

Jeiel Miguel Lopes

**MODELO BASEADO EM ANÁLISE DE AGRUPAMENTOS  
VOLTADO À GESTÃO DE IDEIAS**

Dissertação submetida ao Programa de Pós-Graduação em Tecnologias da Informação e Comunicação da Universidade Federal de Santa Catarina para a obtenção do Grau de Mestre em Tecnologias da Informação e Comunicação.

Orientador: Prof. Dr. Alexandre Leopoldo Gonçalves.

Coorientador: Prof. Dr. Robson Rodrigues Lemos.

Araranguá  
2019

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

Lopes, Jeiel Miguel

MODELO BASEADO EM ANÁLISE DE AGRUPAMENTOS  
VOLTADO À GESTÃO DE IDEIAS / Jeiel Miguel Lopes ;  
orientador, Alexandre Leopoldo Gonçalves,  
coorientador, Robson Rodrigues Lemos, 2019.  
148 p.

Dissertação (mestrado) - Universidade Federal de  
Santa Catarina, Campus Araranguá, Programa de Pós  
Graduação em Tecnologias da Informação e Comunicação,  
Araranguá, 2019.

Inclui referências.

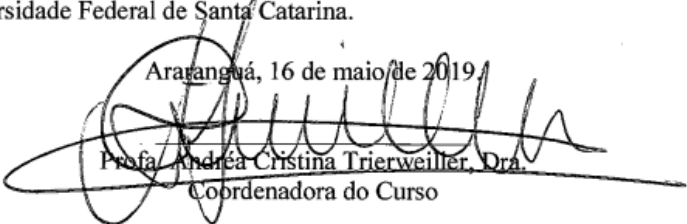
1. Tecnologias da Informação e Comunicação. 2.  
Análise de Agrupamentos. 3. Gestão de Ideias. 4.  
Gestão de Inovação. I. Gonçalves, Alexandre Leopoldo.  
II. Lemos, Robson Rodrigues. III. Universidade  
Federal de Santa Catarina. Programa de Pós-Graduação  
em Tecnologias da Informação e Comunicação. IV. Título.

Jeiel Miguel Lopes

**MODELO BASEADO EM ANÁLISE DE AGRUPAMENTOS  
VOLTADO À GESTÃO DE IDEIAS**


Esta Dissertação foi julgada adequada para obtenção do Título de “Mestre”, e aprovada em sua forma final pelo Programa de Pós-Graduação em Tecnologias da Informação e Comunicação da Universidade Federal de Santa Catarina.

Araçanguá, 16 de maio de 2019.

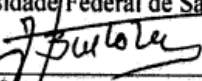


Prof. Andréa Cristina Trierweiler, Dra.  
Coordenadora do Curso

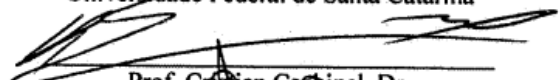
**Banca Examinadora:**



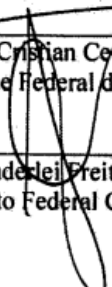
Prof. Alexandre Leopoldo Gonçalves, Dr.  
Orientador  
Universidade Federal de Santa Catarina



Prof. Juarez Bento da Silva, Dr.  
Universidade Federal de Santa Catarina



Prof. Cristian Cochinel, Dr.  
Universidade Federal de Santa Catarina



Prof. Vandelei Freitas Junior, Dr.  
Instituto Federal Catarinense



Este trabalho é dedicado à minha esposa, meu filho, meu mentor e à minha família que me incentivaram e tornaram esta conquista possível.



## AGRADECIMENTOS

Gostaria de primeiramente agradecer ao meu Deus pelas ricas dádivas que tem permitido em minha vida. Agradeço imensamente a minha esposa Aline, que tem sido minha fiel companheira e amiga, e me motiva a superar os desafios da jornada, por me apoiar a seguir em frente, investindo nosso tempo neste projeto. Aos nossos pais, que tem dado total apoio e tem orado por nossas vidas. Aos familiares e amigos, que por muitas vezes compreenderam a necessidade de estarmos ausentes. Ao meu orientador Dr Alexandre Leopoldo Gonçalves, a quem honestamente admiro, sem sombra de dúvidas, sua vida e dedicação são um exemplo a seguir. Aos meus gestores, que compreenderam os desafios deste programa e investiram na minha capacitação acadêmica e profissional. E, não menos importante, ao meu mentor, Dr. Edilson Dalbello, que acompanhou esta minha trajetória, ouviu-me e me aconselhou, impulsionando-me a alcançar meus objetivos. Sem estas pessoas, certamente não teria chegado até aqui. Agradeço também aos membros da banca examinadora que dispuseram de seu tempo e conhecimento para colaborar com este trabalho. Finalmente, ao PPGTIC pela oportunidade de receber e compartilhar conhecimentos neste curso de mestrado. A todos, meu muito obrigado.





“A man never stands so tall as when he stoops to help a boy”

(Johnnie Barnes)



## RESUMO

Com os avanços da *internet* as pessoas passaram a ter maior poder de comunicação produzindo novas demandas para o mercado. Conectadas a isso, muitas empresas passaram a implementar processos de inovação aberta sendo suportados por colaborações externas à organização, principalmente no que tange a captura de ideias. Consequentemente, novos desafios surgiram, entre eles, o aumento no volume de ideias e, a partir disso, a necessidade de apresentar respostas mais rápidas e assertivas visando manter as organizações competitivas. Sendo assim, este trabalho propõe um modelo voltado à avaliação e gestão de ideias, a fim de proporcionar informações que auxiliem especialistas de domínio na seleção de potenciais ideias. Para tanto, o modelo se utiliza de Análise de Agrupamentos combinada à criação de uma ordenação (*ranking*) dos grupos, de acordo com a sua relevância, estabelecida por meio da composição de um índice de atualidade. Com o intuito de avaliar os resultados da aplicação do modelo, foram elaborados diferentes cenários de estudos sobre os dados coletados na comunidade de Gestão de Ideias, Ubuntu Brainstorm<sup>®</sup>, combinados aos dados de artigos coletados na base de dados Semantic Scholar<sup>®</sup>. A partir dos resultados obtidos constata-se que a aplicação do índice de atualidade nos agrupamentos possibilita estabelecer a ordenação e, desta forma, a relevância desses. Em todos os cenários analisados, a ordenação no processo de agrupamento permitiu confirmar que os grupos com maior relevância tendem a conter ideias positivas e implementadas. Por fim, conclui-se que o modelo proposto neste trabalho é capaz de produzir informações para auxiliar especialistas de domínio no processo de Gestão de Ideias.

**Palavras-chave:** Análise de Agrupamentos, Gestão de Ideias, Gestão de Inovação.



## ABSTRACT

With the advances of the internet, people started to have greater power of communication producing new demands for the market. Aiming to expand its competitiveness, many companies started to implement processes of open innovation being supported by collaborations external to the organization, mainly in what concerns the capture of ideas. As a result, new challenges have arisen, among them, the increase in the number of ideas and, hence, the need to present faster and more assertive responses in order to keep organizations competitive. Thus, this work proposes a model focused on the analysis and selection of ideas. To do so, the model suggests the use of Cluster Analysis combined with the creation of a ranking of the groups, according to their relevance, established through the composition of a current index. In order to evaluate the results of the application of the model, different study scenarios were elaborated using the data collected in the Idea Management community, Ubuntu Brainstorm®, combined with data from articles collected from the Semantic Scholar® database. From the results obtained it was found that by using the current index makes possible to establish the significance of the groups, so enabling the ordering of the clusters. In all scenarios analyzed the ordering allowed to confirm that clusters with better ranking tend to contain positive and implemented ideas. Finally, it is concluded that the proposed model in this work is capable of producing information to assist domain specialists in the process of Idea Management.

**Keywords:** Cluster Analysis, Idea Management, Innovation Management



## LISTA DE FIGURAS

Figura 1 – Modelo de processo da metodologia DSR.....	36
Figura 2 – Metodologia x estrutura do trabalho.....	36
Figura 3 – Fluxograma de um <i>Stage-Gate System</i> tradicional .....	42
Figura 4 – <i>New Concept Development Model</i> .....	44
Figura 5 – <i>New Concept Development Model</i> .....	45
Figura 6 – Cinco fatores da Gestão Integrada de Ideias .....	48
Figura 7 – <i>The Technology Development Model</i> .....	49
Figura 8 – Tarefas da Mineração de Dados .....	60
Figura 9 – Fatores relevantes do processo de agrupamento .....	61
Figura 10 – Agrupamento hierárquico aglomerativo e divisivo.....	65
Figura 11 – 5-tree.....	65
Figura 12 – Representação de um dendograma .....	66
Figura 13 – Etapas do agrupamento baseado em grade .....	69
Figura 14 – <i>Cluster</i> baseado em densidade.....	70
Figura 15 - Etapas do modelo proposto .....	82
Figura 16 – Modelo proposto.....	83
Figura 17 – Visualização de ideias no Ubuntu Brainstorm® .....	84
Figura 18 – Diagrama de Entidade-Relacionamento .....	85
Figura 19 – Página de <i>download</i> dos dados da Semantic Scholar®.....	88
Figura 20 – JSON Schema provido pela Semantic Scholar® .....	89
Figura 21 – Comando para obter a base de dados da Semantic Scholar® .....	89
Figura 22 – Diagrama ER proposto para os dados da Semantic Scholar® .....	90
Figura 23 – Tabela de fato de artigos por palavra-chave e ano.....	92
Figura 24 – Tabela de relacionamento de palavras-chave e ideias .....	95
Figura 25 – Fluxograma do algoritmo de análise de agrupamentos.....	98
Figura 26 – Comparação da média de polaridade e ideias implementadas das execuções 13 - 16.....	108
Figura 27 – Comparação do número de clusters das execuções 13 - 16 .....	108
Figura 28 – Comparação do número de clusters das execuções 1 e 17.....	113
Figura 29 – Comparação da média de polaridade e ideias implementadas das execuções 1 e 17.....	113
Figura 30 – Comparação do número de clusters das execuções 2 e 18.....	115
Figura 31– Comparação da média de polaridade e ideias implementadas das execuções 2 e 18 .....	115
Figura 32 – Comparativo entre a execução 6 e a execução 22.....	118
Figura 33 – Os primeiros cinco <i>clusters</i> da execução 18 (polaridade x <i>status</i> ) .	121
Figura 34 – Análise da relação polaridade x <i>status</i> do <i>cluster</i> 18 .....	122
Figura 35 – Somatório do índice de atualidade por polaridade e <i>status</i> do <i>cluster</i> 18.....	123
Figura 36 – Nuvem de termos do <i>cluster</i> 18.....	123
Figura 37 – Análise da relação polaridade x <i>status</i> do <i>cluster</i> 2.....	124
Figura 38 – Somatório do índice de atualidade por polarização e <i>status</i> do <i>cluster</i> 2.....	125

Figura 39 – Nuvem de termos do <i>cluster 2</i> .....	125
Figura 40 – Análise da relação polaridade x <i>status</i> do <i>cluster 39</i> .....	126
Figura 41 – Somatório do índice de atualidade por polarização e <i>status</i> do <i>cluster 39</i> .....	127
Figura 42 – Nuvem de termos do <i>cluster 39</i> .....	128
Figura 43 – Análise da relação polaridade x <i>status</i> do <i>cluster 1</i> .....	129
Figura 44 – Somatório do índice de atualidade por polaridade e <i>status</i> do <i>cluster 1</i> .....	130
Figura 45 – Nuvem de termos do <i>cluster 1</i> .....	130



## LISTA DE TABELAS

Tabela 1 – Exemplo de cálculo do índice de atualidade .....	93
Tabela 2 – Resultado da primeira sessão de agrupamento .....	107
Tabela 3 – Os dez primeiros <i>clusters</i> da Execução 1 ordenados pelo índice de atualidade.....	110
Tabela 4 – Os dez primeiros <i>clusters</i> da Execução 2 ordenados pelo índice de atualidade.....	111
Tabela 5 – Resultado da segunda sessão de agrupamento .....	112
Tabela 6 – Os dez primeiros <i>clusters</i> da Execução 17 ordenados pelo índice de atualidade.....	114
Tabela 7 – Os dez primeiros <i>clusters</i> da execução 18 ordenados pelo índice de atualidade.....	116
Tabela 8 – Os dez primeiros <i>clusters</i> da Execução 6 ordenados pelo índice de atualidade.....	117
Tabela 9 – Os dez primeiros <i>clusters</i> da Execução 22 ordenados pelo índice de atualidade.....	118



## LISTA DE QUADROS

Quadro 1 – Resultado da busca de trabalhos correlatos .....	71
Quadro 2 - Comparação dos trabalhos correlatos .....	76
Quadro 3 – Descrição dos atributos do JSON <i>Schema</i> .....	88
Quadro 4 – Cenários de análise de agrupamento .....	103



## LISTA DE ABREVIATURAS E SIGLAS

DSR – *Design Science Research*  
ER – Entidade-Relacionamento  
FEI – *Front End da Inovação (Front End of Innovation)*  
FFE – *Fuzzy Front End*  
IMS – *Idea Management Systems*  
JSON – *JavaScript Object Notation*  
KROM – *Knowledge Reuse Organizational Meta-Model*  
MDS – *Multidimensional Scaling*  
NCD – *New Concept Development model*  
NPPD – *New Product Process Development*  
PCA – *Principal Component Analysis*  
PDF – *Portable Document Format*  
PD&I – Pesquisa, Desenvolvimento e Inovação  
PMEs – Pequenas e Médias Empresas  
PPGTIC – Programa de Pós-Graduação em Tecnologias da Informação e Comunicação  
P&D – Pesquisa e Desenvolvimento  
RDF – *Resource Description Framework*  
SIGs – Sistemas de Gerenciamento de Ideias  
SOMs – *Self-Organizing Maps*  
SVD – *Singular Value Decomposition*  
TDM – *The Technology Development Model*  
TF-IDF – *Term Frequency-Inverse Document Frequency*  
TIC – Tecnologias da Informação e Comunicação  
UFSC – Universidade Federal de Santa Catarina  
URL – *Uniform Resource Locator*  
VSM – *Vector Space Model*



## SUMÁRIO

<b>1 INTRODUÇÃO .....</b>	<b>25</b>
1.1 DEFINIÇÃO DO PROBLEMA .....	29
1.2 OBJETIVOS .....	32
1.2.1 Objetivo Geral .....	32
1.2.2 Objetivos Específicos .....	32
1.3 JUSTIFICATIVA E RELEVÂNCIA DO TEMA.....	32
1.4 DELIMITAÇÕES DA PESQUISA .....	34
1.5 PROCEDIMENTOS METODOLÓGICOS.....	34
1.6 ADERÊNCIA E INTERDISCIPLINARIDADE .....	36
1.7 ESTRUTURA DO TRABALHO .....	37
<b>2 GESTÃO DE IDEIAS.....</b>	<b>39</b>
2.1 INTRODUÇÃO À GESTÃO DE IDEIAS .....	39
2.2 <i>STAGE-GATE SYSTEMS</i> .....	41
2.3 <i>NEW CONCEPT DEVELOPMENT MODEL</i> .....	43
2.3.1 Os Cinco Elementos do <i>Front End of Innovation</i> .....	46
2.4 <i>INTEGRATED IDEA MANAGEMENT MODEL</i> .....	47
2.5 <i>THE TECHNOLOGY DEVELOPMENT MODEL</i> .....	49
2.5.1 <i>System Boundary</i> .....	50
2.5.2 <i>System Environment</i> .....	50
2.5.3 <i>Inputs</i> .....	50
2.5.4 <i>Outputs</i> .....	50
2.5.5 <i>The Process</i> .....	51
2.6 <i>CROWDSOURCING</i> .....	52
2.7 SISTEMAS DE GESTÃO DE IDEIAS .....	54
<b>3 ANÁLISE DE AGRUPAMENTOS .....</b>	<b>59</b>
3.1 INTRODUÇÃO À ANÁLISE DE AGRUPAMENTOS .....	59
3.2 MÉTODOS DE ANÁLISE DE AGRUPAMENTOS .....	63
3.2.1 <i>Baseados em distância</i> .....	64

3.2.2 Baseados em hierarquia.....	64
3.2.3 Baseados em difusão.....	66
3.2.4 Baseados em centroides.....	66
3.2.5 Baseados em pesquisa.....	67
3.2.6 Baseados em grafos.....	68
3.2.7 Baseados em grade.....	68
3.2.8 Baseados em modelos.....	69
3.2.9 Baseados em densidade.....	69
3.3 TRABALHOS CORRELATOS.....	70
<b>4 MODELO PROPOSTO.....</b>	<b>81</b>
4.1 DESCRIÇÃO DO MODELO DA BASE DE IDEIAS.....	81
4.1.1 Elaboração da estrutura de ideias e carga de dados.....	84
4.1.2 Pré-processamento das Ideias.....	87
4.1.3 Coleta, armazenamento e padronização dos dados de uma base acadêmica.....	87
4.1.4 Cálculo do índice de atualidade.....	92
4.1.5 Aplicação do Índice de Atualidade nas Ideias.....	95
4.1.6 Agrupamento de Ideias Similares.....	95
4.1.7 Análise e explicitação do conhecimento.....	99
4.2 CONSIDERAÇÕES FINAIS.....	100
<b>5 APLICAÇÃO DO MODELO E ANÁLISE DOS RESULTADOS.....</b>	<b>103</b>
5.1 CENÁRIOS DE APLICAÇÃO.....	103
5.2 RESULTADOS DOS EXPERIMENTOS.....	104
5.3 ANÁLISE DOS AGRUPAMENTOS.....	119
5.4 CONSIDERAÇÕES FINAIS.....	131
<b>6 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS.....</b>	<b>133</b>
6.1 CONSIDERAÇÕES FINAIS.....	133
6.2 PERSPECTIVAS DE TRABALHOS FUTUROS.....	135
<b>REFERÊNCIAS.....</b>	<b>137</b>



## 1 INTRODUÇÃO

Nos últimos anos, as economias emergentes tem enfrentado dificuldades para manter sua posição competitiva. Em meio a um mercado taxado como hiperdinâmico, catóxico e incerto, para que se tenha sucesso, os ciclos de inovação tem se tornado cada vez mais curtos (DUIN et al., 2010).

Uma inovação de sucesso inicia com uma ideia, e termina quando um novo produto é disponibilizado ao mercado. No entanto, para se manter a competitividade de uma empresa, tornou-se necessário transformar o processo de inovação numa operação constante, colocando frequentemente novos produtos no mercado. Nesse sentido, um processo de inovação efetivo é considerado crucial para esse sucesso (ANDROSCH; REDL, 2017).

A manutenção do processo de inovação, depende da captação e análise de novas ideias. Essas, por sua vez, precisam do tempo e espaço certo para amadurecerem. Este processo de amadurecimento requer que as empresas desenvolvam uma forte cultura de inovação. Nas fases iniciais, apesar das ideias serem imaturas, possuem alto nível de inovação e criatividade. Portanto, é importante deixá-las maturar e não descartar uma ideia muito diferente ainda no início das avaliações (ANDROSCH; REDL, 2017).

Neste ecossistema, onde a geração contínua de novas ideias é superestimada, o tradicional modelo de inovação interno de uma empresa tornou-se limitado. Esta estratégia de inovação onde os projetos são desenvolvidos utilizando apenas os recursos internos e exclusivamente dentro da empresa denomina-se Inovação Fechada (SILVA; ZILBER, 2013).

Por conta da necessidade de captação de novas ideias, as organizações passaram a compreender a necessidade de interagir entre si. Esse processo iterativo é motivado pela busca de soluções inovadoras, capazes de atender às demandas de seus clientes, oferecendo produtos com valores e serviços agregados (DUIN et al., 2010).

Apesar da necessidade constante de inovação, a qualidade dos produtos e serviços ofertados continua sendo um fator relevante para o sucesso de uma empresa. O investimento em Pesquisa, Desenvolvimento e Inovação (PD&I) é um dos aspectos que resguarda a qualidade dos produtos de uma empresa. Nas empresas do grupo Voestalpine®, o departamento de Pesquisa e Desenvolvimento (P&D) é considerado um elemento extremamente relevante na composição do potencial criativo. Em seu processo de PD&I, as equipes são

descentralizadas, e utilizam redes sociais internas, para coletar ideias de seus clientes mais próximos e parceiros comerciais. A partir disso, compartilham-nas com as demais equipes de inovação. Ainda nesse processo, destaca-se a parceria dos departamentos de PD&I com mais de 80 universidades e institutos de pesquisa (ANDROSCH; REDL, 2017).

Este ecossistema de PD&I possui potencial para gerar um fluxo contínuo de novas ideias. Após a coleta, os melhores textos precisam ser selecionados e adicionados ao *roadmap* (roteiro de desenvolvimento) de pesquisa. A partir do *roadmap*, um rigoroso processo seletivo deve ser executado, uma vez que, os custos de pesquisa são considerados relativamente baratos quando comparados com os custos de implementação. Para tanto, é necessário que a seleção das ideias corretas ocorra no tempo certo, captando inclusive as demandas emergentes (JIMÉNEZ-NARVAEZ; GARDONI, 2015; ANDROSCH; REDL, 2017).

A integração de esforço de diferentes empresas com seus públicos, por meio da rede de alcance mundial (*World Wide Web – WWW*), tem despertado o potencial criativo a nível global. No entanto, apesar de ter se apresentado como um caminho promissor, esse potencial criativo a nível global, conectado por redes de colaborativas, ainda necessita superar desafios, para que o potencial criativo das massas possa ser adequadamente utilizado. (ANDROSCH; REDL; 2017, DUIN et al., 2010).

Esta estratégia de inovação colaborativa é chamada de Inovação Aberta. Essa, por sua vez, está baseada em quatro pontos centrais que precisam ser implementados: desenvolvimento de redes de comunicação; consolidação do conhecimento compartilhado; mecanismos de proteção de propriedade intelectual; e, um modelo de negócio de apoio (MCCORMACK; FALLON; CORMICAN, 2015).

De acordo com McCormack, Fallon e Cormican (2015), no mercado medicinal Irlandês, por exemplo, 75% das empresas estudadas consideram a Inovação Aberta importante ou muito importante e 82% destacam a importância da inovação contínua. As redes sociais, por sua vez, são consideradas extremamente importantes, tanto para os usuários internos, quanto para os externos à organização. O contexto de Inovação Aberta implica desafios substanciais, entretanto, esses geram uma nova interação entre os atores envolvidos em projetos abertos, e na sociedade (JIMÉNEZ-NARVAEZ; GARDONI, 2015).

Entre os principais desafios estão:

- O gerenciamento e a supervisão desses projetos se tornam mais complexos;
- Custos extras para gerenciar a cooperação com parceiros externos;
- Aumento da dependência de recursos externos;
- Perda de algumas competências tecnológicas;
- Aumento do risco de vazamento de propriedade intelectual e transbordamento involuntário (JIMÉNEZ-NARVAEZ; GARDONI, 2015).

Em se tratando de Inovação Aberta, uma das etapas mais importantes é a fase inicial da inovação. Nela, ocorre a ideação, o refinamento da ideia, a construção do conhecimento e a avaliação. Além dessas atividades, ocorrem outras tarefas que constituem aspectos fundamentais para evolução do processo criativo (QIN et al., 2016).

Esse processo de geração de ideias, passou a ser realizado de forma colaborativa, fazendo com que as pessoas, conectadas pela rede e por iniciativa própria, interajam entre si, colaborando no processo de ideação ao compartilhar ideias em torno de um tema. Mediante este processo colaborativo, surgiu a demanda de gerenciamento das ideias geradas, de forma que, dentre as colaborações, fosse possível identificar oportunidades para a etapa de implementação (DUIN et al., 2010).

Num processo iterativo e colaborativo, a inovação pode ser encorajadora. Entretanto, o processo de geração de ideias requer uma estratégia de projetos capaz de envolver diversos tipos de conhecimento, compartilhando ideias e oportunidades (GARDONI, 2015).

As plataformas de *crowdsourcing*, por exemplo, são compatíveis com o modelo de Inovação Aberta. Nestas plataformas são implementados processos de pesquisa coletiva, normalmente realizados a partir de redes de comunicação *online*. Todavia, ainda precisam evoluir na construção de um modelo iterativo que compreenda a preservação da propriedade intelectual e recursos para gerir largas escalas de projetos ou ideias (JIMÉNEZ-NARVAEZ; GARDONI, 2015).

Uma das aplicações dos portais colaborativos de *crowdsourcing* no processo de gerenciamento de ideias é na contribuição para a transferência de tecnologia. O processo de transferência de tecnologia compreende o compartilhamento de informações e recursos entre empresas. Esse compartilhamento contempla desde as dinâmicas

estabelecidas nos primeiros estágios da conceptualização do produto, até os métodos e tecnologias para produção (JIMÉNEZ-NARVAEZ; GARDONI, 2015).

O crescimento da competitividade do mercado tem forçado as organizações a manterem um contínuo e fluido processo de inovação. Nas comunidades de gerenciamento de ideias reside grande potencial de transformar negócios e comunidades por meio da inovação. Para tanto, são coletadas ideias de administradores, colaboradores e agora de pessoas externas à organização. Contudo, construir comunidades colaborativas de sucesso é uma tarefa que exige grandes esforços (QIN et al., 2016).

Aliadas às práticas de gestão de ideias, essas colaborações tem empoderado iniciativas ao redor do mundo. Utilizando-se dos recursos da *internet*, a população em geral tem sido convidada a contribuir no processo de criação de novos produtos e serviços, baseados em suas necessidades (SALDIVAR et al., 2016).

A Gestão de Ideias tem o potencial de descobrir ideias de valor que possam ser conduzidas num processo de inovação. Construir comunidades *online* sustentáveis é um grande desafio que requer o entendimento das pessoas envolvidas e de suas necessidades, bem como a combinação das características dos usuários e do propósito da comunidade.

Realizar a gestão eficiente desses recursos requer a utilização de sistemas de gerenciamento de ideias, os quais são utilizados para coleta, seleção, avaliação e desenvolvimento de novas ideias. Com o intuito de suprir essa demanda, surgiram os Sistemas de Gerenciamento de Ideias (*Idea Management Systems - IMS*). Esses sistemas abrangem desde a possibilidade de registrar ideias e suas combinações para elaboração de conceitos, até o agrupamento de ideias, permitindo organizá-las hierarquicamente em árvores (DUIN et al., 2010).

Algumas das funcionalidades disponíveis nesses sistemas são: a atribuição de rótulos às ideias; avaliações; comentários; dentre outros aspectos, como a elaboração de fóruns de discussão (ANDROSCH; REDL, 2017).

Os sistemas de gerenciamento de ideias possibilitam às organizações estabelecerem relações entre os problemas que desejam resolver, e o processo de inovação. Utilizando esses sistemas, é possível disponibilizar aos usuários mecanismos para sugestão e avaliação de ideias, ou publicação de suas opiniões sobre as ideias de outros usuários (SALDIVAR et al., 2016).

As plataformas de gerenciamento de ideias tem provado a capacidade de realizar campanhas alcançando uma larga escala de pessoas, abrangendo de forma mais efetiva o público interessado (SALDIVAR et al., 2016).

Apesar da facilitação de coleta e organização das ideias proporcionadas pelos IMS, destaca-se, ainda, a necessidade de avaliá-las. Além do grande volume de ideias, um fator dificultador na interpretação destas, é o entendimento de que esses dados, quando analisados individualmente, podem não apresentar características relevantes (CARAYANNIS; MEISSNER; EDELKINA, 2017).

Isso ocorre porque as pessoas tendem a basear suas escolhas e sugestões em suas experiências anteriores. Consequentemente, a primeira versão de uma ideia tende a ser uma elaboração simplória sobre um determinado contexto. No entanto, quando essas ideias são combinadas e visualizadas em grupos similares, podem potencializar umas às outras (MAGNUSSON; NETZ; WÄSTLUND; 2014).

Dentre as vantagens agregadas ao processo de avaliação das ideias, é possível destacar a viabilização da análise de comportamento em redes sociais. Esta, permite o relacionamento entre a quantidade de ideias, a conectividade dos contribuintes e os comentários realizados. Essas características podem possuir influência sobre a identificação de ideias de alto valor agregado (SALDIVAR et al., 2016).

## 1.1 DEFINIÇÃO DO PROBLEMA

Por conta do aumento da capacidade de expressão das pessoas proporcionada pelos meios de comunicação e aliado ao aumento contínuo do número de pessoas conectadas, novos mercados e oportunidades tem surgido. Uma das grandes inquietudes das organizações é explorar oportunidades de negócios que as mantenham competitivas no mercado (SÉRGIO; 2016, ALESSI et al., 2015).

Por vezes, essas oportunidades são oriundas de novas ideias geradas em ambientes externos à organização. Essas ideias podem representar desejos ou insatisfações de clientes, parceiros ou investidores (DUIN et al., 2010).

Para que essas oportunidades possam ser aproveitadas pelas organizações, torna-se evidente a necessidade de realizar a captação dessas novas ideias de forma mais eficiente, possibilitando a seleção e implementação das propostas que melhor se alinham com as diretrizes organizacionais (ALVAREZ, 2018).

A criação de canais colaborativos é uma forma alternativa de captação de ideias. A colaboração implica desafios para todos os atores envolvidos, particularmente para as empresas que estão propensas a abrir sua base de conhecimento e integrar sua força a times externos. Este processo, requer confiança, um aspecto de difícil desenvolvimento, com alto grau de incertezas sobre as premissas de propriedade intelectual (JIMÉNEZ-NARVAEZ; GARDONI, 2015).

Outro desafio substancial é análise de novas ideias, normalmente realizada pela avaliação de especialistas. Esta atividade se torna ainda mais desafiadora mediante o aumento exponencial do número de ideias registradas, dificultando a utilização dessas informações pelas organizações (ALVAREZ, 2018).

Quando um volume expressivo de ideias é coletado, é possível que existam registros semelhantes ou que se repitam. A existência de múltiplas ideias com especificidades diferentes, porém, com propostas similares pode representar uma tendência ou necessidade. Para identificar essas ocorrências, é necessário realizar um trabalho de classificação das mesmas. Esse processo pode tornar possível englobar essas ideias em grupos, potencializando a identificação das mais interessantes (MAGNUSSON; NETZ; WÄSTLUND; 2014).

No entanto, o processo de classificação e análise de ideias demanda um trabalho exaustivo. Muitas vezes, parte das ideias obtidas nas comunidades de *crowdsourcing* são pequenos fragmentos de uma nova ideia ou necessidade. Outras vezes, quando analisadas isoladamente, não aparentam ser significativas. Considerando o potencial da geração de ideias proporcionada pelas comunidades *online* de *crowdsourcing*, é possível sugerir que a capacidade de análise e síntese dessas ideias pelas empresas tem se apresentado obsoleta (ALVAREZ, 2018).

Como forma de amenizar o impacto gerado pelo grande volume de dados, muitas das vezes não estruturados, a literatura apresenta métodos de tratamento de dados. Uma das formas possíveis de contribuir com esse processo, é o agrupamento das ideias por "seguimento de mercado", facilitando a avaliação das ideias. (SALDIVAR *et al.*, 2016).

Entretanto, a diversidade de ideias torna o agrupamento do conjunto de dados uma tarefa desafiadora (MAGNUSSON; NETZ; WÄSTLUND; 2014). Apesar da complexidade agregada, as técnicas de análise de agrupamento de dados, representam uma alternativa importante para o processo de análise de ideias, agrupando o conteúdo segundo seu grau de semelhança (*clustering*) para sintetização de dados (CHO, 2016; SERGIO; DE SOUZA; GONCALVES, 2017).

Uma complexidade comumente encontrada nesse processo é a determinação de uma medida de similaridade adequada. Estabelecer corretamente a medida de similaridade é primordial para a geração de grupos densos que representem, de fato, apenas ideias que tenham uma conexão relevante entre si (ALVAREZ, 2018).

Dessa forma, os métodos de análise de agrupamentos necessitam ser explorados, aplicados, experimentados e ajustados. Assim, é possível estabelecer e otimizar métodos de agrupamento de ideias capazes de contribuir, positivamente, na construção de um modelo para o processo de Gestão de Ideias (SERGIO; DE SOUZA; GONCALVES, 2017; ALVAREZ, 2018).

Uma vez que as ideias obtidas em comunidades *online* apresentam características muito subjetivas, observa-se que o conjunto de dados possui uma característica que impõe grande desafio ao processo de agrupamento, ou seja, tais conjuntos possuem alta esparsidade. Considerando as características do cenário abordado nesse trabalho, em que se deseja obter grupos cujas ideias sejam homogêneas entre si, bem como heterogêneas quando comparadas às ideias de outros grupos (EVERITT et al., 2011); entende-se que esse cenário pode ser atendido mediante a aplicação de algoritmos de agrupamento.

Nesse tipo de algoritmo, quanto maior a semelhança entre dois elementos (ideias) de um vetor (conjunto de dados), mais próximos os elementos estão entre si. Um conjunto de elementos próximos uns dos outros caracteriza um grupo (*cluster*).

Como forma de melhorar o resultado do agrupamento, Alvarez (2018) sugere utilização de mecanismos para estabelecer o *ranking* das ideias. Um dos mecanismos sugeridos é a utilização de filtros de temporalidade. Aliados aos comentários e demais dimensões do vetor de ideias, esses mecanismos podem auxiliar a compreender se as ideias com melhor *ranking* tendem a ser implementadas.

A avaliação da utilização de *rankings* também é sugerida por Weiss (2017), que destaca a oportunidade de realização de trabalhos futuros para estabelecer métodos e padrões para o gerenciamento de ideias. Esses métodos poderiam estabelecer modelos para uma melhor ordenação de ideias e gerenciamento de grandes volumes de dados; como, por exemplo, os dados obtidos por intermédio de projetos de inovação aberta.

A partir do contexto apresentado, esta dissertação objetiva responder a seguinte questão de pesquisa: **“Como disponibilizar grupos de ideias cuja relevância desses grupos possibilite auxiliar especialistas na avaliação e na gestão de ideias?”**.

## 1.2 OBJETIVOS

A seguir são apresentados os objetivos geral e específicos deste trabalho.

### 1.2.1 Objetivo Geral

Propor um modelo voltado à avaliação e gestão de ideias por meio de técnicas de análise de agrupamentos.

### 1.2.2 Objetivos Específicos

- Identificar quais elementos, além do texto da ideia, podem auxiliar na determinação de um índice de relevância para cada grupo de ideias;
- Desenvolver um protótipo que permita avaliar a viabilidade do modelo proposto considerando um cenário no âmbito da Gestão de Ideias;
- Realizar uma análise dos resultados alcançados considerando o cenário formulado.

## 1.3 JUSTIFICATIVA E RELEVÂNCIA DO TEMA

Com o advento da *internet* e os avanços tecnológicos, as organizações passaram a enfrentar novos desafios para se manterem competitivas no mercado. Para isso, fez-se evidente a necessidade de investimento nas áreas de Pesquisa, Desenvolvimento e Inovação (PD&I) (BREM; VOIGT, 2009; ALVAREZ, 2018).

No que tange à Gestão de Inovação tem se tornado necessário promover as inovações, sejam elas de produto, processo, organizacional ou de marketing de forma direcionada à atender as necessidade e expectativas dos *stakeholders* (OECD – MANUAL DE OSLO; 2005, p. 55 - 59).

Na expectativa de atender essa demanda, algumas empresas tem envolvido seus parceiros e clientes no processo de Gestão de Ideias suportado por portais de *crowdsourcing*, por intermédio dos quais as ideias, sugestões, expectativas e necessidades são coletadas e podem ser analisadas (QIN et al., 2016, ).

O conceito de *crowdsoucing* se baseia na ideia de que qualquer pessoa pode contribuir com uma solução potencialmente viável ou sugestão, para solucionar um problema. O *crowdsourcing* pode ser



considerado um modelo baseado na combinação de trabalho humano e computacional (MAO et al., 2017).

Utilizando os portais de *crowdsourcing*, expõe-se um tema para discussão, ideação e avaliação a uma larga escala de usuários. Posteriormente, o resultado dessa exposição requer um exaustivo trabalho avaliativo para selecionar os aspectos colaborativos capazes de compor um conceito implementável (WESTERSKI; DALAMAGAS; IGLESIAS, 2013; GISH; HANSEN, 2013; MAO et al., 2017).

Com o objetivo de sustentar a crescente demanda por inovação, a integração dos modelos de gestão de ideias aos sistemas de informação apresenta ser promissora. Essa integração proporciona a manutenção do alto nível de qualidade e eficiência frente ao alto nível de atividade nas redes colaborativas. O foco na coleta, avaliação, suporte e implementação de estratégias com alcance amplo tende a otimizar o processo de gestão, oferecendo novas perspectivas ao negócio (SADRIEV; PRATCHENKO, 2014).

As plataformas modernas não somente envolvem uma diversidade de usuários, mas possibilitam, ativamente, o controle *online* de comunidades. Esse controle ocorre por intermédio de recursos sofisticados de gerenciamento de modelos, envolvendo compensação colaborativa, estabelecimento de propriedade intelectual, bem como a flexibilidade para construir projetos de direito aberto ou privado (QIN et al., 2016; SALDIVAR et al., 2016; JIMÉNEZ-NARVAEZ; GARDONI, 2015).

Uma das formas de otimizar a seleção as ideias é a indexação das colaborações, de acordo com o perfil do usuário. Uma vez estabelecido o perfil dos colaboradores de uma plataforma colaborativa, é possível também identificar *experts*. Em seguida, pode-se estabelecer a relevância de uma ideia ou comentário. Qualificar a informação, pode potencializar a identificação de colaboradores necessários para o bom desenvolvimento de projetos e serviços (MARTINEZ-TORRES; OLMEDILLA, 2016; QIN et al., 2016).

Utilizando-se de técnicas, métodos e ferramentas da Análise de Agrupamentos, esses dados podem ser processados, produzindo novas informações sumarizadas, de forma que se possa realizar uma análise pontual sobre um conjunto comum de ideias. A elaboração de modelos de análise de agrupamento de ideias tem se tornado uma demanda crescente, e a aplicação desses métodos e técnicas pode contribuir para a interpretação e visualização do conjunto de ideias coletadas em uma organização (SERGIO; DE SOUZA; GONCALVES, 2017).

Contudo, a concepção desses agrupamentos ainda precisa ser explorada em novos estudos que levem em consideração a esparsidade dos dados e a necessidade de geração de *clusters* mais densos, e com limiar de similaridade satisfatório para análise de um especialista.

Uma forma de melhorar o resultado dos agrupamentos para os especialistas de domínio, é ofertar mecanismos que possibilitem a identificação da relevância dos grupos gerados, otimizando o processo de análise. A criação de um *ranking*, ordenando os grupos a partir de critérios que representem sua grandeza ou relevância, é uma possibilidade a ser avaliada.

#### 1.4 DELIMITAÇÕES DA PESQUISA

Delimita-se esta pesquisa com o objetivo de propor um modelo de agrupamento, que possibilite a execução de tarefas de análise e seleção de ideias, nas organizações. O foco deste trabalho reside nas áreas de Gestão de Ideias e Análise de Agrupamentos através da proposição de um modelo, cujo objetivo é colaborar com o processo de Gestão de Ideias nas organizações.

O modelo proposto utiliza técnicas de processamento de texto, normalização de matrizes e análise e *ranking* de agrupamentos, não sendo seu objetivo discutir modelos e técnicas de Visualização de Dados ou o desenvolvimento de uma aplicação direcionada ao usuário final. Disponibiliza, contudo, um protótipo com o intuito de auxiliar na avaliação dos resultados obtidos no processo de agrupamento de dados.

#### 1.5 PROCEDIMENTOS METODOLÓGICOS

Uma metodologia é um sistema de princípios, práticas e procedimentos aplicados a uma área específica do conhecimento. O objetivo dos procedimentos metodológicos é descrever a pesquisa realizada, tornando-a reproduzível e classificando-a em vários aspectos.

Para realizar esta pesquisa optou-se pela utilização da metodologia de Peffers et al. (2007), a *Design Science Research* (DSR). A DSR pode ajudar pesquisadores de sistemas de informação a produzir, e apresentar, pesquisas de alta qualidade (PEFFERS et al., 2007).

Este trabalho está caracterizado na linha de pesquisa Tecnologia Computacional, e tem como propósito, o desenvolvimento de um modelo computacional voltado ao agrupamento de dados e *ranking* dos grupos gerados. Por conta da alta aderência da DSR com as disciplinas relacionadas à ciência da computação, essa metodologia pode ser

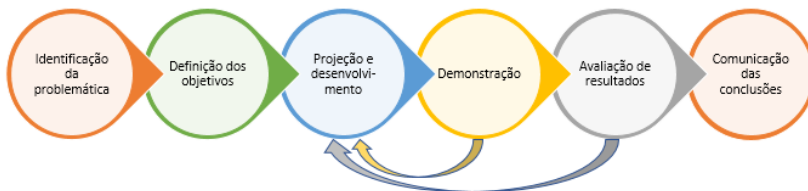
utilizada para a criação de artefatos inovadores na solução de problemas relacionados à área de sistemas de informação (VAISHNAVI; KUECHLER, 2004).

Por intermédio da implementação de sistemas de informação, a metodologia DSR pode ser utilizada para desenvolver mecanismos capazes de justificar, explicar ou prever fenômenos organizacionais e humanos (HEVNER et al., 2014). Ao aplicar DSR na elaboração de sistemas de informação, objetiva-se elaborar sistemas com maior eficiência e eficácia. Para tanto, procura-se estimular a inovação de ideias, produtos, processos e capacidades técnicas.

O modelo representado na Figura 1 está estruturado de forma nominal em seis atividades sequenciais:

- Identificação da problemática: Nessa atividade, busca-se compreender a definição do problema e demonstrar relevância do problema, assegurando a plena compreensão desse;
- Definição dos objetivos: Nessa atividade é realizada a definição dos objetivos do trabalho em relação à problemática apresentada, afim de identificar, analisar e propor soluções capazes de satisfazer à necessidade reportada;
- Projeção e desenvolvimento: Nessa etapa deve ocorrer a elaboração do projeto e o desenvolvimento do objeto de estudo, buscando produzir soluções satisfatórias para este trabalho;
- Demonstração: Nessa atividade é realizada a demonstração do artefato desenvolvido, justificando a aplicação desse utilizando simulações e validações;
- Avaliação de resultados: Nessa atividade é realizada a avaliação dos resultados obtidos, bem como, analisado o comportamento do artefato em relação ao problema definido. Para tanto, são efetuadas análises comparativas entre resultados alcançados e objetivos pré-estabelecidos;
- Comunicação das conclusões: Após a análise dos resultados, é realizada a comunicação das conclusões e apresentação dos trabalhos futuros.

Figura 1 – Modelo de processo da metodologia DSR



Fonte: Autor

Na Figura 2 está apresentada a equivalência das atividades propostas na metodologia DSR; relacionando-as com as seções estabelecidas para a estrutura deste trabalho, evidenciando assim, o uso da metodologia.

Figura 2 – Metodologia x estrutura do trabalho



Fonte: Autor

## 1.6 ADERÊNCIA E INTERDISCIPLINARIDADE

O Programa de Pós-Graduação em Tecnologias da Informação e Comunicação – PPGTIC é um programa em nível de mestrado acadêmico estruturado na área de concentração **Tecnologia e Inovação** (PPGTIC, 2017a). O curso está dividido em três linhas de pesquisa,

sendo: Tecnologia, Gestão e Inovação, Tecnologia Educacional e Tecnologia Computacional (PPGTIC, 2017b).

Este trabalho está em consonância com a linha de pesquisa Tecnologia Computacional, cujo propósito é “desenvolver modelos, técnicas e ferramentas computacionais, auxiliando na resolução de problemas de natureza interdisciplinar. Especificamente, essa linha de pesquisa procura desenvolver novas tecnologias computacionais para aplicação nas áreas de educação e gestão” (PPGTIC, 2017b). Considerando o objetivo da linha de pesquisa em que o trabalho está enquadrado, pode-se julgar que o objeto de estudo e os objetivos propostos neste trabalho atendem aos requisitos solicitados pela linha de pesquisa selecionada. Isso pode ser corroborado, uma vez que o trabalho propõe um modelo de cunho computacional, baseado nos conceitos de análise de agrupamentos, para auxiliar nas etapas de seleção e gestão de ideias, as quais são fundamentais no processo de inovação das organizações.

Dentre os trabalhos já publicados no programa, destaca-se a dissertação de mestrado de ANACLETO (2017) intitulada “Um modelo baseado em Análise de Sentimentos como suporte à Sistemas de Recomendação”, que propunha promover a análise de sentimento em comentários que possam impactar nos resultados oferecidos pelos Sistemas de Recomendação. De acordo com o autor, esta abordagem poderia ser aplicada no contexto de Gestão de Ideias.

## 1.7 ESTRUTURA DO TRABALHO

Esta dissertação está estruturada em seis capítulos. O primeiro capítulo é composto pela introdução, definição do problema e apresentação dos objetivos da pesquisa. Subsequentemente, está apresentada a justificativa e relevância do tema. Apresenta, ainda, delimitações da pesquisa, procedimentos metodológicos, e aderência e interdisciplinaridade referente ao programa de pós-graduação, finalizando com a estrutura do trabalho.

O segundo capítulo apresenta o referencial teórico, abordando o tema “Gestão de Ideias”. O terceiro capítulo apresenta os conceitos principais da área de Análise de Agrupamentos. Na sequência, o quarto capítulo apresenta o modelo proposto, realizando uma descrição detalhada sobre as etapas que o compõem.

O quinto capítulo apresenta uma discussão e avaliação dos resultados obtidos por meio do modelo proposto, respeitando as

delimitações do cenário de estudo. Por fim, o sexto capítulo apresenta as considerações finais e os trabalhos futuros.

## 2 GESTÃO DE IDEIAS

Neste capítulo estão apresentados os conceitos relacionados ao tema “Gestão de Ideias”, apresentando desde os conceitos básicos do termo “ideias”, seu processo de gestão, e a apresentação de métodos para realização dessa gestão. Por fim, tem-se a introdução à temática do uso de TIC na Gestão de Ideias.

### 2.1 INTRODUÇÃO À GESTÃO DE IDEIAS

Nos dias hodiernos, o termo “ideia” é compreendido pelas pessoas como uma imagem mental, existente ou em construção, na mente de um indivíduo. Essas ideias, quando discutidas, podem gerar ou modificar conceitos. Esses, por sua vez, podem ser considerados a base para a construção do conhecimento (DUIN et al., 2010).

No que tange à filosofia, o conceito de ideia pode ser subentendido com base nas perspectivas apresentadas por Platão, Descartes e Locke, como imagens mentais e abstratas, cujo objetivo é representar qualquer objeto do pensamento humano (DUIN et al., 2010).

Segundo Ribeiro (2018, p. 28) a ideia é “uma ocorrência de atividades mentais, baseadas nas habilidades e conhecimentos disponíveis, que resultam em novos pensamentos, conceitos, entendimentos ou atitudes”.

As ideias podem ser representadas através de uma diversidade de formas de comunicação; desde a forma escrita ou desenhada, até a utilização de gestos. Considerando que as representações de uma ideia podem conter poucos ou muitos detalhes, a comunicação e interpretação dessas ideias são fundamentais para a concepção de valor (RIBEIRO, 2018).

A capacidade de inovação de uma organização está relacionada à estrutura organizacional e à forma como as novas ideias são comunicadas e discutidas na empresa diariamente. Esse potencial criativo pode ser expandido, mediante a abertura dos canais de captação de ideias, inspirações e fontes de conhecimento, utilizando redes sociais colaborativas (QIN et al., 2016).

Somente com um grande número de produtos inovadores, as companhias poderão se manter competitivas no mercado. Tornar o processo de inovação efetivo, é crucial para esse sucesso. Para tanto, esses novos produtos precisam ser atrativos e corresponder às expectativas e necessidades dos compradores (QIN et al., 2016; ANDROSCH; REDL. 2017).

Uma inovação de sucesso inicia com uma ideia, e termina quando um novo produto é colocado no mercado. Novas ideias precisam do tempo e espaço certo para amadurecerem, o que requer que a empresa tenha a cultura de inovação. Um dos aspectos que resguarda a qualidade dos produtos da empresa, é o investimento em P&DI (ANDROSCH; REDL, 2017).

As economias emergentes têm passado por sérias dificuldades para manter sua posição no mercado, e tem sido forçadas a desenvolver ciclos curtos de inovação em meio a um mercado taxado como hiperdinâmico, incerto e caótico. Num processo interativo e colaborativo, a inovação pode ser encorajadora. Entretanto, o processo de geração de ideias requer uma estratégia de projetos capaz de envolver diversos tipos de conhecimento, compartilhando ideias e oportunidades (DUIN et al., 2010; JIMÉNEZ-NARVAEZ; GARDONI, 2015; SALDIVAR et al., 2016).

Dentre as ferramentas de Inovação Aberta, destacam-se as ferramentas de *crowdsourcing* que agregam os "caçadores de soluções" aos "solucionadores de problemas" utilizando conexões virtuais de redes adaptadas às necessidades das empresas, estendendo seu alcance às massas de *stakeholders* através de *blogs*, fóruns de discussão, dentre outras ferramentas *online* cabíveis (CARAYANNIS et al., 2017).

Para uma organização, a agregação de valor à uma ideia é uma tarefa imprescindível, de forma que, torna-se necessário a utilização de métodos de Gestão de Ideias para a realização dessa atividade (RIBEIRO; 2018).

A literatura apresenta a Gestão de Ideias como processo de obtenção, análise, sintetização e seleção de ideias. O objetivo desses métodos é aprimorar tanto a identificação de oportunidades e valores num processo criativo de inovação, quanto a entrega de um novo ou melhorado produto, serviço ou processo no mercado (ALVAREZ, 2018, COOPER, 1990).

A Gestão de Ideias pode contribuir no processo de transferência de tecnologia, especialmente, utilizando-se de portais colaborativos (*crowdsourcing*) para contemplar as dinâmicas estabelecidas nos primeiros estágios da conceptualização do produto, mediante a obtenção de ideias externas durante o processo de transferência (JIMÉNEZ-NARVAEZ; GARDONI, 2015).

De acordo com Koehn et al. (2001), a Gestão de Ideias é uma etapa da *front-end* da inovação (Seção 2.3), capaz de potencializar a vantagem competitiva das organizações. Essa colaboração da Gestão de Ideias no



campo de Gestão de Inovação é um dos mais evidentes benefícios para uma organização.

Para realizar uma Gestão de Ideias de forma efetiva, é necessário adotar uma metodologia. A literatura apresenta alguns modelos que podem nortear esse processo. Dentre os modelos identificados na literatura, destacam-se: *Stage-Gate Systems* (COOPER, 1990), *New Concept Development Model* (KOEN et al, 2001), *Integrated Idea Management Model* (BREM; VOIGT, 2007), *The Technology Development Model* (WHITNEY, 2007) e *Crowdsourcing* (QIN et al., 2016).

## 2.2 STAGE-GATE SYSTEMS

De acordo com Cooper (1990), a habilidade de conduzir um bom processo de inovação, colocando no mercado novos produtos originários de uma ideia, e com poucos erros, é a chave do sucesso. Para conduzir esse processo, Cooper (1990) apresenta um modelo denominado *Stage-Gate Systems*. Em seu tempo, os modelos baseados em *Stage-Gate* são apresentados como uma tendência para modelos de gestão de inovação corporativa, agregando efetivas ferramentas de gestão, direção e controle do processo de produção inovadora.

Os sistemas baseados no modelo *Stage-Gate* consideram que a inovação de produtos é um processo, podendo ser, assim, gerenciada. Um processo pode ser subdividido em uma série de estágios ou etapas. Entre cada etapa é necessário estabelecer um controle de qualidade (*gate*).

Os estágios representam onde o trabalho de fato é realizado, e os portões (*gates*), buscam garantir que o trabalho realizado possui os requisitos mínimos de qualidade desejados antes que o processo avance uma etapa. Sempre que um projeto avança de estágio, ele é melhorado; logo, o risco se torna ainda mais gerenciável.

Cada estágio possui objetivos específicos e pressupõe a execução de uma série de atividades, cujos entregáveis se tornam as entradas para o portão subsequente. Além do conjunto de entradas, cada portão é caracterizado por uma série de critérios que correspondem à validação dos objetivos da etapa anterior, bem como uma saída, normalmente a *Go/Kill/Hold/Recycle*.

Nesse sentido, o conceito de *Stage-Gate Systems* objetiva promover a qualidade com foco. Os portões buscam garantir que nenhuma atividade crítica foi esquecida ou ignorada. O resultado, além

de um processo completo, constitui-se da não produção de erros críticos por omissão ou falhas.

Na Figura 3, um tradicional fluxograma de *Stage-Gate Systems* para uma empresa de manufatura é apresentado. Após a concepção de uma nova ideia, é realizada a submissão dessa ao primeiro portão. Nesse primeiro portão, é realizada uma visualização inicial da ideia. Esse é o momento em que, de fato, o projeto nasce. Caso aprovado nessa etapa, a ideia é enviada para o próximo estágio: a fase de avaliação.

A fase de avaliação, não é um estágio com alto custo agregado, porém, é crucial para a identificação dos méritos técnicos e das condições comerciais para o novo produto. Nessa etapa, são realizadas pesquisas bibliográficas, consultas aos usuários centrais, e até mesmo, eventuais provas de conceito. Para avançar à etapa seguinte na qual será realizada a investigação detalhada, a ideia precisa ser aprovada pelo segundo portão. Nesse portão, praticamente ocorre a revalidação dos itens do primeiro portão, contudo, com informações mais consistentes.

Figura 3 – Fluxograma de um *Stage-Gate System* tradicional



Fonte: Adaptado de Cooper (1990)

No segundo estágio, objetiva-se a execução de uma investigação detalhada, verificando estudos de caso, viabilidade econômica e de mercado, e determinando se a ideia proposta atende às necessidades e preferências do consumidor. Essa etapa precisa ser realizada de maneira criteriosa, pois irá fornecer as informações para tomada de decisão baseada em estudos de caso.

A validação baseada em estudos de caso (Portão 3) é a última oportunidade para filtrar as ideias que serão de fato desenvolvidas ou abortadas, antes do início do processo de produção. Nesse portão, é necessário rever todas as atividades do estágio anterior, desde a

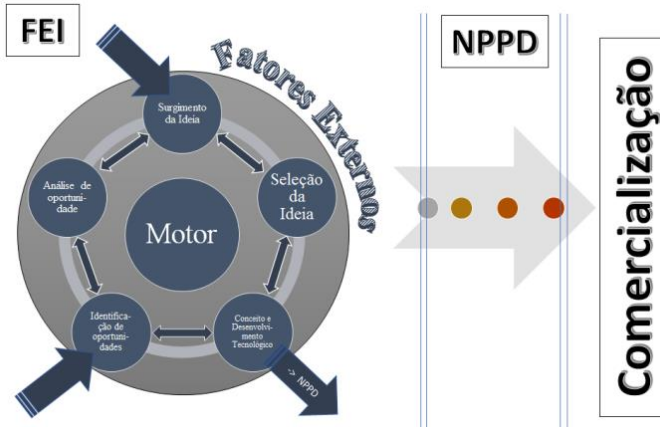
definição do conceito do produto até a especificação do posicionamento estratégico do produto no mercado.

Uma vez aprovada, a ideia avança para o estágio 3: o desenvolvimento. Nessa etapa, o produto é de fato produzido, testado, e são elaborados os planos de lançamento no mercado. Finalizado o desenvolvimento, é realizada uma nova revisão (Portão 4), com o objetivo de assegurar que o produto se mantenha atrativo para o mercado.

O estágio de validação (Estágio 4) serve para testar a completa viabilidade do projeto, do produto, do processo produtivo, da aceitação do consumidor e das finalidades e metas econômicas estabelecidas para o projeto. Uma vez realizadas as validações estabelecidas no projeto, é revisado o estágio de validação, com foco na qualidade das atividades. Esse é o último portão antes da comercialização, portanto, requer, inclusive, a revisão dos planos de mercado e operacionais, conduzindo a ideia ao seu estágio final: a comercialização.

### 2.3 NEW CONCEPT DEVELOPMENT MODEL

Apresentado na Figura 4, o Novo Modelo de Desenvolvimento de Conceito (NCD - *New Concept Development Model*) foi desenvolvido com o objetivo de providenciar uma linguagem comum e ideias para as atividades do *Front End* da Inovação (FEI) demonstrado na Figura 5. De acordo com Koen et al. (2001) as atividades descritas no modelo NCD antecedem às atividades de um *Stage Gate*.

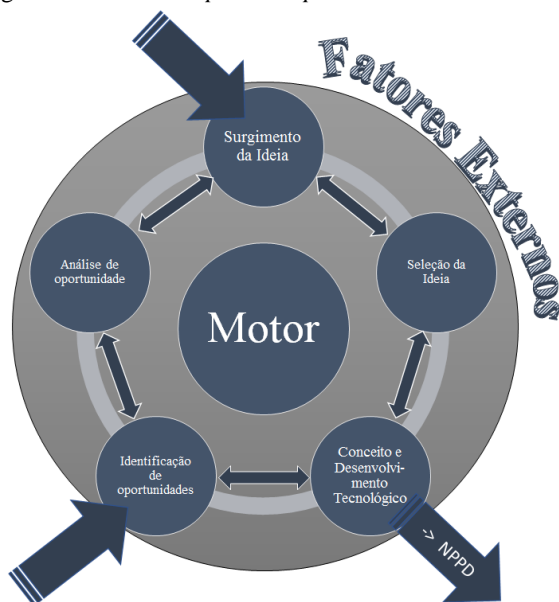
Figura 4 – *New Concept Development Model*

Fonte: Adaptado de Koen et al. (2001)

Para Koen et al. (2001), as atividades do FEI (*Front End of Innovation* - Figura 5), são opostas às descritas no modelo FFE (*Fuzzy Front End*). Esses modelos são opostos pois no FFE o processo de inovação é considerado misterioso e, conseqüentemente, há dificuldade em determinar o responsável por gerenciar as atividades. Já as atividades inerentes ao FEI podem ser devidamente estabelecidas e controladas, pois entende-se que o processo de inovação é gerenciável.

Esse modelo é dividido em três partes: a) os cinco elementos do *Front End*; b) o motor que alimenta os elementos; c) e os fatores de influência externa. Na Figura 5, essas partes estão apresentadas.

Figura 5 – *New Concept Development Model*



Fonte: Adaptado de Koen et al. (2001)

Para Koen et al. (2001), o FFE sugere incorretamente que as atividades do *Front End* não podem ser gerenciadas, portanto, o FEI propõe elementos interativos para o gerenciamento dessa fase inicial do processo de inovação, cujo resultante é a entrada do NPPD (*New Product Process Development*). O NPPD é um processo equivalente ao *Stage Gate*, no entanto apresenta também enfoque na concepção e desenvolvimento de novos processos, enquanto o *Stage Gate* se propõe mais especificamente ao desenvolvimento de produtos.

A área interna em cinza, na Figura 5, representa os cinco elementos que compõe o FEI. A área central representa a força motriz que movimenta os elementos, sendo que este “motor” é abastecido pela liderança e cultura organizacional.

A área ao entorno da figura representa os fatores externos à organização, como o ambiente no qual a organização se encontra, as capacidades organizacionais, estratégias de mercado, dentre outros fatores que podem influenciar tanto na ideação quanto num NPPD.

### 2.3.1 Os Cinco Elementos do *Front End of Innovation*

De acordo com Koen et al. (2001), a área interna do modelo FEI foi especificada como elementos da geração de uma ideia, não como um processo, de forma que as ideias possam circular entre os cinco elementos em qualquer ordem ou direção, podendo passar em um mesmo elemento mais de uma vez.

A identificação da oportunidade (*Opportunity Identification*) é onde a organização identifica as oportunidades para metas ou negócios que deseja alcançar. Esta oportunidade deve representar para a empresa uma vantagem em um ambiente competitivo, ou a simplificação, aceleração ou redução de custos de um processo, ou ainda a proposição de uma direção completamente nova para os negócios da empresa. Neste elemento normalmente são utilizadas técnicas de *brainstorms*<sup>1</sup>, *mind mapping*<sup>2</sup>, *fishbone diagrams*<sup>3</sup>.

A análise da oportunidade (*Opportunity Analysis*) requer informações adicionais pra que uma oportunidade identificada possa ser transformada em uma oportunidade de negócios tangível e mensurável. Neste momento, são realizados esforços para análise de mercado e experimentações científicas. Normalmente são aplicadas técnicas de inteligência competitiva e *trend analysis*. Este elemento pode ser parte de um processo formal ou ocorrer de forma iterativa.

A origem da ideia (*Idea Genesis*) representa o nascimento, desenvolvimento e maturação da oportunidade na concepção de uma ideia. Neste elemento, se propõe a construção, descarte, combinação, reformulação, modificação e evolução de ideias.

A seleção das ideias (*Idea Selection*) realizada no FEI precisa ser menos rigorosa do que em um NPPD, visto que as ideias estão em uma fase inicial, logo, poderão evoluir mediante a circulação entre os elementos.

O elemento final é o Conceito e Desenvolvimento Tecnológico (*Concept and Technology Development*). As atividades deste elemento consistem na elaboração de um *business case* baseado nas estimativas do mercado em potencial, necessidades do consumidor, investimentos necessários, análise de concorrência e dos demais riscos do projeto. Em algumas organizações, esta etapa é considerada a fase inicial do NPPD.

---

<sup>1</sup> <https://www.significados.com.br/brainstorming/>

<sup>2</sup> <https://www.mindmapping.com/>

<sup>3</sup> <https://study.com/academy/lesson/what-is-a-fishbone-diagram-definition-purpose-examples.html>

## 2.4 INTEGRATED IDEA MANAGEMENT MODEL

O Modelo de Gestão Integrada de Ideias (*Integrated Idea Management Model*), apresentado por Brem e Voigt (2007), está baseado em dois pontos principais: de um lado está a visão clássica em que o alinhamento é planejado apenas para os funcionários e gestores da empresa. Do outro lado, a ideia de que a inovação que advém apenas da pesquisa e do desenvolvimento (P&D) é obsoleta.

Assim, a ativação de uma rede de colaboração externa para aquisição de conhecimento, combinada aos esforços estratégicos para a melhoria dos processos internos, está positivamente correlacionada ao crescimento das corporações.

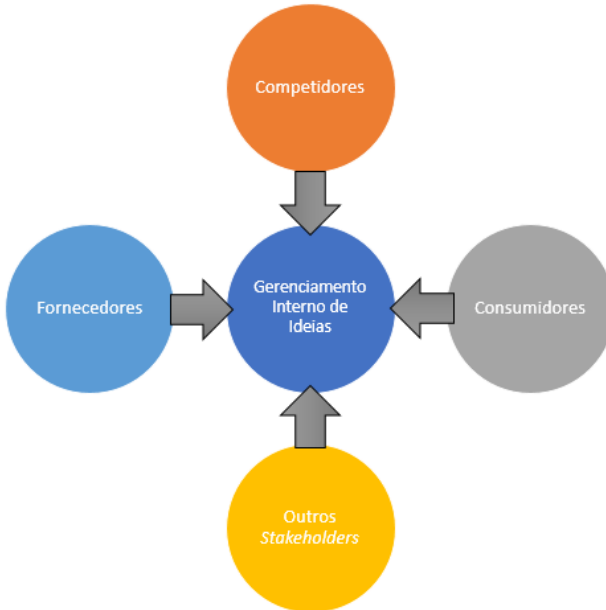
A partir dessas integrações, o principal negócio das grandes companhias se tornou uma fábrica de ideias. Nela, ocorre a combinação de informações e conhecimento, tornando o gerenciamento de ideias um inventário criativo para os processos organizacionais internos.

Para tanto, é necessária a integração de recursos externos à organização como: especialistas externos, clientes, ou usuários finais. O modelo de integração de gestão de ideias propõe um sistema de coordenação e rastreamento de ideias, não apenas para a inovação de processos, mas também, para inovações radicais de produtos, de dentro ou de fora da empresa.

Esse modelo é dividido em áreas conceituais ou “fatores” (Figura 6), sendo: Gestão Interna de Ideias (*Internal Idea Management*); fornecedores (*suppliers*); consumidores (*customers*); concorrentes (*competitors*;) e os demais *stakeholders*.

A principal tarefa do modelo de gestão integrada de ideias consiste em, sistematicamente, integrar ideias de colaboradores internos à ideias geradas pelos consumidores, fornecedores, concorrentes e demais *stakeholders*, de forma estruturada e duradoura.

Figura 6 – Cinco fatores da Gestão Integrada de Ideias



Fonte: Adaptado de Brem e Voigt (2007)

A Gestão Interna de Ideias exerce a função de integração dos demais fatores. Para que essa integração possa ocorrer, é necessário compreender todas as fontes de ideias e mantê-las motivadas, sejam elas internas ou externas à organização.

A participação dos fornecedores influencia positivamente no sucesso do processo de inovação de uma empresa, principalmente, no aspecto de alta competitividade no mercado. Essa contribuição se torna relevante, pois quem fornece detém um profundo conhecimento dos produtos, processos ou serviços ofertados e podem auxiliar na otimização do seu uso.

Além dos fornecedores, os consumidores se apresentam como excelentes parceiros e provedores de ideias. A compreensão de seus desejos, preferências ou necessidades está diretamente relacionada ao sucesso da disponibilização de um produto ou serviço no mercado.

Outro fator determinante para o sucesso do processo de inovação, é conhecer os competidores. A análise do posicionamento da concorrência na área de inovação pode representar uma vantagem



competitiva na seleção de ideias, proporcionando uma análise de inovação estratégica.

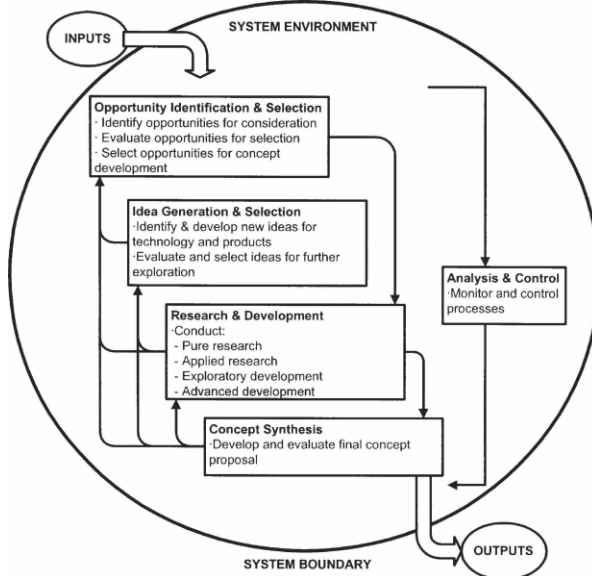
Por fim, os demais *stakeholders* são fontes consideráveis para o processo de gestão integrada de inovação. Eles são representados pela opinião da liderança corporativa, grupos de referência, políticas governamentais, associações, dentre outros. Esse grupo não é apenas importante pelo suporte ao processo de inovação, mas por representarem uma grande fonte de ideias.

## 2.5 THE TECHNOLOGY DEVELOPMENT MODEL

O modelo de Desenvolvimento Tecnológico (*The Technology Development Model - TDM*) apresentado na Figura 7, a seguir, tem seus elementos baseados no Novo Modelo de Desenvolvimento de Conceitos (NCD) (WHITNEY, 2007).

Apesar da semelhança com o FFE, o TDM se diferencia pela inclusão dos processos de P&D, visto que representa um subconjunto do ciclo de invenção, inovação e difusão, iniciando pela origem das ideias, e finalizando na elaboração de conceitos concretos para o NPPD.

Figura 7 – *The Technology Development Model*



Fonte: Whitney (2007)

Nos tópicos abaixo estão apresentados os cinco elementos que compõem o TDM, já apresentado na Figura 7.

### 2.5.1 *System Boundary*

O limite do sistema (*system boundary*) engloba as atividades de P&D, bem como, compreende o escopo de atuação do gerente desse setor numa empresa. Por vezes, requer a interação de outros atores na organização empresarial, como por exemplo, para a identificação, avaliação e seleção de novas oportunidades. Pode ser necessária a participação da equipe de *marketing* e produção por sua *expertise* e pontos de vista.

### 2.5.2 *System Environment*

O ambiente é equivalente à “engrenagem” do NCD, apresentada por Koen et. al. (2001). Nesse ponto, são observados fatores internos, como os objetivos e capacidades da empresa, bem como, o contexto e as limitações do funcionamento do sistema. Muitas limitações se originam de forças externas, como políticas governamentais, leis e regulamentações.

### 2.5.3 *Inputs*

As entradas do sistema (*inputs*) são todo e qualquer estímulo que inicie uma atividade, como por exemplo, novos objetivos de negócio, novas necessidades dos consumidores ou diretivas administrativas da empresa. Normalmente, esses estímulos são orientados à necessidade, em que para responder uma pergunta, são buscadas novas ideias, soluções ou oportunidades. No entanto, esse processo também pode ser estimulado em decorrência de uma nova ideia desenvolvida dentro do próprio sistema.

### 2.5.4 *Outputs*

O resultado final é uma nova ou aprimorada tecnologia pronta para servir como entrada para o NPPD. Esses resultados podem ser uma proposta de projeto, uma estratégia para um novo produto ou tecnologia, ou outro mecanismo que encaminhe uma tecnologia para desenvolvimento.

### 2.5.5 The Process

Os processos do TDM são semelhantes aos elementos do FEI. Eles englobam a Identificação e Seleção de Oportunidades (*Opportunity Identification and Selection*), Geração e Seleção de Ideias (*Idea Generation and Selection*), Pesquisa e Desenvolvimento (*Research and Development*), Síntese de Conceito (*Concept Synthesis*) e Análise e Controle (*Analysis and Control*).

A identificação e seleção de oportunidades é um dos possíveis pontos de entrada. É onde os conceitos orientados à necessidade são gerados e o potencial de mercado é avaliado. A geração e seleção de ideias também é uma possível entrada para esse modelo. Representa a etapa em que, com suporte tecnológico, as ideias são geradas, desenvolvidas ou desenhadas sob a perspectiva dos participantes, sejam eles internos ou externos à organização. O propósito é a geração de uma decisão estratégica para o desenvolvimento da ideia dentro de um conceito aplicável ao mercado.

A etapa de pesquisa de desenvolvimento requer atividades técnicas exploratórias para desenvolver e demonstrar as características de uma nova tecnologia e a visão conceitual do produto. Nessa etapa, são realizadas pesquisas aplicadas, desenvolvimentos exploratórios e desenvolvimentos avançados (prototipação). O objetivo principal é promover a maturação da ideia, interagindo com os demais processos durante o período de ideação.

A síntese de conceito é a última etapa para transformação de ideias e oportunidades em negócio comercializável, ou uma proposta de nova tecnologia econômica e comercialmente viável. Essa é a última etapa de manutenção das ideias antes da saída do sistema. Assim, como nas demais etapas, nesse ponto, o resultado pode ser o envio da ideia recursivamente para as etapas anteriores.

A atividade de análise e controle funciona como um mecanismo de *feedback* para o próprio sistema, em que a administração garante o progresso na direção certa. Esse processo permeia todas as demais etapas do modelo TDM. Nele estão os mecanismos de rastreamento e avaliação de produtividade, tomada de decisão, e influência nos resultados, conforme realizado pelo gerenciamento de P&D ou por um conselho de inovação.

## 2.6 CROWDSOURCING

*Crowdsourcing* é uma prática de inovação de negócios cujo objetivo é obter, de correspondentes internos e externos, serviços, ideias, conteúdos, ou inclusive, recursos financeiros, por meio da solicitação de colaboração a um vasto grupo de pessoas (*crowd*) (QIN et al., 2016).

De acordo com Goodman (2017), a utilização de *crowdsourcing* pode tornar mais fácil, mais rápido e mais barato o processo de coleta de ideias baseado em computador. Para Brabham (2008), *crowdsourcing* é o exercício da inteligência coletiva.

O conceito de *crowdsourcing* está fortemente relacionado ao de Inovação Aberta, cujos objetivos buscam terceirizar a geração de ideias inovadoras e encontrar soluções para as necessidades da Indústria 4.0, a indústria da informação (JIMÉNEZ-NARVAEZ; GARDONI, 2015).

A Inovação Aberta está baseada em quatro pontos centrais que precisam ser implementados: a) desenvolvimento de redes de comunicação; b) consolidação do conhecimento compartilhado; c) mecanismos de proteção de propriedade intelectual e; d) um modelo de negócio de apoio (MCCORMACK; FALLON; CORMICAN, 2015).

Nesse contexto, a Inovação Aberta apresenta desafios substanciais. Entretanto, também proporciona novas formas de interação entre os atores envolvidos em projetos abertos e na sociedade. Baseado na ideia de que qualquer pessoa, potencialmente, pode contribuir com uma solução viável ou sugestão para solucionar um problema, esse método propõe a exposição de um tema, para discussão, ideação e avaliação, a uma larga escala de usuários de uma plataforma de *crowdsourcing* (JIMÉNEZ-NARVAEZ; GARDONI, 2015).

As redes colaborativas de inovação aberta são capazes de amplificar o potencial criativo das companhias, obtendo mais ideias e melhorando, inclusive, a comunicação interna entre os departamentos. A coleta de grandes quantidades de dados pode auxiliar na descoberta de padrões relevantes e refinar hipóteses (GOODMAN, 2017).

Esse modelo possibilita às organizações que maximizem suas capacidades e oportunidades de inovação por intermédio da abordagem colaborativa. Contudo, a colaboração aberta implica desafios para todos os atores envolvidos, particularmente, para as empresas que estão propensas a abrir seu *know how* e integrar sua força a times externos (JIMÉNEZ-NARVAEZ; GARDONI, 2015; QIN et al., 2016).

Estabelecer o perfil dos colaboradores de uma plataforma pode ser uma tarefa vantajosa, uma vez que possibilita a identificação dos *experts* numa determinada temática e, conseqüentemente, a relevância

de uma ideia ou comentário. Identificar essas pessoas pode auxiliar na identificação de colaboradores necessários para o bom desenvolvimento de projetos e serviços (DOAN; RAMAKRISHNAN; HALEVY, 2011).

O processo de seleção de colaboradores ou ideias precisa ser rigoroso, uma vez que os custos tendem a ser elevados na etapa de implementação. Isso requer seleção das ideias corretas no tempo certo, identificando, também, oportunidades emergentes. Esse processo requer confiança, um aspecto de difícil desenvolvimento em que há alto grau de incertezas sobre as premissas de propriedade intelectual (JIMÉNEZ-NARVAEZ; GARDONI, 2015; ANDROSCH; REDL, 2017).

Outro tratamento necessário é o gerenciamento de usuários maliciosos. Ideias maliciosas precisam ser detectadas, bloqueadas e removidas da base. A identificação e punição dessas ideias ou usuários, auxilia na manutenção da confiança das ideias disponíveis na rede (DOAN; RAMAKRISHNAN; HALEVY, 2011).

As plataformas mais modernas, não somente envolvem uma diversidade de usuários, mas possibilitam, ativamente, o controle *online* das comunidades. Esse controle pode ser realizado utilizando recursos sofisticados de gerenciamento de modelos, envolvendo compensação colaborativa, estabelecimento de propriedade intelectual, bem como, a flexibilidade no estabelecimento de projetos de direito aberto ou privado (DOAN; RAMAKRISHNAN; HALEVY, 2011; QIN et al., 2016).

Atualmente, empresas como P&G<sup>®</sup>, Fiat<sup>®</sup>, Amazon<sup>®</sup>, Dell<sup>®</sup>, Starbucks<sup>®</sup> e Boeing<sup>®</sup>, buscam ideias através de redes colaborativas para aperfeiçoamento e desenvolvimento de novos produtos e serviços. Nessas empresas, o *crowdsourcing* normalmente é realizado por avisos com chamadas públicas. Nessas campanhas as pessoas colaboram com suas ideias, respondem a questionários e as ideias melhor pontuadas são premiadas pela empresa (QIN et al., 2016).

De acordo com Qin et al. (2016), num processo de *crowdsourcing*, normalmente, os seguintes passos são executados:

- O requisitante estabelece um problema;
- O desafio é publicado por meio de empresas parceiras e campanhas;
- A comunidade, de forma individual ou coletiva, interage via *internet* realizando comentários, sugestões e avaliações;
- A organização filtra e valida os dados coletados e, eventualmente, a comunidade executa uma nova validação sobre os resultados obtidos;

- Os participantes de determinada proposta escolhida são recompensados.

A aplicação de técnicas de *crowdsourcing* requer, primeiramente, uma mudança de cultura das organizações. Nesse processo, surgem os desafios, os riscos e medos de expor suas estratégias de inovação e necessidades a uma comunidade, bem como, a capacidade de gerir e controlar uma quantidade indeterminada de soluções e propostas, que podem ser sugeridas pela multidão (JENKINS; DEUZE, 2008).

A principal barreira encontrada em empresas que almejam ampliar sua capacidade de inovação é o medo da mudança. Muitas empresas baseadas em inovação fechada, resguardam-se em iniciar um processo de inovação aberta, pois estariam sujeitas a repensarem suas estratégias de mercado (QIN et al.; 2016).

Além da resistência à mudança, outras barreiras são identificadas na literatura, como garantia de propriedade intelectual, suporte governamental, regulamentações legislativas, modelos de negócios, aumento da complexidade de gerenciamento e supervisão dos projetos, dificuldade de identificação de redes de comunicação adequadas e aumento da dependência de recursos externos (JIMÉNEZ-NARVAEZ; GARDONI, 2015; MCCORMACK; FALLON; CORMICAN, 2015; QIN et al., 2016).

Utilizando-se dos recursos da *internet*, a população em geral tem sido convidada a contribuir no processo de criação de novos produtos e serviços, baseados em suas necessidades. A Gestão de Ideias tem o potencial de descobrir ideias de valor que possam ser conduzidas num processo de inovação (SALDIVAR et al., 2016).

Portanto, a utilização de comunidades *online* para potencialização das práticas de Inovação Aberta tem se tornado mais comuns e necessárias. Apesar disso, a garantia da propriedade intelectual, assim como, a identificação de parceiros comerciais devidamente capacitados ou aderentes ao modelo, ainda são barreiras que precisam ser transpostas (MCCORMACK; FALLON; CORMICAN, 2015).

## 2.7 SISTEMAS DE GESTÃO DE IDEIAS

Com os avanços tecnológicos, o potencial de comunicação das massas aumentou de forma significativa. Para realizar uma gestão de ideias eficiente, tornou-se necessário captar os desejos e intenções do mercado. Captar esses desejos, filtrá-los, selecioná-los e desenvolvê-los é, sem dúvida, uma tarefa nada trivial (ALVAREZ, 2018).

A aplicação de TIC nesse processo, tornou-se indispensável. Um exemplo clássico do uso dessas tecnologias é o desenvolvimento de Sistemas de Gestão de Ideias. Com o uso dessas tecnologias, o processo de geração de ideias tem sido realizado de forma colaborativa, em que pessoas conectadas, pela rede e por iniciativa própria, interagem entre si, colaborando no processo de ideação, e compartilhando ideias em torno de um tema (DUIN et al., 2010).

Entretanto, construir comunidades *online* sustentáveis é um grande desafio que requer o entendimento das pessoas envolvidas e de suas necessidades, bem como, a combinação das características do usuários e do propósito da comunidade (SALDIVAR et al., 2016).

Claramente, isso gera uma demanda de gerenciamento dessas ideias, de forma que dentre as colaborações, seja possível identificar oportunidades para a etapa de implementação. Com o intuito de suprir essa demanda, surgiram os Sistemas de Gerenciamento de Ideias (SIGs), do inglês *Idea Management Systems (IMS)* (DUIN et al., 2010).

A implementação de SIGs agrega alguns desafios, como a análise das ideias compartilhadas e as estratégias de colaboração adotadas pelos outros envolvidos que propõem ideias. A análise dos comentários das ideias também pode exercer influência direta na identificação de ideias de alto valor agregado (JIMÉNEZ-NARVAEZ; GARDONI, 2015; SALDIVAR et al., 2016).

Essa gestão possibilita a organização de ideias hierarquicamente em árvores, bem como, a atribuição de rótulos às ideias, avaliações, comentários, dentre outros aspectos, como a elaboração de fóruns de discussão, por exemplo (DUIN et al., 2010).

Salvitar et al. (2016) destacam ainda que os sistemas de gerenciamento de ideias possibilitam às organizações estabelecer uma relação entre os problemas que desejam resolver e o processo de inovação. Ao mesmo tempo, permitem que os usuários sugiram ideias, avaliem-nas e publiquem suas opiniões sobre as ideias de outros usuários.

A partir dos avanços possibilitados pelo desenvolvimento da *internet*, torna-se necessário aprimorar esses recursos para que as ideias coletadas em redes de compartilhamento possam ser previamente processadas, selecionadas e qualificadas (SALDIVAR et al., 2016).

Os sistemas de gerenciamento de ideias integrados às comunidades *online* abrangem não somente a possibilidade do registro de ideias, mas também a realização de combinações para elaboração de conceitos e agrupamento de ideias.

Em um artigo publicado pela Predictive Analytics Today<sup>4</sup> são apresentados os principais provedores de sistemas de gestão de ideias do mercado: Planbox<sup>5</sup>, Aha!<sup>6</sup>, Crowdicity<sup>7</sup>, Spigit<sup>8</sup>, Brightidea<sup>9</sup>, HYPE Innovation<sup>10</sup>, Acuvate Wave<sup>11</sup>, Qmarkets<sup>12</sup>, Wazoku<sup>13</sup>, Idea Drop<sup>14</sup>, IdeaGlow<sup>15</sup>, Wridea<sup>16</sup>, Viima<sup>17</sup>, IdeaScale<sup>18</sup>, Wellspring<sup>19</sup>, HunchBuzz<sup>20</sup>, Imaginatik<sup>21</sup>, Kindling (incorporada à Spigit), Accolade<sup>22</sup>, HYPE GO!<sup>23</sup>, Exago<sup>24</sup>, Ideawake<sup>25</sup>.

Estas plataformas tem provado a capacidade de realizar campanhas, envolvendo uma larga escala de pessoas, alcançando de forma mais abrangente o público interessado. No entanto, apesar da diversidade de soluções disponíveis, ainda são poucos os recursos para a realização do processo de avaliação de ideias.

As campanhas de *crowdsourcing* produzem fontes de dados massivas com muitas ideias de características semelhantes. Muitas das vezes, essas novas ideias são imaturas e incompletas, contudo, podem ser complementadas com outras ideias semelhantes. A medida que esse volume de dados cresce, torna-se latente a necessidade de avaliar novos métodos e recursos para acelerar e otimizar a etapa de ideação (SALDIVAR et al., 2016).

Uma das formas possíveis de auxiliar neste processo é agrupando as ideias por características em comum. A tarefa de agrupar ideias, bem como, a análise do comportamento dos participantes nas redes sociais, pode auxiliar na compreensão do relacionamento entre a quantidade de

---

<sup>4</sup> <https://www.predictiveanalyticstoday.com/top-idea-management-software/>

<sup>5</sup> <https://www.planbox.com/>

<sup>6</sup> <https://www.aha.io/>

<sup>7</sup> <https://crowdicity.com/>

<sup>8</sup> <https://www.spigit.com/>

<sup>9</sup> <https://www.brightidea.com/>

<sup>10</sup> <https://www.hypeinnovation.com/home>

<sup>11</sup> <https://acuvate.com/wave/>

<sup>12</sup> <https://www.qmarkets.net/>

<sup>13</sup> <http://www.wazoku.com/>

<sup>14</sup> <https://ideadrop.co/>

<sup>15</sup> <https://web.ideaglow.com/>

<sup>16</sup> <https://wridea.com/>

<sup>17</sup> <https://www.viima.com/>

<sup>18</sup> <https://ideascale.com/>

<sup>19</sup> <https://www.wellspring.com/>

<sup>20</sup> <https://hunchbuzz.com/>

<sup>21</sup> <https://www.imaginatik.com/>

<sup>22</sup> <https://www.sopheon.com/software/enterprise-innovation-management/>

<sup>23</sup> <https://www.hypego.net/>

<sup>24</sup> <https://www.exago.com/>

<sup>25</sup> <https://ideawake.com/>



ideias e a conectividade com os contruibuintes (ponto central) (SALDIVAR et al., 2016).

O uso de métodos e técnicas de análise de agrupamentos pode ser aplicado para identificação de padrões, agrupando ideias em grupos semelhantes (SERGIO; DE SOUZA; GONCALVES, 2017).

Na próxima sessão, estão apresentados alguns métodos de análise de agrupamentos, identificados na literatura.



### 3 ANÁLISE DE AGRUPAMENTOS

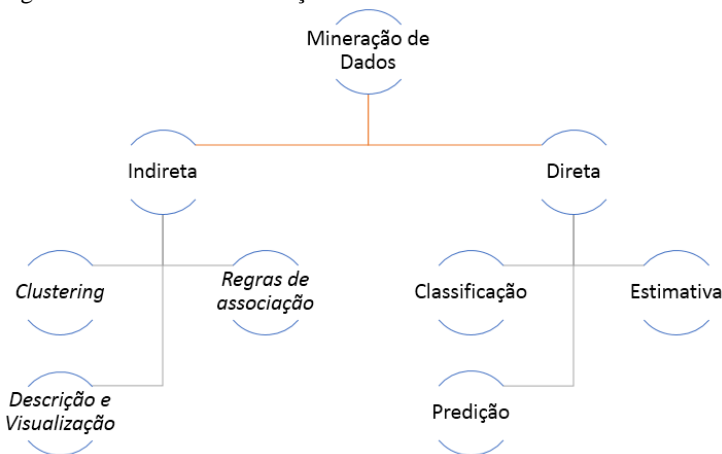
Este capítulo apresenta os conceitos básicos sobre a análise de agrupamentos, suas principais características e aplicações, bem como, os trabalhos correlatos. A análise de agrupamentos pode ser definida como uma forma de criar grupo de objetos (*cluster*), de forma que os objetos sejam semelhantes aos demais de seu agrupamento, e, conseqüentemente, diferentes dos objetos de outros agrupamentos (GAN; MA; WU, 2007).

#### 3.1 INTRODUÇÃO À ANÁLISE DE AGRUPAMENTOS

A análise de agrupamentos é um processo não supervisionado, que divide um conjunto de objetos em grupos homogêneos (GAN; MA; WU, 2007; GOLDSCHMIDT; BEZERRA; PASSOS, 2015; HUANG; et al., 2015). Também pode ser encontrada na literatura como *clustering*, análise de segmentação, análise de taxonomia, ou classificação não supervisionada (SÉRGIO, 2016).

Conforme apresentado na Figura 8, a análise de agrupamentos também é parte relevante do processo de *data mining* (mineração de dados), utilizado para exploração e análise de grandes volumes de dados. Entretanto, é válido ressaltar que, apesar de suas similaridades, as atividades de agrupamento se diferem das atividades de classificação (GAN; MA; WU, 2007).

Figura 8 – Tarefas da Mineração de Dados



Fonte: Adaptado de Gan, Ma, Wu (2007)

Conforme demonstrado na Figura 8, o processo de mineração de dados pode ser dividido em dois grandes grupos: a mineração de dados direta (*direct data mining*) e a mineração de dados indireta (*indirect data mining*).

Na mineração direta, algumas variáveis são determinadas como destinos, formando, assim, classes para categorização dos dados. Diferentemente, na mineração de dados indireta (*indirect data mining*), nenhuma característica dos elementos é pré-determinada como destino, visto que não se sabe, previamente, por qual grupo de dados se está procurando, ou quais elementos formam este grupo e por que. A análise de agrupamentos é, portanto, uma tarefa da mineração de dados indireta.

Para realizar a análise de agrupamento, alguns fatores importantes devem ser observados (Figura 9), como por exemplo: identificar os tipos de dados, realizar a conversão de escalas, padronizar (*standardization*) e transformar os dados, aplicar técnicas de visualização de dados; e determinar as medidas de similaridade e dissimilaridade (distância) que serão utilizadas.

Figura 9 – Fatores relevantes do processo de agrupamento



Fonte: Autor.

A observância dessas características implica na escolha adequada dos algoritmos de *clustering*, proporcionando que o agrupamento dos dados se torne melhor e mais correto. Os tipos de dados podem ser tipificados como discretos ou contínuos, e podem representar o grau de quantização que um atributo pode conter.

Um atributo discreto deve conter um conjunto finito de valores possíveis. Quando esse dado discreto é binário ele pode possuir apenas dois valores: “verdadeiro” ou “falso”.

Um atributo de valor contínuo pode ser definido como aquele que contém um valor entre dois limites quaisquer, como por exemplo, a aplicação de escalas de dados que indicam a importância relativa dos números. Essas escalas podem ser quantitativas, que incluem medidas de intervalo ou de razão; ou qualitativas, que podem ser nominais ou ordinais.

Entretanto, muitas vezes, as características a serem utilizadas para categorizar objetos no processo de agrupamento não estão armazenadas com o mesmo tipo de medidas ou escalas. Essas variáveis podem ser de tipos completamente distintos, de diferentes escalas ou categóricas. Por exemplo, informações de temperatura podem ser medidas em graus Celsius ( $^{\circ}\text{C}$ ), Fahrenheit ( $^{\circ}\text{F}$ ), ou Kelvin ( $\text{K}$ ); enquanto unidades de medidas de distância podem ser em metros ( $\text{m}$ ), milhas ( $\text{mi}$ ) ou polegadas ( $\text{in}$ ).

Para normalizar esses dados, possibilitando assim, o correto agrupamento dos objetos, algumas abordagens estão mencionadas na literatura, por exemplo: aplicar um coeficiente de similaridade; analisar subconjuntos de objetos envolvendo variáveis de um mesmo tipo e, em seguida, realizar a sintetização dos resultados das diferentes análises; e converter variáveis de tipos diferentes em um mesmo tipo (GAN; MA; WU, 2007; HUANG, 2008).

Esse processo de normalização ou padronização dos dados (*standardization*) se faz necessário, pois raramente os dados serão utilizados no processo de agrupamento de forma bruta. Preparar os dados para uma análise de agrupamentos requer que sejam realizadas transformações nos dados. Para realizá-las, alguns métodos conhecidos podem ser utilizados. Entre eles, pode-se utilizar média, mediana, desvio padrão, intervalo, estimativa de Huber, estimativa de peso de Tukey e estimativa de onda de Andrew (GAN; MA; WU, 2007).

Por outro lado, a visualização de dados constitui-se etapa vital do processo de análise de agrupamentos. É através das técnicas de visualização de dados que se torna possível apresentar valores sintetizados nos agrupamentos gerados. Permite, ainda, justificar os motivos pelos quais determinados objetos pertencem a um mesmo grupo, ou mesmo, apresentar uma visão espacial da distribuição dos objetos e grupos em um grafo (KEIM; KRIEGEL, 1996; ALLAN; LEOSKI; SWAN, 1997).

O processo de padronização reduz a dimensionalidade dos dados, diminuindo a complexidade de análise do conjunto. Contudo, após a padronização, é possível considerar que parte do conhecimento de localização e escala dos dados originais pode ser perdido ou alterado. Para realizar essa padronização, diversas técnicas podem ser empregadas, como por exemplo, a aplicação de técnicas de redução de dimensionalidade, escalonamento multidimensional (*multidimensional scaling* - MDS), mapas auto-organizáveis (*self-organizing maps* - SOMs), mapeamentos não lineares, mapas de preservação de classe, coordenadas paralelas e mapas de árvores (*tree maps*) (MERKL; RAUBER, 1997; VESANTO, 1999; PAMPALK; RAUBER; MERKL, 2002; VENNA, 2006).

Por fim, são apresentadas as medidas de similaridade e de distância (medida de dissimilaridade). Na literatura, essas medidas são utilizadas quantitativamente para descrever o quão próximos ou distantes dois objetos ou grupos estão entre si (MILLIGAN; COOPER, 1988).

Existem medidas para indicar a proximidade de matrizes, grafos, matrizes de dispersão, matrizes de covariância; assim como, existem medidas para vários tipos de dados, sejam eles numéricos, categóricos, binários e de tipos mistos. A consideração dos aspectos anteriormente apresentados influencia na boa escolha do método e, conseqüentemente, do algoritmo específico e adequado para a classificação e agrupamento dos dados (GAN; MA; WU, 2007).

Existem muitos algoritmos de agrupamento criados para atenderem demandas de áreas do conhecimento bastante diversificadas, como reconhecimento de padrões, inteligência artificial, tecnologia da informação, processamento de imagens, biologia, psicologia e *marketing* (GAN; MA; WU, 2007; AHMAD; STARKEY, 2017; YUNOH et al., 2016).

Durante a análise de agrupamentos, várias técnicas e métodos distintos são necessários, cada qual modelado para atender necessidades específicas, com tipos de dados de entrada distintos, assim como, as medidas de similaridade entre os elementos e/ou grupos. Em aspectos gerais, utilizando métodos automáticos ou semiautomáticos, é possível obter informações relevantes, a partir de grandes volumes de dados (BERRY; LINOFF, 2000).

### 3.2 MÉTODOS DE ANÁLISE DE AGRUPAMENTOS

A literatura apresenta diferentes formas de classificar os métodos e algoritmos de agrupamento. Dentre as formas identificadas na literatura, é possível mencionar os métodos: hierárquicos, difusos (*fuzzy*), baseados em distância, em centroides (*center-based*), em pesquisa (*search-based*), em grafos (*graph-based*), em partição (*grid-based*), em modelos (*model-based*) e em densidade (*density-based*) (GAN; MA; WU, 2007; FAHAD et al., 2014, GOLDSCHMIDT; BEZERRA; PASSOS, 2015).

### 3.2.1 Baseados em distância

Os métodos baseados em distância utilizam as medidas de similaridade para fornecer, quantitativamente, valores que expressem a correlação ou associação entre objetos. As medidas de similaridade também podem ser identificadas na literatura como coeficiente de similaridade, medidas de dissimilaridade ou distâncias (GAN; MA; WU, 2007).

As medidas de similaridade podem ser expressas de diversas formas, por exemplo, por meio de medidas de distância (euclidiana), medidas correlacionais (Pearson), medidas de associação (índice de Jaccard) e medidas de ângulo vetorial (cosseno).

As medidas de distância representam a dissimilaridade entre dois elementos de um conjunto de dados, nos quais os valores maiores denotam menor similaridade; enquanto os valores menores representam maior equivalência entre os elementos (ERDOĞAN; TIMOR, 2005).

Diferentemente das medidas de distância, as medidas de associação são diretamente proporcionais ao grau de similaridade entre dois elementos. Quanto maior a associação, mais semelhantes os elementos tendem a ser entre si. Por outro lado, quanto menor a medida, mais distintos eles se representam (REAL; VARGAS, 1996).

As medidas correlacionais utilizam a correspondência de padrões ao longo dos atributos para representar a similaridade entre os elementos. Nessa análise, não ocorre a consideração da magnitude ou representatividade dos valores dos atributos, mas o padrão global da ocorrência ou co-ocorrência desses valores, a partir da distribuição das frequências, ou pelo compartilhamento de variância (FIGUEIREDO; SILVA, 2009).

### 3.2.2 Baseados em hierarquia

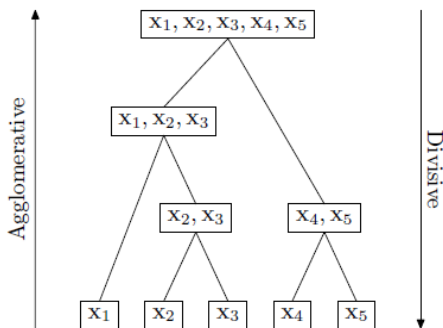
Os algoritmos baseados em hierarquia são uns dos principais algoritmos de agrupamento de dados. Esse tipo de algoritmo permite dividir o *data set* (conjunto de dados), utilizando estratégias aglomerativas ou estratégias divisivas.

Conforme representado na Figura 10, algoritmos aglomerativos iniciam o processo de agrupamento, considerando cada objeto do vetor como um *cluster* (*bottom-top*), agrupando-os por “semelhança” até que exista um grande *cluster*. Diferentemente, os algoritmos divisivos consideram o *data set* inicial como um grande *cluster* (*top-bottom*), e a



cada interação, dividem o conjunto de dados em grupos menores (GAN; MA; WU, 2007).

Figura 10 – Agrupamento hierárquico aglomerativo e divisivo

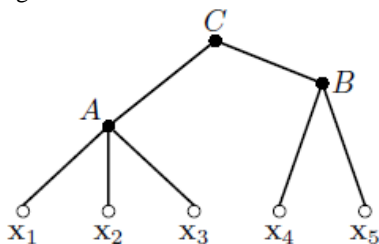


Fonte: Gan, Ma, Wu (2007)

Dentre as formas comuns de representação de *clusters* hierárquicos estão *n-tree* e o dendograma, com formato especial de uma estrutura em árvore.

Na Figura 11, está apresentada uma *n-tree*, em que o número de *n* é igual a cinco ( $n=5$ ). É um diagrama de árvore, aninhado hierarquicamente simples, que pode ser usado para representar um agrupamento hierárquico. Os nós representados por um círculo sem preenchimento simbolizam um objeto do conjunto de dados. Os nós representados por um círculo preenchido se referem a um grupo ou *cluster*, seja ele intermediário ou final.

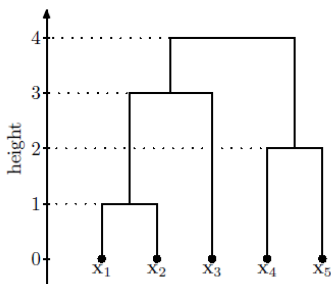
Figura 11 – 5-tree



Fonte: Gan, Ma, Wu (2007)

Um dendograma (Figura 12) é uma  $n$ -tree cujos nós internos estão associados a uma “medida” ou “altura”, que satisfaz uma condição aplicável a todos os subconjuntos da árvore.

Figura 12 – Representação de um dendograma



Fonte: Gan, Ma, Wu (2007)

### 3.2.3 Baseados em difusão

Os algoritmos baseados em difusão (*fuzzy*) como o *Fuzzy c-means* (GUSTAFSON; KESSEL, 1979) e FMLE (*Fuzzy Maximum Likelihood Estimation*) (GATH; GEVA, 1989; XIE; BENI, 1991), se enquadram numa categoria distinta de algoritmos. Esses, estão baseados no conceito apresentado por Zadeh (1965), que busca representar a similaridade entre os conjuntos de dados por meio de uma função de associação, cujos valores estão entre 0 (zero) e 1 (um).

De forma diferente do que se encontra nos métodos tradicionais de agrupamento, em que cada elemento pertence a apenas um *cluster*, nesse método, um objeto pode pertencer a mais de um *cluster*, contudo, com diferentes probabilidades indicadas. Nesse modelo, com o auxílio de uma função de pertinência, é possível aplicar uma avaliação gradual, indicando em valores unitários reais entre 0 e 1 o quanto um elemento pertence a um cluster (GUSTAFSON; KESSEL, 1979; BEZDEK; EHRlich; FULL, 1984; GATH; GEVA, 1989; XIE; BENI, 1991).

### 3.2.4 Baseados em centroides

Esse método de agrupamento pode ser considerado um dos mais tradicionais e efetivos, leia-se adequados, para realizar o agrupamento de grandes volumes de dados, bem como, *data sets* de alta dimensionalidade. Os *clusters* formados por esse modelo de algoritmo

possuem formas convexas e cada *cluster* é representado por um centro (centroide) (RADEV et al., 2004; HAN; KARYPIS, 2000; JAIN, 2010).

Um dos algoritmos mais populares e utilizados deste método é o *k-means*: um algoritmo não hierárquico criado para agrupar dados numéricos, onde cada *cluster* deve possuir um ponto central (centroide) que representa a média do agrupamento. Além de ser bastante difundido, o *k-means* possui uma série de variações, como: *k-modes*, *continuous k-means*, *compare-means*, *sort-mean*, dentre outros (GAN; MA; WU, 2007).

Tradicionalmente, esse algoritmo recebe um valor *k* que será o número fixo de *clusters* gerado. Durante o processo de agrupamento, utilizando medidas de similaridade, os elementos são, sucessivamente, alocados para os *clusters* mais próximos. Com base numa função de erro, o algoritmo executa novas interações sobre o conjunto de dados, até que a função de erro não seja alterada de forma significativa ou a associação dos clusters não sofra mais alterações (JAIN, 2010).

### 3.2.5 Baseados em pesquisa

Esse método de agrupamento busca conduzir uma pesquisa com o objetivo de identificar um *cluster* globalmente ideal, a fim de ajustar o conjunto de dados. A identificação de um *cluster* global ideal é um dos desafios conhecidos, uma vez que os algoritmos tendem a interromper as interações, mediante a identificação de uma partição local ideal do conjunto de dados (AL-SULTAN, 1995).

Para tanto, é necessário estabelecer uma função objetiva (função de avaliação). Essa função é utilizada para definir as partições de dados. Após a definição dessas, são aplicadas técnicas para identificar a partição globalmente ideal para os dados (AL-SULTAN; FEDJKI, 1997).

Apesar do objetivo desses algoritmos ser o de garantir uma solução globalmente ideal, não necessariamente, a melhor partição será encontrada (BASU; BILENKO; MOONEY, 2003; GAN; MA; WU, 2007).

### 3.2.6 Baseados em grafos

Os algoritmos de agrupamento que se utilizam de métodos baseados em grafos, para aglomerar os objetos, normalmente, a partir do *data set* de entrada, constroem esses conjuntos de dados a partir de uma matriz de dissimilaridade (GAN; MA; WU, 2007).

Para realizar o processo de agrupamento, o algoritmo Chameleon controla um grafo esparso a partir do conjunto de dados de entrada e da matriz de similaridade do banco de dados. Uma das estratégias mais utilizadas para estabelecer a representação gráfica, é a abordagem do gráfico de vizinho mais próximo de  $k$ . Posteriormente, são aplicados algoritmos aglomerativos (hierárquicos) para unir ou agrupar os “*subclusters*”, utilizando a interconectividade relativa e a proximidade dos *clusters* (KARYPIS, 1999; GAN; MA; WU, 2007).

O algoritmo CACTUS (Agrupamento categórico usando sumários), proposto por Ganti, Gehrke e Ramakrishnan (1999), utiliza abordagens diferentes para realizar o agrupamento. Inicialmente, é executada uma tarefa de sumarização, na qual são computados os valores interatributos e intra-atributos. Posteriormente, a partir desses sumários, um gráfico de similaridade é definido, por meio do qual os *clusters* são identificados. Nessa etapa, todos os atributos são avaliados para calcular as projeções que os *clusters* podem ter, gerando assim, “*clusters*” candidatos. A partir desse resultado, é executada a etapa de validação para a geração dos agrupamentos finais.

### 3.2.7 Baseados em grade

O método de agrupamento baseado em grade apresenta como vantagem a redução da complexidade computacional. A ideia central desse modelo é realizar o agrupamento utilizando estruturas de grades multidimensionais com células densas (XU et al. , 1998; LIAO; LIU; CHOUDHARY, 2004).

Conforme demonstrado na Figura 13, esse processo de agrupamento normalmente compreende cinco etapas básicas: particionar o espaço de dados em um número finito de células (ou criar a estrutura de grade), estimar a densidade celular de cada célula, classificar as células de acordo com suas densidades, identificar os centroides dos *clusters*, e identificar a passagem entre as células vizinhas (PARK; LEE, 2004; GAN; MA; WU, 2007).

Figura 13 – Etapas do agrupamento baseado em grade



Fonte: Autor

### 3.2.8 Baseados em modelos

Os métodos baseados em modelos, utilizam-se de distribuições de probabilidade para representar agrupamentos de dados distintos, por meio de combinações finitas de densidades de probabilidades, e estabelecendo propriedades geométricas de matrizes de covariância, por exemplo (BANFIELD; RAFTERY, 1993).

Nesse modelo, cada combinação de probabilidade representa um agrupamento diferente a partir dos quais os dados são identificados (FRALEY; RAFTERY, 1998). As duas principais abordagens de agrupamento desse método são: a abordagem de probabilidade de classificação e a abordagem de probabilidade de mistura (FRALEY; RAFTERY, 2002; YEUNG et al., 2001).

### 3.2.9 Baseados em densidade

A abordagem baseada em densidade pode identificar conjuntos de dados de forma arbitrária. Nesse método, a divisão dos *clusters* ocorre mediante a identificação de uma divisão espacial; de forma que os

*clusters* sejam representados por regiões altamente densas; e separados por regiões de baixa densidade (SANDER, 1998; CAO et al., 2006; KRIEGEL et al., 2011).

Figura 14 – *Cluster* baseado em densidade



Diferentemente de um *cluster* baseado em centroide, que tende a ter uma forma circular, os agrupamentos por densidade, representados na Figura 14, podem possuir formas arbitrárias (SANDER, 1998).

Uma das principais vantagens desse método é a sua capacidade de lidar com ruídos de forma eficiente. Por outro lado, há uma dificuldade em estabelecer os parâmetros para o algoritmo, como por exemplo limite de densidade (ESTER et al., 1996).

Dentro os algoritmos identificados na literatura que implementam essa estratégia, estão: BRIDGE (DASH et al, 2001), CUBN (WANG; WANG, 2003), DBCLASD (XU et al., 1998), DBSCAN (ESTER et al., 1996) e DENCLUE (HINNEBURG, 1998).

### 3.3 TRABALHOS CORRELATOS

A partir da pesquisa bibliográfica realizada nas bases de dados Scopus®, Science Direct®, IEEE Xplore® e ACM Portal®, foram analisados estudos relacionados com a problemática desta pesquisa, que também disponibilizaram soluções com aderência ao contexto apresentado neste trabalho.

Para realizar esta pesquisa bibliográfica que está descrita a seguir, foram considerados artigos publicados entre 2009 e 2018. Para elaboração dos filtros de pesquisa, foi estabelecido que os artigos deveriam estar relacionados à análise de agrupamento e gestão de ideias. Para tanto, foram elegidas as palavras-chave “*clustering*” e “*cluster analysis*” para o assunto “análise de agrupamentos” e a palavra-chave “*idea management*” para o assunto “gestão de ideias”, resultando na

seguinte expressão de consulta: ("*clustering*" or "*cluster analysis*") AND ("*idea management*").

Na busca realizada, foram encontrados 20 artigos. Desses, seis foram selecionados como correlatos ao tema deste trabalho. Os critérios de seleção utilizados foram:

- A aplicação de técnicas de análise de agrupamento deveria estar direcionada à gestão de ideias/inaçãoção;
- Disponibilidade do artigo em meio digital.

No Quadro 1 são apresentados os 20 artigos obtidos a partir da busca inicial realizada, utilizando os critérios estabelecidos nas bases de dados supramencionadas.

Dos 20 artigos encontrados, um foi removido por duplicidade, 13 foram removidos por não tratarem especificamente da aplicação de técnicas de análise de agrupamentos como ferramentas para auxiliar no processo de gestão de ideias; e seis artigos foram aceitos para a realização da revisão da literatura de trabalhos correlatos.

Quadro 1 – Resultado da busca de trabalhos correlatos

Ano	Título	Autor	Situação
1982	Scaling of corporate multivariate performance criteria subjective composition versus the analytic hierarchy process	Arrington, Jensen e Tokutani	Recusado
1988	New product forecasting models: Directions for research and implementation	Mahajan e Wind	Recusado
1992	Optimal positioning of new product-concepts: Some analytical implications and empirical results	Eliashberg e Manrai	Recusado
1993	Chapter 10 Conjoint analysis with product-positioning applications	Green e Krieger	Recusado
1994	Using purchase influence niching for better focus in industrial marketing plans: Observations and examples	Weber	Recusado
2008	Idea Management in Creative Lives	Coughlan e Johnson	Recusado
2009	Framework for analyzing and clustering short message database of ideas	Paukkeri e Kotro	Aceito
2009	Open innovation in SMEs: Trends, motives and management challenges	Van De Vrande et al.	Aceito
2012	Collective Intelligence as Community Discourse and Action	De Liddo et al.	Recusado

2012	Designing a Facilitator's Cockpit for an Idea Management System	Baez e Convertino	Recusado
2012	Idea relationship analysis in open innovation crowdsourcing systems	Westerski, Iglesias e Garcia	Aceito
2012	Innovation Cockpit: A Dashboard for Facilitators in Idea Management	Baez e Convertino	Recusado
2016	How Corporate Decisions Force Innovations: Factors and Choices to Act	Viederyte	Recusado
2016	Identifying the configurational paths to innovation in SMEs: A fuzzy-set qualitative comparative analysis	Poorkavoos et al.	Recusado
2017	Crowd-based Feature Selection for Document Retrieval in Highly Demanding Decision-making Scenarios	Pintas, Correia e Garcia	Aceito
2017	Idea Identification Model to Support Decision Making	Sergio, De Souza e Goncalves	Aceito
2017	Idea identification model to support decision making	Sergio, De Souza e Goncalves	Duplicado
2017	Towards a Reference Model for Agile New Service Development Using the Example of E-Mobility Service Systems	Lamberth-Cocca e Meiren	Recusado
2018	An organizational model to understand the creativity workshop	Barrios et al.	Aceito
2018	Supporting Reflective Use of Design Idea Archives Using Email	Inie, Dow e Dalsgaard	Recusado

Fonte: Autor

No trabalho intitulado “*Framework for analyzing and clustering short message database of ideas*”, Paukkeri e Kotro (2009) apresentam um novo *framework* para o desenvolvimento de uma ferramenta de gestão de ideias: o Note. Esse *framework* permite que um colaborador adicione um novo cartão (nota) com uma observação, ideia ou situação problema. A partir das notas adicionadas, a ferramenta proposta é capaz de reunir, promover e gerenciar ideias inovadoras. Essa ferramenta utiliza métodos de mineração de texto no processamento de ideias, gerenciamento e visualização. Para a etapa de visualização, as notas são agrupadas por seguimento. O agrupamento desses cartões possibilita ao especialista responsável pelo desenvolvimento de novos produtos, identificar ideias relacionadas, enriquecendo umas às outras.

O estudo “*Open innovation in SMEs: Trends, motives and management challenges*”, apresentado por Van De Vrande et al. (2009),



investiga se as práticas de inovação aberta, fortemente difundidas nas grandes corporações, também são aplicadas em pequenas e médias empresas (PMEs). Os autores dedicam especial atenção aos benefícios e desafios percebidos pelas PMEs que adotam práticas de inovação aberta. Para compreender melhor a incidência da inovação aberta nas PMEs, os autores utilizaram análise de agrupamentos para reunir os respondentes em grupos de PMEs que são homogêneas tanto em sua estratégia de inovação aberta quanto na organização de práticas de inovação. Para viabilizar a análise de agrupamentos, inicialmente foi realizada a redução da dimensionalidade do *dataset* utilizando a técnica de Análise de Componentes Principais (*principal component analysis* - PCA). A partir da geração dos *clusters* os autores exploraram as diferenças entre as PMEs utilizando testes exploratórios não paramétricos.

Na publicação “*Idea Relationship Analysis in Open Innovation Crowdsourcing Systems*”, os autores buscam aumentar a quantidade de relacionamentos nos dados (ideias), mediante a construção de uma estrutura hierárquica para os relacionamentos e a sumarização do conjunto de dados. Essa estrutura colabora para que novos relacionamentos entre as ideias sejam estabelecidos, tais como, a herança e transitividade dos relacionamentos, proporcionando assim, um aumento significativo na quantidade de ideias similares obtidas (WESTERSKI; IGLESIAS; GARCIA, 2012).

No estudo “*Idea Identification Model to Support Decision Making*”, os autores demonstram a viabilidade do modelo de análise de agrupamentos voltado à gestão de ideias. Para tanto, foram apresentados dois cenários de estudo, a partir dos quais, os autores relataram que padrões e tendências podem ser identificados ao utilizar análise de agrupamentos para geração de grupos de ideias semelhantes. As ideias utilizadas para este trabalho foram coletadas de sites das empresas Starbucks® e Dell®. O método proposto foi baseado em ontologias e análise de agrupamentos com a finalidade de contribuir na tomada de decisão no processo de gestão de ideias (SERGIO; DE SOUZA; GONCALVES, 2017).

Nesse mesmo contexto, Pintas, Correia e Garcia (2017) apresentam o trabalho “*Crowd-based Feature Selection for Document Retrieval in Highly Demanding Decision-making Scenarios*”. No estudo os autores propõem o *CrowdFS* como uma abordagem que utiliza técnicas de inteligência coletiva para selecionar recursos relevantes e específicos sobre um grande volume de dados. Apesar dos resultados positivos obtidos, como trabalho futuro os autores sugerem a aplicação de rótulos na etapa de classificação, ou a utilização de técnicas de

análise de agrupamentos. De acordo com os autores, ao diminuir a dimensionalidade dos dados a serem analisados, o processo de identificação dos fatos relevantes se torna menos moroso.

Semelhantemente, na publicação “*An organizational model to understand the creativity workshop*”, Barrios et al. (2018) propõem a utilização da abordagem organizacional KROM (*Knowledge Reuse Organizational Meta-Model*) para modelar o processo de gerenciamento de ideias, a partir da interação entre os papéis, durante todas as atividades de criatividade. Os autores consideram que cada atividade realizada numa organização deve objetivar um alvo comum: a geração de novas ideias. Portanto, o processo de geração de novas ideias é uma atividade colaborativa que passa por diferentes atividades como a ideação, o agrupamento das ideias semelhantes, a avaliação, dentre outras tarefas.

Além dos artigos identificados na pesquisa bibliográfica, é válido mencionar outros trabalhos que foram utilizados como referencial para esta dissertação. Dentre esses trabalhos destacam-se: “Análise de agrupamentos e mineração de opinião como suporte à gestão de ideias” de Alvarez (2018), “Um modelo baseado em ontologia e análise de agrupamento para suporte à gestão de ideias” de Sérgio (2016) e “Modelo de avaliação de potenciais ideias alinhadas ao contexto organizacional” (SÉRGIO, 2018).

Alvarez (2018) apresenta um método para suporte e avaliação de ideias baseado em métodos e técnicas de Engenharia do Conhecimento. O experimento foi realizado mediante o desenvolvimento de um protótipo funcional para um conjunto de cenários formulados. A partir da importação do conjunto de dados da Ubuntu Brainstorm, os dados foram submetidos ao algoritmo de análise de opinião. Após a qualificação das ideias como positivas, neutras ou negativas, o *dataset* foi submetido ao processo de análise de agrupamentos. Por fim o autor afirma que o agrupamento de ideias similares permite que os especialistas de domínio analisem um maior número de ideias coletadas ao mesmo tempo, tornando o processo de análise das ideias mais fácil e produtivo. Além desta vantagem apresentada, a informação da polaridade das ideias possibilita expressar uma opinião inicial sobre o conteúdo por parte de determinada comunidade. Como trabalhos futuros é sugerida a integração das áreas estudadas, o desenvolvimento de um sistema de *ranking* e recomendação de ideias e a evolução do algoritmo responsável pelo processo de mineração de opinião a partir de um conjunto de ideias.

Sérgio (2016) propõe um modelo baseado em Ontologia e Análise de Agrupamentos para suporte à Gestão de Ideias. O modelo proposto utiliza métodos e técnicas de Engenharia do Conhecimento com o objetivo de auxiliar a tomada de decisão no contexto da Gestão de Ideias. Para o autor a utilização de técnicas de análise de agrupamento é possível evidenciar padrões e tendências com relação às ideias analisadas em um domínio. Os agrupamentos permitem que usuários interpretem de uma melhor maneira as ideias coletadas na organização, facilitando a tomada de decisão. Por fim é recomendado como trabalhos futuros a evolução do algoritmo responsável pelo processo de agrupamento, a inclusão da dimensão tempo no agrupamento e avaliação de ideia pois a divisão do *dataset* em sub conjuntos temporais poderia tornar a análise mais consistente, a exploração das tecnologias presentes na Web Semântica e o desenvolvimento de um sistema de recomendação.

Sérgio (2019) apresenta um modelo de avaliação de potenciais ideias alinhadas ao contexto organizacional utilizando técnicas de Engenharia do Conhecimento. Para atender os objetivos do trabalho são aplicadas técnicas de Mineração de Textos, Análise de Agrupamentos e Ontologias para o domínio de Gestão de Ideias. Por meio do modelo proposto é possível realizar consultas em diferentes fontes de dados gerando insumos para que os especialistas possam usar no processo avaliativo de potenciais ideias a serem desenvolvidas.

No Quadro 2 é apresentada uma comparação entre os trabalhos correlatos e o trabalho atual, formado pelas seguintes colunas:

- Autor/Ano: Esta coluna armazena a informação autoral do trabalho;
- Categoria: Classificação de acordo com o tipo da entrega realizada;
- Áreas: Áreas do conhecimento compreendidas no trabalho;
- Métodos e técnicas: Principais métodos e técnicas utilizados para atender os objetivos do trabalho;
- Entregável: Síntese da resposta produzida pelo trabalho;
- Fonte de dados: Base de dados de origem das informações utilizadas no trabalho.

Quadro 2 - Comparação dos trabalhos correlatos

<b>Autor/ Ano</b>	<b>Categoria</b>	<b>Áreas</b>	<b>Métodos e técnicas</b>	<b>Entregável</b>	<b>Fonte de dados</b>
Paukkeri, Kotro (2009)	Desenvolvimento de Software	Gestão de Ideias, Memória Organizacional	Mineração de Textos	Desenvolvimento de um IMS - Framework	Base de dados privada
Van de Vrande et al. (2009)	Revisão da Literatura	Gestão de Ideias, Inovação Aberta	Análise de Componentes Principais	Identificação de práticas de inovação aberta em PMEs	Não identificada
Westerski, Iglesias, Garcia (2012)	Desenvolvimento de Software	Gestão de ideias, Inovação Aberta, Crowdsourcing	Taxonomia, Classificação de Textos, Métodos Hierárquicos, Análise de Agrupamentos	Geração de melhores agrupamentos por meio do aumento da quantidade de relacionamento entre as ideias.	Ubuntu Brainstorm®
Sergio, de Souza e Goncalves (2017)	Desenvolvimento de Software	Gestão de Ideias, Inovação Aberta	Análise de Agrupamentos, Ontologia	Apresentação de grupos de ideias como suporte à gestão de ideias	Starbucks® Dell®
Pintas, Correia e Garcia (2017)	Desenvolvimento de Software	Crowdsourcing, Inteligência Coletiva, Business Intelligence	Redução de Dimensionalidade, Multi-label Text Learning, Análise de Agrupamentos	Geração de melhores agrupamentos por meio do aumento da quantidade de relacionamento entre as ideias.	Reuters-21578
Barrios et al. (2018)	Creativity workshop	Gestão do Conhecimento, Gestão de Ideias	Descoberta de domínio, <i>Creativity Groups</i>	Solucionar todos os desafios elencados por meio da utilização do	Não identificada

				modelo de negócio KROM	
Alvarez (2018)	Desenvolvimento de Software	Gestão de Ideias	Mineração de Opinião, Análise de Agrupamentos	Agrupamento de ideias similares como suporte à gestão de ideias e tomada de decisão	Ubuntu Brainstorm
Sergio (2016)	Desenvolvimento de Software	Gestão de Ideias	Ontologia, Análise de Agrupamentos	Agrupamento de ideias similares como suporte à gestão de ideias e tomada de decisão	Starbucks Dell
Sergio (2019)	Desenvolvimento de Software	Gestão de Ideias	Mineração de Textos, Análise de Agrupamentos, Ontologia	Base de conhecimento suportada por ontologia e agrupamento de ideias similares voltado à gestão de ideias e tomada de decisão.	Sinapse da Inovação
Autor	Desenvolvimento de Software	Gestão de Ideias	Análise de Agrupamentos	Agrupamento de ideias similares, ordenados por relevância, como suporte à gestão de ideias e tomada de decisão	Ubuntu Brainstorm

Fonte: Autor

Ao analisar o Quadro 2 é possível observar que dentre os 10 trabalhos relacionados, 8 estão concentrados na categoria

“Desenvolvimento de Software”, 1 em “*Creative Workshop*” e 1 em “Revisão da Literatura”.

Quanto às áreas de conhecimento, elas estão distribuídas da seguinte forma (um mesmo trabalho pode estar relacionado a mais de uma área):

- Gestão de Ideias: 9 ocorrências
- Inovação Aberta: 3 ocorrências;
- *Crowdsourcing*: 2 ocorrências;
- Gestão do Conhecimento: 1 ocorrência;
- Memória Organizacional: 1 ocorrência;
- Inteligência Coletiva: 1 ocorrência;
- *Business Intelligence*: 1 ocorrência.

Quanto aos métodos e técnicas utilizados, ocorre a seguinte distribuição (um mesmo trabalho pode estar relacionado a mais de um método ou técnica):

- Análise de Agrupamentos: 9 ocorrências;
- Ontologia: 3 ocorrências;
- Descoberta de Domínio: 1 ocorrência;
- *Creativity Groups*: 1 ocorrência;
- Taxonomia: 1 ocorrência;
- Classificação de Textos: 1 ocorrência;
- Métodos Hierárquicos: 1 ocorrência;
- Redução de Dimensionalidade: 1 ocorrência;
- Mineração de Textos: 1 ocorrência;
- Análise Exploratória: 1 ocorrência.

Quanto às bases de dados de origem, ocorre a seguinte distribuição de uso (um mesmo trabalho pode mencionar mais de uma base de dados):

- Base de dados não identificada: 2 ocorrências;
- Base de dados privada: 1 ocorrência;
- Ubuntu Brainstorm<sup>®</sup>: 3 ocorrências;
- Starbucks: 2 ocorrências;
- Dell: 2 ocorrências;
- Reuters-21578: 1 ocorrência;
- Sinapse da Inovação: 1 ocorrência.

Assim como neste trabalho, todos os trabalhos relacionados possuem uma forte ligação ou dependência com o tema Gestão de Ideias e Análise de Agrupamentos.

De forma semelhante ao trabalho atual, 3 dos trabalhos apresentados propõem a utilização de Análise de Agrupamentos como ferramenta para suporte à Gestão de Ideias, simplificado as informações para a tomada de decisão de um especialista. Estes trabalhos, apesar de convergentes, utilizam-se de diferentes métodos e técnicas para a proposição do modelo e solução dos desafios identificados.

Ao analisar estas características é possível compreender que a utilização de métodos e técnicas de Análise de Agrupamentos, assim como de outras técnicas pertencentes ao contexto da Engenharia do Conhecimento, permite agregar valor a determinado conjunto de ideias. O enriquecimento e a transformação de dados dessa natureza pode promover melhores insumos tendo em vista o suporte à gestão de ideias e a tomada de decisão.





## 4 MODELO PROPOSTO

Neste capítulo é apresentado o modelo proposto voltado ao agrupamento de ideias, descrevendo ainda, as principais características e informações que promovem suporte à proposta. Além disso, estão descritos os procedimentos realizados para tratamento dos dados e as demais etapas.

### 4.1 DESCRIÇÃO DO MODELO DA BASE DE IDEIAS

Com base no processo de aquisição de dados disponibilizados em redes de *crowdsourcing*, criadas com o objetivo de viabilizar a convergência de informações, este trabalho propõe a elaboração de um modelo de agrupamento de dados voltado à gestão de ideias.

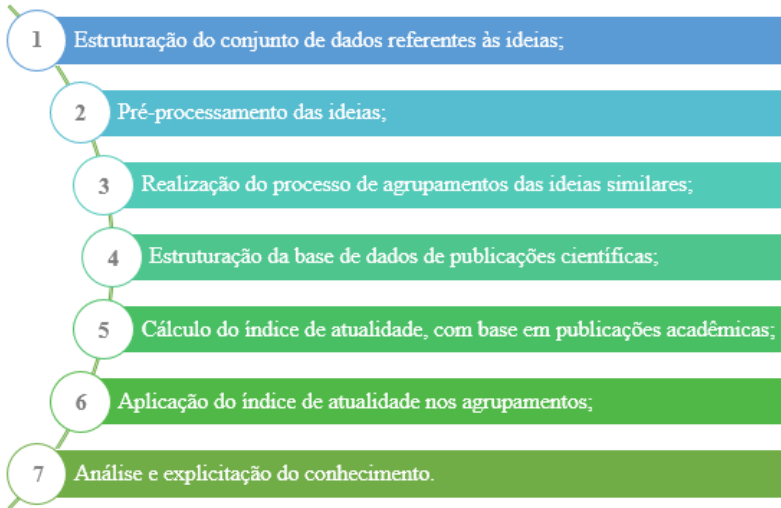
Mediante a obtenção e normalização dos conjuntos de dados (ideias e publicações científicas), o modelo objetiva promover suporte à análise, avaliação e seleção de ideias.

Para alcançar os objetivos deste trabalho, foi estabelecido um modelo para a identificação e seleção das ideias de acordo com seu grupo, considerando um índice de atualidade para cada ideia. O índice de atualidade de uma ideia é um valor atribuído por meio de um cálculo definido na Seção 4.1.4.

Ao somar o índice de atualidade das ideias de um determinado agrupamento é possível estabelecer um índice geral. O índice geral permite a geração de um *ranking* de agrupamento, possibilitando que os agrupamentos mais relevantes sejam analisados primeiro. O estabelecimento de um *ranking* a partir de um índice de atualidade é considerado um elemento fundamental para o modelo.

Para organizar o modelo proposto, foram elaboradas 7 etapas, apresentadas na Figura 15, com o objetivo de promover suporte ao processo de gestão de ideias.

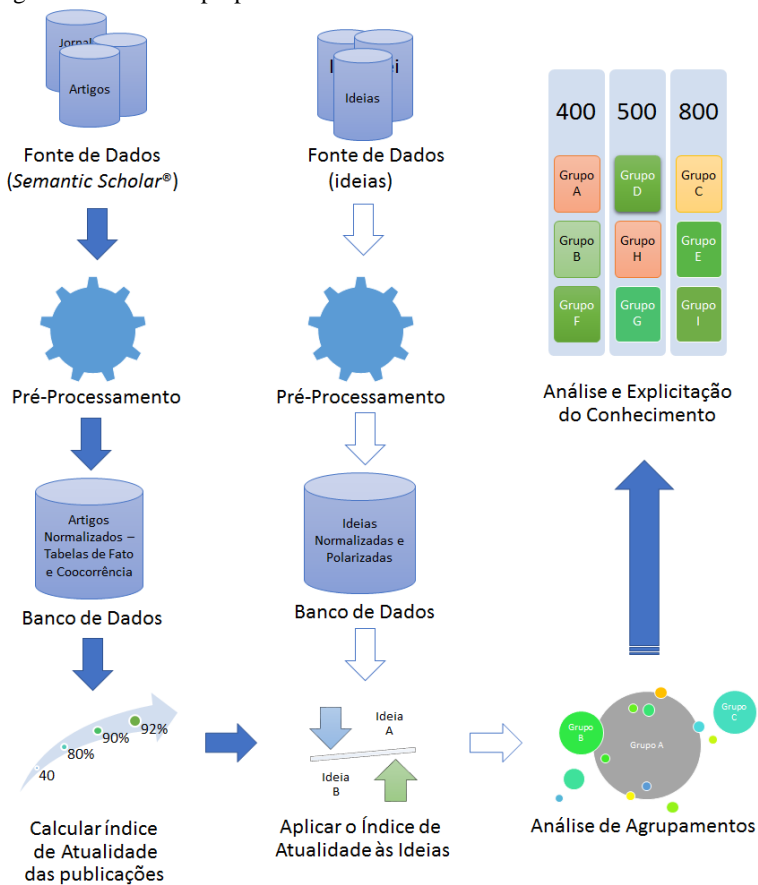
Figura 15 - Etapas do modelo proposto



Fonte: Autor

Analisando a representação do modelo realizado na Figura 16, é possível observar que os dois fluxos de obtenção de dados são executados de forma independente, até o ponto em que ambos os fluxos se conectam para a aplicação do índice de atualidade na base de ideias.

Figura 16 – Modelo proposto



Fonte: Autor

Nos próximos tópicos as etapas do modelo estão detalhadas.

### 4.1.1 Elaboração da estrutura de ideias e carga de dados

A primeira etapa se inicia mediante extração de ideias da fonte de dados selecionada para este trabalho. Após a extração dos dados é necessário realizar a formatação e importação dos dados de acordo com o modelo de dados proposto, computando os valores desejados e preparando o conjunto de dados para o processo de agrupamento.

Para realização deste trabalho foi utilizada a base de dados da Ubuntu Brainstorm<sup>®</sup>. A escolha da base foi determinada por alguns fatores, dentre os quais destacam-se: a característica de publicidade das ideias, a disponibilidade da base para *download* e indicação do estado de cada ideia. O estado da ideia determina se ela foi ou não implementada.

Conforme demonstrado na Figura 17, a ideia registrada na base de dados Ubuntu Brainstorm<sup>®</sup> possui algumas características necessárias para a composição deste modelo, sendo elas: código de identificação, título, nome do autor, data, situação, descrição da ideia, marcadores de classificação (*tags*), além de informar o relacionamento com o projeto.

Figura 17 – Visualização de ideias no Ubuntu Brainstorm<sup>®</sup>

The screenshot displays the Ubuntu Brainstorm interface. At the top, there are four navigation tabs: 'Idea sandbox', 'Popular ideas', 'Ideas in development', and 'Implemented ideas'. The main content area shows an idea titled 'Idea #29850: Simplify and speed up switching windows.' Below the title, there are links for 'Description', 'Report duplicate', and 'Help promote this idea!'. The idea was written by 'al\_chesse' on '14 Jun 12 at 13:30'. The rationale states: 'I say I have 3 firefox windows open. I'm randomly switching between them and find it difficult to use the standard unity effect. The three firefox windows are small and it's sometimes difficult to see which one I want. I've grown to like the extra real estate unity allows, but I don't like having to click/hover/click to switch an app. I simply want to click once. I've also tried the alt tab down arrow method and I can't get used to it.' The tags are 'minimize open switcher ungroup unity'. There are three solutions listed: 'Solution #1: Option to ungroup Icons in the Unity bar.' (written by al\_chesse), 'Solution #2: Lens-like listing from the launcher icon' (written by pascualo), and 'Solution #3: Extension of solution #2' (written by pascualo).

Fonte: Alvarez (2018)

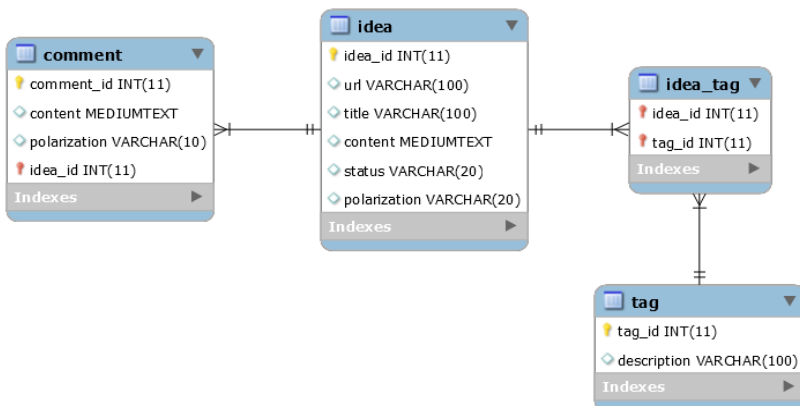
Cada ideia também pode receber comentários, que por sua vez podem receber votos e, conseqüentemente, serem indicados como favoritos. Estes comentários devem possuir título, descrição, nome do autor, data e avaliação.

Tomando essa base como fonte de dados foi realizada a análise da estrutura de dados necessária para suportar a proposta deste trabalho. Para elaboração deste modelo de dados observou-se o trabalho de

Alvarez (2018), adotando o modelo de entidade de relacionamento proposto para a importação e polarização das ideias.

Para tal, foi utilizado o modelo de dados apresentado na Figura 18 representando o diagrama de Entidade-Relacionamento para quatro tabelas: *idea* (ideias), *comment* (comentários), *tag* (marcadores) e “*idea\_tag*” (ideias x marcadores).

Figura 18 – Diagrama de Entidade-Relacionamento



Fonte: Alvarez (2018)

A tabela de *idea* deve armazenar os dados das ideias registradas na base de dados Ubuntu Brainstorm<sup>®</sup>. Os atributos desta tabela são:

- *idea\_id*: identificador da ideia; este campo deve ser único e é utilizado como chave primária da tabela;
- *url*: endereço de localização da ideia na *internet*;
- *title*: título da ideia;
- *content*: conteúdo da ideia, representa o texto descritivo da ideia;
- *status*: situação da ideia, sendo as principais marcações “implementada” e “não implementada”;
- *polarization*: representa a polarização da ideia, esta informação é oriunda do trabalho realizado por Alvarez (2018) levando em consideração a análise de sentimentos. Os valores apresentados nesse atributo podem ser “negativa”, “neutra” ou “positiva”.

Na tabela *comment* são armazenados os dados dos comentários das ideias. Os atributos desta tabela são:

- *comment\_id*: este atributo é o identificador do registro da tabela, ele deve ser, também, único e utilizado como chave primária;
- *content*: este atributo armazena o conteúdo textual do comentário;
- *polarization*: este atributo representa a polaridade da ideia. Para compor este atributo, além da votação da ideia realizada na plataforma Ubuntu Brainstorm<sup>®</sup>, foram aplicados os procedimentos apresentados por Alvarez (2018), para identificação e análise de sentimentos;
- *idea\_id*: este atributo representa o identificador da ideia a que o comentário está relacionado. A coluna é utilizada para compor uma chave estrangeira (*foreign key*), garantindo o relacionamento entre os registros.

A tabela *tag* é utilizada para armazenar distintamente os marcadores registrados em todas as ideias da base. Esta tabela possui apenas dois atributos:

- *tag\_id*: este atributo representa o identificador único e chave primária da tabela;
- *description*: este atributo representa a descrição de cada *tag* (marcador) adicionado à ideia.

De acordo com o diagrama ER apresentado na Figura 18, não há uma relação direta entre a tabela *idea* e a tabela *tag*. Isto ocorre devido ao fato de que os registros de ambas as relações podem se conectar a mais de uma tupla da outra tabela. Para estabelecer este relacionamento foi criada a tabela *idea\_tag*. Os atributos da tabela de relacionamento entre ideias e marcadores são:

- *idea\_id*: este atributo representa o identificador da ideia na tabela *idea*. Nesta relação é utilizado como uma chave estrangeira e compõe a chave primária;
- *tag\_id*: este atributo representa o identificador da *tag* na tabela *tag*. Nesta relação é utilizado como uma chave estrangeira e compõe a chave primária.

#### 4.1.2 Pré-processamento das Ideias

A etapa de pré-processamento inicia-se com a entrada de um documento disponibilizado pelo grupo de pesquisa GI2MO<sup>26</sup> no formato RDF (*Resource Description Framework*). Cada documento está relacionado com uma ideia provida pelos colaboradores da comunidade Ubuntu *Brainstorm*<sup>®</sup> (ALVAREZ; 2018). Mediante a importação e estruturação destes dados, as ideias, comentários e *tags* são armazenadas na base de dados relacional apresentado na Figura 18.

Após a importação de dados no modelo relacional é aplicado o algoritmo de mineração de opinião proposto por Alvarez (2018) para preenchimento do atributo “*polarization*” nas entidades “*idea*” e “*comment*”. A polarização de uma ideia ou comentário resulta na classificação deste como positivo, negativo ou neutro; de forma que, a classificação dos comentários de uma ideia implica diretamente na classificação da ideia. Ou seja, a maior soma dos comentários positivos, negativos ou neutros determina a polaridade final da ideia.

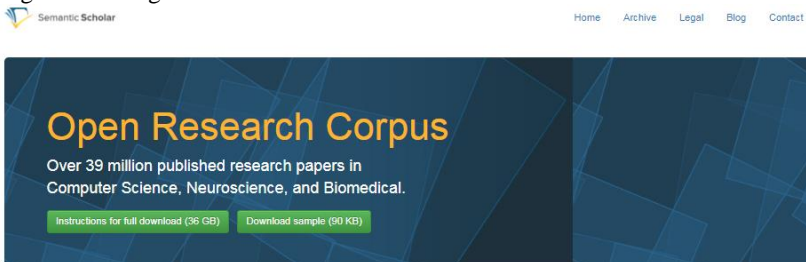
#### 4.1.3 Coleta, armazenamento e padronização dos dados de uma base acadêmica

Os dados de publicações acadêmicas foram obtidos da base de dados disponibilizada pela Semantic Scholar<sup>®</sup>, contendo as informações básicas sobre determinados artigos, autores e palavras-chave. A base possui mais de 45 milhões artigos de pesquisa publicados sendo obtida no formato *Document JSON (JavaScript Object Notation)* disponível no *site* da Semantic Scholar<sup>®27</sup>, conforme demonstrado na Figura 19.

---

<sup>26</sup> <http://www.gi2mo.org/>

<sup>27</sup> <http://labs.semanticscholar.org/corpus/>

Figura 19 – Página de *download* dos dados da Semantic Scholar®

Fonte: Semantic Scholar®

Na página relacionada na Figura 19 está descrito o *JSON Schema* do conjunto de dados a ser obtido, contendo os atributos discriminados no Quadro 3:

Quadro 3 – Descrição dos atributos do *JSON Schema*

<i>Attributes</i>	<i>Type</i>	<i>Description</i>
<i>Id</i>	<i>String</i>	<i>S2 generated research paper ID</i>
<i>Title</i>	<i>String</i>	<i>Research paper title</i>
<i>paperAbstract</i>	<i>String</i>	<i>Extracted abstract of the paper</i>
<i>entities</i>	<i>List</i>	<i>S2 extracted list of relevant entities or topics</i>
<i>s2Url</i>	<i>String</i>	<i>URL to S2 research paper details page</i>
<i>s2PdfUrl</i>	<i>String</i>	<i>URL to PDF on S2 if available</i>
<i>pdfUrls</i>	<i>List</i>	<i>URLs related to this PDF scraped from the web</i>
<i>Authors</i>	<i>List</i>	<i>List of authors with an S2 generated author ID and name</i>
<i>inCitations</i>	<i>List</i>	<i>List of S2 paperId's which cited this paper.</i>
<i>outCitations</i>	<i>List</i>	<i>List of paperId's which this paper cited</i>
<i>Year</i>	<i>Int</i>	<i>Year this paper was published as integer.</i>
<i>Venue</i>	<i>String</i>	<i>Extracted venue published.</i>
<i>journalName</i>	<i>String</i>	<i>Name of the journal that published this paper.</i>
<i>journalVolume</i>	<i>String</i>	<i>The volume of the journal where this paper was published</i>
<i>journalPages</i>	<i>String</i>	<i>The pages of the journal where this paper was published</i>
<i>Sources</i>	<i>List</i>	<i>Identifies papers sourced from DBLP or Medline</i>
<i>Doi</i>	<i>String</i>	<i>Digital Object Identifier registered at doi.org</i>
<i>doiUrl</i>	<i>String</i>	<i>DOI link for registered objects</i>
<i>Pmid</i>	<i>String</i>	<i>Unique identifier used by PubMed</i>

Fonte: Semantic Scholar®.

Na Figura 20 é apresentado um documento JSON que exemplifica o conjunto de dados obtido:



Figura 20 – JSON Schema provido pela Semantic Scholar®

```

1 {
2   "id": "4cd223df721b722b1c40689caa52932a41fcc223",
3   "title": "Knowledge-rich, computer-assisted composition of Chinese couplets",
4   "paperAbstract": "Recent research effort in poem composition has focused on...",
5   "entities": [
6     "Conformance testing",
7     "Natural language generation",
8     "Natural language processing",
9     "Parallel computing",
10    "Stochastic grammar",
11    "Web application"
12  ],
13  "s2Url": "https://semanticscholar.org/paper/4cd223df721b722b1c40689caa52932a41fcc223",
14  "s2PdfUrl": "",
15  "pdfUrls": [
16    "https://doi.org/10.1093/llc/fqu052"
17  ],
18  "authors": [
19    {
20      "name": "John Lee",
21      "ids": [
22        "3362353"
23      ]
24    },
25    "...",
26  ],
27  "inCitations": [
28    "c789e333fdbb963883a0b5c96c648bf36b8cd242"
29  ],
30  "outCitations": [
31    "abe213ed63c426a089bdf4329597137751dbb3a0",
32    "...",
33  ],
34  "year": 2016,
35  "venue": "DSH",
36  "journalName": "DSH",
37  "journalVolume": "31",
38  "journalPages": "152-163",
39  "sources": [
40    "DBLP"
41  ],
42  "doi": "10.1093/llc/fqu052",
43  "doiUrl": "https://doi.org/10.1093/llc/fqu052",
44  "pmid": ""
45 }

```

Fonte: Autor

De acordo com as orientações disponibilizadas no *site*<sup>28</sup> da Semantic Scholar® para obter a base de dados foi utilizado o método de *download* WGET (Figura 21).

Figura 21 – Comando para obter a base de dados da Semantic Scholar®

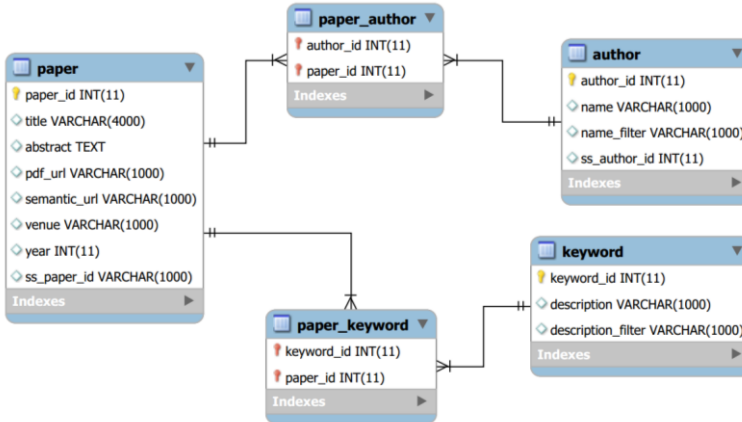
```
wget -i https://s3-us-west-2.amazonaws.com/ai2-s2-research-public/open-corpus/manifest.txt
```

Fonte: Semantic Scholar®

<sup>28</sup> <https://s3-us-west-2.amazonaws.com/ai2-s2-research-public/open-corpus/index.html>

Para suportar os dados obtidos, foi elaborado o modelo de dados representado no diagrama entidade-relacionamento da Figura 22. Esse modelo contém cinco tabelas: *papers* (artigos), *author* (autores), *keyword* (palavras-chave), “*paper\_keyword*” (palavras-chave x artigos) e “*paper\_author*” (autores x artigos).

Figura 22 – Diagrama ER proposto para os dados da Semantic Scholar<sup>®</sup>



Fonte: Autor

A tabela *paper* foi modelada para armazenar os dados dos trabalhos científicos acadêmicos, publicados em periódicos e conferências, registrados na base de dados da Semantic Scholar<sup>®</sup>. Os atributos desta tabela são:

- *paper\_id*: identificador do *paper*, este campo deve ser único e é utilizado como chave primária da tabela;
- *title*: título do *paper*;
- *abstract*: resumo do *paper*, representa o texto descritivo resumido do trabalho;
- *pdf\_url*: endereço de localização do PDF do *paper* na internet;
- *semantic\_url*: endereço de localização do registro do *paper* na internet (Semantic Scholar);
- *venue*: nome da revista, ou cidade onde o artigo foi publicado;
- *year*: ano de publicação do *paper*.

A tabela de *keyword* foi modelada para armazenar distintamente as palavras-chave utilizadas em todos os *papers* registrados na base de dados da Semantic Scholar<sup>®</sup>. Os atributos desta tabela são:

- *keyword\_id*: identificador da palavra-chave. Este campo deve ser único e é utilizado como chave primária da tabela;
- *description*: descrição da palavra-chave;
- *description\_filter*: descrição normalizada (*standardization*) da palavra-chave;
- *current\_index*: este campo indica o índice de atualidade da palavra-chave.

A tabela *author* foi modelada para armazenar distintamente os dados dos autores dos *papers* registrados na base de dados da Semantic Scholar<sup>®</sup>. Os atributos dessa tabela são:

- *author\_id*: identificador do autor, este campo deve ser único e é utilizado como chave primária da tabela;
- *name*: nome do autor;
- *name\_filter*: nome do autor normalizado (*standardization*);
- *ss\_author\_id*: identificador do autor na Semantic Scholar<sup>®</sup>.

De acordo com o diagrama ER apresentado na Figura 22, não há uma relação direta entre as tabelas *paper* e *keyword*, tão pouco entre *paper* e *author*. Isso ocorre devido ao fato de que os registros em ambas as relações podem se conectar a mais de uma tupla da outra tabela.

Para estabelecer este relacionamento foram criadas as tabelas *paper\_keyword* e *paper\_author*. Os atributos da tabela de relacionamento entre *papers* e palavras-chave são:

- *keyword\_id*: esse atributo representa o identificador da palavra-chave na tabela *keyword*. Nesta relação é utilizado como uma chave estrangeira e compõe a chave primária;
- *paper\_id*: esse atributo representa o identificador do *paper* na tabela *paper*. Nesta relação é utilizado como uma chave estrangeira e compõe a chave primária.

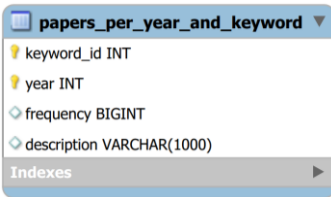
Os atributos da tabela de relacionamento entre *papers* e autores são:

- *author\_id*: este atributo representa o identificador do autor na tabela *author*. Nesta relação é utilizado como uma chave estrangeira e compõe a chave primária;

- *paper\_id*: este atributo representa o identificador do *paper* na tabela *paper*. Nesta relação é utilizado como uma chave estrangeira e compõe a chave primária.

Para viabilizar o cálculo do índice de atualidade foi gerada uma tabela de fato com o objetivo de sumarizar o total de artigos por palavra-chave e por ano. Para compor esta tabela (Figura 23), foi realizada a agregação da quantidade de registros (frequência) da tabela *paper\_keyword* relacionada à tabela *paper*, agrupando-os pelos atributos *keyword\_id* (identificador da palavra-chave) da tabela *paper\_keyword* e *year* (ano da publicação do artigo) da tabela *paper*.

Figura 23 – Tabela de fato de artigos por palavra-chave e ano



Fonte: Autor

O resultante desta projeção foi armazenado na tabela *papers\_per\_year\_and\_keyword* que contém as seguintes colunas:

- *keyword\_id*: este atributo compõe o identificador único (chave primária) da tabela e é referente a palavra-chave relacionada no artigo;
- *year*: este atributo compõe o identificador único (chave primária) da tabela e é referente ao ano em que o artigo foi publicado;
- *frequency*: este atributo representa a quantidade de artigos publicados utilizando a palavra-chave e ano em questão;
- *description*: esse atributo representa a descrição da palavra-chave. A existência deste atributo é uma desnormalização com o objetivo de evitar a realização de uma nova junção na aplicação da tabela de fato.

#### 4.1.4 Cálculo do índice de atualidade

Em se tratando de Gestão de Ideias e Gestão de Inovação, a informação de que uma ideia é positiva pode auxiliar na tomada de

decisão. Entretanto, uma ideia positiva implementada de forma tardia pode não produzir os resultados esperados.

Como forma de agregar valor a uma ideia propõe-se a geração de um atributo que indique o quanto uma ideia é atual. O índice de atualidade visa descrever o quão atual a ideia pode ser considerada.

Para determinar este índice foram utilizados os dados obtidos na base *Semantic Scholar*<sup>®</sup> e sumarizados na tabela *papers\_per\_year\_and\_keyword*. Deste modo, cada termo (palavra-chave) possui um índice de atualidade. Em etapa posterior cada ideia é associada a um conjunto de palavras-chave por meio de uma análise léxica possibilitando que a ideia tenha um índice de atualidade. Um exemplo do cálculo deste índice, proposto por Sérgio (2018), está apresentado na Tabela 1.

Tabela 1 – Exemplo de cálculo do índice de atualidade

Ideia 1										
Termo 1					Termo 2					
Ano	Docs	Tempo	Deflator	Peso	Ano	Docs	Tempo	Deflator	Peso	
2018	2	1	0,900	1,800	2018	0	1	0,900	0,000	
2017	10	2	0,810	8,100	2017	0	2	0,810	0,000	
2016	12	3	0,729	8,748	2016	0	3	0,729	0,000	
2015	8	4	0,656	5,249	2015	0	4	0,656	0,000	
2014	5	5	0,590	2,952	2014	3	5	0,590	1,771	
2013	8	6	0,531	4,252	2013	6	6	0,531	3,189	
2012	5	7	0,478	2,391	2012	5	7	0,478	2,391	
2011	4	8	0,430	1,722	2011	7	8	0,430	3,013	
2010	2	9	0,387	0,775	2010	12	9	0,387	4,649	
2009	3	10	0,349	1,046	2009	10	10	0,349	3,487	
<b>IA (Termo 1)</b>				<b>3,704</b>	<b>IA (Termo 2)</b>				<b>1,850</b>	
									<b>IA (Ideia 1)</b>	<b>2,777</b>

Fonte: Adaptado de Sérgio (2018)

Na Tabela 1 pode-se observar uma representação de uma ideia (**Ideia 1**) que contém dois termos (**Termo 1** e **Termo 2**). A distribuição temporal que denota a ocorrência de cada um dos termos por ano é representada pelas linhas.

A coluna **Ano** identifica o ano em que a ocorrência do termo foi aferida, variando entre 2009 e 2018. A coluna **Docs** contabiliza o número de artigos nos quais o termo faz parte em determinado ano.

A coluna **Tempo** armazena um número sequencial, iniciado com o valor 1 (um) do maior para o menor ano de ocorrência, atribuindo assim, maior peso às ocorrências mais recentes.

A coluna **Deflator** armazena um valor que servirá para reduzir o peso de uma ocorrência em função do tempo decorrido (ano). O valor é

determinado elevando-se o fator de deflação à potência  $n$ , sendo  $n$  o valor da coluna **Tempo**. O fator de deflação é um valor constante e deve ser provido como parâmetro de entrada para o cálculo.

A coluna **Peso** representa o quão relevante um termo é para o cálculo do índice de atualidade de uma determinada ideia.

O conceito de deflator consiste em aplicar um fator de amortecimento, diminuindo a relevância de uma determinada ocorrência. No algoritmo de *PageRank*<sup>®</sup>, por exemplo, é utilizado para determinar a relevância de uma página mediante a utilização de estruturas de *hiperlinks* para navegação na *web* (PAGE, 1999).

O valor atribuído ao peso é o número de documentos que mencionam o termo multiplicado pelo deflator. O peso é calculado para cada ano, sendo que o termo recebe como índice de atualidade a média dos pesos de todos os anos do período avaliado. Por fim, para compor o índice de atualidade de uma ideia é realizada a média dos índices de cada termo da ideia.

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

A Equação 1 ilustra o cálculo, onde:

- $IA_{ideia}$ : resultante constituindo o índice de atualidade;
- $n$ : total de termos (palavras-chave) associadas à ideia;
- $k$ : total de anos considerados no cálculo;
- $docs$ : quantidade de documentos em que há ocorrência do termo em um ano específico;
- $deflator$ : índice de decréscimo temporal, inicialmente definido com valor inferior a 1 (um).

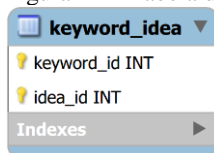
De acordo com os valores apresentados na Tabela 1, foi adotado o valor de 0,9 para a variável *deflator*. Para o atributo  $k$  foi adotado o valor 10, correspondente a diferença entre os anos 2009 e 2018. Para o atributo **Docs** foram considerados os dados obtidos da tabela de fato *papers\_per\_year\_and\_keyword* gerada para otimizar a projeção.

Ao aplicar os dados na equação, o índice médio de atualidade obtido para o **Termo 1** foi de 3,704 e para o **Termo 2**, foi de 1,850. A média calculada para índices de atualidade dos termos da ideia foi 2,777, determinando assim o índice de atualidade da ideia. O resultado foi então armazenado na tabela *keyword*, coluna *current\_index*.

#### 4.1.5 Aplicação do Índice de Atualidade nas Ideias

Para estabelecer o índice de atualidade nas ideias foi elaborado um processo que, aplicando uma análise léxica, identifica o conjunto de palavras-chave associadas com determinada ideia. O resultado desta execução é armazenado em uma nova tabela (Figura 24) denominada *keyword\_idea*.

Figura 24 – Tabela de relacionamento de palavras-chave e ideias



Fonte: Autor

Para realizar este procedimento, o algoritmo leva em consideração as palavras-chave obtidas na base da Semantic Scholar<sup>®</sup>. Mediante a análise léxica, são identificadas as ideias da base Ubuntu<sup>®</sup> cujo texto, composto por título e descrição, contenha a palavra-chave.

Quando uma palavra-chave é localizada no texto da ideia ela é armazenada na tabela *keyword\_idea*, possibilitando que posteriormente o vetor de termos seja acrescido das palavras-chave relacionadas.

Ao finalizar este procedimento, por meio do somatório do índice de atualidade das palavras-chave é possível obter o índice de atualidade da ideia. Este valor é fundamental para a fase final do modelo proposto que envolve a ordenação dos agrupamentos gerados.

#### 4.1.6 Agrupamento de Ideias Similares

Nesta etapa do modelo está apresentado o processo de agrupamento das ideias. Esta atividade objetiva a divisão das ideias em grupos cujas ideias sejam homogêneas entre si e heterogêneas em relação às ideias de outros grupos (EVERITT et al., 2011). Após a realização do agrupamento é criado o *ranking*, de forma que agrupamentos mais atuais tenham maior relevância na ordenação.

O algoritmo utilizado está baseado no Modelo de Espaço Vetorial (*Vector Space Model* – VSM). A entrada deste algoritmo é uma lista de ideias. Cada elemento da lista é transformado em um vetor de espaço  $n$ -dimensional, onde  $n$  representa os termos que compõem o vetor. Os

termos de um vetor são identificados a partir do título e do conteúdo das ideias, com exceção dos termos identificados na lista de *stopwords*.

Para cada um dos termos que compõe a ideia, é atribuído um identificador e um peso, formando a matriz termo-documento. O peso de um termo é definido por meio de um método estatístico denominado *TF-IDF* (*Term Frequency-Inverse Document Frequency*) e representa a relevância deste para o vetor.

Utilizando esta medida numérica é possível calcular o peso de cada palavra em um documento. Basicamente o *TF-IDF* busca determinar o número de ocorrências de um termo num documento (*TF*) em relação a proporção inversa dos documentos em que o termos ocorrem (*IDF*). Termos que aparecem em menos documentos são considerados mais raros e apresentam maior *TF-IDF* do que termos comuns nos documentos. Os termos com maior *TF-IDF* podem ser utilizados para identificar o documento, bem como, compor consultas (RAMOS et al., 2003).

Uma vez calculados os pesos dos termos de cada ideia, são construídos os vetores das ideias, contendo cada um dos termos do documento com seu respectivo peso. A partir destes vetores será determinada a medida de similaridade entre as ideias. Neste trabalho, a medida de similaridade será obtida mediante o cálculo do ângulo do cosseno formado pelos vetores das ideias.

Para realizar o cálculo do cosseno é utilizada a Equação 2, na qual, a partir dos vetores  $t$  e  $q$ ,  $t_i$  e  $t_k$  representam as frequências normalizadas dos  $i$ th e  $k$ th termos do vetor  $t$ , e  $q_i$  e  $q_j$  representam a frequências normalizadas dos termos  $i$ th e  $j$ th do vetor  $q$  (GONÇALVES, 2006).

A equação utilizada para calcular o cosseno está apresentada abaixo:

$$\cos \theta = \frac{\sum_{i=1}^n (t_i \times q_i)}{\sqrt{\sum_{k=1}^n (t_k)^2} \times \sqrt{\sum_{j=1}^n (q_j)^2}} \quad (2)$$

O resultado da equação ocorre mediante o somatório do produto dos pesos dos termos  $t_i$  e  $q_i$ , divididos pela raiz quadrada do somatório de  $t_k$  elevado ao quadrado, multiplicada pela raiz quadrada do somatório de  $q_j$  elevado ao quadrado.

O cosseno resultante é um valor que varia entre 1.0 e -1.0. Quanto mais próximo de 1, maior a similaridade entre os vetores, de forma que



1.0 representa vetores apontado para a mesma direção ( $\cos 0^\circ = 1.0$ ), 0.0 representa vetores ortogonais ( $\cos 90^\circ = 0.0$ ), e -1.0 representa vetores que apontam para direções opostas ( $\cos 180^\circ = -1.0$ ) (JONES; FURNAS, 1987).

O algoritmo utilizado neste trabalho é uma adaptação do algoritmo *K-means*, modificado para utilizar o conceito de similaridade vetorial. Tradicionalmente, o algoritmo *K-means* recebe um valor de entrada  $k$  que determina quantos agrupamentos serão gerados. Com a alteração realizada, o algoritmo passa a receber um valor de entrada que determina o *threshold* (limite) para considerar dois vetores similares. Sempre que um vetor estiver abaixo do limite estabelecido um novo agrupamento é criado.

Na Figura 25 é apresentado o fluxograma do algoritmo de análise de agrupamentos evoluído a partir do trabalho de Alvarez (2018). O processo inicia pela recuperação do conjunto de vetores disponíveis para agrupamento. Após a recuperação dos vetores das ideias é determinado um limiar para os agrupamentos (*threshold*) e então se iniciam as iterações para análise e geração de agrupamentos.

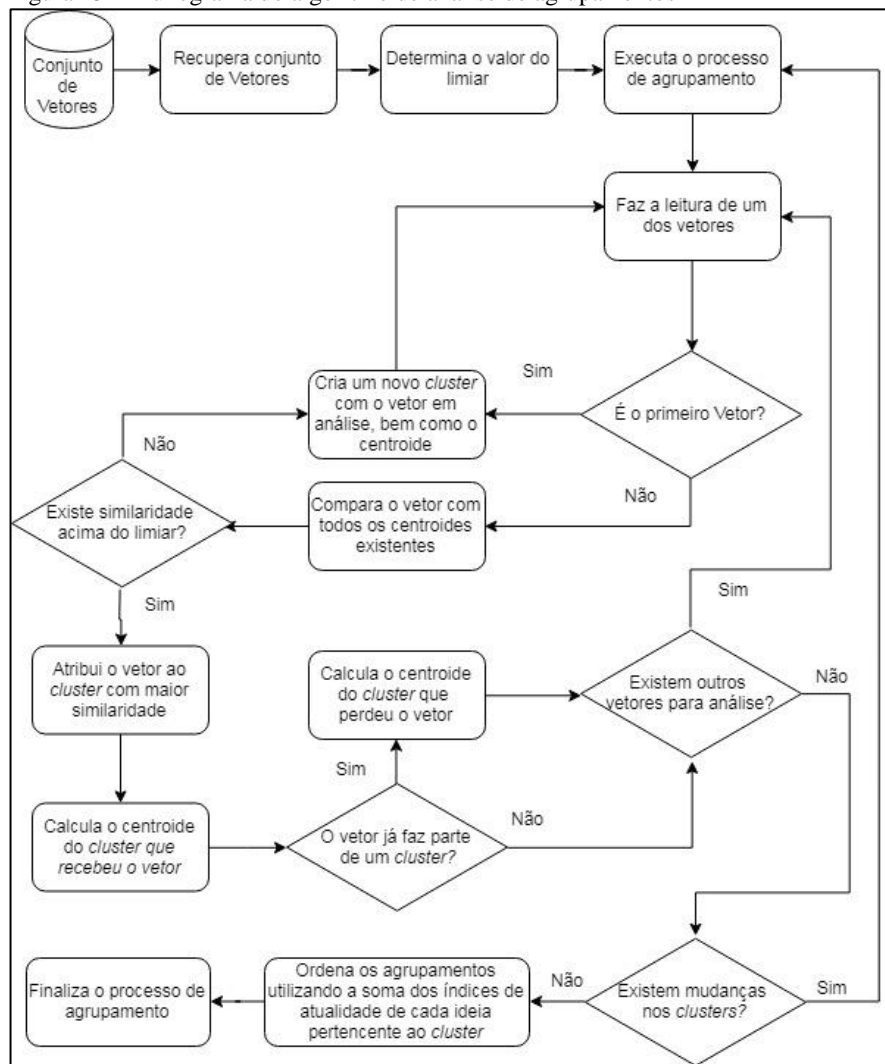
Ao iniciar a iteração, um a um, os vetores são lidos e analisados. Caso seja o primeiro vetor da lista, um novo agrupamento (*cluster*) é criado e o vetor é utilizado para compor o centroide e a análise do vetor em questão é finalizada, iniciando assim, a análise do próximo elemento da lista.

A partir do segundo elemento da lista, cada vetor é comparado com todos os centroides existentes em busca de agrupamentos cuja similaridade (conforme seção 4.1.4) esteja acima do limiar estabelecido. Caso não existam agrupamentos que atendam a busca, um novo agrupamento é gerado utilizando o vetor atual como centroide.

Caso se identifiquem agrupamentos cuja similaridade com o vetor atenda ao limiar estabelecido o vetor é atribuído ao agrupamento com maior similaridade. Após a inclusão de um novo vetor ao agrupamento o centróide do mesmo será recalculado.

Antes de finalizar a iteração é verificado se o vetor pertencia a outro agrupamento. Caso tenha sido parte de outro agrupamento, o agrupamento que perdeu o vetor tem seu centroide novamente calculado.

Figura 25 – Fluxograma do algoritmo de análise de agrupamentos



Fonte: Autor

No final da iteração é verificado se ainda existem outros vetores da lista para análise. Caso exista, o próximo vetor da lista é seleccionado e o procedimento é repetido, caso não, entende-se que chegou-se ao final de uma época.

Ao finalizar a leitura de todos os vetores (chamado de época), se verifica a ocorrência de mudanças nos agrupamentos gerados. Em caso de terem ocorrido movimentações de vetores entre os *clusters*, uma nova época é criada e o procedimento de agrupamento é novamente executado, do contrário o processo é encaminhado para finalização.

Antes de finalizar o processo de agrupamento, com base no somatório do índice de atualidade de cada uma das ideias (representadas pelos vetores) é estabelecido o índice de atualidade de cada agrupamento. Os agrupamentos gerados são ordenados de forma decrescente. Mediante a ordenação, os agrupamentos com maior índice de atualidade são posicionados no topo do *ranking*, e os de menor índice, gradativamente na parte inferior. Essa etapa, bem como as otimizações durante o cálculo do centroide no algoritmo de agrupamento, foram adicionadas em relação à versão do algoritmo proposto por Alvarez (2018).

Na seção a seguir será abordada com mais detalhamento a etapa de análise e explicitação do conhecimento.

#### 4.1.7 Análise e explicitação do conhecimento

Nesta etapa são aplicadas as técnicas de análise de agrupamentos, explorando diferentes medidas para os cenários estabelecidos. De acordo com a literatura, para interpretar os resultados produzidos por um processo de agrupamentos é necessário levar em consideração um grande número de variáveis. Identificar, interpretar e combinar estas interpretações pode agregar complexidade a análise e explicitação do conhecimento (GOLDSCHMIDT; BEZERRA; PASSOS, 2015).

A explicitação dos resultados obtidos remete à identificação de padrões e tendências. Estes padrões devem ser capazes de produzir evidências que atestem a relação da polarização das ideias com os índices de atualidade como ferramenta para o processo de gestão de ideias e tomada de decisão.

Como forma de exemplificar a aplicação do índice de atualidade sobre os agrupamentos, sugere-se um cenário fictício com o seguinte resultado utilizando três grandes grupos de ideias:

- Ideias para melhorar as rodas de uma carroça;
- Ideias para aumentar a autonomia de combustível de veículos;
- Ideias para produção de veículos movidos com combustíveis baseados em energia limpa.

Considerando que todos os grupos de ideias são positivos, torna-se necessário que o especialista de domínio selecione um agrupamento para iniciar a análise das ideias. Em escalas maiores, selecionar os agrupamentos com melhor potencial de implementação pode ser uma tarefa complexa.

No entanto, ao analisar o contexto central dos agrupamentos gerados observa-se que algumas ideias, apesar de positivas não são atuais, como por exemplo, a melhoria das rodas de uma carroça. Outras são contemporâneas, mas não muito recentes, como por exemplo, ideias para aumentar a autonomia de combustível veicular. Há, no entanto, ideias recentes e de alta relevância, como as ideias para utilização de energia limpa na produção de veículos automotores.

A identificação da relevância temporal de uma ideia é um trabalho complexo que exige não apenas conhecimentos prévios sobre a temática, mas uma constante observação dos movimentos de mercado, bem como, de um investimento considerável de tempo.

Ao incluir a informação de temporalidade a um *cluster*, dispondo o resultado dos agrupamentos de forma ordenada (*ranking*), a efetividade do resultado do agrupamento pode ser incrementada, entregando informações adicionais que subsidiem o especialista de domínio na tomada de decisão.

Portanto, a identificação e o mapeamento destes resultados corrobora no processo de alcance dos objetivos específicos estabelecidos para este trabalho, demonstrando que a consideração do índice de atualidade para seleção de ideias positivas potencializa a escolha de boas ideias para implementação.

Entretanto este processo de identificação e mapeamento requer a experimentação de diferentes cenários, tornando a tarefa complexa, principalmente no que se refere ao ato de indicar quais medidas e ações devem ser aplicadas em cada um dos cenários/situações observados.

## 4.2 CONSIDERAÇÕES FINAIS

Este capítulo apresentou o modelo proposto detalhando cada uma das etapas que o constitui. O modelo é composto de algumas etapas fundamentais: a preparação da estrutura de ideias para o modelo proposto, o pré-processamento das ideias, o preenchimento da base com as ideias polarizadas, o pré-processamento de artigos de bases científicas, o preenchimento da base com o índice de atualidade dos termos, a aplicação do índice de atualidade nas ideias, o agrupamento de ideias similares e a análise e explicitação do conhecimento.

Neste sentido, a estrutura do modelo proposto promoveu a integração de técnicas de Tecnologias Computacionais visando a fornecer um ferramental para promover suporte às aplicações relacionadas à Gestão de Ideias.

No próximo capítulo são relatados os cenários de aplicação, assim como, os resultados obtidos a fim de demonstrar a viabilidade do modelo proposto.



## 5 APLICAÇÃO DO MODELO E ANÁLISE DOS RESULTADOS

Neste capítulo é apresentado o cenário de aplicação do modelo proposto voltado à geração e ordenação de agrupamentos de ideias no contexto de Gestão de Ideias. Também são apresentados os resultados e análises dos experimentos realizados.

### 5.1 CENÁRIOS DE APLICAÇÃO

A base de dados utilizada como entrada para cada uma das execuções foi contruída a partir das 698 (seiscentos e noventa e oito) ideias obtidas na base de dados da *Ubuntu Brainstorm*<sup>®</sup>.

Uma vez realizada a extração do conjunto de dados, as ideias passaram por uma etapa de transformação. Nesta etapa de transformação os dados foram adequados para que pudessem ser contidos no modelo proposto, resultando em um conjunto de 498 ideias implementadas e 200 ideias não implementadas.

Os dados resultantes ainda foram submetidos ao algoritmo de análise de sentimento (ANACLETO, 2017; ALVAREZ, 2018) com o intuito de polarizar determinada ideia, indicando se esta é positiva, negativa ou neutra. Como resultado, 137 ideias foram classificadas como negativas, 181 ideias como neutras e 350 ideias como positivas.

Além disso, foram produzidos os índices de atualidade das palavras-chave identificadas na base de dados da *Semantic Scholar*<sup>®</sup>. Após isso, foi realizada a vinculação das palavras-chave às ideias, possibilitando gerar o índice de atualidade de determinada ideia conforme descrito na Seção 4.1.5.

Com o objetivo de avaliar o modelo proposto e demonstrar sua aplicabilidade e viabilidade, foram elaborados quatro cenários de agrupamento apresentados no Quadro 4.

Quadro 4 – Cenários de análise de agrupamento

Cenário	Ideias	Palavras-Chave	Índice de Atualidade
1	Sim	Não	Não
2	Sim	Sim	Não
3	Sim	Não	Sim
4	Sim	Sim	Sim

Fonte: Autor

No primeiro cenário o algoritmo de agrupamento leva em consideração apenas os dados do vetor da ideia, sem a inclusão das palavras-chave relacionadas à ideia e obtidas a partir da base de dados da *Semantic Scholar*<sup>®</sup>, e sem a inclusão do índice de atualidade. O vetor de determinada ideia é composto pelas palavras do título e descrição normalizadas pela medida *tf-idf* conforme Seção 4.1.6.

No segundo cenário são considerados pelo algoritmo de agrupamento os dados do vetor de ideias e as palavras-chave obtidas na base de dados da *Semantic Scholar*<sup>®</sup>, sem a inclusão do índice de atualidade das ideias.

No terceiro cenário além do vetor de ideias, o algoritmo utiliza o índice de atualidade das ideias gerado a partir do índice de atualidade das palavras-chave da base da *Semantic Scholar*<sup>®</sup>.

No quarto cenário todas as três informações são consideradas pelo algoritmo de agrupamento, ou seja, vetor da ideia, palavras-chave relacionadas e índice de atualidade.

De acordo com Alvarez (2018), apenas 273 ideias (39%) possuem *tags*. Neste sentido, o enriquecimento do vetor de ideias mediante a inclusão das palavras-chave identificadas na base de dados da *Semantic Scholar*<sup>®</sup> foi realizado com o propósito de verificar se o algoritmo produziria melhores resultados, gerando uma menor quantidade de grupos. De forma semelhante, foram elaborados cenários onde o índice de atualidade da ideia foi utilizado pelo algoritmo de agrupamento.

Para cada um dos cenários foram realizadas quatro execuções, cada uma delas com um limiar distinto. Por conta da elevada abrangência dos temas constantes na ideia optou-se por utilizar os mesmos limiares apresentados por Alvarez (2018): 0,05, 0,1, 0,15 e 0,20. Quanto maior for o limiar, menor será a similaridade entre os vetores e, portanto, mais agrupamentos são formados.

Conforme descrição realizada na Seção 5.2, a combinação destes limiares com os cenários resultou em 24 execuções, sendo as primeiras 16 execuções nomeadas como **Sessão 1** e as demais execuções como **Sessão 2**.

## 5.2 RESULTADOS DOS EXPERIMENTOS

Nesta sessão, serão apresentados os resultados obtidos mediante a aplicação do modelo proposto neste trabalho. Após determinar os cenários de análise a serem utilizados na avaliação foi realizada a primeira série de execuções (**Sessão 1**). Nesta série foram consideradas



as 698 ideias para cada um dos quatro cenários estabelecidos (Quadro 4).

O conjunto de dados resultantes foi armazenado em uma tabela (“*cluster*”) que possui os seguintes atributos:

- *execution\_id*: este campo é o identificador da execução;
- *cluster\_id*: este campo é o identificador do *cluster* para a execução;
- *idea\_id*: este campo é o identificador da ideia vinculada ao *cluster* da execução;
- *status*: este campo indica se a ideia está ou não implementada;
- *polarization*: este campo indica se a ideia é negativa, neutra ou positiva;
- *current\_index*: este campo indica o índice de atualidade da ideia.

A partir dos dados detalhados gerados para cada uma das execuções foi produzida a sumarização dos resultados por *cluster*. Para tanto foi gerada uma visão temporária (“*sumarized\_clustering\_executions*”) contendo os seguintes atributos:

- *execution\_id*: este campo é o identificador da execução;
- *cluster\_id*: este campo é o identificador do *cluster* para a execução;
- *nr\_ideas*: contagem do número de ideias por *cluster* e execução;
- *nr\_implented*: somatório das ideias implementadas por *cluster* e execução, considerando 1 para implementadas e 0 para não implementadas;
- *nr\_polarization*: somatório das ideias polarizadas por *cluster* e execução, considerando 1 para positivas, 0 para neutras e -1 para negativas;
- *current\_index*: somatório dos índices de atualidade das ideias por *cluster* e execução;
- *perc\_impl*: cálculo realizado para determinar o percentual de ideias implementadas do *cluster*;
- *perc\_pol*: cálculo realizado para determinar o percentual de ideias positivas do *cluster*;
- *session\_id*: dado adicional para identificar a sessão da execução;
- *scenario\_id*: dado adicional para identificar o cenário da execução;

- *threshold*: dado adicional para identificar o limiar aplicado para a execução de acordo com o cenário.

Sobre esta visão (“*sumarized\_clustering\_executions*”) foi gerada uma nova projeção (“*sumarized\_session*”) para sumarizar os dados da sessão. Esta nova projeção contém os seguintes atributos:

- *session\_id*: identificador da sessão da execução;
- *scenario\_id*: identificador do cenário da execução;
- *execution\_id*: identificador da execução;
- *limiar*: valor do limiar utilizado para a execução;
- *nr\_clusters*: contagem de *clusters* por execução;
- *avg\_cluster\_pol*: percentual médio das ideias com polaridade positiva dos *clusters* da execução;
- *avg\_cluster\_impl*: percentual médio das ideias implementadas dos *clusters* da execução
- *current\_index*: somatório do índice de atualidade dos *clusters* da execução.

Na Tabela 2 são apresentados os dados sumarizados da primeira sessão de execuções, cujo somatório dos índices de atualidade é 8908, com as seguintes colunas:

- Cenário: identifica o cenário de execução de acordo com o Quadro 4;
- Execução: identifica o número da execução;
- Limiar: identifica o limiar utilizado pelo algoritmo de agrupamento para determinar o quão semelhante uma ideia precisa ser em relação à outra ideia de um mesmo *cluster*;
- Num. de *Clusters*: registra a quantidade de grupos gerados para a execução;
- Polaridade (Média): identifica o percentual médio de ideias positivas nos *clusters*;
- Implementadas (Média): identifica o percentual médio de ideias implementadas nos *clusters*.

Tabela 2 – Resultado da primeira sessão de agrupamento

Cenário	Execução	Limiar	Num. de Clusters	Polaridade (Média)	Implementadas (Média)
1	1	0,05	76	67,297%	49,230%
1	5	0,1	232	69,210%	52,901%
1	9	0,15	358	71,246%	53,573%
1	13	0,2	457	70,441%	51,776%
2	2	0,05	101	68,719%	52,067%
2	6	0,1	164	67,966%	50,204%
2	10	0,15	211	70,359%	49,424%
2	14	0,2	255	72,655%	49,028%
3	3	0,05	36	63,996%	45,973%
3	7	0,1	93	68,290%	49,659%
3	11	0,15	152	69,043%	47,820%
3	15	0,2	201	71,242%	50,844%
4	4	0,05	73	63,452%	49,870%
4	8	0,1	142	67,814%	51,420%
4	12	0,15	196	71,129%	51,387%
4	16	0,2	245	72,313%	49,282%

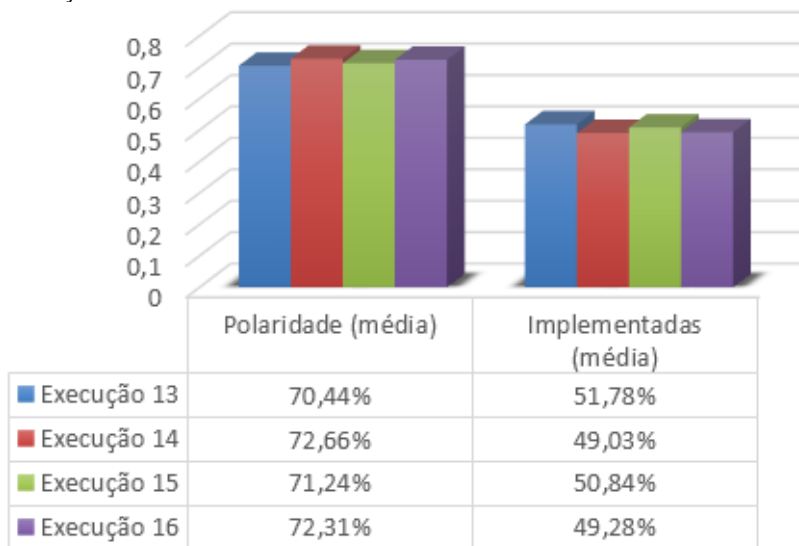
Fonte: Autor

Tomando o resultado sumarizado das 16 execuções da primeira sessão (**Sessão 1**) foi realizada uma análise comparativa entre os diferentes cenários para cada limiar. Para uma análise preliminar adotou-se o limiar 0,20, limitando-se às execuções 13, 14, 15 e 16.

Na Figura 26 é apresentado um gráfico comparativo do percentual médio de polaridade e de ideias implementadas das execuções 13, 14, 15 e 16. Os valores representados neste gráfico possuem pouca variação, pois são uma média sobre a polaridade e implementação das ideias de todos os grupos. A variação ocorre principalmente pela precisão adotada para o campo.

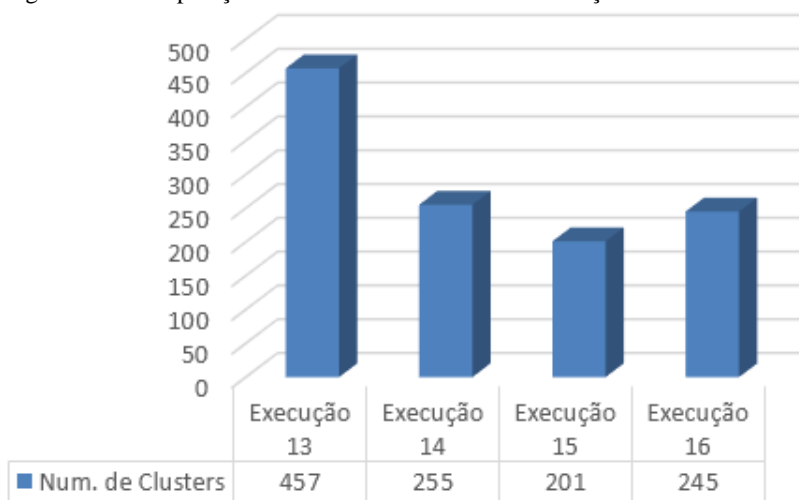
Diferentemente, a quantidade de grupos gerados por execução (Figura 27) apresenta resultados bem distintos. Estes resultados ocorrem por conta das variáveis consideradas na composição do vetor das ideias, conforme descrito em cada um dos cenários utilizados.

Figura 26 – Comparação da média de polaridade e ideias implementadas das execuções 13 a 16



Fonte: Autor

Figura 27 – Comparação do número de clusters das execuções 13 a 16



Fonte: Autor

Na execução 13 foi utilizado o **Cenário 1** que prevê a utilização somente dos dados do vetor original da ideia da base de dados da Ubuntu. Para tal, foram gerados 457 *clusters* sendo em média 70% positivos e com 51% das ideias implementadas. Contudo, considerando a esparsidade do conjunto de dados, ou seja, a diversidade de assuntos tratados nas ideias, muitos *clusters* acabaram sendo gerados com somente uma ideia.

Na execução 14 foi utilizado o **Cenário 2** que prevê o enriquecimento do vetor original com as palavras-chave obtidas a partir da *Semantic Scholar*<sup>®</sup>. Nesta execução, foram gerados 255 grupos, sendo em média 70% positivos e 49% implementados. O enriquecimento do vetor de ideias com as palavras-chave refletiu diretamente no resultado do agrupamento, possibilitando agrupar mais ideias em um mesmo *cluster*.

Na execução 15 foi utilizado o **Cenário 3** que enriquece as entradas do algoritmo de agrupamento (vetores de ideias) com o índice de atualidade das ideias. Nesta execução foram gerados 201 *clusters*, destes em média 71% positivos e 50% implementados.

Na execução 16 foi utilizado o **Cenário 4** onde além do enriquecimento do vetor das ideias com as palavras-chave obtidas a partir da *Semantic Scholar*<sup>®</sup>, o algoritmo considerou o índice de atualidade das ideias. Nesta execução foram gerados 245 *clusters*, sendo em média 72% positivos e 49% implementados.

Este comportamento de variação da quantidade de grupos de acordo com a escolha dos cenários se repete em todos os limiares. No entanto, ao observar o comportamento dos algoritmos de agrupamento para os **Cenários 3 e 4** é possível observar que a inclusão do índice de atualidade no agrupamento compromete sua eficácia.

O algoritmo utilizado baseia-se no conceito de densidade e se propõe a gerar grupos de ideias similares, de forma que as ideias dentro de um grupo sejam mais parecidas entre si e mais distintas possíveis das ideias de outros grupos.

Como mencionado, o agrupamento das ideias utilizando apenas os dados do vetor original, para a **Execução 13**, gerou 457 grupos. Contudo ao agregar o índice de atualidade de cada ideia ao seu vetor este número foi reduzido para 201 grupos. Esta redução expressiva permite concluir que em função do valor elevado do índice de atualidade, a formação dos agrupamentos ocorre basicamente considerando esta dimensão desconsiderando as demais dimensões do vetor das ideias.

Por conta deste comportamento, para as análises subseqüentes foram eliminados os **Cenários 3 e 4**, pois estes utilizam a informação do índice de atualidade das ideias pelo algoritmo de agrupamento.

Uma vez eleitos os cenários a serem analisados e considerando a esparsidade das ideias, bem como, o conjunto de dados reduzido, optou-se por detalhar a análise realizada para os **Cenários 1 e 2** com o limiar **0,05**.

Para tanto foi realizada a projeção dos dados das **Execuções 1 e 2**. Na Tabela 3 são apresentados os 10 *clusters* com maior índice de atualidade produzidos na **Execução 1, Cenário 1, Sessão 1** e limiar **0,05**. Destes, 5 (cinco) *clusters* possuem 50% ou mais das ideias positivas e 5 (cinco) menos de 50%. Dentre os 5 primeiros *clusters*, há uma variabilidade entre 60,87% e 88% das ideias implementadas.

Tabela 3 – Os dez primeiros *clusters* da **Execução 1** ordenados pelo índice de atualidade

Cluster Id	Índice de Atualidade	Num. de Ideias	% implementadas	% positivas
17	524,882	20	75,00%	40,00%
11	460,678	23	60,87%	52,17%
21	387,211	30	63,33%	46,67%
8	355,861	22	86,36%	36,36%
12	305,000	25	88,00%	80,00%
14	301,640	18	33,33%	50,00%
3	285,264	16	75,00%	50,00%
6	263,376	16	81,25%	31,25%
58	249,715	9	66,67%	44,44%
36	246,848	14	78,57%	50,00%

Fonte: Autor

Na Tabela 4 são apresentados os 10 primeiros *clusters* gerados na **Execução 2, Cenário 2, Sessão 1** e limiar **0,05**. Destes, 5 (cinco) *clusters* possuem 50% ou mais das ideias positivas e outros 5 (cinco) menos de 50%. Dentre os primeiros 5 (cinco) *clusters*, há uma variância entre 66,67% e 85,71% das ideias implementadas.

Considerando os objetivos deste trabalho, que consistem na construção de um modelo voltado à gestão de ideias por meio de técnicas de análise de agrupamentos, objetiva-se que após a execução do

algoritmo de agrupamento seja possível indicar grupos de ideias com potencial de implementação considerando o índice de atualidade desses grupos, em que os primeiros são os mais relevantes.

Tabela 4 – Os dez primeiros *clusters* da **Execução 2** ordenados pelo índice de atualidade

Cluster Id	Índice de Atualidade	Num. de Ideias	% implementadas	% positivas
16	869,789	21	76,19%	57.14%
1	583,086	7	85,71%	42.86%
2	564,113	22	81,82%	63.64%
8	444,631	47	76,60%	51.06%
58	438,226	9	66,67%	66.67%
72	419,712	29	75,86%	44.83%
30	387,948	34	55,88%	55.88%
25	352,398	6	100,00%	16.67%
19	322,760	35	51,43%	37.14%
24	285,940	24	75,00%	33.33%

Fonte: Autor

Por conta do resultado obtido nas execuções da **Sessão 1**, foram consideradas as abordagens adotadas por Alvarez (2018) para realizar o aprimoramento do algoritmo de agrupamento. De acordo com Alvarez (2018), a partir da disposição das ideias em grupos considerando a relação estado-polaridade, busca-se analisar se ideias positivas ou negativas tendem ou não a serem implementadas.

Partindo deste princípio, as 211 ideias classificadas como neutras não contribuem para esta análise e, portanto, foram retiradas do estudo. Na segunda sessão (**Sessão 2**), as ideias neutras não foram consideradas no processo de agrupamento.

Na Tabela 5, são apresentados os dados sumarizados da segunda sessão de execuções, cujo somatório dos índices de atualidade de todas as ideias que compõem este estudo é 6131, com as seguintes colunas:

- Cenário: identifica o cenário de execução de acordo com o Quadro 4;
- Execução: identifica o número da execução;

- Limiar: identifica o limiar utilizado pelo algoritmo de agrupamento para determinar o quão semelhante uma ideia precisa ser em relação à outra ideia de um mesmo *cluster*;
- Num. *clusters*: registra a quantidade de grupos gerados para a execução;
- Polaridade (Média): identifica o percentual médio de ideias positivas nos *clusters*;
- Implementadas (Média): identifica o percentual médio de ideias implementadas nos *clusters*.

Tabela 5 – Resultado da segunda sessão de agrupamento

Cenário	Exec.	Limiar	Num. de Clusters	Polaridade (Média)	Implementadas (Média)
1	17	0,05	65	66,450%	74,097%
1	21	0,1	185	72,582%	72,551%
1	25	0,15	292	73,918%	74,138%
1	29	0,2	349	73,544%	72,431%
2	18	0,05	89	70,037%	70,621%
2	22	0,1	144	73,084%	72,537%
2	26	0,15	176	74,758%	69,541%
2	30	0,2	198	76,586%	70,310%

Fonte: Autor

Tomando o resultado sumarizado das 8 execuções da segunda sessão foi realizada uma análise comparativa entre os cenários 1 e 2 para o limiar de **0,05**. Para a segunda sessão de execuções foram desconsiderados os cenários 3 e 4 uma vez que ao considerar o índice de atualidade o processo de agrupamento tende a agregar as ideias considerando somente esta dimensão.

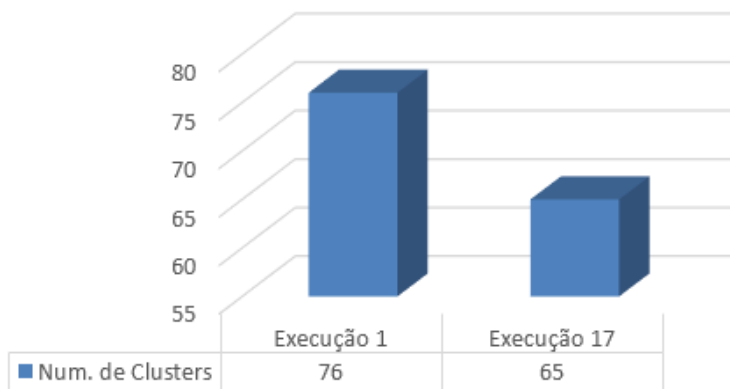
Apesar da redução da soma do índice de atualidade geral, reflexo da remoção das ideias neutras, ao comparar os resultados das execuções referentes a **Sessão 1** (Tabela 2) e **Sessão 2** (Tabela 5), é possível observar uma evidente melhora no resultado final na etapa de agrupamento.

Estabelecendo um comparativo entre as **Sessão 1** e a **Sessão 2**, a execução do **Cenário 1** utilizando o limiar 0,05 (**Execuções 1 (Sessão 1)** e **17 (Sessão 2)**) ocorre uma redução de 76 para 65 agrupamentos (Figura 28), de 67,30% para 66,45% de ideias positivas e produz um



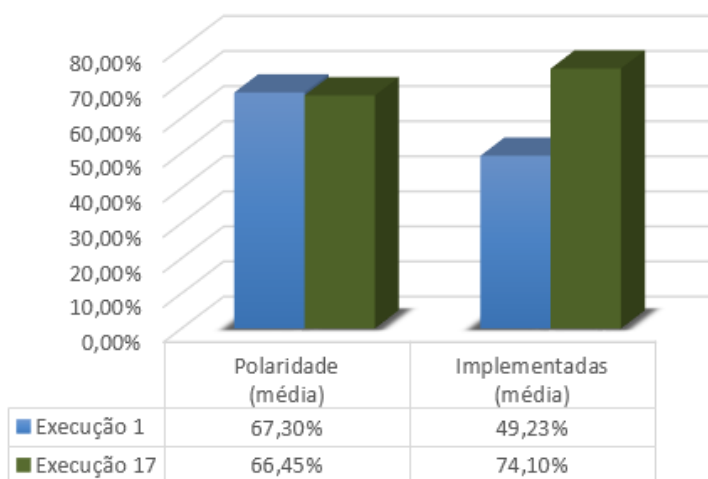
aumento de 49,23% para 74,10% de ideias implementadas, em média por *cluster* (Figura 29).

Figura 28 – Comparação do número de clusters das execuções 1 e 17



Fonte: Autor

Figura 29 – Comparação da média de polaridade e ideias implementadas das execuções 1 e 17



Fonte: Autor

Analisando os resultados da **Execução 17** apresentados na Tabela 6 a melhoria dos resultados da execução 17 em relação aos resultados da execução 1 (Tabela 3) se tornam evidentes.

Na Tabela 6 são apresentados os primeiros 10 *clusters* gerados na **Execução 17, Cenário 1, Sessão 2** e limiar **0,05**. Destes, 9 (nove) *clusters* possuem mais que 50% das ideias positivas, sendo 4 (quatro) acima de 80%, e apenas 1 (um) igual a 50%. Dentre os primeiros 5 (cinco) *clusters*, há uma variabilidade entre 55% e 100% das ideias implementadas.

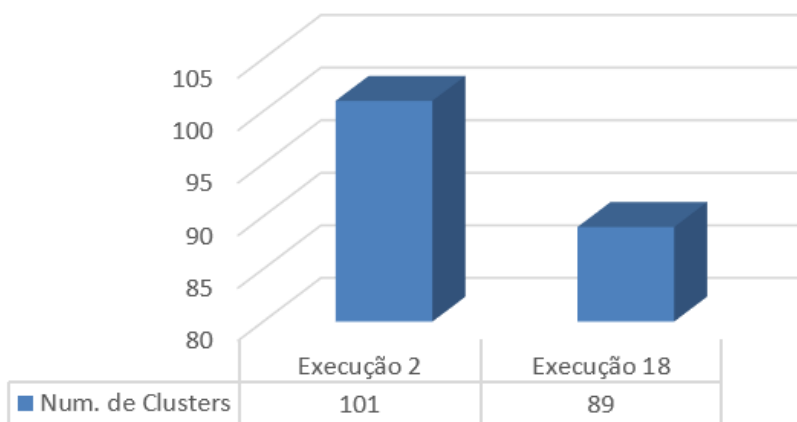
Tabela 6 – Os dez primeiros *clusters* da **Execução 17** ordenados pelo índice de atualidade

Cluster Id	Índice de Atualidade	Num. de Ideias	% implementadas	% positivas
20	399,264	5	100,00%	60,00%
7	324,518	22	95,45%	63,64%
3	323,472	12	75,00%	50,00%
22	317,203	18	55,56%	88,89%
31	288,829	22	72,73%	68,18%
15	232,401	20	65,00%	85,00%
36	217,486	12	91,67%	66,67%
21	210,602	16	62,50%	62,50%
1	202,190	8	87,50%	87,50%
13	198,805	18	72,22%	83,33%

Fonte: Autor

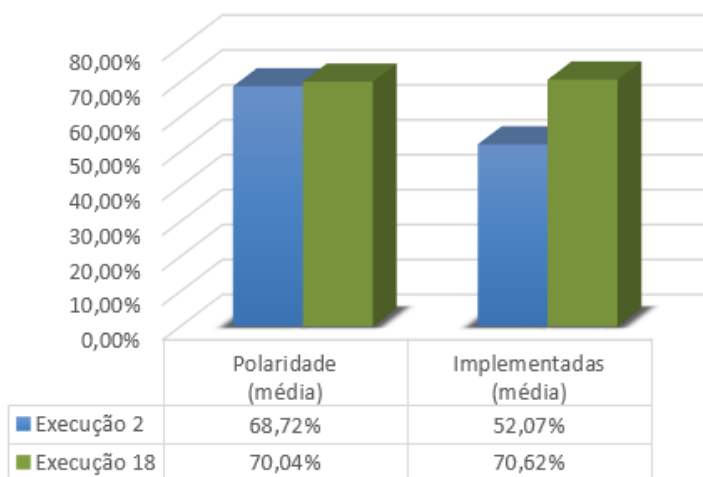
Resultado semelhante ocorre nos resultados sumarizados para o **Cenário 2** e limiar **0,05** considerando a **Execução 2 (Sessão 1)** e **Execução 18 (Sessão 2)**. Há uma redução de 101 para 89 agrupamentos gerados (Figura 30), aumento médio por *cluster* de 68,72% para 70,04% de ideias positivas e de 52,07% para 70,62% de ideias implementadas (Figura 31).

Figura 30 – Comparação do número de clusters das execuções 2 e 18



Fonte: Autor

Figura 31– Comparação da média de polaridade e ideias implementadas das execuções 2 e 18



Fonte: Autor

Analisando a Tabela 7 a melhoria dos resultados da **Execução 18** em relação aos resultados da **Execução 2** (Tabela 4) e **Execução 17** (Tabela 6) se tornam evidentes.

Na Tabela 7 são apresentados os primeiros 10 *clusters* gerados na **Execução 18, Cenário 2, Sessão 2** e limiar 0,05. Destes, todos possuem mais que 50% das ideias positivas, sendo 3 acima de 80%, 3 acima de 70%, 3 acima de 60%, e apenas 1 igual a 50%. Dentre os primeiros 5 *clusters*, há uma variabilidade entre 65% e 100% de ideias implementadas e de 60% a 100% de ideias positivas nos *clusters*.

Tabela 7 – Os dez primeiros *clusters* da **execução 18** ordenados pelo índice de atualidade

Cluster Id	Índice de Atualidade	Num. de Ideias	% implementadas	% positivas
18	763,689	20	65,00%	80,00%
2	584,210	23	73,91%	78,26%
39	582,132	53	60,38%	67,92%
1	385,968	5	100,00%	60,00%
84	277,177	1	100,00%	100,00%
17	265,156	29	51,72%	79,31%
14	234,988	14	85,71%	78,57%
46	229,605	17	82,35%	52,94%
4	225,656	10	90,00%	60,00%
20	186,063	7	14,29%	85,71%

Fonte: Autor

O resultado das comparações das execuções **1, 2, 17 e 18** respaldam a melhoria do algoritmo quando há o enriquecimento do vetor das ideias com as palavras-chave obtidas na base de dados da *Semantic Scholar*<sup>®</sup>. É notável também a melhoria dos resultados das execuções da **Sessão 2**, onde são removidas as ideias neutras.

Buscando entender este comportamento, são apresentadas abaixo duas tabelas (Tabela 8 e Tabela 9) com o resultado dos primeiros 10 *clusters* da **Execução 6 e Execução 22**. A **Execução 6** foi realizada na **Sessão 1** considerando ideias neutras, enquanto que a **Execução 22** foi realizada na **Sessão 2**, sem ideias neutras. Ambas as execuções, **6 e 22**, estão enquadradas no **Cenário 2** e utilizaram o limiar **0,10**.

Na Tabela 8 são apresentados os primeiros 10 *clusters* gerados na **Execução 6, Cenário 2, Sessão 1** e limiar **0,10**. Destes, 5 possuem mais que 50% das ideias positivas, sendo 1 *cluster* com 100% de ideias positivas, 3 acima de 60% e 1 acima de 50%. Dentre os primeiros 5

*clusters*, há uma variabilidade entre 55,26% e 85,71% das ideias implementadas.

Tabela 8 – Os dez primeiros *clusters* da **Execução 6** ordenados pelo índice de atualidade

<b>Cluster Id</b>	<b>Índice de Atualidade</b>	<b>Num. de Ideias</b>	<b>% implementadas</b>	<b>% positivas</b>
19	869,79	21	76,19%	57,14%
1	583,09	7	85,71%	42,86%
86	438,23	9	66,67%	66,67%
60	436,01	38	55,26%	60,53%
32	359,91	14	85,71%	71,43%
30	352,40	6	100,00%	16,67%
145	277,18	1	100,00%	100,00%
67	273,18	7	28,57%	42,86%
29	257,68	23	73,91%	30,43%
96	248,87	27	48,15%	44,44%

Fonte: Autor

Na Tabela 9 são apresentados os primeiros 10 *clusters* gerados na **Execução 22, Cenário 2, Sessão 2** e limiar **0,10**. Destes, todos possuem 50% ou mais das ideias classificadas como positivas. Dentre os *clusters*, 1 possui apenas 1 ideia (100%) positiva, 2 acima de 80%, 3 acima de 70%, outros 3 acima de 60% e apenas 1 com 50%. Dentre os primeiros 5 *clusters*, há uma variabilidade entre 65,26% e 100% das ideias implementadas.

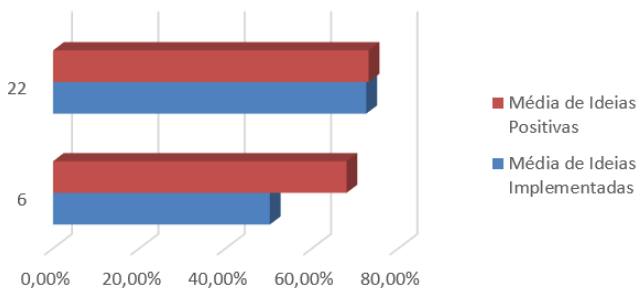
Tabela 9 – Os dez primeiros *clusters* da **Execução 22** ordenados pelo índice de atualidade

Cluster Id	Índice de Atualidade	Num. de Ideias	% implementadas	% positivas
22	763,69	20	65,00%	80,00%
27	419,29	39	69,23%	69,23%
1	385,97	5	100,00%	60,00%
16	289,42	20	65,00%	75,00%
30	289,28	13	84,62%	84,62%
133	277,18	1	100,00%	100,00%
21	250,81	26	50,00%	76,92%
65	223,22	16	87,50%	50,00%
2	211,00	11	90,91%	72,73%
5	193,98	9	100,00%	66,67%

Fonte: Autor

Na Figura 32 é estabelecido um comparativo entre as **Execução 6 (Sessão 1)** e **Execução 22 (Sessão 2)**, ambas enquadradas no **Cenário 1** e utilizando o limiar **0,10**. É possível observar o aumento de 67,97% para 73,08% de ideias positivas e de 50,20% para 72,53% de ideias implementadas, em média por *cluster*. De acordo com as informações obtidas na Tabela 2 e na Tabela 5 ocorre ainda uma redução de 164 para 144 grupos gerados.

Figura 32 – Comparativo entre a **execução 6** e a **execução 22**



Fonte: Autor

Os resultados obtidos na comparação realizada com limiares de **0,10** confirmam o comportamento identificado nas comparações anteriores. As execuções da **Sessão 2** que desconsideram as ideias neutras apresentam melhores resultados. De modo geral, isso significa que os primeiros *clusters* são aqueles que possuem mais ideias identificadas como positivas e que foram implementadas. Vale ressaltar que em um processo real não se sabe *a priori* se determinada ideia será ou não implementada. Todavia, os resultados permitem indicar que um processo de agrupamento utilizando algum mecanismo de ordenação é capaz de apresentar primeiro os melhores agrupamentos de ideias, ou seja, os agrupamentos compostos em sua maioria por ideias positivas e que tendem a ser implementadas.

### 5.3 ANÁLISE DOS AGRUPAMENTOS

No tópico anterior foram apresentados os resultados das execuções das **Sessões 1 e 2**. Após a análise das execuções da primeira sessão, os **Cenários 3 e 4** foram desconsiderados da análise e foi realizada uma nova sessão de execuções com os **Cenários 1 e 2**, contudo somente considerando as ideias positivas e negativas.

A análise comparativa das execuções permitiu perceber melhoria nos resultados nas execuções do **Cenário 2** em relação às execuções do **Cenário 1**. Também apresenta melhor resultado na ordem dos agrupamentos gerados na **Sessão 2**, sem ideias neutras. Por conta da grande profusão de temas nas ideias as análises que produziram agrupamentos mais densos utilizam o limiar 0,05.

Considerando estes fatores são apresentados os resultados dos primeiros 5 agrupamentos da execução 18. A execução 18 está enquadrada no **Cenário 2** (Quadro 4), **Sessão 2** (sem ideias neutras) e com o limiar 0,05.

Na Figura 33 é apresentado um gráfico que demonstra a proporção do *status* pela polaridade dos primeiros 5 *clusters* da **Execução 18**, conforme Tabela 7. Para cada *cluster* existem quatro combinações possíveis:

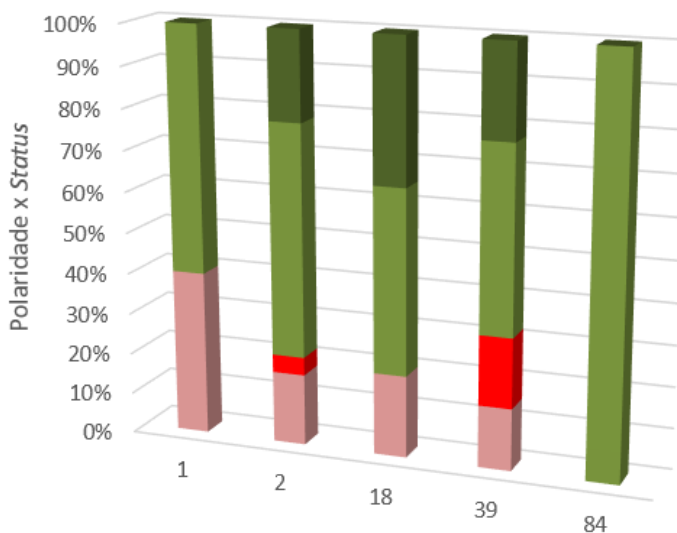
- **POSITIVE – Won't Implemented**: Representa o conjunto de ideias que foram classificadas como positivas, mas não foram implementadas;
- **POSITIVE – Implemented**: Representa o conjunto de ideias que foram marcadas como positivas e foram implementadas. Este conjunto de ideias pode ser considerado por especialistas

como parâmetro para as demais ideias do grupo, ou até mesmo para combinação delas e geração de novas ideias;

- *NEGATIVE – Won't Implemented*: Representa o conjunto de ideias negativas e não implementadas. Espera-se que nos agrupamentos com melhor posição no ranking este conjunto de ideias tenha a menor representação possível;
- *NEGATIVE – Implemented*: Representa o conjunto de ideias que foi classificado como negativo, contudo, foi implementado. Requer análise de especialista para identificar as razões pelas quais foi implementado, mesmo sendo uma ideia negativa. O retorno desta análise pode ser o aprimoramento do algoritmo de análise de sentimentos, ou agrupamento, ou ainda necessidades inerentes ao negócio.



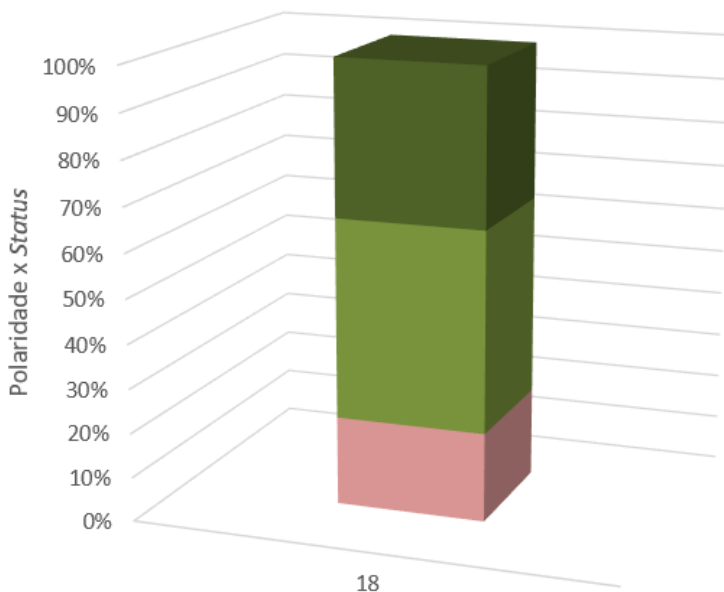
Figura 33 – Os primeiros cinco *clusters* da **execução 18** (polaridade x *status*)



	1	2	18	39	84
■ POSITIVE - Won't Implement	0,00%	21,74%	35,00%	22,64%	0,00%
■ POSITIVE - Implemented	60,00%	56,52%	45,00%	45,28%	100,00%
■ NEGATIVE - Won't Implement	0,00%	4,35%	0,00%	16,98%	0,00%
■ NEGATIVE - Implemented	40,00%	17,39%	20,00%	15,09%	0,00%

Fonte: Autor

Na Figura 34 é apresentada a relação entre a polarização *versus* o *status* das ideias do *cluster* 18 que de acordo com os dados da Tabela 7, ficou classificado como primeiro *cluster* com 763,689 pontos de índice de atualidade, 65% das ideias implementadas e 80% das ideias positivas.

Figura 34 – Análise da relação polaridade x *status* do *cluster* 18

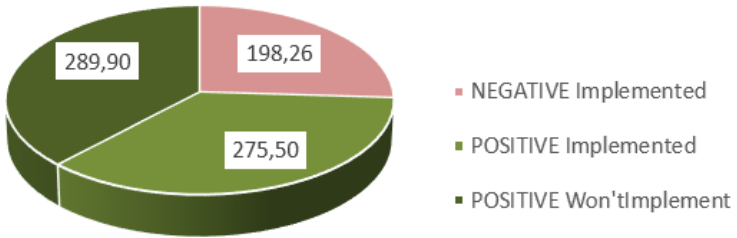
	18
■ POSITIVE - Won'tImplement	35,00%
■ POSITIVE - Implemented	45,00%
■ NEGATIVE - Implemented	20,00%

Fonte: Autor

Este conjunto de dados de ideias positivas (80%) está dividido em dois sub-conjuntos: implementadas (45%) e não implementadas (35%). Dentre as ideias classificadas como negativas, nenhuma foi implementada.

Na Figura 35 é apresentado o somatório do índice de atualidade das ideias, agrupados por polarização e *status* do *cluster* 18. O somatório das ideias positivas contabiliza 565,4 pontos, o equivalente a 74,03% da pontuação total do *cluster*. Este valor por si só é representativo, de forma que, ainda que as ideias negativas fossem removidas do *cluster*, este ainda estaria entre os primeiros 10 *clusters* da execução.

Figura 35 – Somatório do índice de atualidade por polaridade e status do cluster 18



Fonte: Autor

Os termos apresentados na Figura 36 indicam que os principais assuntos relacionados às ideias do cluster 18. Os termos mencionados ao menos 3 vezes são: *Name*: 20 ocorrências; *Ubuntu*: 10 ocorrências; *HTTP*: 4 ocorrências; *Brainstorm*: 3 ocorrências; *Nautilus*: 3 ocorrências; *Version*: 3 ocorrências; *Wiki*: 3 ocorrências.

Figura 36 – Nuvem de termos do cluster 18



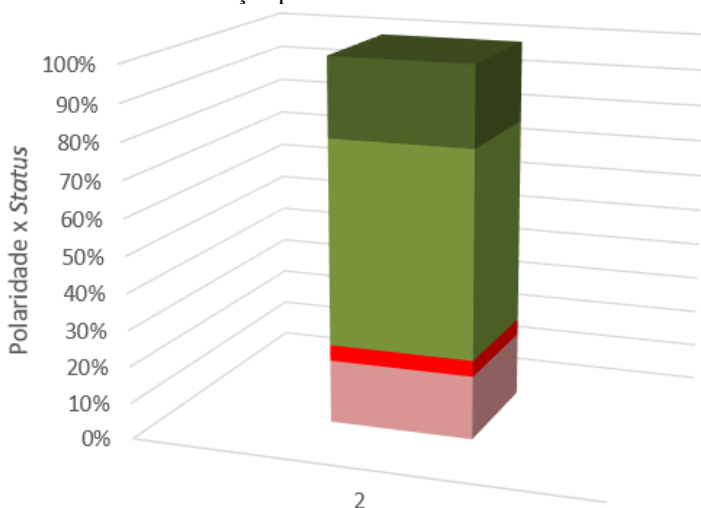
Fonte: Autor

Na Figura 37 é apresentada a relação entre a polarização *versus* o status das ideias do cluster 2 que de acordo com os dados da Tabela 7, ficou classificado como segundo cluster do ranking com 584,210 pontos no índice de atualidade, 73,91% das ideias implementadas e 78,26% das ideias positivas.

Neste conjunto de dados as ideias positivas (78,26%) estão divididas em dois sub-conjuntos: implementadas (56,52%) e não implementadas (21,74%). Dentre as ideias classificadas como negativas, 17,39% foram implementadas e 4,35% não foram implementadas.

De acordo com Alvarez (2018), existem alguns motivos pelos quais uma ideia implementada pode ter sido classificada como negativa, como por exemplo, a marcação indevida da ideia por conta da apresentação de comentários descontentes, mesmo que os comentários fossem sarcásticos ou evidenciem apoio à ideia reportada.

Figura 37 – Análise da relação polaridade x status do cluster 2



	2
■ POSITIVE - Won'tImplement	21,74%
■ POSITIVE - Implemented	56,52%
■ NEGATIVE - Won'tImplement	4,35%
■ NEGATIVE - Implemented	17,39%

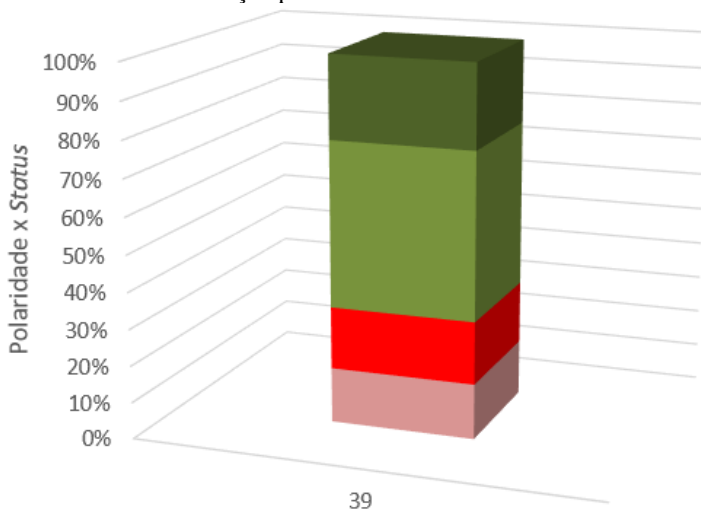
Fonte: Autor



Na Figura 40 é apresentada a relação entre a polaridade *versus* o *status* das ideias do *cluster* 39 que de acordo com os dados da Tabela 7, ficou classificado como terceiro *cluster* do *ranking* com 582,132 pontos de índice de atualidade, 60,38% das ideias implementadas e 67,92% das ideias positivas.

Neste conjunto de dados as ideias positivas (67,92%) estão divididas em dois sub-conjuntos: implementadas (45,28%) e não implementadas (22,64%). Dentre as ideias classificadas como negativas (32,08%), 15,09% foram implementadas e 16,98% não foram implementadas.

Figura 40 – Análise da relação polaridade x *status* do *cluster* 39

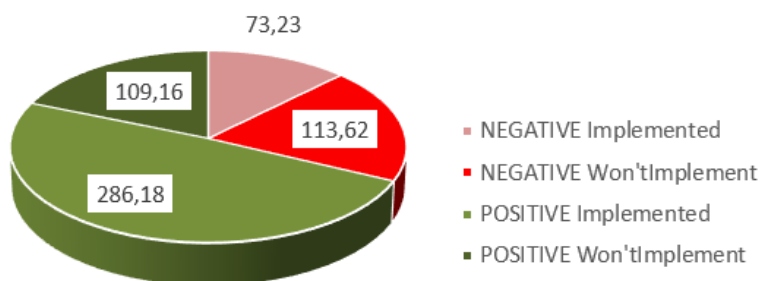


	39
■ POSITIVE - Won'tImplement	22,64%
■ POSITIVE - Implemented	45,28%
■ NEGATIVE - Won'tImplement	16,98%
■ NEGATIVE - Implemented	15,09%

Fonte: Autor

Na Figura 41 é apresentado o somatório do índice de atualidade das ideias, agregados por polaridade e *status* do *cluster* 39. O somatório das ideias positivas contabiliza 395,34 pontos, o equivalente a 67,91% da pontuação total do *cluster*. Este valor pode ser visto como representativo, pois caso as ideias negativas fossem removidas do *cluster* este ainda estaria entre os primeiros 10 *clusters* da execução. Todavia representa uma redução quando comparada à pontuação do *cluster* 2.

Figura 41 – Somatório do índice de atualidade por polarização e *status* do *cluster* 39



Fonte: Autor

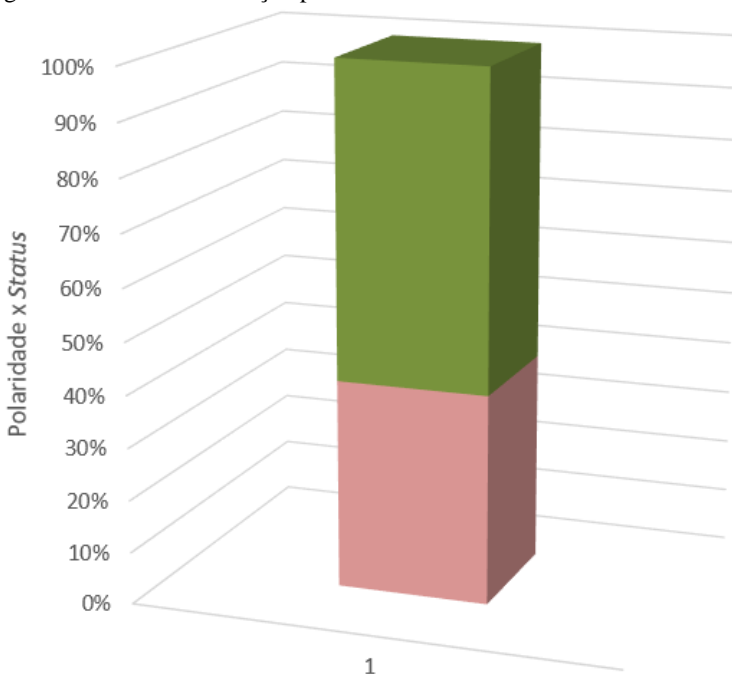
Os termos apresentados na Figura 42 indicam que os principais assuntos relacionados às ideias do *cluster* 39. Os termos mais mencionados são: *Ubuntu*: 33 ocorrências, *Version*: 19 ocorrências, *Repository*: 17 ocorrências, *Download*: 12 ocorrências, *One*: 10 ocorrências; *Information*: 9 ocorrências.

Figura 42 – Nuvem de termos do *cluster* 39

Fonte: Autor

Na Figura 43 é apresentada a relação entre a polaridade *versus* o *status* das ideias do *cluster* 1 que de acordo com os dados da Tabela 7, ficou classificado como quarto *cluster* do *ranking* com 385,132 pontos de índice de atualidade, 100% das ideias implementadas e 60% das ideias positivas.



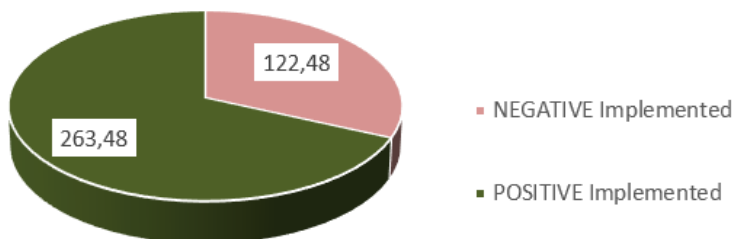
Figura 43 – Análise da relação polaridade x *status* do *cluster* 1

	1
■ POSITIVE - Implemented	60,00%
■ NEGATIVE - Implemented	40,00%

Fonte: Autor

Na Figura 44 é apresentado o somatório do índice de atualidade das ideias, agrupados por polarização e *status* do *cluster* 1. O somatório das ideias positivas contabiliza 263,48 pontos, o equivalente a 68,41% da pontuação total do *cluster*. Este é representativo, de forma que, ainda que as ideias negativas fossem removidas do *cluster*, este ainda estaria entre os primeiros 10 *clusters* da execução.

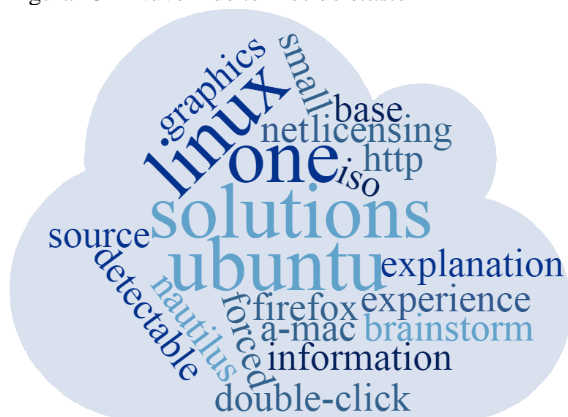
Figura 44 – Somatório do índice de atualidade por polaridade e *status* do *cluster* 1



Fonte: Autor

Os termos apresentados na Figura 45 indicam os principais assuntos relacionados às ideias do *cluster* 1. Os termos mais mencionados são: *Solutions*: 5 ocorrências, *Ubuntu*: 3 ocorrências; *One*: 2 ocorrências; *Linux* 2 ocorrências.

Figura 45 – Nuvem de termos do *cluster* 1



Fonte: Autor.

De acordo com os dados apresentados na Tabela 7, o *cluster* 84 ficou classificado em quinto lugar no *ranking*. Apesar de possuir apenas uma ideia, além de totalizar 277,18 pontos de índice de atualidade ela foi implementada e marcada como positiva.

Este valor expressa a importância da ideia. Mesmo que isoladamente ela possui um alto valor de atualidade, ou seja, tende a ser uma ideia com potencial para implementação.

Os termos mencionados na ideia do *cluster* 84 são: *Smoothing, Computer, Algorithm, Times, Default, Laptop*. O título da ideia em questão é “*Enable subpixel smoothing on LCDs by default*”, sugerindo ativar a suavização de subpixel em LCDs por padrão. Para tanto, o usuário justifica que as fontes padrão são suavizadas usando o algoritmo *Best Shapes*, um vez que os monitores LCDs têm uma resolução horizontal três vezes maior e, portanto, as fontes podem ser três vezes mais lisas se a suavização de *subpixel* estiver ativada.

#### 5.4 CONSIDERAÇÕES FINAIS

Nesse capítulo foram apresentados os resultados do modelo proposto com o objetivo de avaliar sua viabilidade. Para tanto, foram executados 3 grupos de atividades: a) especificação e elaboração dos cenários; b) execução dos algoritmos e apresentação dos resultados dos experimentos; e c) análise dos agrupamentos melhores posicionados na ordenação da execução considerada ideal para os cenários apresentados.

Para avaliação do modelo proposto foi inicialmente realizada uma sessão de execuções com os 4 (quatro) cenários estipulados. Para cada um dos cenários foi realizada uma execução com os limiares 0.05, 0.1, 0.15 e 0.2, resultando nas execuções apresentadas na Tabela 2.

Após a avaliação dos resultados da **Sessão 1**, foi identificado que os cenários 3 e 4, que incluíam o índice de atualidade como entrada no algoritmo de agrupamento distorciam o resultado. Também foi verificada evidente melhora dos resultados do **Cenário 2** em relação aos resultados do **Cenário 1**.

No entanto, ao avaliar os resultados finais dos agrupamentos, o percentual de ideias implementadas nos *clusters* mais bem posicionados no *ranking* ficou abaixo do esperado.

Como o objetivo da avaliação é compreender o *ranking* dos *clusters* utilizando o índice de atualidade e verificar a relação estado-polaridade das ideias melhores classificadas, elaborou-se uma nova sessão de execuções.

A **Sessão 2** passou a considerar apenas os **Cenários 1 e 2**, executando-os para cada um dos limiares anteriormente estabelecido. Ao todo foram 8 (oito) execuções, apresentadas na Tabela 5.

Ao estabelecer um quadro comparativo entre as execuções da **Sessão 1** e **Sessão 2**, foi possível observar uma evolução no resultado final da classificação dos agrupamentos.

Considerando grande quantidade de temas tratados nas ideias, o que pode ser entendido como fator de esparcidade dos dados, o melhor desempenho do **Cenário 2** em relação ao **Cenário 1**, e da **Sessão 2** em relação à **Sessão 1**, foi selecionada a execução 18, enquadrada nestes critérios para realizar a estratificação dos resultados do agrupamento.

Ao realizar a análise dos primeiros 5 agrupamentos gerados na execução 18 foi possível observar que estes tendem a ser altamente positivos e implementados. Foi possível verificar também que uma ideia considerada atual através do índice de atualidade pode figurar no topo do classificação, permitindo sinalizar uma oportunidade efetiva de implementação.

## 6 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

Neste capítulo são apresentadas as considerações finais referentes ao desenvolvimento deste trabalho, bem como, as contribuições geradas e os trabalhos futuros.

### 6.1 CONSIDERAÇÕES FINAIS

Este trabalho apresentou um modelo baseado em Análise de Agrupamento com ordenação de grupos de ideias com o objetivo de auxiliar a tomada de decisão no processo de Gestão de Ideias.

Para construção deste modelo foram utilizados métodos de suporte providos pelas áreas de Mineração de Dados e Análise de Sentimento, com foco específico na Análise de Agrupamentos. O modelo elaborado foi aplicado no contexto de Gestão de Ideias e compreende as seguintes etapas:

- A extração, transformação e carga das fontes de dados de ideias e artigos publicados;
- O cálculo do índice de atualidade das palavras-chave identificadas na base de dados de artigos da Semantic Scholar<sup>®</sup>;
- O enriquecimento das ideias vinculando as palavras-chave a partir da base de artigos;
- O cálculo do índice de atualidade das ideias através do cálculo da média dos índices de atualidade das palavras-chave associadas a determinada ideia;
- A geração e ordenação dos agrupamentos pelo índice de atualidade de cada grupo;
- Análise e explicitação dos agrupamentos de acordo com a ordenação.

A fonte de dados utilizada para obter as ideias foi o Sistema de Gestão de Ideias Ubuntu Brainstorm<sup>®</sup>, uma plataforma de *crowdsourcing* que possibilita aos usuários registrar ideias, sugestões ou reclamações relacionadas ao sistema operacional Ubuntu<sup>®</sup>.

No processo de transformação, os documentos textuais de entrada de dados foram modelados e armazenados no banco de dados conforme estabelecido no diagrama de Entidade-Relacionamento apresentado na Figura 18. Ainda durante a etapa de transformação e carga, foi realizado o enriquecimento dos dados incluindo a informação de polaridade

implementada por Alvarez (2018), classificando as ideias como positivas, negativas ou neutras.

A fonte de dados utilizada para obter os artigos foi a base de dados da *Semantic Scholar*<sup>®</sup>, um serviço de busca para artigos de periódicos e conferências. No processo de transformação, os documentos textuais foram modelados e armazenados no banco de dados conforme estabelecido no diagrama Entidade-Relacionamento apresentado na Figura 22.

Após a etapa de carga dos artigos na base de dados foi realizado o processo de cálculo do índice de atualidade das palavras-chave dos artigos. Os índices de atualidade foram calculados e aplicados na base de ideias conforme descrito nas seções 4.1.4 e 4.1.5.

Realizado o enriquecimento da base de ideias foi executada a etapa de agrupamento de ideias. O agrupamento de ideias similares propõe a divisão do conjunto de ideias em *clusters*, utilizando um método de Análise de Agrupamento baseado em densidade. Para tanto foram gerados vetores para as ideias considerando o título e o conteúdo das ideias.

Estes agrupamentos visam a geração de grupos onde as ideias sejam homogêneas entre si, de forma que sejam semelhantes às demais ideias do seu grupo, e heterogêneas quando comparadas às ideias de outros grupos, ou seja, devem ser distintas.

A partir da geração dos grupos de ideias foram avaliados os diferentes cenários estabelecidos para a ordenação dos agrupamentos. A ordenação objetiva disponibilizar os agrupamentos de forma que, quanto mais relevante o grupo de ideias for, melhor será sua classificação.

Neste trabalho, o critério utilizado para ordenação dos grupos foi o índice de atualidade. O índice de atualidade de um grupo correspondente ao somatório do índice de atualidade das ideias que o compõe.

Mediante a análise das ordenações produzidas em diferentes cenários entende-se que o modelo proposto foi capaz de atender os objetivos estabelecidos para este trabalho e de responder à pergunta de pesquisa. A ordenação, considerando o índice de atualidade, auxilia na identificação de grupos de ideias com potencial de implementação.

Esta informação pode otimizar o processo de identificação de ideias positivas, não implementadas, porém atuais, visto que no processo de Gestão de Ideias e Gestão de Inovação, o fator tempo é uma variável primordial para o sucesso e eficiência da tomada de decisão.

A aplicação do modelo em diferentes cenários proporcionou uma melhor observação sobre o comportamento dos grupos de ideias.

Conseqüentemente, definiu-se que para o estudo realizado o limiar adequado foi de 0,05, promovendo agrupamentos mais densos e com bom resultado após a fase final de ordenação.

A partir dos resultados obtidos é possível constatar que há uma tendência de que as ideias positivas influenciem na geração de grupos altamente positivos. Quando adicionada a variável que representa o índice de atualidade é possível estabelecer a relevância de um grupo permitindo que este seja ordenado. Esta tendência se mostrou verdadeira em todos os limiares e cenários analisados.

Em aspectos gerais, a aplicação do modelo proposto se propõe a auxiliar no processo de Gestão de Ideias. Portanto, a análise de um especialista do domínio ainda é necessária. A análise de um especialista de domínio pode colaborar tanto na confirmação dos resultados obtidos nos processos de agrupamento e ordenação, quanto na melhoria e correção de eventuais pontos de falha do processo, sejam eles na determinação da polaridade das ideias, no agrupamento ou ordenação dos grupos gerados.

Por fim, os resultados obtidos com os experimentos realizados confirmam a capacidade do modelo proposto em prover informações consistentes capazes de auxiliar especialistas de domínio a realizar avaliações das ideias propostas, colaborando para o processo de tomada de decisão.

Neste contexto, o tomador de decisão pode se concentrar em grupos de ideias ao invés de ter que decidir por ideias isoladas. Isto pode maximizar o investimento uma vez que mais ideias podem ser implementadas. Tal fato, pode ainda gerar mais satisfação nos colaboradores internos ou externos à organização, visto que um número maior de ideias tende a ser contemplado.

A sessão seguinte apresenta possíveis trabalhos futuros para o tema abordado neste trabalho.

## 6.2 PERSPECTIVAS DE TRABALHOS FUTUROS

Os resultados obtidos neste trabalho possibilitaram identificar elementos de melhorias no modelo proposto.

No que tange a etapa de ordenação de grupos seria importante considerar outros elementos além do índice de atualidade, por exemplo, a densidade do agrupamento ou outros índices que indicassem a relevância do agrupamento.

Considerando o limiar que atingiu os melhores resultados torna-se necessário pré-processar os dados de maneira mais adequada para

lidar com a esparsidade do conjunto de dados utilizado. Para tal, a aplicação de decomposição matricial, tal como, a *Singular Value Decomposition* (SVD), permite estabelecer relacionamentos latentes entre as dimensões impactando na expressividade dos vetores. No presente trabalho os vetores representam ideias. A partir do aprimoramento da representação dos dados pode-se também evoluir na aplicação de outros algoritmos de agrupamento, como por exemplo, os Mapas Auto-Organizáveis (SOM) associados ao conceito de *Deep Learning*.

Vislumbra-se também a evolução do algoritmo de Mineração de Opinião para compreender as novas terminologias utilizadas em plataformas de *crowdsourcing*. Por exemplo, interpretar comentários falso-negativos que, apesar de usarem termos que tendem a gerar uma classificação negativa, buscam apoiar a ideia apresentada demonstrando descontentamento com o mesmo problema reportado.

Por fim, com o intuito de auxiliar especialistas na tarefa de análise e seleção de ideias com potencial de implementação, as áreas de visualização e explicitação do conhecimento promovem diversas oportunidades. A definição de um fluxo de extração e explicitação dos conceitos de cada agrupamento utilizando técnicas de representação visual e análise de redes pode facilitar a análise e seleção de ideias com potencial de implementação pelos especialistas.



## REFERÊNCIAS

- AHMAD, Aliyu Usman; STARKEY, Andrew. Application of feature selection methods for automated clustering analysis: a review on synthetic datasets. **Neural Computing and Applications**, p. 1-12, 2017.
- ALESSI, M.; CAMILLÒ, A.; CHETTA, V.; GIANGRECO, E.; SOUFIVAND, M.; STORELLI, D. *Applying Idea Management System (IMS) approach to design and implement a collaborative environment in public service related open Innovation processes*. **Complex Systems Informatics and Modeling Quarterly**, n. 5, p. 26-38, 2015.
- ALLAN, James; LEOSKI, Anton V.; SWAN, Russell C. Interactive cluster visualization for information retrieval. In: **submitted for conference publication**. 1997.
- AL-SULTAN, Khaled S. A tabu search approach to the clustering problem. **Pattern recognition**, v. 28, n. 9, p. 1443-1451, 1995.
- AL-SULTAN, Khaled S.; FEDJKI, Chawki A. A tabu search-based algorithm for the fuzzy clustering problem. **Pattern Recognition**, v. 30, n. 12, p. 2023-2030, 1997.
- ANACLETO, Medeiros Matheus. **Um modelo baseado em Análise de Sentimentos como suporte à Sistemas de Recomendação**. Dissertação (Mestrado) - Programa de Pós-Graduação em Tecnologias de Informação e Comunicação, Universidade Federal de Santa Catarina, Florianópolis, 2017.
- ALVAREZ, Guilherme Martins. **Análise de agrupamentos e mineração de opinião como suporte à gestão de ideias**. 2018. Dissertação de Mestrado. Universidade Federal de Santa Catarina.
- ANDROSCH, F. M.; REDL, U. **We Aim to Win – How Innovation is Managed at Voestalpine**. BHM Berg- und Hüttenmännische Monatshefte, v. 162, n. 9, p. 389-396, September 012017.
- ARRINGTON, C. Edward; JENSEN, Robert E.; TOKUTANI, Masao. Scaling of corporate multivariate performance criteria subjective composition versus the analytic hierarchy process. **Journal of Accounting and Public Policy**, v. 1, n. 2, p. 95-123, 1982.

BAEZ, Marcos; CONVERTINO, Gregorio. Designing a facilitator's cockpit for an idea management system. In: **Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work Companion**. ACM, 2012. p. 59-62.

BAEZ, Marcos; CONVERTINO, Gregorio. Innovation cockpit: a dashboard for facilitators in idea management. In: **Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work Companion**. ACM, 2012. p. 47-48.

BANFIELD, Jeffrey D.; RAFTERY, Adrian E. Model-based Gaussian and non-Gaussian clustering. **Biometrics**, p. 803-821, 1993.

BARRIOS, P. C. et al. **An Organizational Model to Understand the Creativity Workshop**. In: **Signal-Image Technology & Internet-Based Systems (SITIS), 2017 13th International Conference on**. IEEE, 2017. p. 496-502.

BARRIOS, Pedro Chávez et al. An Organizational Model to Understand the Creativity Workshop. In: **2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)**. IEEE, 2017. p. 496-502.

BASU, Sugato; BILENKO, Mikhail; MOONEY, Raymond J. Comparing and unifying search-based and similarity-based approaches to semi-supervised clustering. In: **Proceedings of the ICML-2003 workshop on the continuum from labeled to unlabeled data in machine learning and data mining**. 2003. p. 42-49.

BERRY, Michael JA; LINOFF, Gordon S. **Data mining techniques: for marketing, sales, and customer relationship management**. John Wiley & Sons, 2000.

BEZDEK, James C.; EHRlich, Robert; FULL, William. FCM: The fuzzy c-means clustering algorithm. **Computers & Geosciences**, v. 10, n. 2-3, p. 191-203, 1984.

BRABHAM, Daren C. Crowdsourcing as a model for problem solving: **An introduction and cases**. **Convergence**, v. 14, n. 1, p. 75-90, 2008.

BREM, A.; VOIGT, K. I. Innovation management in emerging technology ventures: the concept of an integrated idea management. **International Journal of Technology, Policy and Management**, Olney, v. 7, n. 3, p. 304-321, 2007.

BREM, Alexander; VOIGT, Kai-Ingo. Integration of market pull and technology push in the corporate front end and innovation management—Insights from the German software industry. **Technovation**, v. 29, n. 5, p. 351-367, 2009.

CAO, Feng et al. Density-based clustering over an evolving data stream with noise. In: **Proceedings of the 2006 SIAM international conference on data mining**. Society for Industrial and Applied Mathematics, 2006. p. 328-339.

CARAYANNIS, Elias G.; MEISSNER, Dirk; EDELKINA, Anastasia. Targeted innovation policy and practice intelligence (TIP2E): concepts and implications for theory, policy and practice. **The Journal of Technology Transfer**, v. 42, n. 3, p. 460-484, 2017.

CHO, Y. S., ET AL. **A study of recommending service based on customer inclination using AHP**. IEEE Engineering Management Review 2016.

COUGHLAN, Tim; JOHNSON, Peter. Idea management in creative lives. In: **CHI'08 extended abstracts on Human factors in computing systems**. ACM, 2008. p. 3081-3086.

COOPER, Robert G. Stage-gate systems: A new tool for managing new products, **Business Horizons**, vol. 33, n. 3, p. 44-54, 1990.

DASH, Manoranjan; LIU, Huan; XU, Xiaowei. '1+ 1 > 2': merging distance and density based clustering. In: **Proceedings Seventh International Conference on Database Systems for Advanced Applications. DASFAA 2001**. IEEE, 2001. p. 32-39.

DE LIDDO, Anna et al. Collective intelligence as community discourse and action. In: **Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work Companion**. ACM, 2012. p. 5-6.

DOAN, Anhai; RAMAKRISHNAN, Raghu; HALEVY, Alon Y. Crowdsourcing systems on the world-wide web. **Communications of the ACM**, v. 54, n. 4, p. 86-96, 2011.

DUIN, Heiko et al. **An idea model for distributed Idea Management**. In: Proceedings of the 16th International Conference on Concurrent Enterprising. 2010. p. 1-8.

ELIASHBERG, Jehoshua; MANRAI, Ajay K. Optimal positioning of new product-concepts: Some analytical implications and empirical results. **European Journal of Operational Research**, v. 63, n. 3, p. 376-397, 1992.

ERDOĞAN, Şenol Zafer; TIMOR, Mehpare. A data mining application in a student database. **Journal of aeronautics and space technologies**, v. 2, n. 2, p. 53-57, 2005.

ESTER, Martin et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In: **Kdd**. 1996. p. 226-231.

EVERITT, B. S., LANDAU, S., LEESE, M., STAHL, D. (2011). **Cluster Analysis (5th ed.)**. Chichester, West Sussex, UK: Wiley. <https://doi.org/10.1002/9780470977811>

FAHAD, A.; ALSHATRI, N.; TARI, Z.; ALAMRI, A.; KHALIL, I.; ZOMAYA, A. Y.; FOUFOU, S.; BOURAS, A. A survey of clustering algorithms for big data: Taxonomy and empirical analysis. **IEEE transactions on emerging topics in computing**, v. 2, n. 3, p. 267-279, 2014.

FIGUEIREDO FILHO, Dalson Britto; SILVA JÚNIOR, José Alexandre da. Desvendando os Mistérios do Coeficiente de Correlação de Pearson ( $r$ ). 2009.

FRALEY, Chris; RAFTERY, Adrian E. How many clusters? Which clustering method? Answers via model-based cluster analysis. **The computer journal**, v. 41, n. 8, p. 578-588, 1998.

FRALEY, Chris; RAFTERY, Adrian E. Model-based clustering, discriminant analysis, and density estimation. **Journal of the American statistical Association**, v. 97, n. 458, p. 611-631, 2002.

GAN, Guojun; MA, Chaoqun; WU, Jianhong. Data clustering: theory, algorithms, and applications. **Siam**, 2007.

GANTI, Venkatesh; GEHRKE, Johannes; RAMAKRISHNAN, Raghu. CACTUS-clustering categorical data using summaries. In: **KDD**. 1999. p. 73-83.

GATH, Isak; GEVA, Amir B.. Unsupervised optimal fuzzy clustering. **IEEE Transactions on Pattern Analysis & Machine Intelligence**, n. 7, p. 773-780, 1989.

GISH, Liv; HANSEN, Claus Thorp. A socio-technical analysis of work with ideas in NPD: an industrial case study. **Research in Engineering Design**, v. 24, n. 4, p. 411-427, 2013.

GOLDSCHMIDT, Ronaldo; BEZERRA, E.; PASSOS, E. **Data mining: conceitos, técnicas, algoritmos, orientações e aplicações**. Elsevier Brasil, 2015.

GONÇALVES, Alexandre Leopoldo. **Um modelo de descoberta de conhecimento baseado na correlação de elementos textuais e expansão vetorial aplicado à engenharia e gestão do conhecimento**. Florianópolis, SC, 2006. 196 f. Tese (Doutorado) - Universidade Federal de Santa Catarina, Centro Tecnológico. Programa de Pós-Graduação em Engenharia de Produção.

GOODMAN, Joseph K.; PAOLACCI, Gabriele. **Crowdsourcing consumer research**. **Journal of Consumer Research**, v. 44, n. 1, p. 196-210, 2017.

GREEN, Paul E.; KRIEGER, Abba M. Conjoint analysis with product-positioning applications. **Handbooks in operations research and management science**, v. 5, p. 467-515, 1993.

GUSTAFSON, Donald E.; KESSEL, William C. Fuzzy clustering with a fuzzy covariance matrix. In: **1978 IEEE conference on decision and control including the 17th symposium on adaptive processes**. IEEE, 1979. p. 761-766.

HAN, Eui-Hong Sam; KARYPIS, George. Centroid-based document classification: Analysis and experimental results. In: **European conference on principles of data mining and knowledge discovery**. Springer, Berlin, Heidelberg, 2000. p. 424-431.

HEVNER, Alan; CHATTERJEE, Samir. Design science research in information systems. In: **Design research in information systems**. Springer, Boston, MA, 2010. p. 9-22.

HINNEBURG, Alexander et al. An efficient approach to clustering in large multimedia databases with noise. In: **KDD**. 1998. p. 58-65.

HUANG, Anna. Similarity measures for text document clustering. In: **Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008), Christchurch, New Zealand**. 2008. p. 9-56.

HUANG, Gao et al. Maximin Separation Probability Clustering. In: **AAAI**. 2015. p. 2680-2686.

INIE, Nanna; DOW, Steven; DALSGAARD, Peter. Supporting reflective use of design idea archives using email. In: **Nordic Conference on Human-Computer Interaction**. Association for Computing Machinery (ACM), 2018.

JAIN, Anil K. Data clustering: 50 years beyond K-means. **Pattern recognition letters**, v. 31, n. 8, p. 651-666, 2010.

JAIN, Anil K. Data clustering: 50 years beyond K-means. **Pattern recognition letters**, v. 31, n. 8, p. 651-666, 2010.

JAIN, Anil K.; MURTY, M. Narasimha; FLYNN, Patrick J. Data clustering: a review. **ACM computing surveys (CSUR)**, v. 31, n. 3, p. 264-323, 1999.

JENKINS, Henry; DEUZE, Mark. **Convergence culture**. 2008.

JIANG, Bin et al. Clustering uncertain data based on probability distribution similarity. **IEEE Transactions on Knowledge and Data Engineering**, v. 25, n. 4, p. 751-763, 2013.

JIMÉNEZ-NARVAEZ, L. M.; GARDONI, M. **Harnessing idea management in the process of technology transfer at Canadian Space Agency**. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, v. 9, n. 3, p. 247-252, August 012015.

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.

KARYPIS, George; HAN, Eui-Hong Sam; KUMAR, Vipin. Chameleon: Hierarchical clustering using dynamic modeling. **Computer**, n. 8, p. 68-75, 1999.

KEIM, Daniel A.; KRIEGEL, H.-P. Visualization techniques for mining large databases: A comparison. **IEEE Transactions on knowledge and data engineering**, v. 8, n. 6, p. 923-938, 1996.

KOEN, P.; AJAMIAN, G.; BURKART, R.; CLAMEN, A.; DAVIDSON, J.; D'AMORE, R.; ELKINS, C.; HERALD, K.; INCORVIA, M.; JOHNSON, A.; KAROL, R.; SEIBERT, R.; SLAVEJKOV, A.; WAGNER, K. Providing clarity and a common language to the “fuzzy front end”. **Research-Technology Management**, v. 44, n. 2, p. 46-55, 2001.

KRIEGEL, Hans-Peter et al. Density-based clustering. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, v. 1, n. 3, p. 231-240, 2011.

LAMBERTH-COCCA, Sabrina; MEIREN, Thomas. Towards a reference model for agile new service development using the example of e-mobility service systems. **Procedia CIRP**, v. 64, p. 259-264, 2017.

LIAO, Wei-keng; LIU, Ying; CHOUDHARY, Alok. A grid-based clustering algorithm using adaptive mesh refinement. In: **7th Workshop on Mining Scientific and Engineering Datasets of SIAM International Conference on Data Mining**. 2004. p. 61-69.

MAGNUSSON, Peter R.; NETZ, Johan; WÄSTLUND, Erik. **Exploring holistic intuitive idea screening in the light of formal criteria**. *Technovation*, v. 34, n. 5-6, p. 315-326, 2014.

MAHAJAN, Vijay; WIND, Yoram. New product forecasting models: Directions for research and implementation. **International Journal of Forecasting**, v. 4, n. 3, p. 341-358, 1988.

MAO, Ke et al. A survey of the use of crowdsourcing in software engineering. **Journal of Systems and Software**, v. 126, p. 57-84, 2017.

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.

MCCORMACK, B.; FALLON, E. F.; CORMICAN, K. **An Analysis of Open Innovation Practices in the Medical Technology Sector in Ireland**. *Procedia Manufacturing*, v. 3, p. 503-509, 2015.

MERKL, Dieter; RAUBER, Andreas. Alternative ways for cluster visualization in self-organizing maps. In: **Workshop on Self-Organizing Maps**. 1997. p. 106-111.

MILLIGAN, Glenn W.; COOPER, Martha C. A study of standardization of variables in cluster analysis. **Journal of classification**, v. 5, n. 2, p. 181-204, 1988.

OCDE, **Manual de Oslo**: Diretrizes para coleta e interpretação de dados sobre inovação. 3. Ed. Paris: OCDE, 2005.

ORTIZ, A. et al. MR brain image segmentation by growing hierarchical SOM and probability clustering. **Electronics Letters**, v. 47, n. 10, p. 585-586, 2011.

PAGE, Lawrence et al. **The PageRank citation ranking: Bringing order to the web**. Stanford InfoLab, 1999.

PAMPALK, Elias; RAUBER, Andreas; MERKL, Dieter. Using smoothed data histograms for cluster visualization in self-organizing maps. In: **International conference on artificial neural networks**. Springer, Berlin, Heidelberg, 2002. p. 871-876.

PARK, Nam Hun; LEE, Won Suk. Statistical grid-based clustering over data streams. **ACM Sigmod Record**, v. 33, n. 1, p. 32-37, 2004.



PAUKKERI, Mari-Sanna; KOTRO, Tanja. **Framework for analyzing and clustering short message database of ideas**. na, 2009.

PEFFERS, Ken et al. A design science research methodology for information systems research. **Journal of management information systems**, v. 24, n. 3, p. 45-77, 2007.

PINTAS, Julliano Trindade; CORREIA, Luís; GARCIA, Ana Cristina Bicharra. Crowd-based Feature Selection for Document Retrieval in Highly Demanding Decision-making Scenarios. **Procedia Computer Science**, v. 112, p. 822-832, 2017.

POORKAVOOS, Meysam et al. 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.

PPGTIC (a). Sobre o PPGTIC. 2017. Disponível em: < <http://ppgtic.ufsc.br/sobre-o-ppgtic> >. Acesso em: 17 mai. 2017.

PPGTIC (b). Linhas de Pesquisa. 2017. Disponível em: < <http://ppgtic.ufsc.br/linhas-de-pesquisa/> >. Acesso em: 17 mai. 2017.

QIN, S. et al. **Exploring barriers and opportunities in adopting crowdsourcing based new product development in manufacturing SMEs**. Chinese Journal of Mechanical Engineering, v. 29, n. 6, p. 1052-1066, November 01 2016.

RADEV, Dragomir R. et al. Centroid-based summarization of multiple documents. **Information Processing & Management**, v. 40, n. 6, p. 919-938, 2004.

RAMOS, Juan et al. Using tf-idf to determine word relevance in document queries. In: **Proceedings of the first instructional conference on machine learning**. 2003. p. 133-142

REAL, Raimundo; VARGAS, Juan M. The probabilistic basis of Jaccard's index of similarity. **Systematic biology**, v. 45, n. 3, p. 380-385, 1996.

RIBEIRO, Alessandro Costa. **Modelo de reconhecimento de padrões em ideias usando técnicas de descoberta de conhecimento em textos**. 2018. Dissertação de Mestrado. Universidade Federal de Santa Catarina.

SADRIEV, A. R.; PRATCHENKO, O. V. **Idea management in the system of innovative management**. *Mediterranean Journal of Social Sciences*, v. 5, n. 12, p. 155-158, 2014.

SALDIVAR, J. et al. **Idea Management Communities in the Wild: An Exploratory Study of 166 Online Communities**. 2016 International Conference on Collaboration Technologies and Systems (CTS), 2016. Oct. 31 2016-Nov. 4 2016. p.81-89.

SANDER, Jörg et al. Density-based clustering in spatial databases: The algorithm gbscan and its applications. **Data mining and knowledge discovery**, v. 2, n. 2, p. 169-194, 1998.

SÉRGIO, Marina Carradore. **Modelo de avaliação de potenciais ideias alinhadas ao contexto organizacional**. Qualificação (Doutorado) - Programa de Pós-graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2018.

SÉRGIO, Marina Carradore. **Um Modelo Baseado em Ontologias 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.

SERGIO, Marina Carradore; DE SOUZA, Joao Artur; GONCALVES, Alexandre Leopoldo. Idea Identification Model to Support Decision Making. **IEEE Latin America Transactions**, v. 15, n. 5, p. 968-973, 2017.

VAISHNAVI, V.; KUECHLER, W. Design Research in Information Systems. **Integrated Series in Information Systems**, v. 22, n. 2, p. 1–16, 2004

SILVA, Marcia Vieira; ZILBER, Moises Ari. Benefícios percebidos pela adoção do processo de inovação aberta. **Revista de Administração**, v. 11, n. 3, p. 1-24, 2013.

VAN DE VRANDE, Vareska et al. Open innovation in SMEs: Trends, motives and management challenges. **Technovation**, v. 29, n. 6-7, p. 423-437, 2009.

VENNA, Jarkko; KASKI, Samuel. Local multidimensional scaling. **Neural Networks**, v. 19, n. 6-7, p. 889-899, 2006.

VESANTO, Juha. SOM-based data visualization methods. **Intelligent data analysis**, v. 3, n. 2, p. 111-126, 1999.

VIEDERYTE, Rasa. How corporate decisions force innovations: factors and choices to act. **Procedia Economics and Finance**, v. 39, p. 357-364, 2016.

WALTER, Thomas P.; BACK, Andrea. *A text mining approach to evaluate submissions to crowdsourcing contests*. In: System Sciences (HICSS), 2013 46th Hawaii International Conference on. IEEE, 2013. p. 3109-3118.

WANG, Li; WANG, Zheng-Ou. CUBN: A clustering algorithm based on density and distance. In: **Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No. 03EX693)**. IEEE, 2003. p. 108-112.

WEBER, John A. Using purchase influence niching for better focus in industrial marketing plans: observations and examples. **Industrial Marketing Management**, v. 23, n. 5, p. 419-438, 1994.

WEISS, M. **Patterns for idea management in innovation communities**. 22nd European Conference on Pattern Languages of Programs, EuroPLoP 2017, 2017. Association for Computing Machinery.

WESTERSKI, Adam; DALAMAGAS, Theodore; IGLESIAS, Carlos A. Classifying and comparing community innovation in Idea Management Systems. **Decision Support Systems**, v. 54, n. 3, p. 1316-1326, 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, p. 289-296, 2012.**

WHITNEY, Dwight E. Assemble a technology development toolkit. **Research-Technology Management**, v. 50, n. 5, p. 52-58, 2007.

XIE, Xuanli Lisa ; BENI, Gerardo. A validity measure for fuzzy clustering. **IEEE Transactions on Pattern Analysis & Machine Intelligence**, n. 8, p. 841-847, 1991.

XU, Xiaowei et al. A distribution-based clustering algorithm for mining in large spatial databases. In: **Proceedings 14th International Conference on Data Engineering**. IEEE, 1998. p. 324-331.

YEUNG, Ka Yee et al. Model-based clustering and data transformations for gene expression data. **Bioinformatics**, v. 17, n. 10, p. 977-987, 2001.

YUNOH, M. F. M.; ABDULLAH, S.; SAAD, M. H. M.; NOPIAH, Z. M.; NUAWI, M. Z. K-means clustering analysis and artificial neural network classification of fatigue strain signals. **Journal of the Brazilian Society of Mechanical Sciences and Engineering**, p. 1-8, 2016.

ZADEH, Lotfi A. Fuzzy sets. **Information and control**, v. 8, n. 3, p. 338-353, 1965.