕES FINAIS	212
8	CONSIDERAÇÕES FINAIS.....	214
8.1	LIMITAÇÕES.....	217
8.2	TRABALHOS FUTUROS.....	217
	REFERÊNCIAS	220
	APÊNDICE A – Formulário de avaliação da ontologia.....	239

1 INTRODUÇÃO

Para superar os desafios impostos pelo mercado de negócios, organizações estão em transição na forma como conduzem o processo de inovação. De acordo com Baregheh, Rowley e Sambrook (2009), inovar pode ser caracterizada como a competência de transformar ideias em produtos novos ou melhores, serviços ou processos, com o intuito de alcançar diferenciais capazes de garantir o sucesso de mercado. Criar e colocar em prática ideias inovadoras podem representar uma tática valiosa para fomentar o crescimento econômico (MAHROUM; AL-SALEH, 2013). Fator este, crítico para o sucesso do negócio (POORKAVOOS et al., 2016).

Segundo Poorkavoos et al. (2016), a gestão de ideias reflete de maneira positiva e torna-se uma fonte inimitável e crucial para o desenvolvimento de inovações. Neste sentido, os autores salientam que recentemente as organizações perceberam o potencial das tecnologias em rede para a geração de ideias e desenvolvimento de produtos inovadores. As organizações têm incentivado e investido na participação do cliente no processo de criação de valor do produto por meio de plataformas colaborativas para sugestões de ideias. Estas sugestões são vistas como ativos estratégicos para o sucesso da Gestão de Ideias (SALDIVAR et al., 2019).

Mačiulienė e Skaržauskienė (2016) corroboram ao afirmarem que o envolvimento de clientes, fornecedores e colaboradores na criação de produtos, tornou-se uma estratégia e tendência no mercado de negócios. As plataformas de colaboração permitem que colaboradores façam sugestões de produtos ou serviços, cuja viabilidade será estudada pelas organizações, funcionando como uma ponte entre os desejos dos consumidores e a empresa.

Estas plataformas auxiliam na geração, avaliação e seleção de ideias para serem colocadas em prática, possibilitando a formação de um banco colaborativo de ideias capaz de contribuir no planejamento de ações de melhoria contínua e na promoção de inovações.

A eficiência do processo de gestão de ideias pode gerar:

- Aumento da produtividade organizacional (SINT et al., 2010);
- Otimização dos processos realizados na organização (WESTERSKI; IGLESIAS, 2011);
- Favorecimento na criação de novos produtos (POVEDA; WESTERSKI; IGLESIAS, 2012; XIE; ZHANG, 2010);
- Minimização dos custos (KARIMI-MAJD; MAHOOTCHI, 2015);
- Manutenção da vantagem competitiva empresarial (WESTERSKI; DALAMAGAS; IGLESIAS, 2013);

- Geração de valor comercial por meio das ideias (POVEDA; WESTERSKI; IGLESIAS, 2012; KARIMI-MAJD; MAHOOTCHI, 2015).

Entretanto, a área de Gestão de Ideias possui grandes desafios, tais como:

- O excesso de informação por meio do grande volume de ideias submetidas (SINT et al., 2010; WESTERSKI; DALAMAGAS; IGLESIAS, 2013; MARTINEZ-TORRES; OLMEDILLA, 2016; HOORNAERT, 2017; DZIALLAS; BLIND, 2019);
- Os picos súbitos de submissões de ideias (HOORNAERT, 2017; DZIALLAS; BLIND, 2019);
- A redundância e trivialidade das ideias (SINT et al., 2010);
- O esforço humano necessário durante o processo de gestão de ideias (WESTERSKI; DALAMAGAS; IGLESIAS, 2013; MARTINEZ-TORRES; OLMEDILLA, 2016; HOORNAERT, 2017);
- A dificuldade no processamento interno e execução de ideias por especialistas da organização (DZIALLAS; BLIND, 2019), e;
- A identificação das ideias relevantes perante o volume (HOORNAERT, 2017; DZIALLAS; BLIND, 2019).

Todos estes desafios impactam diretamente na tomada de decisão pelos especialistas da organização no momento da avaliação das ideias, dificultando o gerenciamento das sugestões propostas. Embora haja um número crescente de pesquisas relacionadas à gestão de ideias, a seleção de ideias pragmáticas e úteis continua a ser um grande desafio, faltando pesquisas que utilizem abordagens quantificadas ou orientadas a dados (KARIMI-MAJD; MAHOOTCHI, 2015).

1.1 DEFINIÇÃO DO PROBLEMA

A gestão de ideias representa o ponto inicial do processo de inovação e auxilia na identificação de oportunidades viáveis e prováveis de serem implementadas pela organização (BJÖRK; BOCCARDELLI; MAGNUSSON, 2010; VANDENBOSCH, SAATCIOGLU; FAY, 2006).

Para tal, faz-se necessário o uso de sistemas de gestão de ideias com o intuito de utilizar tecnologias da informação para auxiliar na geração, avaliação e seleção de ideias inovadoras, empregando diferentes processos e critérios dentro do mesmo sistema (LI; LI; CHEN, 2014;

WESTERSKI; DALAMAGAS; IGLESIAS, 2013). Estes sistemas são considerados um ramo promissor do mercado de *software* (FENN; LEHONG, 2011; WESTERSKI; IGLESIAS, 2011).

Por meio de sistemas de gestão de ideias é possível criar comunidades de inovação aberta. Estas podem fornecer uma ampla gama de ideias e reduzir os custos associados à Pesquisa e Desenvolvimento (P&D) (MARTINEZ-TORRES; OLMEDILLA, 2016).

Todavia, os autores Martinez-Torres e Olmedilla (2016); Gabriel et al. (2016) e Dziallas e Blind (2019) salientam que o principal problema destes sistemas é o alto volume de informações geradas, que devem ser internamente processadas e analisadas por especialistas. Para Lee et. al (2018) e Hoornaert et. al (2017), essa dificuldade se agrava devido ao formato não estruturado das ideias, sendo que poucos estudos propuseram um método prático para resolver o desafio do problema de sobrecarga. Isto ocorre porque a formalização e a automatização do processo de avaliação são complexas.

Magnusson, Netz e Wästlund (2014) corroboram e ressaltam que entre os desafios ocasionados pelo alto volume de submissões está a presença de repetições de ideias similares. Para os autores, uma única ideia pode aparecer inúmeras vezes na base de dados, embora com especificidades diferentes, porém relacionadas. Tecnologias baseadas na mineração de dados e texto, como a análise de agrupamento, podem contribuir neste processo ao detectar numa meta-análise de ideias semelhantes, apontando possíveis padrões e tendências (MAGNUSSON; NETZ; WÄSTLUND; 2014).

Essas tecnologias tornam possível processar e visualizar dados de forma mais eficaz, proporcionando aos especialistas explorar o conhecimento de domínio e de domínios adjacentes para encontrar possíveis novas soluções por meio da exploração interativa dos dados (ESCANDÓN-QUINTANILLA; GARDONI; COHENDET, 2018; WESTERSKI; KANAGASABAI, 2019). Permite também uma análise mais eficiente de grandes volumes de dados e uma visualização rápida para a exploração e a interpretação dos resultados (ESCANDÓN-QUINTANILLA; GARDONI; COHENDET, 2018; KÖHL et al., 2019).

Deste modo, pode-se afirmar que o conhecimento envolvido no domínio de gestão de ideias é geralmente latente e deve ser descoberto de forma adequada, usando métodos quantitativos, como a mineração de dados e texto ou abordagens baseadas em dados, amplamente utilizadas nas últimas décadas, colaborando na tomada de decisões (KRIDEL; DOLK, 2013; KARIMI-MAJD; MAHOOTCHI, 2015; WESTERSKI; KANAGASABAI, 2019).

Contudo, o maior desafio ainda é a seleção de ideias úteis e pragmáticas para o desenvolvimento de novos produtos, pois a seleção de uma ideia pode depender de diferentes

fatores, como o número de votos, os comentários, os pontos ganhos, a quantidade de ideias apresentadas, a relevância, a originalidade, o alinhamento de uma ideia à estratégia empresarial da organização e a viabilidade da ideia para implementação (HOSSAIN; ISLAM, 2015a).

A viabilidade da ideia representa o quão provável e facilmente pode ser traduzida em um produto ou serviço comercial. É uma característica central que determina a preferência das empresas tanto do ponto de vista técnico quanto econômico (CHAN; LI; ZHU, 2018). A avaliação da viabilidade tecnológica e econômica das ideias é, portanto, um critério avaliativo saliente que naturalmente vem à mente dos tomadores de decisão (CHAN; LI; ZHU, 2018). As inovações e as novas tecnologias também devem fazer parte da estratégia empresarial e integrar todos os serviços interessados em qualquer ideia ou abordagem inovadora (GROVER, 2017). Assim, em vez de meramente prever a adoção de ideias, seria adequado recomendar as ideias com uma maior possibilidade de serem aceitas pelos tomadores de decisão da organização (YOO et al.; 2015).

Além disso, a eliminação de ideias com baixo potencial pode impedir que as empresas sofram investimentos e falhas dispendiosas (FERIOLI et al., 2010). Os tomadores de decisão, portanto, devem identificar cuidadosamente as ideias potenciais para o desenvolvimento futuro, prevendo seu sucesso potencial e os benefícios que elas trazem para a empresa (BERG, 2016).

Deste modo, deveriam ser recomendadas às ideias que possuem alinhamento ao contexto organizacional (WESTERSKI; KANAGASABAI, 2019). O contexto organizacional pode ser definido como as estruturas abrangentes e sistemas externos ao projeto que influenciam em seu funcionamento (DENISON et al., 1996). Harrison et al. (2016) vão além e definem o contexto organizacional como o domínio relacionado a organização que possui como foco recursos duradouros, incluindo características estruturais, cultura, capacidade de melhoria dos recursos, conhecimento e habilidades e infraestrutura de tecnologia da informação (TI).

A compreensão do contexto organizacional contribui na análise e tomada de decisão no domínio de gestão de ideias (FERIOLI et al., 2010; CHAN; LI; ZHU, 2018; WESTERSKI; KANAGASABAI, 2019), maximizando o valor de novas ideias e aumentando a probabilidade de sucesso da inovação (CHAN; LI; ZHU, 2018; MIKELSONE; VOLKOVA; LIELĀ, 2019b).

Entretanto, decidir sobre a probabilidade de sucesso e o grau de alinhamento com os interesses da organização é uma tarefa difícil e demorada (WESTERSKI; KANAGASABAI, 2019). Portanto, previsões precisas são difíceis no estágio inicial da inovação, uma vez que ideias inovadoras geralmente partem das existentes e estão associadas à incerteza e ao risco (BAER, 2012). Avaliar se uma nova ideia deve ser transformada em um produto é uma decisão crítica que liga a geração e conversão de novas ideias em produtos promissores. Entender o

processo de tomada de decisão é crucial para maximizar o valor de novas ideias e aumentar a probabilidade de sucesso da inovação (CHAN; LI; ZHU, 2018). Para Steele, Johnson e Medeiros (2018), o propósito da avaliação de ideias é selecionar e desenvolver as ideias potenciais, ou seja, com maior potencial de sucesso em uma solução viável.

Nisto, torna-se significativo possuir indicadores precisos para avaliar claramente as ideias, uma vez que melhorar o processo de avaliação de inovações também pode ajudar investidores na decisão de financiar ou não novos empreendimentos (DZIALLAS; BLIND, 2019).

Apesar desses desafios práticos, a literatura relacionada a mecanismos para filtrar e avaliar ideias inovadoras é surpreendentemente escassa (CHAN; LI; ZHU, 2018). Ainda não há um processo ou um modelo único, adequado para avaliar ideias, utilizando abordagens quantificadas ou orientadas a dados. Pode-se ressaltar também a necessidade de novos mecanismos que auxiliam os especialistas a tomarem decisões sobre a potencialidade de implementação de ideias. Assim, uma dimensão de decisão importante a ser considerada ao longo do processo de gestão de ideias se refere sobre quais abordagens de mineração de dados e texto utilizar, resultando na descoberta da lacuna desta pesquisa.

Verifica-se, desta forma, a necessidade de maior aprofundamento no âmbito da identificação de métodos e técnicas de mineração de dados e texto utilizados no domínio de gestão de ideias, assim como de estudos que busquem quantificar o potencial de implementação das ideias.

Considerando que a literatura sobre o tema ainda não abrange toda a complexidade e profundidade das lacunas levantadas, esta tese visa avaliar um conjunto de ideias evidenciando o potencial de implementação de acordo com o contexto organizacional, definido através de temáticas estratégicas pela organização e, além disso, contribuir na visualização destas informações. Neste sentido, objetiva-se a geração de conhecimento científico com o intuito de auxiliar na proposição de soluções para as lacunas aqui evidenciadas.

1.2 PERGUNTA DE PESQUISA

Diante do exposto, surge a seguinte pergunta de pesquisa: Como avaliar o potencial de implementação de ideias levando-se em consideração o contexto organizacional?

1.3 OBJETIVOS

Nesta seção são descritos os objetivos, geral e específicos, que se buscam atingir com o desenvolvimento desta tese.

1.3.1 Objetivo Geral

Propor um modelo de avaliação do potencial de implementação de ideias alinhadas ao contexto organizacional.

1.3.2 Objetivos Específicos

- Identificar métodos e técnicas de mineração de dados e texto para a avaliação de ideias;
- Criar índices que indiquem o potencial de implementação de uma ideia;
- Demonstrar a viabilidade do modelo proposto por meio da construção de um protótipo e da elaboração de cenários de estudo.

Na próxima seção apresenta-se a justificativa e a relevância do tema de pesquisa.

1.4 JUSTIFICATIVA E RELEVÂNCIA DO TEMA

Produzir inovações não é trivial, porém essencial para manutenção da competitividade organizacional. Fato que constitui um dos fatores críticos para o sucesso do negócio (FERAUGE, 2012; POORKAVOOS et al., 2016), tornando uma estratégia poderosa para promoção do crescimento econômico (MAHROUM; AL-SALEH, 2013; KARIMI-MAJD; MAHOOTCHI, 2015).

Nesta perspectiva, as empresas passaram a incentivar e investir na participação do cliente no processo de criação de valor do produto por meio de plataformas colaborativas para gestão de ideias (POORKAVOOS et al., 2016). Plataformas colaborativas são cada vez mais prevalentes e produtivas com os sistemas de gestão de ideias, capturando ideias inovadoras com o potencial para gerar bilhões em valor (TOWNE; ROSÉ, 2016).

O intuito destas plataformas é incentivar os usuários a apresentarem ideias para criar ou melhorar produtos e sua experiência com a organização. Por meio das sugestões de ideias,

organizações exploram a inteligência coletiva dos usuários, fomentando a colaboração para a filtragem de ideias, reduzindo desta forma a quantidade de trabalho necessário para a avaliação (CARBONE et al., 2012).

As empresas exploram a criatividade, as ideias e o conhecimento dos colaboradores em benefício próprio, principalmente no que tange a identificação de melhores soluções para serem implementadas (CARDOSO, 2016). Por meio da combinação de diferentes conhecimentos e habilidades de colaboradores, cria-se valor de forma mais eficaz e eficiente (FROW et al., 2015). Portanto, a inovação é, em grande medida, considerada um processo social e comunicativo, e a entrada de outros indivíduos melhora a geração de ideias nos estágios iniciais da criação e desenvolvimento (BERGENDAHL; MAGNUSSON, 2014).

Para Bergendahl e Magnusson (2014), os sistemas de gestão de ideias passaram a constituir plataformas para ações coletivas de criatividade. Inspirando-se na perspectiva da inovação aberta, muitas empresas têm trabalhado de forma distribuída e em rede. Multinacionais como a Dell[®], Starbucks[®], Cisco[®], Canonical[®], Repsol[®], 3M[®], Procter & Gamble[®] e Google[®] usam sistematicamente princípios de gerenciamento de ideias para capturar e refinar ideias de fontes internas e externas para melhorar os resultados da inovação e direcionaram fortes investimentos para o domínio de Gestão de Ideias (GAMA; FRISHAMMAR; PARIDA, 2019). Elas perceberam não apenas uma oportunidade, mas, sobretudo, uma necessidade para manter a vantagem competitiva no mercado, aproximando-se da comunidade de inovação e estimulando a inovação e a melhoria dos seus produtos (CARBONE et al., 2012; WESTERSKI; DALAMAGAS; IGLESIAS, 2013). Com essas plataformas é possível aprimorar colaborativamente as ideias propostas, avaliá-las e fazer os seus gerenciamentos em sinergia com outros processos empresariais (WESTERSKI; DALAMAGAS; IGLESIAS, 2013).

Neste sentido, Bothos, Apostolou e Mentzas (2012) afirmam que a *internet* contribuiu para criar novos modelos de negócio e inovações com maior participação na utilização, produção e difusão de conhecimento. Com a *web 2.0*, os usuários puderam coletivamente, por meio de serviços e aplicações, gerar conteúdo através de meio virtual (BOTHOS; APOSTOLOU; MENTZAS, 2012; DISSANAYAKE; ZHANG; GU, 2015). Um exemplo deste novo paradigma é a inteligência coletiva.

A inteligência coletiva refere-se à inteligência que emerge das interações entre pessoas. Nas últimas décadas as tecnologias da *web 2.0* possibilitaram novas formas de inteligência coletiva (KHAZAEI; XIAO, 2015). O autor Wang (2013), após realizar uma revisão, destacou as vantagens de se trabalhar com a inteligência coletiva, entre elas: baixo

custo de acesso a um grande número de contribuintes potenciais; geração de soluções; acesso a um grupo maior de colaboradores; especialização externa complementar; diversidade em suposições, soluções, crenças e estímulos; precisão e julgamentos individuais em grande número; diversidade de padrões de perfis de risco e percepções de consumidores sobre a capacidade de inovação de uma empresa.

Elia, Margherita e Passiante (2020) corroboram ao afirmarem que o potencial de adotar uma perspectiva de inteligência coletiva pode residir em duas considerações. Primeiro, a ideia para empreender é uma atividade complexa e articulada, podendo se beneficiar da colaboração de indivíduos e de sistemas que irão contribuir em um nível diferente, compartilhando conhecimento, unindo redes e disponibilizando recursos. Segundo, o uso de tecnologias digitais pode otimizar a produção de bens e serviços inovadores e amplificar o impacto e a escala das redes, além do compartilhamento de conhecimento e experiência, úteis para realizar as diferentes fases do processo empreendedor (ELIA; MARGHERITA; PASSIANTE, 2020).

Assim, o papel das TIC passa a ser não somente o de um artefato por meio do qual o conhecimento pode ser articulado, armazenado, compartilhado e recuperado, mas também uma forma para que as empresas possam ampliar suas redes sociais em busca de novos conhecimentos relevantes (MARTINI; NEIROTTI; APPIO, 2016).

Porém, avaliar e organizar essas contribuições é um desafio devido ao volume de informações e a estrutura dos dados representada por coleções de documentos textuais (HOORNAERT et al., 2017). À medida que essas coleções crescem, demandam maior tempo de avaliação.

Encontrar ou sintetizar valor a partir de um subconjunto de documentos estreitamente relacionados pode ser especialmente difícil, exigindo esforços de pesquisa para produzir tecnologias avançadas que facilitem a análise de texto (LIU et al., 2012). Também para fornecer recomendações para apoiar a navegação e a descoberta de informações desejadas em sites com grandes volumes de conteúdo textuais (ZHAO et al., 2011).

Segundo os autores Hoornaert et al. (2017), é altamente desafiador para as organizações processarem grandes volumes de informações que fluem por meio de comunidades de *crowdsourcing* e identificar as ideias potenciais de investimento. Segundo os autores, as tentativas de automatizar ou formalizar o processamento de ideias são complicadas, pois novas ideias e suas discussões são geralmente fornecidas como texto. Para Niederberger et al. (2012), este é um problema de Recuperação de Informação (RI) difícil, devido os textos serem muito curtos e com pouca estrutura predefinida. Os autores citam vários desafios como

textos usuais e curto, falta de estrutura formal, ideias propensas a erros, escritas desleixadas e multilinguismo (ideias podem conter palavras de diferentes idiomas).

Deste modo, os tomadores de decisão podem não ter tempo suficiente ou recursos cognitivos para avaliar grandes volumes de ideias (ALKSHER et al., 2016). Beretta (2019) corrobora ao afirmar que a geração de uma grande quantidade de ideias cria um problema de seleção, dificultando a identificação de quais ideias devem ser transformadas em projetos formais. Este aumento no número de ideias a serem avaliadas acresce significativamente as despesas das empresas de recursos internos para processar, podendo limitar a abertura da empresa às ideias mais inovadoras (STEELE; JOHNSON; MEDEIROS, 2018).

Extraír a ideia oculta dos textos é essencial para o processo de tomada de decisão em razão do aumento do tempo para processar as informações coletadas que podem causar a perda de produtividade (ALKSHER et al., 2016).

Além disso, menos atenção tem sido dispndida em outras fases gerenciais críticas, como estabelecer prioridades entre as ideias inovadoras que são geradas nos limites organizacionais da empresa e na seleção das mais consistentes com a estratégia de negócios da empresa (SALDIVAR et al., 2016).

Segundo os autores Hoornaert et al. (2017), métodos de avaliação de conteúdo de ideias baseados em mineração de dados e texto são promissores, pois podem ser utilizados para processar grandes quantidades de conteúdo não estruturado em tempo real, podendo fornecer previsões confiáveis do sucesso comercial dos produtos propostos e a intenção de compra dos consumidores.

O modelo proposto buscar contribuir no processo de avaliação de ideias. Para isto, torna-se necessário representar o conhecimento de domínio. Nos Sistemas de Gestão de Ideias, a representação do conhecimento é realizada por meio de ontologias (WESTERSKI et al., 2013). Segundo Westerski et al. (2013, p. 11): “a ontologia estabelece bases para a gestão do conhecimento com base na interligação de sistemas corporativos e ativos da *internet* com o princípio de aumentar a consciência da informação e ajudar na avaliação da inovação”.

Ontologias proporcionam uma estrutura para captar os conhecimentos envolvidos e possibilitam raciocinar sobre o conhecimento coletado, realizando inferências e sugestões, além de permitir recuperar a informação de maneira mais precisa e eficiente (HERNÁNDEZ-GONZÁLEZ et al., 2014; AFACAN; DEMIRKAN, 2011). A formalização da área de conhecimento por meio da ontologia é interessante para a colaboração entre humanos e para a interoperabilidade entre sistemas digitais, criando um modelo de informação explícita que facilita a comunicação (HERNÁNDEZ-GONZÁLEZ et al., 2014).

1.5 ORIGINALIDADE DA PESQUISA

O modelo proposto nesta pesquisa combina técnicas de mineração de dados e texto, utilizando como forma de representação do conhecimento uma ontologia para auxiliar no processo de avaliação de ideias alinhadas ao contexto organizacional. Mais precisamente, busca-se contribuir no processo de tomada de decisão, auxiliando especialistas de domínio na avaliação de ideias alinhadas às temáticas estratégicas relacionadas ao contexto organizacional.

Para identificar o problema e propor uma solução foi necessário realizar uma busca sistemática, além da procura de uma lacuna de conhecimento que correlacione a gestão de ideias e os métodos e técnicas de mineração de dados e texto que podem contribuir ou melhorar a avaliação das ideias. Nesta tese, o termo potenciais ideias representa o potencial de implementação de determinada ideia.

Com o objetivo de obter uma visão mais aprofundada e completa, a revisão sistemática de literatura foi realizada por meio da expressão de busca ("*idea management*" or "*idea generation*" or "*idea mining*" or "*crowdstorm**") AND ("*cluster**" or "*text mining*" or "*data mining*" or "*crowdsourcing*" or "*quatitative analysis*"), nas bases de dados *Scopus*[®], *Web of Science*[®] e *Science Direct*[®]. A escolha das bases ocorreu em função da credibilidade e relevância. Também se pesquisou pela expressão de busca contendo os termos ("*evaluation idea*" and ("*text mining idea*" or "*data mining*")), contudo não foi identificado nenhum documento relacionado.

A revisão realizada inicialmente em 2017 passou por atualizações constantes e buscou sempre por novos artigos relacionados ao tema de pesquisa. A última revisão que objetivou encontrar novos documentos ocorreu em dezembro de 2019. A base de dados da *Science Direct*[®] retornou 24 artigos contendo a expressão nos campos de título, resumo e palavras-chave. Na base *Web Of Science*[®] foram encontrados 106 documentos que continham a expressão no campo tópico. Na base *Scopus*[®] foram identificados 143 documentos condizentes com a expressão nos campos de título, resumo e palavras-chave. A pesquisa totalizou o montante de 163 artigos sem repetição.

Dos 163 artigos analisados e classificados como relevantes para a pesquisa, apenas oito apresentavam métodos ou técnicas de mineração utilizadas no contexto de gestão de ideias. A discussão dos artigos é realizada na Seção 2.9 do Capítulo 2. Durante a revisão sistemática, não foram encontrados artigos que tivessem como foco a avaliação de ideias com base em métodos e técnicas de mineração de dados e texto alinhadas às temáticas estratégicas relacionadas ao contexto organizacional.

Esses números demonstram como são incipientes os estudos relacionados ao tema. Ao analisar as publicações resultantes da revisão bibliográfica apresentada no Quadro 1, pode-se perceber que nenhum dos estudos investigou o potencial de implementação de ideias alinhadas às temáticas estratégicas envolvidas no contexto organizacional por meio de métodos e técnicas de mineração de dados e texto, com o intuito de contribuir na tomada de decisão. Portanto, é possível constatar um campo propício para novas pesquisas.

Quadro 1 - Tecnologias e ferramentas utilizadas no domínio de Gestão de Ideias

Id	Método/Técnica de Mineração de Dados e Texto	Finalidade	Autor(es)	Ano
1	Análise de Agrupamentos	Utilizar um método de agrupamento baseado em uma rede Hebbian, capaz de aprender os principais componentes dos dados de forma contínua.	Niederberger et al.	2012
2	Análise de Agrupamentos	Utilizar um algoritmo de agrupamento para descobrir se os caminhos gerados por máquinas produzem maior fluência e amplitude no processo de ideação.	Chan, Dang e Down	2016
3	Análise de Agrupamentos	Desenvolver um modelo baseado em ontologias e análise de agrupamento para apoio à Gestão de Ideias. A meta é contribuir no processo de tomada de decisão. O algoritmo utilizado no agrupamento foi o Lingo [®] .	Sérgio, Souza e Gonçalves	2017
4	Algoritmo t-SNE	Criar uma plataforma para possibilitar postar ideias, visualizá-las de outros usuários e as postadas pelo solicitante, mover ideias de (B) para (C) para organizar, visualizar um <i>minimap</i> do espaço de trabalho, além de visualizar ideias semelhantes. A plataforma também calcula uma matriz aproximada de similaridade de ideias usando o Algoritmo <i>t-Distributed Stochastic Neighbor Embedding</i> (t-SNE).	Siangliulue et al.	2016
5	Algoritmo <i>tf-idf</i> e Análise de Sentimento	Recomendar ideias utilizando técnicas de mineração de texto, incluindo <i>tf-idf</i> e análise de sentimento.	Lee et al.	2018
6	Método automatizado de Recuperação de Informação	Desenvolver um método que inclui simultaneamente os 3Cs em relação a novas ideias (conteúdo, experiência de colaborador e <i>feedback</i> da multidão). Utilizar também métodos lineares clássicos e métodos não-lineares de	Hoornaert et al.	2017

		aprendizado de máquina para prever a implementação de ideias.		
7	Mineração de dados e texto	Classificação de ideias com base na análise da capacidade de inovação disruptiva das ideias. Assim, utilizar um conjunto de ferramentas analíticas e algoritmos de detecção outlier.	Westerski e Kanagasabai	2019
8	Ferramenta Leximancer®	Utilização do <i>software</i> Leximancer® para recuperar informações dos textos das ideias e obter <i>insights</i> sobre o discurso.	Wehnert et al.	2018

Fonte: Adaptado de Sérgio e Gonçalves (2019)

Conforme pode ser observado, o Quadro 1 está dividido em quatro categorias representadas por tons de cinza e branco. As linhas 1, 2 e 3 correspondem à aplicação da análise de agrupamento no domínio de Gestão de Ideias e *Crowdsourcing*. As linhas 4 e 5 se referem à utilização de algoritmos no contexto de Gestão de Ideias e *Crowdsourcing*, enquanto que as linhas 6 e 7 dizem respeito à introdução de métodos baseados em mineração de dados e texto. E, por último, a utilização da ferramenta Leximancer®.

Este processo possibilitou a definição inicial do problema de pesquisa e de algumas formas preliminares em como solucioná-lo. No que tange ao ineditismo desta tese, verificou-se que as pesquisas relacionadas à avaliação de ideias, em sua maioria, são estudos que configuram levantamentos descritivos de métodos e técnicas orientados a dados.

Nos esboços iniciais do objeto de pesquisa, a temática acerca da avaliação de ideias com o emprego de métodos e técnicas de mineração de dados e texto, mostrou-se promissora. Ao avançar no entendimento do tema, verificou-se que o provimento de temáticas estratégicas vinculadas ao contexto organizacional é essencial para que tais ideias possam ser elucidadas como potenciais para receberem investimentos. Além disso, a criação de índices baseados em fontes externas, oriundas da *web*, como a base de patentes dos Estados Unidos da América (EUA) e a base de artigos científicos *Semantic Scholar*®, é essencial para que o processo de avaliação de ideias efetivamente promova valor à organização. Por fim, a representação semântica da base de conhecimento visa proporcionar a formalização e conceitualização compartilhada do domínio. Neste contexto, foi identificada uma oportunidade para tratar deste tema de forma inédita, com a proposição de um modelo para avaliação de ideias alinhadas às temáticas estratégicas relacionadas ao contexto organizacional. Desta forma, pode-se contribuir na visualização dos dados por meio da análise de agrupamento e outras abordagens de visualização de informação. A instanciação do modelo visa reduzir o tempo dispendido com o

processo de avaliação por especialistas de domínio e contribuir na tomada de decisão de quais ideias devam ser selecionadas, otimizando, desta forma, os recursos organizacionais.

1.5.1 Contribuições

A principal contribuição desta pesquisa consiste na proposição de um modelo para avaliação de ideias alinhadas às temáticas estratégicas relacionadas ao contexto organizacional. O modelo utiliza a representação semântica, bem como métodos e técnicas de mineração de dados e texto para a avaliação do potencial de implementação de ideias.

Além da contribuição principal, são apresentadas outras contribuições da referente pesquisa:

- Definição de uma estrutura de dados para armazenamento, recuperação e análise das ideias;
- Criação de uma ontologia para representação do domínio;
- Implementação de um protótipo que ateste a viabilidade do modelo proposto, bem como a aplicação em cenários de estudo;
- Criação de índices para a avaliação do potencial de implementação das ideias;
- Construção de uma base de conhecimento voltada à representação das ideias, do contexto organizacional e dos resultados obtidos com o modelo.

A próxima seção demonstra o escopo da presente pesquisa.

1.6 ESCOPO DO TRABALHO

Esta tese busca avaliar o potencial de implementação de ideias alinhadas às temáticas estratégicas relacionadas ao contexto organizacional, identificando características e particularidades contidas em uma base de conhecimento. Sendo assim, características como desempenho e interface gráfica não serão levadas em consideração, pois não são essenciais para atingir os objetivos.

O modelo proposto utiliza-se de uma série de métodos e técnicas de mineração de dados e texto, não tem como pretensão trazer uma contribuição direta para determinada técnica ou método em si.

A abordagem utilizada para a tarefa de agrupamento foi proposta por Alvarez (2018), que utilizou uma versão modificada do algoritmo *K-means*.

Para representar a base de conhecimento do modelo foi desenvolvida uma ontologia. Entretanto, não é o foco desta tese apresentar técnicas de modelagem ou criação de ontologias.

Também não é intuito desta tese que o protótipo desenvolvido para suportar o modelo proposto apresente características de um produto final, uma vez que ele será desenvolvido para avaliar o modelo proposto.

Para a avaliação do modelo proposto, serão utilizadas ideias coletadas manualmente de bases de ideias disponíveis na *web*. Também será realizada a avaliação do modelo e ontologia propostas com especialistas da área. Por fim, testes com modelos preditivos supervisionados serão efetuados. Não se constitui como objetivo desta tese o desenvolvimento de um novo método de agrupamento de dados.

Esta tese baseia-se na premissa de que o conhecimento contido nas ideias pode ser analisado, explicitado e visualizado por meio de técnicas de mineração de dados e texto, evidenciando padrões e tendências em ideias alinhadas às temáticas estratégicas relacionadas ao contexto da organização.

1.7 ADERÊNCIA AO PPGEGC

O objetivo desta seção é contextualizar e evidenciar a aderência do trabalho proposto ao objeto de pesquisa do Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento. Para isso, dividiu-se a seção em três subseções que apresentam a identidade da tese, o contexto estrutural da pesquisa no PPGEGC e as referências factuais relacionadas à pesquisa.

1.7.1 Identidade

A presente tese possui como área de concentração a Engenharia do Conhecimento, na linha de pesquisa: Teoria e Prática em Engenharia do Conhecimento. Esta linha trabalha com a criação, o desenvolvimento e a implantação de soluções da Engenharia do Conhecimento em organizações (EGC, 2017).

A área de Engenharia do Conhecimento (EC) determina metodologias e ferramentas para capturar e modelar o conhecimento com a intenção de torná-lo independente de seres humanos, possibilitando a formalização e apropriação deste conhecimento por organizações ou sistemas.

A EC objetiva transformar o processo de construção de sistemas baseados no conhecimento com o desenvolvimento de métodos, linguagens e ferramentas apropriadas (KAHRAMAN, 2014). Segundo Kahraman (2014), torna-se muitas vezes necessário em sistemas de engenharia do conhecimento, solucionar problemas complexos com recursos incompletos, fato que demanda um tratamento adequado da incerteza e confiança nas opiniões dos especialistas. O autor também destaca que as técnicas de modelagem e análise da incerteza na engenharia do conhecimento apresentam ferramentas analíticas e úteis, apropriadas para analistas atuais e futuros e para profissionais que procuram compreender os fundamentos do conhecimento.

Sistemas de conhecimento podem impactar no aumento da produtividade, preservação de conhecimento e melhorar o processo de tomada de decisão. Schreiber et al. (2000) definem sistema de conhecimento como sendo “qualquer sistema de informação que gerencie, armazene e/ou aplique conhecimento organizacional explicitamente representado”.

A EC possui como intuito pesquisar e desenvolver técnicas e ferramentas para a codificação, formalização e gestão do conhecimento; métodos de análise da estrutura e processos em atividades intensivas em conhecimento; sistemas de conhecimento.

O objetivo do PPGEGC é investigar, conceber, desenvolver e aplicar modelos, métodos e técnicas relacionados tanto a processos/bens/serviços como ao seu conteúdo técnico-científico (EGC, 2017). A contribuição da EC vem ao encontro ao objeto de pesquisa do Programa de Pós-Graduação e está relacionado à macroprocessos de explicitação, gestão e disseminação do conhecimento.

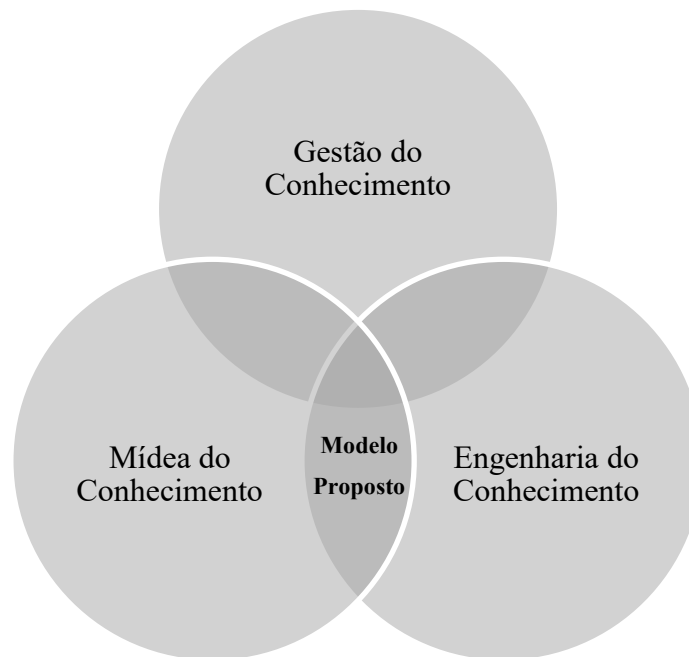
Dentre os macroprocessos estudados pelo Programa, esta tese trabalha na gestão, formalização, codificação e explicitação do conhecimento de domínio. A formalização e codificação presente no modelo desenvolvido estão baseadas em uma ontologia voltada ao domínio em questão e na análise de agrupamento, por meio da formação de grupos de ideias semelhantes alinhadas ao contexto organizacional. A explicitação do conhecimento ocorre por meio de técnicas de mineração de dados e texto, bem como por meio da visualização de informação. Assim, a gestão do conhecimento ocorre na promoção e suporte ao processo de avaliação e auxílio à tomada de decisão para avaliação de ideias.

A pesquisa está situada ainda no âmbito interdisciplinar que, segundo o autor Repko (2008), busca avançar as fronteiras disciplinares para criar novos conhecimentos com base em problemas de pesquisa em comum. Vale salientar que as atividades atribuídas à área de Engenharia do Conhecimento relacionam-se com as demais áreas de concentração do programa, Gestão do Conhecimento e Mídia do Conhecimento nos seguintes aspectos:

- Pesquisa e desenvolvimento de metodologias de identificação, representação e gestão de conhecimento;
- Aplicação de sistemas de conhecimento à gestão do conhecimento organizacional (formalização, memória e tomada de decisão);
- Aplicação de sistemas de conhecimento à interação homem-máquina, como suporte aos trabalhadores de conhecimento, inclusive na educação;
- Aplicação de sistemas de conhecimento em mundos virtuais interativos visando à melhoria da eficácia e eficiência dos processos de treinamento e capacitação (EGC, 2017).

A Figura 1 retrata a integração das áreas de concentração do Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento envolvidas na pesquisa.

Figura 1 – Integração das áreas de concentração com o tema de pesquisa.



Fonte: Autora

1.7.2 Contexto Estrutural do EGC

A Engenharia de Conhecimento surgiu a partir da Inteligência Artificial. Seu principal produto eram os sistemas baseados em conhecimento. Produto este intensificado após a década de 90, momento em que o conhecimento tornou-se um elemento estratégico para as organizações.

Com a popularidade dos sistemas baseados em conhecimento, criou-se a necessidade de uma abordagem sistêmica para o desenvolvimento e criação dos sistemas (STUDER, BENJAMINS; FENSEL, 1998).

Segundo Studer, Benjamins e Fensel (1998), a Engenharia do Conhecimento trabalha com a aplicação de teorias, métodos e ferramentas direcionadas à modelagem e representação do conhecimento de determinado domínio. Tem como objetivo auxiliar em tarefas intensivas em conhecimento e assim fornecer elementos para a Gestão do Conhecimento (SCHREIBER et al., 2000).

Ainda segundo Schreiber (2007), a EC provê orientações sobre qual o momento e os procedimentos para aplicar técnicas de representação de conhecimento para resolver problemas específicos.

Para tal, os objetivos da área de Engenharia do Conhecimento incluem:

A pesquisa e o desenvolvimento de técnicas e ferramentas para a formalização, codificação e gestão do conhecimento; métodos de análise da estrutura e processos conduzidos por profissionais em atividades de conhecimento intensivo; pesquisa e desenvolvimento de sistemas de conhecimento (EGC, 2017).

No que diz respeito ao escopo desta tese, o aspecto que a contextualiza na área de Engenharia do Conhecimento reside no fato do modelo possuir como objetivo a materialização, principalmente dos macroprocessos de explicitação, formalização/codificação e gestão do conhecimento.

Ao desenvolver um sistema de conhecimento voltado ao domínio de gestão de ideias, em que o conhecimento é representado por intermédio de uma ontologia, acredita-se prover auxílio à avaliação de ideias alinhadas ao contexto organizacional, contribuindo na tomada de decisão por especialistas da organização.

1.7.3 Referências Factuais

Na área de Representação Semântica foram encontrados 39 trabalhos, sendo os mais atuais:

- BOTELHO, Maurício. Aplicação de Ontologias na Organização de Conteúdos para Apoio à Equipes de Desenvolvimento de *Software*. Dissertação (Mestrado) - Programa de Pós-graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2015.
- BRAGLIA, Israel. Um Modelo Baseado em Ontologia e Extração de Informação como Suporte ao Design Instrucional na Geração de Mídias do Conhecimento. Tese - Programa de Pós-graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2014.

- CARDENAS, Yuri Gomes. Modelo de Ontologia para Representação de Jogos Digitais de Disseminação do Conhecimento. Dissertação (Mestrado) - Programa de Pós-graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2014.
- CECI, Flavio. Um Modelo Baseado em Casos e Ontologia para Apoio a Tarefa Intensiva em Conhecimento de Classificação com Foco na Análise de Sentimento. Tese- Programa de Pós-graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2015.
- JUNIOR, Vanderlei Freitas. Ontologia para Representação Semântica de Indicadores de Desempenho Considerando Aspectos de Vaguidade, Temporalidade e Relacionamento entre Indicadores. Tese- Programa de Pós-graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2016.
- PEREIRA, Larissa Mariany Freiberger. OGDPub: uma ontologia para publicação de dados abertos governamentais. Dissertação (Mestrado) - Programa de Pós-graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2017.
- SÉRGIO, Marina Carradore. Um Modelo Baseado em Ontologia e Análise de Agrupamento para Suporte à Gestão de Ideias. Dissertação (Mestrado) - Programa de Pós-graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2016.
- SILVA, Thales do Nascimento da. Um Modelo Baseado em Ontologia para Suporte à Tarefa Intensiva em Conhecimento de Recomendação. Dissertação (Mestrado) - Programa de Pós-graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2015.

No domínio de Gestão de Ideias foram encontrados oito trabalhos aderentes ao tema, sendo estes:

- ALVAREZ, Guilherme Martins. Análise de Agrupamentos e Mineração de Opinião como Suporte à Gestão de Ideias. Dissertação (Mestrado) - Programa de Pós-graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2018.

- DOROW, Patrícia Fernanda. O Processo de Geração de Ideias para Inovação: Estudo de Caso em uma Empresa Náutica. Dissertação (Mestrado) - Programa de Pós-graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2013.
- MIGUEZ, Viviane Brandão. Uma Abordagem de Geração de Ideias para o Processo de Inovação. Dissertação (Mestrado) - Programa de Pós-graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2012.
- PRADA, Charles A. Proposta de modelo para o gerenciamento de portfólio de inovação: modelagem do conhecimento na geração de ideias. Dissertação (Mestrado) - Programa de Pós-graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2009.
- RIBEIRO, Alessandro Costa. Modelo de Reconhecimento de Padrões em Ideias usando Técnicas de Descoberta de Conhecimento em Textos. Dissertação (Mestrado) - Programa de Pós-graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2016.
- ROCHADEL, Willian. Identificação de Critérios para Avaliação de Ideias: Um Método Utilizando Folksonomias. Dissertação (Mestrado) - Programa de Pós-graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2016.
- SÉRGIO, Marina Carradore. Um Modelo Baseado em Ontologia e Análise de Agrupamento para Suporte à Gestão de Ideias. Dissertação (Mestrado) - Programa de Pós-graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2016.
- VALDATI, Aline de Brittos. Processo de seleção de ideias em empresas inovadoras. Dissertação (Mestrado) - Programa de Pós-graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2017.

Na área de Análise de Agrupamento os trabalhos relacionados foram:

- ALVAREZ, Guilherme Martins. Análise de Agrupamentos e Mineração de Opinião como Suporte à Gestão de Ideias. Dissertação (Mestrado) - Programa de

Pós-graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2018.

- CECI, Flávio. Um Modelo Semiautomático Para a Construção e Manutenção de Ontologias a partir de bases de documentos não estruturados. Dissertação (Mestrado) - Programa de Pós-graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2010.
- SÉRGIO, Marina Carradore. Um Modelo Baseado em Ontologia e Análise de Agrupamento para Suporte à Gestão de Ideias. Dissertação (Mestrado) - Programa de Pós-graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2016.

Diante das referências apresentadas, a presente tese é aderente ao Programa por propor um modelo baseado em conhecimento. Compreende-se que o trabalho está de acordo com a área de concentração de Engenharia do Conhecimento e tem como objetivo promover suporte a área de Gestão de Ideias, além de possui trabalhos anteriores que abordam temáticas similares.

1.8 ESTRUTURA DO TRABALHO

O presente trabalho é composto por oito capítulos descritos a seguir:

- O primeiro capítulo apresenta o tema e os objetivos desta pesquisa, suas delimitações e aderência ao programa de pós-graduação;
- O segundo capítulo é composto pelo referencial teórico, abordando os assuntos principais relacionados à pesquisa como: Gestão de Ideias, Mineração de Dados e Texto, Ontologia e Trabalhos Correlatos;
- No capítulo 3 é apresentada a metodologia de pesquisa empregada no desenvolvimento do trabalho;
- No capítulo 4 é apresentada a ontologia proposta, com a descrição detalhada das classes, propriedades de dados e propriedades de objetos;
- No quinto capítulo apresenta-se a definição dos índices, com a descrição detalhada dos cálculos;
- No capítulo 6 é apresentado o modelo proposto, com a descrição completa das etapas que o compõem;
- No sétimo capítulo apresenta-se a análise e discussão dos resultados alcançados por meio dos cenários de estudo;

Por fim, as considerações finais desta tese e as perspectivas de trabalhos futuros. Por conseguinte, são disponibilizadas as referências utilizadas na pesquisa e os apêndices.

2 FUNDAMENTAÇÃO TEÓRICA

Os tópicos apresentados neste capítulo fundamentam a pesquisa elaborada, envolvendo: Contextualização, Definição de Ideia, Gestão de Ideias, Sistemas de Gestão de Ideias, Inovação Aberta, Critérios de Seleção, Ontologia, Mineração de Dados e Texto e Trabalhos Correlatos. A seguir será apresentado o referencial teórico que envolve cada um dos tópicos.

2.1 CONTEXTUALIZAÇÃO

Segundo Chesbrough (2003b), Tidd, Bessant e Pavitt (2005) e Koulopoulos (2011), a inovação pode ser caracterizada como um processo de mudança com valor mensurável, assumindo dimensão estratégica, no qual as ideias são transformadas em negócios viáveis.

Aplicar a gestão de ideias, segundo Chesbrough (2012), gera baixo custo, pois é possível buscar uma gama de novos conhecimentos a partir da comunidade global, visto que há disponibilidade de conhecimentos em bancos de dados científicos públicos e artigos *online* em diferentes áreas do conhecimento. Para o autor, existe “um cenário de fartura de conhecimentos, ideias novas e estimulantes que podem surgir a partir de inúmeros lugares” (CHESBROUGH, 2012, p. 127). A utilização do conhecimento disponível em rede configura um novo modelo de gestão, caracterizado como um modelo de inovação aberta (CHESBROUGH, 2003b). A inovação aberta “é um paradigma que assume que as empresas podem e devem usar ideias internas e externas para o mercado” (BÜCHELER; SIEG, 2011, p. 328).

Produzir inovações torna-se uma estratégia fundamental de criação de valor e de competitividade (COHN; KATZENBACH; VLAK, 2008; KOULOPOULOS, 2011; PRIEM; CARR 2012). Esta estratégia intensificou-se com a *web 2.0* e propiciou a criação de um mundo virtual de inteligência coletiva, gerando conexão e colaboração entre multidões para o desenvolvimento do processo inovador. As organizações passaram a desenvolver plataformas para coletar ideias externas, visando impulsionar o processo de inovação interno (ADAMCZYK; BULLINGER; MOESLIN, 2011).

Esta abordagem envolve o conceito de *crowdsourcing* que, segundo Howe (2006), transforma o processo de inovação mais colaborativo. Neste sentido, os atores deste processo participam com sugestões para determinados problemas, contribuindo no processo de inovação (MURAH et al., 2013).

Contudo, o envolvimento de colaboradores no processo de gestão de ideias não é uma prática recente. No passado, as organizações abriam seus processos de inovação, solicitavam sugestões e ideias de clientes, funcionários e membros por meio de “caixas de sugestões” físicas localizadas em áreas comuns (FAIRBANK; WILLIAMS, 2001).

O surgimento de tecnologias sociais e colaborativas baseadas na *web* transformou esses mecanismos para coletar recomendações dos colaboradores em Sistemas de Gerenciamento de Ideias (IMS) ativos, sofisticados e dedicados. Mecanismos estes que permitem que as pessoas proponham ideias, além de classificar e realizar comentários sobre ideias de outros usuários (SALDIVAR et al., 2019).

2.2 DEFINIÇÃO DE IDEIA

Segundo Gurteen (1998), uma ideia consiste em algo que ainda não foi testado, comprovado e realizado. Michalko (2003) defende que uma ideia é caracterizada por um pensamento produtivo, é olhar as mesmas coisas sob uma nova linha de pensamento em busca de novas alternativas. Segundo Tidd, Bessant e Pavitt (2005), são conceitos ou pensamentos transformados em artefato tangível, em invenção. Para Koen et al. (2002), é a solução inicial para um problema elaborado.

A ideia é fruto do pensamento humano e pode ser expressa de diferentes maneiras, tais como: opinião, plano, imagem, prospecção, previsão e avaliação sobre um assunto específico (LEE, 2012). Para Kruse et al. (2013), ideia é uma imagem formada na mente humana, escondida em grandes quantidades de dados.

Thorleuchter, Van Den Poel e Prinzie (2010) definem ideia como uma imagem existente ou formada na mente, mas que pode ser escrita como informação textual. Os autores afirmam que uma ideia é a combinação de um meio e um fim apropriado. Para os autores, ideias podem ser a base de avanços tecnológicos. Thorleuchter e Van Den Poel (2012) entendem como a combinação de várias palavras técnicas que aparecem juntas em um padrão textual.

No contexto da inovação, uma ideia pode ser definida como “uma oportunidade de criar valor para novos investimentos” (KORNISH; ULRICH, 2014, p.15). A ideia, ao ser bem sucedida, pode ser considerada como inovadora (THORLEUCHTER; VAN DEN POEL, 2016). Inovações envolvem ideias de produtos e serviços, novas formas de usá-los, novas aplicações inexploradas, apoiadas pela análise de tendências emergentes e de cenários que estão em mudança (KANG; KANG, 2009). Neste sentido, o processo de inovação inicia com uma nova ideia e a coloca em prática para obter um produto inovador que possa ser introduzido com

sucesso no mercado e que aumente o valor para a organização ou para o cliente (MCKEOWN, 2008).

Infelizmente, o potencial inovador de uma nova ideia só pode ser estimado depois, após o produto correspondente ser bem sucedido e introduzido no mercado (THORLEUCHTER; VAN DEN POEL, 2016). Segundo os autores, atualmente uma alta porcentagem de inovações falha porque as ideias não inovadoras são selecionadas como ponto de partida. Com isso, o processo de inovação consome tempo e custo (DISSELKAMP, 2015) e, portanto, os tomadores de decisão na área da gestão da inovação precisam de estimativas anteriores do potencial inovador de uma ideia (THORLEUCHTER; VAN DEN POEL, 2016).

Entretanto, ideias constituem vários problemas conceituais. Liu, Goulding e Brailsford (2015) realizaram pesquisas e elencaram quatro problemas principais. Primeiramente, o termo ideia em si não é totalmente bem definido a partir de uma perspectiva de compreensão. Em segundo lugar, novas ideias são baseadas no conhecimento de domínio, o que torna difícil a formalização. Em terceiro lugar, as ideias de domínios exibem características amplamente variáveis. Por último, os métodos para ideação são difíceis de codificar. Essas questões implicam que a obtenção de soluções perfeitas para problemas sem a entrada humana não é realista (LIU; GOULDING; BRAILSFORD, 2015).

2.3 GESTÃO DE IDEIAS

O processo de Gestão de Ideias é utilizado pelas organizações para promover a inovação. Segundo Saldivar et al. (2019), a gestão de ideias refere-se ao processo de coleta, desenvolvimento e seleção de ideias para desenvolver produtos ou serviços novos e inovadores, ou para melhorar os existentes, com o objetivo de envolver os colaboradores na inovação. Para Bothos, Apostolou e Mentzas (2012), a gestão de ideias pode ser definida como o “processo de reconhecer a necessidade de ter ideias, gerá-las e avaliá-las”.

O processo de gestão de ideias inclui as fases de preparação, geração de ideias, melhoria, avaliação, implementação e implantação (GOCHERMANN; NEE, 2019). Neste sentido, a gestão de ideias consiste no processo de captura, armazenamento e organização de ideias adotadas no processo avançado de *front-end*. Beretta (2019) destaca que gerar um fluxo sustentável de ideias no *Front End* da inovação é fundamental para o desempenho inovador bem-sucedido das empresas.

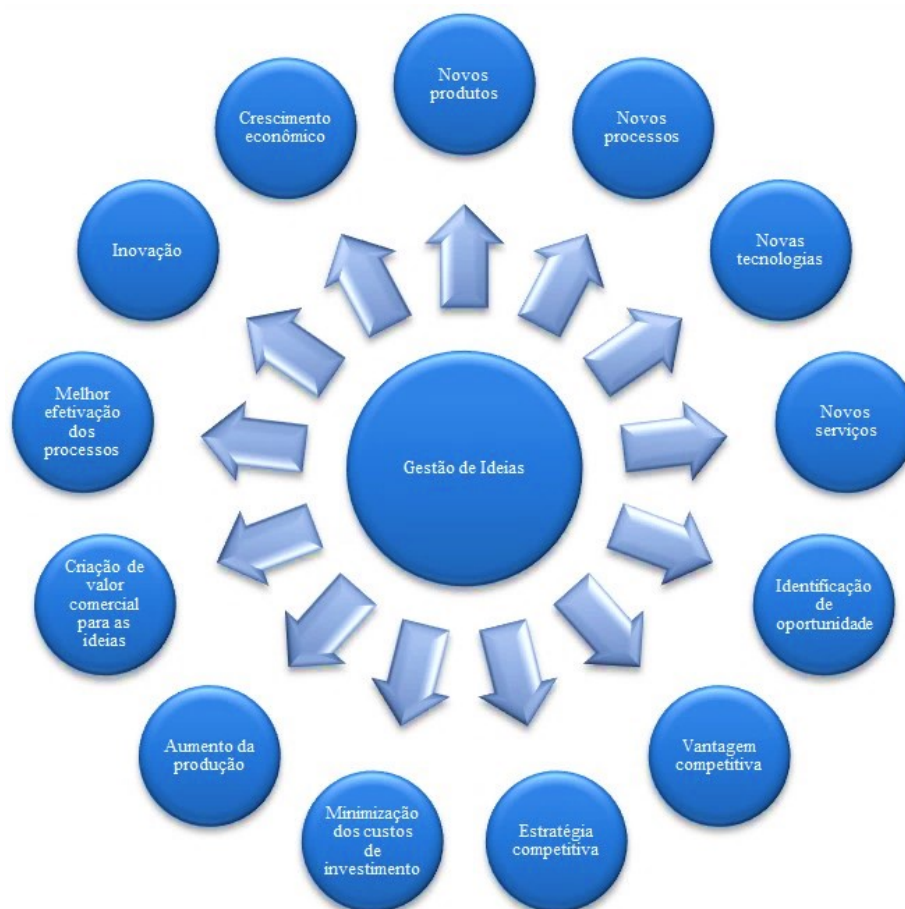
Enríquez et al. (2019) afirmam que as práticas de gestão de ideias favorecem a captura de ideias para análise, executam ações apropriadas e monitoram o progresso da fase de *design* de um produto, servindo para codificar o conhecimento externo, além de envolver questões-chave que tratam da aquisição e processamento de informações. Segundo os autores Martini, Neirotti e Appio (2017), essas práticas permitem as organizações confiarem na criatividade, na atitude empresarial e na motivação de todos os funcionários e colaboradores. Mais especificamente, as práticas de gestão de ideias podem ser conceitualizadas como o processo de reconhecer a necessidade de ideias, subsidiar, gerar, avaliar e priorizar (FLYNN et al., 2003; VANDENBOSCH et al., 2006; GAMLIN; YOURD; PRATICK, 2007).

Segundo Martini, Neirotti e Appio (2017), para que as práticas de busca externas de sugestões de ideias sejam efetivas, as empresas precisam implantar mecanismos internos de integração e sistemas de gestão de ideias. Com isso, funcionários capacitados podem explorar melhor o ambiente, identificar potenciais oportunidades de crescimento de mercado por meio de inovação de produtos e deliberar sobre os processos de colaboração interna e externa que são necessários para desenvolver ideias inovadoras (MARTINI; NEIROTTI; APPIO, 2017).

Em outras palavras, as práticas de gestão de ideias podem fornecer a infraestrutura necessária para executar pesquisas externas adequadamente, e garantir que as empresas se adaptem às oportunidades e tenham processos flexíveis de criação de estratégias (MARTINI; NEIROTTI; APPIO, 2017). Para os autores, a probabilidade de encontrar a recombinação de conhecimento mais vantajosa e competitiva, com base no conhecimento externo, é maior quando as práticas de gestão de ideias funcionam não apenas como catalisadoras, mas também como organizadoras de atenção, *expertise* e incentivos. Neste sentido, a gestão sistemática de ideias fornece o suporte necessário para transformá-las em projetos (VARINY et al., 2019).

A gestão de ideias pode trazer vários benefícios para a organização. Na Figura 2 são apresentados alguns dos benefícios.

Figura 2 – Benefícios da aplicação do processo de Gestão de Ideias



Fonte: Sérgio (2016)

2.4 SISTEMAS DE GESTÃO DE IDEIAS

Muitas organizações utilizam a *internet* para criar plataformas organizacionais que integram conhecimento para produzir inovações relevantes aos negócios. São sistemas automatizados que permitem armazenar, avaliar e analisar ideias (SALDIVAR et al., 2016).

Fundamental neste processo, as TIC passaram a favorecer o surgimento de redes flexíveis em portais hierárquicos e sistemas holísticos de conhecimento. Ao utilizarem tecnologias presentes na *web 2.0*, as empresas podem optar por interagir com comunidades de inovação ou publicar desafios de inovação em plataformas de inovação (FREY; LÜTHJE; HAAG, 2011).

O surgimento de tecnologias sociais e colaborativas baseadas na *web* transformou esses mecanismos para coletar ideias de clientes (por exemplo, caixas de sugestões) em Sistemas de Gerenciamento de Ideias (IMS) ativos, sofisticados e dedicados, permitem assim

que as pessoas proponham ideias, além de classificar e comentar em sugestões de ideias de outros usuários (SALDIVAR, 2019).

Os sistemas de gestão de ideias, vistos como uma evolução dos esquemas clássicos de sugestões de ideias (HOBER; SCHAARSCHMIDT; VON KORFLESCH, 2019), incluem meios para solicitar, capturar e registrar ideias, ferramentas de colaboração, mecanismos de recompensas e ferramentas para avaliação e seleção de ideias, bem como possibilitam gerar indicadores e relatórios (WESTERSKI; DALAMAGAS; IGLESIAS, 2012).

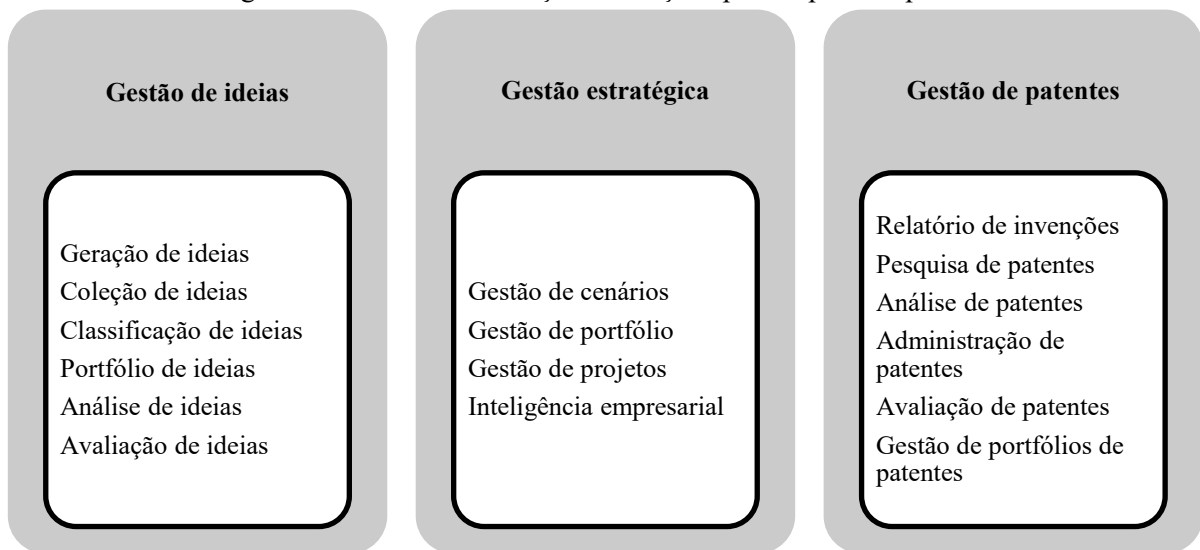
Definidos como parte integrante da gestão da inovação focado na geração, avaliação e seleção de ideias, os sistemas de gestão de ideias contribuem na criação de melhorias e inovações. Um processo estruturado de seleção, avaliação e implementação de ideias possui o potencial de gerar diversos benefícios como maior agilidade e rapidez para levar novos produtos ao mercado (MAUZY, 2006), bem como selecionar as ideias mais promissoras para suas necessidades de inovação ou para resolver um problema (KLEIN; CONVERTINO, 2015).

Adamczyk, Bullinger e Moeslin (2011) reconhecem que todas as funcionalidades reunidas nas plataformas proporcionam maior sustentabilidade econômica às empresas. Para Kohler, Matzler e Füller (2009), as plataformas adicionam novas capacidades, promovendo maior agilidade e redução de custos nos processos de inovação.

Tendo em vista essas vantagens, muitas organizações começaram a utilizar sistemas de gestão de ideias para apoiar atividades de inovação (HOBER; SCHAARSCHMIDT; VON KORFLESCH, 2019) como uma ferramenta sistemática gerenciável para gerar e avaliar ideias (MIKELSONE; VOLKOVA; LIELĀ, 2019a). Assim, diferentes modelos foram propostos na literatura como Hüsigg e Kuhn (2007) e Murah et al. (2013).

Hüsigg e Kuhn (2007) analisaram as inovações apoiadas por sistemas de computadores e propuseram um modelo com diferentes metodologias para identificar novos conhecimentos e capturar ideias. O modelo subdivide-se em Gestão de Ideias, Gestão Estratégica e Gestão de Patentes. Em cada fase do modelo, existem divisões de categorias que tratam a informação de forma detalhada. Segundo os autores, estas três divisões compõem uma visão holística do processo. A Figura 3 apresenta o modelo, sendo a fase de Gestão de Ideias, foco deste trabalho, explanada a seguir.

Figura 3 – Círculo de inovação: inovação apoiada por computador

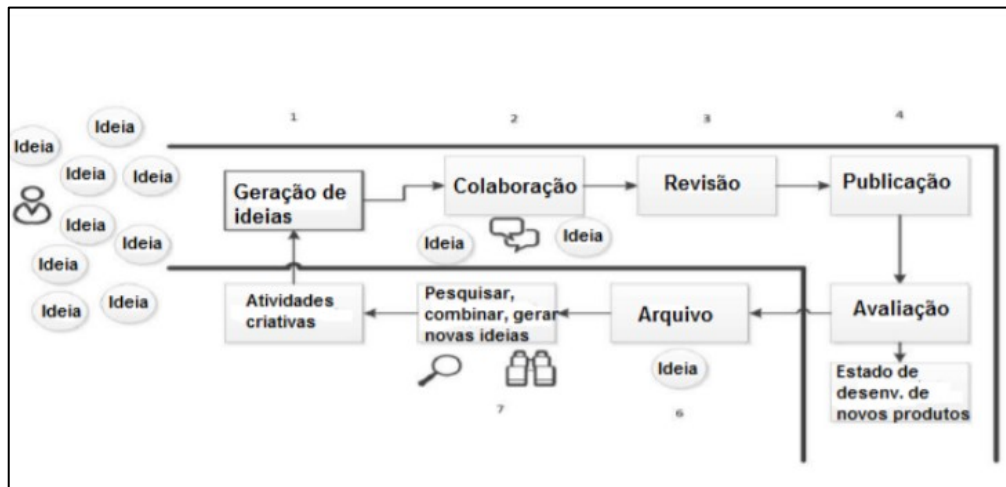


Fonte: Adaptado de Kohn e Hüsigg (2007)

Hüsigg e Kohn (2007) dividiram a fase de Gestão de Ideias em seis subcategorias. Os autores salientam que na subcategoria de geração de ideias são utilizadas, principalmente, ferramentas com base na criatividade como *brainstorming*, *TRIZ* (teoria da solução inventiva de problemas é um exemplo de conhecimento codificado para inovação, que sugere 40 princípios para gerar ideias e encontrar soluções) e ferramentas de mapeamento mental. A segunda subcategoria tem ferramentas para a coleta de ideias, composta principalmente por sistemas de banco de dados baseados na *intranet*, com o objetivo principal de melhorar a eficiência do processo. A terceira subcategoria compreende ferramentas de classificação de ideias que ajudam a classificar as ideias coletadas. O portfólio de ideias inclui ferramentas que oferecem opções sofisticadas para visualizar as ideias. A próxima subcategoria ajuda a analisar as ideias. Essas ferramentas permitem reunir e armazenar dados relevantes. A última categoria contém ferramentas de avaliação de ideias, composta por um sistema integrado que facilita a avaliação por parte de especialistas (HÜSIG; KOHN, 2007).

Os autores Murah et al. (2013), baseados no conceito colaborativo, propuseram a figura de um objeto que pode ser manipulado a partir de uma primeira ideia e reutilizado por um usuário que exerce várias funções. O usuário faz a submissão, convida outros usuários, edita, revisa, avalia e decide se vale a pena publicar ou não a ideia. Assumindo vários papéis, o usuário é fortemente envolvido no processo de Gestão de Ideias, conforme apresentado na Figura 4 (MURAH et al., 2013).

Figura 4 – Arquitetura de sistema conceitual de um sistema de gerenciamento de ideias



Fonte: Stankowitz (2014)

De acordo com Murah et al. (2013), o sistema possui diferentes estados em um único fluxo de trabalho. No estado de edição inicial, o usuário pode editar ou enviar uma ideia. No estado de colaboração, é possível colaborar com outros usuários na criação de novas ideias ou na edição das já existentes e armazenadas no sistema. Na fase de revisão, ocorre o apoio ou rejeição da ideia submetida. Caso a ideia possua avaliação positiva por vários usuários, será publicada, avaliada e arquivada. Se rejeitada, voltará para o estado de colaboração para mudanças e melhorias.

Conceber e desenvolver plataformas para gestão de ideias é complexo, contudo, essas plataformas são fundamentais para apoiar, estimular e facilitar o processo de Gestão de Ideias (MURAH et al., 2013).

Entretanto, apesar da colaboração *online* ter se tornado uma ferramenta poderosa para aumentar a geração de ideias em grandes corporações, o gerenciamento pode sofrer uma sobrecarga de propostas dos funcionários, dificultando o processo de tomada de decisão (LAUTO et al., 2013).

2.5 INOVAÇÃO ABERTA

Os clientes mudam suas necessidades constantemente, o processo tecnológico torna-se mais veloz e a competitividade em nível global faz com que os produtos fiquem obsoletos e vulneráveis mais facilmente (TIDD; BESSANT; PAVITT, 2015). A globalização e a dinamicidade do mercado de negócios exigem investimentos em inovação, mas criar produtos inovadores não é trivial, porém crítico para o sucesso do negócio (KOULOPOULOS, 2011; MOOS et al., 2011; POORKAVOOS et al., 2016).

Para Sener e Hobikoglu (2013), um dos quesitos fundamentais da economia de inovação é permitir que o lucro das ideias externas transformadas em inovações sustente o controle sobre organizações concorrentes. Nessa situação, é necessário criar novos processos ou novas configurações organizacionais voltadas à competitividade. A transformação de ideias em negócios viáveis gera valor mensurável e adquire caráter e rumos estratégicos para as empresas na criação de valor (KOULOPOULOS, 2011).

Contudo, dois fatores travam a habilidade das empresas em promover a inovação: a rapidez da inovação exigida para aumentar e manter o lucro, e a diminuição da capacidade produtiva de inovação interna (NAMBIAN; SAWHNEY, 2011). O conjunto destes fatores produz modificações nos processos internos de inovação. Segundo Chesbrough (2012), existem elementos que contribuem para o que se denomina de erosão da inovação interna, entre eles:

- Alta mobilidade de pessoas capacitadas e experientes, carregando consigo parte do conhecimento adquirido na organização e, possivelmente, compartilhando com outras empresas;
- Crescimento do capital de risco privado, criando novas formas de comercializar ideias;
- Velocidade de comercialização de produtos, reduzindo sua vida tecnológica.

No paradigma de inovação fechada às ideias, invenções, pesquisas e os desenvolvimentos necessários para colocar um produto no mercado, são produzidos dentro da própria organização. Já no paradigma voltado a inovação aberta, pressupõe-se que as organizações podem e devem utilizar ideias externas, além de internas, para facilitar as inovações, compartilhar riscos e melhorar a produtividade e a competitividade (CHESBROUGH; 2006; CHESBROUGH, 2011). Mais recentemente, a ideia de inovação aberta enfatizou ainda mais o valor da aquisição de conhecimento externo, argumentando que é uma estratégia necessária para aumentar a produtividade de P&D diante da crescente concorrência e dos ciclos mais rápidos de desenvolvimento de tecnologia (LACERDA; VAN DEN BERGH, 2020).

A inovação aberta pode ser definida como: “o uso de entradas e saídas de conhecimentos individuais para acelerar a inovação interna, e expandir os mercados para o uso externo da inovação, respectivamente” (CHESBROUGH; VANHAVERBEKE; WEST, 2006). Assim, surgiu da dificuldade das organizações em produzir inovações por meio de seus próprios esforços, emergindo como um paradigma estratégico importante que apoia a inovação contínua dentro da estrutura da organização (PUSTOVRH; RANGUS; DRNOVŠEK, 2020).

Chesbrough (2003a) fez uma comparação entre os modelos de inovação fechada e inovação aberta, e elaborou o seguinte quadro comparativo apresentado no Quadro 2.

Quadro 2 - Comparação entre inovação fechada e inovação aberta

Inovação fechada	Inovação aberta
Pessoas inteligentes trabalham na empresa.	Nem todas as pessoas inteligentes trabalham na empresa, por isso é necessário encontrar e explorar o conhecimento e experiência das pessoas que estão fora da empresa.
Para lucrar com P&D, a empresa necessita pesquisar, e desenvolver o que pesquisa.	P&D externo podem criar valor significativo.
Se a organização realizar suas pesquisas, conseguirá chegar primeiro ao mercado.	Não temos a origem da pesquisa para lucrar com isso.
Se a empresa comercializar primeiramente uma inovação, ela ganha.	Construir um modelo de negócio melhor é mais satisfatório do que chegar primeiro ao mercado.
Se a empresa criar mais e melhores ideias na indústria, ela ganha.	Se a empresa fizer o melhor uso das ideias internas e externas, ela ganha.
A empresa deve controlar a propriedade intelectual (IP) para que os concorrentes não lucrem com suas ideias.	A empresa deve tirar proveito do uso da propriedade intelectual do outro. Comprando a propriedade intelectual do outro, sempre que se puder avançar no modelo de negócio.

Fonte: Adaptado de Chesbrough (2003a)

De acordo com Chesbrough (2012), no modelo tradicional de inovação há pressupostos quanto ao aproveitamento de recursos e de conhecimento advindos da própria empresa, dando ênfase a investimentos em P&D interno e em ideias de profissionais internos à organização. Contudo, a competição no mercado de negócios tornou-se mais complexa, exigindo das organizações maior capacidade de lidar com as incertezas e acompanhar as mudanças rápidas do mercado.

Segundo Sarigianni et al. (2019), a inovação aberta envolve atividades para obter ideias inovadoras de usuários ou agentes fora e dentro das fronteiras organizacionais. Fato que leva as empresas a confiarem no conhecimento externo para desenvolver novos produtos e serviços em vez de executar totalmente o processo de inovação no departamento de P&D.

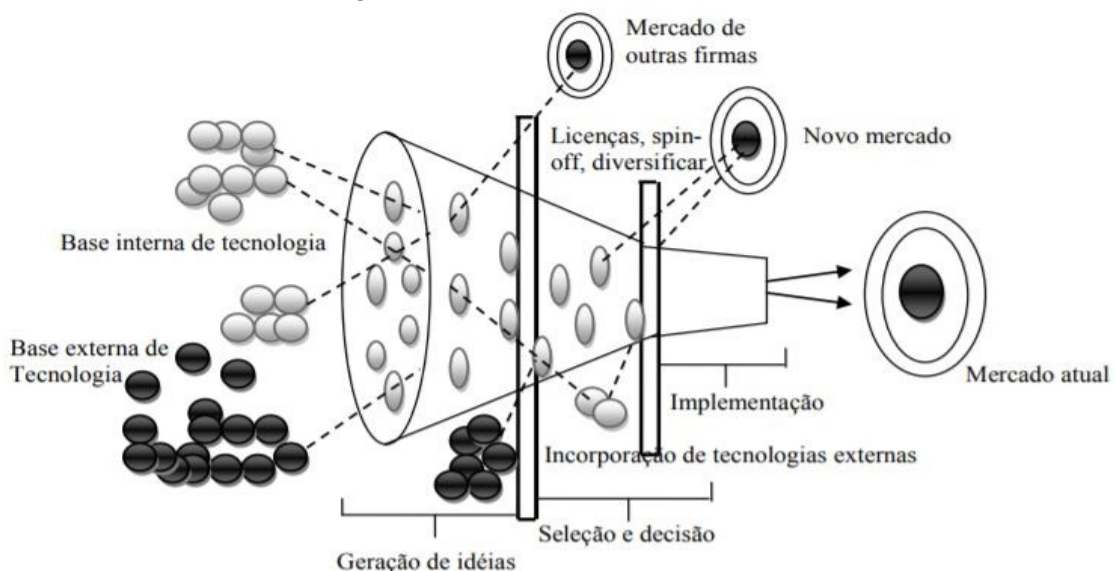
Vários atores podem envolver-se no processo de inovação aberta, incluindo clientes, universidades, fornecedores, concorrentes e usuários (FELIZ; ZENGER, 2014). Apesar disso, quatro elementos-chave são necessários para implementar integralmente os processos de inovação aberta: desenvolvimento de redes, consolidação do intercâmbio de conhecimentos, mecanismos de proteção de propriedade intelectual e modelos de negócios de suporte (MCCORMACK; FALLON; CORMICAN, 2015).

Os benefícios da inovação aberta incluem maior acesso aos especialistas externos, custos mais baixos de desenvolvimento e aprimoramento de tecnologia, menor tempo de

colocação no mercado e produtos de melhor qualidade (PUSTOVRH; RANGUS; DRNOVŠEK, 2020). As organizações abrem seu processo de inovação para integrar a base de conhecimento, acelerar a criatividade, a flexibilidade e obter excelência na produção de conhecimento. Ao explorar a criatividade, as ideias e o conhecimento dos colaboradores em benefício próprio, as organizações identificam melhores soluções para serem implementadas (CARDOSO, 2016). Combinando os diferentes conhecimentos e habilidades dos colaboradores, cria-se valor de forma mais eficaz e eficiente (FROW et al., 2015).

Henry Chesbrough (2012) apresenta um modelo voltado à inovação aberta, cujo conhecimento externo às organizações é utilizado em estratégias de inovação de forma a acelerar o processo de inovação ou prover subsídios para a sua ocorrência, como demonstrado na Figura 5.

Figura 5 - Modelo de inovação aberta



Fonte: Traduzido de Chesbrough (2012)

Wang (2012) salienta que o modelo de inovação aberta representa uma ruptura de valores, pois o conhecimento passa a ser adquirido por meio de parceiros, que em conjunto contraem competências necessárias à inovação em virtude de sua complementaridade.

Os processos de inovação aberta são suportados por uma plataforma *online*. É por meio dela que declarações do problema são distribuídas e os solucionadores de problemas enviam suas soluções (MERGEL, 2015). Os problemas variam de campanhas de *design* simples, por exemplo, projetando um novo logotipo para uma marca ou planejando uma nova campanha publicitária para um produto, até problemas tecnológicos complexos. Fatores que podem levar a sobrevivência de uma organização ou a descoberta de um novo mercado com o objetivo de atrair novos clientes e aumentar a participação de mercado da organização (MERGEL, 2015).

Um exemplo proeminente foi o uso de um concurso com o objetivo de redesenhar o algoritmo de sugestão de filmes do serviço de *streaming* da Netflix®. A equipe vencedora de profissionais estatísticos e matemáticos ganhou o prêmio em dinheiro de 1 milhão de dólares (MERGEL, 2015).

Nos últimos anos, vários métodos e ferramentas foram desenvolvidos, incluindo conceitos de cocriação, inovação baseada no usuário, inteligência coletiva e *crowdsourcing*.

2.5.1 *Crowdsourcing* e Gestão de Ideias

As empresas estão continuamente buscando novas formas de gerar ideias (WESTERSKI; DALAMAGAS; IGLESIAS, 2013) e durante a última década, a propensão a cooperar além das fronteiras corporativas e incorporar conhecimento externo em inovações internas, aumentou na maioria das organizações (MUHDI, 2011). Mudanças no ambiente de negócios, como o aumento da concorrência, a disseminação do trabalho móvel e a necessidade de tomar decisões hábeis rapidamente, alteraram a natureza da tomada de decisões organizacionais (CHIU; LIANG; TURBAN, 2014). Com o intuito de lidar com essas mudanças, novos modelos e ferramentas são empregados nas organizações.

O termo *crowdsourcing* foi cunhado por Jeff Howe em um artigo publicado na revista Wired (HOWE, 2006). Segundo o autor, o termo significa “o ato de fazer um trabalho tradicionalmente realizado por um empregado e terceirizá-lo para um grupo indefinido, geralmente grande, sob a forma de uma chamada aberta”, consolidando-se como uma prática comercial que significa literalmente terceirizar uma tarefa para a multidão. Niederberger et al. (2012) definem *crowdsourcing* como um processo distribuído que envolve tarefas de terceirização para uma rede de pessoas, cada vez mais utilizado por empresas para gerar soluções para problemas de vários tipos.

Renard e Davis (2019) corroboram ao afirmarem que o *crowdsourcing* é uma estratégia organizacional que terceiriza uma função desempenhada pelos funcionários para uma rede indefinida e geralmente grande de pessoas, com o intuito de coletar ideias e contribuições pela internet.

Neste sentido, *crowdsourcing* fornece às empresas uma nova maneira de envolver seus colaboradores em atividades de inovação (KOHLENER; CHESBROUGH, 2019). Ao terceirizar tarefas relacionadas à inovação para grandes multidões externas, as empresas podem acessar ideias criativas diversas (ZOU; ZHANG, LIU, 2015). Além disso, podem alavancar multidões para diferentes aspectos da inovação, incluindo avaliações de conceito, compreensão das

necessidades dos clientes, solução de problemas específicos, levantamento de fundos para a inovação e geração de ideias (PANCHAL, 2015).

Sua utilização na geração de ideias ocorre principalmente por meio de competições de ideias ou *crowdstorm* (HOSSAIN; KAURANEN, 2015). Segundo os autores, nas competições de ideias as organizações ou indivíduos chamam multidões para apresentar novas ideias e as melhores são selecionadas. Para Poetz e Schreier (2012), as multidões podem superar profissionais em muitos níveis de ideação de novos produtos. A diversidade da multidão aumenta significativamente as chances de gerar ideias úteis (HOSSAIN; KAURANEN, 2015). Schweitzer et al. (2012) afirmam que as competições de ideias levam a mais e melhores ideias a um custo menor por ela.

De modo geral, o *crowdstorm* tem sido utilizado pelas empresas para combinar esforços e gerar melhores resultados. O *crowdstorm* é um método derivado da união do *brainstorm* e do *crowdsourcing*, caracterizada como uma técnica de geração de ideias, na qual une os princípios do *brainstorm*, com a ampliação de limites do *crowdsourcing* para estimular o surgimento de inovações (ABRAHAMSON; RYDER; UNTERBERG, 2013). Esta técnica permite a troca de conhecimentos, estimulando a inovação independente de questões geográfica para a captação de ideias. Adicionalmente, pode envolver técnicas alternativas como a gamificação, por meio do uso de jogos (SOUZA et al., 2006).

Coletar ideias através de *crowdsourcing* tornou-se uma prática comum para as empresas se beneficiarem de ideias externas e inovarem (WANG; WANG; TAO, 2017). O aumento das plataformas de *crowdsourcing* para geração e seleção de ideias, atraiu pesquisas sobre suas implicações para resolução de problemas organizacionais e inovação (BRABHAM, 2013). Pesquisadores e profissionais defendem abordagens mais abertas, por meio do envolvimento de indivíduos externos em plataformas de *crowdsourcing* para inovação e ideação (HOSSAIN; ISLAM, 2015a).

Os benefícios destas plataformas incluem: reduções consideráveis de tempo e custos para as organizações, reunião de ideias e apoio à inovação organizacional (CHIU; LIANG; TURBAN, 2014; POETZ; SCHREIER, 2012; RENARD; DAVIS, 2019). Benefícios estes que impactam positivamente na organização, pois negócios envolvendo *crowdsourcing* crescem rapidamente (ESCANDON-QUINTANILLA; JIMENEZ-NARVAEZ; GARDONI, 2015; KOHLER, 2018).

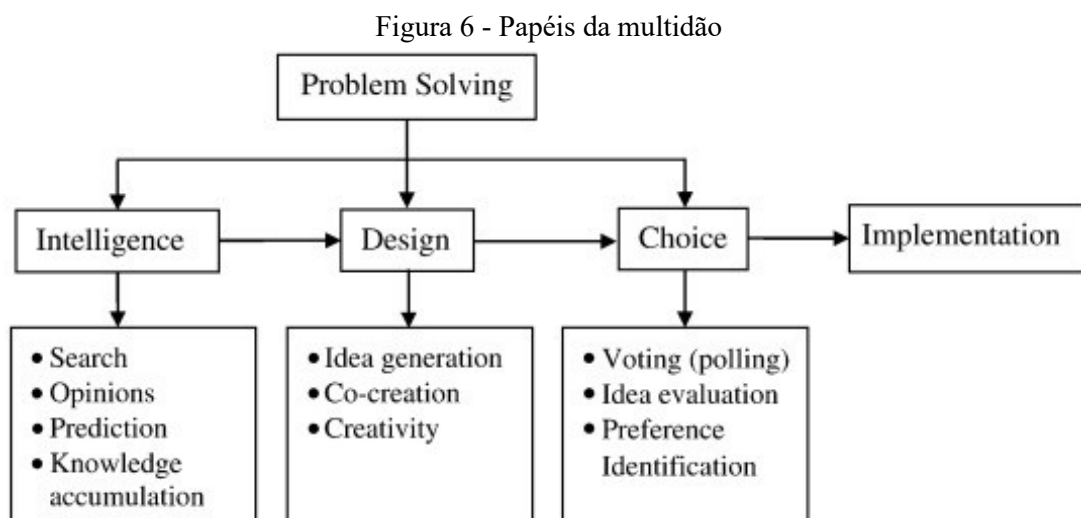
Existem muitas aplicações potenciais de *crowdsourcing*, sendo uma das mais importantes o suporte ao processo organizacional de tomada de decisões (BRABHAM, 2013; HOWE, 2006). Além do auxílio à tomada de decisões, ferramentas de *crowdsourcing* apoiam

a resolução de problemas e a exploração de oportunidades (CHIU; LIANG; TURBAN, 2014). Segundo os autores, um problema de difícil solução para um tomador de decisão pode ser resolvido por uma multidão, contudo, as multidões podem gerar um número elevado de ideias na tentativa de resolver o problema. Para os autores Prpić et al. (2015), as organizações empregam *crowdsourcing* para construir, assimilar e aproveitar o capital da multidão que será utilizado posteriormente para tomar decisões; realizar pequenas tarefas (*microtask*), dividindo um projeto em pequenas partes; inovar e resolver problemas.

Chan, Li e Zhu (2015) destacam que o relacionamento colaborativo entre clientes e empresas presentes nestas ferramentas pode abranger vários processos de negócios, desde ideação de produtos ou serviços até o seu lançamento no mercado, compondo todas as fases do processo de gestão de ideias. A importância da interação dos clientes para a ideação não é um fenômeno novo, mas a disponibilidade generalizada da *internet* aumentou significativamente a capacidade de interação entre empresas e clientes (HOSSAIN; ISLAM, 2015a). Dado o potencial de participação que pode ser alcançado, possibilitando às organizações olharem além das suas fronteiras e assim compartilharem conhecimentos, a abordagem de *crowdsourcing* torna-se cada vez mais interessante (ESCANDON-QUINTANILLA; JIMENEZ-NARVAEZ; GARDONI, 2015).

Bonazzi, Viscusi e Barbey (2017), afirmam que o *crowdsourcing* é tradicionalmente considerado adequado para fornecer diferentes tipos de suporte ao processo de tomada de decisão, especialmente na fase de *design*, geração de ideias e cocriação; na fase de escolha, por meio de votação, bem como na fase de inteligência para a exploração das informações.

A Figura 6 ilustra os exemplos de papéis que uma multidão pode representar em diferentes fases de decisão.



Fonte: Chiu, Liang e Turban (2014)

Muitas empresas, incluindo a Unilever[®], Starbucks[®] e Dell[®], patrocinam comunidades de *crowdsourcing* de ideias para capturar sugestões de ideias diretamente de seus clientes e efetivamente alavancar conhecimentos, habilidades e recursos da multidão (CHAN; LI, ZHU, 2015). Por exemplo, a AT&T[®] recebeu aproximadamente 28 mil ideias entre 2009 e 2014 por meio de seu sistema de gerenciamento de ideias e alocou US \$ 44 milhões para financiar as mais promissoras (HOBER; SCHAARSCHMIDT; VON KORFLESCH, 2019). Outro exemplo foi o lançamento do Fiat Mio[®], com base em mais de 11 mil ideias enviadas por mais de 17 mil usuários em todo o mundo (RENARD; DAVIS, 2019).

Aproveitar a criatividade de uma multidão pode fornecer um meio altamente eficiente e eficaz de adquirir ideias, trabalho e conteúdo para resolver problemas (BEER et al., 2017). A multidão é usada como fonte de trabalho, geração de ideias ou identificação de problemas, sendo um termo geral utilizado para descrever uma variedade de maneiras pelas quais as organizações, particularmente no setor com fins lucrativos, aproveitam os pensamentos, insumos e ideias de seus colaboradores (CLARK et al., 2016). Tem por objetivo melhorar o processo de tomada de decisões, completar tarefas pesadas ou contribuir na cocriação de projetos (CHIU; LIANG; TURBAN, 2014).

Apesar dos potenciais benefícios do *crowdsourcing*, houve pouco progresso nas áreas diretamente relacionadas à pesquisa sobre o suporte à tomada de decisões gerenciais (CHIU; LIANG; TURBAN, 2014). Tomadores de decisão, segundo Alksher et al. (2016), podem não ter tempo suficiente ou recursos cognitivos para examinar centenas de milhares de ideias potenciais.

O grande desafio da área mencionado por Hossain e Islam (2015a) é realmente a tarefa de avaliar as ideias e encontrar uma equipe apropriada para tal tarefa, pois muitas vezes as ideias excelentes que são propostas podem não ser implementáveis por uma empresa (MARTÍNEZ-TORRES, 2013). Isto ocorre especialmente quando um número elevado de ideias é gerado em um curto período de tempo (HOSSAIN; ISLAM, 2015a). Este volume também pode gerar atrasos significativos de projetos. Por exemplo, em seu recente projeto de inovação, a empresa Google[®] teve que recrutar 3 mil funcionários para analisar 150 mil ideias recebidas da multidão, atrasando em nove meses o projeto (ALKSHER et al., 2016). De modo geral, muitas empresas possuem dificuldades em gerenciar todo o processo e identificar quais ideias recebidas das multidões são apropriadas e passíveis de serem implementadas (BOUDREAU; LAKHANI, 2013).

Os autores Chiu, Liang e Turban (2014) ressaltam que a avaliação de ideias inclui três questões principais: a seleção de avaliadores adequados para avaliar a qualidade do resultado

do processo; a aferição de qualidade que envolvem os critérios reais para avaliar as ideias e, por último, as métricas de avaliação. Contudo, para apoiar a avaliação efetiva, o desenvolvimento e o uso apropriado da tecnologia são essenciais (CHIU; LIANG; TURBAN, 2014).

2.6 CRITÉRIOS DE SELEÇÃO DE IDEIAS

Clientes em particular são os principais responsáveis em sustentar qualquer negócio e, portanto, são vitais para qualquer organização (HOUSSAIN; ISLAM, 2015b). Uma pesquisa recente, com grandes empresas situadas nos EUA e na Europa, mostrou a importância dada pelas organizações às plataformas de inovação aberta, envolvendo clientes no processo de geração de ideias (CHESBROUGH; BRUNSWICKER, 2013). As organizações utilizaram plataformas *online* para envolver clientes, fornecedores, funcionários, parceiros e cidadãos, no desenvolvimento de um ecossistema para geração de ideias (GOUILLART; BILLINGS, 2013). Entretanto, o envolvimento entre os clientes exige considerável investimento relacionado ao tempo da equipe de gerenciamento da plataforma (HOUSSAIN; ISLAM, 2015b).

O avanço na utilização de comunidades de inovação para geração de ideias, por meio de uma multidão, emerge como um novo método de inovação (LEE; SEO, 2013). Segundo Lee e Seo (2013), o objetivo da multidão é gerar ideias que possam ser sementes de inovação e entender as características de ideias selecionadas pelo usuário e pela organização, podendo levar ao aumento na eficácia do processo de geração de ideias. Os autores também extraíram características do conteúdo de ideias como subjetividade, negatividade, privacidade e profundidade de ideias; características relacionadas aos usuários como experiência e reputação; examinar quais são os fatores que afetam a seleção de ideias. Uma análise de 71.134 ideias do MyStarbucksIdea.com mostrou que há diferenças significativas entre ideias selecionadas pelo usuário e selecionadas pela organização.

Na maioria dos casos, ideias de contribuintes externos à organização precisam ser refinadas antes de serem colocadas em prática. Identificar ideias implementáveis é desafiador e envolve diferentes etapas, tais como rastreamento, avanço, refino, suporte e testes (GOUILLART; BILLINGS, 2013). O processo de avaliar uma ideia, por exemplo, é extremamente intenso e requer grande quantidade de tempo, energia, paciência e imaginação (JOURET, 2009; HOUSSAIN; ISLAM, 2015b). Um grande fluxo de ideias, em muitos casos, pode comprometer o processo de ideação, pois uma empresa pode não possuir a capacidade de absorção necessária para processar todas as ideias propostas (DI GANGI; WASKO, 2009;

HOUSSAIN; ISLAM, 2015b). Além disso, um desafio crucial é determinar quais ideias selecionar (KING; LAKHANI, 2013).

Stevanović et al. (2012) definiram o conceito de critério de seleção como algo que deve medir alguma coisa importante e ser independente. Os autores mencionam que os critérios são a base para definir regras para um *ranking* de valores aplicáveis a cada uma delas. Critérios de seleção também são denominados indicadores para medir as ideias (LI et al., 2012).

Como critérios para avaliação e seleção de ideias, Asensio (2002) propõe uma abordagem bidimensional baseada em escores comparativos de “valor” e “esforço” para cada ideia apresentada. Os critérios para determinar o valor de uma ideia seriam, por exemplo, alinhamento com a missão da empresa, probabilidade de sucesso, impacto econômico e grau de novidade. Segundo o autor, os critérios de esforço podem ser a quantidade de recursos humanos e financeiros necessários, os efeitos na estrutura da organização e o tempo a ser dedicado. Em termos mais gerais, os critérios para análise de ideias podem incluir o ineditismo da proposta, sua utilidade, viabilidade técnica e econômica e os benefícios potenciais traduzidos em resultados (rentabilidade, eficiência, agilidade, redução de tempo e de custos) para a empresa, seus clientes ou funcionários (BAKKER, 2010).

Schuurman et al. (2012) evidenciaram um exemplo relacionado às cidades. Representantes de uma cidade avaliaram todas as ideias selecionadas em três critérios: inovação, benefício para o usuário e viabilidade (SCHUURMAN et al., 2012). Desta forma, o processo de ideação poderia ser avaliado, bem como os resultados da seleção dos usuários versus os especialistas independentes. A primeira dimensão denominada inovação foi interpretada como o grau em que a ideia é nova em comparação com outros produtos ou serviços existentes em determinado contexto. A segunda dimensão envolveu os benefícios para os usuários, definida como à medida que essa ideia atende às necessidades ou preocupações dos visitantes e habitantes da cidade, oferecendo certo valor agregado a eles. Na terceira e última dimensão denominada viabilidade, considerou-se os aspectos técnicos, questões legais, políticas e os aspectos econômicos da ideia (SCHUURMAN et al., 2012).

Mechant et al. (2012) empregaram critérios de seleção de ideias também para o contexto de cidades inteligentes. Eles utilizaram uma escala de cinco pontos com 12 critérios. Destes, oito critérios se referiam à localização, tempo real, acessibilidade, potencial verde, potencial para tecnologia de sensores, viabilidade, inovação e potencial de valorização que atribuíram um peso maior devido à sua grande relevância para o projeto. Os outros quatro critérios se referiam ao valor agregado para as diversas partes interessadas (residentes, turistas,

usuários empresariais e governos). No total, 38 especialistas preencheram a ferramenta de avaliação, cada um avaliando 28 ideias. A entrada resultou em um índice.

Outro critério é a qualidade da ideia que, segundo Poetz e Schreier (2012), pode ser medida usando três variáveis-chave: a novidade da ideia em comparação com os produtos do mercado alvo existentes; o valor da ideia em termos de sua capacidade de resolver o problema e assim criar benefícios para o cliente; a viabilidade de uma ideia em termos de facilidade que poderia ser traduzida em um produto comercial (os avaliadores consideram os aspectos técnicos e econômicos ao avaliar a viabilidade de uma ideia). Apesar de serem um pouco mais detalhados, esses procedimentos de avaliação reproduzem realisticamente o processo de tomada de decisão, segundo os autores. Já os autores Blohm et al. (2011) propuseram a utilização de quatro dimensões distintas para medir a qualidade da ideia: novidade, viabilidade, relevância e elaboração.

Para Murukannaiah, Ajmeri e Singh (2016), uma ideia pode ser avaliada também quanto a sua criatividade. Esta, por sua vez, de acordo com uma definição amplamente aceita (RUNCO; JAEGER, 2012), traz novidade e utilidade. Contudo, Murukannaiah, Ajmeri e Singh (2016) reconhecem que a clareza é crucial uma vez que as classificações de ideias pouco claras confundem a análise. Segundo os autores pode-se definir:

- Clareza: um requisito claro é inequívoco e fornece um nível apropriado de detalhes;
- Utilidade: um requisito útil leva a produtos que fornecem valor ou utilidade aos seus usuários. Segundo Abdala et al. (2017), o termo utilidade significa: se a solução gerada atende aos requisitos do problema e pode ser facilmente aplicada;
- Novidade: um novo requisito é algo que um usuário encontra de original e inesperado, ou seja, algo que não é comum, mundano ou convencional.

Toubia e Netzer (2016) também elencam a criatividade como critério. Os autores elencaram e definiram os critérios: criatividade (por exemplo, “quão criativo é esta ideia?”); interesse de compra (por exemplo, “qual a probabilidade de você adquirir?”); popularidade (por exemplo, “quão popular é?”); e qualidade da escrita (por exemplo, “a descrição é bem escrita?”). Cada item possui associada uma escala Likert de cinco pontos.

Contudo, a seleção de uma ideia, segundo Hossain e Islam (2015a), pode depender de vários outros fatores, como a quantidade de votos, os comentários, os pontos ganhos, o volume de ideias sugeridas, a relevância e a viabilidade de uma ideia para implementação e, sobretudo, o alinhamento de uma ideia à estratégia empresarial da organização. Entretanto, os fatores

principais que funcionam como catalisadores para selecionar ideias são limitados (HOUSSAIN; ISLAM, 2015b).

Lauto et al. (2013) destacam o papel fundamental da área de P&D na escolha das propostas a serem desenvolvidas com base em sua viabilidade em termos de tecnologia, mercado, ajuste organizacional e recursos de desenvolvimento. Para os autores há alguns critérios fundamentais que são necessários na tomada de decisão como:

- Viabilidade técnica: representa o quanto a ideia parece viável com base em fatos científicos e/ou técnicos comprovados de artigos publicados ou conhecimento inédito;
- Originalidade: condiz com o que a ideia é realmente surpreendente, algo nunca percebido anteriormente. Dean et al. (2006) definem originalidade como o grau em que uma ideia é nova e incomum. Para os autores Abdala et al. (2017) originalidade significa que a solução gerada possui recursos exclusivos ainda não projetado para um problema;
- Necessidade do cliente: diz respeito ao quanto a ideia está relacionada a uma necessidade comprovada do cliente;
- Recursos: relaciona-se com o conhecimento e as competências necessárias para implementar a ideia e se estes estão disponíveis ou se a empresa possui a capacidade de adquiri-los;
- Potencial de vendas: se a ideia tem alto potencial de vendas em longo prazo;
- Vantagem competitiva: A ideia oferece vantagens únicas em relação às soluções dos concorrentes;
- *Gut feel*: o que o sentimento de avaliação do avaliador expressa.

Segundo Hossain e Islam (2015b), as ideias exclusivas são implementadas mais rapidamente do que as relacionadas com outras ideias, no entanto, ideias únicas precisam de mais tempo, mais comentários e pontos de implementação para serem colocadas em prática.

Outra forma se refere ao mapeamento de conceitos que ajuda a solicitar e organizar ideias de especialistas e partes interessadas (VALDES-VASQUEZ; KLOTZ, 2013). Este método pode ser usado para categorizar os resultados em vários *clusters* que podem ter explicações concorrentes ou *backgrounds*. Permite também comparar os resultados entre o grupo de participantes. O método integra tarefas qualitativas de geração e classificação de ideias com duas análises quantitativas, dimensionamento multidimensional e análise de *cluster* (VALDES-VASQUEZ; KLOTZ, 2013). Segundo os autores, uma das descobertas relevantes

do uso dessas análises é a criação de mapas visuais que podem ajudar a desenvolver planos de ação e estruturas. Eles dividiram o método básico de mapeamento de conceitos em seis etapas. Primeiramente, prepara o projeto e estabelece o foco da pesquisa, identificando um grupo preliminar de participantes e agendando a logística do estudo. Em segundo lugar, o método requer participantes para gerar ideias em relação ao tópico da pesquisa. Em terceiro lugar, é necessária a contribuição dos mesmos participantes ou de diferentes para classificar a lista final em grupos conforme o julgamento de cada participante, de acordo com escalas de classificação predeterminadas. Em quarto lugar, os dados obtidos a partir da classificação são usados para gerar resultados visuais bidimensionais, como mapas de pontos e aglomerados, determinados usando escala multidimensional e análise de agrupamentos hierárquico, respectivamente. A combinação dos dados de classificação dos participantes com mapas permite a geração de representações gráficas tridimensionais. Em quinto lugar, esses mapas resultantes são interpretados pelo grupo de participantes ou pelos pesquisadores, dependendo dos objetivos da pesquisa. Segundo os autores, essas interpretações servem como insumos para o desenvolvimento de planos de ações, estudos de avaliação ou estruturas.

2.7 ONTOLOGIA

O conceito original da palavra Ontologia vem da Filosofia e possui relação com a existência do ser ou aos tipos de existência (BORST, 1997). Segundo o dicionário Aurélio “é a parte da filosofia que trata do ser, isto é, do ser concebido como tendo uma natureza comum que é inerente a todos e a cada um dos seres” (FERREIRA, 1986).

Poli e Obrst (2010) corroboram ao salientarem que o termo é visto com duas óticas diferentes: a primeira relacionada à filosofia, mencionada anteriormente e a segunda relacionada à Ciência da Computação, utilizada inicialmente na Inteligência Artificial e nos dias atuais empregada também na Engenharia do Conhecimento.

Na Inteligência Artificial o termo está associado ao raciocínio exercido por um sistema de inteligência artificial para desempenho de determinada tarefa, descrevendo formalmente um domínio de conhecimento.

Gruber (1993) foi um dos pioneiros a atribuir significado ao termo, definindo-o como a exploração e representação dos relacionamentos semânticos e conceituais da informação. Para Guarino (1998), constitui-se em uma especificação das propriedades e relações entre os conceitos de um determinado domínio.

Contudo, uma das principais definições do termo fundamentada em Gruber (1993) e Borst (1997) é proposta por Studer, Benjamins e Fensel (1998) ao afirmarem que: “a ontologia é a representação formal e explícita de uma conceitualização compartilhada”.

Grimm, Hitzler e Abecker (2007) destacam que as ontologias são modelos conceituais para representar determinado domínio, sendo interpretável por máquina com o auxílio das técnicas de representação de conhecimento. Para Suárez-Figueroa et al. (2011), a ontologia possui como foco capturar o conhecimento consensual de forma genérica, podendo ser reutilizada e compartilhada entre aplicações de *software* e por grupos de pessoas.

Chandrasekaran, Josephson e Benjamins (1998) corroboram ao destacarem a aplicabilidade do uso de ontologias, afirmam que elas elucidam a estrutura de conhecimento e permitem o compartilhamento de conhecimento.

Ao desenvolver uma ontologia é necessário explicitar componentes de domínio como: descrição de categorias e objetos e as relações entre os dados envolvidos no processo (LULA; PALIWODA-PEKOSZ, 2008). Segundo os autores Suárez-Figueroa *et al.* (2011), classes, relacionamento, axiomas e instâncias formam o conjunto de componentes que compõem a representação do conhecimento de domínio. Maedche (2002) salienta que a descrição de uma ontologia deve ser composta por cinco tuplas: conceitos, relacionamentos, hierarquia de conceitos, funções que relacionam conceitos e um conjunto de axiomas.

Gómez-Pérez (1999) define e apresenta os cinco elementos da ontologia, conforme consta no Quadro 3.

Quadro 3 - Elementos da ontologia

Elemento	Definição
Classes	Correspondem ao componente básico de uma ontologia. Um conjunto de classes formam uma taxonomia.
Relações	Representam a interação entre as classes de um domínio.
Funções	Caso especial de relacionamento estabelecendo uma única relação com outro elemento.
Axiomas	Regras utilizadas para modelar sentenças sempre verdadeiras.
Instâncias	Representam indivíduos específicos de uma determinada classe, ou seja, os próprios dados.

Fonte: Autora

Ontologias podem assumir variados formatos, incluindo fundamentalmente: um vocabulário de termos e sua especificação, as definições dos relacionamentos entre os conceitos, resultando na estruturação do domínio e suas restrições (JASPER; USCHOLD, 1999).

Neste sentido, segundo Almeida e Bax (2003), ontologias otimizam processos de recuperação da informação. Podem ser divididas de duas formas, conforme o grau de

formalismo: representando uma taxonomia ou um vocabulário controlado (*light-weight*) ou ainda especificando relacionamentos e restrições lógicas entre os conceitos (*heavy-weight*) (BATZIAS, SIONTOROU, 2012).

Os autores Almeida e Bax (2003) propuseram uma especificação detalhada dos tipos de ontologia, apresentada no Quadro 4.

Quadro 4 - Tipos de ontologias

Abordagem	Classificação	Definição
Quanto à função Mizoguchi, Vanwelkenhuysen e Ikeda (1995)	Ontologias de domínio	Reutilizáveis no domínio, fornecem vocabulário sobre conceitos, seus relacionamentos, sobre atividades e regras que os governam.
	Ontologias de tarefa	Fornecem um vocabulário sistematizado de termos, especificando tarefas que podem ou não estar no mesmo domínio.
	Ontologias gerais	Incluem um vocabulário relacionado a coisas, eventos, tempo, espaço, casualidade, comportamento, funções, etc.
Quanto ao grau de formalismo Uschold e Gruninger (1996)	Ontologias altamente informais	Expressa livremente em linguagem natural.
	Ontologias semi-informais	Expressa em linguagem natural de forma restrita e estruturada.
	Ontologias semiformais	Expressa em linguagem artificial definida formalmente.
	Ontologia rigorosamente formal	Os termos são definidos com semântica formal, teoremas e provas.
Quanto à aplicação Jasper e Uschold (1999)	Ontologias de autoria neutra	Um aplicativo é escrito em uma única língua e depois convertido para uso em diversos sistemas, reutilizando-se as informações.
	Ontologias como especificação	Cria-se uma ontologia para um domínio que é usada para documentação e manutenção no desenvolvimento de softwares.
	Ontologia de acesso comum à informação	Quanto ao vocabulário é inacessível, a ontologia torna a informação inteligível, proporcionando vocabulário compartilhado dos termos.
Quanto à estrutura Ontologia de alto-nível Haav e Lubi (2001)	Ontologia de alto-nível	Descrevem conceitos gerais relacionados a todos os elementos da ontologia (espaço, tempo, matéria, objeto, evento, ação, etc.) os quais são independentes do problema ou domínio.
	Ontologia de domínio	Descrevem o vocabulário relacionado ao domínio, como, por exemplo, medicina ou automóveis.
	Ontologia de tarefa	Descrevem uma tarefa ou atividade, como por exemplo, diagnósticos ou compras, mediante

		inserção de termos especializados na ontologia.
Quanto ao conteúdo VanHeijst, Schreiber e Wielinga (2002)	Ontologias terminológicas	Especificam termos que serão usados para representar o conhecimento em um domínio (por exemplo, os léxicos).
	Ontologias de informação	Especificam a estrutura de registros de bancos de dados (por exemplo, os esquemas de bancos de dados).
	Ontologias de modelagem do conhecimento	Especificam conceituações do conhecimento, tem uma estrutura interna semanticamente rica e são refinadas para uso no domínio do conhecimento que descrevem.
	Ontologias de aplicação	Contém as definições necessárias para modelar o conhecimento em uma aplicação.
	Ontologias de domínio	Expressam conceituações que são específicas para um determinado domínio do conhecimento.
	Ontologias genéricas	Similares às ontologias de domínio, mas os conceitos que as definem são considerados genéricos e comuns a vários campos.
	Ontologias de representação	Explicam as conceituações que estão por trás dos formalismos de representação do conhecimento.

Fonte: Almeida e Bax (2003, p. 10)

Este trabalho utiliza o conceito de ontologia baseado na visão da Ciência da Computação e da Engenharia do Conhecimento, visualizando a ontologia como um modelo de dados para modelagem e representação do conhecimento de domínio.

2.7.1 Ontologias e Gestão de Ideias

Sérgio (2016) identificou por meio das bases de dados *Scopus*, *ACM*, *IEEE* e *Springer Link*, os principais estudos envolvendo soluções semânticas e gestão de ideias. Dos 21 artigos coletados, seis apresentavam uma solução semântica para o domínio de Gestão de Ideias. Realizando uma nova revisão da literatura, foram encontrados mais quatro trabalhos e somente um relacionado ao tema. Para esta revisão, as mesmas bases de dados da primeira pesquisa foram utilizadas.

Xie e Zhang (2010) propuseram um sistema de gerenciamento de ideias, promovendo a criação de equipes. Como resultado, os autores desenvolveram um IMS (Sistema de Gestão de Ideias, do inglês, *Idea Management System*) capaz de realizar o reconhecimento de uma ideia criativa por meio de análise semântica latente, da seleção de uma ideia e da evolução, além de sua visualização.

Sint et al. (2010) desenvolveram uma ferramenta denominada *Ideator* para apoiar o processo de gestão de ideias, capaz de prover suporte à geração de ideias de maneira colaborativa, baseada na semântica promovida pelo *framework* KiWi.

No trabalho de Westerski, Iglesias e Rico (2010) utilizam-se tecnologias da *Web Semântica* para promover interoperabilidade entre Sistemas de Gestão de Ideias e *softwares* heterogêneos. O modelo propõe como e quais tipos de anotações de metadados devem ser aplicados no domínio de Sistemas de Gestão de Ideias.

Westerski e Iglesias (2011) criaram um modelo de geração de dados abertos, ligados para a *web*. Como resultado alcançado, existe a promoção da interligação de dados estruturados para o uso em sistemas voltados à Gestão de Ideias, em face da quase inexistência de sistemas de Gestão do Conhecimento dedicados à inovação.

Westerski, Iglesias e Rico (2010) apresentam uma ontologia de opinião, demonstram os benefícios da publicação na *web* e os resultados da mineração de opinião de uma forma estruturada.

Poveda, Westerski e Iglesias (2012) desenvolveram um modelo, projeto e arquitetura baseados em uma busca semântica para sistemas de inovação aberta com foco em sistemas de Gestão de Ideias. O sistema permite coletar, organizar e buscar ideias, melhorando a interação entre usuários e simplificando o processo de análise de ideias (POVEDA; WESTERSKI; IGLESIAS, 2012).

Os quatro últimos trabalhos apresentados fazem parte de um mesmo projeto denominado GI2MO. Esta plataforma corresponde a um sistema de gerenciamento de ideias de código aberto, baseado na arquitetura de componentes do Sistema de Gerenciamento de Conteúdo Drupal® (POVEDA; WESTERSKI; IGLESIAS, 2012). O objetivo é organizar e configurar tecnologias semânticas no contexto de Gestão de Ideias.

Ao analisar a ontologia proposta pelo projeto GI2MO, observou-se que ela não provê o suporte ao alinhamento do contexto organizacional.

Por fim, ao realizar uma nova busca nas bases de dados, identificou-se o trabalho de Sérgio, Souza e Gonçalves (2017). O trabalho propõe um modelo baseado em ontologia e análise de agrupamentos para apoiar o gerenciamento de ideias, com o objetivo de contribuir no processo de tomada de decisão. A demonstração de viabilidade do modelo foi realizada com a construção de um protótipo. A ontologia proposta nesta tese é uma evolução desta proposição, incluindo agora o contexto organizacional e os índices que compõem o processo de avaliação de ideias.

2.8 MINERAÇÃO DE DADOS E TEXTO

Todos os dias, as pessoas produzem e criam ideias que podem contribuir no desenvolvimento de novos produtos, serviços, formação e promulgação de novas políticas (KHAN et al., 2014). Para sobreviver no mercado, os tomadores de decisão, incluindo os investidores e suas equipes gerenciais, devem tentar continuamente perceber as necessidades e exigências não ditas pelos clientes, descobrindo seus padrões comportamentais (KARIMI-MAJD; MAHOOTCHI, 2015). A descoberta de padrões coloca os tomadores de decisão em uma posição melhor, em que os serviços mais qualificados podem ser projetados e fornecidos (KARIMI-MAJD; MAHOOTCHI, 2015).

Para identificar um número relativamente pequeno de potenciais ideias perante o volume, os tomadores de decisão precisam identificar as informações textuais-chave que representam as ideias (DINH et al., 2015). Contudo, devido ao volume de dados textuais gerados, identificar informações textuais de forma manual é considerado um processo dispendioso e demorado (DINH et al., 2015). Poetz e Schreier (2012) corroboram ao afirmarem que organizações como a Dell® ou a Lego®, por exemplo, podem ser confrontadas com um alto volume de ideias advindas da comunidade de usuários e não poder filtrar e selecionar as ideias mais promissoras devido ao esforço de seleção.

Segundo Dinh et al. (2015), a grande massa de dados não estruturados gerados pelas plataformas *online* de inovação aberta, dificulta o processo de avaliação, ressaltando a relativa escassez de estudos de casos envolvendo tais plataformas.

Neste sentido, na última década, as técnicas de mineração de dados emergiram como ferramentas poderosas para descobrir padrões latentes em grandes conjuntos de dados (KARIMI-MAJD; MAHOOTCHI, 2015). O processo de mineração de dados, também chamado de descoberta de conhecimento em bancos de dados (KDD), baseia-se na extração do conceito de palavras ocultas no texto ou documento (ALKSHER et al., 2016).

Caracterizada como subárea da Mineração de Dados, a mineração de texto, que também pode ser conhecida como Análise de Texto Inteligente, Mineração de Dados de Texto ou Descoberta de Conhecimento em Texto (do inglês, *Knowledge Discovered in Texts* - KDT) refere-se ao processo de extração de informações e conhecimento interessante e não trivial de dados não estruturados (GUPTA, 2009). Este processo é baseado na extração de padrões latentes em conteúdo não estruturado (ALKSHER et al., 2016).

A mineração de texto tornou-se uma aliada no processamento de ideias, pois as pessoas são limitadas na capacidade de processar informações, e o número de ideias produzidas por

uma multidão certamente excede essa capacidade, gastando-se muito tempo para analisar grandes volumes de ideias (CHAN; DANG; DOW, 2016). Extrair informações relevantes a partir de textos torna-se uma nova fonte de informação para executivos e tomadores de decisão (WANG; OHSAWA; 2013, THORLEUCHTER; VAN DEN POEL, 2015).

2.8.1 Técnicas de Análise Matricial

Métodos da Álgebra Linear e da Estatística são cada vez mais utilizadas nas mais variadas aplicações computacionais. A ascendência da Álgebra Linear e da Estatística na área de Computação está relacionada à robustez com o qual os métodos abordam problemas computacionais comuns a diferentes domínios, bem como a formalização matemática provida por meio de seus conceitos.

2.8.1.1 Singular Value Decomposition (SVD)

Na álgebra linear, a decomposição de valores singulares (SVD), do inglês (*Singular Value Decomposition*), refere-se à fatoração de uma matriz real ou complexa.

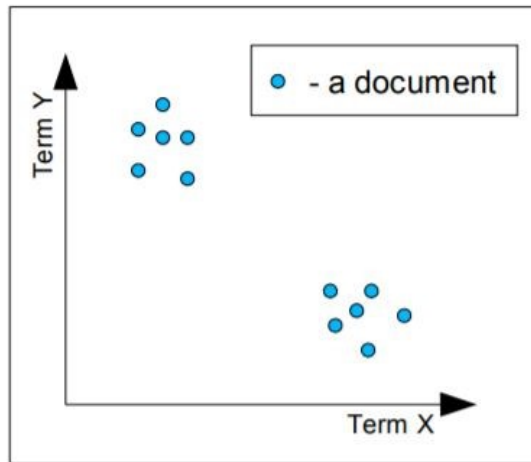
SVD é um método computacional concebido para decompor *arrays* de dados (ZEKRI; MOKHTARI; COHEN, 2016). O método SVD transforma um conjunto de variáveis correlacionadas em um conjunto de variáveis derivadas não correlacionadas, e quantifica as relações (principalmente correlações) entre as variáveis originais e os objetos (STEWART, 1993).

Em razão de suas propriedades algébricas, SVD é comumente utilizado com os objetivos de obter eficiência na manipulação de dados matriciais, redução da dimensionalidade dos dados e remoção de ruídos dos dados originais (WALL et al., 2003).

SVD divide uma matriz A ($t \times d$) em três matrizes U , Σ e V , de modo que $A = U\Sigma V^T$. U representa $t \times t$, a matriz ortogonal cujos vetores de coluna são chamados de vetores singulares esquerdos de A , V é uma matriz ortogonal $d \times d$, cujos vetores de coluna são chamados vetores singulares de A , e Σ representa $t \times d$, matriz diagonal tendo os valores singulares de A em ordem decrescente ao longo de sua diagonal. A classificação r_A da matriz A é igual ao número de valores singulares que não são zero. As primeiras colunas r_A de U formam uma base ortogonal para o espaço de coluna de A (KLOPOTEK; WIERZCHON; TROJANOWSKI, 2013).

Osiński, Stefanowski e Weiss (2004) explicam da seguinte forma: há uma matriz de documentos de termos onde A é a representação dos documentos como vetores de pesos característicos, conforme a Figura 7.

Figura 7 - Matriz termo x documento



Fonte: Osiński, Stefanowski e Weiss (2004)

A matriz A é decomposta usando SVD para adquirir uma base ortogonal no espaço de recursos multidimensional, conforme representa a Figura 8.

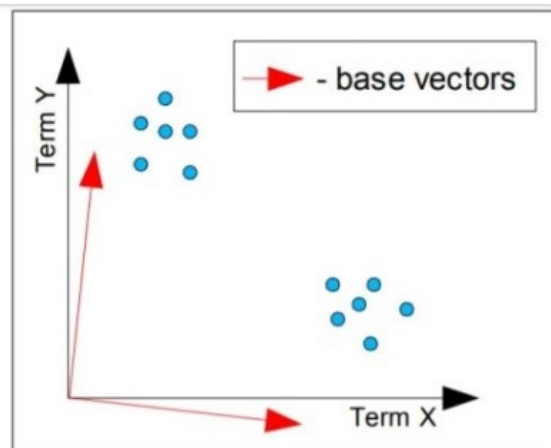
Figura 8 - Matriz decomposta

$$\underbrace{\begin{bmatrix} * & * \\ * & * \\ * & * \end{bmatrix}}_A = \underbrace{\begin{bmatrix} * & * & * \\ * & * & * \\ * & * & * \end{bmatrix}}_U \underbrace{\begin{bmatrix} * & \\ & * \end{bmatrix}}_S \underbrace{\begin{bmatrix} * & * \\ * & * \end{bmatrix}}_{V^T}$$

Fonte: Osiński, Stefanowski e Weiss (2004)

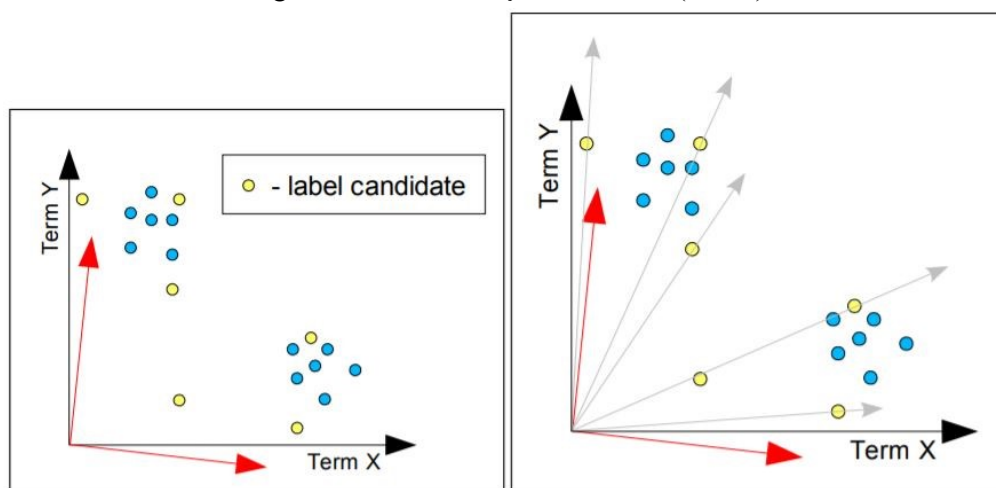
Há a seguinte representação da matriz decomposta na Figura 9.

Figura 9 – Representação da matriz decomposta



Fonte: Osiński, Stefanowski e Weiss (2004)

Os candidatos do rótulo do *cluster* são expressos no mesmo espaço vetorial que os documentos na matriz A . Assim, pode-se calcular facilmente a sua semelhança com os primeiros vetores k da base, conforme mostra a Figura 10.

Figura 10 – Identificação de rótulos (*labels*)

Fonte: Osiński, Stefanowski e Weiss (2004)

2.8.1.2 *Latent Dirichlet Allocation (LDA)*

A origem dos modelos baseados em tópico é na Indexação Semântica Latente (do inglês *Latent Semantic Indexing* - LSI) (DEERWESTER et al., 1990). LSI, segundo Deerwester et al. (1990), serviu de base para o desenvolvimento de um modelo de tópicos. Entretanto, LSI não é um modelo probabilístico, portanto, não é um modelo de tópico autêntico. Em 2001, Hornann, Bach e Blei propuseram a Análise Semântica Probabilística Latente (do inglês *Probabilistic Latent Semantic Analysis* - PLSA), caracterizada por um modelo de tópicos

genuíno. Após a PLSA, Blei et al. (2003) fizeram a proposição da Alocação Latente de *Dirichlet* (do inglês *Latent Dirichlet Allocation* - LDA), um modelo generativo probabilístico mais completo sendo visto como uma extensão do PLSA.

Muitos modelos probabilísticos baseiam-se em LDA por meio da combinação com tarefas específicas (LIU et al., 2016). Segundo Liu et al. (2016), todos os modelos de tópicos acima mencionados foram inicialmente introduzidos na comunidade de análise de texto para descoberta de tópicos não supervisionados em um corpus de documentos.

Pesquisadores introduziram essa abordagem nos campos da mineração de documentos biológicos e médicos, devido sua superioridade na análise de coleções de documentos em larga escala (LIU et al., 2016).

A ideia geral por trás de algoritmos de modelagem de tópicos, como o LDA, baseia-se no pressuposto de que um documento é uma combinação de tópicos. Mais precisamente na análise de texto os documentos geralmente são representados como uma matriz de termos-documentos onde as palavras são os recursos de nível único dos documentos.

Cada documento é modelado como uma mistura de tópicos e cada um deles é uma distribuição de probabilidade discreta que define a probabilidade de cada palavra aparecer em um determinado tópico (ALGHAMDI; ALFALQI, 2015). Estas probabilidades de tópicos fornecem uma representação concisa de um documento (ALGHAMDI; ALFALQI, 2015).

Para obter outro nível de recursos de documentos de texto, o LDA pode ser usado para inferir os tópicos latentes desses documentos. Como resultado, obtém-se uma representação de dois níveis de documentos de texto usando palavras como recursos e tópicos individuais como recursos de grupo. Em outras palavras, enquanto o modelo de espaço vetorial é usado para obter o primeiro nível de representação, um modelo de tópico pode ser usado para obter o segundo nível de representação.

2.8.1.3 *Latent Semantic Indexing (LSI)*

Outra técnica que pode ser utilizada para identificar tópicos latentes (ou conceitos) de uma coleção de documentos é a Indexação Semântica Latente (do inglês *Latent Semantic Indexing* - LSI).

LSI utiliza a decomposição de valor singular da matriz X para identificar um subespaço linear no espaço de características de *tf-idf*, que captura a maior parte da variância na coleção (BLEI et al., 2003). Esta abordagem pode conseguir compactação significativa em grandes coleções de documentos (BLEI et al., 2003). Segundo Deerwester et al. (1990), as

características derivadas do LSI são combinações lineares dos recursos originais de *tf-idf* e podem capturar alguns aspectos das noções linguísticas básicas, como sinonímia e polissemia.

Esta técnica projeta consultas a documentos em um espaço com dimensões semânticas latentes. LSI tenta superar os problemas de combinação léxica usando índices conceituais derivados estatisticamente, em vez de palavras individuais para recuperação (DEERWESTER et al., 1999). Segundo os autores, LSI pressupõe que existe uma estrutura subjacente ou latente no uso das palavras parcialmente obscurecido pela variabilidade na escolha das palavras. Para tal, utiliza-se da técnica de SVD para estimar a estrutura no uso de palavras em todos os documentos. A recuperação é realizada usando o banco de dados de valores e vetores singulares obtidos (ROSARIO, 2000). Os dados de desempenho mostram que esses vetores estatisticamente derivados são indicadores de significado mais robustos do que termos individuais (ROSARIO, 2000).

Segundo Rosario (2000), no espaço semântico latente, uma consulta a um documento pode ter alta semelhança de cosseno mesmo que não compartilhem quaisquer termos, desde que seus termos sejam semanticamente semelhantes em um sentido a ser descrito mais tarde. A métrica de similaridade utilizada em LSI é uma alternativa às medidas de sobreposição de palavras como *tf-idf*, sendo LSI um método para redução da dimensionalidade. Para Rosario (2000), o espaço semântico latente é escolhido de tal forma que as representações no espaço original são alteradas o mínimo possível quando mensurado pela soma dos quadrados das diferenças.

2.8.2 Análise de Agrupamento

Identificar informações relevantes sobre dados não estruturados de alta dimensão como texto é um grande desafio, sendo necessários algoritmos adequados para encontrar ou gerar recursos interpretáveis que possam contribuir de forma significativa durante a análise dos dados (ZHAO et al., 2015). Segundo o autor, isso se deve principalmente à dimensionalidade e ao problema da esparsidade, pois uma mesma coleção de documentos textuais pode conter inúmeras palavras únicas, e essas palavras geralmente são tratadas como recursos para representar documentos de texto pelo modelo de espaço vetorial. Zhao et al. (2015) afirmam que este é um desafio para as técnicas atuais de mineração de texto como a análise de agrupamento.

O intuito na formação de agrupamentos de informações textuais é separar vários documentos dispostos de forma desorganizada em um conjunto de grupos que contenham

documentos com assuntos semelhantes. Isto facilita a localização de documentos em uma dada coleção que estejam relacionados.

A fim de fornecer uma representação significativa dos dados textuais e melhorar a eficiência dos algoritmos de análise de texto, muitas técnicas foram propostas para agrupar palavras que tenham semântica semelhante em conceitos ou tópicos.

A tarefa de agrupamento de texto é um método apropriado utilizado para particionar grandes volumes de documentos de texto em grupos (ABUALIGAH; KHADER, 2017). Estas tarefas têm sido aplicadas em muitos domínios na área de mineração de texto, como recuperação de texto, categorização de texto, entre outras (LI et al., 2014; ABUALIGAH; KHADER, 2016).

Neste sentido, os autores Westerski e Kanagasabai (2019) realizaram pesquisas e definiram que os algoritmos de agrupamento *k-means* e *K-medoids* são adequados ao contexto de Gestão de Ideias. As análises também indicaram a medida estatística *tf-idf* como adequada à normalização dos vetores e à utilização da medida de similaridade do cosseno para estabelecer a pertinência de ideias em um determinado agrupamento. Estes elementos são descritos em detalhes nas seções a seguir.

2.8.2.1 Representação e Normalização de Documentos

Um documento é entendido como “qualquer material que sirva de base para estender nosso conhecimento, disponível para estudo ou comparação” (SCHÜRMEYER, 1935). A crescente importância dos dados não estruturados como uma das formas mais comuns de dados, inevitavelmente os tornou uma importante fonte de análise (KIM; KIM; CHO, 2017).

Kim, Kim e Cho (2017) explicam que para aplicar técnicas de aprendizagem de máquina e mineração de dados, é necessário que os dados brutos dos documentos sejam transformados em vetores numéricos, nos quais as características definidoras de cada documento são capturadas. Se esses vetores de documentos podem preservar a proximidade adequada entre os documentos e suas características únicas, os algoritmos de mineração de texto subsequentes podem extrair informações mais precisas, ocultas e valiosas a partir dos dados (KIM; KIM; CHO, 2017).

Segundo os autores Kim, Kim e Cho (2017) para extrair padrões e *insights* interessantes de dados textuais, o passo fundamental e crucial é a representação de documentos.

2.8.2.2 Vector Space Model (VSM)

O modelo de espaço vetorial (VSM) é um método de representação textual amplamente utilizado na classificação de documentos. No entanto, continua a ser um problema desafiador a questão da alta dimensionalidade. Uma tentativa de reduzir este problema leva em consideração técnicas de redução de dimensionalidade, no entanto, tais técnicas têm deficiências como a perda de algumas informações importantes (AL-ANZI; ABUZEINA, 2018).

VSM é caracterizado como um modelo algébrico utilizado para realizar a representação dos documentos de texto. Durante muito tempo, o VSM foi aplicado em áreas como recuperação de informação e processamento de linguagem natural, a fim de representar documentos como vetores de palavras (OSIŃSKI; STEFANOWSKI; WEISS, 2004).

Esta técnica transforma o problema de comparar dados textuais em um outro de comparação de vetores algébricos em um espaço multidimensional. Após realizar a transformação, as operações de álgebra linear são utilizadas para calcular semelhanças entre os documentos (OSIŃSKI; STEFANOWSKI; WEISS, 2004). Segundo os autores, toda palavra da coleção de documentos compõe uma dimensão separada no VSM e cada documento é representado por um vetor abrangendo todas essas dimensões.

Esta relação é melhor expressa como uma matriz $t \times d$, geralmente chamada de matriz de termo-documento, onde t é o número de termos únicos e d é a quantidade de documentos (OSIŃSKI; STEFANOWSKI; WEISS, 2004). O elemento aij da matriz A é, portanto, uma representação numérica da relação entre o termo i e o documento j . Existem muitos métodos para calcular o aij , comumente referidos como métodos de ponderação do termo (OSIŃSKI; STEFANOWSKI; WEISS, 2004).

Após construir a matriz A , a distância entre os vetores que representam os documentos a e b pode ser calculada de diferentes maneiras (OSIŃSKI; STEFANOWSKI; WEISS, 2004). Segundo Osiński, Stefanowski e Weiss (2004) a medida mais comumente utilizada é o cosseno entre os vetores a e b , sendo definida como:

$$\cos\theta = \frac{\sum_{i=1}^n (w1_i * w2_i)}{\sqrt{\sum_{k=1}^n (w1_k)^2} * \sqrt{\sum_{j=1}^n (w2_j)^2}} \quad (1)$$

A equação mede o ângulo do cosseno formado pelos vetores que representam os documentos (ideias) contendo os termos e seus pesos. Segundo Jones e Furnas (1987), a equação pode variar de peso 1.0 ($\cos 0^\circ = 1.0$) para vetores apontando na mesma direção, peso 0.0 ($\cos 90^\circ = 0.0$) para vetores ortogonais e peso -1.0 ($\cos 180^\circ = -1.0$) para vetores apontando em direções opostas.

2.8.2.3 Normalização utilizando *tf-idf*

A estrutura dos documentos textuais necessita da aplicação de técnicas especializadas para assim poderem ser analisadas por sistemas computacionais, devido ao significado implícito atribuído a cada palavra na linguagem humana (SABOL et al., 2009).

Computacionalmente, um documento é representado como um vetor, e cada dimensão é representada por um termo constante no documento.

Tf-idf significa a frequência do termo–inverso da frequência nos documentos, (do inglês, *term frequency–inverse document frequency*), é uma medida estatística que objetiva indicar a importância de uma palavra em um documento com relação a uma coleção de documentos (LESKOVEC; RAJAMARAN; ULLMAN, 2014). Esta medida estatística é comumente utilizada nas áreas de recuperação de informação e mineração de dados e texto.

Essencialmente, o *tf-idf* funciona determinando a frequência relativa das palavras em um documento específico em comparação com a proporção inversa dessa palavra em todo o corpus do documento. Intuitivamente, esse cálculo determina a importância de uma determinada palavra em um documento específico (RAMOS, 2003). As palavras que são comuns em um único ou um pequeno grupo de documentos tendem a ter números *tf-idf* mais altos do que palavras comuns, como artigos e preposições.

Basicamente, determinada dimensão possui um peso associado obtido a partir da equação *tf-idf*, que estabelece a relação entre a quantidade de vezes que o termo é mencionado no documento pela quantidade de documentos do corpus que mencionam o mesmo. Define-se assim:

- Frequência do termo *i* em um documento *k* (tf_{ik}): número de vezes que o termo é mencionado no documento;
- Frequência nos documentos (*df*): número de documentos da coleção em que o termo é mencionado;
- Frequência no documento invertida (*idf*): $idf = \log (d/df)$ onde *d* representa o número de documentos presentes no banco de dados;

- Após, define-se o peso do termo k do documento i como:

$$W_{ik} = tf_{ik} * idf_k$$

Segundo Ramos (2003) *tf-idf* é um algoritmo eficiente e simples para combinar palavras em uma consulta para identificar documentos relevantes. A codificação *tf-idf* é direta, tornando-o ideal para compor a base de algoritmos mais complicados e sistemas de recuperação de consultas (BERGER et al, 2000).

2.8.2.4 Algoritmos de Análise de Agrupamentos

A tarefa de agrupar documentos advém da aprendizagem de máquina designada como não supervisionada, não possuindo qualquer forma de treinamento dos dados para produzir uma classificação (KONCHADY, 2006).

É amplamente utilizada na exploração e análise de dados com foco na descoberta e interpretação de grupos de objetos tendo propriedades e/ou comportamentos semelhantes (ANDRIENKO; ANDRIENKO, 2009; CARPINETO et al., 2009).

De maneira geral, algoritmos de agrupamento particionam um conjunto de objetos em aglomerações (MANNING; SCHÜTZE, 2003). O intuito é aproximar internamente os objetos, proporcionando a homogeneidade do grupo, definindo uma estrutura para os dados e um distanciamento externo (separação) entre os grupos gerados (HAIR et al., 2010).

Esta técnica contribui significativamente na análise de extensos volumes de dados, possibilitando ao analista considerar grupos de objetos em vez de objetos singulares. Segundo Andrienko e Andrienko (2009), a formação dos grupos de objetos pode auxiliar na detecção de objetos incomuns que exigem uma investigação especial.

Algoritmos capazes de agrupar documentos textuais utilizam a abordagem chamada de modelo de espaço vetorial, na qual cada documento é representado por um vetor de frequências de termos (BEIL; ESTER; XU, 2002), podendo as frequências serem transformadas em pesos. Os vetores de termos devem ser normalizados para medir a similaridade entre dois documentos quaisquer. Para tal, utiliza-se uma medida de similaridade entre documentos, como por exemplo, o cosseno.

Na literatura encontram-se esforços para tentar separar as ideias em grupos. Estes esforços, muitas vezes manuais, envolvem tempo e custo. Os autores Mechant et al. (2012), por exemplo, após a fase de ideação em sessões de *brainstorming*, convidaram os participantes para dividir as ideias em grupos. Estes grupos foram divididos em ideias atuais, com alto potencial

que estejam relacionadas à empresa, ideias com chances de sucesso, entre outras. Segundo os autores, a criação de agrupamentos permitiu avançar em direção à priorização e avaliação.

2.8.2.4.1 K-means

Na mineração de dados o algoritmo *k*-means é um método para agrupar documentos. Introduzido por J. B. MacQueen, em 1967, caracteriza-se por ser um algoritmo simples de aprendizagem não supervisionada para lidar com o problema de agrupamento. Este algoritmo objetiva particionar n observações em k grupos, cada observação pertencerá ao grupo mais próximo à média.

Baseado na tarefa de particionamento, o algoritmo *k*-means é um dos mais conhecidos e utilizados no meio científico e industrial (JAIN; MURTY; FLYNN, 1999). Jain (2010) corrobora ao afirmar que, apesar do algoritmo ter sido publicado pela primeira vez em 1955 e já existirem milhares de algoritmos de agrupamento, *k-means* continua sendo amplamente utilizado devido à dificuldade em projetar um algoritmo de agrupamento de propósito geral. Segundo Jain (2010), a facilidade de implementação, simplicidade, eficiência e sucesso empírico são os principais motivos da popularidade do algoritmo. *K-means* que segue os seguintes passos:

- Inicialmente o algoritmo começa com a informação de quantos grupos serão formados durante o processo de agrupamento. Em geral, segundo Jain (2010), o *k-means* é executado com diferentes valores de k e a partição mais significativa para o especialista do domínio é selecionada. Diferentes inicializações de k podem levar a diferentes agrupamentos finais, uma vez que o algoritmo apenas converge para mínimos locais (JAIN, 2010). Segundo Jain (2010), uma das formas de superar este desafio é executar o algoritmo *k-means* com determinado k em várias partições iniciais diferentes, e escolher a partição com o menor erro quadrático;
- Em seguida, o algoritmo distribui um elemento para cada grupo. No início, estes elementos serão a semente inicial de cada grupo e, conseqüentemente, o centroide.
- Após cada iteração do algoritmo, novos elementos são atribuídos aos grupos e o centroide é recalculado, caracterizado pela média entre os elementos. O algoritmo *k-means* utiliza geralmente a distância euclidiana para calcular a distância entre os elementos.

Para Jain (2010), “*k-means* tem uma história rica e diversificada, tal como foi descoberto de forma independente em diferentes campos científicos por Steinhaus (1956), Lloyd (proposto em 1957, publicado em 1982), Ball e Hall (1965) e MacQueen (1967)”.

2.8.2.4.2 Kohonen

O algoritmo de Kohonen possui inspiração biológica e foi desenvolvido por Teuvo Kohonen, em 1982. O algoritmo consiste em uma rede neural artificial interconectada e não supervisionada que permite um automapeamento. O modelo de rede neural com mapas auto organizáveis de Kohonen (*Self-Organizing Maps - SOM*) tornou-se amplamente utilizado.

Kohonen é considerado relativamente simples e com a capacidade de organizar dimensionalmente dados complexos em grupos (*clusters*), de acordo com as relações (CASTRO; VON ZUBEN, 1999). Possui como característica a aplicação no agrupamento de dados, pois permite representar uma coleção de dados com variados atributos em um mapa de baixa dimensão, gerando *clusters* similares próximos um do outro no mapa (JIANG et al., 2004).

O mapa de características auto-organizado produz mapeamentos de espaços dimensionais elevados para espaços cuja dimensão topológica é geralmente menor que a original (CASTRO; VON ZUBEN, 1999). Os mapeamentos são capazes de preservar as relações de vizinhança entre os dados de entrada. Uma propriedade a ser explorada em uma ampla gama de aplicações como reconhecimento de fala, análise de dados e otimização combinatória (KASKI; KOHONEN, 1996; KASKI, 1997). Mapeamentos similares podem ser identificados em várias áreas do cérebro humano, por exemplo. Também em outros animais, indicando que a preservação topológica entre os espaços de entrada e saída é uma característica em sistemas de processamento de sinais bem desenvolvidos (CASTRO; VON ZUBEN, 1999). No entanto, segundo os autores, verificou-se que o uso de estruturas de dimensões fixas limita severamente os mapeamentos resultantes.

Kohonen necessita apenas dos parâmetros de entrada, mostrando-se adequado a problemas cujo os padrões são desconhecidos ou indeterminados (CASTRO; VON ZUBEN, 1999). Assim, é capaz de diminuir a dimensão de um grupo de dados conseguindo manter a representação real em relação às propriedades relevantes dos vetores de entrada. Portanto, é possível obter como resultado um conjunto das características do espaço de entrada transformando um mapa multidimensional em bidimensional, adicionando os elementos ao novo mapa de tal forma que os objetos similares sejam posicionados próximos uns dos outros.

O algoritmo compõe um grupo de redes geralmente chamadas de redes baseadas em concorrência ou redes competitivas (FAUSETT, 1994). “Essas redes combinam concorrência com uma estratégia de aprendizagem não supervisionada para fazer o ajuste dos pesos, com base em semelhanças existentes nos padrões de entrada” (CASTRO; VON ZUBEN, 1999). Ao agrupar os dados em *clusters*, cada unidade de saída pode representar um *cluster*, limitando o número de *clusters* ao número de unidades de saída (CASTRO; VON ZUBEN, 1999). Segundo os autores, ao decorrer do processo de treinamento, a rede determina “a unidade de saída (chamada vencedora) e que está mais próxima do vetor de entrada. O vetor de peso associado ao vencedor é ajustado de acordo com o algoritmo de aprendizagem”. Assim, as unidades dentro de um determinado limiar do vencedor podem ter seus vetores de peso ajustados de forma semelhante. A rede apresenta características como utilização de aproximação dos pontos similares onde os mesmos são processados separadamente e possibilita ao mapa obter centros em um plano bidimensional, melhorando a compreensão e visualização.

2.8.2.4.3 Lingo

Lingo é um algoritmo baseado em resultados de pesquisa e amplamente utilizado para agrupamento de dados textuais, preocupado em definir as melhores descrições (etiquetas) para os grupos gerados de documentos. De acordo com Osiński (2003), o algoritmo possui como princípio encontrar descrições significativas de *clusters* e, em seguida, com base nestas descrições, determinar o conteúdo pertencente a cada grupo. Estas etiquetas (rótulos) são identificadas com base no espaço do modelo vetorial juntamente com a técnica de indexação semântica latente, do inglês (LSI), buscando identificar os conceitos presentes nos documentos.

Segundo Osiński (2003), o algoritmo é composto pelos seguintes passos:

- Primeiramente realiza-se o pré-processamento. Para cada documento ocorrerá à filtragem de texto, a identificação do idioma do documento e a aplicação da técnica de *stemming* (identificação da raiz de uma palavra);
- Em seguida, realiza-se a extração de características com o objetivo de descobrir termos e frases frequentes;
- O terceiro passo possui como intuito gerar a inferência do rótulo do *cluster*, aplicando LSI para descobrir os conceitos abstratos. Para cada conceito, identificar a frase com melhor correspondência;

- No quarto momento, retiram-se os rótulos de *cluster* semelhantes, objetivando a descoberta de conteúdo. Para isso, para cada rótulo de *cluster* será utilizado o modelo do espaço vetorial para determinar o conteúdo de fragmentação;
- Por último, ocorrerá a formação final do *cluster* e a fusão de agrupamentos similares.

O algoritmo Lingo diferencia-se no momento de escolher a descrição do agrupamento, uma vez que objetiva a identificação do melhor termo que descreva determinado agrupamento (OSINSKI; WEISS, 2004).

Como ponto fraco, pode-se destacar a incapacidade de realizar agrupamento hierárquico e tempo de processamento bastante lento, devido ao uso de decomposição singular do valor (SVD) (WROBLEWSKI, 2003).

2.8.2.5 Métricas de Avaliação de Agrupamentos

A comparação entre tarefas de agrupamento ocorre através da análise do melhor resultado e desempenho. As métricas para avaliar o desempenho advêm da área de Recuperação de Informação (RI) e são baseadas na noção de relevância.

Segundo Goutte e Gaussier (2005), a avaliação empírica desempenha um papel central na estimativa do desempenho dos sistemas de processamento de linguagem natural (PNL) ou de recuperação de informações (RI).

O erro quadrático é uma das medidas mais utilizadas para avaliar a eficiência de modelos de agrupamentos sobre dados numéricos (DUDA; HART, 1973). Constitui a função critério mais utilizada em algoritmos de agrupamento particionais.

Sua aplicabilidade está na avaliação de dados numéricos, podendo ser adaptada para dados textuais (HALKIDI et al., 2001; DUDA; HART, 1973).

Um exemplo é o algoritmo *k-means*, um dos mais simples e conhecidos algoritmos de erro quadrático. O erro quadrático é a menor soma dos quadrados das distâncias entre os centroides e os respectivos pontos pertencentes ao agrupamento correspondente.

O erro quadrático de um agrupamento C com k agrupamentos de um conjunto de padrões E , é definido conforme a equação 2:

$$E = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2 \quad (2)$$

Nesta fórmula, E é a soma do erro quadrado para todos os objetos na base de dados, x_i é o ponto no espaço representando um dado objeto em determinado cluster j , e c_j é o representante do *cluster* C_j . Assim, para um determinado agrupamento C_j , o vetor médio c_j é a melhor representação para os exemplos em C_j , visto que o somatório dos erros quadráticos é minimizado. Um bom agrupamento minimiza o valor E .

2.8.3 Análise de Redes Sociais

Os passos iniciais da teoria de grafos encontram-se principalmente nos trabalhos do matemático Leonhard Euler (ALEXANDERSON, 2006). O artigo escrito por Euler tratava do problema das pontes de Koningberg, cidade famosa por suas sete pontes que unem as margens do rio Pregel a duas ilhas. A partir deste problema, Euler criou o primeiro teorema da teoria dos grafos (ALEXANDERSON, 2006).

Um grafo é definido como um conjunto de nós conectados por arestas, formando uma rede. Logo, uma rede pode ser definida como um conjunto de nodos ou vértices conectados por ligações ou arestas e tem despertado o interesse de diferentes áreas (NEWMAN, 2001). Este interesse ocorre principalmente pelo aumento da disponibilidade de dados relacionados às interações humanas, frente à necessidade de compreender como diferentes atores interagem e para analisar fenômenos singulares. Analistas de redes acreditam que a forma como um sujeito vive depende em grande parte de como ele está vinculado dentro de uma rede de conexões sociais.

Neste sentido, a Análise de Redes Sociais (ARS), originária da Sociologia, da Psicologia Social e da Antropologia, possibilita compreender a dinâmica das relações, fluxos de informação, mecanismos e agentes ao analisar ambientes complexos de interações (FREEMAN, 1996). Segundo a *International Network for Social Network Analysis* (INSNA), a ARS possui como foco descobrir padrões de interações entre pessoas ou unidades de análise (INSNA, 2020). É frequentemente utilizada para compreender as redes informais, espontâneas e não intencionais provenientes das inter-relações (GROSSETTI, 2003).

Por meio das relações, é possível analisar a estrutura das ligações existentes entre os indivíduos que compõem as redes sociais (TOMAEL; MARTELETO, 2013). Os analistas procuram identificar possíveis padrões e determinar as condições em que esses padrões ocorrem para posteriormente descobrir suas consequências. A grande diferença da ARS para outras análises é que a ênfase está nas ligações entre os indivíduos e não nos atributos (características) dos atores (WASSERMAN; FAUST, 1999).

Os autores Cross, Parker e Borgatti (2000) destacam que a ARS é considerada uma ferramenta significativa para analisar relacionamentos que promovem compartilhamento da informação e do conhecimento. Deste modo, a ARS interessa a pesquisadores de diferentes áreas que buscam compreender o impacto sobre a vida social, dando origem a metodologias de análise, numa estrutura em forma de redes.

Logo, muitos sistemas podem ser modelados e analisados como uma rede. Na matemática, por exemplo, baseado na teoria dos grafos, é possível aplicar algoritmos matemáticos às redes para determinar os vértices mais conectados, o caminho mínimo entre os vértices, entre outras medidas (YE; SONG; LI, 2012). As medidas comumente utilizadas durante a análise de redes sociais são: densidade, distância média, diâmetro e grau médio. Medidas como a modularidade também são relevantes (NEWMAN, 2006).

A densidade é utilizada para indicar o nível de conectividade da rede. Redes com alta densidade representam alta conectividade (KATERNDAHL, 2012). A distância média entre nodos da rede indica o número médio de intermediários ou caminhos de possibilidades entre dois nós quaisquer. O diâmetro é o comprimento da maior distância entre dois nós, medidos pelo número de ligações. Já o grau médio de uma rede determina o número médio de relacionamentos entre os nodos (MARTINS et al., 2012).

Outra medida é a modularidade, que objetiva estimar a qualidade dos agrupamentos gerados em uma rede, ou seja, a qualidade da divisão/distribuição realizada na rede (NEWMAN, 2006). Em redes reais, estes valores ocorrem entre 0,3 e 0,7, redes com valores maiores de modularidade possuem conexões densas entre os nós dentro do agrupamento e conexões esparsas entre nós de diferentes agrupamentos (NEWMAN; GIRVAN, 2004).

Também se pode analisar os nodos individualmente por meio das medidas de centralidade. A centralidade de intermediação quantifica o número de vezes que um nó atua como uma ponte ao longo do caminho mais curto entre dois outros nós (FREEMAN, 1977). Em um grafo conectado, a centralidade de proximidade de um nó é o comprimento médio do caminho mais curto entre o nó e todos os outros contidos no grafo. Assim, quanto mais central é um nó, mais próximo ele fica de todos os outros.

Segundo Campos (2000), outra questão importante é a normalização, principalmente quando os requisitos para o emprego da estatística paramétrica, entre eles, normalidade da distribuição dos erros, homogeneidade das variâncias, e aditividade dos efeitos dos fatores de variação não puderem ser preenchidos pelos dados da amostra experimental. Este recurso é evidentemente mais poderoso que a abordagem não paramétrica devido à própria natureza dos dados, muitas vezes não numéricos (CAMPOS, 2000).

2.9 TRABALHOS CORRELATOS

Ao realizar um mapeamento da literatura, conforme descrito na Seção 1.5, sobre os métodos e técnicas de mineração de dados e texto aplicados ao contexto de gestão de ideias, foram identificados trabalhos que serão citados a seguir, conforme o Quadro 1.

Alguns destes trabalhos relacionados ao domínio de estudo, apresentaram abordagens baseadas em análise de agrupamento. Os autores Niederberger et al. (2012), por exemplo, desenvolveram um método de agrupamento baseado em uma rede Hebbian e, Chan, Dang e Down (2016) utilizaram um algoritmo de agrupamento para identificar se os caminhos gerados por máquinas produzem maior fluência e amplitude no processo de ideação. O estudo mais correlacionado a esta pesquisa é o de Sérgio, Souza e Gonçalves (2017), no qual propuseram um modelo baseado em ontologia e análise de agrupamento para apoiar a Gestão de Ideias, além de contribuir no processo de tomada de decisão. Um protótipo foi desenvolvido para demonstrar a viabilidade do modelo e aplicado em dois cenários de estudo. Os autores afirmaram que, ao analisar grupos de ideias semelhantes, padrões e tendências podem ser evidenciados, contribuindo significativamente na visualização das ideias. Porém, não apresentavam indicativos do potencial de implementação de uma ideia nem consideravam as temáticas estratégicas relacionadas ao contexto organizacional.

Pesquisas relacionadas à proposição de algoritmos também foram identificadas na literatura. Dentre os algoritmos está o trabalho de Siangliulue et al. (2016). Os autores criaram uma plataforma denominada *ideaHound*. Esta plataforma permite postar ideias, visualizar ideias de outros usuários e as postadas pelo solicitante. O algoritmo proposto calcula a matriz aproximada de similaridade de ideias usando o algoritmo *t-Distributed Stochastic Neighbor Embedding* (t-SNE). A utilização de técnicas de mineração textual também é proposta por Lee et al. (2018) que analisaram um grande conjunto de dados coletados da plataforma MyStarbucksIdea.com[®], incluindo a medida estatística *tf-idf* e análise de sentimento. O objetivo consistia no desenvolvimento de modelos de classificação para calcular a probabilidade de adoção de cada ideia. Contudo, como esta tese busca avaliar o potencial de implementação de uma ideia quando esta é sugerida, logo modelos baseados em classificação não poderiam ser aplicados. Visto que para a classificação é necessário estabelecer uma classe meta a ser alcançada. Sendo que a análise de agrupamento é realizada de maneira automática, sem intervenção do especialista. Também não são consideradas previamente propriedades características dos grupos, nem a utilização prévia de grupos de teste para direcionar a classificação.

Hoornaert et al. (2017) propuseram um modelo para auxiliar os gerentes na avaliação e seleção de ideias. Inicialmente, os autores elucidaram os principais aspectos das ideias passíveis de implementação futura. Para isto, foram identificados três elementos estruturais disponíveis para uma ideia: o conteúdo da ideia, o colaborador que a propõe e o *feedback* da multidão. Os resultados foram alcançados com a aplicação de métodos automatizados de recuperação de informação (indexação semântica latente) e algoritmos não lineares de aprendizado de máquina. Entretanto, o trabalho não estabeleceu índices sobre o potencial de implementação das ideias com base em temáticas estratégicas relacionadas ao contexto organizacional.

O artigo relacionado mais recente é escrito por Westerski e Kanagasabai (2019). Os autores utilizaram técnicas de mineração de dados para detecção de *outliers*. O objetivo é identificar meios tecnológicos na identificação automática de inovações para auxiliar na tomada de decisão. Após as descobertas sobre a natureza das inovações disruptivas, o artigo apresentou um estudo comparativo de vários algoritmos de detecção de *outliers*, aplicados a dois conjuntos de dados do mundo real que contêm descrições textuais de ideias para diferentes setores. Os resultados obtidos demonstram a capacidade de detecção de valores extremos e mostraram que o algoritmo *k-NN* com *tf-idf* e cosseno são os melhores candidatos para a tarefa de detecção. Apesar do foco dos autores ser a detecção de *outliers*, estes algoritmos se mostraram promissores para análise das ideias disruptivas, podendo ser analisados para o domínio em investigação. Entretanto, o estudo não considerou as principais temáticas relacionadas ao contexto organizacional, tão pouco indicativos do potencial de implementação das ideias. É importante destacar que *outliers* podem ser identificados por meio de técnicas de agrupamentos, mas os autores não estabeleceram um *ranking* de agrupamentos para identificar os mais promissores, ou seja, os grupos de ideias com maior potencial de implementação.

Por fim, os autores Wehnert et al. (2018) utilizaram o *software* Leximancer[®] para recuperar informações dos textos das ideias e assim obter *insights* sobre o discurso. Segundo os autores, a ferramenta Leximancer[®] é adequada para reconhecer relações e estruturas contextuais em grandes quantidades de dados de texto não estruturados. A ferramenta foca apenas na análise automática de documentos textuais para identificar os principais conceitos. Todavia, não atribui índices aos conceitos nem considera as principais temáticas relacionadas ao contexto organizacional, tão pouco promove um aprofundamento nas análises realizadas.

Por fim, o número reduzido de trabalhos correlatos demonstra como são incipientes os estudos relacionados ao tema. Foram identificadas apenas oito publicações em três grandes bases de dados, o que configura um campo aberto para novas pesquisas. Ao analisar as

publicações resultantes da revisão bibliográfica, pode-se perceber que nenhum dos estudos forneceu índices relacionados ao potencial de implementação de ideias. Ideias estas alinhadas às temáticas estratégicas envolvidas no contexto organizacional, com a aplicação de métodos e técnicas de mineração de dados e texto com o intuito de contribuir na seleção de ideias.

2.10 CONSIDERAÇÕES FINAIS

Este capítulo apresentou o referencial teórico que norteia esta tese. A escolha dos elementos que irão compor o modelo proposto está diretamente relacionada aos elementos evidenciados na revisão sistemática, além de combinar métodos e técnicas adotadas na mineração de dados e texto. Ao submeter uma ideia ao processo de avaliação, a primeira tarefa a ser executada é o reconhecimento das informações e representá-la computacionalmente como uma matriz termo-documento, visando criar subsídios para identificar o domínio. Para representar os documentos (ideias) como vetores de palavras será utilizado o modelo algébrico VSM. Para determinar a importância de uma palavra em um documento em relação a uma coleção de documentos, será utilizada a medida estatística comumente usadas nas áreas de recuperação de informação e mineração de dados *tf-idf*. Visando também criar suporte ao modelo, realiza-se o processo de transformação matricial utilizando SVD, ao qual permitirá na aplicação do modelo a integração dos termos-chave que representam as ideias e os índices propostos nesta tese. Por fim, com o intuito de aplicar a tarefa de agrupamento, será utilizada uma versão modificada do algoritmo *k-means* que utiliza o conceito de similaridade vetorial. A representação e o armazenamento do conhecimento de domínio serão realizados por meio de uma ontologia.

3 METODOLOGIA DE PESQUISA

Este capítulo apresenta a metodologia de pesquisa que norteou o trabalho. Inicialmente, serão apresentados os conceitos basilares fundamentais.

Kourganoff (1990) define o termo pesquisa como “o conjunto de investigações, operações e trabalhos intelectuais ou práticos que tenham como objetivo a descoberta de novos conhecimentos, a invenção de novas técnicas e a exploração ou a criação de novas realidades”.

Segundo o autor Gil (2008, p. 17), o termo pesquisa é definido como:

[...] o procedimento racional e sistemático que tem como objetivo proporcionar respostas aos problemas que são propostos. A pesquisa desenvolve-se por um processo constituído de várias fases, desde a formulação do problema até a apresentação e discussão dos resultados.

Wazlawick (2010) salienta que a natureza original da pesquisa procura apresentar conhecimento novo a partir de observações e teorias construídas para explicá-lo.

Segundo DePoy e Gitlin (1994, p.5), a pesquisa consiste em uma:

[...] estratégia múltipla, sistemática de gerar conhecimento sobre o comportamento humano, a experiência humana e o ambiente humano no qual o processo de ação e pensamento do pesquisador são claramente especificados de forma lógica, compreensível, reproduzível e utilizável.

Planejar passo a passo os procedimentos que serão empregados na pesquisa compreende a primeira fase da pesquisa científica, envolvendo desde a escolha do tema até a operacionalização dos métodos.

Tartuce (2006) destaca que a metodologia científica versa sobre método e ciência. Método (do grego *methodos*) está relacionado ao caminho que será percorrido para alcançar um fim. Segundo o autor, o termo metodologia se refere ao estudo do método, tem como objetivo estabelecer regras e procedimentos para realizar uma pesquisa. Já o termo científico deriva de ciência e está relacionado ao conjunto de conhecimentos precisos e ordenados em relação a determinado domínio do saber. Logo, a metodologia científica é o estudo sistemático e lógico dos métodos empregados nas ciências, seus fundamentos, sua validade e sua relação com as teorias científicas (TARTUCE, 2006).

Minayo (2007, p. 44) vai além e define o termo metodologia de forma mais abrangente:

[...] como a discussão epistemológica sobre o “caminho do pensamento” que o tema ou o objeto de investigação requer; como a apresentação adequada e justificada dos métodos, técnicas e dos instrumentos operativos que devem ser utilizados para as buscas relativas às indagações da investigação; como a “criatividade do pesquisador”, ou seja, a sua marca pessoal e específica na forma de articular teoria, métodos, achados experimentais, observacionais ou de qualquer outro tipo específico de resposta às indagações específicas.

Em relação aos objetivos, esta pesquisa é classificada como exploratória porque não pretende descrever os fatos nem buscar suas causas e explicações, mas sim proporcionar maior familiaridade com o problema (GIL, 2008).

Quanto ao enquadramento metodológico, Lacerda et al. (2013) destacam as principais possibilidades:

[...] responder ao problema de pesquisa formulado; ser avaliado pela comunidade científica; evidenciar procedimentos que robusteçam os resultados da pesquisa e, de forma tão importante quanto, não deve ser compreendido como um ato burocrático.

A atividade preponderante da metodologia é a pesquisa. A metodologia empregada para o desenvolvimento desta pesquisa foi a *Design Science Research Methodology* (SIMON, 1996). As próximas seções apresentarão o embasamento teórico acerca da metodologia de pesquisa.

3.1 PESQUISA CIENTÍFICA E PESQUISA TECNOLÓGICA

A modalidade de pesquisa no qual esta proposta de tese está inserida pode ser definida como pesquisa tecnológica e tem como objetivo a criação de um artefato tecnológico. Por meio da tecnologia, é possível desenvolver e projetar artefatos com base no conhecimento científico (CUPANI, 2006). A união da ciência e da tecnologia viabiliza o processo de criação do conhecimento (SILVA; MENEZES, 2005). Para Cupani (2011), é a partir das inovações tecnológicas que revoluções científicas podem ser desencadeadas. Cita-se ainda que a pesquisa tecnológica pode se utilizar do método científico para comprovar seus resultados.

Bunge (1985, p. 231) destaca que a tecnologia pode ser vislumbrada como “o campo do conhecimento relativo ao projeto de artefatos e ao planejamento de sua realização, operação, ajuste, manutenção e monitoramento, à luz do conhecimento científico”. Logo, o conhecimento científico conduz à teorias mais abrangentes. Para Cupani (2006), o conhecimento tecnológico desenvolve teorias mais limitadas para abordar um problema específico, implicando em uma invenção.

A pesquisa científica tem como intuito a descoberta de algo existente, enquanto que a pesquisa tecnológica objetiva produzir algo novo (CUPANI, 2006). A pesquisa científica é motivada por razões de ordem intelectual (“pesquisa pura”) ou de ordem prática (“pesquisa aplicada”), conforme a tendência do pesquisador à busca de conhecimento (por si só) ou de um determinado aprimoramento (GIL, 2010), respectivamente. Em todo caso, ambos os modelos produzem ciência.

Freitas Junior et al. (2014) realizaram uma distinção entre a pesquisa científica e tecnológica, conforme o Quadro 5:

Quadro 5 – Pesquisa científica e pesquisa tecnológica

Caraterística	Pesquisa Científica	Pesquisa Tecnológica
Definição	Conhecimento da natureza e exploração desse conhecimento (KNELLER,1980).	"O estudo científico do artificial". "Tecnologia pode ser vista como o campo do conhecimento relativo ao projeto de artefatos e ao planejamento de sua realização, operação, ajuste, manutenção e monitoramento, a luz do conhecimento científico." (BUNGE, 1985)
Teorias	Amplo alcance e uso de idealizações, o que obriga a adaptar o conhecimento científico para possibilitar sua aplicação (CUPANI, 2006).	Aplicação limitada, pois o conhecimento tecnológico é específico para uma determinada tarefa. Dois tipos: substantivas (conhecimento sobre a ação tecnológica) e operativas (conhecimento sobre as ações de que dependem o funcionamento dos artefatos) (CUPANI, 2006)
Resultado	Descobrimto de algo existente. O produto é neutro (nem bom nem mau) (CUPANI, 2006; CUPANI, 2011).	Criação de algo novo. O produto não é e nem pode ser neutro. É, no mínimo, ambivalente (CUPANI, 2006; CUPANI, 2011).
Conhecimento	Descritivo (CUPANI, 2006).	Prescritivo. Específico. Peculiar. Conhecimento tácito, do saber-como. (CUPANI, 2006; CUPANI, 2011)
Desafios		Factibilidade, confiabilidade, eficiência dos inventos, relação custo-benefício (CUPANI, 2006).
Limitação	Ditada pela teoria. Pode-se explorar livremente as possibilidades (CUPANI, 2006; CUPANI, 2011).	Ditada pela tarefa Imposta (CUPANI, 2006).
Origem dos dados	Científicos (CUPANI, 2006).	Experiência não científica (CUPANI, 2006). Dados relativos às exigências (técnicas, econômicas, culturais) que o artefato deve satisfazer (CUPANI, 2011)
Tipos de leis	Leis que governam os fenômenos naturais (CUPANI, 2006).	Regras de ação para dar origem aos fenômenos artificiais (CUPANI, 2006).
Pensamento	Abstrato e verbal (CUPANI, 2006)	Analógico e visual. (CUPANI, 2006)
Origem das variáveis	Não específico. (CUPANI, 2006).	Metas a alcançar (CUPANI, 2006).
Objetivos dos experimentos	Entender a realidade (CUPANI, 2006).	Conhecimento prático: "o artefato funcionará?", "haverá, acaso, fatores não previstos teoricamente que serão detectados experimentalmente?" (CUPANI, 2006). Controlar a realidade (CUPANI, 2006)

Explicações	Causais (CUPANI, 2006).	Funcionais (CUPANI, 2006)
Noção de conhecimento	Muda de acordo com as teorias (CUPANI, 2006).	Admitem apreciação de sua verdade ou falsidade, podendo afirmar que o artefato desempenha bem ou mal sua função. Superior em relação ao científico por sua certeza e eficácia (CUPANI, 2006).
Mudança de paradigma	Implica em muito exame e discussão (CUPANI, 2011).	Ocorre devido a anomalias funcionais ou presumíveis. A necessidade da mudança é percebida mais diretamente (CUPANI, 2011).
Revoluções	Inovadoras e eliminatórias (CUPANI, 2011).	Não implicam necessariamente em uma seleção radical, não supõem forçosamente uma nova comunidade e são compatíveis com a continuidade da tecnologia "normal" (CUPANI, 2011).

Fonte: Freitas Junior et al. (2014)

Neste aspecto, é apresentado na próxima seção os princípios e fundamentos da *Design Science Research Methodology*, empregados nesta pesquisa como metodologia norteadora do desenvolvimento da pesquisa tecnológica.

3.2 DESIGN SCIENCE RESEARCH METHODOLOGY (DSRM)

Segundo Geerts (2011), os trabalhos de pesquisa das ciências naturais geralmente aderem a uma estrutura que consiste nas seguintes etapas: definição de problema, revisão de literatura, desenvolvimento de hipóteses, coleta de dados, análise, resultados e discussão.

Em contraponto, tem-se as ciências do “artificial”, destacando que elas “se ocupam da concepção de artefatos que realizem objetivos” (SIMON, 1996, p. 198). O propósito do projeto é “mudar as situações existentes em preferências” (SIMON; 1996). A *Design Science Research* cria artefatos, ou seja, algo criado por humanos geralmente para um propósito prático (GEERTS, 2011). Os autores March e Smith (1995) apresentam quatro tipos diferentes de artefatos:

- Conceitos: formam o vocabulário de um domínio, definindo os termos usados para descrever e pensar sobre as tarefas;
- Modelos: expressão do relacionamento entre conceitos, descrevendo ou representando objetos do mundo real;
- Métodos: conjunto de passos utilizados para executar determinada tarefa;
- Instâncias: concretização de um artefato em seu ambiente e tem como objetivo demonstrar a viabilidade e a eficácia dos modelos e métodos. Informam ainda como implementar ou utilizar determinado artefato e seus possíveis resultados.

Geerts (2011) enfatiza que as duas características importantes dos artefatos da *Design Science Research* são relevância e novidade. Os autores Hevner et al. (2004) sugerem que a *Design Science Research* deve abordar um problema não resolvido de forma única e inovadora ou um problema resolvido de forma mais eficaz ou eficiente.

A *Design Science Research* (DSR) objetiva melhorar a compreensão dos fenômenos dos sistemas de informação com a criação de artefatos de tecnologias da informação. Os artefatos criados incorporam a solução para um problema previamente definido (HEVNER; CHATTERJEE, 2010). Em um mundo cada vez mais conectado e globalizado, os artefatos de TI são criados para serem utilizados em uma variedade de contextos culturais (LAWRENCE; TUUNANEN; MYERS, 2010).

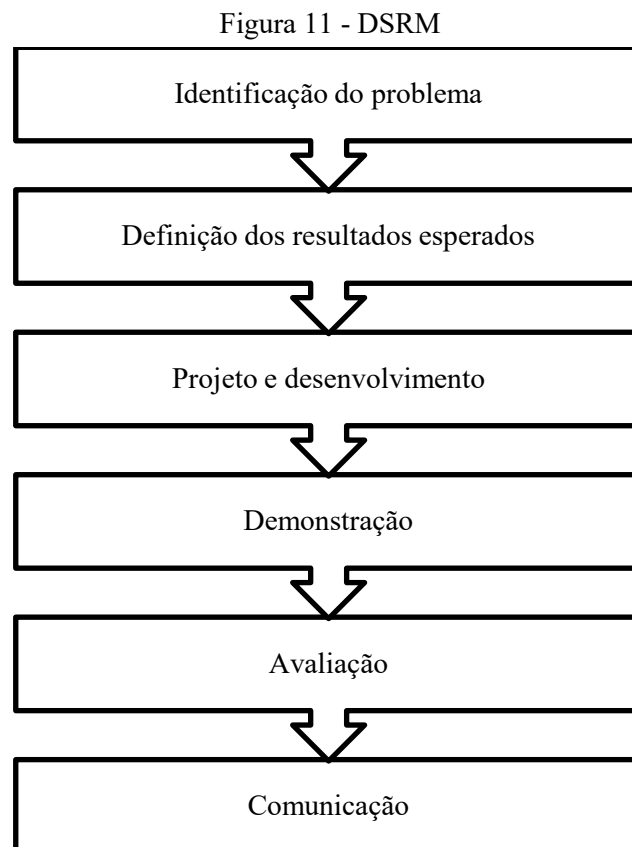
A Metodologia de *Design Science Research* (DSRM) definida por Peffers et al. (2007) é um método voltado à sistemas de informação (GEERTS, 2011). Esta metodologia, segundo Peffers et al. (2007), é desenvolvida a partir de seis etapas procedurais que podem ser executadas de acordo com a necessidade do projeto. As etapas são:

- **Identificação do problema e sua motivação:** nesta etapa é definido o problema de pesquisa e apresenta-se a justificativa para a sua investigação. O autor enfatiza a importância da definição do problema ser empregada na construção de um artefato que pode efetivamente oferecer a solução para este problema. Nesta etapa, elucida-se o estado da arte da dificuldade e a relevância da solução apresentada.
- **Definição dos resultados esperados:** a partir do conhecimento do problema e da noção de viabilidade, bem como se ele é factível, delineiam-se os objetivos da solução a ser desenvolvida. Neste momento, há novamente o estado da arte do problema e a ciência das possíveis soluções já previamente apresentadas.
- **Projeto e desenvolvimento:** esta etapa dedica-se a construção do artefato, determinando sua funcionalidade, sua arquitetura e, em seguida, o desenvolvimento do próprio artefato.
- **Demonstração:** momento de demonstração da utilidade do artefato por meio de um experimento ou simulação, estudo de caso, prova formal ou outra atividade apropriada. Inclui também o conhecimento efetivo em como utilizar o artefato para resolver determinado problema.
- **Avaliação:** objetiva-se observar e mensurar como o artefato atende solução do determinado problema, comparando-se os objetivos propostos para a solução com os resultados obtidos com a utilização do artefato. Nesta etapa, pode-se voltar às

etapas anteriores, projeto e desenvolvimento ou demonstração, de modo a aprimorar o artefato.

- **Comunicação:** nesta etapa divulga-se o problema e a relevância da solução proposta, além da apresentação do artefato desenvolvido.

A Figura 11 detalha a sequência da metodologia *Design Science Research*.



Fonte: Traduzido de Peffers et al. (2007)

A tarefa de avaliação da DSRM dos artefatos recebe especial atenção. Lacerda et al. (2013) enfatizam que, para aumentar a confiabilidade dos resultados da pesquisa, é necessário um conjunto de cuidados e procedimentos especiais para o processo de avaliação. Os autores apresentam alguns métodos de avaliação no Quadro 6.

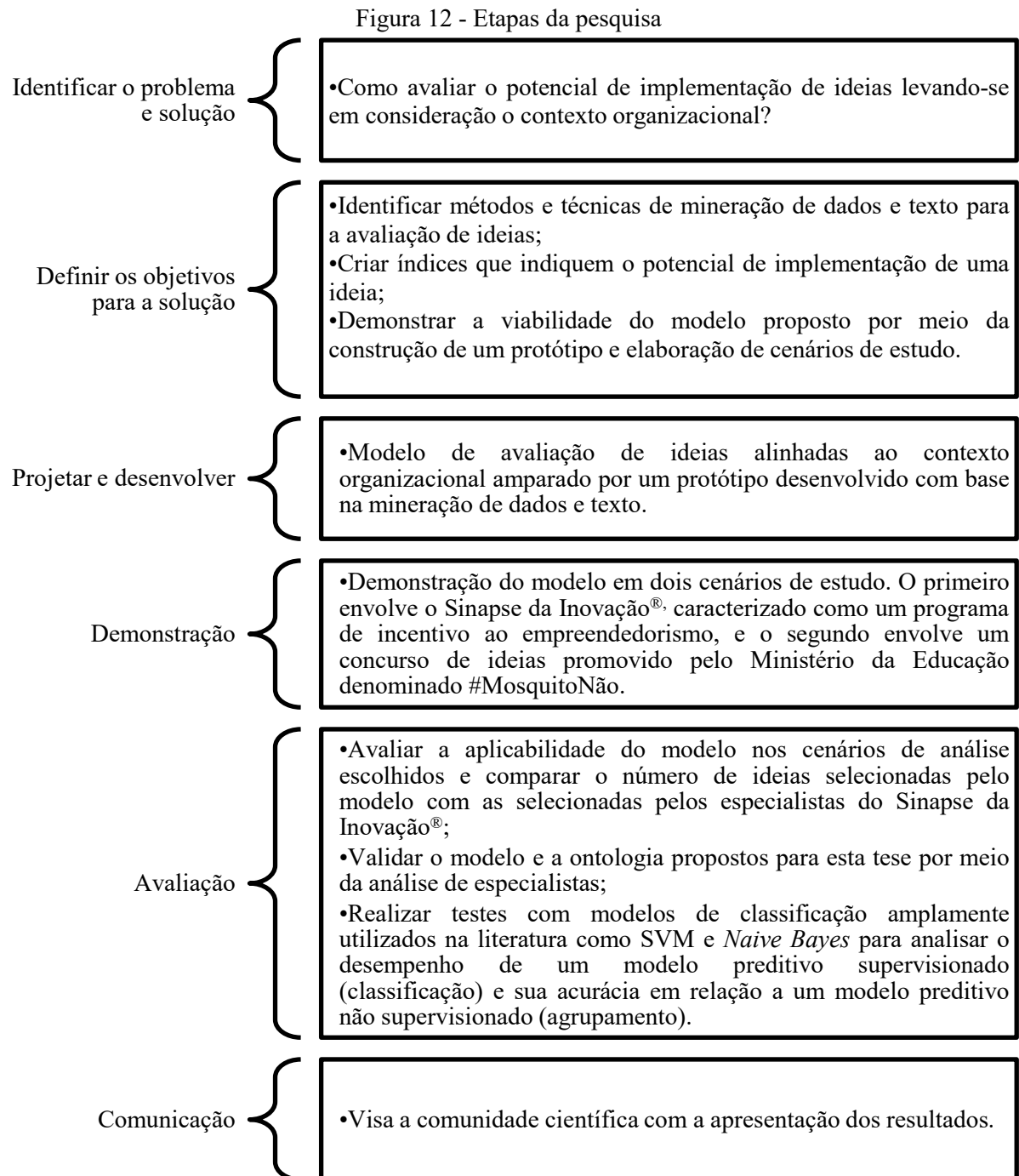
Quadro 6 – Métodos de avaliação da DSRM

Forma de Avaliação	Métodos propostos
Observacional	Estudo de Caso: estudar o artefato existente ou não, em profundidade no ambiente de negócios. Estudo de Campo: monitorar o uso do artefato em projetos múltiplos. Estes estudos podem, inclusive, fornecer uma avaliação mais ampla do funcionamento dos artefatos configurando. Dessa forma, um método misto de condução da pesquisa.
Analítico	Análise Estatística: examinar a estrutura do artefato para qualidades estáticas. Análise da Arquitetura: estudar o encaixe do artefato na arquitetura técnica do sistema técnico geral. Otimização: demonstrar as propriedades ótimas inerentes ao artefato ou então demonstrar os limites de otimização no comportamento do artefato. Análise Dinâmica: estudar o artefato durante o uso para avaliar suas qualidades dinâmicas (por exemplo, desempenho).
Experimental	Experimento Controlado: estudar o artefato em um ambiente controlado para verificar suas qualidades (por exemplo, usabilidade). Simulação: executar o artefato com dados artificiais.
Teste	Teste Funcional (<i>Black Box</i>): executar as interfaces do artefato para descobrir possíveis falhas e identificar defeitos. Teste Estrutural (<i>White Box</i>): realizar testes de cobertura de algumas métricas para implementação do artefato (por exemplo, caminhos para a execução).
Descritivo	Argumento informado: utilizar a informação das bases de conhecimento (por exemplo, das pesquisas relevantes) para construir um argumento convincente a respeito da utilidade do artefato. Cenários: construir cenários detalhados em torno do artefato para demonstrar sua utilidade.

Fonte: Lacerda et al. (2013)

3.3 DESENVOLVIMENTO DA PESQUISA

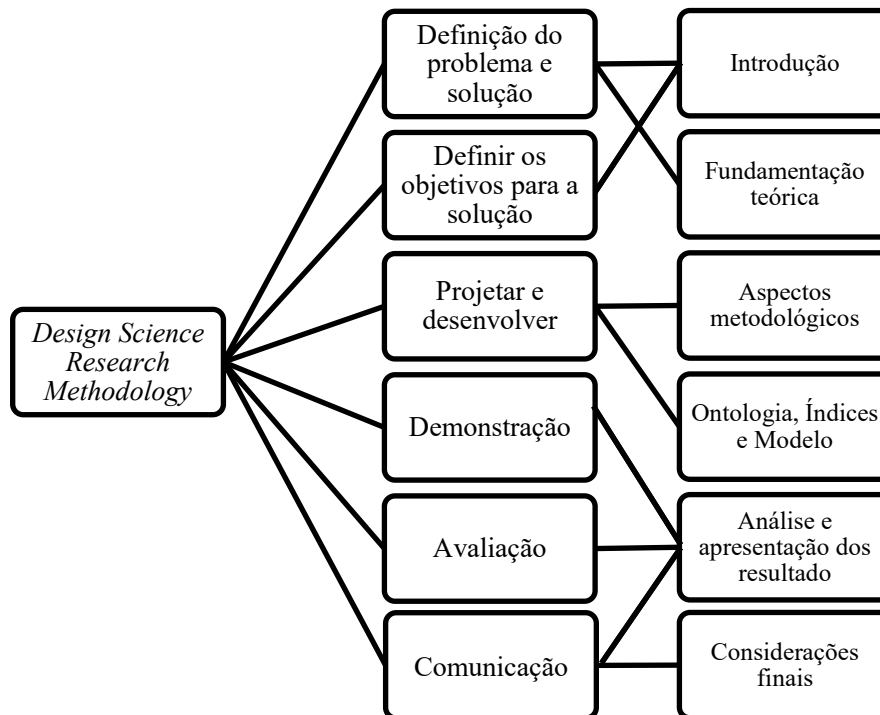
Esta pesquisa de natureza aplicada com objetivos exploratórios e descritivos, possui como principal foco disponibilizar um conjunto de artefatos que forneçam suporte à construção de um protótipo relacionado ao domínio de Gestão de Ideias, mais especificamente ao processo de avaliação de ideias. Com o intuito de construir um protótipo, esta pesquisa teve sua condução amparada na abordagem metodológica da *Design Science Research*, direcionada pela proposição de condução de Peffers et al. (2007). É apresentada na seção anterior e evidenciada na Figura 12. Nesta proposição, cada fase está correlacionada com os passos que a pesquisa talhou em busca de resolver o problema.



Fonte: Autora

As seções desta tese e a etapa metodológica relacionada são apresentadas na Figura

Figura 13 - Desenvolvimento da tese



Fonte: Autora

A seguir, cada uma das etapas presentes no desenvolvimento da tese será detalhada.

3.3.1 Definição do problema e solução

Para identificar o problema e propor uma solução foi necessário realizar uma busca sistemática, além da procura de uma lacuna de conhecimento que correlacione a gestão de ideias e os métodos e técnicas de mineração de dados e texto que podem contribuir na avaliação das ideias. Nesta tese, o termo potenciais ideias representa o potencial de implementação das ideias. Estes elementos são primordiais para a construção desta pesquisa, conforme apresentado no capítulo de fundamentação teórica.

Com o objetivo de obter uma visão mais aprofundada e completa, a revisão sistemática de literatura foi realizada conforme explanada na Seção 1.5. Contudo, não foi identificado nenhum documento com as mesmas características da pesquisa nas bases de dados *Science Direct*[®], *Web Of Science*[®] e *Scopus*[®].

Ao total foram analisados 163 artigos, porém, somente oito artigos foram classificados como relevantes para a pesquisa e apresentavam métodos ou técnicas de mineração de dados e texto utilizados no contexto de gestão de ideias. A discussão dos artigos foi realizada na Seção 2.9 do Capítulo 2. Durante a revisão sistemática, não foram encontrados artigos que tivessem

como foco a avaliação de ideias alinhadas às temáticas estratégicas relacionadas ao contexto organizacional.

Este processo possibilitou a definição inicial do problema de pesquisa e de algumas formas preliminares em como solucioná-lo. Com isto, o problema e a solução foram definidos, conforme explicitado na Figura 12.

3.3.2 Definição dos objetivos para solução

A partir do problema foram traçados os objetivos desta pesquisa e o modelo que serve como base para construção do protótipo.

3.3.3 Projeto e desenvolvimento

A terceira etapa da condução do processo refere-se ao projeto e desenvolvimento do modelo. O artefato (modelo proposto) é aplicado em um ambiente com as características condizentes ao local onde o problema se manifesta, ou seja, em um sistema de Gestão de Ideias onde ocorre a avaliação do potencial de implementação de ideias. A especificação detalhada desta etapa encontra-se no Capítulo 6 deste trabalho. De modo geral, as etapas do desenvolvimento são:

3.3.3.1 Coleta dos dados e preenchimento da base de conhecimento

Na etapa de coleta dos dados são extraídas manualmente as informações referentes às ideias e às temáticas estratégicas relacionadas ao contexto organizacional que, na sequência, serviram de insumo para o preenchimento da base de conhecimento. As temáticas estratégicas podem ser definidas por especialistas ou obtidas a partir de determinado edital de fomento provido pela organização. Para esta tese, o contexto organizacional, devido a amplitude do conceito, é descrito por meio de temáticas estratégicas representadas por termos-chave que descrevem partes do contexto organizacional. Posteriormente, é selecionado o conjunto de ideias para fazer parte da análise.

O formato de apresentação das ideias é semelhante em diferentes plataformas, caracterizado principalmente pelo título, data da postagem, descrição, autor, número de votos e

comentários. A Figura 14 apresenta o formato de disposição das ideias no site Sinapse da Inovação®, um dos cenários de análise utilizado para avaliar a efetividade do modelo proposto.

Figura 14 – Formato de exposição das ideias no site Sinapse da Inovação®

Projeto g-Ideia: Uma ponte entre o Governo e a Sociedade

Oportunidade/Desafio
Ideias representam a matéria-prima do processo de inovação. Porém inovar não é um processo trivial e envolve boas ideias. Ideias podem auxiliar na identificação de oportunidades viáveis e plausíveis de serem executadas nas organizações privadas ou públicas. Apesar da evolução das Tecnologias da Informação e um maior acesso a informação em geral, a sociedade ainda possui poucas ferramentas que possibilitem a proposição de soluções com o objetivo de melhorar a vida dos cidadãos. Ações de Governo Eletrônico podem contribuir neste processo, porém ainda é um desafio no Brasil, apesar do seu grande objetivo ser aproximar o cidadão do governo. As tecnologias da informação e comunicação podem ser empregadas como forma de viabilizar esta aproximação e para administrar recursos e processos.

Solução Proposta
Sistemas de Gestão de Ideias objetivam avaliar, classificar e selecionar ideias para possível implementação. Com a popularização da internet a solução consiste em disponibilizar um sistema online para a sociedade, do modo que a mesma possa opinar, propor e selecionar as ideias e/ou projetos de interesse. Neste sistema, os cidadãos podem sugerir, comentar, visualizar, filtrar e votar nas ideias e projetos. O sistema proporciona também um módulo de visualização de informação por meio de indicadores. Indicadores também são um meio de auxiliar no entendimento das demandas pelos gestores. Outra característica do sistema é o agrupamento dos dados. Através da análise de agrupamento é possível visualizar grupos de ideias. Essa visualização auxilia a gestores na tomada de decisão. Caso uma mesma ideia e/ou projeto ocorra mais de uma vez na base de dados, isto poderá indicar uma tendência, formando grupo de ideias, ou um possível desapego da população por esta ideia/projeto.

Grau de maturidade da solução
Em desenvolvimento, sendo que o modelo de dados, e os serviços de visualização e agrupamento de ideias já foram implementados.

Diferencial Inovativo
Texto publicado com autorização.

O g-Ideia tem como diferencial ser uma proposta inovadora na forma de relacionamento entre o Governo e a Sociedade. Seu maior objetivo é criar uma plataforma de ideias e projetos que possam ser votados, analisados, comentados, e ainda proporcionar indicadores de avaliação para o gestor. Outro ponto é aproximação entre governo e os cidadãos, permitindo que a sociedade contribua no avanço do Estado.

Grau de inovação da solução:
Radical, com mudança drástica na maneira que o produto ou serviço é consumido, trazendo um novo paradigma e modificando o modelo de negócio vigente.

17 versões postadas
Última versão 1.16 postada em 30/10/2015 às 18:19

Sobre a equipe
Cidade: Maracá

Marina Carradore Sérgio
Gênero: Administrativo
Pós-Graduação (concluído): UFSC - Universidade Federal de Santa Catarina.
Marina Carradore Sérgio possui técnico em Administração pelo Centro de Educação Profissional Abílio Paulo (CEDUP), Criciúma, Santa Catarina, Brasil, em 2011. Possui graduação em Tecnologias da Informação e Comunicação pela Universidade Federal de Santa Catarina (UFSC), Araranguá, Santa Catarina, Brasil, em 2013. Atualmente é aluna de Mestrado do Departamento de Engenharia e Gestão do Conhecimento (EGC), na área de Engenharia do Conhecimento da Universidade Federal de Santa Catarina (UFSC), Florianópolis, Santa Catarina, Brasil e pesquisadora da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior.

Suas principais áreas de pesquisas são: Ontologia, Análise de Agrupamentos, Engenharia de Ontologia, Descoberta de Conhecimento em Bases Textuais e Gestão de Ideias.

Alexandre Leopoldo Gonçalves
Gênero: Administrativo
Pós-Graduação (concluído): UFSC - Universidade Federal de Santa Catarina.
Alexandre Leopoldo Gonçalves possui graduação em Ciência da Computação pela Fundação Universidade Regional de Blumenau (1997), mestrado e doutorado em Engenharia de Produção pela Universidade Federal de Santa Catarina em 2000 e 2006. Atualmente é Professor Adjunto lotado no Campus Araranguá/UFSC, e Professor Permanente do Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento/UFSC e do Programa de Pós-Graduação em Tecnologias da Informação e Comunicação/UFSC. Tem experiência nas áreas de Ciência da Computação e Engenharia do Conhecimento atuando principalmente nos seguintes temas: Extração e Recuperação de Informação, Descoberta de Conhecimento e Engenharia de Ontologia.

Compartilhe essa ideia
Facebook, Twitter, LinkedIn, Google+

Fonte: Sinapse da Inovação®

Na Figura 14 é possível visualizar o título da ideia, a oportunidade/desafio, a solução proposta, o grau de maturidade da solução, o diferencial inovativo, a data e hora da última atualização da ideia, o(s) nome(s) do(s) autor(es), bem como as respectivas informações de currículo e *links* para compartilhamento da ideia. Na parte superior ao título é possível visualizar o número de comentários, o número de votos que a ideia recebeu e a categoria a qual a ideia pertence. Nos comentários das ideias são visualizados o nome do autor do comentário e a sua descrição, bem como a data e hora de postagem. Seguindo estes mesmos passos, foram coletadas as ideias da coletânea denominada #MosquitoNão.

Visando contribuir na formalização e conceitualização compartilhada do domínio, a base conhecimento é representada no formato de uma ontologia. A partir do resultado da revisão sistemática da literatura e em pesquisas realizadas na *web*, não se identificou uma ontologia voltada à avaliação de ideias que envolvem o contexto organizacional e suas especificidades.

A ontologia proposta neste trabalho é construída com base na metodologia *Ontology Development 101*, descreve um processo iterativo composto por sete passos para a construção de ontologias (NOY; MCGUINNESS, 2001). Foi adotada esta metodologia por sua praticidade

ao descrever as etapas de desenvolvimento (NOY; MCGUINESS, 2001). Para justificar a existência da ontologia e sua futura avaliação foram estabelecidas 15 perguntas de competência a partir da revisão de literatura. Quando se fala em contexto organizacional, diversas questões foram elaboradas. A ontologia é detalhada no Capítulo 4.

3.3.3.2 Pré-processamento dos dados

As ideias selecionadas necessitam passar por um processo de pré-processamento. Inicialmente ocorre a geração do vetor inicial com os termos-chave das ideias. Neste estudo, os termos-chave foram elencados para todas as ideias coletadas nos dois cenários de estudo. Na sequência, realiza-se a consulta nas bases de fontes externas, visando a definição dos índices propostos neste trabalho.

As fontes externas utilizadas para a consulta foram a base de patentes do Escritório Americano de Marcas e Patentes (USPTO[®] - *United States Patent and Trademark Office*), além da base de dados *Semantic Scholar*[®]. Para tanto, criou-se um processo automático para coleta dos dados e inseriu-se em um banco de dados os documentos coletados para, posteriormente, realizar consultas no momento da criação dos índices das ideias. Ressalta-se que, devido o volume de dados, coletou-se todas as patentes da base registradas nos últimos três anos, mais especificamente referentes ao ano de 2017, 2018 e até maio de 2019 quando o modelo foi executado. Coletou-se também 10 anos da base de dados *Semantic Scholar*[®], entre 2009 e 2018. A descrição detalhada destas coletas está descrita na Seção 5.1.

Para a criação dos índices, além das consultas em fontes externas, também é necessário o retorno à base de conhecimento a fim de recuperar as descrições das temáticas estratégicas envolvendo o contexto organizacional. A definição dos índices, bem como os respectivos cálculos, estão descritos em detalhes no Capítulo 5. Em seguida, ocorre a sumarização das informações retornadas com foco na temporalidade, assim como a recuperação de termos relacionados aos originais do vetor obtido por coocorrência. Para representar os documentos (ideias) como vetores de palavras é utilizado o modelo algébrico VSM. Para determinar a importância de uma palavra em um documento em relação a uma coleção de documentos, é utilizada a medida estatística *tf-idf* comumente utilizada nas áreas de recuperação de informação e mineração de dados e texto. A comparação de similaridade presente nos índices de aderência ao contexto organizacional e de tecnologia é determinada pela equação do cosseno descrita na Seção 2.8.2.2. O índice de originalidade é gerado por meio da equação modificada do cosseno para atender aos requisitos estabelecidos e o índice de atualidade é definido por uma equação

proposta pela tese, levando em consideração o número de documentos que citam determinado termo-chave ao longo dos anos. Por fim, ocorre a composição final do vetor com os termos originais, novos termos advindos das fontes externas e os índices gerados. O vetor final é então submetido à etapa de transformação dos dados, descrita na seção seguinte.

3.3.3.3 *Transformação dos dados*

A transformação é responsável por adequar os dados para serem utilizados em algoritmos de mineração. A partir da matriz obtida no pré-processamento é realizada a decomposição matricial por meio da técnica *Singular Value Decomposition* (SVD), que analisa de maneira iterativa as coocorrências entre os termos dispostos na matriz, visando estabelecer relacionamentos latentes. A etapa de transformação é descrita detalhadamente na Seção 6.1.3.

3.3.3.4 *Agrupamento dos dados*

Com o objetivo de reunir ideias semelhantes em um mesmo agrupamento e contribuir na visualização de informações relacionadas às potenciais ideias para implementação, é utilizada uma versão modificada do algoritmo *k-means*, utilizando o conceito de similaridade vetorial. O algoritmo de agrupamento é baseado no trabalho de Alvarez (2018). Mais detalhes podem ser obtidos na Seção 6.1.4.

3.3.4 **Demonstração**

A etapa de demonstração da viabilidade do modelo proposto ocorreu com a construção de um protótipo e elaboração de cenários. Os cenários escolhidos para aplicação do modelo foram a base do programa de incentivo ao empreendedorismo Sinapse da Inovação[®] e o programa lançado pelo Ministério da Educação #MosquitoNão.

O primeiro cenário de análise é composto por ideias coletadas do programa Sinapse da Inovação[®] que oferece recursos financeiros, capacitações e suporte para transformar ideias inovadoras em negócios de sucesso. A sexta edição do programa, realizada no Estado de Santa Catarina, foi escolhida para aplicação da análise. Ao total, esta edição recebeu 1791 ideias com potencial de inovação de todas as regiões do Estado. Entende-se que o programa Sinapse da Inovação[®] se enquadra como cenário de estudo para aplicação e avaliação do referido modelo, pois recebe ideias de diferentes produtos e/ou serviços. Para este cenário foram coletadas 200

ideias referentes à Fase II do programa. Destaca-se que, destas 200 ideias, apenas 100 foram selecionadas para receberem investimentos.

O segundo cenário de análise é composto por 93 ideias coletadas entre os meses de junho a setembro de 2016 pelo Ministério da Educação, por meio de uma consulta pública. E, corresponde a uma coletânea de ideias denominada #MosquitoNão. Neste cenário, não é considerado o contexto organizacional, pois o intuito é analisar e demonstrar a generalidade do modelo proposto.

Para explicitar o conhecimento obtido é empregada a análise de redes sociais para realizar uma apreciação mais visual do contexto de estudo, constituindo assim um elemento relevante para a compreensão do trabalho. Desta forma, busca-se explorar o fluxo das informações e os relacionamentos existentes na formação de mapas de termos (tópicos), utilizando a análise de redes. Para tal, utilizou-se a ferramenta Gephi[®]. Esta ferramenta é empregada na exploração e compreensão de grafos voltada à análise exploratória de dados, o usuário interage com a representação e manipulação das estruturas, formas e cores, com o intuito de revelar propriedades ocultas dos dados de análise (GEPHI, 2019). De modo geral, auxilia analistas de dados na formulação de hipóteses e na descoberta de padrões permite isolar singularidades de estruturas ou falhas durante a análise (GEPHI, 2019).

A análise de redes, para o primeiro cenário de estudo, levou em consideração as palavras-chave das 200 ideias coletadas no site Sinapse da Inovação[®]. Foram coletados ao todo 715 termos-chave, gerando 9.353 coocorrências. Inicialmente, o peso da coocorrência é determinado pela quantidade de vezes que dois termos são mencionados conjuntamente. Posteriormente, é realizada a normalização dos valores pela raiz quarta. Na sequência, divide-se cada peso pelo valor máximo encontrado. Para formação da rede foram excluídos termos muito genéricos visto que estes tendem a se conectar com muitos outros termos, reduzindo a relevância nos resultados produzidos.

Os parâmetros de entrada utilizados para formação da rede foram: o grau médio, o diâmetro da rede e a modularidade para compor a visão geral da rede. O Algoritmo de distribuição utilizado é o *Force Atlas 2*. O grau é empregado para a partição dos nós. Por fim, a classificação dos nós é realizada por meio do índice de modularidade. A descrição detalhada da etapa de demonstração é apresentada no Capítulo 7, referente à análise e apresentação dos resultados.

Deste modo, o modelo proposto compreende etapas adicionais para que os objetivos do trabalho sejam alcançados, entre elas, a explicitação do conhecimento de domínio que fornece os subsídios para a etapa de avaliação. Esta característica possibilita explicitar visualmente

grupos de ideias, assim como os relacionamentos presentes entre as ideias coletadas e seu potencial de implementação, por meio de técnicas de visualização, entre elas, *treemap*, mapas de tópicos e análise de redes sociais.

3.3.5 Avaliação

A avaliação deste trabalho é constituída por três etapas. A primeira se refere à análise da efetividade e assertividade do protótipo desenvolvido, a segunda etapa se refere à avaliação da ontologia e do modelo por meio de especialistas e, por último, foram realizados testes para comparar o desempenho de um modelo preditivo supervisionado (classificação) e sua acurácia, em relação a um modelo preditivo não supervisionado (agrupamento) para o primeiro cenário de estudo.

Na primeira etapa foram verificadas a efetividade e a assertividade do modelo proposto ao se comparar as ideias selecionadas pelo modelo para o primeiro cenário de análise, além das ideias selecionadas pelos especialistas do Sinapse da Inovação[®]. Portanto, é empregado o método experimental que, segundo Lacerda et al. (2013), busca estudar o artefato em um ambiente controlado para verificar suas qualidades. Também se verificou a generalidade do modelo proposto para o segundo cenário de estudo, envolvendo as ideias da coletânea #MosquitoNão, porém não foi considerado o contexto organizacional na análise.

Na segunda etapa, por intermédio de um questionário, a ontologia e o modelo propostos foram avaliados por um conjunto de especialistas. Foram convidados 20 especialistas, contudo somente seis responderam aos questionamentos. A apresentação consistiu em inicialmente expor um documento contendo os benefícios da gestão de ideias, os conceitos envolvidos no domínio, as vantagens de explorar e avaliar as ideias, bem como os objetivos e diferenciais da ontologia. A ontologia e o modelo foram detalhados e propostos aos especialistas, conforme consta nos Capítulos 4 e 6, com o intuito de facilitar o entendimento a respeito da avaliação.

A avaliação foi composta por docentes doutores e um coordenador de compras e licitações com titulação de mestre e aluno de doutorado, ambos com experiência nas áreas de engenharia e gestão do conhecimento. Consistiu em um questionário (Apêndice A) definido em três dimensões denominadas: Dimensão 1 – Questões envolvidas na pesquisa, Dimensão 2 – Aplicabilidade e Dimensão 3 – Resultados. Inicialmente foram solicitados os dados de identificação, formação e área do conhecimento em que os especialistas possuíam experiência.

As respostas dos especialistas para as questões abordadas no questionário serão explicitadas na Seção 7.3.

Por fim, aplicou-se algoritmos de classificação amplamente utilizados na literatura como SVM e *Naive Bayes* para analisar o desempenho de modelos preditivos supervisionados (classificação) por meio dos índices de acurácia, precisão e revocação (*recall*), além do coeficiente kappa. O *software LightSide*¹ foi empregado para os testes e as configurações padrões de *unigrams* e *cross validation* foram utilizadas. O intuito desta análise é prover subsídios para uma discussão frente aos resultados obtidos nesta tese, com foco em um modelo preditivo não supervisionado (análise de agrupamentos).

3.3.6 Comunicação

A comunicação desta pesquisa é direcionada à comunidade científica com a apresentação dos resultados. Foram escritos e publicados artigos em periódicos e congressos internacionais durante o desenvolvimento deste estudo.

3.4 SÍNTESE DA METODOLOGIA DE PESQUISA

A síntese da metodologia de pesquisa adotada nesta tese e descrita neste capítulo é apresentada na forma de quadro, objetivando uma maior compreensão e sumarização das informações. O Quadro 7 apresenta a sintetização da abordagem da metodologia adotada.

Quadro 7 – Síntese da abordagem metodológica

Descrição	Abordagem
Natureza da pesquisa	Aplicada
Cunho da pesquisa	Tecnológica
Objetivos da pesquisa	Exploratório e descritivo
Forma de abordagem do problema	Quantitativa
Metodologia de desenvolvimento da pesquisa	<i>Design Science Research</i>

Fonte: Autora

Com o intuito de prover maior compreensão das etapas presentes no desenvolvimento desta tese, amparada na abordagem metodológica da *Design Science Research*, é apresentado o Quadro 8 com os materiais e métodos utilizados em cada etapa.

¹ <http://lightsidelabs.com>

Quadro 8 – Síntese da abordagem utilizada nas etapas da pesquisa

Etapa	Descrição
Definição do problema e solução	Revisão sistemática da literatura.
Definição dos objetivos para solução	Revisão sistemática da literatura.
Projeto e desenvolvimento	Base de conhecimento: representada por ontologia proposta com base na metodologia <i>Ontology Development 101</i> .
	Coleta dos dados e preenchimento da base de conhecimento: coleta manual das ideias do site Sinapse da Inovação® e da coletânea #MosquitoNão. Coleta automática dos documentos da base de patentes dos EUA e <i>Semantic Scholar</i> ®. Por fim, é realizada uma inserção manual das temáticas estratégicas relacionadas ao contexto organizacional.
	Pré-processamento: <ul style="list-style-type: none"> • Representação dos documentos: representação vetorial utilizando VSM; • Normalização dos documentos: através da medida estatística <i>tf-idf</i>; • Criação dos índices: <ul style="list-style-type: none"> ○ Índice de aderência ao contexto organizacional: medida de similaridade do cosseno; ○ Índice de atualidade: cálculo proposto pela tese com base na temporalidade; ○ Índice de tecnologia: medida de similaridade do cosseno; ○ Índice de Originalidade: equação modificada do cosseno.
	Transformação dos dados: decomposição matricial utilizando SVD.
	Agrupamento dos dados: algoritmo <i>k-means</i> modificado proposto por Alvarez (2018).
Demonstração	Apresentação dos dois cenários de estudo. <ul style="list-style-type: none"> • Primeiro cenário: 200 ideias coletadas do site Sinapse da Inovação®; • Segundo cenário: 93 ideias coletadas da coletânea #MosquitoNão.
Avaliação	<ul style="list-style-type: none"> • Análise da efetividade e comparação se as ideias com potencial de implementação do primeiro cenário de estudo sugeridas pelo modelo proposto na tese, são as mesmas escolhidas pelos especialistas do Sinapse da Inovação® para receber investimentos.
	<ul style="list-style-type: none"> • Avaliação por especialistas da ontologia e modelo propostos.
	<ul style="list-style-type: none"> • Aplicação de algoritmos de classificação para analisar o desempenho de um modelo preditivo supervisionado (classificação) e sua acurácia, em relação a um modelo preditivo não supervisionado (agrupamento).
Comunicação	Apresentação dos resultados e considerações finais.

Fonte: Autora

Por fim, as informações apresentadas neste capítulo se constituem no arcabouço utilizado no desenvolvimento desta pesquisa.

4 REPRESENTAÇÃO DA BASE DE CONHECIMENTO

Este capítulo apresenta de maneira detalhada a ontologia proposta nesta tese para representar a base de conhecimento, com o intuito de descrever o conhecimento do domínio de gestão de ideias, bem como as temáticas estratégicas relacionadas ao contexto organizacional. A partir do resultado da revisão sistemática da literatura e em pesquisas realizadas na *web*, não se identificou ontologias voltadas à avaliação de ideias que envolve o contexto organizacional e suas especificidades.

4.1 METODOLOGIA ADOTADA

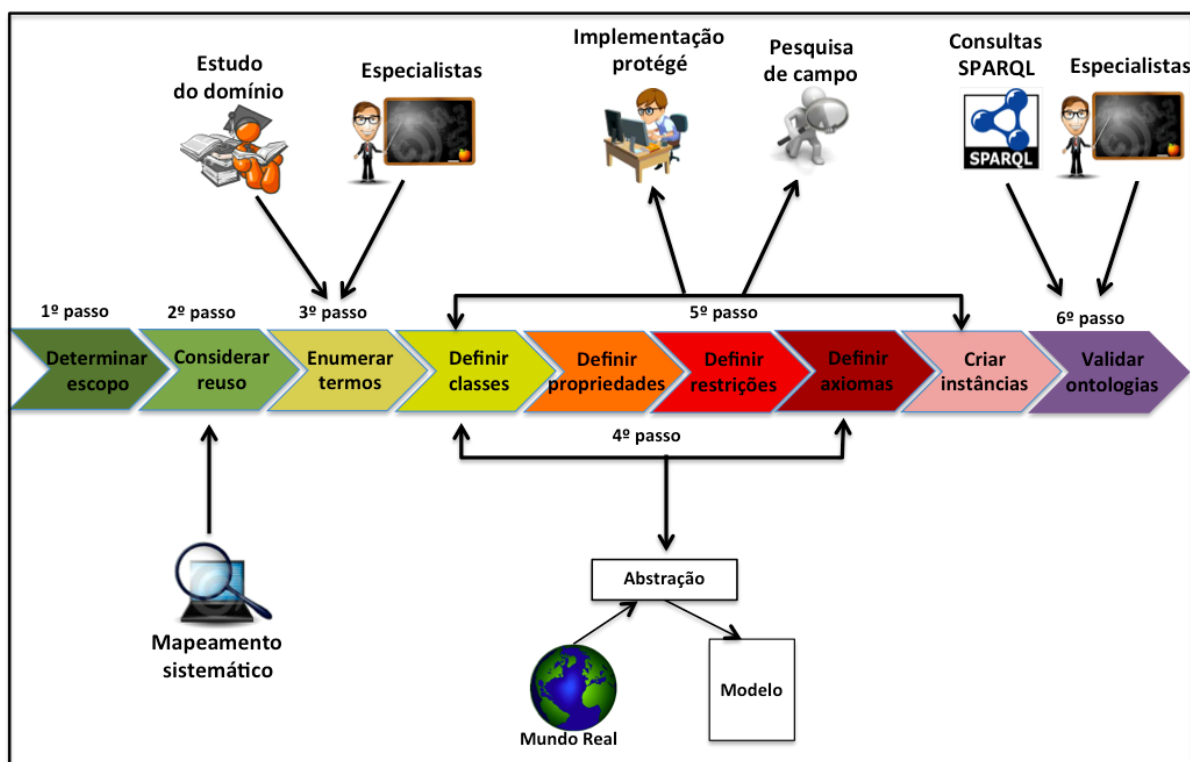
A ontologia proposta neste trabalho é construída com base na metodologia *Ontology Development 101*, que descreve um processo iterativo composto por sete passos para a construção de ontologias (NOY; MCGUINNESS, 2001). Foi adotada esta metodologia por sua praticidade ao descrever as etapas de desenvolvimento (NOY; MCGUINNESS, 2001).

A ferramenta utilizada para criação da ontologia foi o *software* Protégé[®] versão 5.1.01, devido a ampla e fácil utilização em projetos. Esta ferramenta de criação de ontologias contempla linguagens padrões da *web* semântica e possui *plug-ins* que potencializam suas funcionalidades.

A linguagem escolhida foi a OWL (*Web Ontology Language*) por ser uma linguagem robusta recomendada pela W3C, principal organização de padronização da *web* como padrão de representação de ontologias. A linguagem definida para realização de consultas na ontologia foi a SPARQL[®] por ser suportada pelo Protégé[®], recomendada pela W3C, e linguagem padrão de consulta e manipulação de dados para uso no contexto da *web* semântica.

O desenvolvimento da ontologia foi realizado em duas etapas. Na primeira foi construída uma ontologia para formalização dos elementos essenciais que devem ser considerados no domínio de gestão de ideias. A segunda fase foi uma evolução da ontologia com o objetivo de formalizar elementos contextuais das discussões colaborativas, advindas dos questionamentos aos especialistas de domínio. As etapas de implementação da ontologia, de acordo com o Método 101, podem ser visualizadas na Figura 15 e serão descritas a seguir, sendo que a etapa de validação da ontologia é descrita nos Capítulos 4 e 7.

Figura 15 - Método de desenvolvimento de ontologia



Fonte: Souza (2016)

O primeiro passo, conforme observado na Figura 15, se refere a definição do escopo, elucidando os objetivos da ontologia, os atores, sua finalidade e a definição das perguntas de competência que a ontologia deveria responder. O segundo passo envolve o reuso de outras ontologias. No terceiro passo ocorre a definição dos termos com base no estudo de domínio. No quarto passo são definidas as classes, as propriedades e as restrições por meio da construção de um modelo conceitual que representa o mundo real de forma abstrata. O quinto passo objetiva a implementação da ontologia, neste caso a ferramenta Protégé[®] foi utilizada. O último passo envolve a avaliação da ontologia, realizada por meio de consultas SPARQL para responder as perguntas de competência apresentadas na Seção 4.2 deste capítulo, além da avaliação com especialistas de domínio exposta na Seção 7.3.

4.1.1 Escopo da ontologia

A ontologia envolve a representação e formalização do domínio de gestão de ideias voltado à avaliação da ideia. O seu escopo é apoiar a avaliação de ideias alinhadas às temáticas relacionadas ao contexto organizacional, contribuindo no processo de tomada de decisão por especialistas da organização. Neste momento, também foram definidas 15 perguntas de competência, ou seja, questionamentos que a ontologia deve ser capaz de responder. Estas, além

de justificar a existência da ontologia, servem para sua futura avaliação. O Quadro 9 apresenta as perguntas de competência definidas no escopo da ontologia.

Quadro 9 – Perguntas de competência definidas

Perguntas de competência
1. Quais são as potenciais ideias de acordo com o valor de investimento?
2. Quais são as potenciais ideias de acordo com o valor de investimento e áreas de investimentos potenciais?
3. Quais são as potenciais ideias de acordo com o índice de aderência ao contexto organizacional?
4. Quais são as potenciais ideias de acordo com o índice de originalidade e o valor de investimento?
5. Quais são as potenciais ideias de acordo com o índice de originalidade?
6. Quais são as potenciais ideias de acordo com as áreas de investimento potenciais da organização?
7. Quais são as potenciais ideias de acordo com o índice de tecnologia?
8. Quais são as ideias semelhantes?
9. Quais são as ideias pertencentes à determinada categoria?
10. Quais são as ideias que receberam maior número de votos?
11. Quais são as ideias mais recentes?
12. Quais são as ideias que possuem maior valor estimado de investimento?
13. Quais são as potenciais ideias de acordo com o índice de atualidade?
14. Quais são as ideias pertencentes a determinado grau de maturidade?
15. Quais são as ideias em que a organização possui o número de colaboradores necessários para executá-las?

Fonte: Sérgio e Gonçalves (2019)

Estas perguntas não englobam todos os questionamentos, uma vez que a metodologia indica que essa listagem não precisa ser exaustiva (NOY; MCGUINNESS, 2001), pois elas servem para validação e refinamento da ontologia em uma etapa posterior.

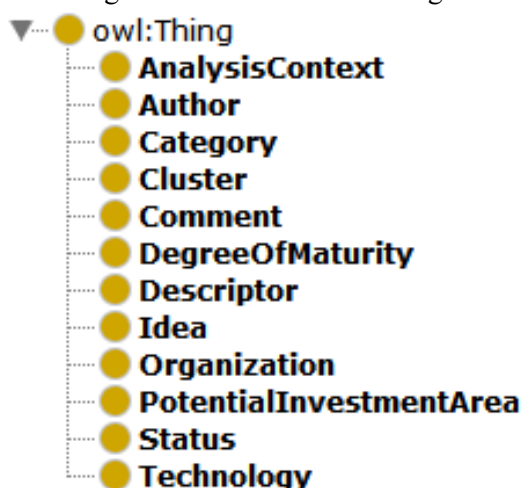
4.1.2 Reuso da ontologia

Para aproveitar ontologias existentes, esta ontologia utilizará parte daquela proposta por Sérgio, Souza e Gonçalves (2017).

4.1.3 Definição dos conceitos e classes

Entre as classes principais presentes na ontologia encontram-se a classe que representa a ideia, classe *Idea*, a classe *Cluster* responsável por representar o resultado do processo de agrupamento e a classe *AnalysisContext* que representa o contexto organizacional. A Figura 16 ilustra as classes que compõem a ontologia de domínio.

Figura 16 – Classes da ontologia



Fonte: Sérgio e Gonçalves (2019)

No Quadro 10 são apresentadas as classes presentes na ontologia de domínio e suas respectivas descrições.

Quadro 10 - Descrição das classes da ontologia

Classe	Descrição da classe
<i>AnalysisContext</i>	Classe responsável por definir o contexto organizacional considerando determinado período.
<i>Author</i>	Classe responsável por definir o autor de uma ideia.
<i>Category</i>	Classe responsável por descrever a categoria a qual a ideia pertence.
<i>Cluster</i>	Classe responsável por reunir em grupos as ideias similares.
<i>Comment</i>	Classe responsável por descrever os comentários que determinada ideia recebeu.
<i>DegreeOfMaturity</i>	Classe responsável por descrever o grau de maturidade da ideia.
<i>Descriptor</i>	Classe responsável por descrever os termos de determinada ideia, áreas de investimento ou tecnologias.
<i>Idea</i>	Classe responsável por descrever a ideia.
<i>Organization</i>	Classe responsável por descrever a organização.
<i>PotentialInvestmentArea</i>	Classe responsável por descrever as áreas com maior potencial de investimento pela organização de acordo com a atribuição de um <i>score</i> .
<i>Status</i>	Classe responsável por descrever o status de determina ideia.
<i>Technology</i>	Classe responsável por descrever as tecnologias que a organização possui ou pretende investir/possuir em algum momento.

Fonte: Sérgio e Gonçalves (2019)

4.1.4 Definição das propriedades e restrições

Com o intuito de especificar as classes e suas propriedades de dados, o Quadro 11 é apresentado.

Quadro 11 - Descrição das classes da ontologia e suas propriedades de dados

Classe	Propriedades de dado	Descrição da Propriedade
AnalysisContext	<i>estimatedInvestmentValue</i>	Valor estimado de investimento nas ideias
	<i>date</i>	Data de análise do contexto
Author	<i>name</i>	Nome do autor
	<i>id</i>	Número de identificação
	<i>resume</i>	Descrição do currículo
Category	<i>description</i>	Descrição da categoria da ideia
	<i>id</i>	Número de identificação
Cluster	<i>description</i>	Descrição do agrupamento
	<i>id</i>	Número de identificação
	<i>score</i>	Peso ou relevância atribuído ao agrupamento
Comment	<i>description</i>	Descrição do comentário
	<i>id</i>	Número de identificação
	<i>date</i>	Data de postagem
	<i>rating</i>	Peso atribuído ao comentário
	<i>polarity</i>	Responsável por determinar se o comentário é positivo ou negativo
DegreeOfMaturity	<i>description</i>	Descrição do grau de maturidade da ideia. Caracterizados por uma das seguintes instâncias: Ideia, Protótipo conceitual, Protótipo testado, Protótipo finalizado e Produto sendo comercializado
	<i>id</i>	Número de identificação
Descriptor	<i>description</i>	Descrição do termo
	<i>Score</i>	Valor responsável por determinar o grau de atualidade do termo
Idea	<i>degreeOfInnovation</i>	Descrição do grau de inovação da ideia
	<i>description</i>	Descrição da ideia
	<i>Date</i>	Data de sugestão
	<i>economicSector</i>	Descrição do setor econômico
	<i>estimatedInvestmentValue</i>	Valor estimado de investimento da ideia
	<i>Id</i>	Número de identificação
	<i>innovativeDifferential</i>	Diferencial inovativo
	<i>Title</i>	Título
	<i>Version</i>	Versão da ideia
	<i>scoreOrganizationalContextAdherence</i>	Índice de alinhamento da ideia, atribuído de acordo com a aderência ao contexto organizacional
	<i>scoreNovelty</i>	Índice de atualidade da ideia
	<i>scoreTechnology</i>	Índice de tecnologia
	<i>scoreOriginality</i>	Índice de originalidade
	<i>numberEmployees</i>	Número de empregados
	<i>numberEmployeesEstimated</i>	Número de empregados estimados
<i>votesAffirmative</i>	Número de votos positivos que a ideia recebeu	
<i>votesNegative</i>	Número de votos negativos que a ideia recebeu	
<i>implemented</i>	Representa um booleano responsável por informar se a ideia foi implementada	
Organization	<i>description</i>	Nome da organização
	<i>id</i>	Número de identificação
PotentialInvestmentArea	<i>id</i>	Número de identificação
	<i>score</i>	Grau de importância da área para a organização
Status	<i>description</i>	Descrição do Status. Caracterizados por uma das seguintes instâncias: Revisada, Arquivada, Em revisão, Lançada e Em breve
Technology	<i>id</i>	Número de identificação
	<i>score</i>	Grau de importância da tecnologia para a organização
	<i>status</i>	Responsável por indicar se a tecnologia já está presente na organização ou se pretende investir/possuir

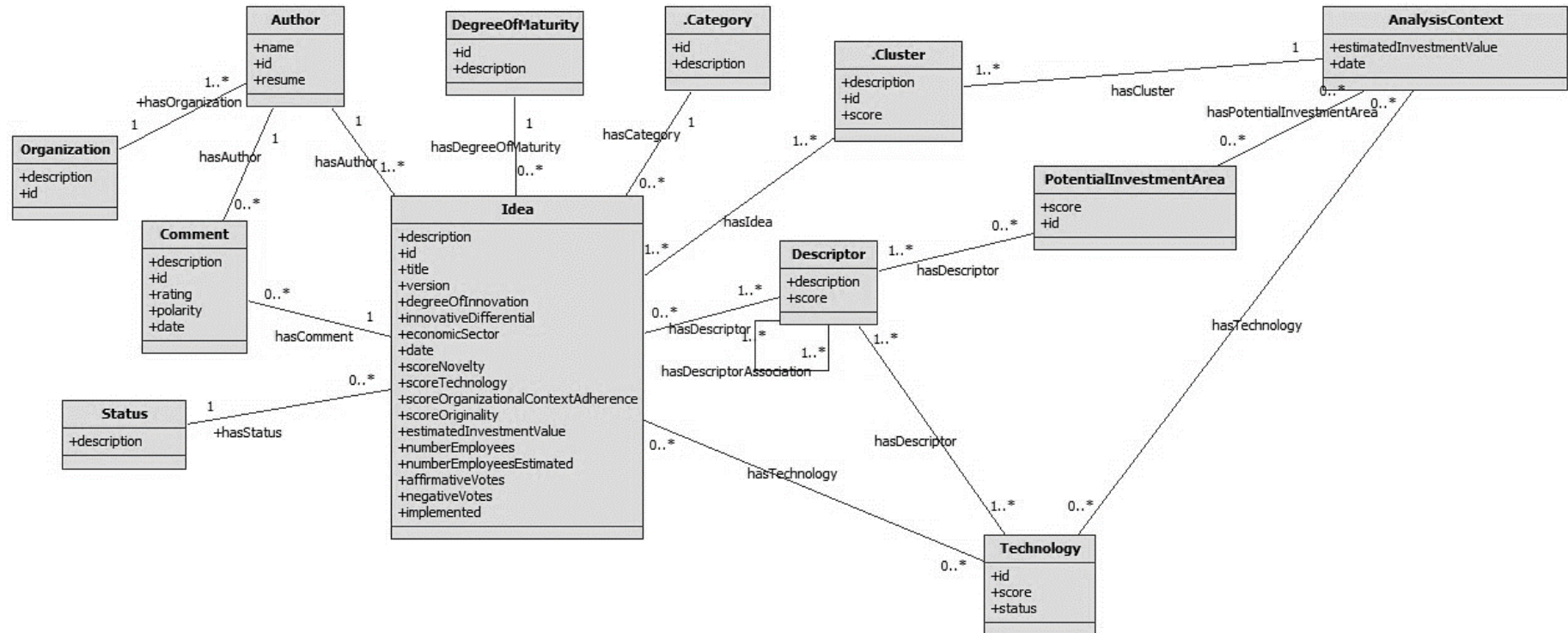
Fonte: Sérgio e Gonçalves (2019)

Cabe salientar que algumas propriedades da ontologia de domínio voltadas à caracterização do contexto organizacional necessitam ser preenchidas manualmente. Estas propriedades estão presentes na classe *AnalysisContext*. As propriedades que necessitam ser preenchidas são: *estimatedInvestmentValue* que corresponde ao valor aproximado de

investimento da ideia e *date*, correspondente a data de realização da análise. A classe *AnalysisContext* relaciona-se com a classe *PotentialInvestmentAreas* que possui como propriedade de dado o campo *description*, correspondente a área de investimento que a organização pretende investir. Também é necessário preencher a classe *Technology* que se relaciona com a classe *AnalysisContext*. A classe *Technology* possui a descrição das tecnologias utilizadas na organização ou que pretende investir.

Com o intuito de exemplificar a ontologia, cardinalidades e relacionamentos, é apresentada a Figura 17.

Figura 17 – Detalhamento das classes presentes na ontologia



Fonte: Sérgio e Gonçalves (2019)

Na Figura 17 é possível identificar o nome da classe na parte superior, bem como suas respectivas propriedades de dados. As ligações entre as classes representam as propriedades de objetos.

Uma das classes fundamentais da ontologia é a classe *Idea*, responsável por descrever as características das ideias e os índices que serão atribuídos às ideias com a execução do modelo. As principais propriedades de dados da classe *Idea* são: o título da ideia, a descrição da ideia, a data da última atualização da ideia, um identificador, o índice de maturidade da ideia, o grau de inovação e o diferencial inovativo (descrição do diferencial inovativo da ideia indicada pelo colaborador). Ela também estará associada: ao número de votos e aos respectivos comentários que recebeu, assim como ao(s) autor(es), ao valor de investimento estimado, entre outras informações.

É importante destacar que a ideia possui ainda índices calculados durante a avaliação de potencialidade: índice de aderência ao contexto organizacional (elucida o contexto de análise e o contexto organizacional); índice de atualidade (o quanto atual determinada ideia pode ser considerada); índice de originalidade (o quanto original determinada ideia pode ser considerada); índice de tecnologia (representa o grau de alinhamento da ideia às tecnologias organizacionais). A classe *Idea* também se relaciona com a classe *Author* via propriedade de objeto *hasAuthor*, a classe *Comment* via propriedade *hasComment*, a classe *Status* via propriedade *hasStatus*, *Category* via propriedade *hasCategory*, a classe *Technology* por meio da propriedade *hasTechnology*, *Descriptor* por meio da propriedade *hasDescriptor* e a classe *DegreeOfMaturity*, através da propriedade *hasDegreeOfMaturity*. Existe também o relacionamento da classe *Idea* com a classe *Cluster*.

A classe *Cluster* é responsável por representar os agrupamentos de ideias gerados. Ela possui um número identificador (*id*), uma descrição (*description*) e um *score*. Determinado *cluster* poderá ter associado uma ou *n* ideias através da propriedade *hasIdea*.

A classe *AnalysisContext* é responsável por representar o contexto organizacional em determinado momento. O objetivo é criar um contexto de análise delimitando quais ideias farão parte de determinada avaliação. As propriedades da classe *AnalysisContext* são: *estimatedInvestmentValue*, que corresponde ao valor aproximado de investimento da ideia e *date*, que corresponde a data de realização da análise. A classe *AnalysisContext* relaciona-se com a classe *PotentialInvestmentAreas*, via propriedade *hasPotentialInvestmentAreas*; *Technology* por meio da propriedade *hasTechnology* e *Cluster*, via propriedade *hasCluster*. As informações que irão compor a classe *AnalysisContext* podem vir por meio de especialistas da organização ou por fontes externas. Desenvolver o pensamento estratégico, assim como o

planejamento de ações de investimento em diferentes áreas alinhadas ao contexto pode contribuir significativamente para o crescimento organizacional. O estudo pode levar em consideração outras fontes de informações, sejam nacionais ou internacionais.

A classe *PotentialInvestmentAreas* é responsável por descrever as áreas de investimento potenciais e possui como propriedades de dados um identificador (*id*) e um *score* indicando a importância da área para a organização. A classe relaciona-se com as classes *Descriptor*, *Technology* e *AnalysisContext*, conforme mencionado.

A classe *Author* é responsável por descrever o autor de determinada ideia ou comentário. Possui como propriedade de dados (*description*, *id*, *resume*) e está associada via propriedade de objeto, a classe *Organization*, que representa a organização através da propriedade *hasOrganization*, a classe *Comment*, por meio da propriedade *hasAuthor* e a classe *Idea*, por meio da propriedade *hasAuthor*.

A classe *Descriptor* descreve as palavras-chave relacionadas à determinada classe. No caso desta ontologia servirá para descrever os termos-chave das classes *Cluster*, *PotentialInvestmentAreas* e *Technology*.

A classe *Technology* possui um identificador (*id*), um atributo chamado *score* indicando a importância da tecnologia para a organização, e um atributo chamado *status* para indicar se a tecnologia já está presente na organização ou se a mesma pretende investir/possuir. Esta classe relaciona-se com a classe *Descriptor* com o objetivo de descrever todos os termos relacionados à tecnologia de interesse.

A classe *Comment* descreve os comentários que determinada ideia recebeu, e relaciona-se com a classe *Idea* e *Author*.

Por fim, a classe *Status* é responsável por descrever o *status* de determinada ideia, indicando se a mesma foi implementada ou está em revisão, por exemplo.

Cabe destacar que as classes *AnalysisContext*, *PotentialInvestmentArea*, *Technology* e *DegreeOfMaturity* foram criadas para descrever o contexto organizacional e o grau de maturidade das ideias, e as demais classes foram reusadas da ontologia proposta por Sérgio, Souza e Gonçalves (2017). A classe *ClusterIdea* alterou o nome para *Cluster*.

4.2 LEVANTAMENTO DE INFORMAÇÕES PARA FORMALIZAÇÃO DA ONTOLOGIA

Esta seção aborda a formalização da ontologia proposta nesta pesquisa. Para tanto, a partir das 15 perguntas de competência definidas no escopo da ontologia e apresentadas no

Quadro 9, objetivou-se avaliar após a construção da ontologia se esta respondia às perguntas citadas. Para cada pergunta de competência foi elaborada uma consulta SPARQL. O resultado é apresentado a seguir.

Para responder a pergunta de competência número 1 “quais são as potenciais ideias de acordo com o valor de investimento?”, foi elaborada a consulta SPARQL, representada na Figura 18. O resultado da consulta é apresentado na Tabela 1.

Figura 18 – Pergunta de competência 1

```
private static String queryString1 = "PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>"
+ "PREFIX owl: <http://www.w3.org/2002/07/owl#>"
+ "PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>"
+ "PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>"
+ "PREFIX ontology: <http://www.semanticweb.org/marina/ontologies/2018/1/untitled-ontology-369#>"
+ "SELECT ?title ?value"
+ " WHERE { ?y ontology:title ?title."
+ "?y ontology:estimatedInvestmentValue ?value."
+ "?x rdf:type ontology:AnalysisContext."
+ "?x ontology:estimatedInvestmentValue ?estimatedInvestmentValueO."
+ "Filter (?value <= ?estimatedInvestmentValueO)}"
+ "ORDER BY DESC(?value)";
```

Fonte: Autora

Tabela 1 - Retorno da pergunta de competência 1

Título	Valor
<i>Active Vest</i> - Conforto ao sentar	R\$ 9000,00
Agrotechlink: Agricultura e tecnologia conectadas.	R\$ 9000,00
ALDA - Assistente de Letramento a Deficientes Auditivos	R\$ 9000,00
<i>BeePay</i> - O melhor de dois mundos	R\$ 9000,00
<i>Cheap2Ship</i> - Sistema inovador de cotação de fretes	R\$ 9000,00
<i>Code Money</i>	R\$ 9000,00
Diário da Alergia	R\$ 9000,00
Financia Inovação - Plataforma de projetos de fomento	R\$ 9000,00
<i>Hero HMB-FA</i> - Suplemento de resposta ativa.	R\$ 9000,00
<i>Illegic</i> - Saúde além do rótulo	R\$ 9000,00
Impressora 3d sem filamento	R\$ 9000,00
Inovação tecnológica na produção de mudas de macieira	R\$ 9000,00
<i>KneeMob</i> - Mobilizador patelar multidirecional	R\$ 9000,00
NIDE- Monitoramento de morros e encostas	R\$ 9000,00
PagueDepois - Marcou hoje!?! PagueDepois	R\$ 9000,00
PARE- Proteção de Animais em Rodovias e Estradas	R\$ 9000,00
<i>PillsController</i> - Dosador automático de medicamentos	R\$ 9000,00
Plataforma <i>UpMe!</i>	R\$ 9000,00
RoPE - Robô Programável Educacional	R\$ 9000,00
SIGPEP (Sistema Integrado de Prontuário Eletrônico de Pacientes)	R\$ 9000,00
Simplex - Sistema de planejamento e plano de negócios	R\$ 9000,00
SIS-MOBAL - Sistema de Monitoramento de Baixa Altitude	R\$ 9000,00
Sistema de detecção e bloqueio de drone	R\$ 9000,00
Sistema de monitoração e controle inteligente para chuveiros	R\$ 9000,00
Sistema inteligente híbrido foto-termo-voltaico	R\$ 9000,00
Use inteligência artificial para saber onde abrir seu novo negócio	R\$ 9000,00
Welob - <i>Software</i> para planejamento e gerenciamento de obras	R\$ 9000,00

Fonte: Autora

Por meio do resultado é possível verificar que a ontologia proposta responde quais são as ideias potenciais relacionadas a um determinado valor de investimento. O valor de investimento utilizado para apresentação foi fixado em R\$ 9.000,00 (nove mil reais).

A consulta SPARQL da pergunta de competência número 2 “quais são as potenciais ideias de acordo com o valor de investimento e áreas de investimentos potenciais?” está ilustrada na Figura 19. O resultado é apresentado na Tabela 2, é possível verificar que a ontologia proposta responde quais são as ideias potenciais relacionadas tanto a um determinado valor de investimento quanto as áreas potenciais de investimentos. O valor de investimento estimado foi de até R\$ 10.000,00 (dez mil reais) e a categoria selecionada foi “Tecnologia Social”.

Figura 19 – Pergunta de competência 2

```
private static String queryString2 = "PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>"
+ "PREFIX owl: <http://www.w3.org/2002/07/owl#>"
+ "PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>"
+ "PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>"
+ "PREFIX ontology: <http://www.semanticweb.org/marina/ontologies/2018/1/untitled-ontology-369#>"
+ "SELECT ?title ?category ?value"
+ " WHERE { ?y ontology:title ?title."
+ "?y ontology:estimatedInvestmentValue ?value."
+ "?y ontology:hasCategory ?x."
+ "?x ontology:description ?category."
+ "?j rdf:type ontology:PotentialInvestmentArea."
+ "?j ontology:description ?description."
+ "?k rdf:type ontology:AnalysisContext."
+ "?k ontology:estimatedInvestmentValue ?estimatedInvestmentValueO."
+ "Filter (?value <= ?estimatedInvestmentValueO && sameTerm(?category, ?description))}"
+ "ORDER BY DESC(?value)";
```

Fonte: Autora

Tabela 2 - Retorno da pergunta de competência 2

Título	Categoria	Valor
<i>Active Vest</i> - Conforto ao sentar	Tecnologia social	R\$ 9000
ALDA - Assistente de Letramento a Deficientes Auditivos	Tecnologia social	R\$ 9000
Diário da Alergia	Tecnologia social	R\$ 9000
Financia Inovação - Plataforma de projetos de fomento	Tecnologia social	R\$ 9000
NIDE- Monitoramento de Morros e Encostas	Tecnologia social	R\$ 9000
PARE- Proteção de Animais em Rodovias e Estradas	Tecnologia social	R\$ 9000
<i>PillsController</i> - Dosador automático de medicamentos	Tecnologia social	R\$ 9000
Plataforma <i>UpMe!</i>	Tecnologia social	R\$ 9000
RoPE - Robô Programável Educacional	Tecnologia social	R\$ 9000
Porquinho Virtual	Tecnologia social	R\$ 8000
<i>RuralMarket</i> - Inteligência de mercado aplicada ao meio rural	Tecnologia social	R\$ 8000
Trazzelo - Temos os melhores profissionais para o bem-estar do idoso	Tecnologia social	R\$ 8000
Balizar - Assistente de estacionamento	Tecnologia social	R\$ 7000
SNAZZY - Compartilhe e ganhe dinheiro com suas roupas guardadas	Tecnologia social	R\$ 7000
OiMaggie - A assistente virtual estudantil	Tecnologia social	R\$ 6000
<i>FindMed</i> - A saúde em suas mãos	Tecnologia social	R\$ 5000
Redação Online - Plataforma virtual de revisões textos	Tecnologia social	R\$ 5000
SalvoVidas.com -Seja avisado quando um banco de sangue precisa de você	Tecnologia social	R\$ 5000
<i>Makeup PRO</i> - Conectando maquiadores e clientes	Tecnologia social	R\$ 4000
<i>Punch Arena</i>	Tecnologia social	R\$ 4000
C4NDIROO - Sistema de reaproveitamento de embalagens longa	Tecnologia social	R\$ 3000

<i>Posithividades</i> - App para pessoas que vivem e convivem com o HIV-AIDS.	Tecnologia social	R\$ 3000
Reconsidere - Gestão sustentável de resíduos sólidos	Tecnologia social	R\$ 3000
ROLABOLA - Individualização do futebol	Tecnologia social	R\$ 3000
SaladaFácil	Tecnologia social	R\$ 3000
Vizinhança Protegida - Sistema de Segurança Colaborativo	Tecnologia social	R\$ 3000
Conect@projetos	Tecnologia social	R\$ 2000
<i>EletoBlocks</i> - Blocos eletrônicos modulares e programáveis	Tecnologia social	R\$ 2000
FarolVR - Realidade virtual para novas conexões	Tecnologia social	R\$ 2000
<i>Lifter</i> de transferência multifuncional para pessoas com deficiência	Tecnologia social	R\$ 2000
<i>LINKMESH</i> - Mapa de recursos estratégicos	Tecnologia social	R\$ 2000
Sistema recirculador de água fria do banho	Tecnologia social	R\$ 2000
Somos um iFood para cerveja	Tecnologia social	R\$ 2000
Adubooo - Plataforma de localização de resíduos orgânicos	Tecnologia social	R\$ 1000
<i>CITYLUX</i> - Cidade Geradora	Tecnologia social	R\$ 1000
HeyHo Green Protein - Suplementos proteicos práticos e sem monotonia	Tecnologia social	R\$ 1000

Fonte: Autora

Com o intuito de responder a pergunta de competência número 3 “quais são as potenciais ideias de acordo com o índice de aderência ao contexto organizacional?”, elaborou-se a consulta SPARQL, representada na Figura 20. O resultado desta consulta é apresentado na Tabela 3 permitindo verificar que a ontologia proposta responde quais são as ideias potenciais, de acordo com o índice de aderência ao contexto organizacional. Destaca-se que algumas consultas SPARQL foram limitadas a 10 retornos, visando reduzir a quantidade de dados retornada. Neste sentido, o conceito de relevância temporal foi definido por um delimitador fixo por questões de simplificação. Contudo, na interação com o especialista, este deveria informar um valor para o índice de aderência que julgar mais adequado. O índice de aderência ao contexto organizacional foi estipulado em maior ou igual a 0,3.

Figura 20 – Pergunta de competência 3

```
private static String queryString3 = "PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>"
+ "PREFIX owl: <http://www.w3.org/2002/07/owl#>"
+ "PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>"
+ "PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>"
+ "PREFIX ontology: <http://www.semanticweb.org/marina/ontologies/2018/1/untitled-ontology-369#>"
+ " SELECT ?title ?score"
+ " WHERE { ?y ontology:title ?title ."
+ " ?y ontology:scoreOrganizationalContextAdherence ?score }"
+ " ORDER BY DESC(?score)" + "LIMIT 10";
```

Fonte: Autora

Tabela 3 - Retorno da pergunta de competência 3

Título	Índice de aderência ao contexto organizacional
C4NDIROO - Sistema de reaproveitamento de embalagens longa	0.319830895761228
Eco Alliance - Recicladora de resíduos orgânicos	0.315352460917944
Farmstec - Inovações tecnológicas	0.311584301492395
Aquecedores por indução industriais	0.301338323974139
ManejeBem: Rede social para agricultores	0.338059668421820

Fonte: Autora

Para responder a pergunta de competência número 4 “quais são as potenciais ideias de acordo com o índice de originalidade e o valor de investimento?” foi elaborada a consulta SPARQL, representada na Figura 21. O resultado desta consulta é apresentado na Tabela 4 e por meio dele é possível verificar que a ontologia proposta responde quais são as ideias potenciais relacionadas a um determinado valor de investimento e o índice de originalidade. O valor de investimento estimado foi de até R\$ 10.000,00 (dez mil reais).

Figura 21 – Pergunta de competência 4

```
private static String queryString4 = "PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>"
+ "PREFIX owl: <http://www.w3.org/2002/07/owl#>"
+ "PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>"
+ "PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>"
+ "PREFIX ontology: <http://www.semanticweb.org/marina/ontologies/2018/1/untitled-ontology-369#>"
+ "SELECT ?title ?score ?value"
+ " WHERE { ?y ontology:title ?title."
+ "?y ontology:estimatedInvestmentValue ?value."
+ "?y ontology:scoreOriginality ?score."
+ "?x rdf:type ontology:AnalysisContext."
+ "?x ontology:estimatedInvestmentValue ?estimatedInvestmentValueO."
+ "Filter (?value <= ?estimatedInvestmentValueO)}"
+ "ORDER BY DESC(?score)" + "LIMIT 10";
```

Fonte: Autora

Tabela 4 - Retorno da pergunta de competência 4

Título	Índice de originalidade	Valor
Simple-sistema de planejamento e plano de negócios	1	9000
<i>Hero</i> HMB-FA - Suplemento de resposta ativa.	1	9000
<i>PillsController</i> - Dosador automático de medicamentos	1	9000
Financia Inovação - Plataforma de projetos de fomento	1	9000
Impressora 3d sem filamento	1	9000
APATUIA- Frutas estruturadas com umidade	1	8000
Chuveiro <i>Invert</i>	1	8000
<i>Crave Platform</i> - Industria 4.0	1	8000
Rede de Investimento	1	8000
Tour 360SC	1	8000

Fonte: Autora

A consulta SPARQL da pergunta de competência número 5 “quais são as potenciais ideias de acordo com o índice de originalidade? está ilustrada na Figura 22. O resultado é apresentado na Tabela 5.

Figura 22 – Pergunta de competência 5

```
private static String queryString5 = "PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>"
+ "PREFIX owl: <http://www.w3.org/2002/07/owl#>"
+ "PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>"
+ "PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>"
+ "PREFIX ontology: <http://www.semanticweb.org/marina/ontologies/2018/1/untitled-ontology-369#>"
+ " SELECT ?title ?score"
+ " WHERE { ?y ontology:title ?title ."
+ " ?y ontology:scoreOriginality ?score }"
+ " ORDER BY DESC(?score)" + "LIMIT 10";
```

Fonte: Autora

Tabela 5 - Retorno da pergunta de competência 5

Título	Índice de originalidade
AMAIE Consultoria de imagem	1
APATUIA- Frutas estruturadas com umidade	1
APP gestão de propriedades rurais familiares	1
Aquecedores por indução industriais	1
ARMTECH - Assistente cirúrgico robótico	1
<i>Azimov Trade</i>	1
Bica de caminhão graneleiro automática	1
Biorrefinaria de microalgas	1
<i>Bodout. Mais esporte, menos cheiro.</i>	1
C4NDIROO - Sistema de reaproveitamento de embalagens longa	1

Fonte: Autora

Por meio do resultado é possível verificar que a ontologia proposta responde quais são as ideias potenciais que possuem o maior índice de originalidade.

Com o intuito de responder a pergunta de competência número 6 “Quais são as potenciais ideias de acordo com as áreas de investimento potenciais da organização?”, elaborou-se a consulta SPARQL, representada na Figura 23. O resultado desta consulta é apresentado no Quadro 12. Por meio dele é possível verificar que a ontologia proposta responde quais são as ideias potenciais de acordo com as áreas potenciais de investimento. A área selecionada foi “Nanotecnologia”.

Figura 23 – Pergunta de competência 6

```
private static String queryString6 = "PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>"
+ "PREFIX owl: <http://www.w3.org/2002/07/owl#>"
+ "PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>"
+ "PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>"
+ "PREFIX ontology: <http://www.semanticweb.org/marina/ontologies/2018/1/untitled-ontology-369#>"
+ "SELECT ?title ?category"
+ " WHERE { ?y ontology:title ?title."
+ "?y ontology:hasCategory ?x."
+ "?x ontology:description ?category."
+ "?j rdf:type ontology:PotentialInvestmentArea."
+ "?j ontology:description ?description."
+ "Filter (sameTerm(?category, ?description))}";
```

Fonte: Autora

Quadro 12 - Retorno da pergunta de competência 6

Título	Categoria
Aplicabilidade de <i>Golden-Max</i> em diferentes produtos da área da saúde	Nanotecnologia
Biomateriais compósitos nanoestruturados para manufatura aditiva	Nanotecnologia
Equipamento de eletrofiação com foco na engenharia de tecido	Nanotecnologia
Inovação em nanotecnologia verde	Nanotecnologia
INTEGRA - Bioativação de implantes osseointegráveis	Nanotecnologia
Nanocosméticos com antioxidante obtido de organismo aquático	Nanotecnologia
<i>TechPain</i> : A revolução do tratamento da dor crônica na artrite	Nanotecnologia
<i>THERMO-OFF</i> - Eficiência térmica	Nanotecnologia

Fonte: Autora

Para responder a pergunta de competência número 7 “quais são as potenciais ideias de acordo com o índice de tecnologia?” foi elaborada a consulta SPARQL, representada na Figura 24. O resultado desta consulta é apresentado na Tabela 6 e, por meio dele, é possível verificar que a ontologia proposta responde quais são as ideias potenciais com maior índice de tecnologia.

Figura 24 – Pergunta de competência 7

```
private static String queryString7 = "PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>"
+ "PREFIX owl: <http://www.w3.org/2002/07/owl#>"
+ "PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>"
+ "PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>"
+ "PREFIX ontology: <http://www.semanticweb.org/marina/ontologies/2018/1/untitled-ontology-369#>"
+ " SELECT ?title ?scoreTechnology"
+ " WHERE { ?y ontology:title ?title ."
+ " ?y ontology:scoreTechnology ?scoreTechnology }"
+ " ORDER BY DESC(?scoreTechnology)" + "LIMIT 10";
```

Fonte: Autora

Tabela 6 - Retorno da pergunta de competência 7

Título	Índice de tecnologia
<i>Seashell - Powering innovation</i>	1.00000000000000000000
<i>Farmstec - Inovações tecnológicas</i>	0.96565613982713762565
<i>Keeps Mobile Game Learning</i>	0.96392293628552411925
<i>Leoa - Plataforma SAAS com chatbot para facilitar a declaração de IR</i>	0.96380259886777651075
<i>EasyHealth - Plataforma IoT para realização de exames</i>	0.96141917714682937121
<i>ARMTECH - Assistente cirúrgico robótico</i>	0.89652835320247608821
<i>Azimov Trade</i>	0.88649413864543541288
<i>ModuHub Core - Aceleração da IoT no Brasil</i>	0.83798330123194902862
<i>Dynamique - Smart Business</i>	0.82905465132479196005
<i>Abellion - A inteligência que sua plantação precisa</i>	0.82684309231979313572

Fonte: Autora

A consulta SPARQL da pergunta de competência número 8 “quais são as ideias semelhantes?” está ilustrada na Figura 25. O resultado da consulta é apresentado na Tabela 7 e corresponde a todas as ideias pertencentes ao *cluster* de número 5.

Figura 25 – Pergunta de competência 8

```
private static String queryString8 = "PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>"
+ "PREFIX owl: <http://www.w3.org/2002/07/owl#>"
+ " PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>"
+ " PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>"
+ " PREFIX ontology: <http://www.semanticweb.org/marina/ontologies/2018/1/untitled-ontology-369#>"
+ " SELECT ?title ?cluster"
+ " WHERE { ?y ontology:title ?title ."
+ " ?y ontology:isIdeaOf ?x."
+ " ?x ontology:description ?cluster}"
+ " ORDER BY (?cluster)"
+ " LIMIT 10";
```

Fonte: Autora

Tabela 7 - Retorno da pergunta de competência 8

Título	Identificador do <i>cluster</i>
Casa de hóspedes	5
<i>Inottel: Envio de dados e troca de informações entre hotéis e hóspedes</i>	5

Fonte: Autora

Para responder a pergunta de competência número 9 “quais são as ideias pertencentes à determinada categoria?” foi elaborada a consulta SPARQL, representada na Figura 26. O resultado desta consulta é apresentado no Quadro 13 e, por meio dele, é possível verificar que a ontologia proposta responde quais são as ideias relacionadas à determinada categoria. A categoria selecionada foi “Gestão”.

Figura 26 – Pergunta de competência 9

```
private static String queryString9 = "PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>"
+ "PREFIX owl: <http://www.w3.org/2002/07/owl#>"
+ "PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>"
+ "PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>"
+ "PREFIX ontology: <http://www.semanticweb.org/marina/ontologies/2018/1/untitled-ontology-369#>"
+ "SELECT ?title ?category"
+ "WHERE { ?y ontology:title ?title ."
+ " ?y ontology:hasCategory ?x."
+ " ?x ontology:description ?category."
+ " FILTER (regex(?category, '^gestao'))}";
```

Fonte: Autora

Quadro 13 - Retorno da pergunta de competência 9

Título	Categoria
<i>Abellion</i> - A inteligência que sua plantação precisa	Gestão
APP gestão de propriedades rurais familiares	Gestão
BuscaCity - Inteligência Artificial	Gestão
Casa de hóspedes	Gestão
<i>CHAWork</i> - Sistema inteligente de recrutamento e seleção	Gestão
<i>Code Money</i>	Gestão
<i>Dynamique - Smart Business</i>	Gestão
<i>E-licencie</i> : Garantindo a sustentabilidade das empresas	Gestão
Elo Torcedor - Modelo de gestão para torcedores esportivos	Gestão
<i>FIRE</i> - Gestão de clubes de caça e tiro	Gestão
<i>Kluc</i> - Obtenha <i>insights</i> dos seus dados através de simples perguntas	Gestão
<i>LabCloud</i> - Inovação para processos laboratoriais veterinários	Gestão
Mais Leite	Gestão
Máquina portátil para gestão de estacionamento. (<i>Top CloudPark</i>)	Gestão
Mate in Box - O clube de assinaturas de erva-mate	Gestão
Monefica - Educação financeira	Gestão
Organização Ágil - Metodologia de transformação ágil para empresas	Gestão
PagueDepois - Marcou hoje!?! PagueDepois	Gestão
Plataforma gestão por ideias	Gestão
Rede de investimento	Gestão
SCFA - Sistema de Certificação e Fiscalização de Acessibilidade	Gestão
SIMPLES - Sistema de planejamento e plano de negócios	Gestão
Sistema INFRAVIA - Sistema de infraestrutura urbana	Gestão
TERTIUM - Gestão de mão de obra terceirizada empresas públicas e privadas	Gestão
Use inteligência artificial para saber onde abrir seu novo negocio	Gestão
Viaja SC (Descobertas que encantam)	Gestão
Vou - Eventos e descontos	Gestão
Welob - <i>Software</i> para planejamento e gerenciamento de obras	Gestão

Fonte: Autora

A consulta SPARQL da pergunta de competência número 10 “quais são as ideias que receberam maior número de votos?” está ilustrada na Figura 27. O resultado é apresentado na Tabela 8, por meio dele é possível verificar que a ontologia proposta responde em formato decrescente às ideias com maior número de votos.

Figura 27 – Pergunta de competência 10

```
private static String queryString10 = "PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>"
+ "PREFIX owl: <http://www.w3.org/2002/07/owl#>"
+ " PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>"
+ " PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>"
+ " PREFIX ontology: <http://www.semanticweb.org/marina/ontologies/2018/1/untitled-ontology-369#>"
+ " SELECT ?title ?affirmativeVotes"
+ " WHERE { ?y ontology:title ?title ."
+ " ?y ontology:affirmativeVotes ?affirmativeVotes.}"
+ "ORDER BY DESC(?affirmativeVotes)"
+ " LIMIT 10";
```

Fonte: Autora

Tabela 8 - Retorno da pergunta de competência 10

Título	Número de votos
SIGPEP (Sistema Integrado de Prontuário Eletrônico de Pacientes)	9
<i>Dynamique - Smart Business</i>	9
<i>Keeps Mobile Game Learning</i>	9
<i>Fleet Manager - IOT aplicado a veículos</i>	9
Medição dimensional através de sistema óptico	9
Bioacelerador de autodepuração de drenagem acida de mina	9
Inteligência Artificial para análise de dívidas com a União	9
Controle de irrigação por pivô central	9
<i>Fertileasy - Monitoramento e controle de produção agrícola</i>	9
Elo Torcedor - Modelo de gestão para torcedores esportivos	9

Fonte: Autora

Com o intuito de responder a pergunta de competência número 11 “quais são as ideias mais recentes?”, elaborou-se a consulta SPARQL, representada na Figura 28. O resultado desta consulta é apresentado no Quadro 14.

Figura 28 – Pergunta de competência 11

```
private static String queryString11 = "PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>"
+ "PREFIX owl: <http://www.w3.org/2002/07/owl#>"
+ "PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>"
+ "PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>"
+ "PREFIX ontology: <http://www.semanticweb.org/marina/ontologies/2018/1/untitled-ontology-369#>"
+ " SELECT ?title ?date"
+ " WHERE { ?y ontology:title ?title ."
+ " ?y ontology:date ?date.}"
+ " ORDER BY DESC (?date)"
+ " LIMIT 10";
```

Fonte: Autora

Quadro 14 - Retorno da pergunta de competência 11

Título	Data de submissão
Liopet ração	03/12/2017
Conect@projetos	03/12/2017
<i>Fruitkeep - Conservação inteligente de frutos</i>	03/12/2017

Fonte: Autora

Por meio do resultado obtido é possível verificar que a ontologia proposta responde quais são as ideias mais recentes.

Para responder a pergunta de competência número 12 “quais são as ideias que possuem maior valor estimado de investimento?” foi elaborada a consulta SPARQL, representada Figura 29. O resultado desta consulta é apresentado na Tabela 9 e, por meio dele, é possível verificar que a ontologia proposta responde quais são as ideias com maior valor estimado de investimento.

Figura 29 – Pergunta de competência 12

```
private static String queryString12 = "PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>"
+ "PREFIX owl: <http://www.w3.org/2002/07/owl#>"
+ "PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>"
+ "PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>"
+ "PREFIX ontology: <http://www.semanticweb.org/marina/ontologies/2018/1/untitled-ontology-369#>"
+ " SELECT ?title ?value"
+ " WHERE { ?y ontology:title ?title ."
+ " ?y ontology:estimatedInvestmentValue ?value .}"
+ " ORDER BY DESC (?value)"
+ " LIMIT 10";
```

Fonte: Autora

Tabela 9 - Retorno da pergunta de competência 12

Título	Valor
<i>Active Vest</i> - Conforto ao sentar	9000
<i>Agrotechlink</i> : agricultura e tecnologia conectadas.	9000
ALDA - Assistente de Letramento a Deficientes Auditivos	9000
<i>BeePay</i> - O melhor de dois mundos	9000
<i>Cheap2Ship</i> - Sistema inovador de cotação de fretes	9000
<i>Code Money</i>	9000
Diário da Alergia	9000
Financia Inovação - Plataforma de projetos de fomento	9000
<i>Hero HMB-FA</i> - Suplemento de resposta ativa	9000
<i>Ilergic</i> - Saúde além do rotulo	9000

Fonte: Autora

A consulta SPARQL da pergunta de competência número 13 “quais são as potenciais ideias de acordo com o índice de atualidade?” está ilustrada na Figura 30.

Figura 30 – Pergunta de competência 13

```
private static String queryString13 = "PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>"
+ "PREFIX owl: <http://www.w3.org/2002/07/owl#>"
+ "PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>"
+ "PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>"
+ "PREFIX ontology: <http://www.semanticweb.org/marina/ontologies/2018/1/untitled-ontology-369#>"
+ " SELECT ?title ?score"
+ " WHERE { ?y ontology:title ?title ."
+ " ?y ontology:scoreNovelty ?score }"
+ " ORDER BY DESC(?score)" + "LIMIT 10";
```

Fonte: Autora

Tabela 10 - Retorno da pergunta de competência 13

Título	Índice de atualidade
Joga+Perfil	1.00000000000000000000
MS <i>CondoHub</i> - Gestão de serviços para administradoras de condomínio	0.80220016873058795126
Plataforma integrada de gestão e atenção a saúde	0.78956350352234239278
<i>TechPain</i> : A revolução do tratamento da dor crônica na artrite	0.74305373006525928881
<i>Markets</i> - sistema para administração de produtos em grandes <i>Marketplaces</i>	0.74070868968474751813
Sistema INFRAVIA - Sistema de infraestrutura urbana	0.74059531627946667031
<i>FindMed</i> - A saúde em suas mãos	0.69411303631924470101
Glicosímetro não-invasivo Inteligente	0.68002954319299439154
<i>MediBridge</i> - Conectado você com sua saúde	0.49280732763004107635
<i>Fleet Manager</i> - IOT aplicado a veículos	0.42593088282895715066

Fonte: Autora

Por meio do resultado apresentado na Tabela 10 é possível verificar que a ontologia proposta responde quais são as ideias que possuem o maior índice de atualidade.

Para responder a pergunta de competência número 14 “quais são as ideias pertencentes a determinado grau de maturidade?” foi elaborada a consulta SPARQL representada na Figura 31.

Figura 31 – Pergunta de competência 14

```
private static String queryString14 = "PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>"
+ "PREFIX owl: <http://www.w3.org/2002/07/owl#>"
+ "PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>"
+ "PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>"
+ "PREFIX ontology: <http://www.semanticweb.org/marina/ontologies/2018/1/untitled-ontology-369#>"
+ " SELECT ?title ?degreeOfMaturity"
+ " WHERE { ?y ontology:title ?title ."
+ " ?y ontology:hasDegreeOfMaturity ?x."
+ " ?x ontology:description ?degreeOfMaturity."
+ " FILTER regex(?degreeOfMaturity, '^prototipo finalizado!')}"
+ "ORDER BY (?title)";
```

Fonte: Autora

Quadro 15 - Retorno da pergunta de competência 14

Título	Grau de maturidade
<i>Abellion</i> - A inteligência que sua plantação precisa	Protótipo finalizado
<i>Agrotechlink</i> : agricultura e tecnologia conectadas.	Protótipo finalizado
ALDA - Assistente de Letramento a Deficientes Auditivos	Protótipo finalizado
Apatuia- Frutas estruturadas com umidade	Protótipo finalizado
Aplicabilidade de <i>Golden-Max</i> em diferentes produtos da área da saúde	Protótipo finalizado
<i>Azimov Trade</i>	Protótipo finalizado
Biorrefinaria de Microalgas	Protótipo finalizado
<i>Chawork</i> - Sistema inteligente de recrutamento e seleção	Protótipo finalizado
<i>Cheap2Ship</i> - Sistema inovador de cotação de fretes	Protótipo finalizado
<i>Connect Smart</i> - Acionamento inteligente para coletores de pó	Protótipo finalizado
<i>E-licencie</i> : Garantindo a sustentabilidade das empresas	Protótipo finalizado
Ecoar eficiência energética e conforto térmico em ar	Protótipo finalizado
Filme de prova para prótese odontológica	Protótipo finalizado
<i>Findmed</i> - A saúde em suas mãos	Protótipo finalizado
<i>FIRE</i> - Gestão de clubes de caça e tiro	Protótipo finalizado
<i>Hero</i> HMB-FA - Suplemento de resposta ativa.	Protótipo finalizado

Impressora 3d sem filamento	Protótipo finalizado
<i>Kluc</i> - Obtenha insights dos seus dados através de simples perguntas	Protótipo finalizado
KOHMNECT - Sistema de controle e telemetria de baús	Protótipo finalizado
Liopet ração	Protótipo finalizado
Manejebem: Rede social para agricultores	Protótipo finalizado
Máquina de desrama de eucaliptos e pinus	Protótipo finalizado
Máquina para homogeneizar tintas	Protótipo finalizado
<i>NIDE</i> - Monitoramento de morros e encostas	Protótipo finalizado
Painéis Autoportantes em madeira	Protótipo finalizado
<i>Petrostats - Petrophysical Statistics</i>	Protótipo finalizado
<i>Posithividades</i> - App para pessoas que vivem e convivem com o HIV-AIDS.	Protótipo finalizado
SIGPEP (Sistema Integrado de Prontuário Eletrônico de Pacientes)	Protótipo finalizado
Sistema INFRAVIA - Sistema de infraestrutura urbana	Protótipo finalizado
<i>Smart Energy</i>	Protótipo finalizado
West Carijó - Dispositivo estimulador de aves de corte	Protótipo finalizado

Fonte: Autora

Por meio do resultado da consulta apresentado no Quadro 15 é possível verificar que a ontologia proposta responde quais são as ideias relacionadas a determinado grau de maturidade.

A consulta SPARQL da pergunta de competência número 15 “quais são as ideias em que a organização possui o número de colaboradores necessários para executá-las?” está ilustrada na Figura 32. O resultado é apresentado na Tabela 11.

Figura 32 – Pergunta de competência 15

```
private static String queryString15 = "PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>"
+ "PREFIX owl: <http://www.w3.org/2002/07/owl#>"
+ "PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>"
+ "PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>"
+ "PREFIX ontology: <http://www.semanticweb.org/marina/ontologies/2018/1/untitled-ontology-369#>"
+ " SELECT ?title ?numberEmployeesEstimated ?numberEmployees"
+ " WHERE { ?y ontology:title ?title ."
+ " ?y ontology:numberEmployeesEstimated ?numberEmployeesEstimated."
+ " ?y ontology:numberEmployees ?numberEmployees."
+ " FILTER (?numberEmployees >= ?numberEmployeesEstimated)}"
+ "ORDER BY (?title)";
```

Fonte: Autora

Tabela 11 - Retorno da pergunta de competência 15

Título	Número de colaboradores necessários	Número de colaboradores
APATUIA- Frutas estruturadas com umidade	4	6
<i>Crave Platform</i> - Indústria 4.0	3	8
Financia Inovação - Plataforma de projetos de fomento	1	3
<i>Hero</i> HMB-FA - Suplemento de resposta ativa.	6	8
Impressora 3d sem filamento	3	7
Máquina de desrama de eucaliptos e pinus	8	10
Painéis Autoportantes em madeira	6	7
<i>PillsController</i> - Dosador automático de medicamentos	5	6
Simple-sistema de planejamento e plano de negócios	2	4
Tour 360SC	1	2

Fonte: Autora

Por meio do resultado obtido é possível verificar que a ontologia proposta responde quais são as ideias que a organização possui colaboradores suficientes para executar as ideias.

Ressalta-se que a ontologia não se limita às questões de competência inicialmente formuladas para a sua especificação e construção. Considerando a abrangência dos conceitos e os índices produzidos na etapa de pré-processamento, diversas outras consultas são possíveis, dependendo do interesse dos especialistas.

4.3 CONSIDERAÇÕES FINAIS

Este capítulo apresentou a ontologia proposta para representação da base de conhecimento e da sua formalização. Também foi detalhado cada uma das etapas constituintes na metodologia *Ontology Development 101*. As classes principais da ontologia são: a classe *Idea*, que representa a ideia, a classe *Cluster*, responsável por representar o resultado do processo de agrupamento e a classe *AnalysisContext*, que representa as temáticas estratégicas relacionadas ao contexto organizacional. A representação da base de conhecimento no formato de uma ontologia proporciona vantagens na especificação formal e explícita de conceitos compartilhados.

5 DEFINIÇÃO DOS ÍNDICES

Este capítulo apresenta de forma detalhada os índices propostos pelo trabalho, com o intuito de contribuir na promoção e suporte ao processo de avaliação de ideias. Neste sentido, esta tese propõe a criação de quatro índices que irão compor ao final o índice de potencialidade de implementação da ideia.

5.1 ANÁLISE DE FONTES EXTERNAS

A análise de fontes externas permite auxiliar especialistas na tomada de decisão. Estas buscas contribuem na definição de possíveis investimentos ao realizar estudos em fontes da *web* ou relatórios.

Neste sentido, desenvolver o pensamento estratégico e o planejamento de ações de investimento em diferentes áreas alinhadas ao contexto da organização pode contribuir significativamente no crescimento da mesma. O estudo leva em consideração outras fontes de informações, sejam nacionais ou internacionais, extraíndo as áreas promissoras e representando-as na ontologia de domínio. Em uma etapa posterior irá servir como filtro ou elemento de suporte ao processo de decisão estratégica.

Além disso, as fontes externas auxiliam na etapa de pré-processamento, possibilitando a definição dos índices discutidos a seguir por meio de pesquisas realizadas em bases de patentes e bases de artigos científicos. Por exemplo, para identificar o grau de originalidade de uma ideia são realizadas buscas em bases de patentes e bases de artigos científicos. A busca em bancos de patente nacionais e internacionais pode ser motivada por vários interesses, tais como, verificar se um produto está protegido por patente no país, indicar o estado da técnica, analisar a originalidade das propostas, pesquisar oportunidades de exploração, buscar por alternativas técnicas para a solução de problemas, entre outros.

Na página do Escritório Americano de Marcas e Patentes (USPTO[®]), por exemplo, é permitido pesquisar pedidos de patente e também patentes concedidas nos Estados Unidos. A base possibilita a busca no texto completo das patentes concedidas a partir do ano de 1976 e o acesso às imagens dos documentos a partir de 1790. A base conta com mais de 90 milhões de documentos de patentes. De modo geral, cada registro de patente é composto por um índice (*id*) que representa um identificador único. Também possui um título, uma data referente à patente, um resumo da patente e um conjunto de palavras-chave (termos) que a caracterizam. Possui ainda um espaço destinado à alegação de relevância e justificativa da patente, bem como uma

descrição detalhada, podendo conter figuras, textos e quadros explicativos. Ressalta-se que, devido o volume de dados, coletou-se todas as patentes da base registradas nos três últimos anos.

A atualidade da ideia é definida a partir de buscas realizadas na base de dados *Semantic Scholar*[®]. Esta base foi desenvolvida em 2015 pelo *Allen Institute for Artificial Intelligence*, projetada para ser um serviço de busca inteligente em artigos de periódicos. Atualmente, a base conta com mais de 180 milhões de artigos. Cada registro na base da *Semantic Scholar*[®] representa um artigo que possui um identificador único. Atributos comuns a este tipo de conteúdo são: o título, o resumo, o ano e o local de publicação do artigo. Em cada registro existe também um atributo para armazenar o *url* do artigo que estará preenchido quando este for *open access*. Caso não esteja preenchido, existe um *link* para o artigo no site do periódico que o disponibiliza. Por fim, existem duas relações adicionais: uma para indicar os autores do artigo e outra para indicar as palavras-chave do artigo.

Todos os atributos descritos tanto para base de patentes quanto para a base *Semantic Scholar*[®] foram coletados por um processo automático. Foram salvos em um banco de dados para facilitar as consultas geradas no momento da criação dos índices que irão auxiliar os especialistas no processo de tomada de decisão.

5.2 ÍNDICE DE ADERÊNCIA AO CONTEXTO ORGANIZACIONAL

O primeiro índice é o de aderência ao contexto organizacional. Este índice é obtido analisando a ocorrência dos termos das ideias e dos termos descritos na classe *PotentialInvestmentArea*. Esta classe descreve as áreas potenciais de investimentos que a organização deseja investir e relaciona-se com a classe *AnalysisContext*, responsável por definir determinado contexto de análise.

Para estabelecer similaridades entre os vetores de ideias e o vetor de potenciais áreas de investimento, a literatura apresenta como possíveis equações: produto interno, cosseno, pseudo-cosseno, *dice*, *overlap* (Jones e Furnas, 1987). Uma das mais citadas e utilizadas é a equação do cosseno, explicitada na Seção 2.8.2.2, utilizada na tese.

Para normalização dos documentos optou-se por utilizar a medida estatística *tf-idf*, apresentada na Seção 2.8.2.3. A Tabela 12 apresenta o cálculo do referido índice.

Tabela 12 – Exemplo de cálculo do índice de aderência ao contexto organizacional

Vetor da Ideia 1							Vetor da Organização					
Termo	<i>tf</i>	<i>tf/max(tf)</i>	<i>df</i>	<i>idf</i>	<i>d</i>	<i>Peso</i>	<i>Peso</i> ^2	Termo	<i>Peso</i>	<i>Peso</i> ^2	Produto dos Pontos	
Termo 1	2	1	1	2,30	200	2,30	5,29	Termo 1	1	1	2,30	
Termo 2	0	0	0	0	200	0	0,00	Termo 2	1	1	0,00	
Termo 3	0	0	0	0	200	0	0,00	Termo 3	1	1	0,00	
Termo 4	2	1	1	2,30	200	2,30	5,29	Termo 4	1	1	2,30	
Termo 5	0	0	0	0	200	0	0,00	Termo 5	1	1	0,00	
Termo 6	1	0,5	1	2,30	200	1,15	1,32	Termo 6	1	1	1,15	
Termo 7	1	0,5	1	2,30	200	1,15	1,32	Termo 7	1	1	1,15	
Termo 8	0	0	0	0	200	0	0,00	Termo 8	1	1	0,00	
Peso total:							3,638248			2,828427	6,90309	
Cosseno:												0,670820

Fonte: Autora

Na Tabela 12 *tf* representa o número de vezes que o termo é mencionado no documento, *tf/max(tf)* é a normalização de *tf* pelo máximo *tf* do vetor, *df* é o número de documentos da coleção em que o termo é mencionado, *idf* é a frequência no documento invertida e *d* é o número de documentos (ideias) presentes no banco de dados. Após, é realizado o cálculo do *Peso* que corresponde a coluna *tf/max(tf)* multiplicado pela coluna *idf*. O Vetor da Ideia 1 é comparado ao Vetor da Organização. O peso dos termos do vetor da organização será sempre 1, pois corresponde ao vetor de referência. Após a normalização dos documentos, aplica-se o cálculo de similaridade, no qual utilizou-se o cosseno explicitada na Seção 2.8.2.2, equação 1. Para o cálculo, o peso dos termos de cada vetor foi elevado ao quadrado. O Peso total que corresponde a raiz quadrada da soma da coluna *Peso* ^2 do Vetor da Ideia 1 foi de 3,638248 e do Vetor da Organização foi de 2,828427. Posteriormente, realizou-se o produto dos pontos e o somatório da coluna foi 6,90309. Por fim, o valor do cosseno foi de 0,670820, foi dividido o somatório da coluna Produto dos Pontos pela multiplicação do Peso total da coluna *Peso* ^2 do Vetor da Ideia 1 e do Vetor da Organização.

5.3 ÍNDICE DE ATUALIDADE

O segundo índice gerado é o de atualidade. Ele é responsável por descrever o quanto a atual determinada ideia pode ser considerada. O índice de atualidade é determinado por consultas realizadas na base *Semantic Scholar*[®]. Com o intuito de clarificar o entendimento deste índice é apresentada a Tabela 13.

Tabela 13 – Exemplo de cálculo do índice de atualidade

Ideia 1										
Termo 1						Termo 2				
Ano	Docs	Tempo	Deflator	Peso		Ano	Docs	Tempo	Deflator	Peso
2018	10	1	0,900	9,000		2018	8	1	0,900	7,200
2017	2	2	0,810	1,620		2017	5	2	0,810	4,050
2016	13	3	0,729	9,477		2016	2	3	0,729	1,458
2015	7	4	0,656	4,593		2015	3	4	0,656	1,968
2014	6	5	0,590	3,543		2014	5	5	0,590	2,952
2013	5	6	0,531	2,657		2013	1	6	0,531	0,531
2012	9	7	0,478	4,305		2012	2	7	0,478	0,957
2011	1	8	0,430	0,430		2011	0	8	0,430	0,000
2010	2	9	0,387	0,775		2010	0	9	0,387	0,000
2009	4	10	0,349	1,395		2009	0	10	0,349	0,000
IA (Termo 1)				3,779		IA (Termo 2)				1,912
						IA (Ideia 1)				5,691

Fonte: Autora

Como pode ser observado na Tabela 13, o cálculo de atualidade se refere a *Ideia 1* que possui dois termos (*Termo 1* e *Termo 2*). Cada linha da tabela representa a distribuição temporal da frequência do termo ao longo dos anos, variando de 2009 a 2018.

A primeira coluna da tabela representa a distribuição temporal por ano (2009 a 2018). A segunda coluna *Docs* informa o número de documentos que mencionam o termo. A terceira coluna *Tempo* possui um número sequencial para atribuir peso maior aos anos mais atuais. A quarta coluna representa o *Deflator* que é determinado elevando-o à potência n , sendo n o valor da coluna *Tempo*. O conceito de deflator é derivado de algoritmos como o *PageRank*[®]. No caso do algoritmo *PageRank*[®] existe um fator de amortecimento indicando a perda de relevância de determinada página à medida que se utiliza da estrutura de *hiperlinks* para visitar páginas na *web*. O deflator é definido por meio da análise dos agrupamentos utilizando diferentes valores, visando identificar o mais adequado para o cenário estabelecido na tese. A quinta coluna corresponde ao *Peso*, que é o número de documentos que mencionam o termo multiplicado pelo deflator. O índice de atualidade de determinado termo da ideia é obtido por meio da média dos pesos de cada ano. Ao final, o índice de atualidade é estabelecido pela soma dos índices de cada termo da ideia. A equação 3 ilustra o cálculo.

$$IA_{ideia} = \frac{1}{n} \sum_{i=1}^n \frac{1}{k} \sum_{j=1}^k docs \times deflator^j \quad (3)$$

Na equação, n representa o total de termos da ideia, k representa o total de anos considerados no cálculo, $docs$, a quantidade de documentos que mencionam o termo em um ano específico e $deflator$, um índice de decréscimo temporal que deve ser inicialmente definido com um valor inferior a 1.

Neste exemplo, o valor atribuído ao deflator foi de 0,9. O índice de atualidade para o Termo 1 foi de 3,779 e para o Termo 2 foi de 1,912. O índice de atualidade da ideia, considerando os Termos 1 e 2 foi de 5,691. Como o índice é maior do que 1, torna-se necessária uma normalização entre 0 e 1, compatibilizando com os demais índices. A normalização ocorreu dividindo-se este valor pelo valor máximo do índice de atualidade presente na base de dados. Pressupõe-se a partir deste índice que, quanto mais atual determinada ideia for, maior será o interesse por ela.

5.4 ÍNDICE DE TECNOLOGIA

O terceiro índice é o de tecnologia. A classe *Technology* é responsável por descrever as tecnologias disponíveis na organização e as que a organização pretende investir/possuir. O cálculo deste índice é semelhante ao apresentado no índice de aderência ao contexto organizacional. Recupera-se a descrição das tecnologias que a organização possui e as tecnologias que a organização gostaria de investir a partir da base de conhecimento, estabelecendo-se um vetor com os termos e compara-o ao vetor da ideia. A normalização dos documentos foi através da medida estatística *tf-idf*, apresentada na Seção 2.8.2.3. além da similaridade determinada pelo cosseno, apresentado na Seção 2.8.2.2, equação 1.

Tabela 14 – Exemplo de cálculo do índice de tecnologia

Vetor da Ideia 1							Vetor da Organização					
Termo	tf	tf/max(tf)	df	idf	d	Peso	Peso ^2	Termo	Peso	Peso ^2	Produto dos Pontos	
Termo 1	3	1	1	2,30	200	2,30	5,29	Termo 1	1	1	2,30	
Termo 2	2	0,66666667	1	0	200	0	0,00	Termo 2	1	1	0,00	
Termo 3	0	0	0	0	200	0	0,00	Termo 3	1	1	0,00	
Termo 4	1	0,33333333	1	2,30	200	0,77	0,59	Termo 4	1	1	0,77	
Termo 5	0	0	0	0	200	0	0,00	Termo 5	1	1	0,00	
Termo 6	2	0,66666667	1	2,30	200	1,53	2,35	Termo 6	1	1	1,53	
Termo 7	2	0,66666667	1	2,30	200	1,53	2,35	Termo 7	1	1	1,53	
Termo 8	0	0	0	0	200	0	0,00	Termo 8	1	1	0,00	
Peso total:							3,254148			2,828427	6,13608	
Cosseno:												0,666667

Fonte: Autora

Na Tabela 14 tf representa o número de vezes que o termo é mencionado no documento, $tf/\max(tf)$ é a normalização de tf pelo máximo, df é o número de documentos da coleção em que o termo é mencionado, idf é a frequência no documento invertida e d é o número de documentos presentes no banco de dados. Após, é realizado o cálculo do peso que corresponde a coluna $tf/\max(tf)$ multiplicado pela coluna idf . O Vetor da Ideia 1 é comparado ao Vetor da Organização. No Vetor da Organização o peso dos termos será sempre 1, pois corresponde ao vetor de referência. Após a normalização dos documentos, aplica-se o cálculo de similaridade, no qual utilizou-se a equação do cosseno, explicitada na Seção 2.8.2.2, equação 1. Para o cálculo, o peso dos termos de cada vetor foi elevado ao quadrado. O Peso total que corresponde a raiz quadrada da soma da coluna $Peso^2$ do Vetor da Ideia 1 foi de 3,254148 e do Vetor da Organização foi de 2,828427. Posteriormente, realizou-se o produto dos pontos e o somatório da coluna foi 6,13608. Ao final, o valor do cosseno foi 0,666667, dividiu-se o somatório da coluna *Produto dos Pontos* pela multiplicação do Peso total da coluna $Peso^2$ do Vetor da Ideia 1 e do Vetor da Organização.

5.5 ÍNDICE DE ORIGINALIDADE

Por fim, define-se o índice de originalidade da ideia. Ele tem como objetivo localizar os termos da ideia que constam na base de patentes do Escritório Americano de Marcas e Patentes (USPTO[®])² e na *Semantic Scholar*[®] e determinar o grau de originalidade.

Ao realizar a pesquisa na base de patentes com os termos da ideia pode-se obter a quantidade de patentes que mencionam estes termos. Por exemplo, a ideia de desenvolvimento de um Produto X possui os seguintes termos: A, B e C. Ao pesquisar pelos termos individualmente, todos ocorrem na base de patentes. O mesmo ocorre quando a pesquisa é realizada combinando-se dois termos (A e B; A e C, e B e C). Contudo, ao se pesquisar a combinação entre todos os termos A, B e C, não é possível encontrá-la na base. Esta característica permitirá estabelecer o índice proposto nesta tese como um indicativo de originalidade. À medida que ocorra a evolução das combinações de termos e estas combinações não estejam presentes na base de patentes, o índice é incrementado.

Para compor o índice de originalidade é realizada uma pesquisa com a combinação dos termos nas bases de dados *Semantic Scholar*[®] e USPTO[®]. A busca é realizada combinando-se os termos-chave que definem uma ideia, sendo que os termos-chave em inglês são

² <https://www.uspto.gov/>

submetidos a uma busca nas duas bases. O índice de originalidade utilizado na tese é composto pela média dos índices obtidos nas duas bases de dados.

A Tabela 15 apresenta apenas os resultados identificados distribuídos por ano, comparando inicialmente o Termo 1 e o Termo 2, posteriormente o Termo 1 e o Termo 3, o Termo 2 e o Termo 3 e os três termos conjuntamente.

Tabela 15 – Exemplo de cálculo do índice de originalidade

Termo 1 x Termo 2									
Ano	Docs total	Docs	Tempo	Peso dos Termos (T)	Peso dos Elementos (E) (Vetor de Referência)	Deflador (D)	Produto dos pontos (T * E * D)	Potência de (T)	Potência de (E)
2018	2500	15	1	2,221849	1	0,900	2,000	4,936612	1
2017	2000	10	2	2,30103	1	0,810	1,864	5,294739	1
2016	1000	12	3	1,920819	1	0,729	1,400	3,689545	1
2015	2000	8	4	2,39794	1	0,656	1,573	5,750116	1
2014	3000	5	5	2,778151	1	0,590	1,640	7,718124	1
2013	1500	8	6	2,273001	1	0,531	1,208	5,166535	1
2012	1890	5	7	2,577492	1	0,478	1,233	6,643464	1
2011	1200	4	8	2,477121	1	0,430	1,066	6,13613	1
2010	1000	2	9	2,69897	1	0,387	1,046	7,284439	1
2009	400	3	10	2,124939	1	0,349	0,741	4,515365	1
Somatório							13,771	57,13507	10

IO (Termo 1 X Termo 2) determinado pela equação 4	0,423871
--	----------

Termo 1 x Termo 3									
Ano	Docs total	Docs	Tempo	Peso dos Termos (T)	Peso dos Elementos (E) (Vetor de Referência)	Deflador (D)	Produto dos pontos (T * E * D)	Potência de (T)	Potência de (E)
2018	2500	1	1	3,39794	1	0,900	3,058	11,546	1
2017	2000	1	2	3,30103	1	0,810	2,674	10,8968	1
2016	1000	1	3	3	1	0,729	2,187	9	1
2015	2000	2	4	3	1	0,656	1,968	9	1
2014	3000	1	5	3,477121	1	0,590	2,053	12,09037	1
2013	1500	10	6	2,176091	1	0,531	1,156	4,735373	1
2012	1890	2	7	2,975432	1	0,478	1,423	8,853194	1
2011	1200	1	8	3,079181	1	0,430	1,325	9,481357	1
2010	1000	1	9	3	1	0,387	1,162	9	1
2009	400	1	10	2,60206	1	0,349	0,907	6,770716	1
Somatório							17,915	91,37381	10

IO (Termo 1 X Termo 3) determinado pela equação 4	0,407336
--	----------

Termo 2 x Termo 3									
Ano	Docs total	Docs	Tempo	Peso dos Termos (T)	Peso dos Elementos (E) (Vetor de Referência)	Deflador (D)	Produto dos pontos (T * E * D)	Potência de (T)	Potência de (E)
2018	2500	0	1	0	1	0,900	0,000	0	1
2017	2000	0	2	0	1	0,810	0,000	0	1
2016	1000	0	3	0	1	0,729	0,000	0	1
2015	2000	0	4	0	1	0,656	0,000	0	1
2014	3000	1	5	3,477121	1	0,590	2,053	12,09037	1
2013	1500	1	6	3,176091	1	0,531	1,688	10,08756	1
2012	1890	1	7	3,276462	1	0,478	1,567	10,7352	1
2011	1200	1	8	3,079181	1	0,430	1,325	9,481357	1
2010	1000	1	9	3	1	0,387	1,162	9	1
2009	400	1	10	2,60206	1	0,349	0,907	6,770716	1
Somatório							8,703	58,1652	10

IO (Termo 2 X Termo 3) determinado pela equação 4	0,63913
--	---------

Termo 1 x Termo 2 x Termo 3									
Ano	Docs total	Docs	Tempo	Peso dos Termos (T)	Peso dos Elementos (E) (Vetor de Referência)	Deflador (D)	Produto dos pontos (T * E * D)	Potência de (T)	Potência de (E)
2018	2500	0	1	0	1	0,900	0,000	0	1
2017	2000	0	2	0	1	0,810	0,000	0	1
2016	1000	0	3	0	1	0,729	0,000	0	1
2015	2000	0	4	0	1	0,656	0,000	0	1
2014	3000	0	5	0	1	0,590	0,000	0	1
2013	1500	0	6	0	1	0,531	0,000	0	1
2012	1890	0	7	0	1	0,478	0,000	0	1
2011	1200	0	8	0	1	0,430	0,000	0	1
2010	1000	0	9	0	1	0,387	0,000	0	1
2009	400	1	10	2,60206	1	0,349	0,907	6,770716	1
Somatório							0,907	6,770716	10

IO (Termo 1 X Termo 2 X Termo 3) determinado pela equação 4	0,889738
--	----------

Índice de Originalidade da Ideia 1 determinado pela média	0,590019
--	----------

Fonte: Autora

A primeira coluna da tabela representa a distribuição temporal por ano (2009 a 2018). A segunda coluna *Docs Total*, a quantidade de documentos armazenados na base de dados no ano em específico. A terceira coluna *Docs*, o número de documentos que mencionam o termo. A quarta coluna *Tempo* representa um número sequencial para atribuir peso maior aos anos mais atuais. A quinta coluna *Peso dos Termos (T)* representa a frequência invertida, o valor será calculado através de $\log(\text{Docs Total} / \text{Docs})$. A sexta coluna representa o *Peso dos Elementos (E) (Vetor de Referência)*. A sétima coluna representa o *Deflador (D)*, que é determinado elevando-se o deflador à potência n , sendo n o valor da coluna *Tempo*. Este conceito também foi apresentado no índice de atualidade. O deflador será definido através da análise dos agrupamentos utilizando diferentes valores. Visa identificar o mais adequado para o cenário

estabelecido na tese. A oitava coluna corresponde ao *Produto dos Pontos* ($T * E * D$), multiplica-se o *Peso dos Termos* (T) pelo *Peso dos Elementos* (E) (*Vetor de Referência*) e pelo deflator (D). A multiplicação pelo *Peso dos Elementos* (E) é importante, pois representa o vetor de referência, visto como um vetor em que todos os anos possuem a mesma importância, ou seja, cada ano possui peso 1. Desta forma, não pode ser considerado original, uma vez que qualquer que seja a combinação de termos, esses já constam na base de pesquisa para todos os anos. A nona coluna Potência de (T) é o *Peso dos Termos* (T) elevado a potência 2. A última coluna corresponde à Potência de (E), ou seja, *Peso dos Elementos* (E) (*Vetor de Referência*) elevado a potência 2. A equação 4 ilustra o cálculo utilizado para determinar o índice de originalidade propondo uma versão modificada do cosseno. A equação parte do pressuposto que, quanto maior a similaridade entre os vetores envolvidos (combinação de termos e referência), menor é a originalidade da combinação de termos pesquisados que compõem determinada ideia.

$$\cos\theta = 1 - \frac{\sum_{i=1}^n (w1_i * w2_i * deflator^{tempo})}{\sqrt{\sum_{k=1}^n (w1_k)^2} * \sqrt{\sum_{j=1}^n (w2_j)^2}} \quad (4)$$

Nesta equação, n representa o total de termos da ideia; k o total de anos considerados no cálculo; $w1_i$ corresponde à frequência invertida em um ano específico; o valor $w2_i$ e $w2_j$ será sempre 1, pois corresponde ao vetor de referência, ou seja, o vetor dito perfeito; *deflator*, um índice de decréscimo temporal que deve ser inicialmente definido com valor inferior a 1. Por fim, o peso será decrescido de 1. O decréscimo ocorre com o objetivo de tornar o valor mais próximo a 1, uma vez que, quanto menor for a similaridade do vetor dos termos em relação ao vetor de referência, maior será a sua originalidade. Após o cálculo, o índice de originalidade é estabelecido pela média dos índices de cada combinação de termo da ideia. No exemplo apresentado, o índice de originalidade da Ideia 1 foi de 0,59002.

5.6 ÍNDICE DE POTENCIALIDADE DE IMPLEMENTAÇÃO DA IDEIA

No final, o índice de potencialidade (IP) da ideia será a média dos quatro índices apresentados, conforme equação (5).

$$IP = \frac{1}{4} (IA + IAD + IO + IT) \quad (5)$$

Este índice possibilita criar um *ranking* pela ordem de classificação das ideias, estabelecido pelo modelo.

5.7 CONSIDERAÇÕES FINAIS

Este capítulo apresentou os índices propostos para esta tese, buscando fornecer indícios do potencial de implementação de uma ideia. Detalharam-se também as consultas realizadas em fontes externas utilizadas na definição dos índices. Os índices gerados são cruciais para determinar o potencial de implementação das ideias, sendo eles: o índice de atualidade baseado em artigos científicos da base *Semantic Scholar*[®]; o índice de aderência ao contexto organizacional, que utiliza em sua composição as informações advindas da base de conhecimento; o índice de originalidade com base em dados coletados da base norte americana de patentes USPTO[®] e da *Semantic Scholar*[®] e o índice de tecnologia baseado nas informações presentes na base de conhecimento. Por fim, a média da soma dos índices compõe o índice de potencialidade da ideia.

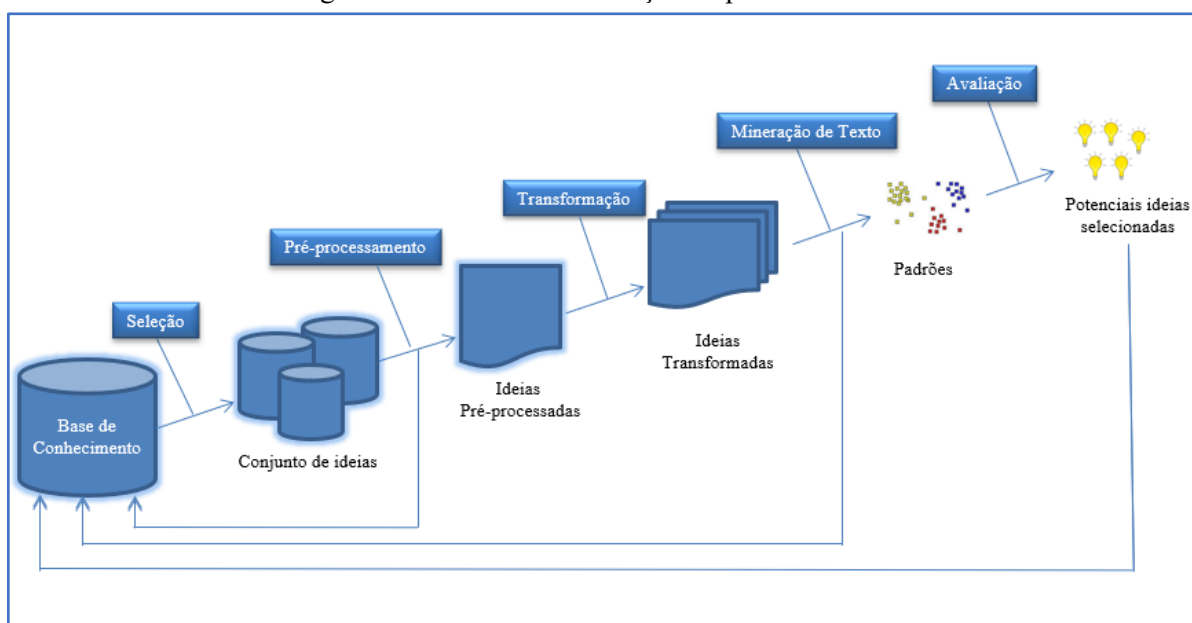
6 MODELO PROPOSTO

Este capítulo apresenta de maneira detalhada o modelo proposto nesta tese. Todas as fases do modelo são detalhadas a seguir, sendo que esta pesquisa projeta e desenvolve a estrutura e a arquitetura em si, como suporte ao processo de avaliação de ideias.

6.1 APRESENTAÇÃO DO MODELO

Quando uma ideia é submetida ao processo de avaliação, a primeira tarefa a ser executada é o reconhecimento das informações, visando criar subsídios para identificar o domínio. A representação e o armazenamento do conhecimento de domínio são realizados através de uma ontologia. O modelo proposto consiste em 5 etapas fundamentais: a população da base de conhecimento e seleção de ideias, o pré-processamento, a transformação, a aplicação dos métodos e técnicas de mineração de texto e a avaliação das ideias. Apesar de seguir a abordagem tradicional de mineração de dados e texto, existem diferenças fundamentais agregadas, visando gerar valor quando aplicadas ao contexto de gestão de ideias. A Figura 33 apresenta o modelo proposto.

Figura 33 – Modelo de avaliação de potenciais ideias



Fonte: Autora

O modelo inicia pelo provimento de **ideias**, caracterizadas por textos curtos e com poucas informações que as detalham. Nesta etapa de **população da base de conhecimento e de seleção de ideias** preenche-se a ontologia que representa a base de conhecimento com as

ideias e com as informações sobre o contexto organizacional. Após, realiza-se a **seleção** do conjunto de ideias que fará parte da análise. Por exemplo, selecionam-se as ideias do último edital, as que pertencem à determinada categoria, entre outras.

Na etapa seguinte, é realizado o **pré-processamento das ideias**, que possui como objetivo prover a correta formatação e a representação do formato textual das ideias. Esta etapa é onerosa, pois envolve a aplicação de algoritmos que consomem boa parte do tempo do processo de extração de conhecimento. Nesta etapa também ocorre a interação com a base de conhecimento, visando compreender o contexto organizacional.

Posteriormente, acontece a etapa de **transformação das ideias**. Ela consiste na adequação dos dados para serem utilizados na etapa de mineração de texto. Estas transformações podem variar conforme o domínio da aplicação. Em seguida, aplicam-se **métodos e técnicas de mineração de dados e texto**, objetivando descobrir padrões úteis que possam contribuir no processo de avaliação e seleção de potenciais ideias. A etapa de mineração propriamente dita é responsável pelo desenvolvimento de cálculos e algoritmos, tem como intuito a descoberta de padrões e a extração de conhecimento.

Finalmente, realiza-se a **avaliação** das ideias por especialistas da organização, visa identificar as potenciais ideias alinhadas às temáticas estratégicas relacionadas ao contexto organizacional. Como resultado, são identificadas e selecionadas as ideias passíveis de implementação, e esta informação retroalimenta a base de conhecimento.

Com o intuito de clarificar o entendimento do modelo proposto, as etapas de **População da Base de Conhecimento e Seleção de Ideias, Pré-processamento, Transformação, Mineração de Texto e Avaliação de Ideias** são detalhadas.

6.1.1 População da Base de Conhecimento e Seleção de Ideias

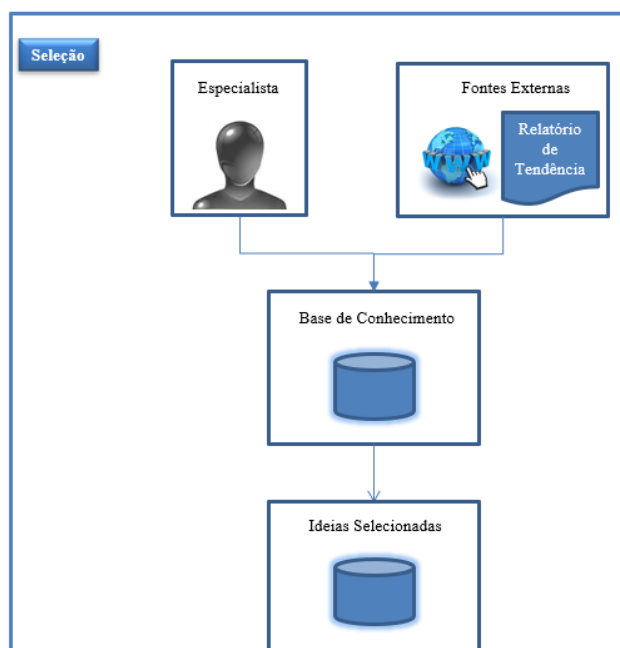
O modelo inicia com a população da base de conhecimento com as ideias e as informações referentes ao contexto organizacional. A base mantém as informações que guiam determinado processo de avaliação das ideias.

Na ontologia, o especialista irá preencher informações referentes às temáticas estratégicas relacionadas ao contexto organizacional, como: o valor estimado de investimento, a data de realização da análise, as áreas potenciais de investimento que a organização gostaria de investir e a descrição das tecnologias que a organização possui ou deseja investir. As áreas de investimentos potenciais também poderão ser preenchidas através de relatórios de tendências

e informações especializadas a partir da *web*, consideradas fontes externas. Estas informações armazenadas na base de conhecimento serão utilizadas na etapa de pré-processamento.

A etapa de seleção é responsável por selecionar as ideias de acordo com os filtros estabelecidos pelos especialistas da organização. A Figura 34 ilustra os passos desta etapa.

Figura 34 – População da base de conhecimento e seleção de ideias



Fonte: Autora

Como pode ser observado na Figura 34, o objetivo é selecionar as ideias correspondentes a determinados filtros estabelecidos de acordo com critérios selecionados por especialistas. Estes critérios podem variar como: ideias do último edital, área de investimento, categoria, enfim, reduzindo o conjunto inicial de ideias para um conjunto menor direcionado a um contexto de análise.

6.1.2 Pré-processamento

Esta etapa objetiva a correta formatação e representação do formato textual das ideias para que sejam transformadas na etapa seguinte. Além disso, através da análise de fontes externas à organização e considerando as temáticas relacionadas ao contexto organizacional disponível na base de conhecimento, ocorre a criação de índices: o índice de atualidade, o índice de aderência ao contexto organizacional, o índice de originalidade e o índice de tecnologia, que são agregados à determinada ideia e indica o índice de potencialidade de implementação da mesma.

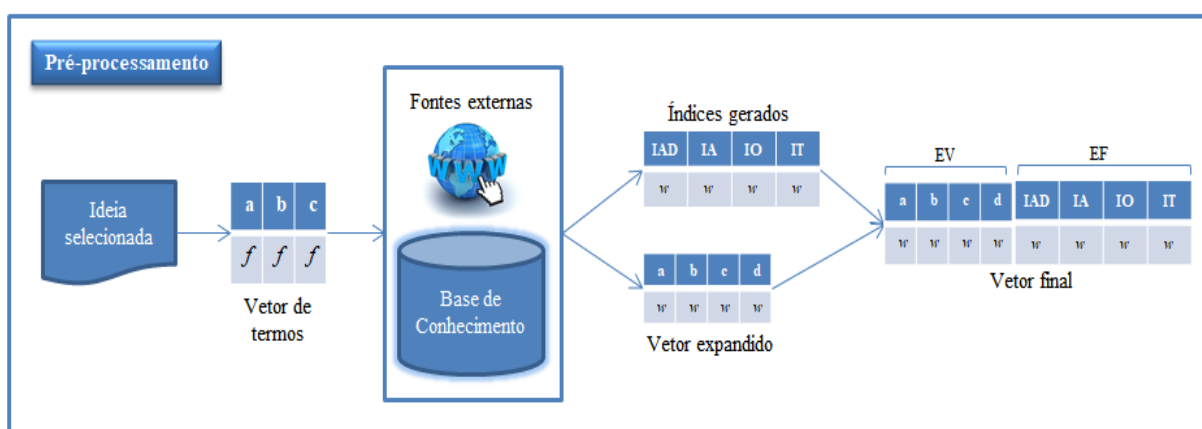
Inicialmente, cada ideia é composta por um conjunto de termos (palavras-chave), representada através de um vetor. Os termos podem ser identificados/extraídos utilizando-se métodos automáticos ou semiautomáticos com processamento de linguagem natural, ou mesmo por meio da utilização de dicionários, tesouros ou ontologias. Nesta tese, os termos-chave foram definidos manualmente com a leitura das ideias coletadas.

Posteriormente, cria-se um vetor expandido que leva em consideração as temáticas relacionadas ao contexto organizacional armazenado na ontologia, as ideias e as informações advindas de fontes de informação externas. As fontes externas podem ser representadas por meio de bases de dados públicos disponíveis, tais como: bases de patentes e artigos de produção científica e tecnológica. Exemplos dessas bases são: a base norte americana de patentes (USPTO®) e a base de artigos científicos *Semantic Scholar*® utilizadas nesta pesquisa. Essas bases são coletadas e indexadas com o objetivo de permitir sua utilização na etapa de pré-processamento. Maior detalhamento das consultas em fontes externas é apresentado na Seção 5.1.

A consulta utiliza o vetor com os termos-chave de cada ideia e realiza uma busca nas bases de dados presentes e indexadas no banco de dados. Além disso, a informação do ano de publicação é de suma importância, possibilitando a criação de distribuições temporais dos termos de modo a auxiliar na composição dos índices de atualidade e originalidade.

Nesta etapa também ocorre a análise das ideias e do contexto organizacional, gerando um índice de aderência da ideia ao contexto organizacional e um índice de tecnologia (considerando as tecnologias já disponíveis na organização ou de interesse dela, que estão declarados na ontologia). A Figura 35 apresenta o detalhamento da etapa de pré-processamento.

Figura 35 – Pré-processamento de ideias



Fonte: Autora

De acordo com a Figura 35, o fluxo de pré-processamento se inicia com a ideia selecionada. Após, ocorre a geração do vetor inicial com os termos da ideia. Em seguida se realiza a consulta nas bases de fontes externas e, após isto, a sumarização das informações retornadas com foco na temporalidade, bem como a recuperação de termos relacionados aos termos originais do vetor obtidos por coocorrência. Posteriormente, são criados os índices de atualidade, de aderência ao contexto organizacional, de originalidade e de tecnologia. Por fim, ocorre a composição final do vetor com os termos originais, novos termos advindos das fontes externas e índices. Para a composição dos índices é necessário o retorno à base de conhecimento para análise do contexto organizacional.

O vetor final resultante representando uma ideia é determinado através da composição de duas variáveis, sendo constituído por elementos variáveis (EV) e elementos fixos (EF). Os elementos variáveis são determinados pelos termos que representam a ideia, bem como a relevância de cada termo. A relevância pode ser determinada pela frequência em que o termo aparece na ideia ou nos artigos relacionados aos termos que a compõem. Ao final, os elementos variáveis podem ser normalizados utilizando abordagens estatísticas.

Já os elementos fixos são representados por quatro índices, sendo: o índice de aderência ao contexto organizacional (IAD), o índice de atualidade (IA), o índice de originalidade (IO) e o índice de tecnologia (IT). O Capítulo 5 apresenta detalhadamente estes índices. A composição de todos os vetores é representada por uma matriz exposta na Tabela 16 que servirá de entrada para a etapa seguinte.

Tabela 16 - Vetor final da etapa de pré-processamento

Id	T₁	T₂	T_N	IAD	IA	IO	IT
1	$f_{1,t1}$	$f_{1,t2}$	$f_{1,tn}$	$w_{1,iad}$	$w_{1,ia}$	$w_{1,io}$	$w_{1,it}$
2	$f_{2,t1}$	$f_{2,t2}$	$f_{2,tn}$	$w_{2,iad}$	$w_{2,ia}$	$w_{2,io}$	$w_{2,it}$
3	$f_{3,t1}$	$f_{3,t2}$	$f_{3,tn}$	$w_{3,iad}$	$w_{3,ia}$	$w_{3,io}$	$w_{3,it}$

Fonte: Autora

Cada linha da tabela representa a ideia. A primeira coluna representa o número de identificação do termo. Na segunda, terceira e quarta coluna são estabelecidas a frequência dos termos. As demais colunas representam os índices gerados.

6.1.3 Transformação

A etapa de transformação das ideias é responsável por adequar os dados para serem utilizados em algoritmos de mineração de texto. A partir dos vetores de ideias obtidos na etapa

de pré-processamento compõe-se uma matriz permitindo a realização da transformação por meio de técnicas, tais como, *tf-idf*, *SVD*, *LSI*.

Para apresentar em linhas gerais a etapa de transformação é realizado um exemplo considerando uma matriz formada por um conjunto de ideias e os termos e índices relacionados a elas. A Tabela 17 ilustra as ideias com seus respectivos termos e índices.

Tabela 17 - Matriz obtida na etapa de pré-processamento

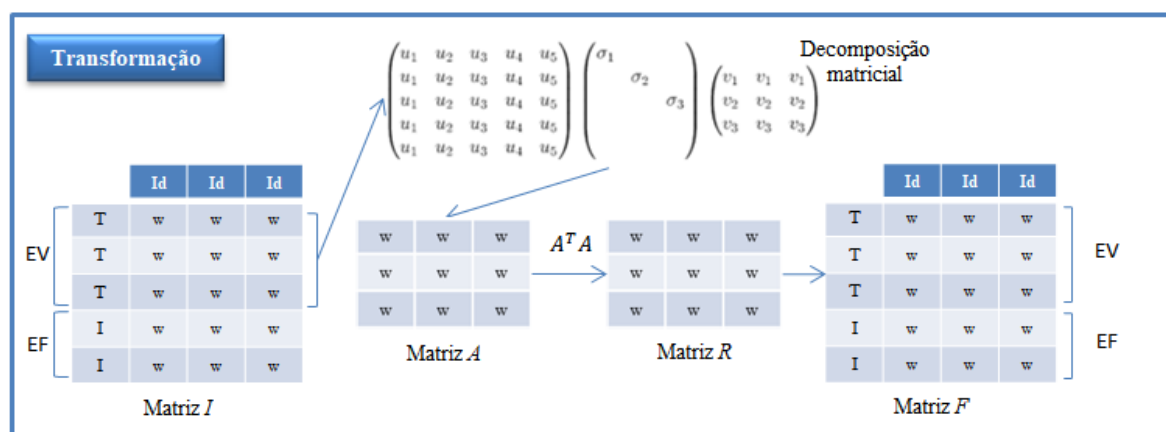
Id	Ideia ₁	Ideia ₂	Ideia ₃	Ideia ₄	Ideia ₅
Produção	5	1	0	0	0
Agrícola	2	0	0	0	0
Plataforma	2	3	0	0	0
Resíduo orgânico	0	10	0	0	0
Adubo	0	7	0	0	0
Dor	0	0	5	0	0
Agricultura	3	2	0	5	0
Tecnologia	7	1	0	3	5
Automação de sistema de cultivo	5	0	0	2	0
Jogo	0	0	0	0	7
Libras	0	0	0	0	3
IAD	W _{1,IAD}	W _{2,IAD}	W _{3,IAD}	W _{4,IAD}	W _{5,IAD}
IA	W _{1,IA}	W _{2,IA}	W _{3,IA}	W _{4,IA}	W _{5,IA}
IO	W _{1,IO}	W _{2,IO}	W _{3,IO}	W _{4,IO}	W _{5,IO}
IT	W _{1,IT}	W _{2,IT}	W _{3,IT}	W _{4,IT}	W _{5,IT}

Fonte: Autora

A partir da matriz é realizada a decomposição matricial por meio da técnica *Singular Value Decomposition* (SVD), que analisa de maneira iterativa as coocorrências entre os termos dispostos na matriz, visando estabelecer relacionamentos latentes. Desta forma, objetiva-se aumentar a semântica de representação de determinado espaço de análise. A matriz possui também os índices referentes a cada ideia. Contudo, estes índices não fazem parte do processo de decomposição matricial.

A Figura 36 representa os passos envolvidos na etapa de transformação.

Figura 36 – Transformação de ideias



Fonte: Autora

Após a recomposição da matriz original decomposta através do SVD, conforme descrito na Seção 2.8.1.1, tem-se uma matriz dita completa. Analisando individualmente as dimensões do vetor (representado pelas colunas da matriz, em que cada coluna pode ser entendida como uma ideia), todas são preenchidas com valores que contribuem (valores acima de zero), não contribuem (valores abaixo de zero) ou não promovem qualquer diferença (valor zero) no aumento da semântica do vetor. Ou seja, determina qual é a influência da dimensão (vista como um termo) considerando as relações diretas e indiretas com outros vetores (ideias).

Ao utilizar a matriz resultante, torna-se possível gerar uma matriz que represente os pares de ideias através de $R = A^T A$, permitindo que as ideias sejam comparáveis novamente entre si. Conforme apresentado na Figura 36, após a multiplicação ocorre a integração da matriz resultante R com a parte da matriz original I (parte EF) que se refere aos índices. Ao final da integração a matriz F pode ser utilizada na tarefa de agrupamento.

6.1.4 Mineração de Texto

A etapa de mineração de texto é responsável pela descoberta de padrões e extração de conhecimento. A Mineração de Texto, também conhecida como Análise de Texto, Análise de Texto Inteligente ou Descoberta de Conhecimento de Texto (KDT), refere-se ao processo automático de extração de informações e conhecimentos novos e não triviais a partir de dados textuais não estruturados (GUPTA; LEHAL, 2009).

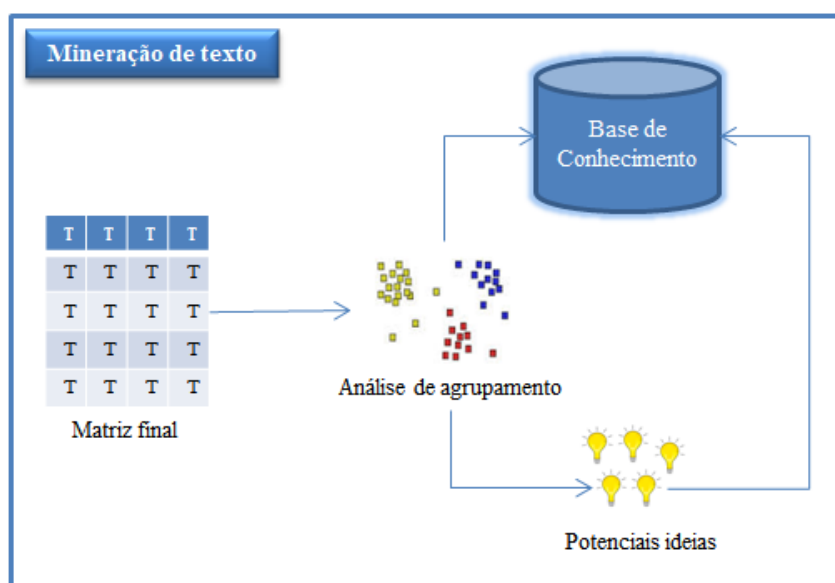
Os avanços tecnológicos levaram ao rápido desenvolvimento do campo de Mineração de Texto, constituindo-o como um campo interdisciplinar relacionado ao aprendizado de máquina, mineração de dados, recuperação de informação, estatística, raciocínio e linguística computacional. A mineração de texto geralmente emprega técnicas de mineração de dados em um conjunto de documentos textuais para identificar conhecimento oculto, e os resultados podem ser não apenas a situação geral de um único documento, mas também a classificação de conjuntos de documentos ou o agrupamento dos mesmos (GUPTA; LEHAL, 2009).

A análise de agrupamento é utilizada para reunir um conjunto de dados no contexto do trabalho representadas por ideias, em grupos. O objetivo é agrupar ideias similares que compartilham um conjunto de propriedades comuns em grupos. Com o intuito de atingir o objetivo, primeiramente o VSM foi utilizado em conjunto com o algoritmo de agrupamento. VSM é amplamente utilizado, tratando a proximidade semântica como proximidade espacial, iniciando com uma lista de termos gerada a partir de determinado documento textual e

transformando-o em um vetor de espaço n -dimensional, no qual n representa os termos que compõe o vetor (RUSSEL; NORVIG, 1995).

Os vetores são formados considerando principalmente o título e o conteúdo de cada ideia. Outros conteúdos textuais também podem ser utilizados. Os termos que compõem o vetor da ideia possuem um identificador e um peso que representa a importância do termo, formando a matriz termo-documento. O peso é definido com a aplicação da medida estatística *tf-idf*. A Figura 37 apresenta as etapas envolvidas na mineração de texto.

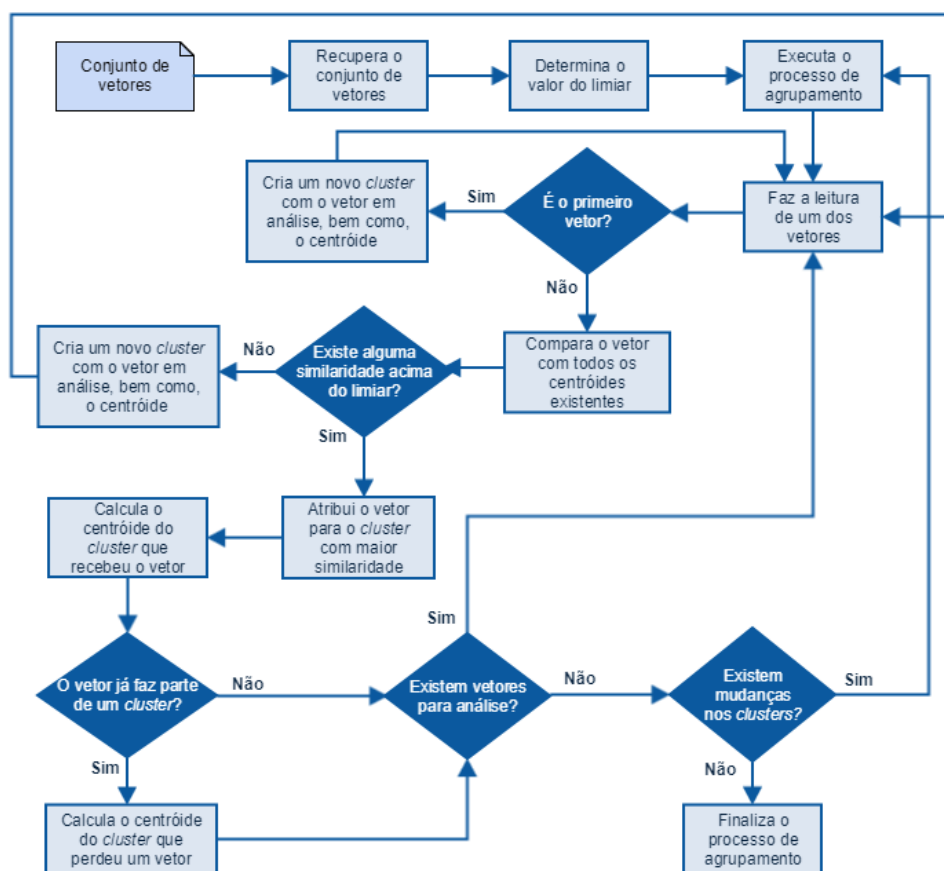
Figura 37 – Mineração de texto



Fonte: Autora

Com o objetivo de aplicar a tarefa de agrupamento, uma versão modificada do algoritmo *k-means* utilizando o conceito de similaridade vetorial foi utilizada. O algoritmo de agrupamento é baseado no trabalho de Alvarez (2018), conforme a Figura 38. Contudo, existem diferenças que serão discutidas a seguir, relacionadas à etapa final de mineração.

Figura 38 – Fluxograma do algoritmo de agrupamento



Fonte: Alvarez (2018)

O algoritmo inicia com a recuperação do conjunto de vetores de ideias previamente pré-processados e transformados. A partir disso, é determinado o limiar (*threshold*) indicando o nível de similaridade entre os vetores de determinado agrupamento. Após a leitura de um vetor de ideia, o agrupamento inicial é determinado pelo primeiro vetor a ser analisado, definido também como o centróide.

Para os demais vetores, é comparado à similaridade do vetor de acordo com o limiar definido (*threshold*) com todos os centróides identificados, e atribui-se o vetor para o agrupamento com maior similaridade. Recalcula-se o centróide do agrupamento que recebeu o novo vetor. Para definir a similaridade das ideias durante a etapa de agrupamento, utiliza-se a medida do cosseno (apresentada na Seção 2.8.2.2).

Caso o vetor ainda não esteja associado a um dos agrupamentos, define-se um novo agrupamento com o vetor corrente, ou seja, o vetor que está sendo analisado. Contudo, ao definir um vetor como sendo um novo agrupamento, é necessário verificar se o vetor de análise fazia parte de outro agrupamento. Se a condição for verdadeira, calcula-se novamente o centróide do agrupamento que perdeu o vetor. Do contrário, é verificado se ainda existem

vetores para análise. Se existirem, efetua-se a leitura dos novos vetores para análise e o fluxo se repete. Assim que todos os vetores forem lidos e analisados, o resultado da etapa de agrupamento é armazenado na base de conhecimento, ou seja, instâncias da classe *Cluster* são criadas e ocorre a associação com as ideias que pertencem ao agrupamento.

6.1.5 Avaliação das Ideias

Nesta etapa é realizada a avaliação dos agrupamentos e dos índices obtidos durante a aplicação do modelo pelos especialistas da organização.

Para tal, a tese propõe um modelo que é caracterizado pelos passos envolvidos no enriquecimento das ideias com novos atributos/índices, bem como a utilização da tarefa de agrupamento para sugerir grupos de potenciais ideias, levando em consideração as temáticas estratégicas relacionadas ao contexto organizacional.

A avaliação das ideias é realizada por especialistas da organização, possibilitando identificar o potencial de implementação de cada ideia e do próprio agrupamento formado. Os índices calculados durante a avaliação de potencialidade da ideia são: índice de aderência ao contexto organizacional; índice de atualidade; índice de originalidade; índice de tecnologia. Todos contribuem significativamente nesta fase de avaliação ao compor o índice de potencialidade da ideia. O índice de potencialidade é utilizado na apresentação dos agrupamentos. Ao analisar a formação dos agrupamentos mais relevantes, o especialista poderá analisar as ideias com maior potencial de implementação e a similaridade do grupo.

Neste sentido, os dados coletados e armazenados na base de conhecimento contribuem na visualização das ideias alinhadas às temáticas estratégicas relacionadas ao contexto organizacional. Inicialmente, o tomador de decisão, visualizando o conjunto de ideias, poderá não saber em qual ideia investir. Contudo, ao visualizar o agrupamento de ideias similares, poderá estudar a semântica do grupo e extrair maiores informações sobre as ideias, de tal maneira que se possa explicitar um entendimento a respeito do agrupamento. Assim, ao identificar ideias semelhantes, a organização poderá investir em grupos de ideias.

Deste modo, o modelo proposto inclui etapas adicionais para que os objetivos do trabalho sejam alcançados, entre elas, a explicitação do conhecimento de domínio que fornece os subsídios para a etapa de avaliação. Possibilita assim, explicitar visualmente grupos de ideias, assim como os relacionamentos presentes entre as ideias coletadas e seu potencial de implementação, por meio de técnicas de visualização, entre elas, *treemap*, mapas de tópicos e análise de redes sociais. Com o modelo proposto, o objetivo é contribuir no processo de análise

e avaliação de ideias, auxiliando na tomada de decisão sobre quais ideias serão ou não implementadas.

Como resultado, são identificadas e selecionadas as ideias passíveis de implementação por meio dos grupos de ideias, sendo que esta informação retroalimenta a base de conhecimento.

6.2 CONSIDERAÇÕES FINAIS

Este capítulo apresentou o modelo proposto, detalhou cada uma das etapas constituintes. Os pontos centrais do modelo são compostos pela população da base de conhecimento e seleção das ideias, pré-processamento, transformação, aplicação de técnicas de mineração de dados e texto e a avaliação das ideias. Na etapa de avaliação ocorre a explicitação do conhecimento em que recursos de visualização de informação promovem subsídios para que os objetivos do modelo proposto sejam atingidos. Nesse sentido, a integração desses métodos e técnicas no modelo proposto visa fornecer um ferramental que promova suporte à avaliação e identificação de potenciais ideias para a implementação.

7 ANÁLISE E APRESENTAÇÃO DOS RESULTADOS

Este capítulo tem como objetivo apresentar o processo de operacionalização (instanciação), demonstração e avaliação, bem como a discussão dos resultados obtidos pelo modelo proposto.

7.1 APRESENTAÇÃO DO CENÁRIO

Os cenários escolhidos para aplicação do modelo foram a base do programa de incentivo ao empreendedorismo Sinapse da Inovação[®] e o programa lançado pelo Ministério da Educação #MosquitoNão, conforme foi descrito na Seção 3.3.4.

O primeiro cenário de análise é composto por ideias coletadas do programa Sinapse da Inovação[®]. Para este cenário, foram coletadas 200 ideias referentes à Fase II do programa, sendo que apenas 100 ideias foram aprovadas e receberam financiamento.

Dentre as características do cenário de estudo, as ideias estão divididas por categorias. Ao total são nove categorias: Nanotecnologia, Gestão, Tecnologia Social, Biotecnologia, Tecnologia da Informação e Comunicação, Química e Materiais, Eletrônica, *Design* e Mecatrônica.

Ao analisar as ideias aprovadas no programa Sinapse da Inovação[®], verificou-se a distribuição destas ideias perante as categorias, conforme a Tabela 18. Observa-se que dentre as ideias aprovadas na fase final do programa Sinapse da Inovação[®], a categoria Tecnologia da Informação e Comunicação possui 28 das ideias aprovadas, seguida por Eletrônica, com 18, e Tecnologia Social, com 17 ideias aprovadas.

Tabela 18 - Distribuição das ideias coletadas por categoria

Categoria	Quantidade de ideias aprovadas
Tecnologia da Informação e Comunicação	28
Tecnologia Social	17
Eletrônica	18
Gestão	14
Biotecnologia	08
Mecatrônica	06
Nanotecnologia	04
Química e Materiais	03
Design	02

Fonte: Autora

Cada ideia proposta possui determinado Grau de Maturidade. Ela pode estar no nível inicial de Ideia, Protótipo conceitual, Protótipo testado, Protótipo finalizado e Produto sendo comercializado. A Tabela 19 apresenta a distribuição por grau de maturidade das ideias aprovadas.

Tabela 19 - Distribuição das ideias coletadas por grau de maturidade

Categoria	Quantidade de ideias aprovadas
Protótipo testado	49
Protótipo finalizado	19
Protótipo conceitual	16
Produto sendo comercializado	13
Ideia	3

Fonte: Autora

Das 100 ideias aprovadas na etapa final do programa, as que possuíam o protótipo testado obtiveram 49 aprovações. As que possuíam o protótipo finalizado atingiram 19 aprovações. No nível conceitual foram 16 ideias aprovadas e produto sendo comercializado alcançaram 13. As que se encontravam apenas no nível de ideia obtiveram 3 aprovações.

O segundo cenário de análise é composto por 93 ideias e corresponde a uma coletânea de ideias denominada #MosquitoNão. As ideias foram coletadas a por meio de uma consulta pública lançada pelo Ministério da Educação para identificar ações que poderiam ser realizadas por estudantes, professores, gestores públicos e sociedade, a fim de combater o mosquito *Aedes Aegypti* e enfrentar a microcefalia.

No total, 1.662 propostas de ideias foram sugeridas e receberam curtidas e comentários. Essas ações, que podem ser implementadas por estudantes, professores, técnicos e gestores públicos ou demais membros da sociedade civil, foram agrupadas em nove temas: “Utilização de Tecnologias”, “Prêmios, Bonificações e Incentivos”, “Capacitação”, “Conscientização, Orientações e Mutirão de Combate”, “Vistoria, Alertas e Punições”, “Armadilhas e Métodos de Prevenção”, “Pesquisa”, “Criação e Evolução” e “Parcerias”.

Neste cenário não será considerado o contexto organizacional, pois o intuito é analisar e demonstrar a generalidade do modelo proposto.

7.1.1 População da Base de Conhecimento e Seleção das Ideias

O procedimento de coleta das ideias ocorreu manualmente gerando um documento para cada ideia e tem como objetivo popular em um momento posterior a ontologia de domínio que representa a base de conhecimento. No primeiro cenário de estudo, as temáticas estratégicas

relacionadas ao contexto organizacional foram descritas através da leitura do edital de fomento e inseridas na ontologia. Temáticas estas relacionadas às áreas potenciais de investimento e descrição das tecnologias que a organização pretende investir.

Após, na seleção do conjunto de ideias que fará parte da análise, foram selecionadas todas as ideias para os dois cenários.

7.1.2 Pré-processamento e Transformação

As ideias selecionadas e as temáticas estratégicas relacionadas ao contexto organizacional, ambas armazenadas na base de conhecimento, são pré-processadas objetivando a correta formatação e a representação do formato textual das ideias e das informações de consulta.

Na etapa de pré-processamento são gerados os seguintes índices: o índice de atualidade, o índice de aderência ao contexto organizacional, o índice de originalidade e o índice de tecnologia.

7.1.3 Transformação

A etapa de transformação possui o intuito de adequar os dados na etapa de mineração de texto, por intermédio da decomposição matricial.

7.1.4 Mineração de Texto

Nesta etapa, por meio da análise de agrupamento as ideias semelhantes são reunidas em grupos, e a partir dos índices das ideias que compõem cada grupo será gerado um índice geral. Isto permite produzir um *ranking* dos grupos de ideias. Os índices gerados são importantes para esta etapa: o índice de atualidade gerado com base em dados de bases de artigos científicos, no caso a *Semantic Scholar*[®]; o índice de aderência ao contexto organizacional, que utiliza em sua composição as informações advindas da base de conhecimento; o índice de originalidade fundamentado na base de patentes USPTO[®] e na base de artigos científicos *Semantic Scholar*[®], e o índice de tecnologia com base nas informações presentes na base de conhecimento.

7.2 ANÁLISE DOS RESULTADOS OBTIDOS

Nesta sessão, serão apresentados os resultados obtidos com a aplicação do modelo proposto nesta tese.

7.2.1 Primeiro cenário de análise

Após determinar os cenários de análise a serem utilizados na avaliação foi realizada a primeira série de execuções. Em todas as execuções do protótipo neste primeiro cenário foram consideradas as 200 ideias coletadas no programa Sinapse da Inovação[®].

Cabe ressaltar que as 200 ideias coletadas já passaram por diferentes fases do programa até chegar a etapa semifinal. O que indica que das 1791 ideias inscritas no programa, estas 200 ideias selecionadas possuem grande potencial de implementação e originalidade.

Como mencionado anteriormente, o contexto organizacional pode ser determinado por especialistas da organização ou por editais de fomento. O Sinapse[®] possui um edital que especifica as áreas de interesse, sendo o contexto organizacional descrito por meio da leitura do referido edital. O Quadro 16 apresenta as atividades de impacto identificadas no edital.

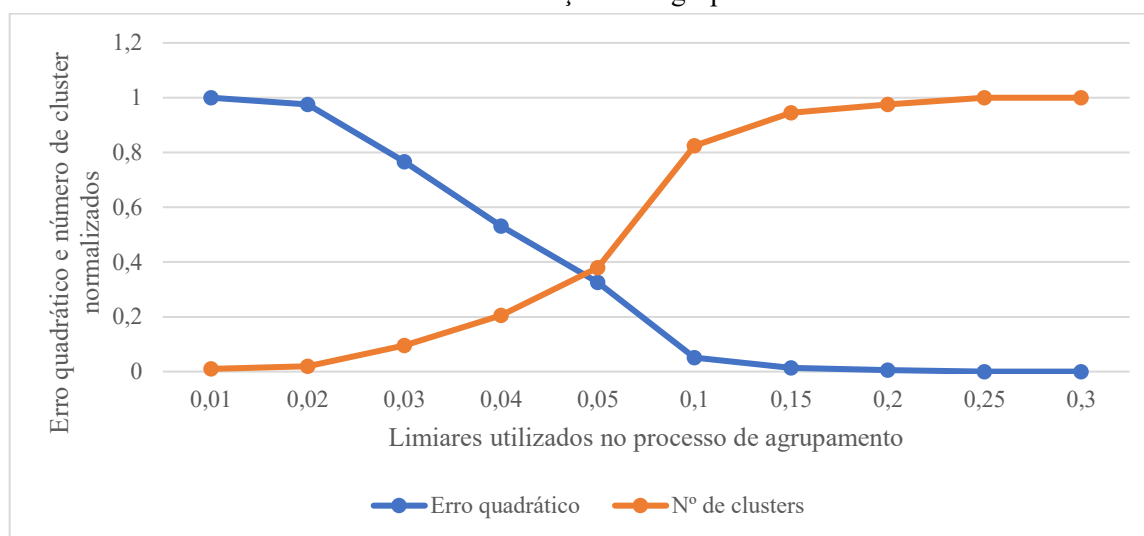
Quadro 16 - Atividades de impacto

Atividades de impacto
Racionalização e humanização da mão de obra
Captação
Armazenamento e uso da água em atividades agrícolas
Máquinas e equipamentos para produção de pequena escala
Redução e mitigação de impactos ambientais das atividades
Automação de processos produtivos agropecuário e pesqueiro
Energias renováveis aplicadas a produção de pequena escala
Informática aplicada nos processos de produção

Fonte: Autora

Também se identificou as áreas de conhecimento descritas no edital: Eletrônica, Mecatrônica, Química e Materiais, Tecnologia da Informação e Comunicação, Nanotecnologia, Biotecnologia, Tecnologia Social, Gestão e Design. A partir das atividades de impacto e das categorias em que os projetos deveriam ser enquadrados de acordo com sua área de conhecimento, identificou-se os principais termos relacionados às potenciais áreas de investimentos e à área de tecnologia, inserindo-os na base. A Figura 39 ilustra alguns dos termos-chave utilizados para compor o vetor de contexto da organização relacionado às potenciais áreas de investimento.

Gráfico 1 - Formação dos agrupamentos

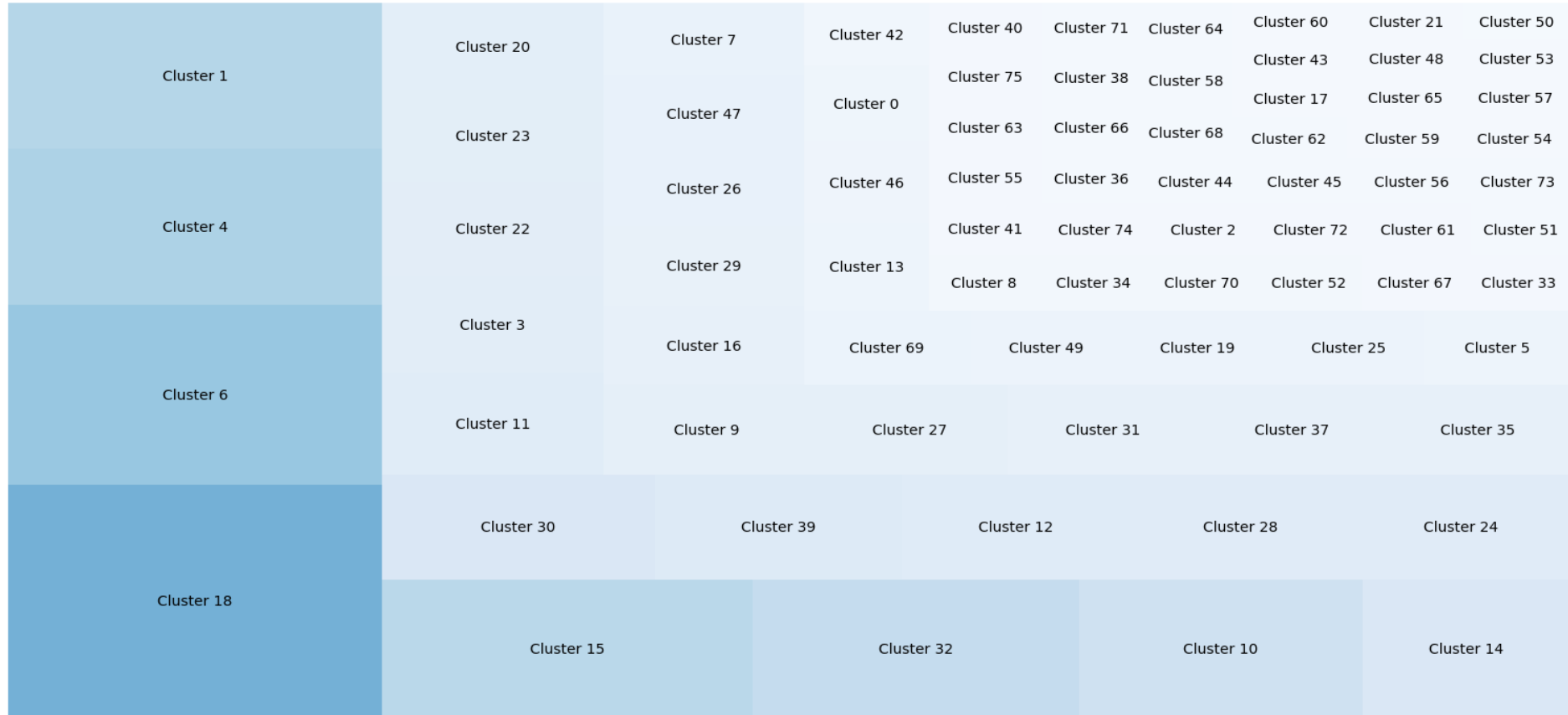


Fonte: Autora

O Gráfico 1 apresenta no eixo x o limiar de similaridade utilizado, e no eixo y a variação tanto para o erro quadrático quanto para o número de agrupamentos formados em cada limiar. Ambas as dimensões foram normalizadas pelo máximo valor, tanto do erro quanto do número de agrupamentos. Neste sentido, a linha azul determina o erro quadrático calculado na formação dos agrupamentos e a linha laranja, o número de *clusters* gerados. Analisando o gráfico, observa-se que no fator de similaridade 0,05 as duas linhas estão próximas, o que na análise do cenário indica um melhor resultado com a aproximação entre o número de *clusters* formados e o erro quadrático calculado, ambos normalizados. Quando se aumenta o limiar de similaridade, consequentemente o número de *clusters* aumenta tendendo ao número de documentos (ideias) no conjunto de dados e o erro quadrático diminui tendendo a zero. Tal situação pode conduzir a um cenário de análise inadequado.

Em decorrência do resultado optou-se por utilizar o fator de similaridade 0,05. Neste cenário foram gerados 76 *clusters* (conforme algoritmo descrito na Seção 6.1.4) com erro quadrático normalizado 0,326. Cabe ressaltar que o fator de similaridade foi mais baixo devido às ideias serem muito diferentes umas das outras, pois o Sinapse[®] aceita ideias de diferentes categorias. A Figura 41 apresenta a *treemap* com os *clusters* formados para o primeiro cenário de análise, partindo do canto inferior esquerdo para o canto superior direito. A dimensionalidade da *treemap* é determinada pelo peso do agrupamento (somatório dos índices de potencialidade das ideias que compõem o agrupamento). A tonalidade do agrupamento ilustra a quantidade de ideias que foram selecionadas na última etapa do concurso. Ou seja, quanto mais escura é a tonalidade do retângulo que representa o agrupamento, maior o número de ideias aprovadas.

Figura 41 – Visualização dos agrupamentos através de *treemap* para o primeiro cenário de análise

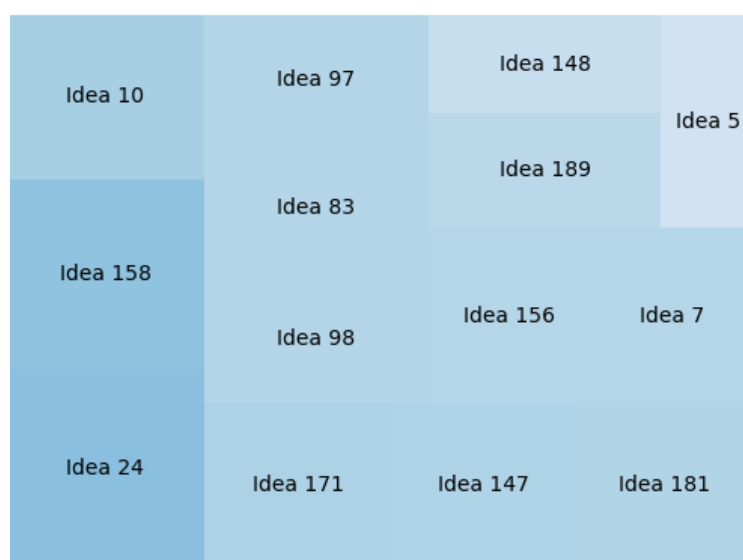


Fonte: Autora

Os *clusters* mais relevantes apresentados na Figura 41 são os que possuem os identificadores 18, 6 e 4. A seguir serão analisados os dez agrupamentos com maior relevância e os identificadores das ideias pertencentes aos mesmos.

O primeiro agrupamento com maior relevância pertencente ao primeiro cenário é o *cluster* 18. A Figura 42 apresenta os identificadores das ideias pertencentes ao agrupamento e a relevância (calculada pelo índice de potencialidade) de cada ideia é determinada pelo seu tamanho na *treemap*. Ou seja, quanto mais escura a tonalidade do retângulo, maior o índice de potencialidade da ideia.

Figura 42 – *Treemap* para o *cluster* 18



Fonte: Autora

A Tabela 20 ilustra as 14 ideias pertencentes ao agrupamento.

Tabela 20 – Ideias que compõem o *cluster* 18

Nº	Id	Título	Selecionada	Índice de Potencialidade
1	24	Glicômetro não-invasivo Inteligente	SIM	0.494677381853069
2	158	Markets Sistema Para Administração de Produtos Em Grandes <i>Marketplaces</i>	SIM	0.475602089465181
3	10	MediBridge - Conectado você com sua saúde	SIM	0.407371547945428
4	171	ilergic - Saúde além do rótulo	SIM	0.390622203799662
5	147	EasyHealth - Plataforma IoT para realização de exames	NÃO	0.376198030008146
6	181	BeePay - O melhor de dois mundos	SIM	0.371825061096217
7	98	BuscaCity - Inteligência Artificial	SIM	0.369867060977018
8	83	Welob - <i>Software</i> para planejamento e gerenciamento de obras	SIM	0.369491646473389
9	97	e-licencie: Garantindo a Sustentabilidade das Empresas	SIM	0.363741998746376
10	156	Goonbo - Gestão das implantações de clientes de tecnologia	NÃO	0.363017500495438

11	7	VetOrtho 3D-Orteses e Próteses Personalizadas em Animais de Companhia	SIM	0.362748629791084
12	189	Sistema de Cotações de Material de Construção	NÃO	0.338090153052392
13	148	Sistema de levantamento automatizado de materiais para construção	NÃO	0.289048785150491
14	5	Sistema Não-Destrutivo para Detecção e Aquisição de Microvestígios	NÃO	0.250099271090608
Soma dos Índices de Potencialidade				5.2224
Total de Ideias Selecionadas				9
Total de Ideias não Selecionadas				5

Fonte: Autora

A primeira coluna corresponde a um número sequencial, a segunda coluna determina o identificador da ideia, a terceira coluna o título da ideia, a quarta se a ideia foi selecionada ou não, e a última coluna ilustra o índice de potencialidade calculado pelo protótipo. Dentre as 14 ideias pertencentes ao *cluster*, nove foram selecionadas na etapa final e cinco não foram selecionadas. O *cluster* 18 possui 64,24% de ideias implementadas, sendo que entre as dez primeiras ideias, somente duas não foram implementadas. A Figura 43 apresenta a nuvem de termos do *cluster* 18.

Figura 43 – Nuvem de termos do *cluster* 18

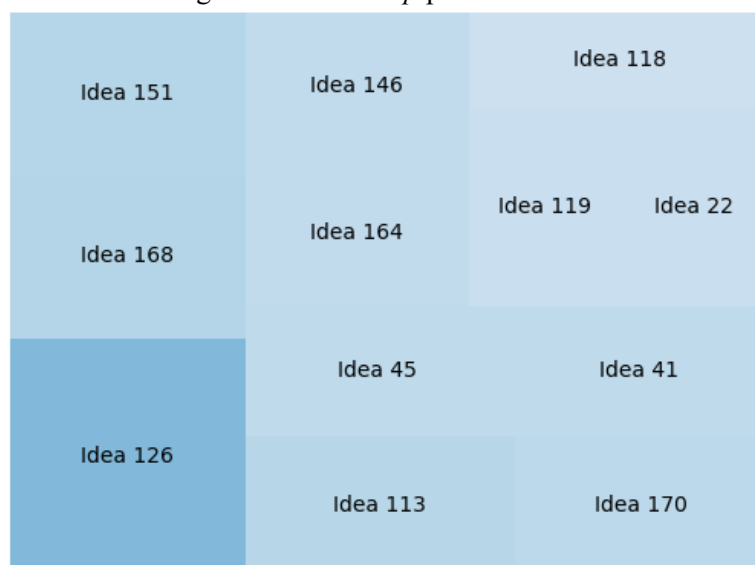


Fonte: Autora

Os principais termos mencionados foram: produtos com 24 ocorrências, usuário com 14, plataforma com 11, cliente com 11 e aplicativo com nove.

O *cluster* 6 corresponde ao segundo *cluster* com maior relevância no cenário de estudo. A Figura 44 apresenta a representação no formato *treemap* com os identificadores das ideias.

Figura 44 – Treemap para o cluster 6



Fonte: Autora

Com o objetivo de descrever o agrupamento, a Tabela 21 apresenta as ideias pertencentes ao mesmo.

Tabela 21 – Ideias que compõem o cluster 6

Nº	Id	Título	Selecionada	Índice de Potencialidade
1	126	FindMed - A saúde em suas mãos	SIM	0.526032533527417
2	168	Azimov Trade	SIM	0.365119348937897
3	151	Inteligência Artificial para Análise de Dívidas com a União	SIM	0.360689145213271
4	113	Lifter de transferência multifuncional para pessoas com deficiência	SIM	0.352881640709508
5	170	Regulariza - Tranquilidade no pagamento de débitos de veículos	NÃO	0.333619374303773
6	45	Equipamento automático de lavagem de automóveis sem escovas.	SIM	0.324721474401247
7	41	Connect Smart Acionamento inteligente para coletores de pó	SIM	0.318893033348983
8	164	Plataforma de Empréstimo P2P - Din Dim <i>Smart Money</i>	SIM	0.312991954052258
9	146	Sistema de monitoração e controle inteligente para chuveiros	NÃO	0.312726126604293
10	119	Balizar - assistente de estacionamento	SIM	0.283714986246055
11	22	Chuveiro Invert	SIM	0.283078724771654
12	118	Sistema recirculador de água fria do banho	NÃO	0.268861344565462
Soma dos Índices de Potencialidade				4.0433
Total de Ideias Selecionadas				9
Total de Ideias não Selecionadas				3

Fonte: Autora

As ideias que possuem maior relevância são as ideias 149, 169 e 150. A Tabela 22 explicita as ideias pertencentes ao agrupamento com os respectivos índices de potencialidade gerados pelo protótipo.

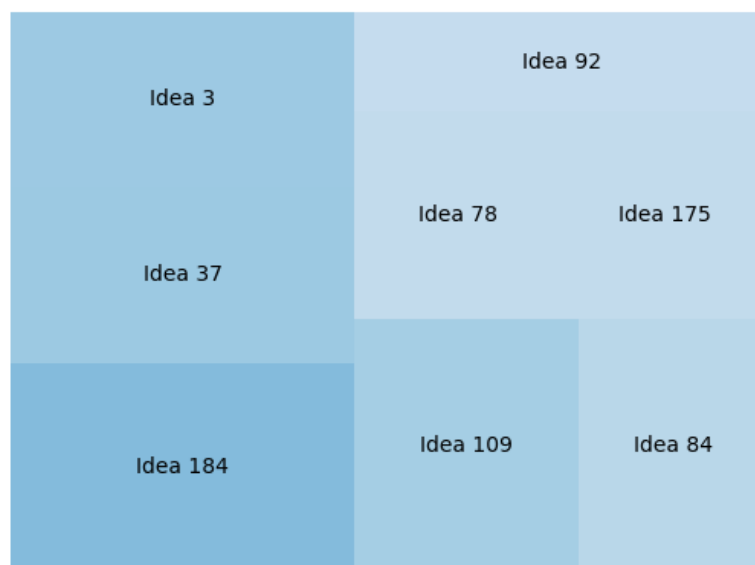
Tabela 22 – Ideias que compõem o *cluster* 4

Nº	Id	Título	Selecionada	Índice de Potencialidade
1	149	Plataforma Integrada de Gestão e Atenção à Saúde	NÃO	0.454353521885293
2	169	Registro de prontuário medico utilizando a tecnologia <i>blockchain</i>	NÃO	0.394013718582871
3	150	LoyalNow - Promova a satisfação de clientes em tempo real	NÃO	0.385322181589884
4	80	LabCloud - Inovação para Processos Laboratoriais Veterinários	SIM	0.379421066855758
5	101	Dynamique - <i>Smart Business</i>	NÃO	0.358085736974817
6	105	EletroBlocks - Blocos eletrônicos modulares e programáveis	SIM	0.341186084378641
7	86	PagueDepois - Marcou hoje!?! PagueDepois	SIM	0.328289940434362
8	161	SinSalarial - Remuneração Inteligente	SIM	0.312784090979846
9	93	TERTIUM gestão de mão de obra terceirizada empresas pública e privadas	SIM	0.297037606839737
10	194	SIGPEP (Sistema Integrado de Prontuário Eletrônico de Pacientes)	NÃO	0.236123295492491
Soma dos Índices de Potencialidade				3.4866
Total de Ideias Selecionadas				5
Total de Ideias não Selecionadas				5

Fonte: Autora

O terceiro *cluster* com maior relevância, apresentado na Tabela 22 possui a soma dos índices de potencialidade das ideias de 3,4866. Na formação do agrupamento há a presença de cinco ideias selecionadas e cinco não selecionadas, correspondendo a um total de 50% de assertividade. A nuvem de termos do agrupamento 4 é apresentado na Figura 47.

Figura 50 – Treemap para o cluster 15



Fonte: Autora

Com o intuito de especificar as ideias pertencentes ao agrupamento, a Tabela 24 apresenta as ideias do quinto *cluster* com maior relevância no cenário de análise.

Tabela 24 – Ideias que compõem o *cluster* 15

Nº	Id	Título	Selecionada	Índice de Potencialidade
1	184	Joga+Perfil	SIM	0.517440008606402
2	37	Farmstec - Inovações Tecnológicas	SIM	0.439191529207651
3	3	O3System - A Garantia de Sanidade no Agronegócio.	SIM	0.434540381148166
4	109	RuralMarket - Inteligência de mercado aplicada ao meio rural	SIM	0.411225189699982
5	84	App Gestão de Propriedades Rurais Familiares	NÃO	0.347026576591041
6	78	FIRE - Gestão de Clubes de Caca e Tiro	NÃO	0.308903966152792
7	175	Vitula - sistema de gestão de pequenas propriedades	NÃO	0.308638568841463
8	92	Mais Leite	NÃO	0.297428902507722
Soma dos Índices de Potencialidade				3.0643
Total de Ideias Selecionadas				4
Total de Ideias não Selecionadas				4

Fonte: Autora

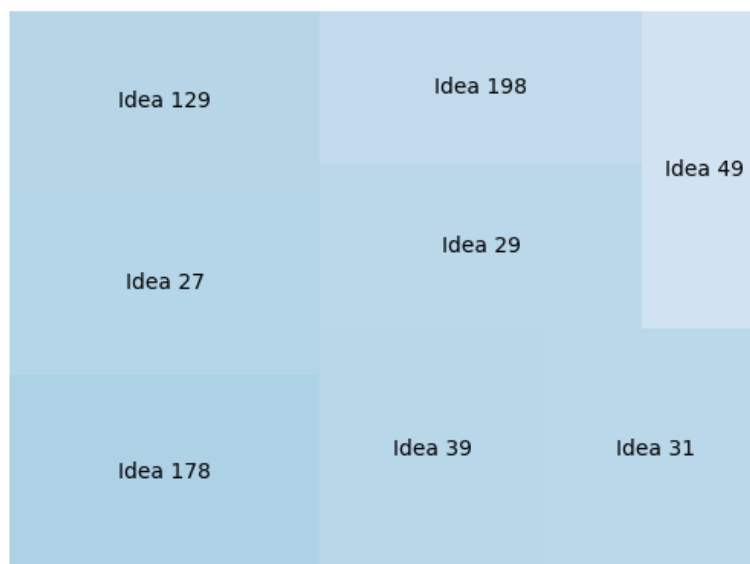
Na Tabela 24 novamente há um empate, quatro ideias foram selecionadas para a etapa final e quatro não foram. Contudo, as quatro com maiores índices de potencialidade permaneceram no topo do agrupamento. A soma dos índices de potencialidade das ideias pertencentes ao agrupamento foi de 3,0643. A Figura 51 apresenta a nuvem com os termos mais citados.

Figura 51 – Nuvem de termos do *cluster* 15

Fonte: Autora

Os termos mais citados foram: gestão com 16 ocorrências, produção com dez, desempenho e dados com nove ocorrências, acompanhamento e rural com sete.

A Figura 52 apresenta a *treemap* correspondente ao sexto agrupamento.

Figura 52 – *Treemap* para o *cluster* 32

Fonte: Autora

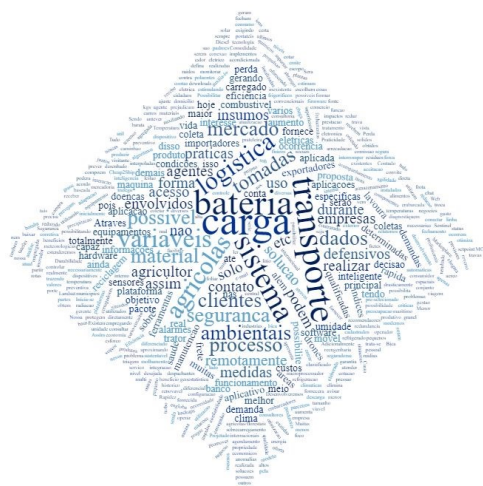
A Tabela 25 ilustra as ideias pertencentes ao *cluster* de número 32.

Tabela 25 – Ideias que compõem o *cluster* 32

Nº	Id	Título	Selecionada	Índice de Potencialidade
1	178	Quiron - Sistema de Gestão Remota da Sanidade Agrícola e Florestal	SIM	0.388450718824773
2	27	proFARM auxilia agricultores a tomar as melhores decisões	NÃO	0.360224150113768
3	129	Reconsidere - Gestão sustentável de resíduos sólidos	SIM	0.356290899500202
4	39	AtlasPower - Tornando baterias inteligentes	SIM	0.347930343711726
5	31	YAK - Trator Elétrico	SIM	0.343453089128944
6	29	KOHNNECT - Sistema de Controle e Telemetria de Baús	SIM	0.340032433108844
7	198	Cheap2Ship - Sistema inovador de cotação de fretes	SIM	0.312965325283814
8	49	Bica de caminhão Graneleiro automática	NÃO	0.251902170842536
Soma dos Índices de Potencialidade				2.7012
Total de Ideias Selecionadas				6
Total de Ideias não Selecionadas				2

Fonte: Autora

O *cluster* 32 é composto por oito ideias, sendo seis selecionadas e duas não selecionadas, um total de 75% de assertividade. A soma dos índices do agrupamento foi de 2,7012. A Figura 53 apresenta a nuvem de termos do *cluster* 32.

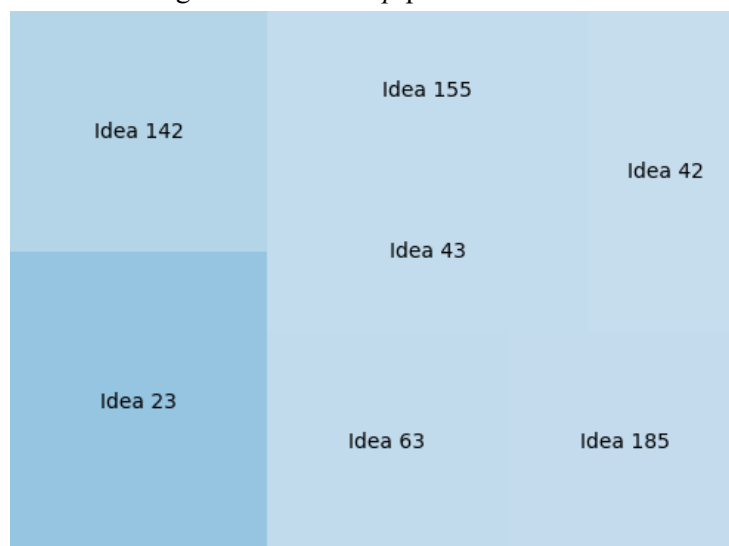
Figura 53 – Nuvem de termos do *cluster* 32

Fonte: Autora

Os termos mais citados no agrupamento foram: carga mencionada 12 vezes, bateria mencionada oito vezes, sistema e transporte sete vezes, logística, agrícolas e variáveis mencionadas cinco vezes.

O *cluster* 10 corresponde ao sétimo *cluster* com maior relevância no cenário de estudo. A Figura 54 apresenta *treemap* correspondente ao agrupamento gerado.

Figura 54 – *Treemap* para o *cluster* 10



Fonte: Autora

As ideias com maior relevância foram as ideias com os identificadores 23, 143 e 63. A Tabela 26 apresenta a formação do *cluster* 10.

Tabela 26 – Ideias que compõem o *cluster* 10

Nº	Id	Título	Selecionada	Índice de Potencialidade
1	23	Fertileasy - Monitoramento e controle de produção agrícola	SIM	0.457499398721715
2	142	Fleet Manager - IOT Aplicado a Veículos	SIM	0.367408473734708
3	63	Nanocosméticos com antioxidante obtido de organismo aquático	SIM	0.312755300489122
4	185	MobiFleet - Plataforma de Gerenciamento de Frotas e Ativos	SIM	0.308092355092034
5	43	Controle de Irrigação por pivô central	NÃO	0.303810253088373
6	155	FIRECALL - <i>Software</i> de Alerta Inteligente De Incêndio	NÃO	0.302027437760918
7	42	Qendu - Segurança Eletrônica	SIM	0.294132936701603
Soma dos Índices de Potencialidade				2.3457
Total de Ideias Selecionadas				5
Total de Ideias não Selecionadas				2

Fonte: Autora

A Tabela 26 apresenta o agrupamento composto por sete ideias. Somente duas não foram selecionadas para a etapa final. Ao total, 71,42% das ideias foram aprovadas. As ideias aprovadas e com maiores índices de potencialidade permaneceram no topo. A Figura 55 apresenta os termos mais citados no agrupamento: sistema 14 vezes, ativos nove vezes,

informações e nuvem sete vezes, incêndio e usuário seis vezes, irrigação, solução e dados cinco vezes.

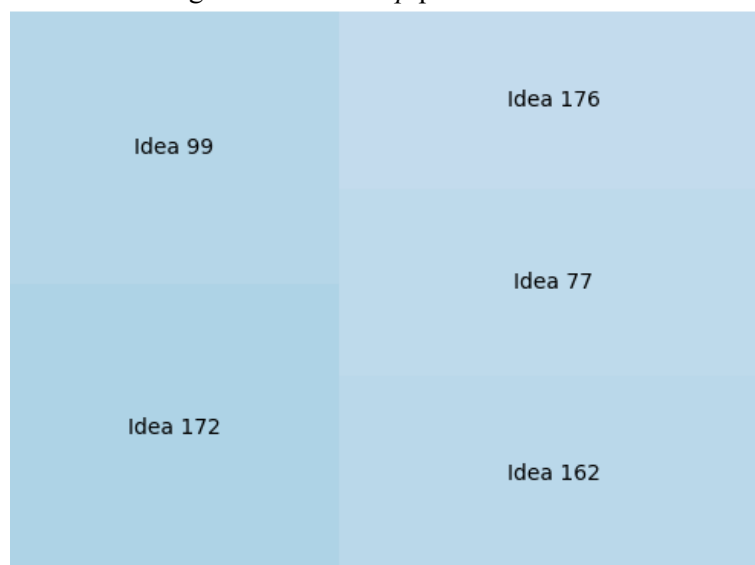
Figura 55 – Nuvem de termos do *cluster* 10



Fonte: Autora

A Figura 56 apresenta a representação no formato *treemap* do agrupamento 14.

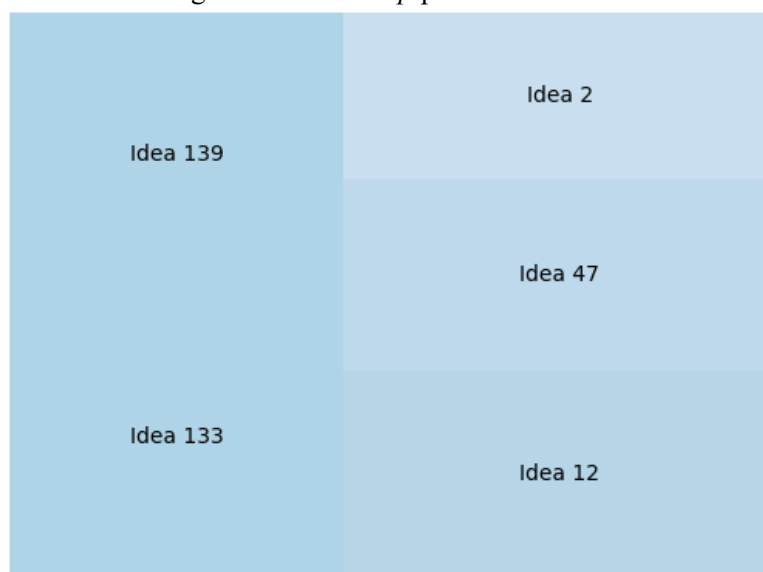
Figura 56 – *Treemap* para o *cluster* 14



Fonte: Autora

Objetivando detalhar as ideias pertencentes ao agrupamento, a Tabela 27 ilustra o oitavo *cluster* formado.

Figura 58 – Treemap para o cluster 30



Fonte: Autora

Na Tabela 28 é possível visualizar as ideias pertencentes ao penúltimo agrupamento analisado.

Tabela 28 – Ideias que compõem o cluster 30

Nº	Id	Título	Selecionada	Índice de potencialidade
1	133	Adubooo - Plataforma de localização de resíduos orgânicos	SIM	0.376043059517708
2	139	SaladaFácil	NÃO	0.37560577764316
3	12	APATUIA- Frutas Estruturadas com Umidade	SIM	0.35732261345226
4	47	Eco Alliance - Recicladora de resíduos orgânicos	NÃO	0.328981025584187
5	2	Utilização da Polpa de Pinhão como Substrato para Produtos Lácteos	NÃO	0.285072456111231
Soma dos Índices de Potencialidade				1.7230
Total de Ideias Selecionadas				2
Total de Ideias não Selecionadas				3

Fonte: Autora

No penúltimo agrupamento analisado, apresentado na Tabela 28, há a presença de cinco ideias. A partir deste momento observa-se que os *clusters* começam a perder relevância. Neste agrupamento apenas duas ideias foram selecionadas. A nuvem com os termos mais citados no agrupamento é ilustrada na Figura 59, e os termos mais coocorrentes foram: resíduos, orgânicos, alimentos, localização, plataforma, material, frutas, entre outros.

Ao sumarizar as análises acima há os seguintes dados apresentados na Tabela 30.

Tabela 30 – Sumarização dos agrupamentos

Nº	Cluster	Ideias selecionadas	Ideias não selecionadas	Porcentagem de selecionadas
1	18	9	5	64,28%
2	6	9	3	75,00%
3	4	5	5	50,00%
4	1	6	3	66,66%
5	15	4	4	50,00%
6	32	6	2	75,00%
7	10	5	2	71,42%
8	14	3	2	60,00%
9	30	2	3	40,00%
10	39	1	4	20,00%

Fonte: Autora

A primeira coluna refere-se a um sequencial, a segunda coluna a identificação do *cluster*, a terceira coluna ao número de ideias selecionadas na etapa final, a quarta apresenta o número de ideias não selecionadas e, por fim, a última coluna ilustra a porcentagem de ideias selecionadas no *cluster*.

Os dois primeiros *clusters* possuem 18 ideias selecionadas de um total de 26, o que corresponde a aproximadamente 70% de assertividade. Esse percentual corresponde a cerca de 64%, quando são considerados os três primeiros *clusters*. Considerando os dez *clusters* apresentados, a assertividade foi de 80%. Este valor corresponde ao número de ideias selecionadas maior ou a ocorrência de empate entre selecionadas ou não.

Outra possibilidade de análise envolve a recuperação de todas as ideias ordenadas de maneira decrescentemente pelo índice de potencialidade. A Tabela 31 apresenta as 30 primeiras ideias.

Tabela 31 – Ideias ordenadas pelo índice de potencialidade

Nº	Id	Título	Selecionada	Índice de Potencialidade
1	126	FindMed - A saúde em suas mãos	SIM	0.526032533527417
2	184	Joga+Perfil	SIM	0.517440008606402
3	94	Sistema Infravia - Sistema de Infraestrutura Urbana	SIM	0.505149673143376
4	24	Glicosímetro não-invasivo Inteligente	SIM	0.494677381853069
5	158	Markets Sistema Para Administração de Produtos Em Grandes <i>Marketplaces</i>	SIM	0.475602089465181
6	173	MS CondoHub - Gestão de Serviços para Administradoras de Condomínio	SIM	0.468094200940046
7	67	TechPain: A revolução do tratamento da dor crônica na artrite	NÃO	0.467139262542341
8	23	Fertileasy - Monitoramento e controle de produção agrícola	SIM	0.457499398721715
9	149	Plataforma Integrada de Gestão e Atenção a Saúde	NÃO	0.454353521885293
10	37	Farmstec - Inovações Tecnológicas	SIM	0.439191529207651
11	3	O3System, A Garantia de Sanidade no Agronegócio.	SIM	0.434540381148166
12	1	Petiscos para pet, a partir de resíduos e agregando o fator probiótico	SIM	0.422893984378469
13	109	RuralMarket - Inteligência de mercado aplicada ao meio rural	SIM	0.411225189699982
14	91	Abellion - A inteligência que sua plantação precisa	SIM	0.407644472999114
15	10	MediBridge - Conectado você com sua saúde	SIM	0.407371547945428
16	90	Kluc - obtenha insights dos seus dados através de simples perguntas	SIM	0.39944920038607
17	33	agrotechlink: agricultura e tecnologia conectadas.	SIM	0.396501366684544
18	160	Aplicativo Situation - Criar uma interação entre o virtual e o real	NÃO	0.395302236239693
19	174	ManejeBem: Rede Social para Agricultores	SIM	0.39460242969367
20	169	Registro de prontuário médico utilizando a tecnologia <i>blockchain</i>	NÃO	0.394013718582871
21	171	ilergic - Saúde além do rotulo	SIM	0.390622203799662
22	4	Vestal Oncogenomics - Plataforma de análise molecular oncológica	NÃO	0.389704331438049
23	196	SoftSport -Modelo de treino para escola de base de futebol e futsal	NÃO	0.388525219499796
24	178	Quiron - Sistema de Gestão Remota da Sanidade Agrícola e Florestal	SIM	0.388450718824773
25	18	Termobook - Tomada de decisão baseada em registros de temperatura	SIM	0.385701993445489
26	150	Loyalnow - Promova a satisfação de clientes em tempo real	NÃO	0.385322181589884
27	172	IA para reconhecimento de fraudes na rede de distribuição de energia	NÃO	0.383397631169208
28	153	Leoa - Plataforma SaaS com <i>ChatBot</i> para Facilitar a Declaração de IR	SIM	0.382051129670249
29	48	ARMTECH - Assistente Cirúrgico Robótico	SIM	0.379580523020649
30	80	LabCloud - Inovação para Processos Laboratoriais Veterinários	SIM	0.379421066855758

Fonte: Autora

Das dez primeiras ideias apresentadas, oito foram selecionadas na etapa final. Ao analisar-se as 20 primeiras ideias ordenadas pela média dos índices, observa-se que apenas quatro não foram selecionadas pelos especialistas do programa, ao total 80% das ideias foram selecionadas.

Ao analisar as 30 primeiras ideias ordenadas, observa-se que 22 delas foram selecionadas pelos especialistas do programa na fase final, totalizando a assertividade em 73,33%. Das 100 primeiras ideias ordenadas pela média dos índices gerados pelo protótipo, 63 foram selecionadas na etapa final.

Por fim, o algoritmo proposto melhorou em aproximadamente 70% o resultado final se considerado apenas a ordenação pelo índice de potencialidade da ideia e se ela foi implementada ou não. Ressalta-se que a base de ideias para análise é composta por apenas 200 ideias pré-selecionadas e que pertencem a etapa semifinal do programa. Tendo em vista as considerações acerca das ideias, pode-se dizer que estas 200 já são originais e com altas perspectivas de implementação.

7.2.1.1 Análise da formação da rede de palavras do primeiro cenário

Com o objetivo de realizar uma análise mais visual do contexto de estudo, optou-se pela aplicação da análise de redes, constituindo assim um elemento relevante para avaliação do trabalho.

O intuito é explorar o fluxo das informações e os relacionamentos existentes na formação da rede de termos, utilizando a análise de redes. Para tal, utilizou-se a ferramenta Gephi[®]. Esta ferramenta é empregada na exploração e compreensão de grafos voltados à análise exploratória de dados. O usuário interage com a representação e manipulação das estruturas, formas e cores, com o intuito de revelar propriedades ocultas dos dados de análise (GEPHI, 2019). O objetivo da ferramenta é “ajudar os analistas de dados a fazerem hipóteses, intuitivamente descobrir padrões, isolar singularidades de estruturas ou falhas durante a terceirização de dados” (GEPHI, 2019).

Segundo os autores Ye, Song e Li (2012), as medidas comumente utilizadas durante a análise são a densidade, a distância média, o diâmetro e o grau médio. Medidas como a modularidade também são relevantes (NEWMAN, 2006).

A densidade é utilizada para indicar o nível de conectividade da rede. Redes com alta densidade representam alta conectividade (KATERNDAHL, 2012). A distância média entre nodos da rede indica o número médio de intermediários ou caminho de possibilidades entre

dois termos quaisquer. O diâmetro indica o caminho mais longo. O grau médio de uma rede determina o número médio de relacionamentos entre os nodos (MARTINS et al., 2012). A modularidade é uma medida que objetiva estimar a qualidade dos agrupamentos gerados em uma rede, ou seja, a qualidade da divisão/distribuição realizada na rede (NEWMAN, 2006). Em redes reais estes valores ocorrem entre 0,3 e 0,7; pois redes com valores maiores de modularidade possuem conexões densas entre os nós dentro do agrupamento e conexões esparsas entre nós de diferentes agrupamentos (NEWMAN; GIRVAN, 2004).

A análise presente neste trabalho utilizou como parâmetros de entrada:

- Visão geral da rede: grau médio, o diâmetro da rede e modularidade;
- Algoritmo de distribuição: *Force Atlas 2*;
- Partição dos nós: realizada por meio do grau;
- Classificação dos nós: realizada por meio do índice de modularidade;

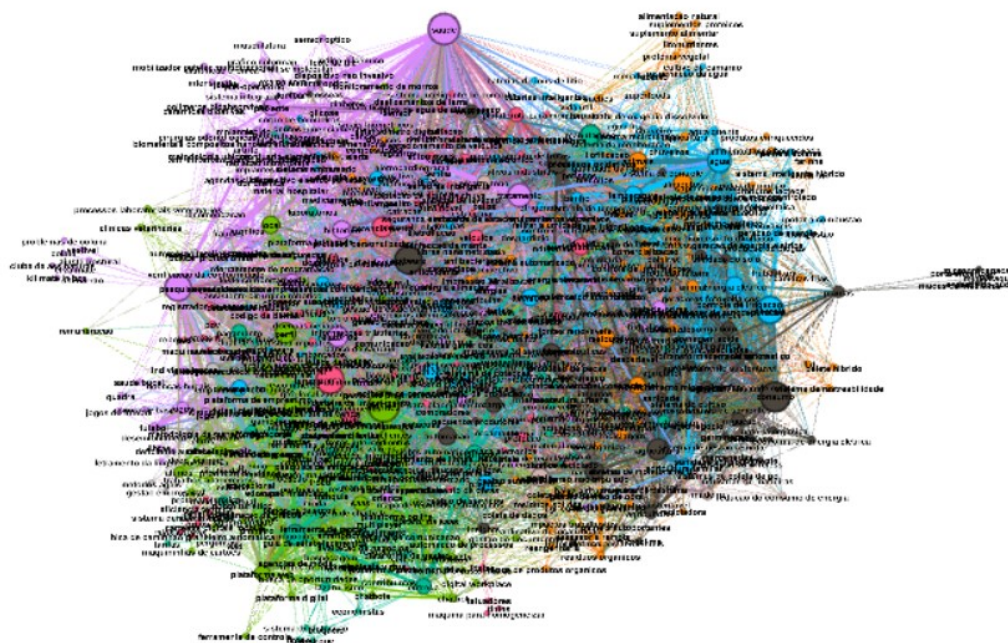
A avaliação do cenário de estudo levou em consideração as palavras-chave (termos) das 200 ideias coletadas no site Sinapse da Inovação[®]. Foram coletados ao todo 715 termos, gerando 9353 coocorrências. Inicialmente o peso da coocorrência é determinado pelo número de vezes que dois termos são mencionados em conjunto nas ideias. Posteriormente, é realizada a normalização dos valores pela raiz quarta, dividindo-se cada peso pelo valor máximo encontrado.

Segundo Campos (2000), a normalização é importante quando os requisitos para o emprego da estatística paramétrica, entre eles, normalidade da distribuição dos erros, homogeneidade das variâncias e aditividade dos efeitos dos fatores de variação, não puderem ser preenchidos pelos dados da amostra experimental. Este recurso é evidentemente mais poderoso que a abordagem não paramétrica devido à própria natureza dos dados, muitas vezes não numéricos (CAMPOS, 2000).

Após a normalização dos dados, foram retirados os termos mais genéricos como: sistema, produtos, processos, empresas, plataforma, produção e negócio. Diversas análises foram realizadas considerando os pesos das arestas de 0,4, 0,5, 0,6, 0,7 e 0,8.

A primeira análise ocorreu com o valor 0,4. Contudo, devido à densidade da rede, a análise visual torna-se difícil. A Figura 62 apresenta a rede.

Figura 62 - Formação da rede com peso das arestas igual a 0,4



Fonte: Autora

Ao total, a rede possui 662 nodos e 7258 arestas. O grau médio da rede é 21,927. Neste sentido, como o grau médio determina o número médio de relacionamentos entre os nodos, observa-se certa dificuldade de separabilidade, indicando uma rede com alta conectividade. O diâmetro da rede é 5, a densidade é 0,330 e a modularidade 0,312. Os nodos com maior grau são: saúde com 191, clientes com 190, consumo com 179, aplicativo com 176, conta com 166, *software* com 163, energia com 161, pesquisa com 157, comunicação com 135, água com 134, tratamento com 134. A partir da análise inicial aplicou-se um algoritmo de organização da rede. Entre as diversas opções disponíveis na ferramenta Gephi[®] optou-se pelo *Force Atlas 2* devido a qualidade da distribuição produzida. O *Force Atlas 2* utiliza um algoritmo baseado no conceito da física de atração e repulsão. De certo modo, objetiva resolver as deficiências do algoritmo *Force Atlas*, fazendo um equilíbrio entre a qualidade do *layout* final e a velocidade do algoritmo computacional. Seu desempenho para redes maiores é melhor quando comparado ao algoritmo *Force Atlas*.

Na sequência, visando um entendimento visual mais adequado da estruturação da rede na formação de grupos, o algoritmo de modularidade foi utilizado. Este algoritmo analisa o nível de conectividade (grau) de determinado nó para determinar a sua coloração. Ou seja, um nó em específico terá determinada cor caso este esteja mais conectado a nós vinculados a uma cor em particular. Ao final, há uma perspectiva de divisão dos nós através de diferentes colorações, em que cada coloração determina um possível grupo/classe.

Por fim, foi ajustado o tamanho dos nós que pode ser obtido considerando, por exemplo, o seu grau ou informações de centralidade. Para a rede acima se utilizou o grau, contudo, em função da densidade da rede, não houve alterações significativas quando comparado à centralidade.

Entretanto, para os objetivos desta análise, torna-se mais adequado a avaliação dos nodos mais representativos de cada grupo. Neste sentido, a análise será realizada por meio do nodo que possui maior grau em cada agrupamento. Termos com grau abaixo de 50 não serão considerados. A Tabela 32 apresenta os termos centrais de cada agrupamento com seus respectivos graus.

Tabela 32 - Termos pesquisados no cenário com peso 0,4

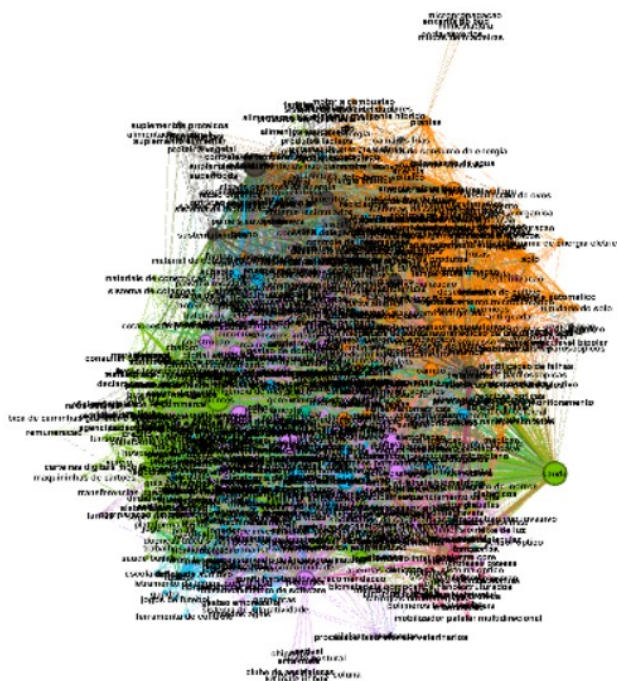
Termo	Grau	Coloração do Grupo
Clientes	190	Verde claro
Comunicação	135	Cinza claro
Conta	166	Rosa
Alimentos	100	Laranja
Saúde	191	Roxo
Consumo	179	Cinza escuro
Negócios	113	Verde escuro
Energia	161	Azul
Coleta	96	Marrom claro

Fonte: Autora

Ao realizar a pesquisa, considerando os termos acima, foram retornadas 103 ideias distintas, sendo que 56 (54,36%) ideias foram selecionadas para implementação. Analisando a classificação das ideias ordenadas pelo índice de potencialidade, verificou-se que, das dez primeiras ideias, nove foram selecionadas (90%). Entre as 20 primeiras, 16 foram selecionadas (80%).

A segunda análise ocorreu com o limiar de 0,5. Todavia, a rede também ficou densa, dificultando a análise visual devido ao número elevado de nodos e, principalmente, ao número de arestas. A Figura 63 apresenta a formação da rede.

Figura 63 - Formação da rede com peso das arestas igual a 0,5



Fonte: Autora

Ao total, a rede foi composta por 662 nodos e 7.258 arestas produzindo um grau médio da rede de 21,927, densidade de 0,330, diâmetro da rede de cinco e modularidade 0,321. Os nodos com maior grau são: saúde com 191, clientes com 190, consumo com 179, aplicativo com 176, conta com 166, *software* com 163, energia com 161, pesquisa com 157, comunicação com 135, água com 134, tratamento com 134. A Tabela 33 apresenta os termos centrais de cada agrupamento com seus respectivos graus.

Tabela 33 - Termos pesquisados no cenário com peso 0,5

Termo	Grau	Coloração do Grupo
Comunicação	135	Azul
Energia	161	Laranja
Saúde	191	Verde claro
Entrega	74	Verde escuro
Aplicativo	176	Roxo
Coleta	96	Cinza claro
Consumo	179	Cinza escuro
Paciente	64	Rosa

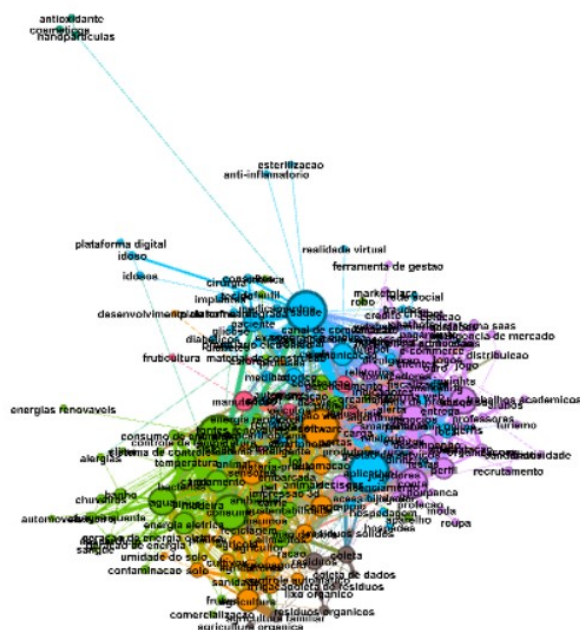
Fonte: Autora

Ao realizar a pesquisa, considerando os termos acima, foram retornadas 101 ideias distintas, sendo que 59 (58,41%) destas foram selecionadas para implementação. Ao analisar a classificação das ideias ordenadas pelo índice de potencialidade, foi verificado que das dez

primeiras ideias, oito foram selecionadas (80%). Entre as 20 primeiras ideias, 15 foram selecionadas (75%).

A terceira análise ocorreu com limiar de 0,6, recuperando ainda muitos termos, o que dificulta uma análise visual. Ao total, a rede foi composta por 205 nodos e 930 arestas. A Figura 64 apresenta a formação da rede.

Figura 64 - Formação da rede com peso das arestas igual a 0,6



Fonte: Autora

O grau médio da rede é 9,073, densidade de 0,044, o diâmetro da rede é 5 e a modularidade é 0,353. Os nodos com maior grau são: saúde e aplicativo com 61, energia com 59, *software* com 57, clientes com 55, consumo com 52, conta com 43, água com 38, pesquisa com 37, negócios com 34 e sensores com 31.

A análise também é realizada por meio do nodo que possui maior grau em cada agrupamento. O algoritmo identificou sete agrupamentos e a Tabela 34 apresenta os termos.

Tabela 34 - Termos pesquisados no cenário com peso 0,6

Termo	Grau	Coloração do Grupo
Energia	59	Verde claro
Resíduos	19	Cinza escuro
Software	57	Laranja
Manutenção	24	Rosa
Nanopartículas	3	Verde escuro
Clientes	55	Roxo
Saúde	61	Azul

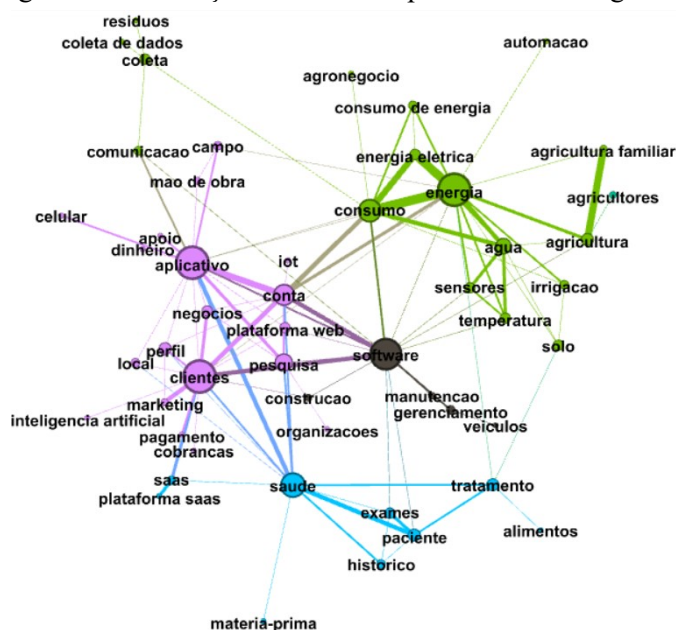
Fonte: Autora

Ao realizar a pesquisa em questão foram retornadas 81 ideias distintas, sendo que 48 (59,25%) foram selecionadas. Ao analisar a classificação das ideias, ordenadas pelo índice de potencialidade, foi verificado que das dez primeiras ideias, nove foram selecionadas (90%). Entre as 20 primeiras ideias, 14 foram selecionadas (70%).

O penúltimo cenário analisado ocorreu com limiar de 0,7. Devido à presença de menos nodos e arestas, a visualização foi facilitada. A Figura 65 apresenta a formação da rede.

O grau médio da rede é 3,864, a densidade é 0,067 o diâmetro da rede é 6,0 e a modularidade é 0,442. Pode-se perceber a partir dos valores das métricas uma rede mais adequada para avaliação, visto que o diâmetro se manteve próximo dos demais estudos e o índice de modularidade foi superior aos estudos anteriores. Os nodos com maior grau são: aplicativo, energia e clientes com grau de 17, *software* com 16, saúde com 12, consumo com 11, conta com dez, pesquisa com oito, água com sete e agricultura, perfil, tratamento e paciente com cinco.

Figura 65 - Formação da rede com peso das arestas igual a 0,7



Fonte: Autora

A análise a seguir considera o nodo que possui maior grau em cada agrupamento. O algoritmo de modularidade identificou oito agrupamentos, contudo, três não serão analisados, pois possuem grau 1. A Tabela 35 apresenta os termos.

Tabela 35 - Termos pesquisados no cenário com 0,7

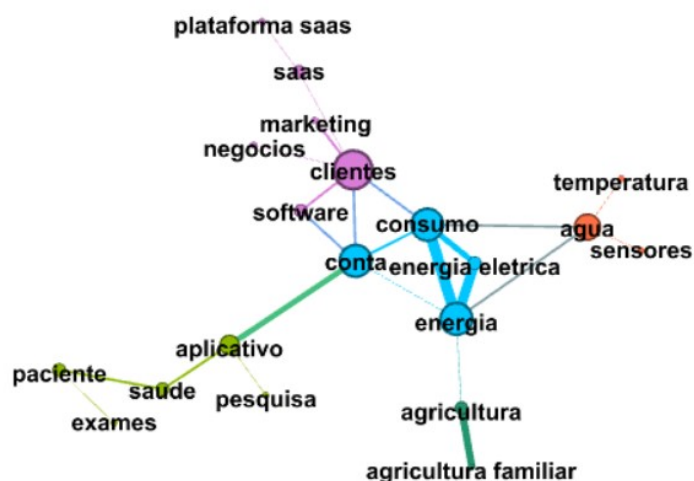
Termo	Grau	Coloração do grupo
Aplicativo	17	Roxo
Energia	17	Verde
Saúde	12	Azul
Software	16	Preto
Agricultores	2	Verde

Fonte: Autora

Ao realizar a pesquisa foram retornadas 82 ideias distintas, sendo que 44 foram selecionadas (53,65%). Ao analisar o *ranking* das ideias ordenadas pelo índice de potencialidade, verificou-se que das dez primeiras ideias, oito foram selecionadas (80%). Entre as 20 primeiras ideias, 14 foram selecionadas (70%).

O último cenário considerou o peso maior ou igual 0,8. A Figura 66 apresenta a formação da rede.

Figura 66 - Formação da rede com peso das arestas igual a 0,8



Fonte: Autora

O grau médio da rede é 2,4, a densidade é 0,126, o diâmetro da rede é 7 e a modularidade é 0,453, sendo estes valores similares ao estudo anterior com limiar de 0,7. Os nodos com maior grau são: clientes com seis, energia, consumo e conta com cinco, água com quatro e aplicativo com três.

A análise a seguir é realizada por meio do nodo que possui maior grau em cada agrupamento. O algoritmo identificou cinco agrupamentos com os termos principais apresentados na Tabela 36.

Tabela 36 - Termos pesquisados no cenário com 0,8

Termo	Grau	Coloração do Grupo
Energia	5	Azul
Clientes	6	Roxo
Aplicativo	3	Verde claro
Agricultura	2	Verde escuro
Água	4	Laranja

Fonte: Autora

Ao realizar a pesquisa foram retornadas 77 ideias distintas, sendo que 43 foram selecionadas (55,84%). Ao analisar o *ranking* das ideias, ordenado pelo índice de potencialidade, verificou-se que entre as dez primeiras ideias, oito foram selecionadas (80%). Já entre as 20 primeiras ideias, 14 foram selecionadas (70%).

Com o intuito de facilitar a visualização e comparação dos resultados apresentados, a Tabela 37 ilustra os valores obtidos para cada limiar.

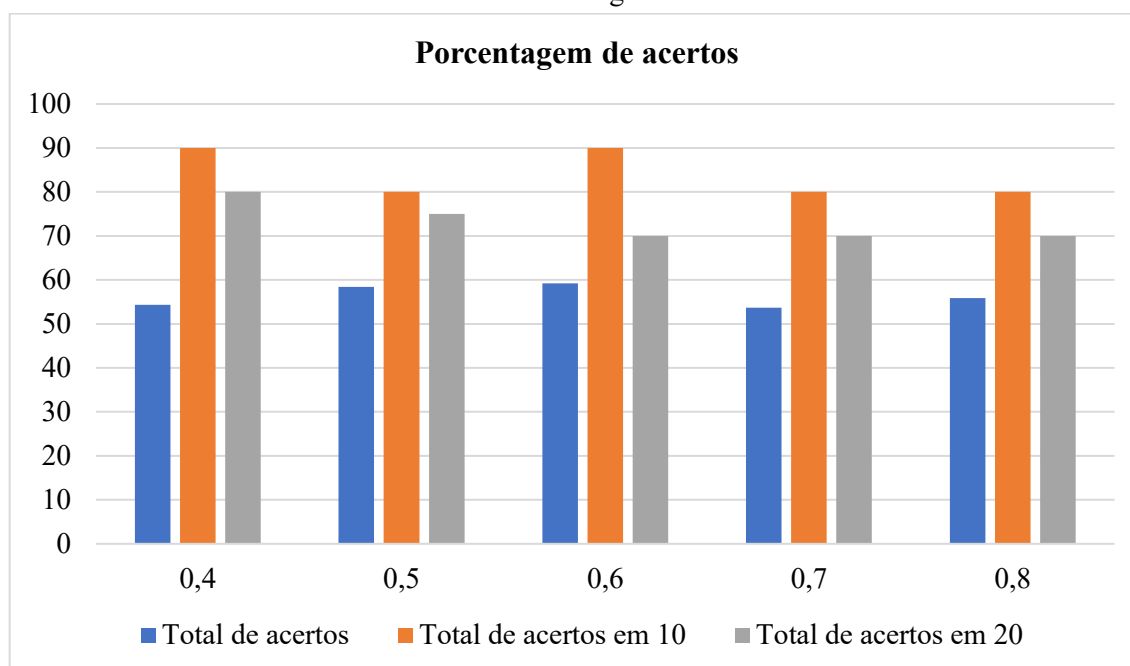
Tabela 37 - Síntese dos resultados

Cenário de Análise	0,4	0,5	0,6	0,7	0,8
Grau	21,927	21,927	9,073	3,864	2,4
Grau médio	147,88	239,11	39,71	12,8	4
Densidade	0,033	0,033	0,044	0,067	0,126
Diâmetro	5	5	5	6	7
Modularidade	0,312	0,321	0,353	0,442	0,453
Percentual de acertos entre as 10 primeiras	90%	80%	90%	80%	80%
Percentual de acertos entre as 20 primeiras	80%	75%	70%	70%	70%
Percentual total de acertos	54,36%	58,41	59,25%	53,65%	55,84

Fonte: Autora

Analisando a porcentagem de acertos do total de ideias selecionadas e considerando as dez primeiras ideias retornadas e as 20 primeiras ideias retornadas ordenadas pelo índice de potencialidade, verificou-se que os valores são próximos. Contudo, devido à expressividade visual da rede e valores retornados, o índice 0,7 demonstra ser o mais adequado. Cabe ressaltar que o edital recebeu aproximadamente 1.800 ideias e a análise ocorreu com as 200 ideias pertencentes à etapa semifinal do processo de seleção, ou seja, são ideias que possuem alto grau de potencialidade e originalidade. O Gráfico 2 apresenta o comparativo de acertos.

Gráfico 2 - Porcentagem de acertos



Fonte: Autora

Também foram analisadas as atividades de impacto relacionadas ao edital, conforme apresentadas no Quadro 17.

Quadro 17 - Atividades de impacto

Nº	Atividades de Impacto	Termo-chave da Análise
1	Racionalização e humanização da mão de obra	Mão de obra
2	Captação	Captação
3	Armazenamento e uso da água em atividades agrícolas	Água
4	Máquinas e equipamentos para produção de pequena escala	Produção
5	Redução e mitigação de impactos ambientais das atividades	Impactos ambientais
6	Automação de processos produtivos agropecuário e pesqueiro	Automação
7	Energias renováveis aplicadas a produção de pequena escala	Energia renovável
8	Informática aplicada nos processos de produção	Produção
9	Agricultura familiar	Agricultura

Fonte: Autora

Para cada termo-chave foi recuperado o número de ideias selecionadas e não selecionadas. A Tabela 38 apresenta a sumarização dos resultados obtidos, bem como o percentual de ideias selecionadas por termo.

Tabela 38 – Sumarização dos resultados por termo-chave

Nº	Termo-chave	Nº de ideias selecionadas	Nº de ideias não selecionadas	Percentual de ideias selecionadas
1	Mão de obra	3	1	75,00%
2	Captação	2	1	66,66%
3	Água	10	3	76,92%
4	Produção	17	14	54,83%
5	Impactos ambientais	3	1	75,00%
6	Automação	4	1	80,00%
7	Energia renovável	2	0	100,00%
8	Agricultura	3	4	42,85%

Fonte: Autora

Ao termo mão de obra, relacionado à atividade 1, foram identificadas quatro ideias, sendo que três foram selecionadas. O termo energia renovável (atividade 7) está associado a duas ideias selecionadas. Ao termo agricultura (Atividade 9), foco do edital de fomento, foram identificadas sete ideias, e destas, três foram selecionadas. No total foram identificadas 54 ideias distintas, sendo 32 selecionadas (59%). Esta evidência é importante, pois está alinhada diretamente ao contexto organizacional, representado pelo edital de fomento.

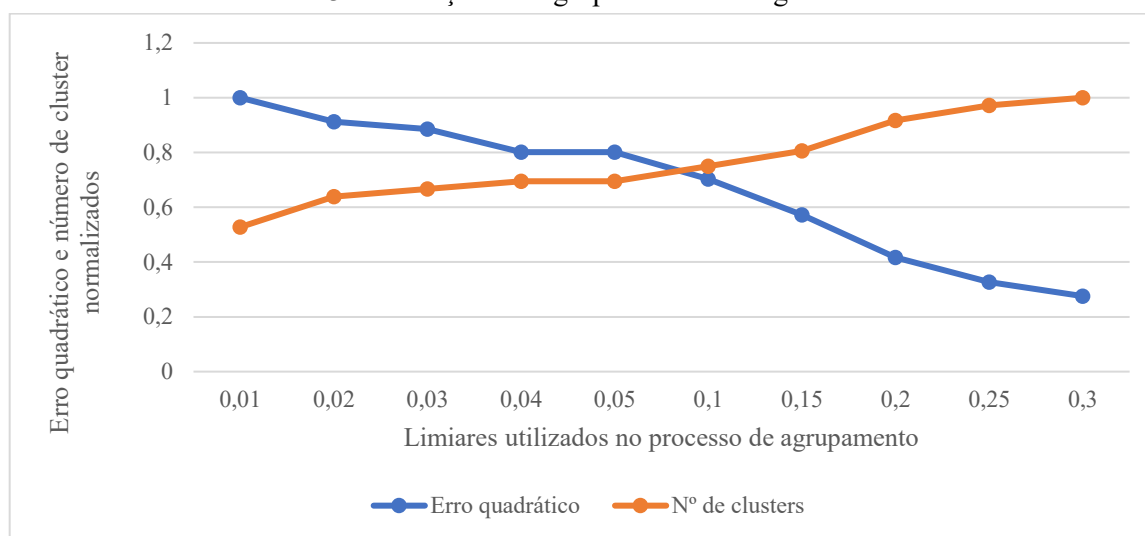
7.2.2 Segundo cenário de análise

O segundo cenário de análise é composto pelas ideias coletadas da campanha #MosquitoNão promovida pelo Ministério da Educação. Ao total, o documento final contém 93 ideias selecionadas. O objetivo desta análise é identificar as características presentes na formação dos agrupamentos, bem como os índices gerados pelo protótipo, independente do contexto organizacional.

Ressalta-se que neste cenário não será discutida a diferença entre ideias selecionadas e não selecionadas, pois a análise é composta por 93 ideias selecionadas. O objetivo é passar uma visão geral do cenário de análise e mostrar a generalidade do modelo proposto.

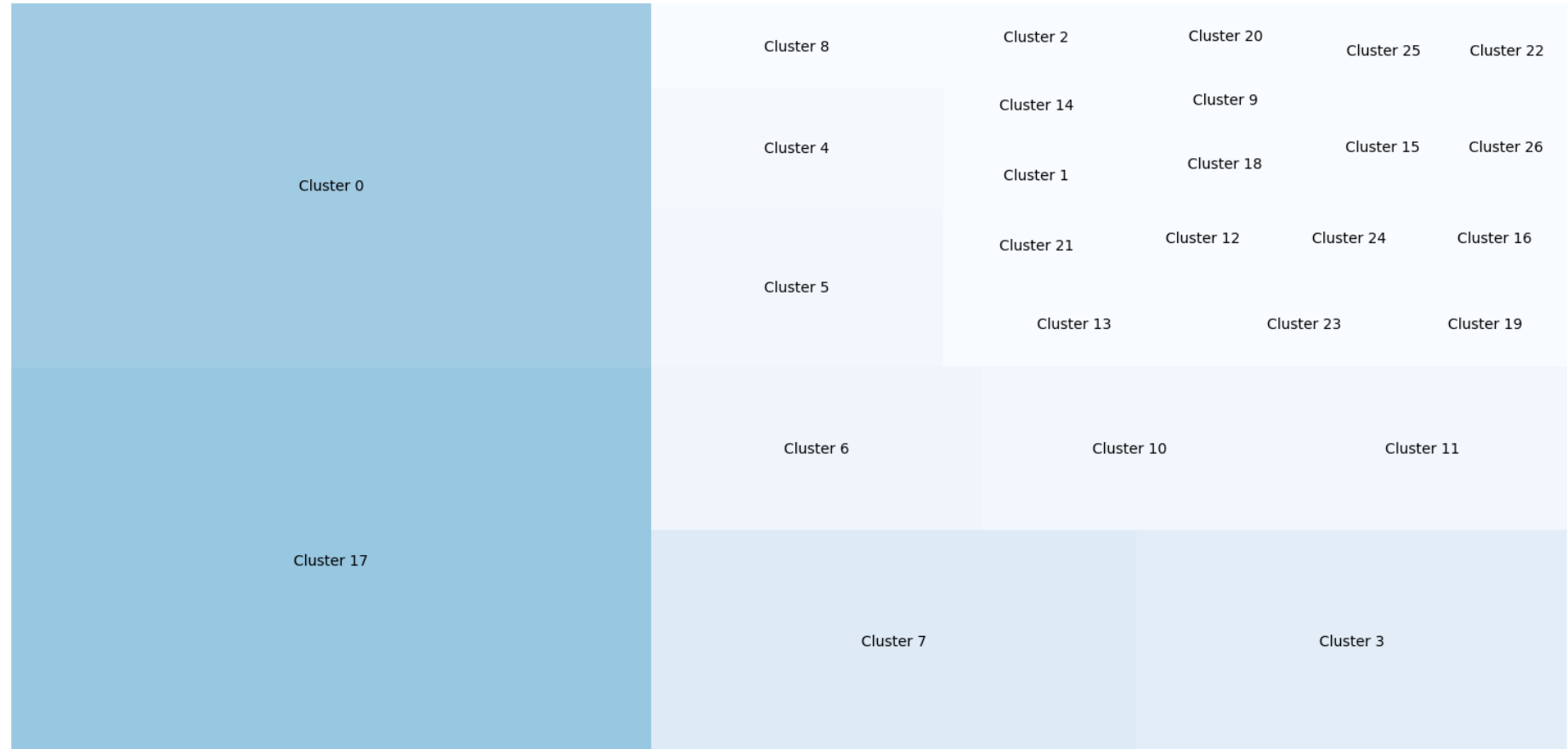
Foram geradas dez execuções do algoritmo de agrupamento de ideias. O fator de similaridade na formação dos grupos variou de 0,01 a 0,3, conforme apresentado no Gráfico 3.

Gráfico 3 - Formação dos agrupamentos no segundo cenário



Fonte: Autora

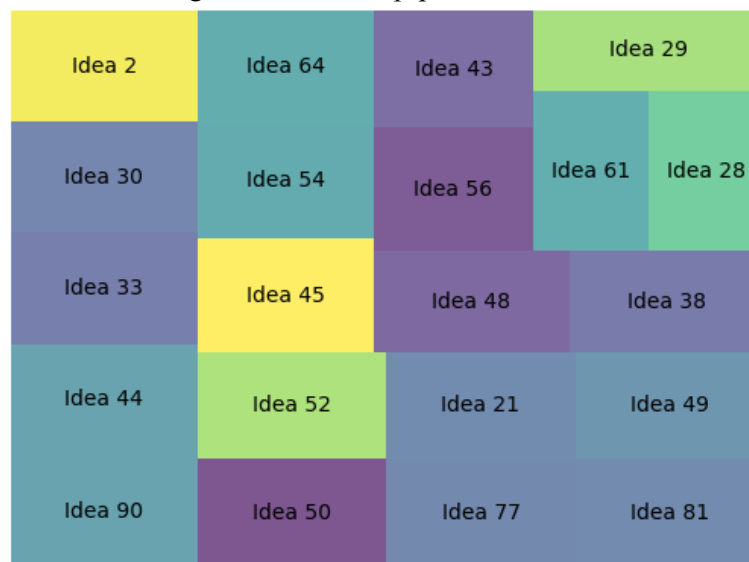
O Gráfico 3 apresenta no eixo x o limiar de similaridade utilizado e no eixo y a variação, tanto para o erro quadrático quanto para o número de agrupamentos formados em cada limiar. Ambas as dimensões foram normalizadas pelo máximo valor, tanto do erro quanto do número de agrupamentos. Neste sentido, a linha azul determina o erro quadrático calculado na formação dos agrupamentos e a linha laranja o número de *clusters* gerados. Analisando o gráfico, observa-se que no fator de similaridade 0,1 as duas linhas estão próximas, o que na análise do cenário indica um melhor resultado com a aproximação entre o número de *clusters* formados e o erro quadrático calculado, ambos normalizados. Neste cenário, foram gerados 27 *clusters* e o erro quadrático normalizado foi 0,703. A Figura 67 apresenta os principais *clusters* gerados, sendo a dimensão dos retângulos da *treemap* determinada pela soma dos índices de potencialidade de cada ideia pertencente a determinado agrupamento.

Figura 67 – Visualização dos agrupamentos através de *treemap* para o segundo cenário de análise

Fonte: Autora

Conforme a Figura 67, os agrupamentos com maiores índices de potencialidade foram os *clusters* 17, 0 e 7. A seguir serão apresentados os cinco *clusters* mais relevantes do segundo cenário de análise. O primeiro agrupamento com maior relevância no cenário de análise é o *cluster* 17, apresentado na Figura 68.

Figura 68 – Treemap para o *cluster* 17



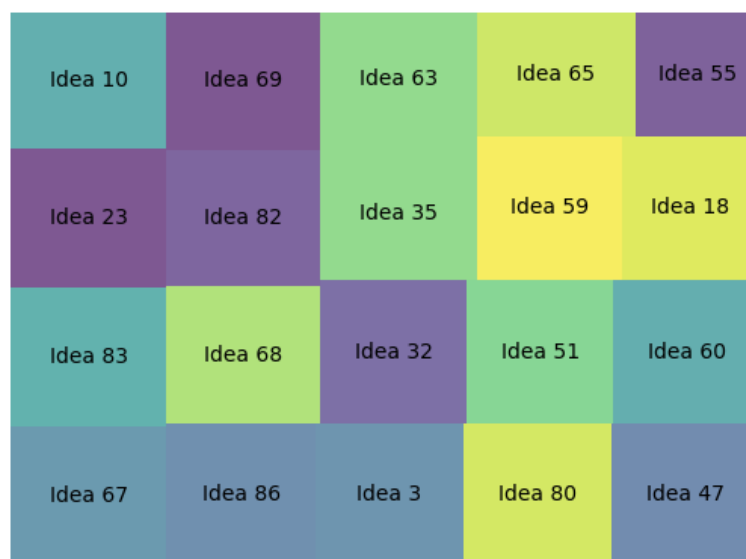
Fonte: Autora

As ideias com maior representatividade foram as que possuíam os identificadores 90, 44 e 33. A Tabela 39 apresenta um número sequencial, o identificador da ideia, o título e o índice de potencialidade.

Tabela 39 – Ideias que compõem o *cluster* 17

Nº	Id	Título	Índice de Potencialidade
1	90	União dos cursos de agronomia e técnico agropecuário no combate ao mosquito	0.498527959636863
2	44	Plano interno de combate ao mosquito para cada instituição	0.493410046596415
3	33	Dia do combate ao mosquito	0.493367436244904
4	30	Escola de samba contra o Zika	0.493320548460057
5	2	Borrifadores eletrônicos em praças públicas com repelentes naturais e sensores de movimento	0.492954638180595
6	50	Sistema de visitação em áreas de risco	0.488352351185327
7	77	Emprego de métodos legislativos para o controle da entrada e disseminação do mosquito <i>Aedes aegypti</i> nos estados e seus respectivos municípios	0.488070977061144
8	81	Fazer grupo de apoio às mães de bebês com microcefalia	0.486860406789373
9	52	Uso e distribuição da planta Lavanda (<i>Lavandula sp.</i>) como repelente natural no combate ao mosquito	0.480675196036826
10	21	Oficina de arte para reciclar materiais que possam virar possíveis criadores de mosquito	0.480548226716724
11	49	Cuidados com jarros de flores em túmulos	0.48051471696627

Figura 70 – Treemap para o cluster 0

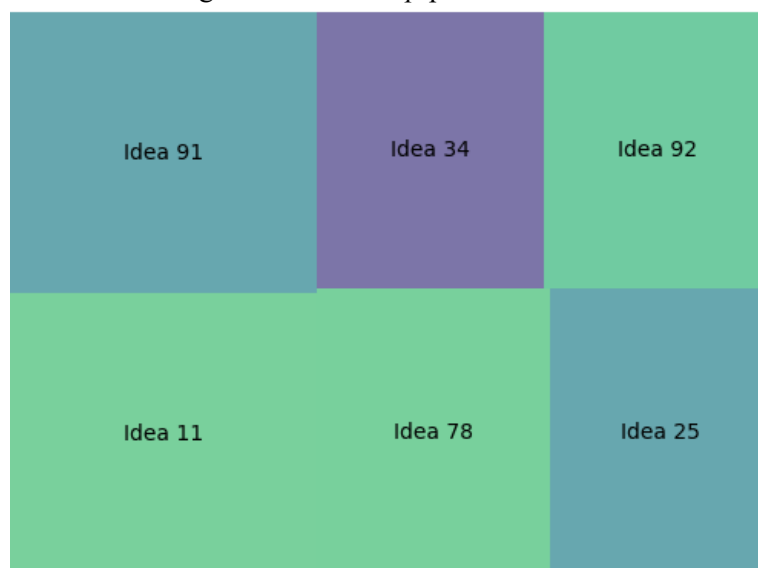


Fonte: Autora

É apresentada a Tabela 40 que tem como objetivo especificar as ideias que compõem o agrupamento.

Tabela 40 – Ideias que compõem o cluster 0

Nº	Id	Título	Índice de Potencialidade
1	67	Propor a construção de uma equipe multidisciplinar para angariar recursos no exterior para auxiliar em pesquisas vinculadas aos Institutos de pesquisa e divulgação desse conhecimento para a população	0.498166854273117
2	83	Concurso musical ZikaMusic	0.497759360632181
3	23	Curso gratuito para produção caseira de repelentes caseiros para a população menos favorecida	0.493376798879843
4	10	Mapeamento de focos	0.486476516804149
5	86	Fazer compostagem urbana	0.485912981670064
6	3	Uso de drones para apontar foco de poluição em lotes e casas de difícil acesso	0.484924901913737
7	80	Tudo que vai tem que voltar - logística reversa	0.483805681297258
8	47	Distribuição de kits de combate ao mosquito	0.482814016246802
9	68	Aumento de linhas de pesquisa relacionadas ao mosquito vírus e a microcefalia em centros de pesquisa	0.481248255163503
10	82	Produzir hortaliças folhosas em garrafas pet	0.480623912592125
11	69	Aumento do número de pesquisas aprovadas com recursos para pesquisa básica sobre o mosquito transmissor	0.480598633273358
12	32	Projeto orientações contra o mosquito Aedes aegypti	0.480505179156931
13	51	Construção de uma mosquiteira a fim de combater a proliferação do mosquito Aedes aegypti	0.480447924987573
14	60	Essência de cedro e alecrim no combate ao mosquito Aedes aegypti	0.479649290539863
15	35	Mutirão de combate ao mosquito	0.479355492742965
16	63	Repelente caseiro de citronela	0.477415649069326

Figura 72 – Treemap para o *cluster 7*

Fonte: Autora

A seguir é apresentada a Tabela 41 com a descrição das ideias pertencentes ao agrupamento.

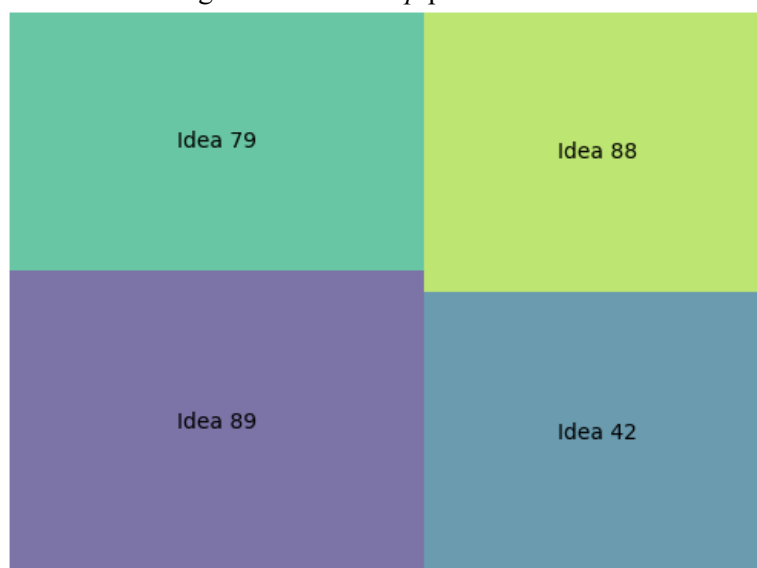
Tabela 41 – Ideias que compõem o *cluster 7*

Nº	Id	Título	Índice de Potencialidade
1	11	Criação de um aplicativo de celular que contenha informações de como combater o mosquito e como tratar os doentes	0.882717304458997
2	91	Criar rede de contato entre hospitais estaduais	0.882285754681142
3	78	Criar espaços destinados ao tratamento de bebês com microcefalia	0.679872529148396
4	25	Informações nas contas de água e luz sobre o combate ao mosquito	0.646550783465342
5	34	Trabalhar na feira de ciências em todas as escolas do seu município o tema #MosquitoNao	0.642641801948564
6	92	Eliminação dos focos de maior dispersão através do rastreamento de mosquitos transmissores	0.641221720623717
Soma dos Índices de Potencialidade			4.3752

Fonte: Autora

O terceiro *cluster* apresentado na Tabela 41 possui seis ideias, cuja soma dos índices das ideias foi 4,3752. A Figura 73 apresenta a nuvem de termos do *cluster 7*.

Figura 76 – Treemap para o cluster 6



Fonte: Autora

A Tabela 43 apresenta as ideias pertencentes ao agrupamento.

Tabela 43 – Ideias que compõem o cluster 6

Nº	Id	Título	Índice de Potencialidade
1	89	Parceria entre universitários da área de saúde e ambiental	0.652259674777843
2	79	Estagio contra o mosquito Aedes aegypti	0.549863105400228
3	42	Levar alunos para vivenciar situações de aplicação prática de conhecimentos adquiridos na teoria sobre o mosquito Aedes aegypti	0.506093817424863
4	88	Ampliar a oferta de saneamento básico a todas as comunidades	0.496478456851111
Soma dos Índices de Potencialidade			2.2046

Fonte: Autora

O cluster apresentado na Tabela 43 possui quatro ideias e corresponde ao cluster com a quinta maior relevância. A soma dos índices das ideias foi 2.2046. A Figura 77 apresenta a nuvem de termos do cluster 6.

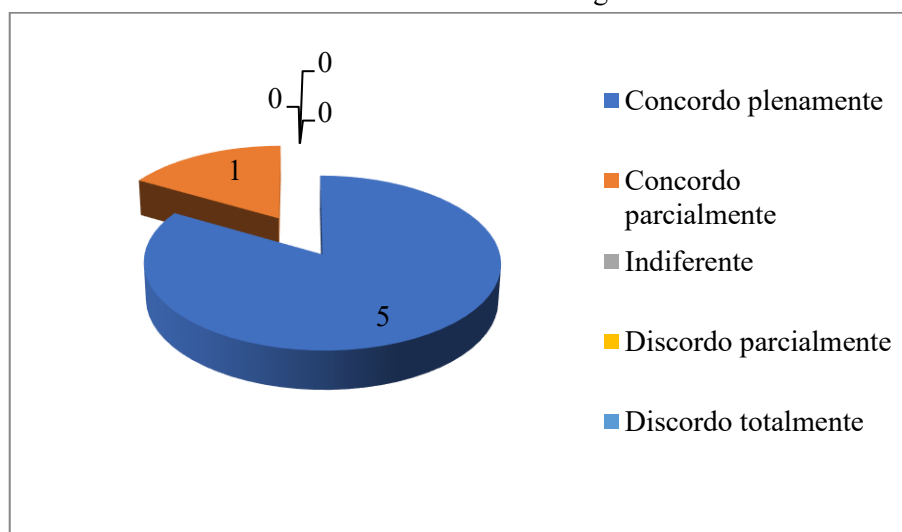
Vale reforçar que inicialmente foram solicitados dados de identificação, formação e área do conhecimento que os especialistas possuíam experiência.

7.3.1 Análise das Respostas às Perguntas da Dimensão 1

A primeira dimensão corresponde às questões envolvendo a pesquisa. Nesta dimensão foram apresentadas três questões.

Na questão 1 o respondente foi questionado se considerava a identificação de potenciais ideias com base no contexto organizacional, uma atividade importante no processo de tomada de decisões estratégicas. O Gráfico 1 apresenta as respostas.

Gráfico 4 – Dimensão 1 - Pergunta 1

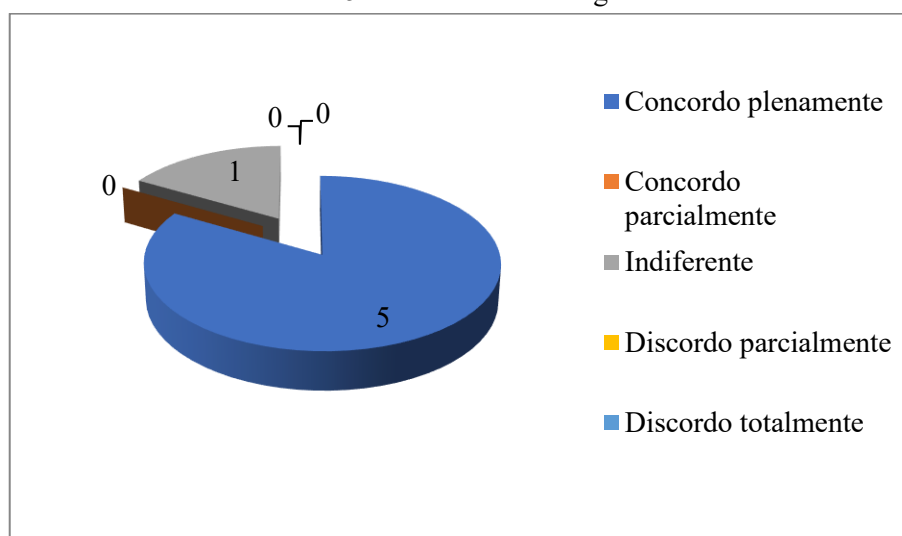


Fonte: Autora

Observa-se que 5 (cinco) respondentes concordam plenamente que a identificação de potenciais ideias relacionadas ao contexto organizacional corresponde a uma atividade importante no processo de tomada de decisão.

Na questão 2 o respondente foi questionado se considerava que a análise e elucidação de potenciais ideias podiam trazer subsídios para a tomada de decisão. O Gráfico 5 ilustra os resultados.

Gráfico 5 – Dimensão 1 - Pergunta 2

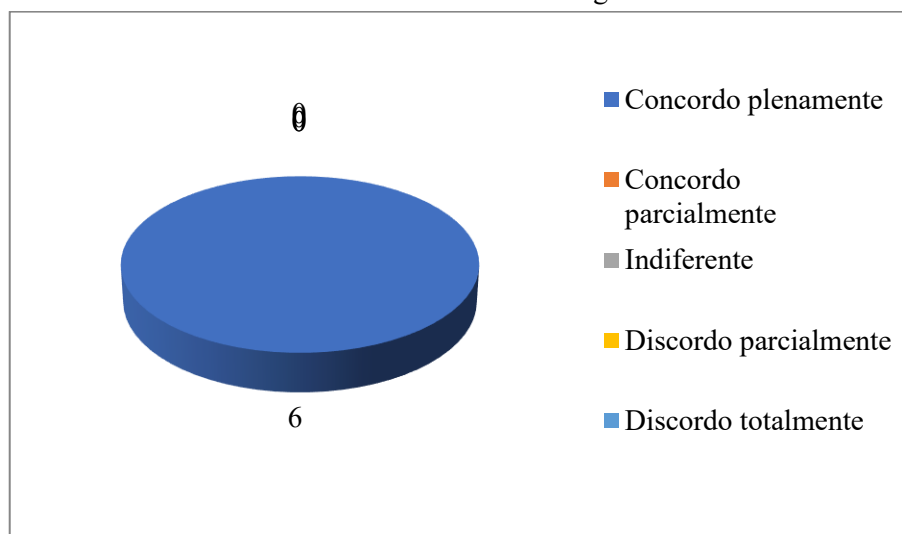


Fonte: Autora

Dos respondentes, cinco concordam totalmente que a análise e elucidação de potenciais ideias possam trazer subsídios para a tomada de decisão. Apenas um respondente posicionou-se indiferente, contudo, o entrevistado parece não ter compreendido a pergunta em virtude do comentário que expôs.

Na pergunta três os especialistas foram questionados em relação aos dados. Antes da pergunta propriamente dita existe um texto explicativo: ““Identificar informações relevantes sobre dados não estruturados como texto é um grande desafio, sendo necessários algoritmos confiáveis para encontrar ou gerar recursos (padrões) interpretáveis para analisar esses dados de forma significativa. Uma possibilidade consiste na geração de agrupamentos. O intuito na formação de agrupamentos a partir de informações textuais é separar vários documentos dispostos de forma desorganizada em um conjunto de grupos que contenham documentos com assuntos semelhantes. Isto permite localizar documentos rapidamente em uma dada coleção. A fim de fornecer uma representação mais significativa dos dados textuais e melhorar a eficiência dos algoritmos de análise textuais, a Análise de Agrupamento foi utilizada. Nesse caso, o processo de identificação de potenciais ideias com base no contexto organizacional é modelado de forma a permitir o agrupamento de ideias semelhantes. Você considera que essa abordagem complementa as possibilidades de identificação de potenciais ideias?” O Gráfico 6 apresenta os resultados.

Gráfico 6 – Dimensão 1 - Pergunta 3



Fonte: Autora

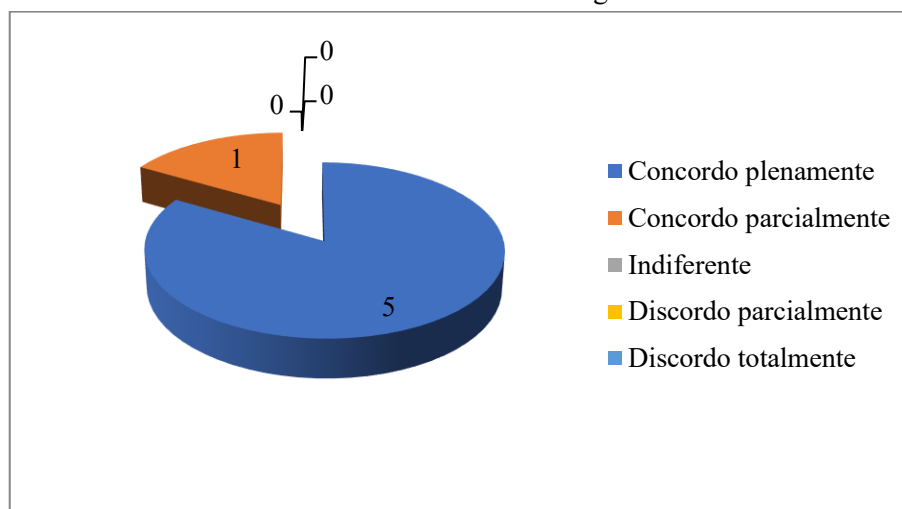
Todos os respondentes acreditam que a abordagem proposta pela tese poderá contribuir na identificação de potenciais ideias.

7.3.2 Análise das Respostas às Perguntas da Dimensão 2

A segunda dimensão trata sobre o documento de apresentação da ontologia proposta e os procedimentos para sua operacionalização, abordando a aplicabilidade e dimensão da pesquisa.

A pergunta 1 questionou: “A modelagem e representação de conhecimentos envolvidos no processo de identificação de potenciais ideias com base no contexto organizacional, propostos e definidos neste modelo, auxiliam na tomada de decisão pela organização”? O Gráfico 7 apresenta as respostas obtidas.

Gráfico 7 – Dimensão 2 - Pergunta 1

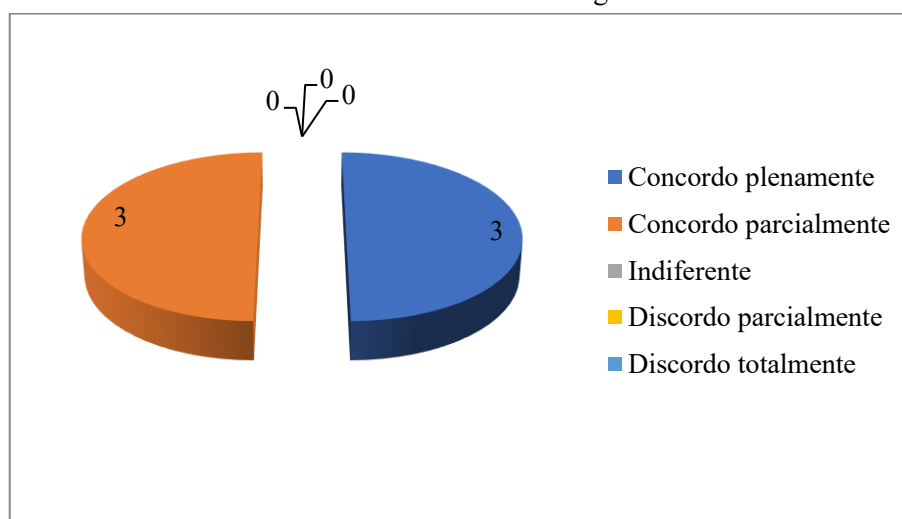


Fonte: Autora

Todos os respondentes concordam plenamente ou parcialmente que a modelagem e representação de conhecimento proposta pelo trabalho auxiliam no processo de tomada de decisão pela organização.

A pergunta dois questionou se os entrevistados consideravam os conceitos (representados pelas classes) e relacionamentos apresentados nesta representação suficientes para permitir que vários aspectos do processo de elucidação de potenciais ideias fossem explorados. O Gráfico 8 apresenta os resultados.

Gráfico 8 – Dimensão 2 - Pergunta 2

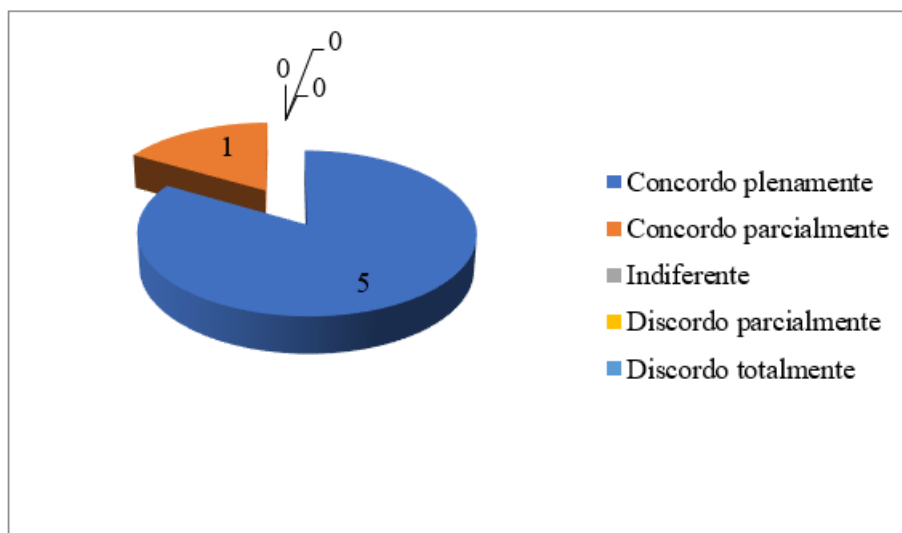


Fonte: Autora

Todos os respondentes concordam plenamente ou parcialmente que a modelagem dos conceitos e relacionamentos apresentados nesta proposta são suficientes para elucidar e explorar potenciais ideias. Contudo, acreditava-se que a ontologia poderia evoluir, pois trata-se de um trabalho que estava em andamento no momento em que foi realizada a avaliação.

Na pergunta 3 os especialistas responderam se: “as perguntas de competências são suficientes e cobrem os aspectos mais importantes relacionados à identificação de potenciais ideias alinhadas ao contexto organizacional”? O Gráfico 9 ilustra os dados obtidos.

Gráfico 9 – Dimensão 2 – Pergunta 3

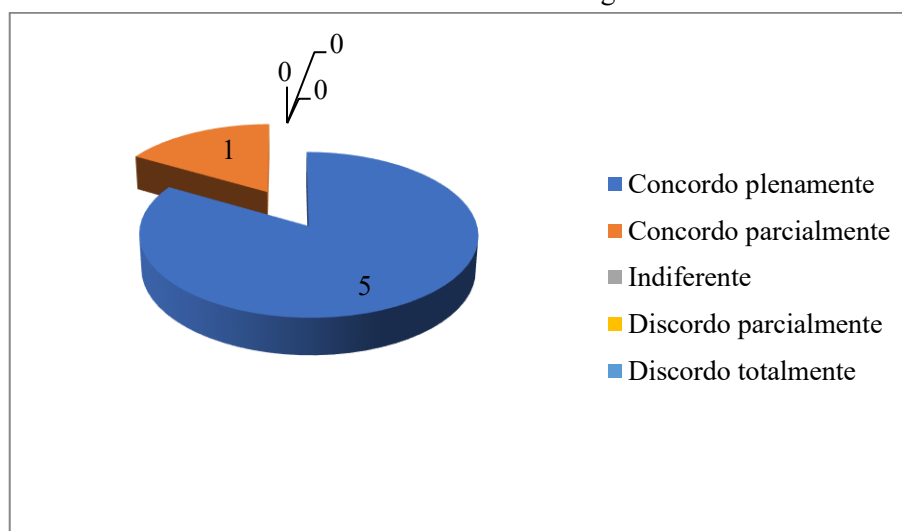


Fonte: Autora

Dos especialistas, cinco acreditam que as perguntas de competência cobrem os aspectos mais importantes relacionados à identificação de potenciais ideias.

A pergunta 4 questionou os entrevistados a respeito do modelo baseado em conhecimento para identificação de potenciais ideias (vide Figura 33): “A partir do seu entendimento a respeito dos elementos que compõem o modelo baseado em conhecimento, você considera que o mesmo pode contribuir significativamente no processo de tomada de decisão”? O Gráfico 10 apresenta as respostas obtidas.

Gráfico 10 – Dimensão 2 - Pergunta 4



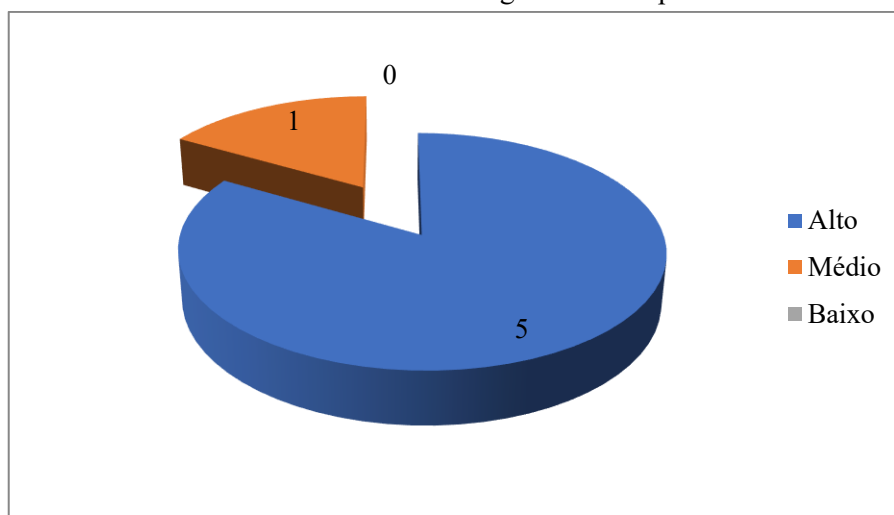
Fonte: Autora

Vale destacar que todos os respondentes concordam plenamente ou parcialmente que o modelo proposto pode contribuir significativamente no processo de tomada de decisão.

7.3.3 Análise das Respostas às perguntas da Dimensão 3

Na terceira dimensão, os especialistas avaliaram os resultados das perguntas de competência, apresentadas no escopo da ontologia (subseção 4.1.1), em relação ao critério “grau de contribuição” no processo de identificação de potenciais ideias alinhadas ao contexto organizacional. Foi escolhida uma entre as três opções: alto, médio, baixo. Na pergunta de competência 1 os especialistas foram questionados sobre: “quais são as potenciais ideias de acordo com o valor de investimento”? O Gráfico 11 ilustra as respostas.

Gráfico 11 – Dimensão 3 - Pergunta de competência 1

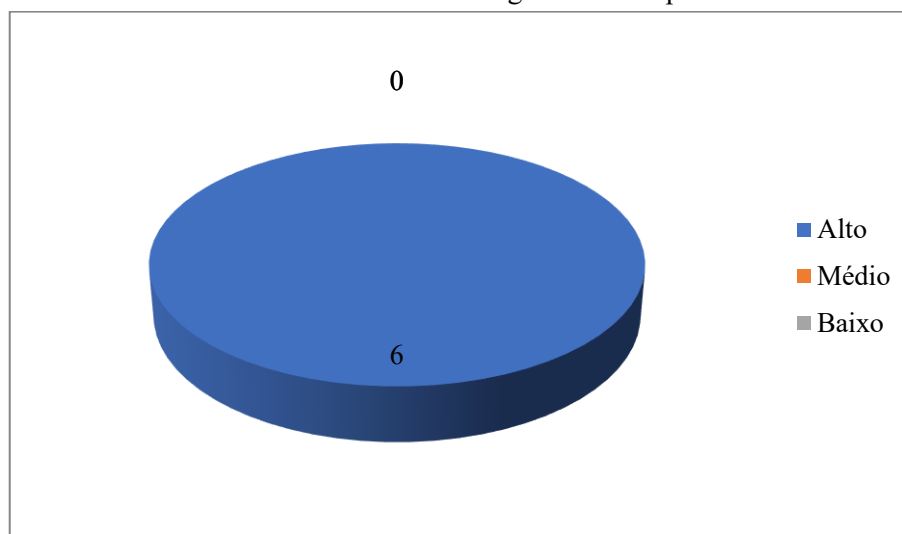


Fonte: Autora

Dos especialistas participantes da pesquisa, cinco acreditam que o valor de investimento pode influenciar na seleção de potenciais ideias.

Na pergunta de competência 2 dois os especialistas foram questionados sobre: “quais são as potenciais ideias de acordo com o valor de investimento e áreas de investimentos potenciais”? O Gráfico 12 expõe os dados obtidos.

Gráfico 12 – Dimensão 3 – Pergunta de competência 2

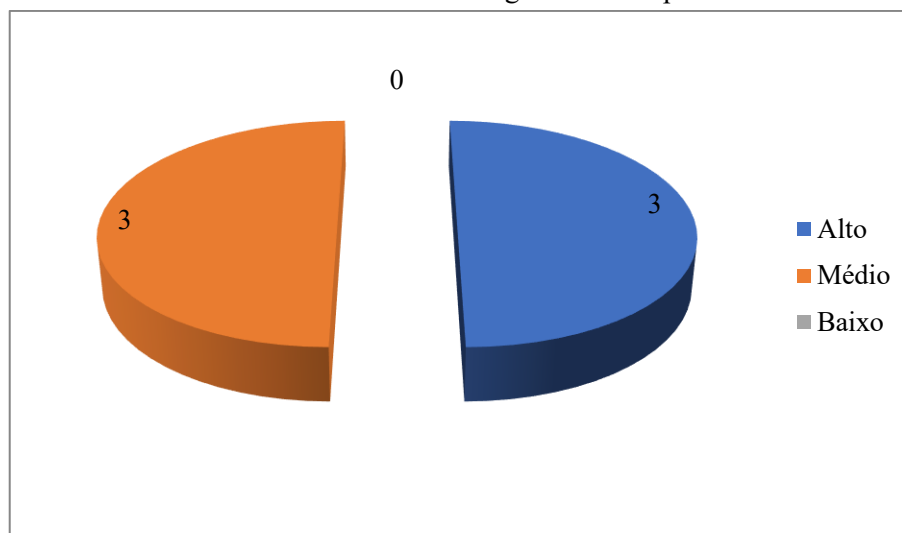


Fonte: Autora

Observa-se que todos os especialistas avaliaram como nível alto os resultados da pergunta de competência 2 em relação ao critério “grau de contribuição”.

A pergunta de competência 3 é: “quais são as potenciais ideias de acordo com o índice de aderência ao contexto organizacional”? O Gráfico 13 ilustra os resultados.

Gráfico 13 – Dimensão 3 - Pergunta de competência 3

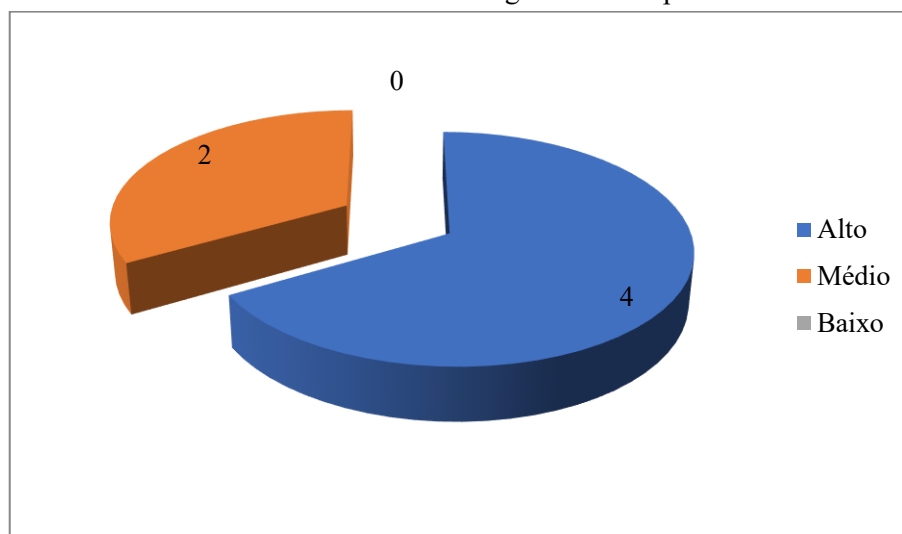


Fonte: Autora

Dos especialistas, três responderem ao grau de contribuição da pergunta de competência 3 como médio. Contudo, a combinação dos índices poderá trazer melhores resultados na avaliação das ideias.

Na pergunta de competência 4 os especialistas foram questionados sobre: “quais são as potenciais ideias de acordo com o índice de originalidade e o valor de investimento”? O Gráfico 14 apresenta as respostas obtidas.

Gráfico 14 – Dimensão 3 - Pergunta de competência 4

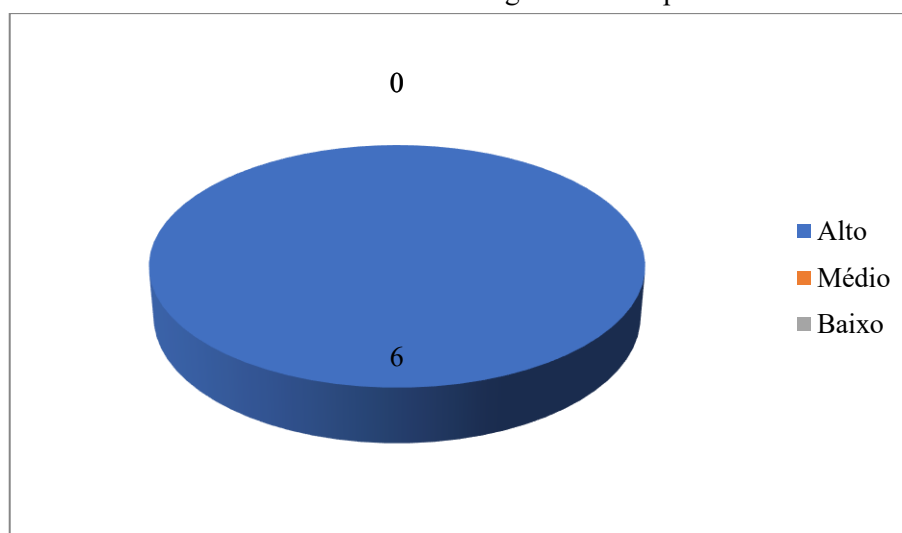


Fonte: Autora

Dos especialistas participantes da pesquisa, 4 acreditam que o índice de originalidade e o valor de investimento podem ser um diferencial na seleção de potenciais ideias.

Na pergunta de competência 5, os especialistas foram questionados sobre: “quais são as potenciais ideias de acordo com o índice de originalidade”? O Gráfico 15 expõe os resultados.

Gráfico 15 – Dimensão 3 - Pergunta de competência 5

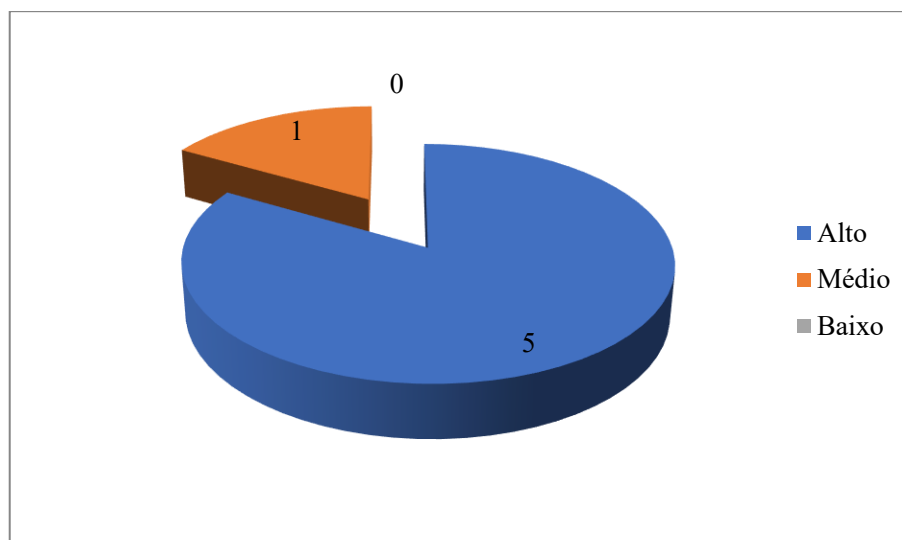


Fonte: Autora

Observa-se que todos os especialistas atribuíram como grau de contribuição nível alto à pergunta de competência 5.

A pergunta de competência 6 questionou: “Quais são as potenciais ideias de acordo com as áreas de investimento potenciais da organização”? O Gráfico 16 ilustra a visão dos especialistas.

Gráfico 16 – Dimensão 3 - Pergunta de competência 6

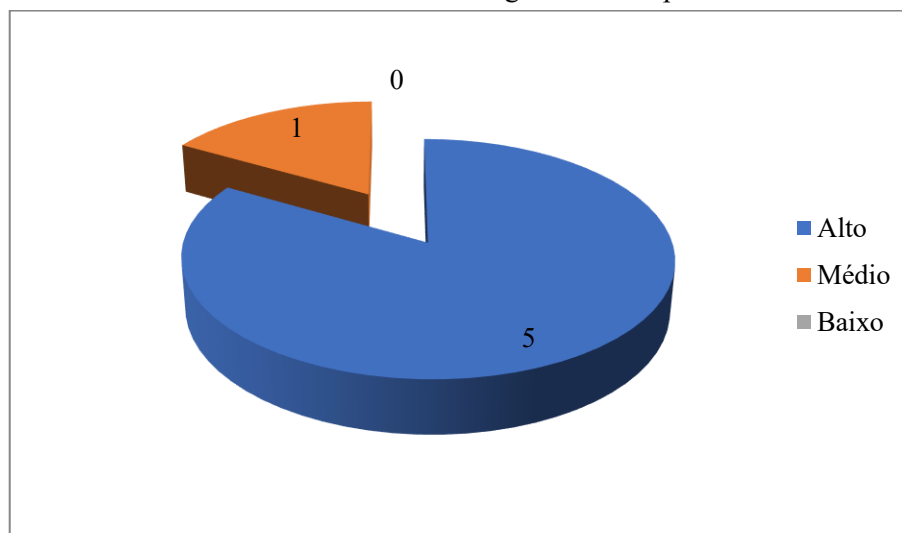


Fonte: Autora

Do total de especialistas, cinco atribuíram como nível alto o grau de contribuição da pergunta.

A pergunta de competência 7 questionou: “quais são as potenciais ideias de acordo com o índice de tecnologia”? O Gráfico 17 apresenta os resultados obtidos.

Gráfico 17 – Dimensão 3 - Pergunta de competência 7

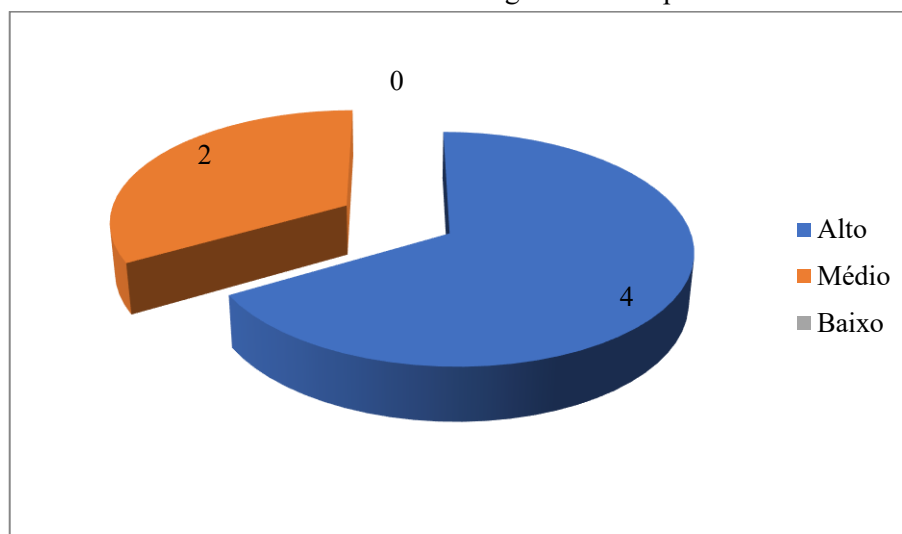


Fonte: Autora

Destaca-se que cinco especialistas atribuíram grau de contribuição nível alto à pergunta de competência 7.

Na pergunta de competência 8 os entrevistados foram questionados sobre: “quais são as ideias semelhantes”? O Gráfico 18 ilustra as respostas.

Gráfico 18 – Dimensão 3 - Pergunta de competência 8

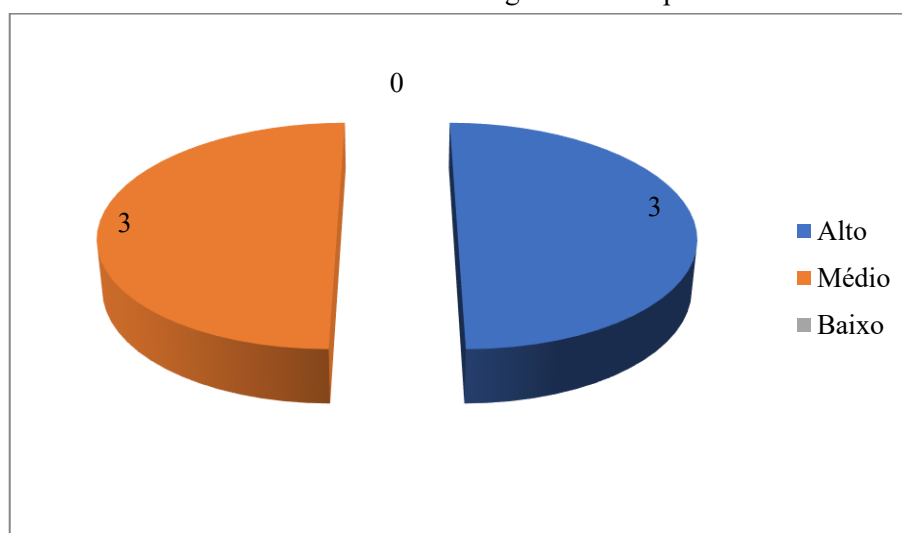


Fonte: Autora

Dos especialistas participantes da pesquisa, quatro acreditam que identificar as ideias semelhantes possa contribuir na seleção de potenciais ideias. Ao identificar ideias semelhantes à organização, pode-se investir em grupos de ideias e mais colaboradores/idealizadores podem ser contemplados. Isto pode inclusive gerar equipes de trabalho mais engajadas, uma vez que mais ideias, em função da similaridade, podem ser contempladas.

A pergunta de competência 9 questionou sobre: “quais são as ideias pertencentes à determinada categoria”? O Gráfico 19 ilustra os resultados.

Gráfico 19 – Dimensão 3 - Pergunta de competência 9

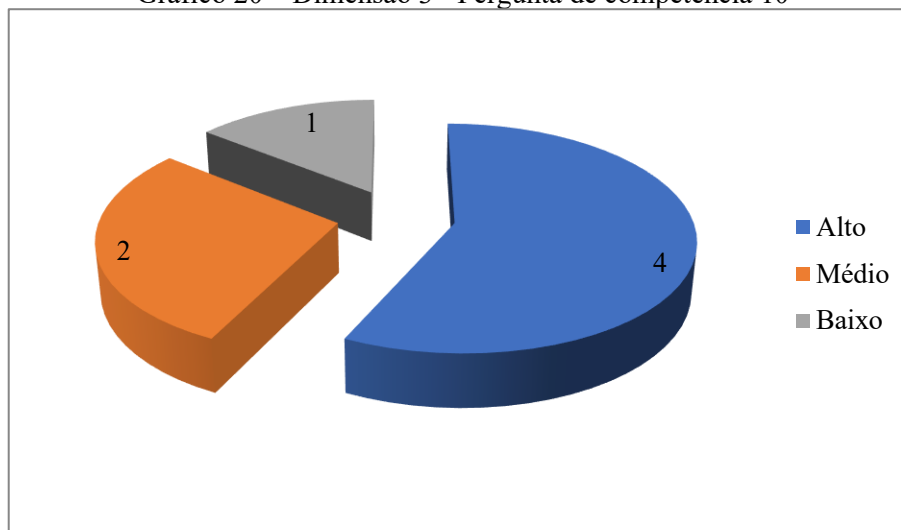


Fonte: Autora

Apenas três especialistas atribuíram grau de contribuição nível alto à pergunta de competência 9. Cabe ressaltar que o item categoria pode ser utilizado pela organização no momento de selecionar as ideias.

Na pergunta de competência número 10 os especialistas foram questionados sobre: “quais são as ideias que receberam maior número de votos”? O Gráfico 20 apresenta os resultados obtidos.

Gráfico 20 – Dimensão 3 - Pergunta de competência 10

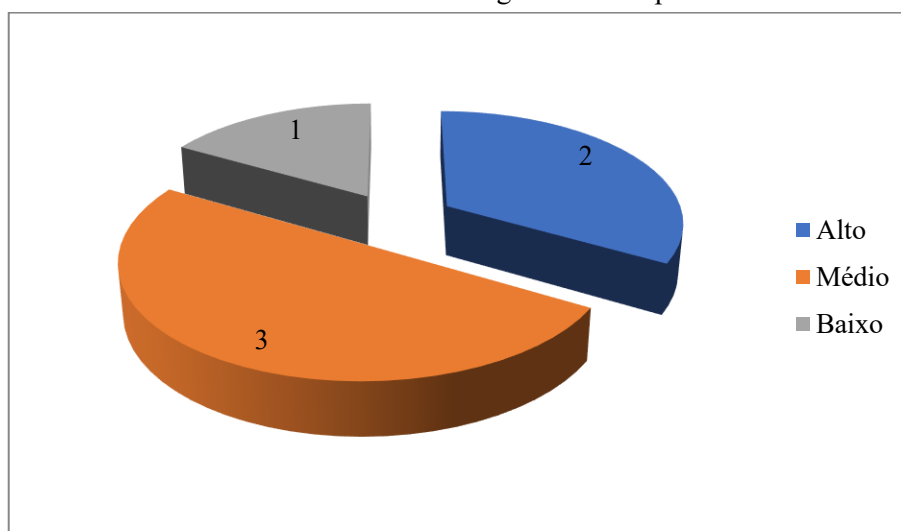


Fonte: Autora

Dos especialistas participantes da pesquisa, quatro acreditam que o número de votos que a ideia recebeu possa contribuir na seleção de potenciais ideias.

A pergunta de competência de número 11 questionou sobre: “quais são as ideias mais recentes”? O Gráfico 21 expõe os dados.

Gráfico 21 – Dimensão 3 - Pergunta de competência 11

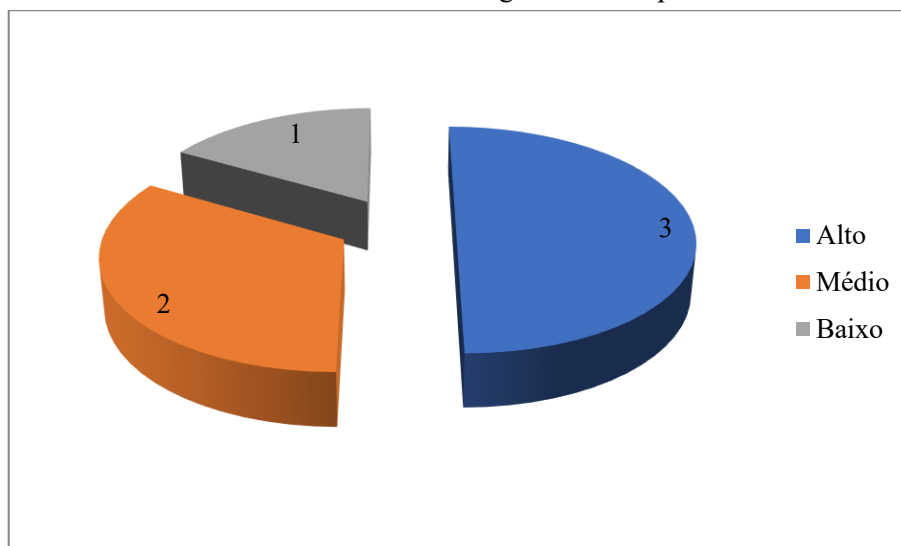


Fonte: Autora

Apenas três especialistas atribuíram grau de contribuição nível alto à pergunta de competência 11. Contudo, esta característica pode ser combinada com outras perguntas de competência para auxiliar especialistas na tomada de decisão.

Na pergunta de competência 12 os especialistas foram questionados sobre: “quais são as ideias que possuem maior valor estimado de investimento”? O Gráfico 22 apresenta as respostas dos especialistas.

Gráfico 22 – Dimensão 3 – Pergunta de competência 12

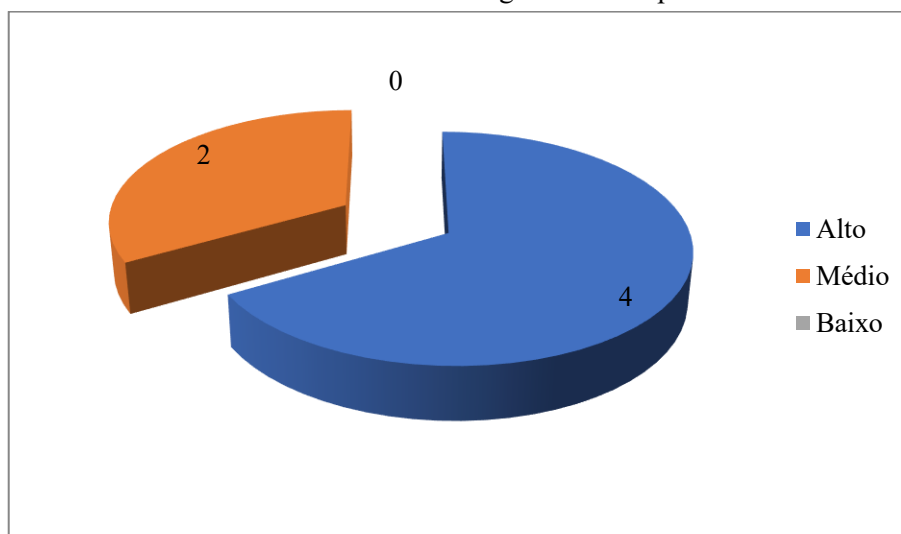


Fonte: Autora

Dos entrevistados, cinco acreditam em nível alto ou médio no grau de contribuição da pergunta 12. Apenas um especialista atribuiu nível baixo.

A pergunta de competência 13 questionou sobre: “quais são as potenciais ideias de acordo com o índice de atualidade”? O Gráfico 23 apresenta as respostas dos especialistas.

Gráfico 23 – Dimensão 3 – Pergunta de competência 13

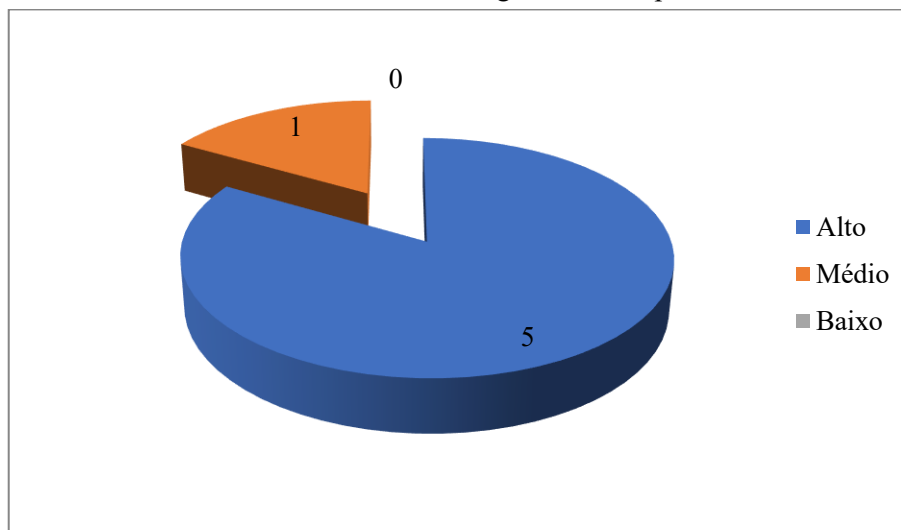


Fonte: Autora

Todos os especialistas envolvidos na pesquisa atribuíram grau de contribuição nível alto ou médio à pergunta de competência número 13.

A pergunta de competência 14 questionou os especialistas a respeito de: “quais são as ideias pertencentes a determinado grau de maturidade”? O grau de maturidade da ideia representa o nível que ela se encontra, por exemplo, no nível de ideia, protótipo finalizado, entre outros. O Gráfico 24 ilustra as respostas.

Gráfico 24 – Dimensão 3 - Pergunta de competência 14

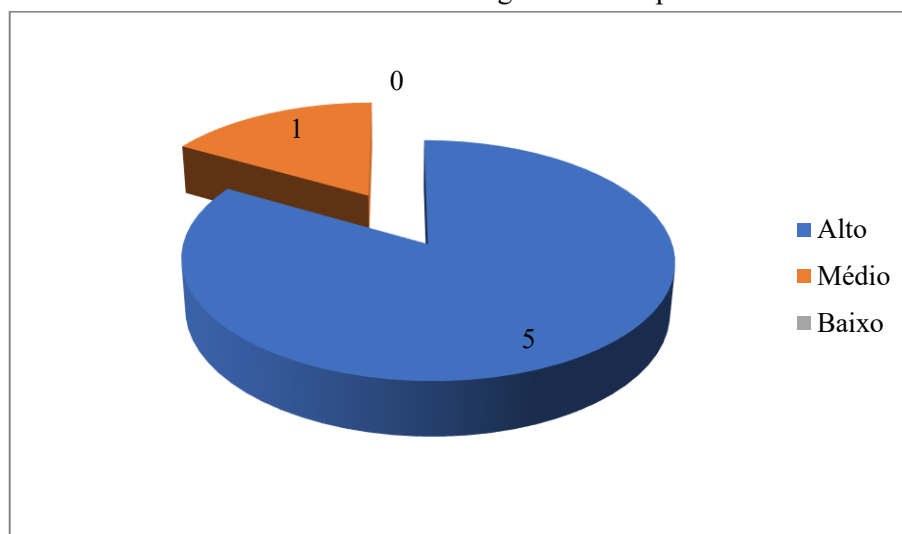


Fonte: Autora

Dos especialistas, cinco atribuíram grau de contribuição nível alto à pergunta de competência número 14. Cabe destacar que, de acordo com as estatísticas apresentadas no início do capítulo na Tabela 19, apenas 3% das ideias com grau de maturidade “Ideia” foram implementadas. Os avaliadores optaram por selecionar ideias que não estavam no estágio inicial e que demonstram algo prático, bem como viabilidade e lucratividade do negócio. Neste sentido, o grau de maturidade pode impactar na avaliação/recomendação da ideia.

A última pergunta de competência questionou os entrevistados sobre: “quais são as ideias em que a organização possui o número de colaboradores necessários para executá-las”?

Gráfico 25 – Dimensão 3 - Pergunta de competência 15



Fonte: Autora

Destaca-se que cinco especialistas participantes da pesquisa atribuíram nível alto ao grau de contribuição da pergunta de competência número 15.

Quanto às sugestões e comentários, os especialistas consideraram relevante a proposta de pesquisa. Os especialistas também destacaram que as perguntas de competências podem ser combinadas como, por exemplo: “quais são as ideias pertencentes à determinada categoria e com maior número de votos”? Outra combinação possível seria: “quais são as ideias mais recentes e que possuem maior valor estimado de investimento”?

7.4 COMPARAÇÃO COM MODELOS PREDITIVOS SUPERVISIONADOS

A análise preditiva tem como meta prever tendências (BABU; VASAVI; NAGARJUMA, 2017). Segundo Nyce (2007), este termo é considerado abrangente, pois engloba uma série de técnicas estatísticas e analíticas para o desenvolvimento de modelos capazes de prever acontecimentos futuros.

A análise preditiva, mais especificamente na área de mineração de dados e texto, se refere a um processo de análise de dados voltado a identificação de tendências, padrões ou relações entre os dados, podendo esta informação ser posteriormente utilizada para desenvolver modelos preditivos (NYCE, 2007). O princípio básico é visualizar os acontecimentos passados de maneira clara, tornar possível fazer previsões para o futuro, prevendo cenários e auxiliando as organizações a tomarem decisões estratégicas que alcancem resultados positivos.

Segundo os autores Mishra e Nishchol (2012), há dois tipos de modelos preditivos: os supervisionados e os não supervisionados. No modelo supervisionado é realizada uma busca

por padrões e relações entre os dados (MISHRA; NISHCHOL, 2012). Essa relação possibilita identificar um nível de acerto. Por outro lado, o modelo não supervisionado concentra-se em explorar as estruturas intrínsecas, correlações, associações e padrões nos dados e, portanto, possui a capacidade de descobrir conhecimento potencialmente útil, mas anteriormente desconhecido (FAN et. al, 2018). Como o conjunto de dados não está rotulado, o sistema tenta classificar estes conjuntos agrupando os semelhantes em determinadas classes (RUSSEL; NORVIG, 2004).

Para realizar as previsões torna-se necessário estabelecer a matriz de confusão. A função desta matriz é analisar quantas vezes o modelo acertou as previsões, definida conforme apresenta a Figura 78.

Figura 78 - Matriz de confusão

	Predito positivo	Predito Negativo
Exemplos positivos	TP	FN
Exemplos negativos	FP	TN

Fonte: Autora

Conforme a Figura 78 TP significa os positivos verdadeiros, FN significa os falsos negativos, FP são os falsos positivos e TN representa os negativos verdadeiros. A partir da matriz de confusão pode-se estabelecer medidas, com o intuito de determinar a acurácia, a precisão (*precision*), a revocação (*recall*), entre outras.

A acurácia refere-se ao grau de efetividade do classificador, definida pela fórmula $R = \frac{TP + TN}{TP + FN + FP + FN}$. *Precision* refere-se à proporção de itens corretamente classificados, definida pela fórmula $Precision = \frac{TP}{TP + FP}$. E, *Recall* é a proporção classificada corretamente como positivo, definida pela fórmula $Recall = \frac{TP}{TP + FN}$.

Outra medida amplamente utilizada é o coeficiente de concordância de Kappa, proposto por Jacob Cohen, em 1960. O coeficiente é uma medida robusta que leva em consideração os acertos e erros apresentados pela Matriz de Confusão (MORAES; MACHADO, 2014). Este coeficiente é comumente utilizado na etapa de avaliação do treinamento em processo de aprendizado de máquina (DUDA; HART; STORK, 2000). A partir dele é avaliado a concordância das avaliações em relação a um cenário de análise. De acordo com Landis e Koch (1977), o coeficiente de Kappa pode ser interpretado através da Tabela 44.

Tabela 44 – Análise do coeficiente de Kappa

Coeficiente de Kappa	Grau de concordância
< 0,00	Concordância pobre
0,00 – 0,20	Concordância pequena
0,21 – 0,40	Concordância regular
0,42 – 0,60	Concordância moderada
0,61 – 0,80	Concordância considerável
0,81 – 1,00	Concordância quase perfeita

Fonte: Landis e Koch (1977)

Após esta breve introdução, a presente tese busca realizar testes considerando o primeiro cenário de análise contendo as 200 ideias coletadas do site Sinapse da Inovação®. Assim, promove subsídios para uma comparação entre o modelo proposto, mais especificamente, a etapa de agrupamento de ideias, em relação à classificação de ideias.

Para esta etapa, as análises foram realizadas por meio do *software LightSide* em que as configurações padrões de *unigrams* e *cross validation* foram utilizadas, conforme mencionado no Capítulo 3. *Unigrams* significa um *n*-grama, ou seja, uma sequência contígua de *n* itens de uma determinada amostra de texto ou fala, neste caso um *token*. Já a *cross validation* é uma técnica para avaliar a capacidade de generalização de um modelo, a partir de um conjunto de dados. Esta técnica é definida como um método estatístico, usada para estimar a habilidade dos modelos de aprendizado de máquina. É amplamente empregada em problemas em que o objetivo da modelagem é a predição. Foi utilizado o valor de *cross validation* igual a 10. Para tanto, foi aplicado dois algoritmos amplamente usados na área de Aprendizado de Máquina, *Naive Bayes* e *Support Vector Machine*.

O algoritmo *Naive Bayes* atingiu 0,485 de acurácia e, após realizar as predições, foi possível analisar a matriz de confusão. A Tabela 45 apresenta a matriz de confusão gerada pelo algoritmo *Naive Bayes*.

Tabela 45 – Matriz de confusão do algoritmo *Naive Bayes*

Act/Pred	Não	Sim
Não	38	58
Sim	45	59

Fonte: Autora

Na Tabela 45 é possível visualizar que a *Precision*, proporção de itens corretamente classificados, foi de 45,7% e *Recall*, a proporção classificada corretamente foi de 39,5%. O algoritmo *Naive Bayes* atingiu o coeficiente do índice Kappa de 0,031, indicando um pequeno grau de concordância.

Também se aplicou o algoritmo *Support Vector Machine* no primeiro cenário de análise. Este algoritmo atingiu 0,540 de acurácia. A Tabela 46 apresenta a matriz de confusão gerada pelo algoritmo.

Tabela 46 – Matriz de confusão do algoritmo *Support Vector Machine*

Act/Pred	Não	Sim
Não	54	48
Sim	50	54

Fonte: Autora

Na Tabela 46 é possível visualizar que a proporção de itens corretamente classificados foi de 51,9%, e a proporção classificada corretamente como positivo foi de 52,9%. O coeficiente do índice Kappa do algoritmo *Support Vector Machine* atingiu 0,081, ficando próximo de zero. Ele aponta também um pequeno grau de concordância.

Em ambos os testes, o índice Kappa ficou próximo de zero, indicando que não existe concordância na classificação além do que seria esperado pelo acaso.

7.5 DISCUSSÃO DOS RESULTADOS

No primeiro cenário de estudo foram avaliadas 200 ideias coletadas no site Sinapse da Inovação®. Entre as 200 ideias analisadas pelo modelo proposto, apenas 100 ideias foram selecionadas pelos especialistas do Sinapse® e receberam financiamento. É importante destacar que as 200 ideias representam a etapa semifinal de seleção do programa, que recebeu ao todo 1.791 ideias. Ou seja, as ideias analisadas pelo protótipo já passaram por um processo de pré-seleção e podem ser consideradas originais e inéditas.

Após a inserção das ideias na base de conhecimento, representada pela ontologia de domínio, o contexto organizacional foi descrito por intermédio das informações do edital de

fomento. Com as devidas informações já preenchidas na base, iniciou-se o pré-processamento das ideias. Ele representa uma das fases mais importantes do modelo, pois é responsável por executar as consultas em fontes externas e organizacionais para estabelecer os índices referentes às ideias. Ao final, a média destes índices determina o índice de potencialidade das ideias.

Observou-se que o índice de potencialidade foi o que apresentou maior assertividade, ao comparar as ideias selecionadas pelo modelo com as selecionadas pelos especialistas do Sinapse[®], atingindo aproximadamente 70% de acertos. Esta informação é significativa, pois indica que, ao se analisar somente os índices individualmente, os resultados não são satisfatórios. Por exemplo, ao examinar a classificação apenas pelo índice de originalidade, a taxa de acerto ficou abaixo de 60%.

A próxima etapa do modelo envolveu a transformação matricial com SVD, possibilitando na aplicação do modelo a integração dos termos-chave que representam as ideias e os índices propostos nesta tese. Por conseguinte, a etapa de mineração de texto foi responsável por reunir as ideias semelhantes no mesmo agrupamento, utilizando uma versão modificada do algoritmo *k-means* baseada no conceito de similaridade vetorial. O grande objetivo foi contribuir na visualização das ideias ao apresentar os agrupamentos mais relevantes.

A partir dos grupos formados foram elencados os dez agrupamentos mais relevantes. Foi verificado que no primeiro agrupamento se atingiu uma assertividade de aproximadamente 65% (nove ideias selecionadas do total de 14), quando comparadas às ideias selecionadas pelo modelo com as escolhidas pelos especialistas do Sinapse da Inovação[®]. No segundo agrupamento a assertividade foi de 75% (nove ideias selecionadas do total de 12). A soma do número de ideias dos dois agrupamentos mais relevantes foi 26, e o percentual de acerto conjunto foi de 70%, totalizando 18 ideias selecionadas tanto pelo modelo quanto pelos especialistas nos dois primeiros agrupamentos.

Ao se analisar os dez agrupamentos mais significativos, o algoritmo atingiu uma taxa de assertividade de 80%, quando considerados os grupos onde o número de ideias selecionadas foi maior ou houve a ocorrência de empate entre selecionadas e não selecionadas. Um ponto fundamental desta análise é a possibilidade dos tomadores de decisão estudarem a semântica e as características do grupo, bem como a formação dos índices e, deste modo, obter mais informações que os auxiliem na tomada de decisão. Permite assim, ordenar decrescentemente as ideias pertencentes a um grupo pelo índice de potencialidade. Logo, pode-se estabelecer um *ranking* de classificação das ideias por agrupamento, facilitando a análise e visualização das ideias.

Outra possibilidade de análise é recuperar todas as ideias ordenadas de maneira decrescentemente pelo índice de potencialidade, independente dos grupos formados. A taxa de assertividade do modelo, para este caso, foi de aproximadamente 70% como mencionado anteriormente.

Adicionalmente, o modelo incluiu outras etapas para que os objetivos do trabalho fossem alcançados, entre elas, a explicitação do conhecimento de domínio que fornece os subsídios para a etapa de avaliação. A explicitação do conhecimento e do fluxo de informações ocorreu através do uso de técnicas de visualização de informação, entre elas, nuvem de termos, *treemaps* e mapas de tópicos.

Nesta perspectiva, a análise de redes sociais contribuiu na explicitação do conhecimento no primeiro cenário avaliado. A formação da rede foi composta por 82 ideias com limiar 0,7. Observou-se que, ao formar o *ranking* das ideias ordenadas pelo índice de potencialidade, os seguintes valores foram obtidos: entre as 10 primeiras ideias, oito foram selecionadas (80%); entre as 20 primeiras ideias, 14 foram selecionadas (70%). Esta análise teve como objetivo realizar uma apreciação visual do cenário de estudo de modo a demonstrar a relevância da análise de redes na etapa final do modelo proposto.

Pode-se afirmar ainda que, ao explorar-se o fluxo das informações e os relacionamentos existentes na formação da rede, as sugestões de ideias estavam aderentes às temáticas estratégicas do contexto organizacional, representadas pelo edital de fomento. Para cada termo-chave, conforme a Tabela 38, se recuperou o número de ideias selecionadas e não selecionadas. Deste modo, se evidencia que a partir das atividades de impacto, as propostas estão alinhadas diretamente às temáticas estratégicas relacionadas ao contexto organizacional.

Para verificar a efetividade do modelo foram aplicados testes supervisionados ao primeiro cenário. Eles envolveram a utilização de dois algoritmos de classificação amplamente utilizados na área de Aprendizado de Máquina, o algoritmo *Naive Bayes* e o algoritmo *Support Vector Machine*. Os testes mostraram que a assertividade dos algoritmos foi baixa, e em ambos os testes o índice Kappa ficou próximo de zero, indicando que não existe concordância na classificação dos algoritmos além do que seria esperado ao acaso. Ou seja, não existe qualquer padrão na classificação que possa ser utilizado para alcançar uma acurácia mais elevada. Os textos no formato como se encontram tornaria impossível descobrir um padrão para um próximo edital de avaliação de ideias, por exemplo, pois representam apenas ideias variadas. De modo geral, as ideias não possuem características que indiquem quais possuem ou não potencial de serem selecionadas e implementadas.

O modelo proposto também foi aplicado a um segundo cenário de análise visando demonstrar sua generalidade e aplicabilidade. O segundo cenário de análise foi composto por ideias coletadas da consulta pública promovida pelo Ministério da Educação denominada #MosquitoNão. O objetivo da análise era identificar as características presentes na formação dos agrupamentos, bem como os índices gerados pelo protótipo, independente do contexto organizacional. Por serem ideias relacionadas ao mesmo tema, referentes ao combate ao mosquito da dengue, pode-se observar que ambas eram muito próximas na descrição e os termos-chave mais citados eram comuns, como: *Aedes*, *Aegypti*, combate e mosquito. Como o cenário não foi composto por ideias selecionadas e não selecionadas, ou seja, a base de conhecimento foi preenchida apenas com as 93 ideias selecionadas, não foi possível gerar uma análise mais apurada e testes classificatórios para este cenário. Apesar disso, o fluxo de análise proposto pelo modelo possibilita que os tomadores de decisão tenham informações que possibilitam estabelecer a ordem de classificação de ideias e, por meio de representações visuais, facilitar a seleção de potenciais ideias.

Apesar do modelo não ter sido avaliado levando-se em conta um número significativo de ideias, é importante ressaltar que ele provê suporte à análise para grandes volumes de dados. O número de patentes coletadas da USPTO[®] e artigos científicos coletados da *Semantic Scholar*[®] foi expressivo em volume, necessitando de capacidade de processamento.

Por fim, a avaliação do modelo e da ontologia evidenciam a contribuição do trabalho. Os resultados obtidos com a aplicação do protótipo são promissores, visto que o modelo proposto pela tese atingiu uma assertividade de aproximadamente 80% nos dez principais agrupamentos formados e 70% em comparação com o conjunto de ideias selecionadas pelos especialistas. Deste modo, pode-se concluir que o modelo proposto permite auxiliar especialistas na avaliação de ideias com potencial de implementação e, consequentemente, na tomada de decisão.

7.6 CONSIDERAÇÕES FINAIS

Este capítulo apresentou e discutiu a análise do modelo proposto por meio da aplicação do protótipo desenvolvido em dois cenários de análise. Foi considerado o programa Sinapse da Inovação[®] e a coletânea de ideias #MosquitoNão. Inicialmente, a base de conhecimento foi populada com as ideias e com a descrição das temáticas estratégicas relacionadas ao contexto organizacional. Na etapa de pré-processamento, foram gerados quatro índices para estimar o potencial de implementação de uma ideia. Em seguida, foram realizadas as devidas

transformações nos dados visando a preparação para a etapa seguinte. Na etapa de mineração de texto ocorreu a formação dos agrupamentos e utilizou-se o erro quadrático com o objetivo de estabelecer o limiar mais adequado para determinar a formação dos grupos. A partir dos agrupamentos formados apresentou-se um *ranking* das ideias pertencentes ao agrupamento, ordenadas pelo índice de potencialidade. Assim, o índice de potencialidade permite ordenar determinada ideia intra-agrupamento, bem como determinar a ordem de importância dos agrupamentos produzidos. Os resultados obtidos possibilitaram explicitar visualmente os relacionamentos presentes entre as ideias coletadas e seu potencial de implementação com base nas características definidas. As ideias e seus relacionamentos, assim como os índices gerados, promovem suporte à tomada de decisão por especialistas da organização, fornecendo indícios sobre o potencial de implementação das ideias analisadas. Os testes para comparar o desempenho de um modelo preditivo supervisionado (classificação) e sua acurácia, em relação a um modelo preditivo não supervisionado (agrupamento), mostraram que a assertividade dos algoritmos foi baixa em ambos os testes. O índice Kappa ficou próximo de zero, indicando que não existe concordância na classificação realizada pelos algoritmos além do que seria esperado ao acaso. Por fim, a avaliação dos especialistas evidenciou a significância do trabalho, e os testes mostraram a efetividade e assertividade do modelo proposto.

8 CONSIDERAÇÕES FINAIS

Inovar é um fator diferencial em um mercado competitivo, e a principal matéria-prima deste processo são as ideias. A gestão de ideias pode trazer inúmeros benefícios para a organização como: inovações, crescimento econômico, novos produtos, otimização de processos, identificação de oportunidades, vantagem competitiva, minimização de custos de investimento e aumento de produtividade. Contudo, devido à seleção de ideias não inovadoras, altas taxas de inovações falham, impactando diretamente no processo, gerando desperdício de tempo e custo.

Um processo estruturado de seleção, avaliação e implementação de ideias produz maior agilidade e rapidez para levar novos produtos ao mercado. Assim, sistemas de gestão de ideias se tornam plataformas fundamentais para apoiar, estimular e facilitar o processo de gestão de ideias e promover inovações.

Todavia, o principal problema destes sistemas é avaliar e organizar as sugestões de ideias devido ao volume de informações e a estrutura dos dados, representada por coleções de documentos textuais. Além disso, os especialistas podem ter tempo reduzido para avaliar adequadamente inúmeras ideias. A avaliação destas informações, realizada de forma manual, é considerada demorada e dispendiosa. É necessário direcionar esforços em fases gerenciais críticas, tais como estabelecer prioridades entre as ideias inovadoras que são geradas nas organizações e, a partir disso, selecionar as mais condizentes com o contexto organizacional. Entretanto, a formalização e automatização do processo de avaliação são atividades complexas.

Nesta perspectiva, o trabalho iniciou pela revisão sistemática da literatura sobre métodos e técnicas de mineração de dados e texto para auxiliar no processo de avaliação de ideias, visando atender o primeiro objetivo específico. Evidenciou-se na literatura equações, métodos, técnicas e algoritmos utilizados no domínio de gestão de ideias como o cosseno, *tf-idf*, SVD e algoritmos não supervisionados como o *k-means*. Contudo, não foram identificados trabalhos com foco na avaliação de ideias e com a proposição de índices para estimar o potencial de implementação. Tão pouco o estabelecimento de um *ranking* de classificação envolvendo as ideias. Após a revisão, realizou-se a representação do conhecimento de domínio e a proposição do modelo.

O modelo proposto, objetivo geral deste trabalho, possibilita analisar ideias com base em diferentes fontes de dados externas e/ou organizacionais, gerando insumos para que especialistas possam avaliar potenciais ideias a serem desenvolvidas pela organização.

Por meio da aplicação de métodos e técnicas de mineração de dados e texto, estabeleceram-se índices que auxiliam no segundo objetivo específico deste trabalho, tais como: o índice de atualidade da ideia obtido por meio de buscas realizadas na base de artigos científicos *Semantic Scholar*[®]; o índice de aderência ao contexto organizacional com base nas informações repassadas pelos especialistas da organização e/ou fontes externas, mantidas na base de conhecimento; o índice de originalidade que consideram buscas realizadas na base de patentes USPTO[®] e na base *Semantic Scholar*[®]; o índice de tecnologia baseado nas informações presentes na base de conhecimento, informadas por especialistas.

A média dos quatro índices gerados compõe o índice geral de potencialidade da ideia. Este índice indica o grau de alinhamento da ideia às temáticas estratégicas relacionadas ao contexto organizacional e o seu potencial de implementação. Assim, fornece insumos para a avaliação das ideias, possibilitando criar um *ranking* de classificação das ideias.

Também se constituiu em uma etapa fundamental do modelo a análise de agrupamento, importante tarefa para as áreas de mineração de dados e texto e aprendizado de máquina. A análise de agrupamento permite avaliar um conjunto de características em relação à população em análise, com alta empregabilidade, consiste em agrupar objetos com base em suas semelhanças. A ideia central na formação dos grupos é manter a homogeneidade dentro do agrupamento formado e a heterogeneidade entre eles. Tal estratégia reduz o conjunto de análise de n ideias para um novo conjunto de k grupos. Por meio da análise de agrupamento é possível evidenciar padrões e tendências em relação às ideias pertencentes a determinado domínio.

O ponto central é que, em um primeiro momento, o especialista, ao analisar determinado conjunto de ideias, por vezes expressivo em volume, terá dificuldade para definir em qual investir. Todavia, o modelo contribui na visualização dos agrupamentos de ideias e na definição dos índices gerados. Os tomadores de decisão poderão estudar a semântica e as características do grupo, bem como a formação dos índices e, deste modo, obter mais informações que os auxiliem na tomada de decisão. Portanto, pode-se afirmar que os grupos formados permitem interpretar de uma maneira mais adequada as ideias coletadas. Com isto, o processo de tomada de decisão tende a ser facilitado, colaborando na decisão dos especialistas sobre quais ideias são passíveis de serem implementadas e requerem futuros investimentos.

Deste modo, ao se identificar agrupamentos de ideias aderentes às temáticas estratégicas relacionadas ao contexto organizacional, pode-se investir em grupos de ideias. Isso pode inclusive gerar equipes de trabalho mais engajadas, uma vez que mais ideias, em função

da similaridade, podem ser contempladas ou mesmo criar a sumarização do grupo e propor uma nova ideia central.

Adicionalmente, visando contribuir na explicitação do conhecimento produzido pelo modelo, foi utilizado a análise de redes sociais com o intuito de realizar uma apreciação mais visual dos casos de estudo, constituindo em um elemento relevante para a compreensão do trabalho. Desta forma, buscou-se explorar o fluxo das informações e os relacionamentos existentes na formação das redes.

Outra característica importante do modelo é a presença da base de conhecimento representada no formato de uma ontologia. Definida como um modelo conceitual, a ontologia é empregada para caracterizar determinado domínio. São interpretáveis por máquinas e capturam o conhecimento consensual de forma genérica, podendo ser reutilizadas e compartilhadas entre aplicações de *software* e por grupos de pessoas. A ontologia proposta nesta tese é responsável por representar o conhecimento de domínio e características que contribuem na tomada de decisão, bem como os índices gerados durante as etapas do modelo.

Para demonstrar sua efetividade, o modelo proposto foi aplicado em dois cenários de estudo, terceiro objetivo específico desta tese. O primeiro cenário de análise foi composto por ideias coletadas do programa Sinapse da Inovação[®]. Ao total foram coletadas 200 ideias referentes à Fase II do programa. A aplicação do modelo no primeiro cenário objetivou extrair informações do edital para compor o contexto organizacional e, assim, determinar as ideias mais aderentes. A identificação das atividades de impacto mostrou que a proposta está alinhada diretamente ao contexto organizacional, representado pelo edital de fomento. Já o segundo cenário de estudo envolveu ideias coletadas da coletânea #MosquitoNão, promovida pelo Ministério da Educação. Ao total foram 93 ideias selecionadas. No segundo cenário, o objetivo da análise foi identificar as características presentes na formação dos agrupamentos, bem como os índices gerados pelo protótipo, independente do contexto organizacional, com o objetivo de demonstrar a generalidade e aplicabilidade do modelo.

A avaliação do modelo e da ontologia propostos por meio de especialistas de domínio, evidenciou a significância do trabalho. Também foram realizados testes para comparar o desempenho de um modelo preditivo supervisionado (classificação) e sua acurácia, em relação a um modelo preditivo não supervisionado (agrupamento) para o primeiro cenário de estudo. Para tanto, foram aplicados dois algoritmos muito utilizados na área de Aprendizado de Máquina: o algoritmo *Naive Bayes* e o algoritmo *Support Vector Machine*. O algoritmo *Naive Bayes* atingiu 0,485 de acurácia e o algoritmo *Support Vector Machine*, 0,540. Outro ponto relevante do teste é que o coeficiente de Kappa ficou próximo de zero, indicando que não existe

concordância na classificação além do que seria esperado pelo acaso, ou seja, não há qualquer padrão na classificação que possa ser utilizado para alcançar uma acurácia mais elevada.

Esta informação é significativa, uma vez que o modelo proposto aplicado ao primeiro cenário atingiu uma acurácia de 70%, se considerado apenas a ordenação decrescente das ideias pelo índice de potencialidade, ou seja, entre as 100 primeiras ideias ordenadas de maneira decrescente pelo índice de potencialidade, aproximadamente 70 foram implementadas. Se observado os dez agrupamentos de maior relevância, houve 80% de assertividade na formação dos agrupamentos. Ressalta-se que a base de ideias para análise é composta por apenas 200 pré-selecionadas e que pertencem a etapa semifinal do programa. Tendo em vista as considerações acerca das ideias, pode-se dizer que estas 200 ideias já possuíam uma perspectiva de implementação.

Por fim, os resultados obtidos nos cenários de estudo permitem concluir que o modelo proposto tem a capacidade de promover vantagens às organizações em aspectos como a redução do tempo de avaliação das ideias e a disponibilização de indícios sobre o potencial de implementação. Assim, auxilia os tomadores de decisão na definição de quais ideias serão colocadas em prática. Deste modo, o objetivo geral foi atingido e a pergunta de pesquisa respondida através do arcabouço desenvolvido, aplicado e avaliado neste trabalho.

8.1 LIMITAÇÕES

Como limitação, a pesquisa apresenta o cenário de estudo a partir de informações coletadas na *web*. Portanto, é reconhecida a necessidade de criação do contexto organizacional a partir de informações advindas de especialistas da organização. Também se considera como limitação a descrição do contexto organizacional, visto que este conceito é amplo e no trabalho foi descrito a partir de temáticas estratégicas relacionadas a ele.

Outro ponto é o número de ideias avaliadas. Contudo, o modelo proposto fornece suporte a grandes volumes de dados. Mesmo que o número de ideias não seja expressivo devido à falta de acesso à base completa de ideias, foram coletados e analisados um número expressivo de patentes e artigos científicos para definição dos índices.

8.2 TRABALHOS FUTUROS

Durante a realização deste trabalho, diversos pontos de evolução e novas possibilidades de pesquisas e desenvolvimento foram observados. Entre eles, vislumbra-se a

partir dos agrupamentos formados, a sugestão de novas ideias baseadas em pesquisas na base de conhecimento e na *web*, utilizando-se o conceito de mineração de ideias. Outros algoritmos de agrupamento também podem ser aplicados com o objetivo de analisar qual a abordagem mais adequada para lidar com informações textuais. O avanço em formas de normalização e representação de documentos, bem como da análise matricial, podem contribuir na avaliação do modelo proposto na tese.

Cabe sugerir ainda avaliar o modelo em diferentes organizações com contextos diversificados. Além disso, o aprimoramento de análises, por meio de redes sociais, pode se tornar uma ferramenta relevante na gestão de ideias.

O avanço em pesquisas envolvendo indicadores para o domínio de gestão de ideias pode contribuir significativamente na área e complementar o índice de potencialidade proposto nesta tese. Como exemplo, cita-se a criação de um índice com base nas avaliações dos colaboradores sobre as ideias presentes na plataforma.

A ampliação do conceito de contexto organizacional, bem como sua descrição na ontologia, pode contribuir ainda mais com a análise, avaliação e seleção das potenciais ideias para implementação.

O gerenciamento de críticas e sugestões recebidas de colaboradores internos ou externos apresenta-se como um cenário interessante para o modelo. Permitiria, por exemplo, gerar um novo indicador ao analisar e explorar os impactos referentes ao desempenho da organização no mercado junto ao seu público alvo.

Outra possibilidade se refere à temporalidade da ideia, ou seja, a data de sua submissão. Uma ideia pode não ter sido aprovada em determinado período por motivos diversos. Contudo, em um novo ciclo de submissão, novas ideias com potencial de investimento podem ser similares à ideias já existentes na base, mas que não foram implementadas. O processo de agrupamento poderia levar em consideração as ideias atuais, e também analisar as demais ideias similares que não foram implementadas no passado.

É possível também explorar o potencial da análise semântica presente na ontologia proposta, contribuindo na análise das ideias. Entre as alternativas, destacam-se a utilização de raciocínio lógico já disponível em linguagens atuais voltadas à representação de conhecimento e o conceito de Métodos de Resolução de Problemas (PSM - *Problem-Solving Methods*). Por meio de PSMs, torna-se possível a solução de um problema em nível de conhecimento, abstraindo-se os detalhes das ações de inferência que necessitam de conhecimentos específicos para atingir determinados objetivos. Assim, PSM pode ser utilizado como um guia para

aquisição de conhecimento em um domínio específico, ou seja, pode direcionar o processo de elicitación de conhecimento.

Com o objetivo de atingir resultados mais consistentes no processo de avaliação de ideias e contribuir na visualização dos resultados, haveria o desenvolvimento de um sistema de recomendação. Tal sistema, com base em algumas regras e pré-classificações sugeridas por especialistas de domínio, bem como no conjunto de índices gerados, poderia sugerir ideias a serem implementadas.

Por fim, recomenda-se que diferentes tamanhos de organizações e contextos sejam levados em consideração em futuras aplicações. Que mais pesquisas sejam realizadas a fim de complementar a descrição do contexto organizacional.

REFERÊNCIAS

- ABDALA, Lucas N; FERNANDES, R. B.; OGLIARI, A.; LÖWER, M.; FELDHUSEN, J. Creative Contributions of the Methods of Inventive Principles of TRIZ and BioTRIZ to Problem Solving. **Journal of Mechanical Design**, v. 139, n. 8, p. 082001, 2017.
- ABRAHAMSON, Shaun; RYDER, Peter; UNTERBERG, Bastian. **Crowdstorm: the future of innovation, ideas, and problem solving**. John Wiley & Sons, 2013.
- ABUALIGAH, Laith Mohammad; KHADER, Ahamad Tajudin. Unsupervised text feature selection technique based on hybrid particle swarm optimization algorithm with genetic operators for the text *Clustering*. **The Journal of Supercomputing**, p. 1-23, 2017.
- ABUALIGAH, Laith Mohammad; KHADER, Ahamad Tajudin; AL-BETAR, Mohammed Azmi. Unsupervised feature selection technique based on genetic algorithm for improving the Text *Clustering*. In: **International Conference on Computer Science and Information Technology (CSIT)**, IEEE, 2016. p. 1-6.
- ADAMCZYK, S.; BULLINGER, A. C.; MOESLEIN, K. M. Commenting for new ideas: insights from an open innovation platform. **International Journal of Technology Intelligence and Planning**, v. 7, n. 3, p. 232-249, 2011.
- AFACAN, Yasemin; DEMIRKAN, Halime. An ontology-based universal design knowledge support system. **Knowledge-based systems**, v. 24, n. 4, p. 530-541, 2011.
- AL-ANZI, Fawaz S.; ABUZEINA, Dia. Beyond vector space model for hierarchical Arabic text classification: A Markov chain approach. **Information Processing & Management**, v. 54, n. 1, p. 105-115, 2018.
- ALEXANDERSON, Gerald. About the cover: Euler and Königsberg's Bridges: A historical view. **Bulletin of the american mathematical society**, v. 43, n. 4, p. 567-573, 2006.
- ALGHAMDI, Rubayyi; ALFALQI, Khalid. A survey of topic modeling in text mining. I. J. ACSA, v. 6, n. 1, p. 147-153, 2015.
- ALKSHER, M. A; AZMAN, A.; YAAKOB, R.; KADIR, R. A.; MOHAMED, A.; ALSHARI, E. M. A review of methods for mining idea from text. In: **2016 Third International Conference On Information Retrieval And Knowledge Management (camp)**, p. 88-93, 2016.
- ALMEIDA, Mauricio B.; BAX, Marcello P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. **Ci. Inf.**, Brasília, v. 32, n. 3, Dec. 2003.
- ALVAREZ, Guilherme Martins. ANÁLISE DE AGRUPAMENTOS E MINERAÇÃO DE OPINIÃO COMO SUPORTE À GESTÃO DE IDEIAS. **Dissertação** (Mestrado) - Programa de Pós-graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2018.
- ANDRIENKO, G., ANDRIENKO, N. Interactive *Cluster* analysis of diverse types of spatiotemporal data. **Acm Sigkdd Explorations Newsletter**, New York, v. 11, nº 2, p. 19-28, 2009.

- ASENSIO, S. The process of idea management. **Executive Update**, v. 54, 2002.
- BABU, S. Kishore; VASAVI, S.; NAGARJUNA, K. Framework for Predictive Analytics as a Service using ensemble model. In: **2017 IEEE 7th International Advance Computing Conference (IACC)**. IEEE, 2017. p. 121-128.
- BAER, Markus. Putting creativity to work: The implementation of creative ideas in organizations. **Academy of Management Journal**, v. 55, n. 5, p. 1102-1119, 2012.
- BAKKER, Hendrik Jan. **Idea management: unravelling creative processes in three professional organizations**. 2010.
- BAREGHEH, A.; ROWLEY, J.; SAMBROOK, S. Towards a multidisciplinary definition of innovation. **Management Decision**, United Kingdom, v. 47, n. 8, p.1323-1339, 2009.
- BATZIAS, F. A.; SIONTOROU, C. G. Creating a specific domain ontology for supporting R&D in the science-based sector – The case of biosensors. **Expert Systems with Applications**, v. 39, n. 11, p. 9994– 10015, 2012. Elsevier Ltd.
- BEER, J.; MCCARTHY, I. P.; SOLIMAN, A.; TREEN, E. Click here to agree: Managing intellectual property when *crowdsourcing* solutions. **Business Horizons**, v. 60, n. 2, p. 207-217, 2017.
- BEIL, Florian; ESTER, Martin; XU, Xiaowei. Frequent term-based text *Clustering*. In: **Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining**. ACM, 2002. p. 436-442.
- BERETTA, Michela. Idea Selection in Web-Enabled Ideation Systems. **Journal of Product Innovation Management**, v. 36, n. 1, p. 5-23, 2019.
- BERGENDAHL, Magnus; MAGNUSSON, Mats. Combining collaboration and competition: a key to improved idea management? **European Journal of International Management**, v. 8, n. 5, p. 528-547, 2014.
- BERGER, A. CARUANA, R.; COHN, D.; FREITAG, D.; MITTAL, V. Bridging the Lexical Chasm: Statistical Approaches to Answer Finding. In: **Proc. Int. Conf. Research and Development in Information Retrieval**, p. 192-199, 2000.
- BJÖRK, J.; BOCCARDELLI, P.; MAGNUSSON, M. G. Ideation capabilities for continuous innovation. **Creativity & Innovation Management**, Malden, v. 19, n. 4, p. 385-396, 2010.
- BLEI, David M.; NG, Andrew Y.; JORDAN, Michael I. Latent dirichlet allocation. **Journal of machine Learning research**, v. 3, n. Jan, p. 993-1022, 2003.
- BLOHM, I.; RIEDL, C.; LEIMEISTER, J. M.; KRCMAR, H. Idea evaluation mechanisms for collective intelligence in open innovation communities: Do traders outperform raters? In: **Proceedings of 32nd International Conference on Information Systems**, p. 1-24, 2011.
- BONAZZI, Riccardo; VISCUSI, Gianluigi; BARBEY, Valérie. Crowd and experts' knowledge: Connection and value through the notion of prism. In: **European, Mediterranean, and Middle Eastern Conference on Information Systems**. Springer, Cham, 2017. p. 646-654.

BORST, W. N. **Construction of engineering ontologies for knowledge sharing and reuse**. Universiteit Twente, 1997.

BOTHOS, E.; APOSTOLOU, D.; MENTZAS, G. Collective intelligence with web-based information aggregation markets: The role of market facilitation in idea management. **Expert Systems with Applications**, v. 39, n. 1, p. 1333-1345, 2012.

BOUDREAU, Kevin J.; LAKHANI, Karim R. Using the crowd as an innovation partner. **Harvard business review**, v. 91, n. 4, p. 60-9, 140, 2013.

BRABHAM, Daren C. **Crowdsourcing**. Mit Press, 2013.

BÜCHELER, J. H.; SIEG, J. H. Understanding Science 2.0: Crowdsourcing and Open Innovation in the Scientific Method. **Procedia Computer Science**, v. 7, n. 1, p. 327- 329, 2011.

BUNGE, M. **Treatise on Basic Philosophy. Part II: Life Science, Social Science and Technology**. Boston: D. Reidel, 1985.

CAMPOS, Geraldo Maia. **Estatística Prática para Docentes e Pós-Graduandos**. 2000. Disponível em: http://www.forp.usp.br/restauradora/gmc/gmc_livro/gmc_livro_cap14.html. Acesso em: 22 dez. 2019.

CARBONE, F.; CONTRERAS, J.; HERNÁNDEZ, J. Z.; GOMEZ-PEREZ, J. M. Open Innovation in an Enterprise 3.0 framework: Three case studies. **Expert Systems with Applications**. 2012.

CARDOSO, Gonçalo Oom Alves. **Who is co-creating?: the importance of the co-creator in product preference**. Tese de Doutorado. 2016.

CARPINETO, C. OSIŃSKI, S.; ROMANO, G.; WEISS, D. A survey of web *Clustering* engines. **ACM Computing Surveys (CSUR)**, v. 41, n. 3, p. 17, 2009

CASTRO, L. Nunes; VON ZUBEN, Fernanto José. An improving pruning technique with restart for the Kohonen self-organizing feature map. In: **International Joint Conference on Neural Networks, 1999. IJCNN'99**. IEEE, 1999. p. 1916-1919.

CHAN, Joel; DANG, Steven; DOW, Steven P. Comparing Different Sensemaking Approaches for Large-Scale Ideation. In: **Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems**. ACM, 2016. p. 2717-2728.

CHAN, Kimmy Wa; LI, Stella Yiyang; ZHU, John Jianjun. Fostering customer ideation in *crowdsourcing* community: the role of peer-to-peer and peer-to-firm interactions. **Journal of Interactive Marketing**, v. 31, p. 42-62, 2015.

CHAN, Kimmy Wa; LI, Stella Yiyang; ZHU, John Jianjun. Good to Be Novel? Understanding How Idea Feasibility Affects Idea Adoption Decision Making in Crowdsourcing. **Journal of Interactive Marketing**, v. 43, p. 52-68, 2018.

CHANDRASEKARAN, B., JOSEPHSON, J. R., BENJAMINS, R. Ontology of Tasks and Methods. In: **Workshop on Knowledge Acquisition, Modeling and Management (KAW'98)**, Banff, Canada, 1998.

CHESBROUGH, H. **Inovação aberta: como criar e lucrar com a tecnologia**. Porto Alegre: Bookman, 2012.

CHESBROUGH, H. W. **Open innovation**. Harvard Business School Press, Boston, MA, 2003 (a).

CHESBROUGH, H. Open platform innovation: creating value from internal and external innovation. **Intel® Technology Journal**, v. 7, n. 3, p. 5-9, 2003 (b).

CHESBROUGH, Henry W. Bringing open innovation to services. **MIT Sloan Management Review**, v. 52, n. 2, p. 85, 2011.

CHESBROUGH, Henry. Open innovation: a new paradigm for understanding industrial innovation. **Open innovation: Researching a new paradigm**, v. 400, p. 0-19, 2006.

CHESBROUGH, Henry; BRUNSWICKER, Sabine. **Managing open innovation in large firms**. Berkeley, CA: Fraunhofer Verlag, 2013.

CHESBROUGH, Henry; VANHAVERBEKE, Wim; WEST, Joel (Ed.). **Open innovation: Researching a new paradigm**. Oxford University Press on Demand, 2006.

CHIU, Chao-Min; LIANG, Ting-Peng; TURBAN, Efraim. What can *crowdsourcing* do for decision support?. **Decision Support Systems**, v. 65, p. 40-49, 2014.

CLARK, B. Y.; ZINGALE, N.; LOGAN, J.; BRUDNEY, J. A framework for using *crowdsourcing* in government. **International Journal of Public Administration in the Digital Age**, v. 3, n. 4, p. 57-75, 2016.

COHEN, J. A coefficient of agreement for nominal scales. **Educat. Psyc. Measurement**, 1960.

COHN, J.; KATZENBACH, J.; VLAK, G. Finding and grooming breakthrough innovators. **Harvard Business Review**, v. 86, n. 12, p. 62-69, 2008

CONVERTINO, Gregorio; SÁNDOR, A.; BAEZ, Marcos. Idea spotter and comment interpreter: Sensemaking tools for idea management systems. In: **ACM Communities and Technologies Workshop: Large-Scale Idea Management and Deliberation Systems Workshop**. 2013.

CROSS, Rob; PARKER, Andrew; BORGATTI, Steve. Strategic collaboration: Using social network analysis to integrate human networks. **White paper**, 2000.

CUPANI, A. **Filosofia da Tecnologia: um convite**. Florianópolis: Editora da UFSC, 2011.

CUPANI, A. La peculiaridad del conocimiento tecnológico. **ScientiaeStudia**, São Paulo, v. 4, n. 3, p. 353-71, 2006.

DEAN, D. L.; HENDER, J.; RODGERS, T.; SANTANEN, E. Identifying good ideas: constructs and scales for idea evaluation. **Journal of the Association for Information Systems**, 7, 10, 646-699. 2006.

DEERWESTER, S.; DUMAIS, S. T.; FURNAS, G. W.; LANDAUER, T. K.; HARSHMAN, R. Indexing by latent semantic analysis. **Journal of the American society for information science**, v. 41, n. 6, p. 391, 1990.

DENISON, Daniel R.; HART, Stuart L.; KAHN, Joel A. From chimneys to cross-functional teams: Developing and validating a diagnostic model. **Academy of Management journal**, v. 39, n. 4, p. 1005-1023, 1996.

DEPOY, Elizabeth; GITLIN, Laura N. **Introduction to research: Multiple strategies for health and human services**. Mosby, 1994.

DI GANGI, Paul M.; WASKO, Molly. Steal my idea! Organizational adoption of user innovations from a user innovation community: A case study of Dell IdeaStorm. **Decision Support Systems**, v. 48, n. 1, p. 303-312, 2009.

DINH, T. C.; BAE, H.; PARK, J.; BAE, J. A framework to discover potential ideas of new product development from *crowdsourcing* application. **International Conference on Computer, Networks, Systems, and Industrial Applications**, 2015.

DISSANAYAKE, Indika; ZHANG, Jie; GU, Bin. Virtual Team Performance in Crowdsourcing Contest: A Social Network Perspective. In: **System Sciences (HICSS), 2015 48th Hawaii International Conference on**. IEEE, 2015. p. 4894-4897.

DISSELKAMP, Marcus. **Innovations management: Instrumente und Methoden zur Umsetzung im Unternehmen**. Springer-Verlag, 2015.

DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern Classification*. 2nd. ed. Wiley Interscience, 2000.

DUDA, R.; HART, P. **Pattern classification and scene analysis**, New York: Wiley, 1973.

DZIALLAS, Marisa; BLIND, Knut. Innovation indicators throughout the innovation process: An extensive literature analysis. **Technovation**, v. 80, p. 3-29, 2019.

EGC. **EGC - Áreas de Concentração**. 2017. Disponível em: <http://www.egc.ufsc.br/pos-graduacao/programa/areas-de-concentracao/>. Acesso em: 30 Mai. 2017. 2016.

ELIA, Gianluca; MARGHERITA, Alessandro; PASSIANTE, Giuseppina. Digital entrepreneurship ecosystem: How digital technologies and collective intelligence are reshaping the entrepreneurial process. **Technological Forecasting and Social Change**, v. 150, p. 119791, 2020.

ENRÍQUEZ, José Gonzalez et al. An approach to characterize and evaluate the quality of Product Lifecycle Management Software Systems. **Computer Standards & Interfaces**, v. 61, p. 77-88, 2019.

ESCANDÓN-QUINTANILLA, Ma-Lorena; GARDONI, Mickaël; COHENDET, Patrick. Improving concept development with data exploration in the context of an innovation and technological design course. **International Journal on Interactive Design and Manufacturing (IJIDeM)**, v. 12, n. 1, p. 161-172, 2018.

ESCANDON-QUINTANILLA, Ma-Lorena; JIMENEZ-NARVAEZ, Luz-Maria; GARDONI, Mickael. Strategies to employ social networks in early design phases (idea generation). In: **DS 80-8 Proceedings of the 20th International Conference on Engineering Design (ICED 15) Vol 8: Innovation and Creativity**, Milan, Italy, 27-30.07. 15. 2015.

FAIRBANK, James F.; WILLIAMS, Scott David. Motivating creativity and enhancing innovation through employee suggestion system technology. **Creativity and innovation management**, v. 10, n. 2, p. 68-74, 2001.

FAN, Cheng et al. Unsupervised data analytics in mining big building operational data for energy efficiency enhancement: A review. **Energy and Buildings**, v. 159, p. 296-308, 2018.

FAUSETT, Laurene. **Fundamentals of neural networks: architectures, algorithms, and applications**. Prentice-Hall, 1994.

FENN, Jackie; LEHONG, Hung. Hype cycle for emerging technologies, 2011. **Gartner**, July, 2011.

FERAUGE, Perrine. A conceptual framework of corporate social responsibility and innovation. **Global Journal of Business Research**, v. 6, n. 5, p. 85-96, 2012.

FERIOLI, M.; DEKONINCK, E.; CULLEY, S.; ROUSSEL, B.; RENAUD, J. Understanding the rapid evaluation of innovative ideas in the early stages of design. **International Journal of Product Development**, v. 12, n. 1, p. 67-83, 2010.

FERREIRA, Aurélio Buarque de Holanda. **Novo dicionário de língua portuguesa**. 2.ed. Rio de Janeiro : Nova Fronteira, 1986.

FLYNN, M.; DOOLEY, L.; O'SULLIVAN, D.; CORMICAN, K. IDEA MANAGEMENT for organisational innovation. **International Journal of innovation management**, v. 7, n. 04, p. 417-442, 2003.

FREEMAN, Linton C. A set of measures of centrality based on betweenness. **Sociometry**, p. 35-41, 1977.

FREEMAN, Linton C. Some antecedents of social network analysis. **Connections**, v. 19, n. 1, p. 39-42, 1996.

FREITAS JUNIOR, V.; WOSZEZENKI, C.; ANDERLE, D. F.; SPERONI, R.; NAKAYAMA, M. K. A pesquisa científica e tecnológica. **Espacios**, v. 35, n. 9, p. 12, 2014

FREY, Karsten; LÜTHJE, Christian; HAAG, Simon. Whom should firms attract to open innovation platforms? The role of knowledge diversity and motivation. **Long Range Planning**, v. 44, n. 5, p. 397-420, 2011.

FROW, P.; NENONEN, S.; PAYNE, A.; STORBACKA, K. Managing co-creation Design: A Strategic Approach to Innovation. **British Journal of Management**, v. 26, n. 3, p. 463-483, 2015.

GABRIEL, A. MONTICOLO; D., CAMARGO, M.; BOURGAULT, M. Ontology to Represent the Knowledge Domain of a Creative Workshop. In: **12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)**. IEEE, p. 618-623, 2016.

GAMA, Fábio; FRISHAMMAR, Johan; PARIDA, Vinit. Idea generation and open innovation in SMEs: When does market-based collaboration pay off most?. **Creativity and Innovation Management**, v. 28, n. 1, p. 113-123, 2019.

GAMLIN, Janet N.; YOURD, Raymond; PATRICK, Valerie. Unlock creativity with “active” idea management. **Research-Technology Management**, v. 50, n. 1, p. 13-16, 2007.

GEERTS, Guido L. A design science research methodology and its application to accounting information systems research. **International Journal of Accounting Information Systems**, v. 12, n. 2, p. 142-151, 2011.

GEPHI. **Features**. Disponível em: <https://gephi.org/features/>. Acesso em: 22 dez. 2019.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4th ed. São Paulo: Atlas, 2008.

GIL, Antônio Carlos. **Como elaborar projetos de pesquisa**. 5. ed. São Paulo: Atlas, 2010.

GOCHERMANN, Josef; NEE, Ingo. The Idea Maturity Model—A Dynamic Approach to Evaluate Idea Maturity. **International Journal of Innovation and Technology Management**, v. 16, n. 05, p. 1950030, 2019.

GOUILLART, Francis; BILLINGS, Douglas. Community-powered problem solving. **Harvard business review**, v. 91, n. 4, p. 70-7, 140, 2013.

GOUTTE, Cyril; GAUSSIÉ, Eric. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In: **ECIR**. 2005. p. 345-359.

GRIMM, S., HITZLER, P., ABECKER, A. Knowledge Representation and Ontologies: Logic, Ontologies and Semantic Web Languages. In: Rudi Studer, Stephan Grimm, Andreas Abecker (Eds). **Semantic Web Services: Concepts, Technologies, and Applications**. Springer, 2007

GROSSETTI, M. Dynamiques des réseaux et des cercles. Encastremets et découplages. **Revue d'économie industrielle**, v. 2, p. 327-355, 2003.

GROVER, Veena Mehta. Alignment of Business Strategy and Innovation Strategy: Strategic Benefits. **Imperial Journal of Interdisciplinary Research**, v. 3, n. 1, 2017.

GRUBER, T. **What is an Ontology?** 1993. Disponível em: <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>. Acesso em: Jan. 2015.

GUARINO, N. Formal Ontology in Information Systems: **Proceedings of the 1st International Conference** June 6-8, 1998, Trento, Italy. 1998.

GUPTA, Vishal et al. A survey of text mining techniques and applications. **Journal of emerging technologies in web intelligence**, v. 1, n. 1, p. 60-76, 2009.

GURTEEN, D. Knowledge, creativity and innovation. **Journal of Knowledge Management**, v.2, n.1, p.5-13, 1998

HAIR, J. F.; ANDERSON, R. E.; TATHAM, R. L.; WILLIAM, C. **Multivariate Data Analysis**. 7. ed. Pearson Prentice Hal, 2010. 593 p.

HARRISON, Michael I. et al. Effects of organizational context on Lean implementation in five hospital systems. **Health care management review**, v. 41, n. 2, p. 127-144, 2016.

HERNÁNDEZ-GONZÁLEZ, Y.; GARCÍA-MORENO, C.; RODRÍGUEZ-GARCÍA, M. Á.; VALENCIA-GARCÍA, R.; GARCÍA-SÁNCHEZ, F. A semantic-based platform for R&D project funding management. **Computers in Industry**, v. 65, n. 5, p. 850-861, 2014.

HEVNER, A.; CHATTERJEE, S. Design science in information systems research. **MIS quarterly**, v. 28, n. 1, p. 75-105, 2010.

HOBBER, Bjoern; SCHAARSCHMIDT, Mario; VON KORFLESCH, Harald. Internal idea contests: Work environment perceptions and the moderating role of power distance. **Journal of Innovation & Knowledge**, 2019.

HOORNAERT, S.; BALLINGS, M.; MALTHOUSE, E. C.; VAN DEN POEL, D. Identifying new product ideas: waiting for the wisdom of the crowd or screening ideas in real time. **Journal of Product Innovation Management**, v. 34, n. 5, p. 580-597, 2017.

HOSSAIN, Mokter; ISLAM, KM Zahidul. Generating ideas on online platforms: A case study of "My Starbucks Idea". **Arab Economic and Business Journal**, v. 10, n. 2, p. 102-111, 2015 (a).

HOSSAIN, Mokter; ISLAM, KM Zahidul. Ideation through online open innovation platform: dell IdeaStorm. **Journal of the Knowledge Economy**, v. 6, n. 3, p. 611-624, 2015 (b).

HOSSAIN, Mokter; KAURANEN, Ilkka. Crowdsourcing: a comprehensive literature review. **Strategic Outsourcing: An International Journal**, v. 8, n. 1, p. 2-22, 2015.

HOWE, J. The Rise of Crowdsourcing. Disponível em: <http://www.wired.com/wired/archive/14.06/crowds.html>, 2006.

INSNA (**International Network for Social Network Analysis**). Disponível em: <http://www.insna.org>. Acesso em: 01/05/2020.

JAIN, A. K.; MURTY, M. N.; FLYNN, P. J. Data *Clustering*: a review. **Acm Computing Surveys**, v. 31 n° 3, p. 264-323, 1999.

JAIN, Anil K. Data *Clustering*: 50 years beyond K-means. **Pattern recognition letters**, v. 31, n. 8, p. 651-666, 2010.

JASPER, R; USCHOLD, M. A Framework for Understanding and Classifying Ontology Applications. **Proceedings of the IJCAI-99 Workshop on Ontologies and Problem Solving Method (KRR5)**, Stockholm, Sweden, 1999.

JIANG, D; TANG, C; ZHANG, A. *Cluster analysis for gene expression data: a survey*. **IEEE Transactions on Knowledge and Data Engineering**, v. 16. 11, p. 1370-1386, 2004.

JONES, W.P.; FURNAS, G.W. Pictures of relevance: A geometric analysis of similarity measures. **Journal of American Society for Information Science**, v. 38, n. 6, p. 420-442, 1987.

JOURET, Guido. Inside Cisco's search for the next big idea. **Harvard Business Review**, v. 87, n. 9, p. 43-45, 2009.

KAHRAMAN, Cengiz. Preface for the special issue: Uncertainty modeling and analytics in knowledge engineering. **Knowledge-Based Systems**, n. 70, p. 1-2, 2014.

KARIMI-MAJD, Amir-Mohsen; MAHOOTCHI, Masoud. A new data mining methodology for generating new service ideas. **Information Systems and e-Business Management**, v. 13, n. 3, p. 421-443, 2015.

KASKI, Samuel. Data exploration using self-organizing maps. In: **Acta polytechnica scandinavica: Mathematics, computing and management in engineering series**, n. 82. 1997.

KASKI, Samuel; KOHONEN, Teuvo. Exploratory data analysis by the self-organizing map: Structures of welfare and poverty in the world. In: **Neural networks in financial engineering. Proceedings of the third international conference on neural networks in the capital markets**. 1996.

KATERNDAHL, D. Evolution of the research collaboration network in a productive department. **Journal of Evaluation in Clinical Practice**, v. 18, n. 1, p. 195-201, 2012.

KHAN, K.; BAHARUDIN, B.; KHAN, A.; ULLAH, A. Mining opinion components from unstructured reviews: A review. **Journal of King Saud University-Computer and Information Sciences**, v. 26, n. 3, p. 258-275, 2014.

KHAZAEI, Taraneh; XIAO, Lu. Computational Analysis of Collective Intelligence in Conversational Text. In: **System Sciences (HICSS), 2015 48th Hawaii International Conference on. IEEE**, 2015. p. 1596-1605.

KIM, Han Kyul; KIM, Hyunjoong; CHO, Sungzoon. Bag-of-Concepts: Comprehending Document Representation through *Clustering Words* in Distributed Representation. **Neurocomputing**, 2017.

KING, Andrew; LAKHANI, Karim R. Using open innovation to identify the best ideas. **MIT Sloan management review**, v. 55, n. 1, p. 41, 2013.

KLEIN, Mark; CONVERTINO, Gregorio. A roadmap for open innovation systems. **Journal of Social Media for Organizations**, v. 2, n. 1, p. 1, 2015.

KLOPOTEK, Mieczyslaw A.; WIERZCHON, Slawomir T.; TROJANOWSKI, Krzysztof (Ed.). Intelligent Information Processing and Web Mining: **Proceedings of the International IIS: IIPWM '04 Conference held in Zakopane**, Poland, May 17–20, 2004. Springer Science & Business Media, 2013.

KOEN, P. A.; AJAMIAN G. M.; BOYCE, S.; CLAMEN, A.; FISHER, E.; FOUNTOULAKIS, S.; SEIBERT, R. Fuzzy front end effective methods, tools, and techniques. In: BELLIVEAU; A. P.; GRIFFIN; S. SOMERMEYER (Eds); **The PDMA toolbox for new product development**. New York: John Wiley, 2002.

KÖHL, Anna et al. How Text Mining Algorithms for Crowdsourcing Can Help Us to Identify Today's Pressing Societal Issues. In: **Proceedings of the 52nd Hawaii International Conference on System Sciences**. 2019.

KOHLER, Thomas; CHESBROUGH, Henry. From collaborative community to competitive market: the quest to build a *crowdsourcing* platform for social innovation. **R&D Management**, v. 49, n. 3, p. 356-368, 2019.

KOHN, Stefan; HÜSIG, Stefan. Development of an empirical-based categorisation scheme for CAI software. **International Journal of Computer Applications in Technology**, v. 30, n. 1-2, p. 33-46, 2007.

KOHONEN, Teuvo. Self-organized formation of topologically correct feature maps. **Biological cybernetics**, v. 43, n. 1, p. 59-69, 1982.

KONCHADY, Manu. **Text mining application programming**. Massachusetts: Charles River Media, 2006.

KORNISH, Laura J.; ULRICH, Karl T. The importance of the raw idea in innovation: Testing the sow's ear hypothesis. **Journal of Marketing Research**, v. 51, n. 1, p. 14-26, 2014.

KOULOPOULOS, T. M. **Inovação com resultado**. São Paulo: Editora Gente ; Editora Senac, 2011.

KRIDEL, D.; DOLK, D. Automated self-service modeling: predictive analytics as a service. **Information Systems and e-Business Management**. vol. 11, p. 119–140, 2013.

KRUSE, P.; SCHIEBER, A.; HILBERT, A.; SCHOOP, E. **Idea mining–text mining supported knowledge management for innovation purposes**. 2013.

LACERDA, D. P.; DRESCH, A.; PROENÇA, A.; ANTUNES JÚNIOR, J. A. V. Design Science Research: método de pesquisa para a engenharia de produção. **Gest. Prod.**, São Carlos, v. 20, n.4, p. 741-761, 2013.

LACERDA, Juliana Subtil; VAN DEN BERGH, Jeroen CJM. Effectiveness of an ‘open innovation’ approach in renewable energy: Empirical evidence from a survey on solar and wind power. **Renewable and Sustainable Energy Reviews**, v. 118, p. 109505, 2020.

LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. **Biometrics**, 1977. p. 159–174

LAUTO, G.; VALENTIN, F.; HATZACK, F.; CARLSEN, M. Managing front-end innovation through idea markets at Novozymes. **Research-Technology Management**, v. 56, n. 4, p. 17-26, 2013.

LAWRENCE, Carl; TUUNANEN, Tuure; MYERS, Michael. Extending Design Science Research Methodology for a Multicultural World. **Human Benefit through the Diffusion of Information Systems Design Science Research**, p. 108-121, 2010.

LEE, H.; CHOI, K.; YOO, D.; SUH, Y.; LEE, S.; HE, G. Recommending valuable ideas in an open innovation community: A text mining approach to information overload problem. **Industrial Management & Data Systems**, v. 118, n. 4, p. 683-699, 2018.

LEE, Hanjun; SEO, Soyoung. What Determines an Agreeable and Adoptable Idea? A Study of User Ideas on MyStarbucksIdea. com. In: **System Sciences (HICSS)**, 2013 46th Hawaii International Conference on. IEEE, 2013. p. 3207-3217.

- LEE, Tae-Young. A study on extracting ideas from documents and webpages in the field of idea mining. **Journal of the Korean Society for information Management**, v. 29, n. 1, p. 25-43, 2012.
- LEE, Won Sang; CHOI, Hyo Shin; SOHN, So Young. Forecasting new product diffusion using both patent citation and web search traffic. **PloS one**, v. 13, n. 4, p. e0194723, 2018.
- LESKOVEC, Jure; RAJARAMAN, Anand; ULLMAN, Jeffrey David. **Mining of massive datasets**. Cambridge university press, 2014.
- LI, C.; LIN, M.; YANG, L. T.; DING, C. Integrating the enriched feature with machine learning algorithms for human movement and fall detection. **The Journal of Supercomputing**, v. 67, n. 3, p. 854-865, 2014.
- LI, F.; YANG, Y.; XIE, J.; ZHAO, X.; MA, J.; LI, Y. A Dynamic and Fuzzy Modeling Approach for Multi-attribute New Product Idea Screening and Portfolio. **International Journal Of Advancements In Computing Technology(ijact)**, v. 14, p. 96-103, 2012.
- LI, Xingsen; LI, Liping; CHEN, Zhengxin. Toward extenics-based innovation model on intelligent knowledge management. **Annals of Data Science**, v. 1, n. 1, p. 127-148, 2014.
- LIU, Haixia; GOULDING, James; BRAILSFORD, Tim J. Towards Computation of Novel Ideas from Corpora of Scientific Text. In: **ECML/PKDD (2)**. 2015. p. 541-556.
- LIU, L.; TANG, L.; DONG, W.; YAO, S.; ZHOU, W. An overview of topic modeling and its current applications in bioinformatics. **SpringerPlus**, v. 5, n. 1, p. 1608, 2016.
- LIU, S.; ZHOU, M. X.; PAN, S.; SONG, Y.; QIAN, W.; CAI, W.; LIAN, X. Tiara: Interactive, topic-based visual text summarization and analysis. **ACM Transactions on Intelligent Systems and Technology (TIST)**, v. 3, n. 2, p. 25, 2012.
- LULA, P.; PALIWODA-PEKOSZ, G. An Ontology-Based *Cluster* Analysis Framework. **Proceedings Of The First International Workshop On Ontology-supported Business Intelligence**, New York, p.1-6, 2008.
- MAČIULIENĖ, Monika; SKARŽAUSKIENĖ, Aelita. Evaluation of co-creation perspective in networked collaboration platforms. **Journal of Business Research**, 2016.
- MACQUEEN, J. B. Some methods for classification and analysis of multivariate observations. **Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability**, vol. 1: Statistics, p. 281-297, University of California Press, Berkeley, Calif., 1967.
- MAEDCHE, Alexander. **Ontology Learning for the Semantic Web**, 2002.
- MAGNUSSON, Peter R.; NETZ, Johan; WÄSTLUND, Erik. Exploring holistic intuitive idea screening in the light of formal criteria, **Technovation**, vol. 34, n. 5–6, May–June 2014, Pages 315-326, 2014.
- MAHROUM, Sami; AL-SALEH, Yasser, Towards a functional framework for measuring national innovation efficacy, **Technovation**, vol. 33, n° 10–11, 2013, p. 320-332, ISSN 0166-4972.

MARCH, Salvatore T.; SMITH, Gerald F. Design and natural science research on information technology. **Decision support systems**, v. 15, n. 4, p. 251-266, 1995.

MARTÍNEZ-TORRES, M. Rocío. Application of evolutionary computation techniques for the identification of innovators in open innovation communities. **Expert Systems with Applications**, v. 40, n. 7, p. 2503-2510, 2013.

MARTINEZ-TORRES, Rocio; OLMEDILLA, Maria. Identification of innovation solvers in open innovation communities using swarm intelligence. **Technological Forecasting and Social Change**, v. 109, p. 15-24, 2016.

MARTINI, Antonella; NEIROTTI, Paolo; APPIO, Francesco Paolo. Knowledge searching, integrating and performing: always a tuned trio for innovation?. **Long Range Planning**, v. 50, n. 2, p. 200-220, 2017.

MARTINS, M. E. et al. Service's scientific community: A social network analysis (1995-2010). **Journal of Service Management**, 23(3), 455-469, 2012.

MAUZY, Jeffrey H. Managing Personal Creativity. **Design Management Review**, v. 17, n. 3, p. 64-72, 2006.

MCCORMACK, Bridget; FALLON, Enda F.; CORMICAN, Kathryn. An Analysis of Open Innovation Practices in the Medical Technology Sector in Ireland. **Procedia Manufacturing**, v. 3, p. 503-509, 2015.

MCKEOWN, Max. **The truth about innovation**. Pearson Education India, 2008.

MECHANT, P.; STEVENS, I.; EVENS, T.; VERDEGEM, P. E-deliberation 2.0 for smart cities: a critical assessment of two 'idea generation' cases. **International Journal of Electronic Governance**, v. 5, n. 1, p. 82-98, 2012.

MERGEL, Ines. Opening government: Designing open innovation processes to collaborate with external problem solvers. **Social Science Computer Review**, v. 33, n. 5, p. 599-612, 2015.

MICHALKO, M. From bright ideas to right ideas: capturing the creative spark. **The Futurist**, v. 37, n. 5, p. 52-56, 2003.

MIKELSONE, Elīna; VOLKOVA, Tatjana; LIELĀ, Elita. Potential Benefits of Web-based Idea Management System Based on Practical Evidence. In: **Proceedings of the 12th International Scientific and Practical Conference**, v. 2, 2019 (a).

MIKELSONE, Elīna; VOLKOVA, Tatjana; LIELĀ, Elita. Practical Evidence of Web-Based Idea Management Systems: Classification and Application. **Research for Rural Development**, v. 2, 2019 (b).

MINAYO, Maria Cecília de Souza. O desafio da pesquisa social. **Pesquisa social: teoria, método e criatividade**, v. 27, p. 9-29, 2007.

MISHRA, Nishchol; SILAKARI, Sanjay. Predictive analytics: A survey, trends, applications, opportunities & challenges. **International Journal of Computer Science and Information Technologies**, v. 3, n. 3, p. 4434-4438, 2012.

MOOS, B.; WAGNER, H.-T.; BEIMBORN, D.; WEITZEL, T., "The Role of Innovation Governance and Knowledge Management for Innovation Success, **44th Hawaii International Conference on System Sciences (HICSS)**, Jan. 2011.

MORAES, R. M.; MACHADO, L. S. Psychomotor skills assessment in medical training based on virtual reality using a weighted possibilistic approach. **Knowledge Based Systems**, 2014. p. 97-62.

MUHDI, L.; DAIBER, M.; FRIESIKE, S.; BOUTELLIER, R. The *crowdsourcing* process: an intermediary mediated idea generation approach in the early phase of innovation. **International Journal of Entrepreneurship and Innovation Management**, v. 14, n. 4, p. 315-332, 2011.

MURAH, M. Z.; ABDULLAH, Z.; HASSAN, R.; BAKAR, M. A.; MOHAMED, I.; AMIN, H. M. Kacang cerdas: A conceptual design of an idea management system. **International Education Studies**, v. 6, n. 6, p. 178-184, 2013.

MURUKANNAIAH, Pradeep K.; AJMERI, Nirav; SINGH, Munindar P. Acquiring creative requirements from the crowd: Understanding the influences of personality and creative potential in Crowd RE. In: **Requirements Engineering Conference (RE), 2016 IEEE 24th International**. IEEE, 2016. p. 176-185.

NAMBISAN, S.; SAWHNEY, M. **Cérebro Global**: como inovar em um mundo conectado por redes. São Paulo: Évora, 2011.

NEWMAN, Mark EJ. The structure of scientific collaboration networks. **Proceedings of the national academy of sciences**, v. 98, n. 2, p. 404-409, 2001.

NEWMAN, M. E. J. Modularity and community structure in networks. **Proc.Natl.Acad.Sci.Usa**, 103, 2006.

NEWMAN, M. E. J.; GIRVAN, M. Finding and evaluating community structure in networks. **Physical Review E, American Physical Society**, v. 69, n. 2, 2004.

NIEDERBERGER, T.; STOOP, N.; CHRISTEN, M.; OTT, T. Hebbian principal component *Clustering* for information retrieval on a *crowdsourcing* platform. In: **Nonlinear Dynamics of Electronic Systems**, Proceedings of NDES 2012. VDE, 2012. p. 1-4.

RUSSEL, S.; NORVIG, P. **Inteligência artificial**. Editora Campus, v. 20, 2004.

NOY, N. F.; MCGUINNESS, D. L. **Ontology Development 101: A Guide to Creating your First Ontology**. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March 2001.

NYCE, Charles; CPCU, A. Predictive analytics white paper. **American Institute for CPCU Insurance Institute of America**, p. 9-10, 2007.

OSIŃSKI, Stanisław. **An algorithm for Clustering of web search results**. Tese de Doutorado. Poznań University of Technology, Poland. 2003.

OSIŃSKI, Stanisław; STEFANOWSKI, Jerzy; WEISS, Dawid. Lingo: Search results *Clustering* algorithm based on singular value decomposition. In: **Intelligent information processing and web mining**. Springer, Berlin, Heidelberg, 2004. p. 359-368.

PANCHAL, Jitesh H. Using Crowds in Engineering Design—Towards a Holistic Framework. In: **International Conference on Engineering Design**, Design Society, Milan, Italy, July. 2015. p. 27-30.

PEFFERS, K.; TUUNANEN, T.; ROTHENBERGER, M. A.; CHATTERJEE, S. A design science research methodology for information systems research. **Journal of management information systems**, v. 24, n. 3, p. 45-77, 2007.

POETZ, Marion K.; SCHREIER, Martin. The value of *crowdsourcing*: can users really compete with professionals in generating new product ideas?. **Journal of Product Innovation Management**, v. 29, n. 2, p. 245-256, 2012.

POLI, Roberto; OBRST, Leo. The interplay between ontology as categorial analysis and ontology as technology. In: **Theory and applications of ontology: Computer applications**. Springer Netherlands, 2010. p. 1-26.

POORKAVOOS, M.; DUAN, Y.; EDWARDS, J. S.; RAMANATHAN, R. Identifying the configurational paths to innovation in SMEs: A fuzzy-set qualitative comparative analysis. **Journal of Business Research**, v. 69, n. 12, p. 5843-5854, 2016.

POVEDA, G.; WESTERSKI, A.; IGLESIAS, C. A. Application of semantic search in Idea Management Systems. **International Conference for Internet Technology And Secured Transactions**, 2012, vol., no., p.230 - 236, 10-12 Dec. 2012

PRIEM, R. L.; LI, S. L.; CARR, J. C. Insight and new directions from demand-side approaches to technology innovation, entrepreneurship, and strategic management research. **Journal of Management**, v. 38, n. 1, p. 346-374, 2012.

PRPIĆ, J.; SHUKLA, P. P.; KIETZMANN, J. H.; MCCARTHY, I. P. How to work a crowd: Developing crowd capital through *crowdsourcing*. **Business Horizons**, v. 58, n. 1, p. 77-85, 2015.

PUSTOVRH, Ales; RANGUS, Kaja; DRNOVŠEK, Mateja. The role of open innovation in developing an entrepreneurial support ecosystem. **Technological Forecasting and Social Change**, v. 152, p. 119892, 2020.

RAMOS, Juan. Using tf-idf to determine word relevance in document queries. In: **Proceedings of the first instructional conference on machine learning**. 2003. p. 133-142.

RENARD, Damien; DAVIS, Joseph G. Social interdependence on *crowdsourcing* platforms. **Journal of Business Research**, v. 103, p. 186-194, 2019.

REPKO, A. F. **Interdisciplinary research**. Thousand Oaks: Sage Publications. 2008.

RIEDL, C.; MAY, N.; FINZEN, J.; STATHEL, S.; KAUFMAN, V.; KRCCMAR, H. **An idea ontology for innovation management**. 2009.

ROSARIO, Barbara. Latent semantic indexing: An overview. **Techn. rep. INFOSYS**, v. 240, p. 1-16, 2000.

RUNCO, Mark A.; JAEGER, Garrett J. The standard definition of creativity. **Creativity Research Journal**, v. 24, n. 1, p. 92-96, 2012.

RUSSEL, S.; NORVIG, P. **Artificial intelligence: a modern approach**. Prentice-Hall: New Jersey, 1995. 932p.

SABOL, V.; KIENREICH, W.; MUHR, M.; KLIEBER, W.; GRANITZER, M. Visual Knowledge Discovery in Dynamic Enterprise Text Repositories. **13th International Conference Information Visualisation**, vol., n., p. 361-368, July 2009.

SALDIVAR, J.; BÁEZ, M.; RODRIGUEZ, C.; CONVERTINO, G.; KOWALIK, G. Idea Management Communities in the Wild: An Exploratory Study of 166 Online Communities. In: **Collaboration Technologies and Systems (CTS)**, 2016 International Conference on. IEEE, 2016. p. 81-89.

SALDIVAR, Jorge et al. Online Idea Management for Civic Engagement: A Study on the Benefits of Integration with Social Networking. **ACM Transactions on Social Computing**, v. 2, n. 1, p. 1, 2019.

SARIGIANNI, Christina et al. Innovation Contests: How to Design for Successful Idea Selection. In: **Proceedings of the 53rd Hawaii International Conference on System Sciences**. 2020.

SCHREIBER, A. T.; SCHREIBER, G.; AKKERMANS, H.; ANJEWIERDEN, A.; SHADBOLT, N.; DE HOOG, R.; WIELINGA, B. **Knowledge engineering and management: The CommonKADS Methodology**. MIT Press, Cambridge, Massachusetts, 2000.

SCHREIBER, G. Knowledge engineering. In V. Lifschitz, F. van Harmelen, B. Porter (Eds). **Handbook of Knowledge Representation**. Elsevier, 2007. p. 929-946.

SCHUURMAN, D.; BACCARNE, B.; DE MAREZ, L.; MECHANT, P. Smart ideas for smart cities: Investigating *crowdsourcing* for generating and selecting ideas for ICT innovation in a city context. **Journal of theoretical and applied electronic commerce research**, v. 7, n. 3, p. 49-62, 2012.

SCHWEITZER, F. M.; BUCHINGER, W.; GASSMANN, O.; OBRIST, M. Crowdsourcing: leveraging innovation through online idea competitions. **Research-Technology Management**, v. 55, n. 3, p. 32-38, 2012.

SERGIO, M. C.; SOUZA, J. A.; GONCALVES, A. L. Idea identification model to support decision making. **IEEE Latin America Transactions**, v. 15 n. 5, 2017.

SÉRGIO, Marina Carradore. Um Modelo Baseado em Ontologia e Análise de Agrupamento para Suporte à Gestão de Ideias. **Dissertação (Mestrado)** - Programa de Pós-graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2016.

SÉRGIO, Marina Carradore; GONÇALVES, Alexandre Leopoldo. Análise e interpretação de ideias: proposta de um modelo. **Perspectivas em Ciência da Informação**, v. 24, n. 2, p. 54-71, 2019.

SÉRGIO, Marina Carradore; GONÇALVES, Alexandre Leopoldo. Representação Semântica do Domínio de Gestão de Ideias: Proposta de uma Ontologia. **16ª Conferência Internacional Sobre Sistemas de Informação e Gestão de Tecnologia: CONTECSI**. São Paulo, junho 2019. Disponível em:

<<http://www.contecsi.tecsi.org/index.php/contecsi/16CONTECSI/schedConf/presentations>>. Acesso em: 27 de out. 2019.

SIANGLIULUE, P.; CHAN, J.; DOW, S. P.; GAJOS, K. Z. IdeaHound: Improving Large-scale Collaborative Ideation with Crowd-powered Real-time Semantic Modeling. In: **Proceedings of the 29th Annual Symposium on User Interface Software and Technology**. ACM, 2016. p. 609-624.

SILVA, Edna Lúcia da; MENEZES, Estera Muszkat. Metodologia da pesquisa e elaboração de dissertação. 4. ed. rev. **Atual**, Florianópolis: UFSC, 2005, 138p

SIMON, H. A. **The Sciences of the Artificial**. 3rd ed. Cambridge: MIT Press, 1996.

SINAPSE DA INOVAÇÃO. **Projeto g-Ideia: Uma ponte entre o Governo e a Sociedade**. 2015. Disponível em: <<http://sc5.sinapsedainovacao.com.br/sc5/ideia/projeto-g-ideia>>. Acesso em: 10 jan. 2018.

SINT, R.; MARKUS, M.; SCHAERT, S.; KURZ, T. Ideator - a collaborative enterprise idea management tool powered by KiWi. **Fifth Workshop Semantic Wikis. Linking Data and People**. Hersonissos, Greece, 2010.

SOUZA, Adriana Silva. **PBLontology: Uma Ontologia de Domínio com uma Abordagem Contextual para Aprendizagem Baseada em Problemas**. Dissertação (Mestrado). Universidade Federal da Bahia, Salvador, 2016.

SOUZA, Wagner Vilas Boas et al. Using crowdstorm to prospect innovations in federal institutions of education in Brazil to reduce its consumption of electric energy. In: **2016 49th Hawaii International Conference on System Sciences (HICSS)**. IEEE, 2016. p. 2819-2828.

STANKOWITZ, Rosângela de Fátima. **GESTÃO DE IDEIAS: estrutura de referência para inovação aberta**. 2014. Tese (Doutorado em Tecnologia e Sociedade). Universidade Federal de Santa Catarina (UFSC).

STEELE, Logan M.; JOHNSON, Genevieve; MEDEIROS, Kelsey E. Looking beyond the generation of creative ideas: Confidence in evaluating ideas predicts creative outcomes. **Personality and Individual Differences**, v. 125, p. 21-29, 2018.

STEVANOVIĆ, M; MARJANOVIĆ, D; FTORGA, M. Decision Support System for Idea Selection. **International Design Conference - Design 2012**. Dubrovnik, p. 1951-1960. 2012.

STEWART, Gilbert W. On the early history of the singular value decomposition. **SIAM review**, v. 35, n. 4, p. 551-566, 1993.

STUDER, R.; BENJAMINS, V. R.; FENSEL, D. Knowledge engineering: principles and methods. **Data & knowledge engineering**, v. 25, n. 1, p. 161-197, 1998.

SUÁREZ-FIGUEROA, M. C.; GARCÍA-CASTRO, R.; VILLAZÓN TERRAZAS, B.; GÓMEZ-PÉREZ, A. Essentials In Ontology Engineering: Methodologies, Languages, and Tools. In: **CIB conference W078-W012 - 2nd Workshop on eeBuildings Data Models**, 26-28 out. 2011, Sophia Antipolis, França, 2011.

TARTUCE, T. J. A. **Métodos de pesquisa**. Fortaleza: UNICE–Ensino Superior, v. 7, 2006.

THORLEUCHTER, Dirk; VAN DEN POEL, Dirk. Extraction of ideas from microsystems technology. **Advances in Computer Science and Information Engineering**, p. 563-568, 2012.

THORLEUCHTER, Dirk; VAN DEN POEL, Dirk. Idea mining for web-based weak signal detection. **Futures**, v. 66, p. 25-34, 2015.

THORLEUCHTER, Dirk; VAN DEN POEL, Dirk. Identification of interdisciplinary ideas. **Information Processing & Management**, v. 52, n. 6, p. 1074-1085, 2016.

THORLEUCHTER, Dirk; VAN DEN POEL, Dirk; PRINZIE, Anita. Mining ideas from textual information. **Expert Systems with Applications**, v. 37, n. 10, p. 7182-7188, 2010.

TIDD, J.; BESSANT, J.; PAVITT, K. **Gestão da inovação**. 3ª ed. Porto Alegre: Bookman, 2005.

TIDD, J.; BESSANT, J.; PAVITT, K. **Gestão da Inovação**. 5. Ed. Porto Alegre: Bookman, 2015.

TOMAEL, M. I.; MARTELETO, R. M. Redes sociais de dois modos: aspectos conceituais. **Transinformação**, v. 25, n. 3, p. 245-253, 2013.

TOUBIA, Olivier; NETZER, Oded. Idea Generation, Creativity, and Prototypicality. **Marketing Science**, v. 36, n. 1, p. 1-20, 2016.

TOWNE, W. Ben; ROSÉ, Carolyn P.; HERBSLEB, James D. Measuring Similarity Similarly: LDA and Human Perception. **ACM Transactions on Intelligent Systems and Technology (TIST)**, v. 8, n. 1, p. 7, 2016.

VALDES-VASQUEZ, Rodolfo; KLOTZ, Leidy E. Using the concept-mapping method for empirical studies in construction research. **Journal of Construction Engineering and Management**, v. 139, n. 10, p. 04013002, 2013.

VANDENBOSCH, Betty; SAATCIOGLU, Argun; FAY, Sharon. Idea management: a systemic view. **Journal of Management Studies**, v. 43, n. 2, p. 259-288, 2006.

VARINY, Miroslav et al. Energy saving measures from their cradle to full adoption with verified, monitored, and targeted performance: a look back at energy audit at Catalytic Naphtha Reforming Unit (CCR). **Energy Efficiency**, p. 1-23, 2019.

WALL, Michael E.; RECHTSTEINER, Andreas; ROCHA, Luis M. Singular value decomposition and principal component analysis. In: **A practical approach to microarray data analysis**. Springer, Boston, MA, 2003. p. 91-109.

WANG, Hao; OHSAWA, Yukio. Idea discovery: A scenario-based systematic approach for decision making in market innovation. **Expert Systems with Applications**, v. 40, n. 2, p. 429-438, 2013.

WANG, Kai. Collective innovation: A literature review. In: **2013 Proceedings of PICMET'13: Technology Management in the IT-Driven Services (PICMET)**. IEEE, 2013. p. 608-615.

WANG, Kai; WANG, Hui; TAO, Yu. Combining Ideas in Crowdsourced Idea Generation. **Foundations of Management**, v. 9, n. 1, p. 203-212, 2017.

WANG, M. Exploring potential R&D collaborators with complementary technologies: The case of biosensors. **Technological Forecasting & Social Change**, v. 79, p. 862-874. 2012.

WASSERMAN, Stanley; FAUST, Katherine. Social Network Analysis: methods and applications. **Structural analysis in social the social sciences series**. Cambridge: Cambridge University Press, (1994) 1999. v. 8. 857 p. ISBN 0-521-38707-8.

WAZLAWICK, Raul Sidnei. Uma reflexão sobre a pesquisa em ciência da computação à luz da classificação das ciências e do método científico. **Revista de Sistemas de Informação da FSMA**, v. 6, p. 3-10, 2010.

WEHNERT, P.; KOLLWITZ, C.; DAIBERL, C.; DINTER, B.; BECKMANN, M. Capturing the Bigger Picture? Applying Text Analytics to Foster Open Innovation Processes for Sustainability-Oriented Innovation. **Sustainability**, v. 10, n. 10, p. 3710, 2018.

WESTERSKI, A.; DALAMAGAS, T.; IGLESIAS, C. A. Classifying and comparing community innovation in Idea Management Systems, **Decision Support Systems**, 2013.

WESTERSKI, Adam. **Semantic technologies in idea management systems: a model for interoperability, linking and filtering**. Adam Westerski, 2013.

WESTERSKI, Adam; IGLESIAS, Carlos A.; GARCIA, Javier Espinosa. Idea relationship analysis in open innovation *crowdsourcing* systems. **In: Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom)**, 2012 8th International Conference on. IEEE, 2012. p. 289-296.

WESTERSKI, Adam; IGLESIAS, Carlos Angel. Exploiting Structured Linked Data in Enterprise Knowledge Management Systems: An Idea Management Case Study. EDOCW. p. 395-403, **IEEE Computer Society**, 2011.

WESTERSKI, Adam; IGLESIAS, Carlos Angel; RICO, Fernando Tapia. A Model for Integration and Interlinking of Idea Management Systems. In: **MTSR**. 2010. p. 183-194.

WESTERSKI, Adam; KANAGASABAI, Rajaraman. In search of disruptive ideas: outlier detection techniques in crowdsourcing innovation platforms. **International Journal of Web Based Communities**, v. 15, n. 4, p. 344-367, 2019.

WISDOM, Thomas N.; GOLDSTONE, Robert L. Innovation, Imitation, and Problem Solving in a Networked Group. **Nonlinear Dynamics-Psychology and Life Sciences**, v. 15, n. 2, p. 229, 2011.

WROBLEWSKI, MICHAL. **A hierarchical www pages Clustering algorithm based on the vector space model**. Tese de Doutorado. Poznań University of Technology, Poland. 2003.

XIE, L., ZHANG, P. Idea Management System for Team Creation. **Journal of Software**, North America, 5, nov. 2010.

YE, Q.; SONG, H.; LI, T. Cross-institutional collaboration networks in tourism and hospitality research. **Tourism Management Perspectives**, v.2, n.3, p.55-64, 2012.

YOO, D.; CHOI, K.; LEE, H.; SUH, Y. An Ontology-based Co-creation Enhancing System for Idea Recommendation in an Online Community. **ACM SIGMIS Database**, v. 46, n. 3, p. 9-22, 2015.

ZEKRI, Hamid; MOKHTARI, Ahmad Reza; COHEN, David R. Application of singular value decomposition (SVD) and semi-discrete decomposition (SDD) techniques in *Clustering* of geochemical data: an environmental study in central Iran. **Stochastic environmental research and risk assessment**, v. 30, n. 7, p. 1947-1960, 2016.

ZHAO, H.; SALLOUM, S.; CAI, Y.; HUANG, J. Z. Ensemble subspace *Clustering* of text data using two-level features. **International Journal of Machine Learning and Cybernetics**, p. 1-16, 2015.

ZHAO, S.; ZHOU, M. X.; ZHANG, X.; YUAN, Q.; ZHENG, W.; FU, R. Who is doing what and when: Social map-based recommendation for content-centric social web sites. **ACM Transactions on Intelligent Systems and Technology (TIST)**, v. 3, n. 1, p. 5, 2011.

ZOU, Lingfei; ZHANG, Jinlong; LIU, Wenxing. Perceived justice and creativity in *crowdsourcing* communities: Empirical evidence from China. **Social Science Information**, v. 54, n. 3, p. 253-279, 2015.

APÊNDICE A – Formulário de avaliação da ontologia

Informações iniciais

Nome completo:

E-mail:

Titulação:

Área de concentração:

Função: () Docente () Coordenador de área de concentração () Coordenador de PPG ()
Coordenador de Departamento

Dimensão 1 - Questões de pesquisa

1) Você considera a identificação de potenciais ideias com base no contexto organizacional uma atividade essencial no processo de tomada de decisões estratégicas?

- () Concordo plenamente
- () Concordo parcialmente
- () Indiferente
- () Discordo parcialmente
- () Discordo totalmente

2) Você considera que a análise e elucidação de potenciais ideias possam trazer subsídios para a tomada de decisão?

- () Concordo plenamente
- () Concordo parcialmente
- () Indiferente
- () Discordo parcialmente
- () Discordo totalmente

3) Identificar informações relevantes sobre dados não estruturados como texto é um grande desafio, sendo necessários algoritmos confiáveis para encontrar ou gerar recursos (padrões) interpretáveis para analisar esses dados de forma significativa. Uma possibilidade consiste na geração de agrupamentos. O intuito na formação de agrupamentos, a partir de informações textuais, é separar vários documentos dispostos de forma desorganizada em um conjunto de grupos que contenham documentos com assuntos semelhantes. Isto permite localizar documentos rapidamente em uma dada coleção. A fim de fornecer uma representação mais significativa dos dados textuais, e para melhorar a eficiência dos algoritmos de análise textuais, a Análise de Agrupamento foi utilizada. Nesse caso, o processo de identificação de potenciais ideias com base no contexto organizacional é modelado de forma a permitir o agrupamento de ideias semelhantes com base no contexto organizacional. Você considera que essa abordagem complementa as possibilidades de identificação de potenciais ideias?

- () Concordo plenamente
- () Concordo parcialmente
- () Indiferente
- () Discordo parcialmente
- () Discordo totalmente

Dimensão 2 - A partir da documentação que inclui a apresentação da ontologia proposta e os procedimentos para sua operacionalização, responda as seguintes perguntas:

1) A modelagem e representação de conhecimentos distintos, envolvidos no processo de identificação de potenciais ideias com base no contexto organizacional, propostos e definidos neste modelo, auxiliam no processo de tomada de decisão pela organização?

- Concordo plenamente
- Concordo parcialmente
- Indiferente
- Discordo parcialmente
- Discordo totalmente

2) Você considera os conceitos (representados pelas classes) e relacionamentos apresentados nesta representação suficientes para permitir que vários aspectos do processo de elucidação de potenciais ideias sejam explorados?

- Concordo plenamente
- Concordo parcialmente
- Indiferente
- Discordo parcialmente
- Discordo totalmente

3) As perguntas de competências são suficientes e cobrem os aspectos mais importantes relacionados à identificação de potenciais ideias alinhadas ao contexto organizacional?

- Concordo plenamente
- Concordo parcialmente
- Indiferente
- Discordo parcialmente
- Discordo totalmente

4) Observe a figura da documentação que representa o modelo baseado em conhecimento para identificação de potenciais ideias. A partir do seu entendimento a respeito dos elementos que compõem o modelo baseado em conhecimento, você considera que ele pode contribuir significativamente no processo de tomada de decisão?

- Concordo plenamente
- Concordo parcialmente
- Indiferente
- Discordo parcialmente
- Discordo totalmente

Dimensão 3 - Resultados obtidos a partir do modelo proposto

Avalie os resultados das perguntas de competência em relação ao critério "grau de contribuição" no processo de identificação de potenciais ideias alinhadas ao contexto organizacional. Escolha uma entre três opções: alto, médio, baixo.

1. Quais são as potenciais ideias de acordo com o valor de investimento? () Alto () Médio () Baixo
2. Quais são as potenciais ideias de acordo com o valor de investimento e áreas de investimentos potenciais? () Alto () Médio () Baixo
3. Quais são as potenciais ideias de acordo com o índice de aderência ao contexto organizacional? () Alto () Médio () Baixo
4. Quais são as potenciais ideias de acordo com o índice de originalidade e o valor de investimento? () Alto () Médio () Baixo
5. Quais são as potenciais ideias de acordo com o índice de originalidade? () Alto () Médio () Baixo
6. Quais são as potenciais ideias de acordo com as áreas de investimento potenciais da organização? () Alto () Médio () Baixo
7. Quais são as potenciais ideias de acordo com o índice de tecnologia? () Alto () Médio () Baixo
8. Quais são as ideias semelhantes? () Alto () Médio () Baixo
9. Quais são as ideias pertencentes à determinada categoria? () Alto () Médio () Baixo
10. Quais são as ideias que receberam maior número de votos? () Alto () Médio () Baixo
11. Quais são as ideias mais recentes? () Alto () Médio () Baixo
12. Quais são ideias que possuem maior valor estimado de investimento? () Alto () Médio () Baixo
13. Quais são as potenciais ideias de acordo com o índice de atualidade? () Alto () Médio () Baixo
14. Quais são as ideias pertencentes a determinado grau de maturidade? () Alto () Médio () Baixo

15. Quais são as ideias em que a organização possui o número de colaboradores necessários para executá-las?

Alto Médio Baixo

Sugestões ou comentários: