

UNIVERSIDADE FEDERAL DE PELOTAS
Centro de Desenvolvimento Tecnológico
Programa de Pós-Graduação em Computação



Tese

Metaciência da Computação: Um Estudo Sobre Taxonomia de Pesquisa

Alessander Osorio

Pelotas, 2024

Alessander Osorio

Metaciência da Computação: Um Estudo Sobre Taxonomia de Pesquisa

Tese apresentada ao Programa de Pós-Graduação em Computação da Universidade Federal de Pelotas, como requisito parcial à obtenção do título de Doutor em Ciência da Computação

Orientador: Prof. Dr. Gerson Geraldo H. Cavalheiro
Coorientador: Prof. Dr. Paulo Roberto Ferreira Jr.

Pelotas, 2024

Universidade Federal de Pelotas / Sistema de Bibliotecas
Catalogação da Publicação

O81m Osorio, Alessander

Metaciência da computação [recurso eletrônico] : um estudo sobre taxonomia de pesquisa / Alessander Osorio ; Gerson Geraldo Homrich Cavalheiro, orientador ; Pauro Roberto Ferreira Jr., coorientador. — Pelotas, 2024.

117 f. : il.

Tese (Doutorado) — Programa de Pós-Graduação em Computação, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, 2024.

1. Taxonomia de pesquisa. 2. Avaliação de pesquisa. 3. Metodologia de pesquisa. I. Cavalheiro, Gerson Geraldo Homrich, orient. II. Jr., Pauro Roberto Ferreira, coorient. III. Título.

CDD 005

Elaborada por Simone Godinho Maisonave CRB: 10/1733

Alessander Osorio

Metaciência da Computação: Um Estudo Sobre Taxonomia de Pesquisa

Data da Defesa: 9 de abril de 2024

Banca Examinadora:

Prof. Dr. Alfredo Goldman

Profa. Dra. Tatiana Tavares

Prof. Dr. Rafael Fonseca de Castro

Prof. Dr. Paulo Roberto Ferreira Jr

Prof. Dr. Gerson Geraldo H. Cavalheiro (Presidente)

AGRADECIMENTOS

Agradeço DEUS, pelos desafios na minha trajetória, e as pessoas certas para me ajudar a transpô-los. A ELE, e a você que esteve comigo até aqui o meu muito obrigado!

RESUMO

OSORIO, Alessander. **Metaciência da Computação: Um Estudo Sobre Taxonomia de Pesquisa.** 2024. 118 f. Tese (Doutorado em Ciência da Computação) – Programa de Pós-Graduação em Computação, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, Pelotas, 2024.

O problema na apresentação e comprovação de resultados de publicações científicas vem sendo estudado há muito tempo e a literatura cronologicamente vem apontando estas evidências. Tal problema ainda perdura nos dias atuais. Evidências recentes comprovam tal fato, e isto mostra que algo precisa ser feito para qualificar a pesquisa científica na área da computação. Um dos motivos destacados pela literatura é a falta de conhecimento sobre a pesquisa em si. As possibilidades, nuances e diferentes matizes que permitirão seu enriquecimento. Fugindo à tradicional pesquisa experimental sem a efetiva comprovação de resultados. O presente trabalho apresenta um meta estudo da Ciência em Computação, a partir do qual é apresentada uma Taxonomia de Pesquisa que se apresenta como um guia à pesquisa e ao ensino. Os resultados obtidos são os 11 passos que compreendem as etapas da pesquisa científica, especializadas no contexto da Computação, e seu ciclo lógico de refinamento sucessivo. Os pontos de inserção da pesquisa em Computação dentro da Ciência como um todo, sendo estes Meta, Meio e Fim. A taxonomia sistematiza os conceitos de 96 tipos de classificações identificadas na literatura, agrupadas em 10 categorias. O processo de validação desta taxonomia identificou sua aplicabilidade e completude. Neste processo, primeiro foi elaborada uma ontologia com base nos descritores da taxonomia, sendo realizada uma pesquisa textual em 2.508 publicações científicas de 68 periódicos publicados entre 2018 e Ago/2022. As 124.601 ocorrências resultantes desta pesquisa tiveram seu contexto analisado e classificado pelo algoritmo semi-supervisionado k-NN, buscando identificar se o contexto de uso do descritor se encontra em conformidade com a taxonomia proposta. Os resultados apontam que este fato aconteceu em 73,37% dos contextos de uso dos descritores da taxonomia, mostrando assim sua aplicabilidade. Para avaliar sua completude, foram analisadas as áreas dos periódicos dos quais as publicações foram extraídas. Tal informação segue a classificação da SCOPUS, e para esta análise foram consideradas também as classificações da ACM e GAs da SBC, para as quais foi traçado um paralelo de equivalência. Dentro deste exercício é possível verificar que todas as sub-áreas da Computação segundo as 3 classificações são contempladas nos resultados, mostrando assim sua completude em relação à Computação.

Palavras-Chave: taxonomia de pesquisa; avaliação de pesquisa; metodologia de pesquisa

ABSTRACT

OSORIO, Alessander. **Computer Metascience: A Study on Research Taxonomy.** 2024. 118 f. Tese (Doutorado em Ciência da Computação) – Programa de Pós-Graduação em Computação, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, Pelotas, 2024.

The problem in presenting or validating results of scientific publications has been studied for a long time, and the literature chronologically has been pointing out these pieces of evidence. Such an issue still persists in the present days. Recent evidence confirms this fact, indicating that something needs to be done to enhance scientific research in the field of computing. One of the reasons highlighted by the literature is the lack of knowledge about the research itself. The possibilities, nuances, and different shades that will allow its enrichment. Departing from traditional experimental research without effective validation of results. This paper presents a meta-study of Computer Science, from which a Research Taxonomy is proposed to guide research and teaching. The obtained results are the 11 steps that comprise the stages of scientific research, specialized in the context of computer science, and its logical cycle of successive refinement. The points of insertion of computer research within science as a whole, these being Meta, Middle, and End. The proposed taxonomy systematizes the concepts of 96 types of classifications identified in the literature, grouped into 10 categories. The validation process of this taxonomy identified its applicability and completeness. In this process, an ontology was first developed based on the descriptors of the taxonomy, conducting a textual search in 2,508 scientific publications from 68 journals published between 2018 and Aug/2022. The 124,601 occurrences resulting from this research had their context analyzed and classified by the semi-supervised k-NN algorithm, seeking to identify whether the context of the descriptor's usage conforms to the proposed taxonomy. The results indicate that this happened in 73.37% of the contexts of descriptor usage, demonstrating its applicability. To assess its completeness, the areas of the journals from which the publications were extracted were analyzed. This information follows the SCOPUS classification, and for this analysis, the ACM and SBC's GAS classifications were also considered, for which an equivalence parallel was drawn. Within this exercise, it is possible to verify that all sub-areas of computing according to the three classifications are covered in the results, demonstrating its completeness concerning Computing.

Keywords: research taxonomy; research evaluation; research methodology

LISTA DE FIGURAS

1	Divisão das Ciências	21
2	Etapas da Revisão Sistemática de literatura.	29
3	Método Experimental	37
4	Intersecção dos Métodos (TEDRE, 2007)	39
5	Eixos de estudo da Ciência em Computação	48
6	Taxonomia de Pesquisa em Ciência em Computação	55
7	Estratégia de pesquisa	58
8	Marco Metodológico Parte 1	60
9	Marco Metodológico Parte 2	61
10	Médias de Ocorrências por Categoria	66
11	Etapas da Pesquisa Científica	77

LISTA DE TABELAS

1	Resultado Pesquisa nas Bases Indexadas	32
2	Divisão da Ciência	33
3	Divisão de Áreas da Computação	35
4	Divisões e categorias da pesquisa científica em Computação segundo a RSL	45
5	Categorias da pesquisa científica em Computação segundo a RSL	46
6	Divisões da Pesquisa em Computação	53
7	Ontologia de pesquisa	64
8	Aceite por categorias de descritores da ontologia	66
9	Descritores sem ocorrência	67
10	Aceite por descritores da ontologia	68
11	Contagem de artigos por área e associação de áreas segundo SCOPUS, ACM e SBC	73
12	Distribuição dos valores da Tabela de Aprendizagem do k-NN.	74

LISTA DE ABREVIATURAS E SIGLAS

RSL	Revisão Sistemática de Literatura
SIG	Special Interest Groups
IDR	Interdisciplinary Research
ACM	Association for Computing Machinery
IPIN	International Conference on Indoor Positioning and Indoor Navigation
CCS	Computing Classification Scheme
ERC	European Research Council
GA	Grande Área
CE	Comissões Especiais
SBC	Sociedade Brasileira de Computação
PDF	Portable Document Format
API	Application Programmin Interface
IBL	Instance Based Learning
JSON	JavaScript Object Notation
CSV	Comma Separated Values
ASJC	All Science Journal Classification Codes
ACM-CCS	ACM Computing Classification System

SUMÁRIO

1 INTRODUÇÃO	13
1.1 Motivação	13
1.2 Definição do Problema	15
1.3 Questão de Pesquisa	16
1.4 Objetivos	17
1.5 Resultados Alcançados	17
1.6 Contribuições	17
1.7 Organização do Trabalho	18
2 REVISÃO CONCEITUAL	19
2.1 O que é Ciência	19
2.2 Divisões da Ciência	20
2.3 O Método Científico	22
2.4 Classificação	23
2.5 Teórica e Experimental	25
2.6 O Resultado Científico	25
2.7 Multidisciplinaridade da Ciência	26
2.8 Ciência e Tecnologia	26
2.9 Considerações Finais do Capítulo	27
3 REVISÃO SISTEMÁTICA DE LITERATURA	28
3.1 Metodologia	28
3.1.1 Questões de pesquisa	29
3.1.2 Chave de pesquisa	29
3.1.3 Bases Indexadas	30
3.1.4 Critérios de seleção	30
3.2 Resultados	31
3.3 Desenvolvimento / Resumo Referências	31
3.3.1 Conceitos e divisões da Ciência	32
3.3.2 Etapas da pesquisa científica	34
3.3.3 Divisões das áreas de conhecimento da Computação	34
3.3.4 A pesquisa científica em Computação segundo diversos olhares	35
3.4 Discussão	43
3.4.1 Respondendo às Questões de Pesquisa	44
3.5 Ameaças à Validade	50
3.6 Considerações Finais do Capítulo	50

4 TAXONOMIA DE PESQUISA EM COMPUTAÇÃO	51
4.1 Validação	56
4.2 Método de Classificação	56
4.3 Método	57
4.4 Estratégia de Pesquisa	58
4.5 Processo	59
4.6 Resultados	63
4.6.1 A ontologia de pesquisa	63
4.6.2 Considerações iniciais dos resultados	63
4.6.3 Resultados por categoria da ontologia	65
4.6.4 Resultados por descritores da ontologia	66
4.6.5 Análise da cobertura das áreas do conhecimento	71
4.7 A tabela de aprendizagem do k-NN	72
4.8 Ameaças à Validade	72
4.9 Considerações Finais do Capítulo	74
5 COMPLEMENTOS À TAXONOMIA	75
5.1 Etapas da Pesquisa Científica	75
5.2 O Protocolo RODA	78
5.3 Protótipo de Extensão para Navegador	79
6 DISCUSSÃO SOBRE OS RESULTADOS	80
7 CONCLUSÃO	83
REFERÊNCIAS	85
APÊNDICE A RELATÓRIO DA ANÁLISE DE DADOS	90

1 INTRODUÇÃO

Questões relacionadas à apresentação e comprovação de resultados de experimentos em relatos científicos, como artigos em periódicos ou conferências, têm sido objeto de estudo ao longo do tempo ((TICHY et al., 1995), (PRECHELT, 1994)). As evidências apontadas pela literatura mais recente na área da Computação (TEDRE; MOISSEINEN, 2014; ADLER et al., 2015; MUNAFÒ et al., 2017; OSORIO; DIAS; CA-VALHEIRO, 2018) indicam que existem lacunas ainda a serem cobertas para qualificação dos relatos apresentados. Analisando esses estudos, identifica-se que um dos motivos apontados para justificar a falta de clareza na apresentação dos resultados é a falta de conhecimento por parte dos pesquisadores sobre o próprio processo de pesquisa. Dessa forma, de maneira geral, os relatos científicos apresentam deficiências em abordar todas as possibilidades, nuances e matizes diversos que possibilitariam seu fortalecimento e enriquecimento como material de divulgação científica e disseminação de conhecimento e, em consequência, fortalecendo e enriquecendo a contribuição dada à Computação enquanto grande área do conhecimento.

O objetivo do presente trabalho é conduzir um estudo da Metaciência em Computação e apresentar uma Taxonomia de Pesquisa à área. Tal taxonomia é apresentada para servir de guia no processo de caracterização das pesquisas realizadas e principalmente servir como ferramenta de apoio à formação de pesquisadores. Sua contribuição é estender as discussões sobre metaciência na Computação, fomentando o amadurecimento da área na produção de relatos científicos. Com isso o resultado esperado é que os pesquisadores produzam artigos científicos com técnicas diversas às tradicionais e que possam ser melhor absorvidos pelos demais pesquisadores.

1.1 Motivação

A inadequação na apresentação dos resultados em trabalhos científicos na área da Computação encontra-se relatada em alguns estudos encontrados na literatura. (PRECHELT, 1994) avaliou 190 artigos publicados em quatro revistas de renome na área de redes neurais nos anos de 1993 e 1994, evidenciando que 1/3 dos artigos não

apresentavam comparação quantitativa dos resultados relatados com os obtidos com técnicas previamente conhecidas. (TICHY et al., 1995) analisou 400 artigos publicados pela *Association for Computing Machinery* (ACM) para determinar se cientistas da Computação apoiam seus resultados com avaliação experimental. O estudo descobriu que 40% dos artigos não possuíam qualquer tipo de avaliação. (WAINER et al., 2009) replicou a pesquisa de (TICHY et al., 1995) analisando 147 artigos publicados na ACM no ano de 2005 selecionados aleatoriamente, concluindo que 33% dos artigos encontram-se na mesma situação.

O trabalho de (ADLER et al., 2015) investigou 183 artigos do *International Conference on Indoor Positioning and Indoor Navigation* (IPIN) e concluiu que, embora em muitas das publicações houvesse alguma preocupação na avaliação dos resultados, a qualidade da descrição dos métodos de análise era pobre. Apenas 35% relatam de maneira clara não só a metodologia do experimento em si, mas o que efetivamente os resultados querem estatisticamente dizer. Em (OSORIO; DIAS; CAVALHEIRO, 2018) foi realizada uma meta-avaliação de 12 anos de uma conferência de escopo nacional, e constatou que apenas 3% dos trabalhos nela publicados utilizavam algum teste estatístico na validação e comprovação dos resultados.

Observa-se que na bibliografia citada há evidências desde 1993 de que há problemas nas publicações científicas em Computação. Estes problemas são decorrentes de uma série de fatores, como falta da correta validação dos resultados, uso de comprovação estatística de resultados e de lacunas na caracterização do processo de avaliação experimental ou da metodologia empregada ou ainda de problemas associados a má definição da metodologia utilizada no estudo. Os problemas se sucedem em diferentes subáreas da Computação e perduram ao longo do tempo, como é identificado pelos estudos independentes relatados nos parágrafos anteriores. Uma das causas destes problemas é o pouco conhecimento dos pesquisadores da área da ciência da Computação com os procedimentos metodológicos para a correta condução da pesquisa desde seu início até a comprovação de resultados (TEDRE; MOISSEIEN, 2014). Poucos estudos de fato utilizam a metodologia adequada e raras são as pesquisas em Computação que formulam e validam suas hipóteses de uma maneira iterativa dentro de um ciclo de refinamento. A maioria das proposições são melhoramentos de soluções já propostas (KOZIOLEK, 2005).

Ainda que apontem para tais problemas, a bibliografia não apresenta uma alternativa prática com diretrizes objetivas nas quais pesquisadores possam se apoiar para projetar de maneira prévia e metodologicamente correta o desenvolvimento de suas pesquisas. A adoção e utilização de protocolos e taxonomias de pesquisa auxiliam na: melhora da qualidade dos resultados das pesquisas científicas (MALTA et al., 2010), nas questões de reproduzibilidade (MUNAFÒ et al., 2017), na diminuição da subjetividade conceitual (CUNHA; ARAÚJO, 2019) e na indexação e recuperação da

informação (SOUZA; ARAÚJO JÚNIOR, 2013).

Nessa linha de raciocínio, a adoção de uma conceituação unificada sobre a pesquisa científica em Computação, juntamente com guias de pesquisa aplicados à área, se apresenta como sendo uma ferramenta de grande ajuda na melhoria do projeto, condução e relato e do resultado das pesquisas realizadas e, consequentemente, de sua qualidade. Sob este aspecto é necessário que o pesquisador conheça bem a teoria da pesquisa científica, suas divisões, conceitos e a inter-operabilidade destes. Caso contrário, o resultado tende a ser imperfeito sob a ótica metodológica (TEDRE, 2007).

Para avançar nesse propósito, é crucial que, como educadores, saibamos como formar futuros pesquisadores. Isso envolve não apenas concentrar a formação na parte técnica, mas também nos aspectos da formação metodológica de novos indivíduos. O ensino deve proporcionar a estes pesquisadores em formação a capacidade de conduzir pesquisas de maneira adequada, utilizando os conceitos corretos, e ampliar o horizonte das possibilidades de pesquisa que se abrem para a Ciência em Computação.

Todavia, há uma dificuldade de encontrar uma conceituação unificada para a pesquisa científica em Computação dada sua multi-disciplinaridade (TEDRE, 2011a), tanto no sentido de incorporar áreas de conhecimento externas quanto no sentido de ser incorporada por outras áreas. Isso destaca a necessidade premente de obter um entendimento mais aprofundado da pesquisa realizada em Computação, visando enriquecer o diálogo com outras áreas do conhecimento. Como área recente em relação a demais Ciências Exatas e da Terra, a área da Computação precisa ser epistemologicamente melhor estudada (GHEZZI, 2020; DODIG-CRNKOVIC, 2023), assim como, como participante interdisciplinar em outras áreas (GHEZZI, 2020; HASSANI, 2017), também precisa ter sua contribuição estudada.

1.2 Definição do Problema

A pesquisa científica em Computação deve ser melhor compreendida. Tal compreensão além de ser utilizada para avaliar as técnicas desenvolvidas, irá auxiliar no entendimento de requisitos, práticas de trabalho e nas demonstrações sobre tais técnicas (WAINER et al., 2007).

Em (TEDRE, 2011a), há uma pequena discussão sobre o que efetivamente é ciência em Computação. Ainda que a natureza experimental a faça propícia à ciência, algumas atividades mesmo levando a resultados não podem ser consideradas como ciência. Colocar os componentes de pesquisa juntos para ver o que acontece sem uma metodologia ou sistematização não pode ser considerado ciência. Em uma interpretação moderna, o conhecimento científico gerado pela pesquisa científica não é

somente aquilo que se descobre sobre os fatos e as coisas. Mas aquilo que empodera as pessoas e as permite progredir enquanto sociedade (GHEZZI, 2020), sendo esse a maior contribuição esperada da pesquisa científica.

A generalização de resultados para os objetos de pesquisa é o objetivo principal da pesquisa científica. No entanto, o ponto de vista do pesquisador é significativamente diferente de acordo com a sua área ou subárea de atuação (HASSANI, 2017). Por esse motivo é extremamente importante estudar sob um olhar mais atento a pesquisa científica e suas nuances em Computação.

Melhorar a qualidade da apresentação dos resultados da pesquisa científica aumentará a credibilidade das publicações científicas em Computação. A adoção de medidas que otimizam os principais elementos do processo científico é fortemente recomendada (MUNAFÒ et al., 2017). Há evidências que apoiam a eficácia dessas medidas, mas sua ampla adoção na comunidade exige trabalho colaborativo da comunidade científica e a Universidade deve assumir um papel de protagonismo no processo de descoberta, idealização e disseminação de tais medidas.

Assim, é necessário realizar um estudo epistemológico sobre a taxonomia de pesquisa em Computação. O objetivo é compreender as diversas ideias existentes e unificar a conceituação, visando aprimorar a qualidade dos resultados. Esse aprimoramento ocorrerá por meio da otimização nas práticas de concepção, elaboração, obtenção e demonstração dos resultados. Essa melhoria abrange os processos de formação de novos pesquisadores e contribui para o desenvolvimento da ciência em Computação em prol da Sociedade.

1.3 Questão de Pesquisa

Em função do exposto na seção anterior, as questões de pesquisa aqui consideram a ciência em Computação como o próprio objeto de estudo. Este estudo da ciência, podendo ser identificado como um meta-estudo uma vez que refere-se a um estudo que analisa e sintetiza os resultados de múltiplos estudos individuais realizados no contexto da Computação, é realizado sob uma ótica epistemológica. De maneira abrangente, tem-se a seguinte formulação:

Como a pesquisa em Computação pode ser dividida, classificada, caracterizada, descrita ou conceituada de maneira a refletir todas as etapas do processo de produção do conhecimento, bem como suas diferentes inserções na ciência?

Da questão de pesquisa, dentro do contexto do presente trabalho, desdobrar-se nos objetivos geral e específico. Pretende-se manter o foco na pesquisa e no ensino durante a elaboração do presente trabalho e ao procurar respostas às questões de pesquisa.

1.4 Objetivos

O objetivo geral do presente trabalho é a proposição de uma Taxonomia de Pesquisa em Computação por meio de uma compilação de conceitos que sirva não apenas como ferramenta pedagógica, mas também como apoio às práticas de condução de pesquisas científicas.

Como objetivos secundários, busca-se evidenciar as etapas da produção do conhecimento científico e a inserção da Computação dentro da ciência. Essa abordagem visa proporcionar uma compreensão mais aprofundada das diversas formas que a pesquisa científica em Computação pode assumir, enfatizando a importância de ter esses conceitos complementares em mente. Dessa forma, busca-se preencher a lacuna da falta de conhecimento científico apontada pela bibliografia como uma das possíveis causas de problemas nos resultados de pesquisas científicas em Computação

1.5 Resultados Alcançados

Para atingir os objetivos deste trabalho, realizou-se uma revisão sistemática de literatura (RSL). Esta revisão culmina na proposição de uma taxonomia de pesquisa composta por 96 descritores distribuídos em 10 categorias de técnicas aplicáveis à pesquisa científica em Computação.

A taxonomia foi validada por meio da criação de uma ontologia com base em seus descritores e conceitos correspondentes. Essa ontologia foi investigada em 2.508 artigos científicos publicados entre 2018 e agosto de 2022 em 68 periódicos da ACM. As 124.601 ocorrências foram analisadas quanto ao contexto de uso, utilizando o algoritmo semi-supervisionado k-NN. O resultado final indicou uma conformidade de 73,37% do uso no texto com os conceitos subjacentes aos descritores da taxonomia.

Numa segunda etapa da análise dos dados, foram consideradas as divisões da Computação como área do conhecimento, levando em conta as classificações da SCOPUS, ACM e Sociedade Brasileira de Computação (SBC). Os resultados abrangem todas as divisões propostas pelas três entidades.

Ao final, os resultados indicam a completude e aplicabilidade da taxonomia à área da Computação.

1.6 Contribuições

A principal contribuição desta tese é uma taxonomia de pesquisa em Computação, apresentada em termos de suas categorias, divisões e conceitos. Acompanha a apresentação desta taxonomia um organograma, desenvolvido para que uma representação gráfica forneça um mapa mental para auxiliar na memorização durante o

processo pedagógico da taxonomia. Tal organograma também ser utilizado como um guia rápido de consulta durante a elaboração do processo metodológico em pesquisa.

O ciclo evolutivo das etapas da pesquisa científica e os pontos de inserção da computação dentro da ciência como um todo são também contribuições deste estudo. Para estas etapas não apenas conceitos associados são apontados, mas também foram elaboradas representações gráficas para auxiliar na visualização de tais conceitos.

Contribuições adicionais dizem respeito à RSL, um estudo de caso do RODA (OSORIO; DIAS; CAVALHEIRO, 2018), e um arcabouço de software para realizar a extração das informações e classificação destas. Este arcabouço de software encontra-se disponibilizado no GitHub, para reutilização e captação de contribuições públicas (<https://github.com/metasciencestudies/taxonomy>).

1.7 Organização do Trabalho

O texto desta tese está organizado da seguinte forma. Após o presente capítulo introdutório, no Capítulo 2, são revisados diversos conceitos relacionados à pesquisa científica. No Capítulo 3 é apresentado o resultado de uma revisão sistemática da literatura conduzida sobre o tema. A revisão realizada forneceu embasamento para a proposição da taxonomia de pesquisa em Computação, apresentada no Capítulo 4, juntamente com seu processo de validação. No Capítulo 5, então, são apresentados complementos à taxonomia visando uma melhor contextualização pedagógica da pesquisa científica em Computação como um todo. No Capítulo 6, é documentada uma argumentação sobre os resultados obtidos. Finalmente, no Capítulo 7, são apresentadas as conclusões alcançadas com o presente trabalho e desdobramentos de ações vislumbradas para sua continuidade.

2 REVISÃO CONCEITUAL

O ser humano, desde seus primeiros anos de idade, instintivamente busca a classificação e organização dos objetos e elementos que compõem seu ambiente. A elaboração e aplicação de tais mecanismos fazem parte do processo natural de aprendizagem. Inicialmente cores, formas e tamanhos são explorados como atributos para categorização. Na elaboração de tais processos, quaisquer características podem ser aplicadas, de uma forma ou outra, como critérios de similaridade para se organizar as *coisas da vida* (VESSEY; RAMESH; GLASS, 2005).

Classificar fornece para a área em questão um conjunto de construtores comuns para que possa ser descrita de maneira sistemática, assim como pode-se perceber o desenvolvimento de nichos futuros por meio da interseção de dois ou mais critérios de organização (VESSEY; RAMESH; GLASS, 2005). Antes de iniciar a consideração de uma forma de classificar ou organizar a Pesquisa em Computação, é fundamental organizar inicialmente o pensamento sobre a conceituação do que será estudado e as diferentes perspectivas envolvidas.

A ciência nasce a partir do pensamento humano buscando explicações sobre os fenômenos e acontecimentos que o cercam e de si próprio. O domínio da ciência busca satisfazer as necessidades cotidianas e se oferece como instrumento para controlar a natureza (KÖCHE, 2016). Dessa forma o conhecimento científico acaba por tomar conta das decisões cotidianas, de forma semelhante a uma consulta à previsão do tempo para escolha da roupa ao sair de casa.

2.1 O que é Ciência

Conceitualmente não há apenas uma definição do que é ciência (KÖCHE, 2016). Historicamente, pode ser classificada em ciência *grega*, *moderna* e *contemporânea*. Essa divisão ocorre não só pelo caráter temporal óbvio, mas também em função do próprio entendimento do que é ciência.

A ciência *grega*, conhecida como filosofia da natureza, buscava apenas explicar o homem e seu meio. Baseando-se no pensamento sobre aquilo que é percebido pelos

sentidos de maneira qualitativa. Não se distinguia filosofia de ciência para se fazer de maneira silogística¹ generalizações (KÖCHE, 2016).

Na categoria *moderna* de ciência desta classificação histórica é introduzida a experimentação baseada nos métodos matemáticos existentes em uma abordagem quantitativa. As generalizações são quantitativamente indutivas por meio de um método científico indutivo-confirmado (KÖCHE, 2016). A concepção da ciência contemporânea se distancia das categorias anteriores, pois não basta apenas ser provado verdadeiro para ser aceito como científico. Pelo contrário, ela envolve uma contínua investigação, resultando em uma constante reconstrução de teorias e métodos investigatórios (KÖCHE, 2016). Esse processo caracteriza-se como um meta-análise sistemática do próprio conhecimento científico.

(WAZLAWICK, 2010) apresenta o conceito de ciência como sendo o esforço para descobrir ou aumentar o conhecimento humano, indicando que a pesquisa deve ser realizada de acordo com os princípios do método científico e ainda discute seu impacto em trabalhos de iniciação científica, mestrado e doutorado.

2.2 Divisões da Ciência

(WAZLAWICK, 2010) classifica a ciência em duas categorias: Formal e Empírica. A Ciência Formal estuda ideias independentes de sua aplicabilidade, sendo o objeto de estudo ligado a processos lógicos ou matemáticos. Já a Ciência Empírica estuda os fenômenos do mundo real, sendo obrigatório o uso de observações para embasar as conclusões sobre um determinado objeto de estudo.

A Ciência Empírica é ainda subdividida por (WAZLAWICK, 2010) em Ciências Naturais e Ciências Sociais. As Ciências Naturais exploram o universo, independente da existência do fator humano. As Ciências Sociais consideram o fator humano e suas inter-relações.

A Figura 1, proposta por (DODIG-CRNKOVIC, 2002), apresenta hierarquicamente a divisão da ciência em diversas áreas temáticas. Quanto mais interno, maior é a relação entre as divisões. A figura também expressa o nível de complexidade associado ao tipo de ciência: quanto mais interno, mais simples são os objetos de pesquisa e a prova de resultados.

Lógica e Matemática são as ciências nas quais se obtém o maior grau de certeza nos resultados. Os objetos de pesquisa nestes casos consistem em conjuntos de elementos matemáticos extraídos da linguagem natural, sendo a prova/método do resultado científico geralmente obtida por indução e dedução.

¹O silogismo se apresenta como uma forma de raciocínio dedutivo estruturado formalmente, composto por duas premissas. A partir dessas premissas, é inferida uma terceira proposição que representa a conclusão obtida das duas primeiras [p. ex.: “todos os homens são mortais; os gregos são homens; logo, os gregos são mortais”] (KODRATOFF, 2014).

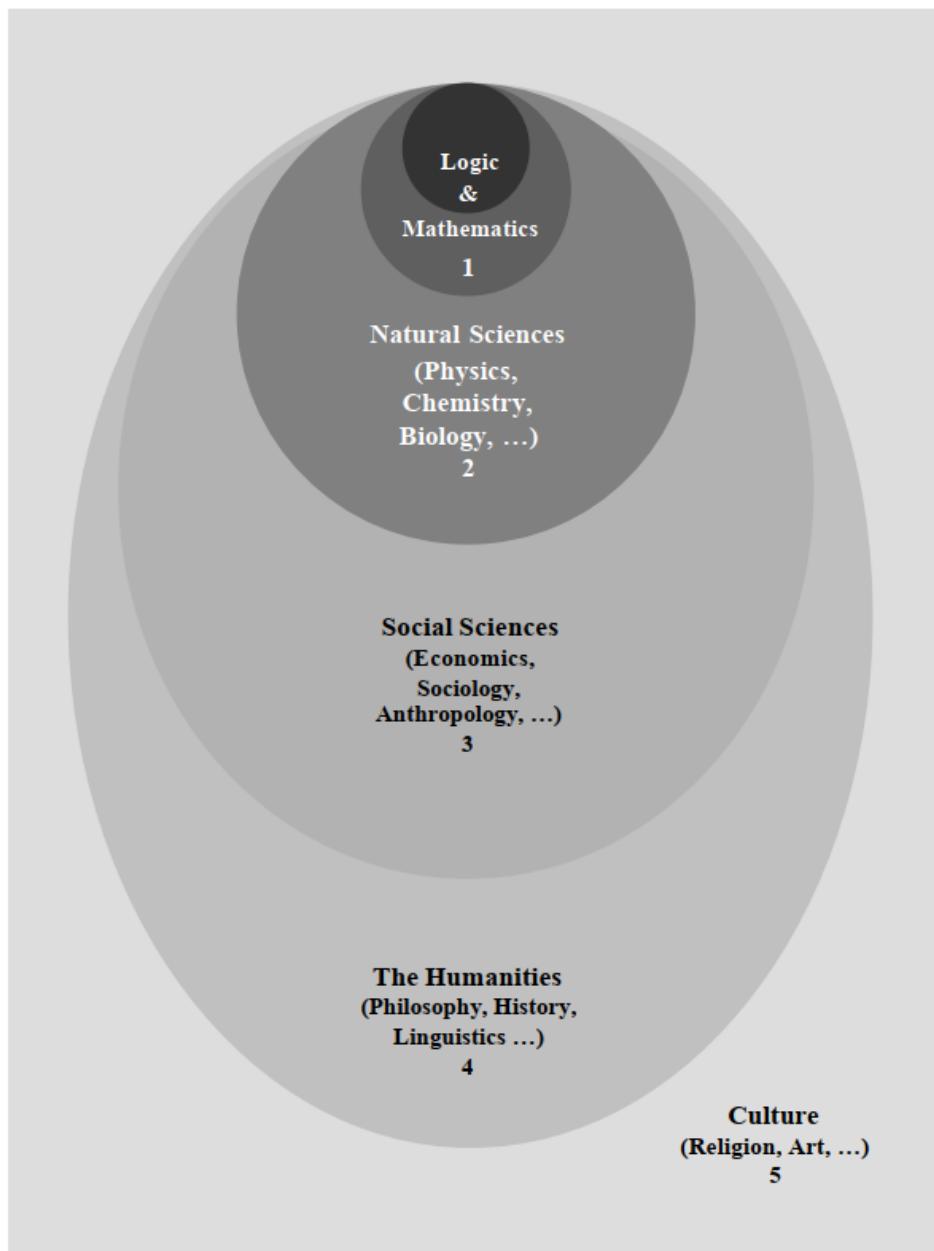


Figura 1 – Divisão da ciência em diversas áreas temáticas (DODIG-CRNKOVIC, 2002).

As Ciências Naturais abrangem disciplinas de natureza mais dura e temas empíricos experimentais, aplicando o método hipotético-dedutivo a objetos do cotidiano, organismos vivos e fenômenos naturais.²

Nas Ciências Sociais, o foco recai sobre o homem e suas inter-relações. Aplicando métodos qualitativos/quantitativos associados à estatística. Já as Ciências Humanas estudam aspectos do coletivo humano, predominando o método qualitativo e rara-

²O método hipotético dedutivo refere-se ao processo de pesquisa científica amplamente utilizada, partindo da formulação de hipóteses são previstos resultados considerando que estas hipóteses sejam verdadeiras, partindo à experimentação ou observação e então analisando os resultados, permitindo apresentar conclusões que validem ou refutem as hipóteses inicialmente apresentadas.

mente necessitando do uso da estatística.

A ciência pode ainda ser classificada quanto à sua aplicabilidade: puras ou aplicadas (WAZLAWICK, 2010). As ciências puras estudam conceitos básicos ou fundamentais relacionados a ideias ou fenômenos físicos, sem necessidade de aplicação imediata. As ciências aplicadas têm em vista a descoberta para aplicação imediata a um processo, visando alguma melhoria deste.

Exatas e Inexatas representam outra classificação no campo científico, conforme destacado por (WAZLAWICK, 2010) ao abordar a reflexão sobre a natureza dos resultados. As ciências exatas, de acordo com essa categorização, apresentam resultados que são precisos, completamente mensuráveis e previsíveis. Em contraste, as ciências inexatas possibilitam a previsão de um comportamento geral do objeto de estudo, embora não assegurem resultados sempre esperados.

A classificação da ciência em relação ao rigor metodológico separa os estudos em duros ou moles. Duros são as ciências que utilizam rigor científico em observações, experimentos e deduções. Moles são ciências cujas evidências baseiam-se em generalizações de casos específicos sem rigor científico³ (WAZLAWICK, 2010).

Em relação à periodicidade dos eventos científicos, a ciência é classificada como nomotética e ideográfica. As ciências nomotéticas estudam eventos que se repetem, levando à descoberta e formulação de evidências. Enquanto as ciências ideográficas estudam eventos que não se repetem, mas ainda assim podem ter validade científica (WAZLAWICK, 2010).

2.3 O Método Científico

O método científico é o mecanismo utilizado para produzir e validar teorias científicas (AYASH, 2014). A produção não se limita apenas à formulação de teorias sobre determinados temas, mas abrange um espectro mais amplo. Isso inclui a elaboração de teorias acerca do processo teórico em si ou o desenvolvimento de ferramentas que facilitam a produção e aprimoramento contínuo das teorias científicas.

A natureza recursiva da ciência é evidenciada pelo método científico. Uma vez formulada uma hipótese e a partir dela feitas as previsões e observações, essas etapas permanecem em um sucessivo ciclo de refinamento. Este refinamento persiste até que se obtenha a confirmação da hipótese (DODIG-CRNKOVIC, 2002).

O empirismo estabelece que toda teoria científica deve ser fundamentada em observações que possam ser testadas para produzir generalizações. O positivismo por sua vez, preconiza que a ciência deve basear-se apenas em valores humanos. Ignorando valores metafísicos, teológicos e místicos (WAZLAWICK, 2010).

³Dados ou Evidências Anedotais – Baseadas em experiências, testemunhos ou acontecimentos pessoais específicos para generalizar evidências sobre um determinado objeto de estudo

O pragmatismo é uma corrente filosófica que assume que não é possível explicar ou saber exatamente o que é a realidade e o papel da ciência é o de explicar as observações a ela relacionadas. Em contrapartida o realismo defende exatamente o oposto. Afirmado que a ciência explica a realidade (WAZLAWICK, 2010).

Outra característica do método científico é a objetividade, que representa a possibilidade de duas pessoas com qualquer nível de conhecimento chegarem às mesmas conclusões ao analisarem os mesmos conjuntos de observações. Isso desconsidera opiniões e subjetividades pessoais (WAZLAWICK, 2010).

A indução é outro princípio do método científico no qual uma situação que se sustenta todos os casos observados permanece válida até que seja provado o contrário. A indução pode demonstrar que um conjunto infinito de observações apresenta uma determinada característica por meio de uma regra que consiga gerar todos os elementos desse conjunto (WAZLAWICK, 2010).

O princípio da refutação estabelece que qualquer teoria científica que busca explicar as observações é sempre possível de ser invalidada ou contradita quando não consegue explicar novas observações. Um estudo de caso pode não conseguir provar uma regra geral, mas pode refutá-la, mostrando que a regra geral não consegue explicar a situação particular das observações estudadas (WAZLAWICK, 2010).

O coherentismo é um princípio intimamente ligado ao pragmatismo. Em nenhum momento, por este princípio, se afirma que uma teoria científica explique a realidade. Pelo contrário, ela se destina a esclarecer as observações estudadas, e, com base no princípio da indução, na ausência de refutação, considera-se essa teoria válida até que seja demonstrado o contrário (WAZLAWICK, 2010).

O princípio da navalha de Occam (BADIUS; TRECHSEL, 1495) diz que, quando há mais de uma explicação para uma mesma teoria científica, a mais simples dentre aquelas que efetivamente explicam as observações deve ser preferida em detrimento de outras (WAZLAWICK, 2010).

2.4 Classificação

A classificação da pesquisa científica é algo controverso devido aos vários aspectos epistemológicos da ciência. Cada autor ao classificá-la apresenta a ciência sob seu entendimento. Não há consenso sobre a classificação, tampouco sobre os aspectos dos quais ela teria origem. Enquanto alguns autores sintetizam tal classificação, outros o fazem com maior detalhamento, ou este é feito apenas sob determinado ponto de vista.

Quanto à natureza, a pesquisa científica pode ser classificada como Trabalho Original, que apresenta conhecimento novo a partir de observações, teorias; *Surveys*, trabalhos científicos que sistematizam um determinado nicho do conhecimento científico.

fico (WAZLAWICK, 2010).

Quanto aos objetivos, a classificação da pesquisa científica é descrita como Exploratória, Descritiva e Explicativa. Exploratórias são aquelas pesquisas em que necessariamente não há uma hipótese a ser testada ou provada e servem como levantamento do estado da arte nos estágios iniciais de um processo de pesquisa. Descritivas são aquelas pesquisas que buscam dados sobre um determinado objeto de pesquisa e que não formulam teorias sobre as observações, sendo meramente descritas. Explicativas são aquelas pesquisas que, após coletar e analisar dados, buscam-se explicações e formulam-se teorias, se caracterizando, neste entendimento, como a pesquisa científica por excelência (WAZLAWICK, 2010).

Quanto aos procedimentos técnicos, a pesquisa científica pode ser classificada como bibliográfica, documental, experimental, de levantamento, pesquisa-ação e exploratória.

- A pesquisa bibliográfica implica no estudo de outras publicações científicas (WAZLAWICK, 2010) para tentar explicar um determinado problema (KÖCHE, 2016). Seu objetivo é conhecer as principais contribuições sobre um determinado tema (KÖCHE, 2016).
- A pesquisa documental consiste na análise de documentos ou dados que ainda não foram publicados (WAZLAWICK, 2010).
- A pesquisa experimental é aquela pesquisa caracterizada pela manipulação de um aspecto da realidade (WAZLAWICK, 2010) (AYASH, 2014) em uma situação específica com alta independência entre variáveis, visando estudar a relação entre elas (KÖCHE, 2016). Sua condução implica na mensuração, por meio de variáveis, da manipulação realizada e a observação das variações desta mensuração pode levar à conclusão. Tal mensuração é realizada por meio de critérios estatísticos rigorosos para que possa ser aceita e generalizável (WAZLAWICK, 2010) (AYASH, 2014).
- A pesquisa de levantamento trabalha basicamente com dados já existentes para chegar às conclusões (WAZLAWICK, 2010). A coleta de dados é feita dentro do ambiente estudado por meio de um instrumento padronizado.
- A pesquisa exploratória, ainda elencada por (KÖCHE, 2016), é aquela em que não se têm as informações ou conhecimentos prévios, sendo necessário um processo investigatório para o levantamento das características e possíveis variáveis de interesse.
- A pesquisa-ação caracteriza-se pela iteração participativa entre pesquisador e pesquisando com um propósito específico (WAZLAWICK, 2010).

2.5 Teórica e Experimental

A Ciência Teórica está intimamente ligada à Lógica e Matemática, seguindo a maneira clássica de construir teorias⁴ como sistemas lógicos com definições rigorosas de objetos e operações para derivar/provar teoremas (DODIG-CRNKOVIC, 2002). Procura entender os limites e poderes computacionais, tendo como principais fundamentos Modelos formais/conceituais, Níveis de Abstração e Eficiência.

A Ciência Experimental não se aplica apenas a determinadas divisões da ciência (Figura 1). Sua utilização está associada à solução de problemas onde necessariamente há o desenvolvimento de uma solução complexa (DODIG-CRNKOVIC, 2002) (AYASH, 2014). A Ciência Experimental é amplamente utilizada para identificar conceitos que facilitam soluções para um problema e em seguida avaliar essas soluções por meio da construção de protótipos.

(DODIG-CRNKOVIC, 2002) ainda cita as simulações como uma terceira ramificação metodológica. Porém destaca que as simulações são guiadas pela teoria por resultados experimentais os quais podem frequentemente sugerir novos experimentos e modelos teóricos. Logo, a simulação é uma aplicação das Ciências Teórica e Experimental.

Não há vantagem entre um tipo ou outro, contanto que a pesquisa seja realizada com todo o rigor científico e a natureza do tipo adotado seja adequada ao objeto estudado (KÖCHE, 2016).

Vantagens da pesquisa experimental segundo (KÖCHE, 2016). A primeira é a fácil manipulação das variáveis proporcionando elevado controle dos experimentos e do detalhamento da relação de dependência entre variáveis. Já a segunda está na a flexibilidade das situações experimentais que facilitam a testagem dos vários aspectos das hipóteses. Por último a possibilidade de replicação do estudo é característica fundamental da pesquisa experimental.

(KÖCHE, 2016) cita como desvantagem da pesquisa experimental a falta de generalização. Uma vez que um resultado reflete uma situação experimental, eventualmente realizada em bancada ou, ao menos, em ambiente controlado, nem sempre reflete uma situação da vida real.

2.6 O Resultado Científico

A aceitação de uma teoria se dá por meio dos resultados obtidos, por meio do raciocínio lógico e das observações e/ou experimentos (DODIG-CRNKOVIC, 2002). Tais

⁴Na lógica tradicional, um axioma ou postulado é uma sentença ou proposição que não é provada ou demonstrada e é considerada como óbvia ou como um consenso inicial necessário para a construção ou aceitação de uma teoria. Por essa razão, é aceito como verdade e serve como ponto inicial para dedução de outras verdades (dependentes de teoria) (BEZIAU, 2020).

resultados precisam, necessariamente, ser reproduzíveis para que possíveis discrepâncias possam ser estudadas. Conferindo-lhes força e confiabilidade, o que transforma uma hipótese em teoria efetivamente (AYASH, 2014).

No início do século XX quando publicada, a teoria da relatividade de Albert Einstein era vista como um conjunto de especulações coerentes, porém sem valor algum. Sob a análise do entendimento corrente na área a nova teoria contrariava outros princípios científicos considerados como exatos à época, e não apresentavam evidência empírica, tampouco testes experimentais (KÖCHE, 2016).

2.7 Multidisciplinaridade da Ciência

Nos dias atuais, há uma tendência da ciência moderna circular por diversas áreas do conhecimento (DODIG-CRNKOVIC, 2002) em uma abordagem multidisciplinar ou transversal (AYASH, 2014). Isto se deve ao fato de que a comunicação entre as diferentes áreas acontece de maneira muito mais intensa, assim como à necessidade de apropriação de métodos externos à área estudada para obtenção dos resultados.

Ao analisar a Computação como uma grande área de conhecimento, é observável a inclusão de elementos de outras áreas da ciência além da Lógica e Matemática. O inverso também é verdadeiro: a Computação como uma área de pesquisa recente em relação às demais está paulatinamente se inserindo no contexto das demais. Efetivamente se observa uma relação simbiótica entre as áreas, onde uma se utiliza da outra como método para chegar aos resultados. Neste sentido, a necessidade de um método novo em uma área pode fazer outra avançar em seu conhecimento, a fazer ciência (GHEZZI, 2020).

2.8 Ciência e Tecnologia

A Computação é uma área intimamente ligada à tecnologia, sendo necessário posicionar Ciência e Tecnologia. Enquanto a Tecnologia representa a aplicação dos conhecimentos, a Ciência a compreensão e explicações dos conhecimentos explorados. A Ciência busca explicações às quais a Tecnologia vai se utilizar. Em Computação, é difícil distinguir uma da outra, uma vez que quase sempre elas andam juntas. Dentro destes conceitos (WAZLAWICK, 2010) reflete dizendo que um trabalho é efetivamente científico se nele está contida Ciência.

Distinguir Ciência e Tecnologia segundo a ótica contemporânea da Ciência pode-se mostrar uma tarefa não trivial. Hoje a Ciência é multidisciplinar e mais complexa do que aquela dos tempos de Aristóteles e Platão (DODIG-CRNKOVIC, 2002).

2.9 Considerações Finais do Capítulo

Neste capítulo foram abordadas conceitos sobre a ciência e seus diferentes paradigmas. Como ponto de partida do referencial bibliográfico, assume-se que Ciência primeiro categoriza-se como formal ou empírica. Discutiu-se a complexidade do método científico, algumas de suas correntes filosóficas e princípios, como empirismo, positivismo e pragmatismo. Explorou-se a diversidade na classificação da pesquisa científica e suas características, desde trabalhos com relatos de temas originais originais até compilações destes, os chamados *surveys*. Abordou-se a Ciência Teórica, ligada a Lógica e Matemática, e a Ciência Experimental, aplicada na solução de problemas complexos. Destacou-se a multidisciplinaridade atual na Ciência, especialmente na Computação, e a relação entre Ciência e Tecnologia, ressaltando a complexidade contemporânea da Ciência.

No próximo capítulo é apresentada uma revisão sistemática de literatura realizada com o propósito de elucidar informações, divisões, classificações e conceitos relacionados à pesquisa científica específica a área da Computação. Esta revisão busca evidenciar os diferentes aspectos e abordagens que permeiam a pesquisa nesta área específica. Fornecendo uma base para a análise crítica e a identificação de lacunas no conhecimento atual sobre a pesquisa científica em Computação.

3 REVISÃO SISTEMÁTICA DE LITERATURA

Neste capítulo é apresentada a Revisão Sistemática de Literatura (RSL) realizada com o objetivo de levantar informações, divisões, classificações e conceitos sobre a pesquisa científica em Computação documentadas em trabalhos já publicados. A partir desse levantamento, empreende-se um meta-estudo da ciência em Computação, visando compilar todas essas informações em um único documento. A investigação sobre a pesquisa científica em Computação, suas possíveis classificações ou divisões, é conduzida por meio da RSL apresentada neste capítulo.

3.1 Metodologia

Para elaborar a presente revisão sistemática da literatura, utilizou-se das orientações do protocolo RODA (OSORIO, 2020). O protocolo RODA complementa a proposta do presente trabalho e encontra-se descrito na Seção 5.2.

As seções que sucedem dentro do texto deste capítulo estão estruturadas segundo as etapas mostradas na Figura 2. O Planejamento é onde as questões de pesquisa são elaboradas baseadas na revisão conceitual da Capítulo 2. A chave de pesquisa é elaborada, refinada e a versão final elaborada. São escolhidas as bases indexadas com base nos testes efetuados na montagem da chave de pesquisa levando em consideração a relevância de cada uma dentro da Computação. Os critérios de seleção são estipulados e aplicados na etapa do Desenvolvimento da RSL. Nesta a pesquisa por referências nas bases indexadas é efetuada tendo seus resultados sistematizados em uma ferramenta específica onde é feita a remoção das referências duplicadas. Na etapa final é elaborado o Relatório da RSL. O qual comprehende os resultados numéricos distribuídos por base indexada, a síntese das referências selecionadas para compor a RSL, a tabulação dos conceitos evidenciados na síntese e finalmente as questões de pesquisa são respondidas.

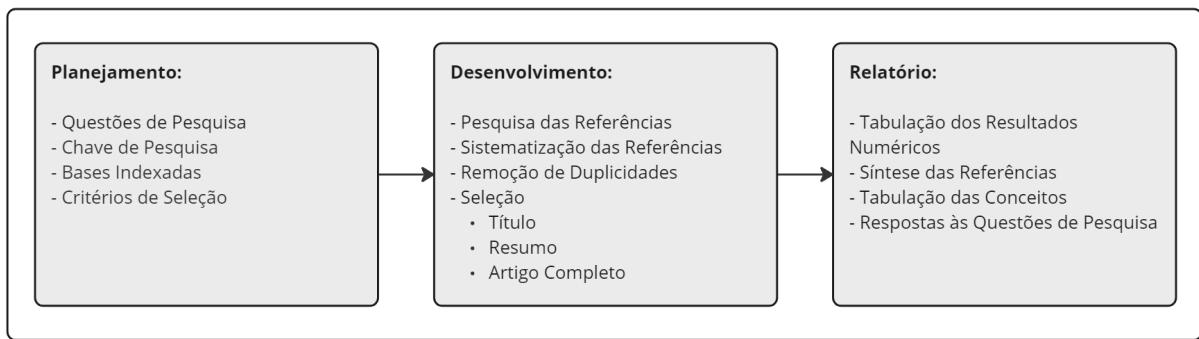


Figura 2 – Etapas da Revisão Sistemática de literatura.

3.1.1 Questões de pesquisa

Dentro do contexto de um meta-estudo da pesquisa científica em Computação, buscando uma melhor compreensão do que é estudado e como é realizado, as questões que se pretende responder com esta revisão são:

- Q1** - Existe alguma divisão explicativa para a pesquisa científica em Computação, considerando tipos de estudos, objetos de pesquisa e métodos avaliativos aplicáveis à área, independentemente da denominação utilizada, seja esta classificação, divisão ou taxonomia?
- Q2** - Quais são quais os aspectos, características e/ou conceitos a serem considerados para se montar uma classificação ou divisão da pesquisa científica em Computação visando sua melhor compreensão?
- Q3** - Qual a inserção da Computação dentro da Ciência e onde esta acontece, considerando aspectos da multidisciplinaridade.
- Q4** - Quais as divisões de estudo atuais da Computação como área de conhecimento, as quais devem ser consideradas na montagem de um panorama abrangente da área?

3.1.2 Chave de pesquisa

A chave de pesquisa derivada das questões de pesquisa propostas foi concebida para trabalhar com as associações entre a grande área da Computação e suas técnicas científicas. Suas classificações, possíveis taxonomias, terminologias ou divisões e também questões relativas à natureza da pesquisa científica em Computação.

Os termos foram utilizados individualmente para fins de testes. Os resultados obtidos para cada palavra chave foram analisadas com a finalidade de refinar a chave de pesquisa para esta revisão sistemática. Na medida em que uma nova palavra-chave de interesse era verificada nos testes, esta também foi submetida a testes e

dependendo do número de resultados obtidos (preferencialmente alto), ela foi definitivamente adicionada aos termos da chave de pesquisa.

Após definidos os termos que comporiam a chave de pesquisa, a eles foi adicionado o termo *computer science* como designador e restritor de área do conhecimento. Os operadores lógicos *AND* e *OR* foram adicionados à montagem final da sentença de busca para garantir a associação lógica de maneira que todos os resultados estivessem ligados a Computação, bem como conteriam pelo menos um dos termos pesquisados.

“computer science” AND (“research methodologies” OR “research classification” OR “research taxonomy” OR “research terminology” OR “research divisions” OR “research approaches” OR “research nature” OR “research history”)

3.1.3 Bases Indexadas

A escolha das bases teve como orientação aquelas onde havia a possibilidade de obter o resultado da pesquisa diretamente em formato *BibTeX*, ou em algum formato que permitisse automatizar a conversão para este formato. Foram escolhidas e utilizadas como mecanismos de busca, as seguintes bases indexadas:

- Springer Link (<http://link.springer.com>)
- IEEE Digital Library (<http://ieeexplore.ieee.org>)
- ACM Digital Library (<http://portal.acm.org>)
- Science@Direct (<http://www.sciencedirect.com>)
- Google Acadêmico (<https://scholar.google.com.br/>)

3.1.4 Critérios de seleção

Após a submissão da chave de pesquisa às bases indexadas, foi feito o *download* e a conversão para o formato *BibTeX* (quando necessário) dos resultados obtidos. Esses arquivos de referências bibliográficas foram importados para o site *Parsifal*, que se oferece como uma ferramenta para realização de revisões sistemáticas¹.

A fase de seleção dos artigos começa pela exclusão das referências duplicadas, iniciou logo em seguida à importação das referências. A ferramenta utilizada (*Parsifal*) permite a identificação automática destes casos. Sendo adotada como estratégia de remoção das publicações duplicadas, o critério de marcar como duplicado os registros provenientes do Google Acadêmico em primeiro lugar. Quando esta base contivesse

¹Parsifal: <https://parsif.al>, acesso em 12 de março de 2024.

um dos registros duplicados. Em segundo lugar a base com maior número de registros, no caso do Google Acadêmico não estar entre as bases envolvidas.

A opção por remover os registros do Google Acadêmico se justifica pelo fato desta base de dados ser mais um mecanismo de busca do que uma base indexada propriamente dita. Seus registros são concentradores de várias outras bases indexadas. Porém, nele foram encontrados registros que não estavam em nenhuma outra base indexada daquelas selecionadas para esta RSL.

O critério de descarte da referência pela leitura do título se deu nos casos em que o título indicava claramente que o artigo não abordava o tema estudado. Caso restasse alguma dúvida a referência era mantida para posterior leitura do resumo (*abstract*), onde novamente o descarte leva em consideração a adequação da referência ao tema estudado.

As referências remanescentes tiveram seu conteúdo lido na íntegra como etapa final do processo de seleção da revisão sistemática. Desta forma foi composto o material a ser utilizado como referencial bibliográfico. Os artigos selecionados deveriam abordar temas condizentes às questões de pesquisa bem como à chave de pesquisa. A temática deveria conter conceitos, possíveis divisões e técnicas de pesquisa de toda natureza relacionados à computação. Apenas artigos que abordam o tema “pesquisa científica em Computação”, aqueles cujo olhar coloca em foco a metodologia científica e que estudam a Ciência em si.

3.2 Resultados

No total foram encontradas 1.389 publicações científicas contemplando a chave de pesquisa utilizada. A Tabela 1 sumariza esses resultados. Do total após importação, a ferramenta *Parsifal* apontou 143 referências duplicadas que foram descartadas segundo os critérios já mencionados.

Após a etapa de seleção das referências de interesse pela análise dos títulos, restaram 172 trabalhos para leitura dos seus resumos. Após essa etapa de leitura, permaneceram 34 artigos para leitura na íntegra. Destes 34 artigos, 21 foram selecionados e 13 considerados sem interesse para o trabalho conduzido.

3.3 Desenvolvimento / Resumo Referências

Nesta seção são compiladas e sistematizadas as ideias dos autores cujos trabalhos foram selecionados para esta revisão sistemática da literatura. Estas ideias estão dispostas nas subseções seguintes, agrupadas segundo categorias afins descritas nos trabalhos originais.

Tabela 1 – Resultado Pesquisa nas Bases Indexadas

Base	Total	Duplicados	Selecionados
ACM Digital Library	80	10	4
Google Acadêmico	956	117	13
IEEE Digital Library	87	0	1
Science@Direct	100	7	1
Springer Link	166	9	2
Total:	1.389	143	21

3.3.1 Conceitos e divisões da Ciência

(GHEZZI, 2020) adota o conceito de Ciência como sendo o conjunto de conhecimentos organizados sobre determinado assunto. Sob essa ótica poderia ser classificada em tantos quantos forem os temas, de formas diferentes e para diferentes propósitos. O autor elenca como uma possível divisão para as Ciências: as Ciências Sociais e Humanas, Ciências da Vida e Ciências da Física e Engenharia, conforme mostra a Tabela 2.

As Ciências Sociais e Humanas compreendem o estudo das pessoas e suas relações. Sua contextualização política, histórica, cultural e econômica enquanto sociedade. Ciências da Vida abrangem todos os aspectos da vida e sua formação química e evolucionária. Já as Ciências da Física e Engenharia compreendem o estudo da matéria e das leis da natureza e do universo que as formam ou regem, assim como aquela construída pelo próprio homem, a terra e o cosmos.

Em resumo tem-se os seres vivos (Ciências da Vida), o universo em que vivem (Ciências da Física e Engenharia) e sua relação (Ciências Sociais) (GHEZZI, 2020). Essa visão sobre a Ciência simplifica, mas ainda contempla, a classificação de (DODIG-CRNKOVIC, 2002) ao mesmo tempo que incorpora os conceitos que (KÖCHE, 2016) e (WAZLAWICK, 2010) dão para Ciência. Este últimos autores argumentam que a Ciência nasce do pensamento humano buscando explicações sobre o meio que o cerca em um esforço para aumentar o conhecimento.

A maneira de se fazer Ciência se dá por meio da pesquisa científica. (GHEZZI, 2020) ainda caracteriza a Ciência como sendo o trabalho criativo e sistemático realizado com o objetivo de aumentar o conhecimento humano, cultural e da sociedade. O autor justifica que a criatividade leva a novas descobertas e a sistematização guia a criatividade a novos resultados. Esses novos resultados, por sua vez, precisam ser obtidos por meio de um processo rigoroso que os confira validade e significância para que efetivamente possam gerar conhecimento.

Outro conceito de pesquisa científica é definido por (HASSANI, 2017) como sendo as diversas estratégias para a geração de conhecimento sobre o meio ambiente, o comportamento e as experiências humanas. Essas estratégias objetivam evidenciar o

Tabela 2 – Divisão da Ciência por (GHEZZI, 2020)

Ciências Sociais e Humanas	Indivíduos, Mercados e Organizações Instituições, Valores, Crenças e Comportamento Sociedade, Diversidade e População A mente humana e sua complexidade Culturas e Produção Cultural O estudo do passado humano
Ciências da Vida	Bioquímica e Biologia Molecular/Estrutural Genética, Genoma, Bioinformática e Sistemas Biológicos Biologia Celular e do Desenvolvimento Fisiologia Patofisiologia e Endocrinologia Neurociência e Distúrbios Neurais Imunologia e Infectologia Ferramentas de Diagnóstico, Terapêuticas e Saúde Pública Biologia Evolucionária, Populacional e Ambiental Ciências da Vida Aplicadas e Biotecnologia não-médica
Ciências da Física e Engenharia	Matemática Constituição Fundamental da Matéria Física da Matéria Condensada Física e Química Analíticas Química Sintética e Materiais Ciência da Computação e Informática Engenharia de Sistemas e Comunicação Engenharia de Produtos e Processos Ciências do Universo Sistemas da Terra

pensamento do pesquisador de maneira lógica, comprehensível, confirmável e útil.

3.3.2 Etapas da pesquisa científica

(HASSANI, 2017) sugere que, de um modo geral, todos os conceitos de pesquisa científica possuem em comum três componentes: uma questão de pesquisa, uma abordagem ou método para resolver ou investigar a respeito da questão de pesquisa, e os resultados das abordagens e métodos utilizados.

A pesquisa científica para (GHEZZI, 2020) não é um processo linear, mas um processo iterativo entre etapas. Cita como etapas a escolha do problema, questões de pesquisa, formulação das hipóteses, desenvolvimento de teorias, desenvolvimento de ferramentas, o projeto e execução de experimentos. A cada etapa vencida, há a possibilidade de retorno às anteriores para refino ou redefinição das ideias.

As abordagens destes dois autores, Hassany e Ghezzi, não indicam conflito ou visões diferentes do que se refere à caracterização das etapas científicas. O primeiro autor oferece uma visão de módulos arquiteturais das etapas envolvidas na pesquisa científica, enquanto o segundo oferece uma visão dos processos envolvidos na interação entre tais módulos. Assim, pode-se entender as duas abordagens como complementares.

3.3.3 Divisões das áreas de conhecimento da Computação

Segundo (GHEZZI, 2020) a Computação como ciência e a própria a pesquisa científica em Computação, são recentes. Dentro de suas reflexões, (GHEZZI, 2020) sugere a existência de uma perspectiva multifacetada para a pesquisa científica na área, uma vez que Computação possui forte interação multidisciplinar com outras áreas. Para melhor compreensão (GHEZZI, 2020) destaca a classificação segundo o *European Research Council* (ERC) para subáreas de Computação, conforme a Tabela 3.

(HASSANI, 2017) compartilha da mesma opinião de (GHEZZI, 2020) a respeito da multidisciplinaridade da Computação. (HASSANI, 2017) ainda vai além ao apresentar em seu trabalho o conceito de Pesquisa Interdisciplinar (IDR) como sendo um modo de pesquisa onde pesquisadores integram contextos de duas ou mais áreas do conhecimento para avançar a compreensão fundamental ou resolver problemas que estão fora do escopo de uma única disciplina ou área de pesquisa.

No trabalho de (HOLZ et al., 2006) é utilizada a lista dos Special Interest Groups (SIG) (COMPUTER MACHINERY, 2021), os quais foram propostos e são mantidos² pela ACM como ponto de partida para representar as principais áreas de pesquisa em Computação atendendo aos interesses da comunidade científica da área.

(RAMESH et al., 2004) utiliza a versão de 1998 da ACM *Computing Classifica-*

²Esta lista pode ser encontrada em <https://www.acm.org/special-interest-groups/alphabetical-listing> (acessado em 08/07/2021).

Tabela 3 – Divisão de Áreas da Computação segundo (GHEZZI, 2020)

-
- Arquitetura de computadores, Computação pervasiva e ubíqua
 - Sistemas computacionais, sistemas paralelos/distribuídos, redes de sensores, sistemas embarcados, sistemas ciber-físicos
 - Engenharia de *software*, sistemas operacionais, linguagens de computadores
 - Teoria da Computação, métodos formais e Computação quântica
 - Criptografia, segurança, privacidade e criptografia quântica
 - Algoritmos, algoritmos paralelos/distribuídos/redes, teoria dos jogos algorítmicos
 - Inteligência artificial, sistemas inteligentes, sistemas multi-agente
 - Computação gráfica, Computação visual, multi-media e jogos
 - Iteração homem-máquina, interfaces, visualização e processamento de linguagem natural
 - *Web* e sistemas de informação, banco de dados, recuperação da informação e fusão de dados
 - Aprendizado de máquina, processamento de dados estatísticos, processamento de sinais
 - Computação científica, simulação e ferramentas de modelagem
 - Bioinformática, Biocomputação, DNA e Computação molecular
-

tion Scheme (CCS)³ como um dos critérios de estudo na caracterização da pesquisa científica em Computação. Tal classificação atualmente encontra-se na última atualização realizada em 2012, e seu objetivo é refletir o estado da arte da pesquisa em Computação.

3.3.4 A pesquisa científica em Computação segundo diversos olhares

Nesta seção são descritas as divisões e critérios de divisão que foram encontrados na leitura dos artigos selecionados nesta RSL. A Tabela 4 sumariza as percepções de cada um dos autores sobre a pesquisa científica. Nela é claramente possível perceber as diferenças e convergências entre elas, as quais são discutidas na sequência desta subseção.

3.3.4.1 Natureza

As Engenharias são abordadas na classificação de (GHEZZI, 2020) como áreas que se destacam na inventividade e na criação de produtos, processos e métodos. Isso se deve, em parte, à concepção de Engenharia, que é frequentemente associada mais à tecnologia do que à Ciência em si. Nesse contexto, a tecnologia é vista como a aplicação prática dos princípios científicos. Este é o mesmo entendimento de (WAZ-LAWICK, 2010), o que leva à separação da Ciência em **Pura** e **Aplicada**. A primeira

³Disponível em <https://dl.acm.org/ccs> (acessado em 08/07/20/2021)

é aquela que cria e gera conhecimento independentemente da aplicação (HUSTADT, 2012), e a última é aquela que aplica esse conhecimento para os mais diversos fins (GHEZZI, 2020). Estendendo esses conceitos, (HASSANI, 2017) afirma que a pesquisa pura não espera uma mudança imediata na situação do campo de pesquisa, enquanto a aplicada precisa necessariamente modificar o cenário.

3.3.4.2 *Propósito Paradigma*

Cada paradigma de pesquisa fundamenta uma abordagem científica na Computação. Oferecendo uma perspectiva única sobre como o conhecimento é construído e investigado.

O paradigma **comportamental**, também conhecido como positivista, tem como objetivo encontrar a verdade. Ele testa hipóteses e teorias por meio da coleta e análise de dados propondo a generalização dos resultados utilizando técnicas estatísticas. Neste paradigma há clara separação dos fatos, os quais consistem na base do conhecimento científico. Sua ontologia assume que a realidade é independente da interpretação humana. Metodologicamente, utiliza critérios quantitativos e contexto-independentes ao estudar a relação entre fenômenos na busca pela verdade (PSCHEIDT, 2012).

O paradigma **interpretativo**, por sua vez, busca a compreensão do objeto de pesquisa. Está fundamentado na crença de que dados sem valor agregado não podem ser obtidos, pois os pesquisadores são influenciados por suas próprias percepções e acabam influenciando o objeto de pesquisa. Neste paradigma considera-se que a realidade é construída de maneira subjetiva e independente pelos humanos. Ontologicamente, assume a existência de múltiplas realidades contexto-dependentes, sendo a análise dessas realidades realizada por meio de critérios qualitativos.

O paradigma **crítico** questiona os porquês na tentativa de reduzir as desigualdades sociais. Analisando as barreiras do desenvolvimento e melhorando as condições sociais e econômicas com o objetivo de alcançar a emancipação. Com uma abordagem pelo paradigma crítico parte-se do pressuposto das desigualdades e as questiona. As evidências são obtidas por meio de critérios tanto qualitativos quanto quantitativos (PSCHEIDT, 2012).

Finalmente, o paradigma **Design Science** concentra-se em processos e produtos com alguma ênfase social. Seu principal propósito é a utilidade (PSCHEIDT, 2012). Este paradigma amplia as visões de mundo dos outros paradigmas. Cada um dos três primeiros (comportamental, interpretativo e crítico) possui pontos fortes e fraquezas. A combinação desses paradigmas com o *Design Science*, em algumas situações, pode abranger completamente o que está sendo investigado (PSCHEIDT, 2012).

3.3.4.3 Método

Métodos são ações humanas realizadas de maneira sistemática e em uma determinada ordem. Dentro da pesquisa, visam atingir os resultados (GHEZZI, 2020). O mesmo autor cita como principal método, e mais utilizado em Computação, o método **Experimental**, no qual a dedução e a indução são aplicáveis como principais mecanismos de prova de resultados e refinamento cíclico do método em si. A Figura 3, reproduzida a partir de (GHEZZI, 2020), ilustra esse processo cíclico de refinamento conforme sugerido pelo autor.

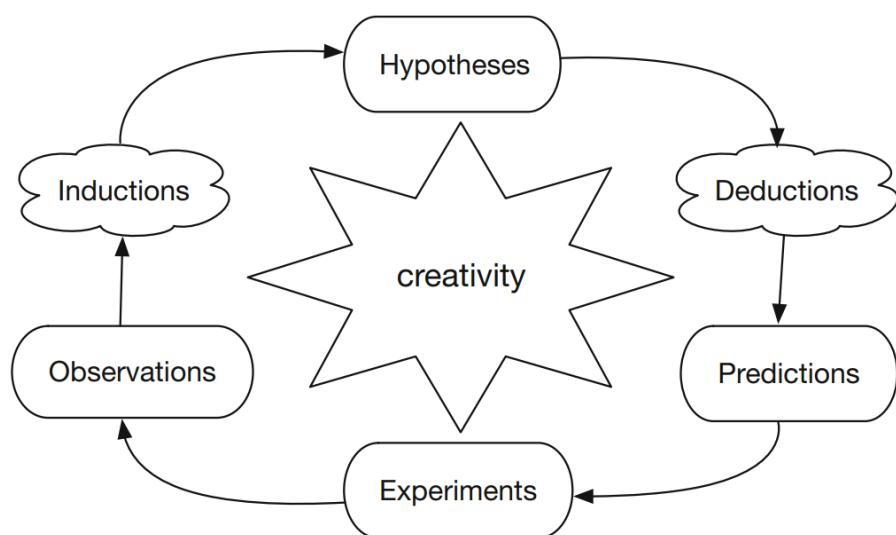


Figura 3 – Método Experimental (GHEZZI, 2020).

Segundo (GHEZZI, 2020), a pesquisa científica pode ser classificada de três formas de acordo com o método. **Formal**, aquela que desenvolve teorias e prova teoremas. Mais comum na matemática, lógica, estatística e em alguns temas na Computação. **Analítica**, aquela que observa e estuda o mundo. Mais comumente encontrada na Física, Química, Biologia e Medicina. A pesquisa experimental é componente chave neste estilo. **Construtiva** aquela que entrega algo concreto. Um novo protótipo ou processo são construídos para mostrar sua usabilidade, funcionamento ou comportamento. Experimentos também são cruciais para este estilo de pesquisa.

No trabalho de (RAMESH et al., 2004) são utilizadas as divisões **Descriativa** e **Avaliativa** em um sentido complementar ao estilo Analítico, e **Formulativo** com o mesmo sentido Construtivista de (GHEZZI, 2020).

No trabalho de (SUDHAKAR et al., 2021) são mostradas as pesquisas Conceitual, Empírica e Experimental. Pesquisa **Conceitual** é aquela caracterizada pela identificação dos construtores elementares na base de conhecimento. Procura explicar a relação destes construtores com fatores e variáveis de um determinado objeto de pesquisa. A pesquisa conceitual é o primeiro passo para a pesquisa empírica e expe-

rimental. Pesquisa **Empírica** é aquela que prova suas hipóteses por meio da coleta e análise estatística de dados. Coleta esta que pode ser feita a partir de questionários, entrevistas ou estudos de caso. Pesquisa **Experimental** é caracterizada como o método de pesquisa mais comum. Em (TEDRE, 2011a) é argumentado que ambas são sinônimos, porém experimental se refere a um nível mais profundo de pesquisa empírica, não só coletando observações mas construindo e testando hipóteses.

No trabalho (KOZIOLEK, 2005) o autor aborda a pesquisa científica como estando dividida entre dois grandes paradigmas metodológicos: Experimental e Analítico. A pesquisa experimental inclui o método científico e suas subdivisões **Engenharia** e **Empírico**. Já a pesquisa analítica compreende o método **Matemático**.

O método científico compreende a observação e a proposição de um modelo de comportamento, a medição, análise e validação de hipóteses do modelo ou teoria estudados (KOZIOLEK, 2005) (HUSTADT, 2012). A repetição é uma premissa do método científico. O autor conceitua o método Engenharia como a observação das soluções já propostas e a construção ou desenvolvimento de soluções melhores e seu refinamento até que nenhuma melhoria seja possível. O método Empírico compreende a proposição de um modelo e sua validação estatística (quantitativa) ou qualitativa. O método Matemático compreende a proposição de uma teoria formal ou um conjunto de axiomas dos quais se derivam resultados podendo ou não serem comparados com observações empíricas.

Corroborando com a mesma ideia dos demais autores (TEDRE, 2007) também divide em três os métodos científicos particularmente usados em Computação: Teórico, Engenharia e Empírico. Na Figura 4, reproduzida de (TEDRE, 2007), o autor conceitua cada uma das divisões bem como ilustra a inter-operacionalidade metodológica. De acordo com o objeto estudado é possível utilizar uma metodologia comum, porém não se aplicam a todos os objetos de pesquisa. Não sendo possível por exemplo, segundo o autor, por meio do método formal evidenciar a qualidade de um produto produzido.

Em (WAINER et al., 2007) são utilizada as definições do trabalho de (TICHY et al., 1995). Pesquisa **Teórica Formal**, aquela que prova sua proposição utilizando uma prova formal ou teorema. **Design e Modelagem**, a pesquisa em que o propósito é desenvolver algo cuja prova não consegue ser de maneira formal. **Empírico**, aquela que coleta dados, analisa e os interpreta. Quer sejam teorias abstratas, modelos ou sistemas onde o foco é a avaliação e não o objeto de pesquisa em si. Finalizando o **Teste de Hipótese**, onde enquadram-se aquelas pesquisas que definem uma hipótese e os experimentos para testá-la. Ainda é utilizada uma quarta classificação nos trabalhos supramencionados, todavia é de natureza genérica visando aquelas pesquisas que não se enquadram nas demais classificações específicas utilizadas. Logo não há lógica em trazê-la para o presente trabalho neste momento.

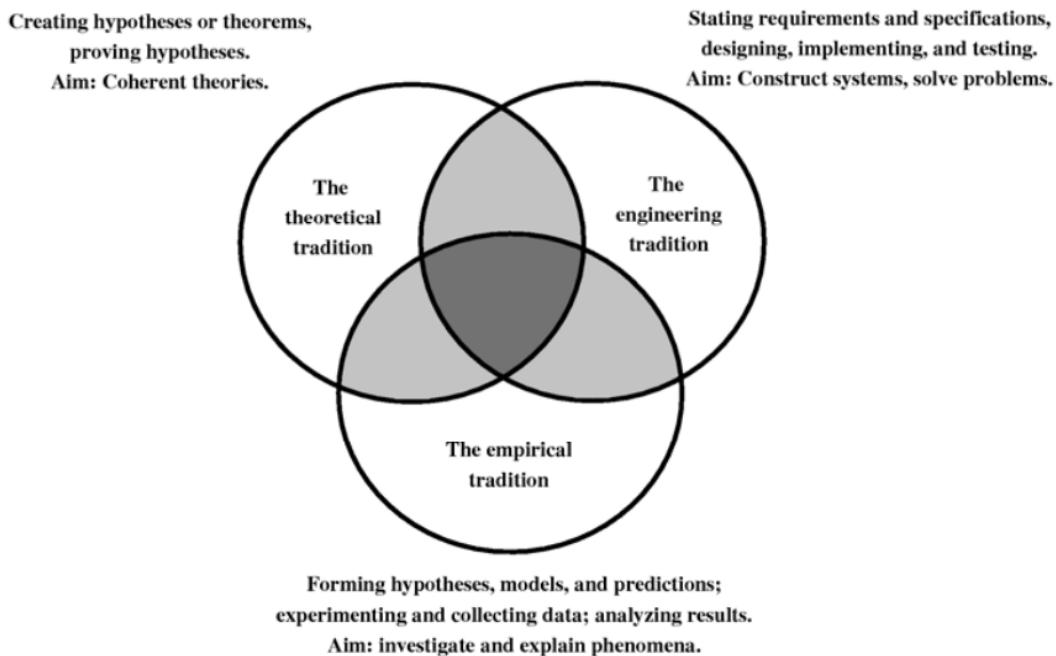


Figura 4 – Intersecção dos Métodos (TEDRE, 2007)

No texto de (HUSTADT, 2012) são descritos alguns tipos de estudos cujas ideias dizem respeito ao método de estudo. **Descritivo** aqueles trabalhos que visam revisar ou avaliar resultados já existentes, incluindo a descrição do estado da arte e a comparação de resultados por meio de experimentos. **Exploratórios** como sendo aqueles estudos que buscam investigar uma nova área ou campo ou problema. **Explanatórios** os estudos que buscam esclarecer os fatos sobre um determinado fenômeno, identificando a relação entre as coisas. **Causal** aquele que avalia a relação entre os objetos de pesquisa. **Normativos** aqueles que visam estabelecer normas, tais regras ou padrões. **Problema-orientado** aqueles estudos que buscam solução para um determinado problema específico.

3.3.4.4 Metodologia

A metodologia é o estudo dos métodos e está intimamente ligada à pesquisa científica. O termo **Protótipo** conforme definição por (SUDHAKAR et al., 2021), refere-se à elaboração de um artefato ainda não existente ou de características inexistentes. Já a **Modelagem** é a representação ou definição de um sistema complexo. Os modelos são úteis para outros métodos científicos (experimental, empírico), não sendo, por vezes, o objeto final da pesquisa, mas sim um meio para atingir esse objetivo. O **Processo**, segundo o autor, é definido como as ações humanas envolvidas na produção de algo. A **Simulação** corresponde à implementação ou execução do objeto de pesquisa em ambientes não reais, por razões de custo, complexidade ou fatores de tempo.

Sob a ótica da origem dos dados, conforme abordado por (KOZIOLEK, 2005) em três metodologias, temos os seguintes aspectos: **Observacionais**, nos quais os dados derivam da observação do objeto de pesquisa e sua subsequente anotação; **Históricos**, nos quais os dados são coletados com base em eventos ou materiais que já aconteceram ou foram publicados; e, finalmente, **Controlados**, nos quais os dados provêm de um meio de coleta preparado para esse fim.

3.3.4.5 Modelos de Prova

Os modelos de prova constituem a base lógica sobre a qual se constrói o conhecimento científico. A metodologia aplicada a tais modelos compreende diversos métodos de raciocínio, como dedução e indução, aplicados no processo de validação e refinamento de hipóteses, e abdução. Esta última consiste em um modelo de prova que se vale das observações sobre os resultados obtidos.

A **Dedução** é utilizada para fazer previsões a respeito de uma hipótese, as quais são validadas por meio de experimentos (GHEZZI, 2020) (HUSTADT, 2012). Estes, por sua vez, geram observações que fornecem evidências para a validação indutiva da hipótese ou o refinamento ou a derivação de novas hipóteses (GHEZZI, 2020)(HUSTADT, 2012). Desse modo fechando o ciclo do método experimental (Figura 3) ou iniciando um novo até que ocorra a aceitação ou refutação completa da hipótese. Ela também pode se basear no conhecimento sobre o mundo para fazer previsões sobre determinado fenômeno ($se\ a = b\ e\ b = c\ ent\ \tilde{a}o\ a = c$).

A **Indução** é um método de raciocínio no qual as observações fornecem evidências para a validade das conclusões. Utiliza evidências empíricas para formular generalizações sobre os resultados (KOZIOLEK, 2005) (HUSTADT, 2012). A dedução é um processo de raciocínio que, a partir de uma ou mais premissas ou teorias, chega a uma conclusão logicamente certa: infere-se a conclusão a partir das premissas. Ambas podem ser formalizadas por meio da lógica matemática (GHEZZI, 2020). Pode-se assim considerar que o raciocínio dedutivo trabalha a partir de premissas gerais para explicar uma observação específica, caso as premissas observadas neste caso específico sejam verdadeiras. Quanto ao raciocínio intuitivo, muitas vezes busca-se obter a conclusões com base em conhecimento prévio ou experiência (PRAJAPATI et al., 2015).

No modelo experimental de acordo com (GHEZZI, 2020) são empregados dois mecanismos de raciocínio: a abstração, que envolve a exclusão de detalhes para se concentrar no fenômeno em estudo; e a generalização, que busca aplicar uma conclusão ao todo por meio de premissas individuais.

A **Abdução** é o método de raciocínio que procede das observações às causas. Útil para a geração de hipóteses, as quais são provadas ou não por observações (HUSTADT, 2012).

Ainda relacionados à prova dos resultados, são arrolados por (GHEZZI, 2020) **Prova e Contra-prova**. A prova mostraria um resultado possível e factível; a contra-prova mostra a impossibilidade de se obter o resultado para o tema estudado. Se a prova é a *verdade*, a contra-prova é a *não verdade*. É a prova pela contradição da verdade.

A relação entre as variáveis de estudo dentro do modelo de prova e a dependência entre elas foi estabelecida por (KOZIOLEK, 2005) como: **Descritiva**, onde o relacionamento entre as variáveis não é considerado; **Correlacional**, em que a variação de variáveis dependentes está relacionada à variação das variáveis independentes; e **Causa-Efeito**, onde as variáveis independentes são a única possível causa de variação das variáveis dependentes.

3.3.4.6 Ambiente de Execução

Os experimentos podem ser classificados como **Reais** ou **Simulados**, sendo estes últimos realizados por meio de algum artefato de software. No caso da simulação, o ambiente e as variáveis são gerados com base em diferentes estratégias para garantir a aleatoriedade e abranger o maior número possível de situações (GHEZZI, 2020). Entretanto, uma desvantagem apontada pelo autor é que nas simulações algumas condições podem ser simplificadas demais tornando-as não realistas. Para contornar essa limitação, é sugerido realizar experimentos que combinem partes simuladas e partes reais.

Seguindo a mesma linha, (KOZIOLEK, 2005) menciona **In Vivo** para experimentos realizados em campo, em condições realísticas, e **In Vitro** para aqueles conduzidos em laboratório, sob condições controladas.

3.3.4.7 Natureza dos Dados

A natureza da pesquisa em relação aos dados e sua análise pode assumir dois aspectos: **Qualitativo** e **Quantitativo**. O Qualitativo refere-se a resultados cuja análise é subjetiva (GHEZZI, 2020), buscando entender as razões por trás de determinado acontecimento (KOZIOLEK, 2005) ou aumentar o conhecimento sobre um fenômeno ao invés de explicá-lo (HUSTADT, 2012). Além disso, qualitativo abrange a análise das características do objeto de estudo quando as variáveis são desconhecidas, podendo-se mesmo utilizar a abordagem Qualitativa para descobri-las (PRAJAPATI et al., 2015), considerando as observações realizadas.

Os dados quantitativos referem-se ao que é mensurado e analisado objetivamente, sem subjetividades (KOZIOLEK, 2005), quando a precisão e a repetibilidade são fatores críticos (HUSTADT, 2012). Essa abordagem está frequentemente associada a testes de hipóteses e ao raciocínio indutivo (HUSTADT, 2012).

Para ambas as divisões (PSCHEIDT, 2012) destaca a necessidade de seguir cri-

térios e orientações de análise. Embora seja possível realizar uma análise mista em uma abordagem “quanti/quali” (GHEZZI, 2020; PRAJAPATI et al., 2015).

No que diz respeito à natureza dos dados, mais especificamente à forma como são coletados e, por consequência analisados, as divisões **Horizontal** e **Vertical** são destacadas no trabalho de (PSCHEIDT, 2012). Ambas se referem às características de nível e temporalidade dos dados. A abordagem Horizontal trata de uma sequência de acontecimentos ao longo do tempo. Enquanto a Vertical está relacionada ao nível do contexto sobre o objeto estudado contido nos dados. Em Computação, segundo o autor, a melhor abordagem é aquela que estabelece a melhor relação entre as duas.

3.3.4.8 Validação / Coleta de dados

(GHEZZI, 2020) lista estratégias de validação de resultados. Estudos **Observacionais** são aqueles que permitem verificar o comportamento do objeto de estudo em cenários reais ou simulados. Os estudos **Piloto** permitem demonstrar o funcionamento do objeto de estudo. Os estudos **Comparativos** comparam o objeto de estudo com pares semelhantes. Na área da Computação apontam-se ainda os estudos **Comunitários** ou **Benchmarks**. Nestes estudos, a comunidade científica trabalha colaborativa ou competitivamente em um mesmo objeto de pesquisa. Ainda há o caso em que estudos independentes aplicam as mesmas entradas de dados (*benchmarks*) para aferir diferenças nos resultados obtidos.

A pesquisa científica, como uma estratégia específica de validação, é representada pelos estudos **Reprodutivos** ou **Replicativos**. Nestes tipos de estudo, busca-se reproduzir os experimentos nas mesmas condições para obter novos resultados. Isso confere força e validade aos resultados em ambas as pesquisas.

Outra estratégia mencionada por (GHEZZI, 2020) são os estudos **Controlados** ou **Caso Controle**. Nessa abordagem, a amostra é dividida em dois grupos: experimental e controle. O grupo experimental é exposto ao objeto de estudo. Ao contrário do grupo controle. O objetivo é a verificação das relações de causa e efeito entre os dois grupos.

(GHEZZI, 2020) enfatiza a importância de os resultados terem validade *Interna* e *Externa*. A Validade Interna refere-se à representação adequada dos resultados em relação ao objeto de estudo. Enquanto a Validade Externa diz respeito à generalização dos resultados para outras situações.

No trabalho de (RAMESH et al., 2004), uma variedade de estratégias de validação de dados foi utilizada para caracterizar a pesquisa em Computação. Algumas dessas estratégias incluem **Pesquisa Ação**, **Análise Conceitual**, **Prova de Conceito**, **Estudo de Caso**, **Análise de Dados**, **Análise de Discurso**, **Etnografia**, **Experimento de Campo**, **Estudo de Campo**, **Teoria fundamentada**, **Hermenêutica**, **Desenvolvimento de Instrumentos**, **Experimento de Laboratório**, **Revisão de Literatura**, **Experiência de Laboratório (software)**, **Meta-análise**, **Prova matemática**, **Análise**

de Protocolo, Fenomenologia, Simulação e Survey, sem uma caracterização individual.

Com relação à coleta de dados, (KOZIOLEK, 2005) cita sem caracterizá-las as divisões como Observacionais (**Monitoramento de Projeto, Estudo de Caso e Estudo de Campo**), Históricos (**Revisão de Literatura, Dado Legado, Lesson-Learned e Análise Estática**), e Controlados (**Experimento de Ambiente Sintético, Análise Dinâmica e Simulação**).

Em (PSCHEIDT, 2012), sem caracterizá-los na totalidade, **Dados Secundários, Survey, Estudo de Campo, Pesquisa Ação, Experimento de Campo, Etnografia, Teoria Embasada e Estudo de Caso**. Pesquisa ação é a integração da teoria e prática para oferecer uma solução prática a um problema real. Etnografia caracteriza-se pela imersão no meio para estudar e explicar o fenômeno estudado em seu contexto (social e cultural). Teoria embasada busca desenvolver uma teoria com base em dados empíricos coletados e analisados sistematicamente, buscando por padrões que sustentem a elaboração da teoria. O estudo de caso, embora tenha como objetivo objetos de pesquisa da vida real, propõe uma das possíveis soluções ou tem escopo limitado a apenas um aspecto do objeto de estudo. Não tem a intenção de propor mudanças no contexto.

3.3.4.9 *Tipos de Validação dos Resultados*

Sob a perspectiva da validação dos resultados, ou seja, do que foi medido, e o que se pretende com essas medições, (TEDRE, 2007) menciona alguns critérios. **Operacionalidade** refere-se à simples capacidade de construir o objeto de pesquisa. **Usabilidade** avalia o quanto fácil é utilizar o objeto de pesquisa. **Efetividade** mede a capacidade do objeto de pesquisa em resolver o problema ao qual se propõe. **Custo-Efetividade** analisa o custo de desenvolvimento do objeto de pesquisa em relação à sua efetividade. **Eficiência** avalia a capacidade do objeto de pesquisa em resolver o problema da melhor maneira possível.

3.4 Discussão

A seguir são apresentados os resultados da Revisão Sistemática da Literatura (RSL) realizada. Em destaque nesta discussão encontram-se observações sobre as divisões da pesquisa científica em Computação identificadas. Os aspectos, características e conceitos, a inserção da Computação na Ciência, e as subdivisões da Computação como grande área do conhecimento. Esses tópicos são abordados em resposta às questões apresentadas na Subseção 3.1.1.

3.4.1 Respondendo às Questões de Pesquisa

No início desta revisão sistemática da literatura, na Seção 3.1.1, foram elaborados quatro questionamentos os quais serão respondidos nesta seção utilizando os resultados sistematizados.

A chave de pesquisa utilizada, apresentada na página 30, assegura uma lógica forte na associação do termo principal, “computer science”, aos termos secundários (aqueles que estão encadeados entre parênteses). Isso significa que o resultado lógico será verdadeiro apenas se o termo principal estiver associado a pelo menos um termo secundário. Isso garante que os resultados sejam todos pertinentes à Computação e estejam relacionados ao meta-estudo da pesquisa científica.

Q1 - Existe alguma divisão explicativa para a pesquisa científica em Computação, considerando tipos de estudos, objetos de pesquisa e métodos avaliativos aplicáveis à área, independentemente da denominação utilizada, seja esta classificação, divisão ou taxonomia?

A Tabela 4 sistematiza as divisões, classificações ou categorias da pesquisa científica em Computação encontradas nesta RSL. As divisões foram agrupadas segundo a conceituação dada pelos autores, assim como a afinidade das abordagens dadas por diferentes autores para a mesma divisão e conceito.

Para responder à Questão 1 (Q1), primeiro é necessário estabelecer dois aspectos. O primeiro é em relação à existência propriamente dita. Sob esse aspecto o próprio fato da Tabela 4 estar inserida neste trabalho indica que a resposta é sim. Tais informações sobre a pesquisa científica em Computação existem. No entanto, é preciso analisar outro aspecto: a questão da completude ou da sistematização dessas informações de forma centralizada. Sob este aspecto, a resposta é não. Não existe uma publicação única que contenha todas essas informações, ou um documento que aborde todos esses conceitos. Embora abrangentes, os trabalhos encontrados não abordam a totalidade dos conceitos identificados.

Ao analisar os dados da Tabela 4, é possível perceber que vários autores abordam a pesquisa científica em Computação como objeto de estudo. Alguns de maneira mais abrangente dentro do tema e outros de maneira mais superficial. Nota-se que não há um único trabalho descrevendo todas essas divisões e conceitos (conforme coluna “Referência” da Tabela 4), bem como não há uma uniformidade no que diz respeito a esses conceitos (conforme Seção 3.3.4). Também é importante destacar que dentro de uma mesma categoria existem autores cujos conceitos e nomenclaturas são convergentes, embora tenha sido verificado diferentes perspectivas destes autores sobre a mesma categoria de conceitos.

Q2 - Quais são os aspectos, características e/ou conceitos a serem considerados para se montar uma classificação ou divisão da pesquisa científica em Computação visando sua melhor compreensão?

Tabela 4 – Divisões e categorias da pesquisa científica em Computação segundo a RSL

Categoria	Divisões	Referências
Natureza	Pura e Aplicada	(GHEZZI, 2020) (HASSANI, 2017) (HUSTADT, 2012)
Propósito/Paradigma	Comportamental, Interpretativa, Crítica, <i>Design Science</i>	(PSCHEIDT, 2012)
Método	Formal, Analítico e Construtivo, Teórica, Experimental	(HARTMANIS, 1993) (GHEZZI, 2020) (TEDRE, 2007)
	Conceitual, Empírico, Experimental	(SUDHAKAR et al., 2021)
	Descriptivo, Avaliativo, Formulativo	(RAMESH et al., 2004)
	Científico, Engenharia, Empírico, Matemático	(KOZIOLEK, 2005) (TEDRE, 2007)
	Formal, <i>Design</i> , Empírico, Teste de Hipóteses	(WAINER et al., 2009)
	Descriptivo, Exploratório, Explanatório, Causal, Normativo, Problema-Orientado	(HUSTADT, 2012)
Metodologia	Protótipo, Modelagem, Processo, Simulação	(SUDHAKAR et al., 2021)
	Observacional, Histórico, Modelo Controlado	(KOZIOLEK, 2005)
Prova científica	indução, dedução	(GHEZZI, 2020) (KOZIOLEK, 2005) (PRAJAPATI et al., 2015) (HUSTADT, 2012)
	Abdução, Indução, Dedução	(HUSTADT, 2012)
Prova científica secundária	Prova, Contra-prova	(GHEZZI, 2020)
Ambiente	Real, Simulado	(GHEZZI, 2020) (HARTMANIS, 1993) (HAZZAN et al., 2006) (SUDHAKAR et al., 2021)
	In Vivo, In Vitro	(KOZIOLEK, 2005)
Dados	qualitativo, quantitativo	(GHEZZI, 2020) (KOZIOLEK, 2005) (PRAJAPATI et al., 2015) (PSCHEIDT, 2012) (HUSTADT, 2012)
	Horizontais, verticais	(PSCHEIDT, 2012)
Estratégia de Validação / Coleta de Dados	Observacional, piloto, comparativo, comunitário, replicação, caso/controle	(GHEZZI, 2020)
	Pesquisa Ação, Análise Conceitual, Prova de Conceito, Estudo de Caso, Análise de Dados, Análise de Discurso, Etnografia, Experimento de campo, Estudo de campo, Teoria fundamentada, Hermenêutica, Desenvolvimento de instrumentos, Experimento de laboratório, Revisão de literatura, Experiência de laboratório (software), Meta-análise, Prova matemática, Análise de protocolo, Fenomenologia, Simulação e Survey	(RAMESH et al., 2004)
	Monitoramento de Projeto, Estudo de Caso e Estudo de Campo, Revisão de Literatura, Dado Legado, <i>Lesson-Learned</i> , Análise estética, Experimento de Ambiente Sintético, Análise Dinâmica, Simulação	(KOZIOLEK, 2005)
	Narrativa, Etnográfico	(TEDRE, 2007)
	Dados Secundários, <i>Survey</i> , Estudo de Campo, Pesquisa Ação, Experimento de Campo, Etnográfica, Teoria embasada, Estudo de Caso	(PSCHEIDT, 2012)
	Pesquisa-Ação, Estudo de caso, <i>Survey</i> , Experimento	(HUSTADT, 2012)
Tipo de Validação do Resultado	Operacionalidade, Usabilidade, Efetividade, Custo-Efetividade, Eficiência	(TEDRE, 2007)
Objetos de estudo	Teoremas, Algoritmos	(GHEZZI, 2020) (HASSANI, 2017) (TEDRE, 2007)
	Processos	(KOZIOLEK, 2005)
	Modelos, Algoritmos	(TEDRE, 2007)
	Algoritmo, Abordagem, Técnica, Métrica, Arquitetura, Protocolo, Estrutura de Dados, Conceito	(PRAJAPATI et al., 2015)

Tabela 5 – Categorias da pesquisa científica em Computação segundo a RSL

Categoria	Citações
Natureza	3
Propósito/Paradigma	1
Método	9
Metodologia	2
Prova científica	5
Prova científica secundária	1
Ambiente	4
Natureza dos Dados	6
Estratégia de Validação / Coleta de Dados	6
Tipo de Validação do Resultado	1

Retiradas da Tabela 4 e listadas separadamente na Tabela 5, estão as categorias encontradas nesta RSL. Elas representam os aspectos da pesquisa científica segundo cada trabalho selecionado. Em alguns deles o autor é explícito ao enumerar essas categorias em outros o autor não o faz mas deixa clara a ideia a que ela se refere.

Enquanto categorias da pesquisa científica em Computação, Natureza refere-se às características essenciais do conhecimento: como o conhecimento é produzido e qual sua aplicabilidade. Em relação à Natureza, a pesquisa pode vir a representar a criação de algo novo ou a produção de uma abordagem inovadora sobre algo que já existe. Já a categoria Propósito, ou Paradigma, compreende o objetivo do conhecimento produzido, seja este objetivo do conhecimento direcionado à verdade, à compreensão ou ainda à crítica do objeto de estudo. Método representa o conjunto de ações e a sistematização ordenada destas utilizadas a produção do conhecimento. Metodologia, por sua vez, compreende a forma como a pesquisa é desenvolvida visando a coleta de dados, ou seja, aquilo que é desenvolvido dentro da pesquisa e se sobre os quais se coletam dados para gerar conhecimento. Modelos de Prova e Modelos de Prova Secundária representam a maneira como os resultados mostram que o conhecimento gerado é verdadeiro ou válido. Os Modelos de Prova podem propor prova direta ou prova pela negação da verdade dos resultados. Ambiente é o aspecto da pesquisa científica que representa a situação em que os dados serão coletados para um objeto de pesquisa desenvolvido. A Natureza dos Dados representa a diferença na forma e na temporalidade em que são coletados. O Tipo de Validação do Resultado refere-se ao aspecto do objeto de estudo ao qual o conhecimento gerado se refere, incluindo a criação do objeto de estudo, seu uso, experiência de uso e custos.

A categoria estratégia de Validação ou Coleta de Dados é a maior em termos de número e mais complexa categoria encontrada na RSL. Essa complexidade decorre do fato de que essa estratégia determina a abordagem da pesquisa científica em relação a outras categorias mencionadas anteriormente abrangendo amplamente quase

todas as etapas, especialmente a coleta e validação dos dados.

Por exemplo a realização da reprodução de uma pesquisa em busca da verdade sobre o conhecimento gerado sobre o objeto de pesquisa conduzida anteriormente. Ao definir essa diretriz de estudo replicativo para a pesquisa, aceita-se toda uma estrutura já determinada para condução da reprodução da pesquisa. A pesquisa desenvolvida terá, assim, a mesma Natureza, Paradigma, Método, Metodologia e demais características da pesquisa original.

Por exemplo um estudo replicativo, que reproduz uma pesquisa realizada anteriormente em busca da verdade sobre o conhecimento gerado sobre o objeto de pesquisa. No momento em que se define essa diretriz para a pesquisa, se está aceitando todo uma estrutura já determinada. A pesquisa desenvolvida terá a mesma natureza, paradigma, método, metodologia e assim por diante.

Aponta-se ainda a estratégia ou categoria *Survey*. Na literatura, assim como no meio acadêmico internacional, esta estratégia está associada a aplicação de entrevistas por meio de questionários. No entendimento a nível Brasil, esta estratégia pode ser confundida com um tipo de estudo bibliográfico não sistematizado visando levantar informações sobre o objeto de pesquisa.

Q3 - Qual a inserção da Computação dentro da Ciência e onde esta acontece dentro de uma análise cartesiana multi eixo da Computação face as divisões da Ciência? Sendo de um lado a Computação como fonte geradora de ciência de outro ela como ferramenta para tal.

Atualmente é impossível imaginar pesquisa, independentemente da área, que não faça uso da Computação. Alguns trabalhos revisados argumentam sobre a multidisciplinaridade da Computação e da sua inserção em outras áreas (GHEZZI, 2020; TEDRE, 2007, 2011a; HASSANI, 2017). Uma vez inserida em outras áreas, surge a inevitável pergunta: a pesquisa em Computação existe por si só, ou ela é apenas um meio para outras áreas? Qual é a fronteira efetiva entre a área fim e a área meio?

Há argumentos que se referem à Computação como sendo uma ciência tão especial que compreenderia um novo e quarto domínio da Ciência, ao lado das Ciências da Física, Vida e Sociais (TEDRE, 2011a; DODIG-CRNKOVIC, 2023). Da mesma forma, em comparação com as demais, a Computação é uma Ciência recente ainda em desenvolvimento e que tem propiciado mudanças significativas na sociedade (GHEZZI, 2020). Todavia este não é o foco do presente trabalho e tais argumentos são citados apenas para caracterizar o quão complexa é a tarefa de estudar a própria ciência em Computação.

Três eixos de estudo da Computação são evidenciados nesta RSL conforme a Figura 5: a Computação como meio, a Computação como fim e o próprio estudo da ciência em Computação. No entanto, não foi encontrada uma classificação ou divisão que contemple esses eixos de estudo. Devido a esse fato, julga-se necessário uma

divisão que classifique essas inserções.

A Computação como meio é aquela que está inserida em outras áreas de pesquisa, as quais se utilizam da mesma para obter seus resultados.⁴ Como fim, é aquela que desenvolve a técnica computacional e a valida aplicando o método científico. Como próprio objeto de estudo ou metaciência, é aquela que faz da ciência da Computação seu objeto de estudo, desenvolvendo e aprimorando o método científico e estudando as relações entre os três eixos.

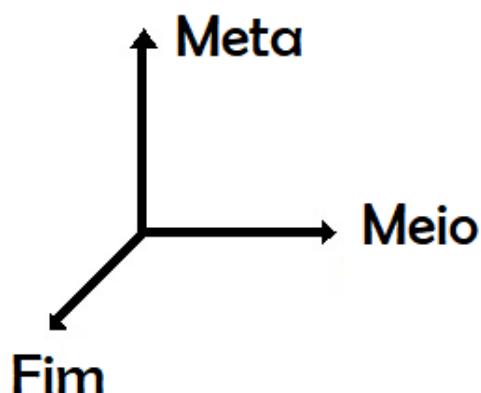


Figura 5 – Eixos de estudo da Ciência em Computação

Dentro do cenário representado pela Figura 5 é possível visualizar que a pesquisa em Computação existe por si só como uma disciplina única capaz de gerar conhecimento no momento em que se assume valores próximos a zero para os eixos Meio e Meta. A variação destes determinará o quão multidisciplinar será a Computação empregada na pesquisa.

Ainda cabe destacar que se visualiza um possível ciclo envolvendo os três eixos. Dentro do conceito de refinamento evolutivo (PRAJAPATI et al., 2015; HASSANI, 2017; GHEZZI, 2020), a Computação, ao ser empregada como meio inovador, demanda resultados inovadores provenientes da Computação como fim. Esses, por sua vez, requerem novos métodos ou esclarecimentos sobre os já existentes para serem aplicáveis, o que pode ser alcançado por meio de estudos metacentíficos dentro da área de Ciência da Computação.

Respondendo à Questão 3 (Q3), a Computação está plenamente inserida na Ciência, por vezes de maneira ubíqua, por vezes de forma mais evidente. A Ciência a partir dela, nela própria e dela mesma. É ao mesmo tempo fonte geradora de conhecimento, ferramenta e objeto de estudo.

Q4 - Quais as divisões de estudo atuais da Computação como área de conhecimento?

⁴Nas discussões apresentadas neste texto não encontram-se levadas em consideração a microinformática, ou informática de escritório.

mento, as quais devem ser consideradas na montagem de um panorama abrangente da área?

A divisão criada pelo ERC e utilizada em (GHEZZI, 2020) foi elaborada com o propósito de auxiliar no planejamento e na avaliação de propostas de financiamento. Ela faz parte de um documento⁵ contendo 25 subdivisões dos grandes grupos das Ciências e Engenharia. Nessa classificação, a Computação está inserida nas Ciências da Física e Engenharias. Todavia são facilmente perceptíveis os três eixos de estudo da ciência em Computação citados anteriormente. A Bioinformática, por exemplo, possui uma classificação própria como área nas Ciências da Vida e está também inserida como uma das subáreas de estudo dentro da Computação, sugerindo que toda a classificação da Bioinformática está englobada na Computação. De maneira semelhante o mesmo acontece com o estudo de DNA e Genética.

Os SIG utilizados em (HOLZ et al., 2006) são grupos criados e mantidos pela ACM para representar as principais áreas de pesquisa, ensino e pensamento, bem como áreas emergentes dentro da Computação. Seu propósito não é estudar a Ciência em Computação, mas estabelecer critérios para agrupar pessoas e periódicos mantidos pela ACM e proporcionar vantagens aos seus associados e participantes destes grupos de interesse. Não obrigatoriamente, têm a necessidade de cobrir todas as áreas dentro da Computação, pois parte do interesse (ou não) de um grupo de pessoas em trabalhar determinado tema de pesquisa.

Mencionados em (RAMESH et al., 2004), os CCS são também de iniciativa da ACM, porém o propósito é diferente. A principal finalidade é servir como uma classificação definitiva para a área da Computação⁶. A editora estimula fortemente os autores a categorizarem suas publicações segundo esta classificação. O argumento é o benefício que a publicação terá em função de que uma vez categorizada de maneira precisa ela fornece ao leitor uma referência rápida do conteúdo. Facilitando a busca de literatura relacionada, bem como a busca online.

Finalizando e respondendo à Questão 4 (Q4), a iniciativa que melhor atende ao que se espera considerar na formação de um panorama da Computação como grande área do conhecimento é a CCS da ACM. As demais, mesmo sendo agrupamentos da área, não devem ser desconsideradas pois representam os nichos de maior interesse dentro da Computação. Cabe destacar que no Brasil, de iniciativa da SBC, há uma separação de interesse da Computação. Inicialmente em Grandes Áreas (GA) e estas divididas em Comissões Especiais (CE).

⁵https://erc.europa.eu/sites/default/files/document/file/ERC_Panel_structure_2020.pdf

⁶<https://dl.acm.org/ccs08/03/2022>

3.5 Ameaças à Validade

A revisão sistemática apresentada nesta seção está sujeita a ameaças à sua validade como qualquer outro trabalho científico. Dois possíveis problemas de validade se destacam aqui, ambos relacionados a vieses de seleção considerando duas diferentes perspectivas.

O primeiro refere-se aos termos utilizados como chave de pesquisa e as bases de dados utilizadas na pesquisa bibliográfica. Embora a *string* de busca tenha sido concebida para serem abrangente na identificação de trabalhos de interesse à pesquisa realizada, pode acontecer que alguma referência não tenha sido encontrada devido a uma palavra-chave diferente ter sido utilizada pelo autor ou por ter sido publicada em uma base indexada diferente das utilizadas na RSL. Entende-se que as bases indexadas escolhidas correspondem as mais representativas na área e com maior número de periódicos e publicações.

O segundo aspecto é que, embora de maneira isenta, apenas um revisor fez a leitura e seleção dos artigos, tendo como supervisor o orientador desta tese. Neste item, ainda podem existir referências que poderiam ser anexadas de acordo com a interpretação de um segundo revisor. No entanto, alguns cuidados foram tomados para tentar evitar esse problema. Quando havia dúvida em relação a uma publicação, a mesma era mantida na lista das selecionadas para uma nova avaliação em uma etapa posterior da RSL. Caso, após a leitura na íntegra, ainda restasse dúvida sobre a possível seleção ou não, o trabalho era mantido na lista para uma segunda leitura em momento distinto e não subsequente à primeira.

3.6 Considerações Finais do Capítulo

Como foi evidenciado neste capítulo, não há uma taxonomia de pesquisa formal para a área da Computação. Vários autores abordam o tema, mas o fazem segundo diferentes óticas e com conceituação diferente. Tais conceitos ficam dispersos entre vários trabalhos científicos. A compilação destas ideias mostra um panorama bastante rico e abrangente sobre a pesquisa em Computação, e ainda com a praticidade de estar em um único lugar. Também é evidenciada as diferentes perspectivas que os objetos de pesquisa em Computação podem assumir dentro da Ciência como um todo. Nos capítulos que seguem, essas informações são sistematizadas.

4 TAXONOMIA DE PESQUISA EM COMPUTAÇÃO

Como exercício de montagem da taxonomia, a Tabela 6 foi elaborada para sintetizar as categorias, divisões e conceitos encontrados na RSL. É possível perceber que para algumas divisões da pesquisa a conceituação é basicamente a mesma, mudando apenas o descritor da classificação, assim como a incorporação do conceito de uma divisão em outra. Nestes casos, é preciso simplificar a classificação, eliminando as duplicidades e ambiguidades, deixando apenas o conceito principal para os casos em questão.

Os grupos menores em termos de número de divisões de classificação não apresentaram maiores problemas no que diz respeito à duplicidade de conceitos. Também não foram visualizados problemas no que diz respeito à falta de classificação, todavia, o processo de validação é onde será confirmada esta observação preliminar.

No grupo Método, as classificações (i) Empírico, Experimental e Teste de Hipóteses, (ii) Construtivo e Design, (iii) Formal e Teórica, (iv) Causal e Explanatório referem-se individualmente aos mesmos conceitos. Na taxonomia, permanecerão “Experimental”, “Construtivo” e “Causal”.

A classificação Empírica não consta na taxonomia categorizada na Tabela 6, uma vez que esta já encontra-se compreendida na pesquisa Experimental e na Indução. Francis Bacon (BACON, 1878) propõe o método Empírico, Experimental e Indutivo; da mesma forma que René Descartes (DESCARTES; CRESS, 1998) o faz com o Racionalismo, Matemático, Lógico e Dedutivo. Todavia, este último não aparece na RSL, somente sua conceituação. Neste sentido, é metodologicamente mais correto suprimir “empírico” do que incluir o Racionalismo.

No grupo Estratégia de Validação ou Coleta de Dados, os conceitos das classificações Experiência de laboratório (software), Experimento, Experimento de Ambiente Sintético, Experimento de Campo, estudo de Campo e Simulação são basicamente os mesmos, mudando apenas o ambiente de execução do objeto de pesquisa. Existe uma divisão na taxonomia que utiliza o ambiente de execução como pivô para classificação. Não justificando ter anexa à Estratégia de Validação. Desta forma, apenas uma classificação “experimento” permanecerá.

De maneira semelhante, as classificações Teoria Embasada e Teoria Fundamentada, Dado Legado e Dados Secundários, Desenvolvimento de Instrumentos e Survey referem-se ao mesmo conceito individualmente. Permanecerão as classificações “Teoria Fundamentada”, “Dados Secundários” e “Aplicação de Instrumento” respectivamente. Sendo esta última alterada, em função da não representatividade das nomenclaturas encontradas a ideia a que elas representam. Assim como o fator de confusão com o conceito de Survey, já mencionado anteriormente neste trabalho.

Ao se pensar no conceito de Dados Secundários, invariavelmente se pergunta sobre os dados primários e a necessidade de ter-se uma classificação que o contemple dentro do grupo Estratégia de Validação ou Coleta de Dados. Analisando as demais classificações deste grupo presentes na Tabela 6, todas elas se referem à geração de dados primários a partir do objeto de estudo, não justificando assim a inserção na taxonomia desta classificação em específico.

Os conceitos de Piloto e Prova de Conceito, ainda que não sejam especificamente os mesmos (inclusive na grafia), eles são muito parecidos e um está inserido dentro do outro. A ideia de “testar o objeto de pesquisa ou metodologia” é muito mais ampla do que simplesmente “testar a viabilidade”. Ao se visualizar a taxonomia como um todo, há um grupo que abrange os tipos de validação sobre o objeto de pesquisa e seu resultado e estes vão além do simples teste de viabilidade. Por esta razão, permanecerá apenas Piloto, sendo seu conceito alterado para “Aquela cuja função é testar o objeto de pesquisa ou metodologia e sua factibilidade”.

Ainda dentro do mesmo grupo, as classificações Etnográfica, Lesson-Learned, Narrativa e Observacional possuem uma mesma visão sobre o objeto de estudo. A de observar e descrever o objeto de estudo, seu uso e a maneira como se relaciona com o meio de execução qualquer que seja este. Deste modo permanecerá apenas a classificação “Observacional” e sua conceituação será alterada de maneira a incorporar os diferentes olhares sobre o objeto de estudo. O novo conceito para “Observacional” será: “Aquela cujos dados são obtidos a partir da observação, uso do objeto de estudo, o resultado deste e sua relação com o meio”.

Tabela 6 – Divisões da Pesquisa em Computação

Grupo	Classificação	Conceito
Natureza	Pura Aplicada	A que cria o conhecimento sem um propósito específico A que aplica o conhecimento
Propósito paradigma	Comportamental Interpretativa Crítica Design Science	Aquela que busca encontrar a verdade Aquela que procura compreender a realidade Aquela que questiona a realidade e as diferenças sociais Aquela que está focada no desenvolvimento de processos e produtos
Método	Analítico Avaliativo Causal Científico Conceitual Construtivo Descriptivo Engenharia Experimental Exploratório Formal Formulativo Matemático Normativo Problema-Orientado Teórica	Aquela que observa e analisa Aquela que observa e avalia Aquela que avalia a relação entre as coisas Aquela que propõe um modelo e o mede, analisa e valida Aquela que busca a base elementar do conhecimento Aquela que entrega algo concreto Aquela que observa, ou revisa já existentes e descreve Aquela que propõe uma versão melhorada daquilo que já existe Aquela que formula hipóteses e as prova estatisticamente com a coleta de dados Aquela que investiga uma nova área ou problema levantando informações sobre eles Aquela que desenvolve teorias e prova teoremas Aquela que desenvolve algo Aquela que propõe teorias formais e axiomas Aquela que estabelece normas, padrões ou regras. Aquela que estuda ou busca solucionar um problema em específico Aquela que não entrega algo concreto
Metodologia	Protótipo Modelagem Processo Simulação Observacional Histórico Modelo Controlado	Aquela que cria algo ainda não existente ou com características não existentes Aquela que mostra uma representação de algo complexo Aquela que determina ações necessárias à produção de um resultado Aquela que é executada ou desenvolvida em um ambiente não real Aquela que é executada com base apenas na observação do objeto de estudo Aquela que é executada levando em consideração aquilo que já foi publicado Aquela que é executada em um ambiente em que se tem o total controle
Prova Científica	Abdução Indução Dedução	Aquela que a prova é baseada em uma provável melhor explicação do resultado Aquela que as observações fornecem validade à prova Aquela que a prova é baseada em previsões das observações
Prova Científica Secundária	Contra-Prova Prova	Aquela que mostra a impossibilidade de se obter um resultado (a não verdade) Aquela que mostra a viabilidade de se obter um resultado (a verdade)
Ambiente	In Vivo In Vitro Real Simulado	Aquela realizada em condições realistas a campo Aquela realizada em condições controladas em laboratório Aquela realizada em ambiente e condições reais Aquela realizada em ambiente e condições sintéticas
Dados	Qualitativo Quantitativo Horizontais Verticais	Aquela em que os dados podem ser expressos numericamente Aquela em que os dados não podem ser expressos numericamente Aquela que analisa o objeto de estudo ao longo da linha temporal Aquela que analisa o contexto do objeto de estudo em um ponto da linha temporal

Tabela 6 – Divisões da Pesquisa em Computação (continuação)

Grupo	Classificação	Conceito
Estratégia de Validação/Coleta de Dados	Análise Estática	Aquela que analisa o código fonte ao invés de sua execução
	Análise Conceitual	Aquela que analisa atributos essenciais para formação de uma ideia
	Análise de Discurso	Aquela que analisa a construção de uma produção textual
	Análise de protocolo	Aquela cujos dados são coletados a partir das etapas das possíveis soluções para um dado objeto de pesquisa e a partir deles propor um resultado final
	Análise Dinâmica	Aquela que analisa o código fonte durante a execução
	Caso-Controle	Aquela em que duas amostras são analisadas, sendo uma exposta ao objeto de estudo e outra não
	Comparativo	Aquela cujo objetivo é comparar o objeto de pesquisa com seus pares
	Comunitário	Aquela em que a comunidade científica trabalha no mesmo objeto de pesquisa ou com o mesmo conjunto de dados
	Dado Legado	Aquela que utiliza dados já coletados anteriormente por outras pesquisas
	Dados Secundários	Aquela que utiliza dados já coletados anteriormente por outras pesquisas
	Aplicação de instrumento	Aquela cujos dados são provenientes da aplicação de instrumento de coleta de dados
	Estudo de campo	Aquela cujos dados coletados no real ambiente do objeto de estudo
	Estudo de Caso	Aquela que propõe uma solução ou nuance dentre as possíveis para o mesmo objeto de estudo
	Etnográfica	Aquela que busca dentro do meio dados para explicar o objeto estudado
	Experiência de laboratório (software)	Aquela cujos dados são obtidos por meio execução do objeto de estudo em ambiente não real
	Experimento	Aquela cujos dados são obtidos por meio execução do objeto de estudo
	Experimento de Ambiente Sintético	Aquela cujos dados são obtidos por meio execução do objeto de estudo em ambiente não real
	Experimento de Campo	Aquela cujos dados são obtidos por meio execução do objeto de estudo em ambiente real
	Fenomenologia	Aquela cujos dados são obtidos pela percepção a um determinado fenômeno
	Hermenêutica	Aquela cujos dados são obtidos pela interpretação de textos
	Lesson-Learned	Aquela cujos dados são obtidos do aprendizado em relação ao uso do objeto de estudo
	Meta-análise	Aquela cujos dados são obtidos dos resultados de estudos
	Monitoramento de Projeto	Aquela cujos dados são obtidos do processo de implementação do objeto de estudo
	Narrativa	Aquela que fornece uma descrição detalhada sobre a experiência com o objeto de estudo
	Observacional	Aquele que os dados são obtidos pela observação do objeto de estudo
	Pesquisa Ação	Aquela que une teoria e prática e produz algo durante o andamento da pesquisa
Tipo de Validação do Resultado	Piloto	Aquela cuja função é testar o objeto de pesquisa ou metodologia
	Prova de Conceito	Aquela que analisa a viabilidade do objeto de estudo
	Prova matemática	Aquela que analisa um modelo ou teorema matemático
	Replicação	Aquela que reproduz a metodologia de outra pesquisa
	Revisão de Literatura	Aquela cujos dados são obtidos na literatura de maneira sistemática
	Simulação	Aquela cujos dados são obtidos por meio execução do objeto de estudo em ambiente não real
	Survey	Aquela cujos dados são provenientes da aplicação de instrumento de coleta de dados
	Teoria embasada	Aquela que busca padrões de comportamento nos dados para fundamentar uma teoria
	Teoria fundamentada	Aquela que busca padrões de comportamento nos dados para fundamentar uma teoria
	Efetividade	Aquela que testa o correto funcionamento do objeto de estudo
	Eficiência	Aquela que testa que além de funcionar, o objeto de estudo o faz de maneira melhor
	Custo-Efetividade	Aquela que testa a efetividade do objeto de estudo em relação aos custos
	Operacionalização	Aquela que testa aspectos da implantação do objeto de estudo em seu ambiente de produção
	Usabilidade	Aquela que testa o quanto simples é o uso do objeto de estudo

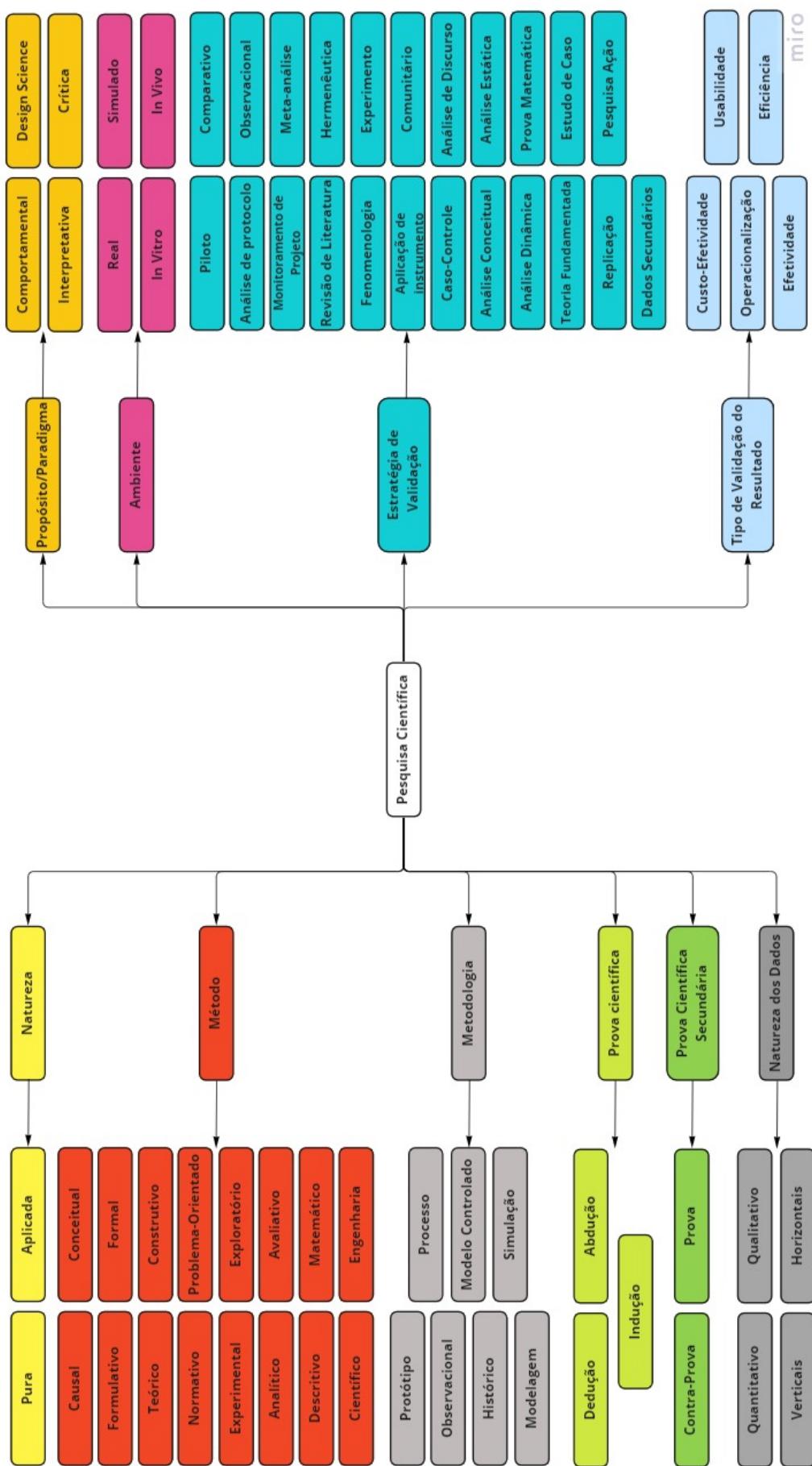


Figura 6 – Taxonomia de Pesquisa em Ciência em Computação

4.1 Validação

O processo de validação da taxonomia busca identificar sua aplicabilidade e completude na área da Computação. Para isso foi elaborada uma ontologia baseada na taxonomia proposta com a qual foi realizada uma pesquisa textual em publicações científicas de periódicos da ACM. O objetivo deste processo foi encontrar as ocorrências dos descritores da ontologia que efetivamente se referem ao contexto do trabalho realizado, ao objeto de estudo da pesquisa, à própria pesquisa em si, ao artigo escrito e aos resultados obtidos.

4.2 Método de Classificação

O volume de dados esperado no experimento que se pretende executar é grande sendo representado por uma grande coleção de artigos científicos. Dado ao grande número de artigos a serem manipulados, realizar a categorização destes na taxonomia proposta de forma manual consumiria grande quantidade de tempo além de, sob a ótica científica e seus preceitos, tal metodologia não seria reproduzível de maneira prática. Para realizar tarefas que precisem manipular grande quantidade de dados, são comumente utilizadas técnicas automatizadas (ou semi-automatizadas) de classificação de dados, conhecidas como aprendizado de máquina. O objetivo desse campo é desenvolver técnicas computacionais capazes de adquirir conhecimento e tomar decisões com base em experiências bem-sucedidas de problemas anteriores (MONARD; BARANAUSKAS, 2003). Essas técnicas também são capazes de generalizar padrões extraídos de conteúdos textuais (ROSSI, 2016).

O aprendizado dessas técnicas pode ocorrer de maneira indutiva, desde que o número de exemplos ou casos seja suficiente para a comprovação da hipótese (MONARD; BARANAUSKAS, 2003; ROSSI, 2016). Ampliando esse conceito, tem-se o aprendizado supervisionado. Onde é fornecido um conjunto de exemplos de treinamento dos padrões conhecidos e aceitos como válidos para a comprovação da hipótese, e também o aprendizado não supervisionado, no qual não são fornecidos exemplos, e a técnica procura encontrar padrões por conta própria (FERREIRA, 2018). A combinação desses dois últimos é denominada aprendizado semi-supervisionado (FERREIRA, 2018), o qual consiste na utilização de um pequeno conjunto de exemplos que permitam à técnica classificar os demais casos da população (MONARD; BARANAUSKAS, 2003; FERREIRA, 2018).

Dentre as diferentes categorias que as técnicas de aprendizado podem assumir estão os algoritmos baseados em instâncias (*Instance Based Learning – IBL*) (KOTSIANTIS et al., 2007; SEN; HAJRA; GHOSH, 2020). Esses algoritmos IBL são considerados técnicas “preguiçosas”, uma vez que a indução do resultado é feita uma vez existindo o caso, não predizendo qual será o próximo caso. Basicamente, são uma

forma de classificar, “lembrando” de outro caso semelhante (MONARD; BARANAUSKAS, 2003).

Os métodos *Nearest Neighbour* (NN) são os mais simples entre os algoritmos IBL (RASCHKA, 2018; SEN; HAJRA; GHOSH, 2020). Tais métodos são amplamente utilizados e com resultados satisfatórios na tarefa de classificar dados (FERREIRA, 2018; MONARD; BARANAUSKAS, 2003; ROSSI, 2016). A ideia central dos métodos NN está na análise dos dados vizinhos à observação considerando um número ou distância “k” para determinar sua classe.

Essa distância representa essencialmente o número de casos vizinhos que devem ser considerados na classificação de um novo caso (SEN; HAJRA; GHOSH, 2020). Quando essa distância k é igual a um ($k=1$), o método é chamado simplesmente de NN (RASCHKA, 2018); caso contrário, é denominado k-NN. Na prática, a literatura geralmente utiliza a nomenclatura k-NN independentemente do valor de k. Além disso, a complexidade do k-NN é $O(n)$ no caso de uma dimensão apenas no conjunto de exemplos, ou $O(n \times m)$, onde n representa o número do conjunto de exemplos e m o número de dimensões da tabela de exemplos (RASCHKA, 2018).

Dentro do presente trabalho, o k-NN (ou NN) é utilizado para classificar os resultados obtidos no experimento, visando validar e consolidar a taxonomia proposta. Optou-se por atribuir a k o valor um (1), considerando apenas os vizinhos imediatamente anterior e posterior a cada observação dos dados, levando em consideração a análise de contexto, conforme será melhor descrito na seção seguinte.

O objetivo desta seção não é esgotar a conceituação sobre aprendizado de máquina, mas contextualizar o tema e padronizar os conceitos utilizados dentro do presente trabalho. A escolha dessa técnica (k-NN) e não de outra se dá, conforme já mencionado, pela sua simplicidade de implementação aliada a um bom desempenho durante o processo de classificação e pela complexidade algorítmica.

4.3 Método

Para realizar a validação da taxonomia proposta foi elaborada uma estratégia de pesquisa textual em publicações científicas. Tal estratégia encontra-se de maneira macro apresentada na Seção 4.4. Para sua execução foi elaborada a automatização dos processos de pesquisa com a construção de um aplicativo, composto de uma série de procedimentos para extração, filtro e registro de informações a partir de uma base de artigos. Este aplicativo é apresentado a seguir na Seção 4.5.

4.4 Estratégia de Pesquisa

Em linhas gerais a estratégia de pesquisa utilizada, mostrada na Figura 7, inicia com a realização da pesquisa por referências bibliográficas e processá-las. Deste processamento resultam informações as quais foram armazenadas em uma base de dados. Também durante o processamento foi feito o *download* dos arquivos das publicações em formato *Portable Document Format* (PDF) e sua conversão para formato texto puro, ambos armazenados em disco. Utilizando os arquivos em formato texto puro, foi realizada a busca textual pelos descritores da ontologia, e o resultado foi novamente armazenado na base de dados.

De posse do resultado da busca textual, as observações obtidas foram submetidas a um algoritmo de classificação semi-supervisionado. O processo de classificação analisou o contexto de uso do descritor no texto, levando em consideração o conceito por trás do descritor levantado pela RSL. As observações classificadas positivamente ou aceitas (como são referenciadas no processo de análise dos dados) são marcadas como tal nos respectivos registros da base de dados. A regra que gerou essa classificação é anotada na tabela de aprendizagem do algoritmo para dar sequência na classificação dos demais casos. Ao final, após a classificação das observações tem-se a taxonomia consolidada.

Para por em prática a estratégia de pesquisa, foram elaborados processos automatizados dentro de um aplicativo desenvolvido para o presente trabalho, o qual é descrito a seguir.

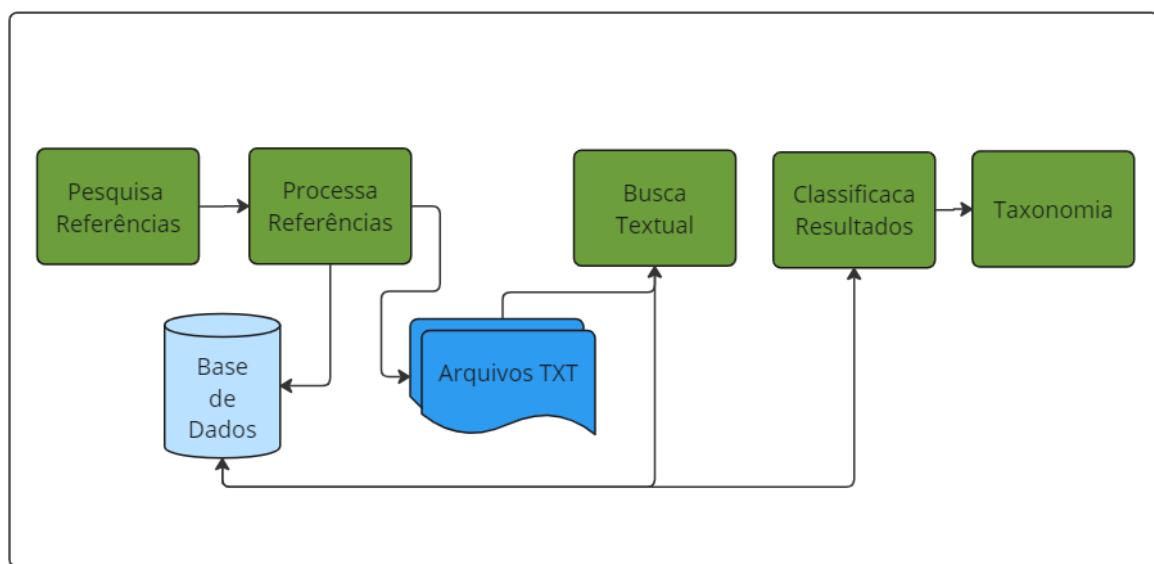


Figura 7 – Estratégia de pesquisa

4.5 Processo

As Figuras 8 e 9 representam o fluxograma do marco metodológico utilizado, complementando-se mutuamente e sendo descritas a seguir. Para facilitar a compreensão, foi adotado um esquema de cores para distinguir os diversos elementos do marco metodológico. Na cor amarelo são identificadas as entradas de dados para as rotinas de processamento, estas identificadas na cor verde. Tais rotinas de processamento foram desenvolvidas na forma de um artefato de software. Elementos em vermelho representam os processos manuais realizados, enquanto os na cor azul escuro indicam os diferentes formatos de arquivos utilizados. Por fim, os elementos marcados em azul claro representam os processos de armazenamento de dados.

Com base na taxonomia elaborada, foi construída uma ontologia de busca (ver Tabela 7). Essa ontologia foi empregada na pesquisa textual em publicações científicas para avaliar a ocorrência e contexto das classificações. Ela consiste em uma lista com até quatro descritores afins para cada uma das classificações da taxonomia. Incluindo o descritor principal e mais três termos relacionados. Essa abordagem justifica-se pela riqueza da linguagem escrita, que inclui sinônimos e expressões mais complexas que podem representar a mesma ideia.

O repositório Crossref¹ serviu como fonte para seleção dos artigos para pesquisa. Para acesso a esse repositório foi utilizada sua interface de programação de aplicativos (API)². O Crossref foi utilizado em detrimento das bases de dados oficiais dos editores por ser um repositório que permite agilidade em encontrar, acessar, conectar, citar e reutilizar publicações científicas na forma de artigos. A API deste repositório disponibiliza um extenso conjunto de metadados de milhares de publicações científicas de centenas de periódicos em todo o mundo. Entre os metadados escolhidos que fazem parte dos dados deste trabalho estão: periódico de publicação, data de publicação, título, indicador de restrição de acesso, link para acesso ao PDF, idioma em que foi publicado, tipo de publicação, resumo/abstract e áreas de conhecimento de referência do periódico onde o artigo foi publicado.

Cabe destacar que até o ano de 2017 os membros participantes do consórcio não eram obrigados a submeter o conjunto completo de metadados propostos pelo repositório. Havia a possibilidade de que metadados fossem suprimidos na submissão, e, a partir de 2018 todo o conjunto de metadados passou a ser obrigatório. Este fato fez com que 2018 fosse marco temporal para seleção de artigos para o processo de validação da taxonomia.

Durante os testes e análise da documentação da interface de acesso ao Crossref, verificou-se que a paginação dos resultados é de no máximo 1000 referências por página. Entre as opções de pesquisa disponíveis na API, foi utilizada aquela que permite

¹<https://www.crossref.org/>

²<https://api.crossref.org/swagger-ui/index.html>

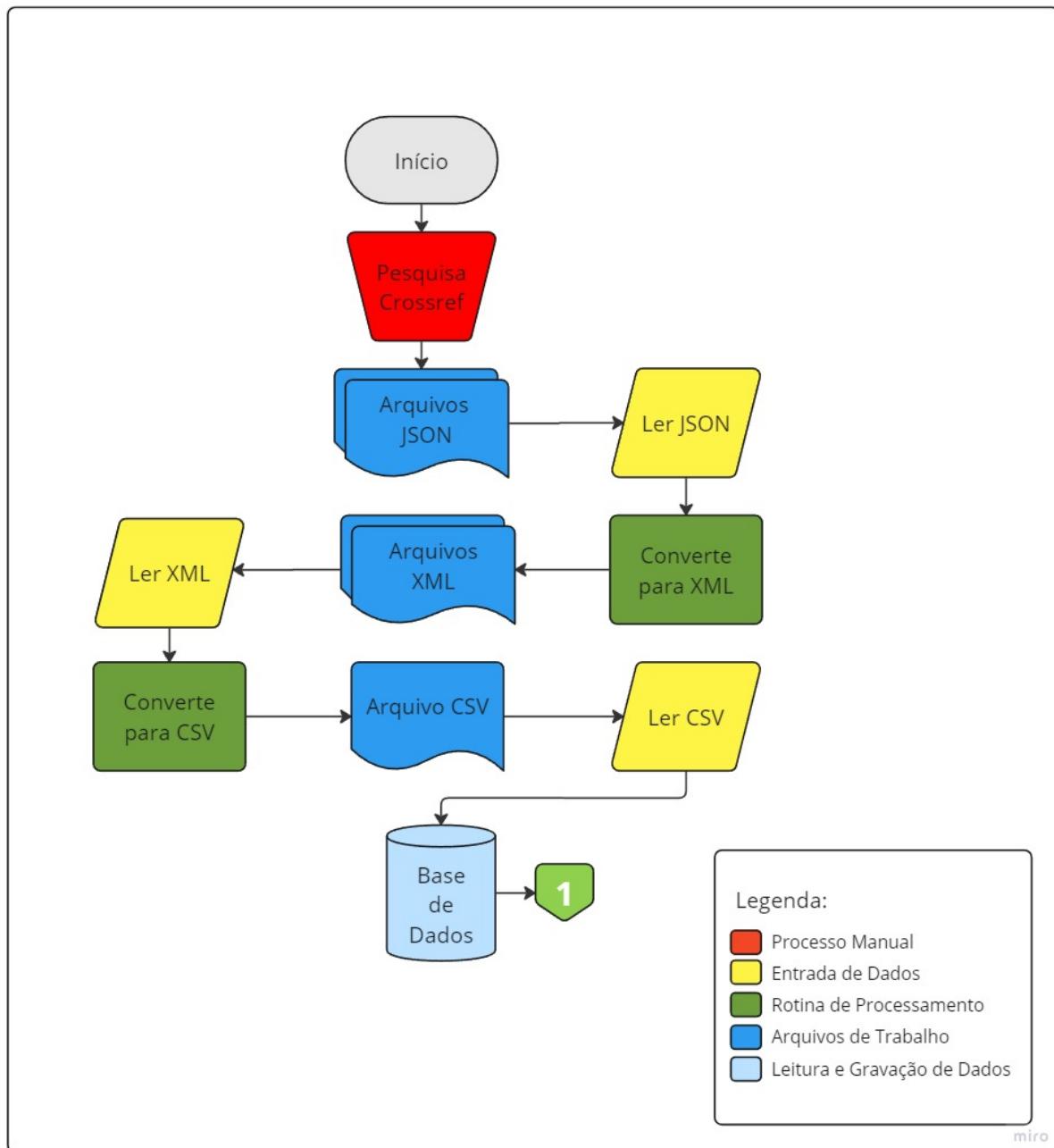


Figura 8 – Marco Metodológico Parte 1

realizar pesquisas em uma determinada base indexada ou membro do repositório.

A pesquisa foi realizada em artigos científicos publicados pela ACM de janeiro de 2018 a agosto de 2022 (aproximadamente 5 anos a partir da data da pesquisa) que continham a chave de pesquisa “ciência da Computação” e eram de acesso aberto. O endereço de busca utilizado na API foi:

https://doi.crossref.org/members/320/works?query=computer+science&rows=1000&filter=reference-visibility:open,from-pub-date:2018-01-01,type:journal-article&facet=type-name:*&mailto=alessander.osorio@inf.ufpel.edu.br

Certos requisitos são solicitados, mas não obrigatórios, sendo fortemente sugeridos.

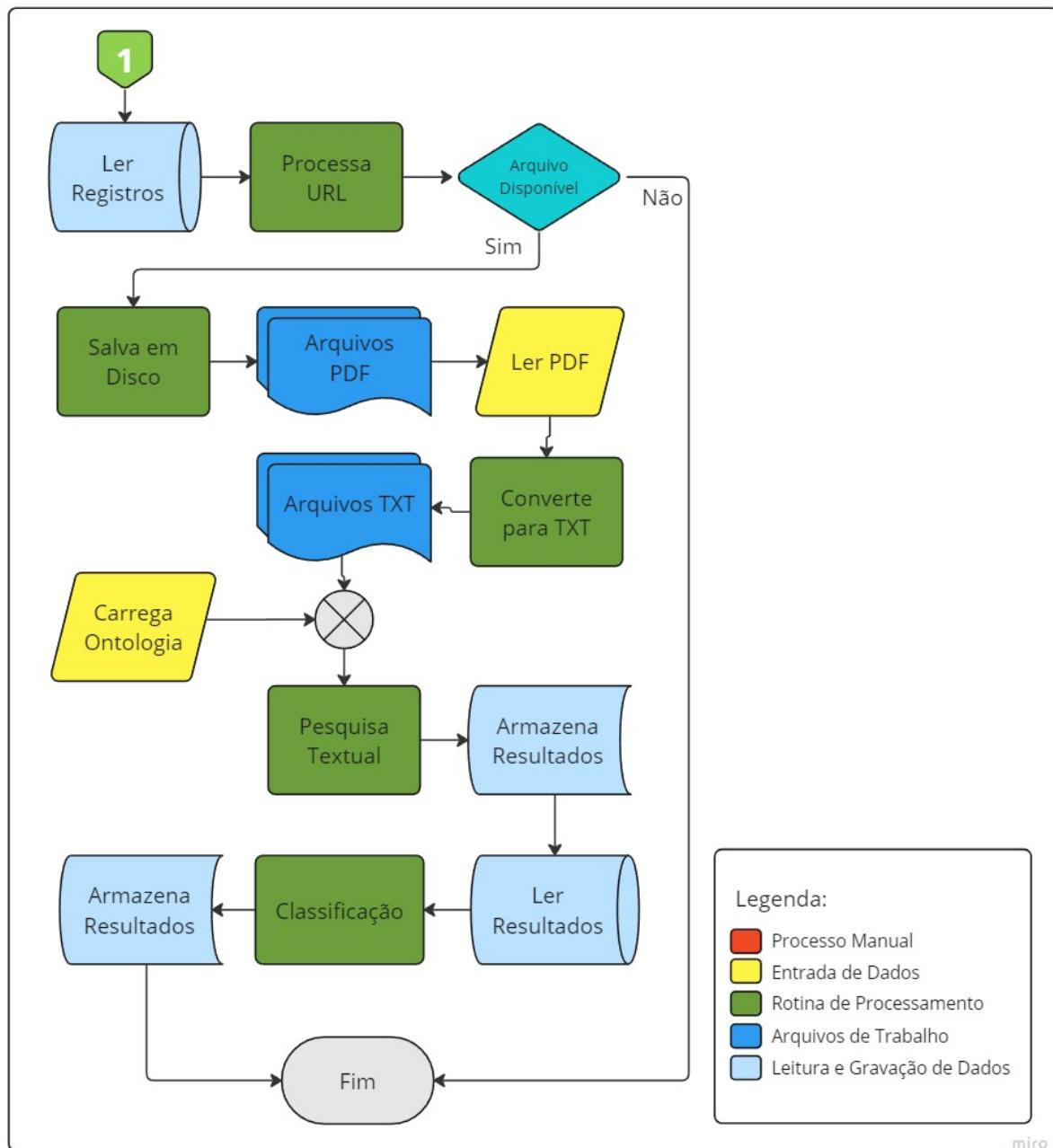


Figura 9 – Marco Metodológico Parte 2

dos para uso na documentação da API. O parâmetro *mailto* identifica aquele que está realizando a consulta à API. Segundo a documentação, o uso deste parâmetro pode aumentar a prioridade das requisições e diminuir o intervalo de tempo entre elas. Além disso, ele serve como um meio de contato em caso de problemas.

O resultado encontrado utilizando a API com os parâmetros mencionados compreendeu um total de 6799 referências. Esses resultados foram paginados em 7 arquivos no formato *JavaScript Object Notation* (JSON), os quais foram convertidos para o formato XML. Embora o XML tenha um tamanho maior, mostrou-se mais prático para buscar e recuperar as *tags* internas referentes aos metadados. A partir disso, foi reali-

zada a montagem de um arquivo *Comma Separated Values* (CSV) com os metadados selecionados. Estes arquivos foram importados em uma base de dados MS Access para sistematização das referências e demais processamentos. Conforme mostrado na Figura 8.

Após a sistematização na base de dados dos arquivos de retorno e suas respectivas referências (conforme Figura 9), começou o processamento dos 6799 registros contidos na base de dados. Do total de referências, foi feito o *download* de 2530 artigos de acesso aberto. Mesmo utilizando a opção de restrição de acesso ao efetuar a pesquisa com a API, não foi possível obter todos os arquivos das referências encontradas. Alguns mesmo com a *tag* de restrição tendo valor “no”, indicando obviamente não haver restrições no acesso à publicação, foram encontrados restritos. Para as referências disponíveis os artigos foram convertidos de formato PDF para formato texto possibilitando a busca textual da ontologia organizada.

Com os arquivos em formato texto puro, foi realizada a pesquisa textual da ontologia, que buscou todos os descritores dentro de cada um dos arquivos de maneira individual, conforme a organização dentro do banco de dados.

Neste ponto do processo foram obtidas cerca de 120 mil ocorrências da ontologia as quais foram armazenadas em uma tabela no banco de dados. Essas ocorrências representam a taxonomia dentro dos 2530 artigos. Entretanto, a simples ocorrência dos descritores da ontologia no texto não necessariamente representa efetivamente o correto contexto de uso do descritor. Um descritor pode ser usado dentro de um contexto com um sentido diferente daquele que ele está sendo referenciado neste trabalho, dentro da ontologia e da sua respectiva representação da taxonomia. Ou seja, não representa efetivamente o sentido textual que se espera empregar.

Com o propósito de aprofundar a análise do contexto de cada ocorrência do descritor no artigo e classificá-las corretamente, foi escrita uma sub-rotina de seleção semi-supervisionada. As ocorrências da pesquisa textual foram summarizadas de acordo com a quantidade de ocorrências, ordenadas de maneira decrescente, e para cada descritor, essas ocorrências foram individualmente analisadas.

O processo foi baseado no algoritmo *k-Nearest Neighbor* (k-NN) de análise semi-supervisionada. Ele começa com a seleção, no fragmento de texto onde a ocorrência do descritor se encontra, das palavras imediatamente vizinhas, tanto anterior quanto posterior. Essas palavras são destacadas para o revisor que pode aceitar que os termos vizinhos associados ao descritor dão o contexto correto de uso ou não, conforme o conceito do descritor na taxonomia proposta. O revisor pode aceitar ambas as vizinhanças ou apenas uma delas, da mesma forma que pode rejeitar.

Uma vez aceito o contexto de uso o revisor confirma tal ação para a sub-rotina. A qual grava o contexto em um arquivo de aprendizagem e passa a procurar nas demais ocorrências contextos semelhantes para aquele descritor e contexto em específico. Os

contextos no arquivo de aprendizagem são válidos para os descritores individualmente dentro de uma mesma categoria de descritores. Pois um contexto válido para um descritor, que o torna aceito como válido, pode rejeitar outro descritor. Os vizinhos associados ao descritor produzem contextos diferentes conforme o descritor.

4.6 Resultados

Nesta seção, são apresentados e descritos os resultados do processo de validação da taxonomia. Devido ao grande volume de dados e informações computadas na análise, o Apêndice A será utilizado como apoio em algumas situações. No corpo do texto, são destacados os pontos principais dos resultados e quando necessário este capítulo em questão será referenciado.

4.6.1 A ontologia de pesquisa

Para uma melhor compreensão da análise de dados aqui descrita, inicia-se apresentando a ontologia que representa a taxonomia e que foi utilizada na busca textual no processo de validação. Na Tabela 7, é possível visualizar as categorias e seus respectivos descritores. Conforme já mencionado, algumas categorias possuem mais de um descritor. O primeiro descritor (#1) é aquele que foi encontrado na RSL e que foi traduzido para a língua inglesa. Os demais são variações de grafia ou diferentes formas de se referenciar ao conceito por trás do primeiro.

4.6.2 Considerações iniciais dos resultados

A presente análise dos resultados diz respeito aos dados oriundos do processo de validação da taxonomia proposta segundo à ontologia dela derivada. Tal ontologia, composta por 96 descritores divididos em 10 categorias, foi pesquisada no texto de 2.530 artigos de 68 periódicos da ACM publicados entre 2018 e agosto de 2023, totalizando 124.601 observações. Estas foram submetidas ao algoritmo semi-supervisionado de classificação k-NN para verificar sua adequação aos conceitos referentes a cada descritor levantados pela revisão sistemática de literatura.

Os dados foram analisados no *software RStudio*³ e ao final, foram considerados aceitos como dentro do contexto do conceito que é proposto pela taxonomia, 91.419 observações representando 73,37% do total. Em contraposição, dentro dos mesmos padrões, 33.182 observações não foram aceitas ou classificadas, as quais representam 26,63% do total, conforme a Tabela 1 no Apêndice A.

As Tabelas 2 e 3 do Apêndice A mostram os totais de ocorrências por Categoria e Descritor, respectivamente. Note que dado o número total de observações, alguns

³Os *scripts* utilizados na análise bem com os arquivos de dados encontram-se disponíveis em: <https://github.com/metasciencestudies/taxonomy>

Tabela 7 – Ontologia de pesquisa

Categoría	Descriptor #1	Descriptor #2	Descriptor #3	Descriptor #4
Natureza	applied pure behavioral critique design science interpretive			
Propósito paradigma				
Método	analytical causal conceptual constructive descriptive engineering evaluative experimental exploratory formal formulative math normative problem-oriented scientific theoretical controlled model historic modeling observational processual prototype simulation abduction deduction induction proof	cause-and-effect concept problem-solving illustrative evaluation empiric investigate		investigative
Metodologia		standarts problem oriented		doctrinal
		theoretical control model historical		
			process analysis	
Prova Científica				
Prova Científica Secundária				
Ambiente	rebuttal in vitro in vivo real simulated	real-world		real world
Dados	horizontals qualitative quantitative verticals action research	longitudinal		
Estratégia de Validação/Coleta de Dados		cross-section		cross-sectional
	case study case-control community rese- arch comparative conceptual analysis dynamic analysis experiment grounded theory hermeneutics instrument applica- tion math test meta-analysis phenomenology pilot project monitoring protocol analysis replication secondary data speech analysis static analysis systematic literature review	benchmark		workload
			interviews	survey
				interview
		pilot study		
Tipo de Validação do Resultado	cost-effectiveness effectiveness efficiency operationalization usability	RSL		

descritores percentualmente apresentam valores não passíveis de representação matemática com o número de casas decimais utilizadas. Todavia, não é possível desconsiderar tais valores, pois estes compõem o total da categoria, bem como o total geral.

A variável de desfecho analisada é de natureza categórica, tendo o valor “Sim” para os aceitos ou classificados, e “Não” para os não aceitos ou não classificados. Os valores apresentados aqui representam a contagem de cada uma delas individualizada. As categorias e seus descritores foram analisados individualmente sem procurar estabelecer uma relação entre elas. Os histogramas de frequência de valores por Categoria e Descritor mostram que os dados não seguem a distribuição normal, sendo o teste do Qui-Quadrado o recomendado para verificação estatística dos valores desta análise.

Ao ser aplicado o teste do Qui-Quadrado, para toda a população das observações, o p-valor encontrado é inferior a 0,01 ($p<0,01$), o que rejeita a Hipótese Nula e aceita a Hipótese Alternativa para os dados analisados.

4.6.3 Resultados por categoria da ontologia

Para analisar os resultados por categoria da ontologia (Tabela 8) não considerou-se o “n” puro, uma vez que as categorias possuem internamente números distintos de descritores. Isto faria com que uma categoria com maior número de descritores, logicamente, obtivesse maior número de ocorrências e por consequência valores enviados em seus números. Nesta parte da análise foram considerados os número de descritores por categoria. Sendo o número de ocorrências dividido por tais valores compõe assim um valor para ‘n’ normalizado ou a média interna por categoria.

Exemplificando, como pode ser visto na Tabela 8, a categoria com maior ocorrências brutas é “Método” com um total de 40.292 ocorrências e 28 descritores. Ao se normalizar este valor obtém-se um “n” normalizado de 1439. Comparando-se com a categoria “Natureza” onde esta possui apenas 2 descritores resultando em um “n” normalizado de 3690. Note que o objetivo desta análise não é estabelecer uma comparação entre as categorias, tal ação é apenas para exemplificar e justificar a forma como os números foram analisados.

Assim sendo, os demais dados da Tabela 8 foram calculados levando-se em consideração o “n” normalizado ou média por categoria. Nesta tabela não foi calculada a predominância dos valores, uma vez que, exceto a categoria “Propósito Paradigma”, todas as demais possuem predominância positiva. A categoria em questão possui uma diferença entre os valores para aceitos e não aceitos percentualmente nula inferior a 0,01% ou 19 ocorrências.

A Figura 10 sintetiza visualmente a Tabela 8. Nela as colunas “Total Sim” e “Total Não” da referida tabela são consideradas para comporem as colunas do gráfico. Da

Tabela 8 – Aceite por categorias de descritores da ontologia

Categoría	N.Desc.	n	%	n.Norm.	Total nSim	%SCat.	%STot.	Total Não	%NCat.	%NTot.
Ambiente	6	11693	9,38	1949	1404	12,0	1,13	544	4,65	0,44
Dados	7	5576	4,48	797	691	12,4	0,55	105	1,88	0,08
Estratégia de Validação/Coleta de Dados	29	26014	20,88	897	695	2,7	0,56	201	0,77	0,16
Metodologia	10	15261	12,25	1526	998	6,5	0,80	527	3,45	0,42
Método	28	40292	32,34	1439	1026	2,5	0,82	412	1,02	0,33
Natureza	2	7380	5,92	3690	2707	36,7	2,17	982	13,31	0,79
Propósito paradigma	4	2776	2,23	694	337	12,1	0,27	356	12,82	0,29
Prova Científica Secundária	2	6977	5,60	3488	3372	48,3	2,71	116	1,66	0,09
Prova Científica	3	1234	0,99	411	348	28,2	0,28	63	5,11	0,05
Tipo de Validação do Resultado	5	7398	5,94	1480	943	12,8	0,76	536	7,25	0,43
Total	96	124601	100,00	1298	12521	NA	10,05	3842	NA	3,08

mesma forma, no Apêndice A, nas Figuras 4 a 13 estão individualizados por categoria com seus respectivos descritores.

Médias de Ocorrências por Categoria

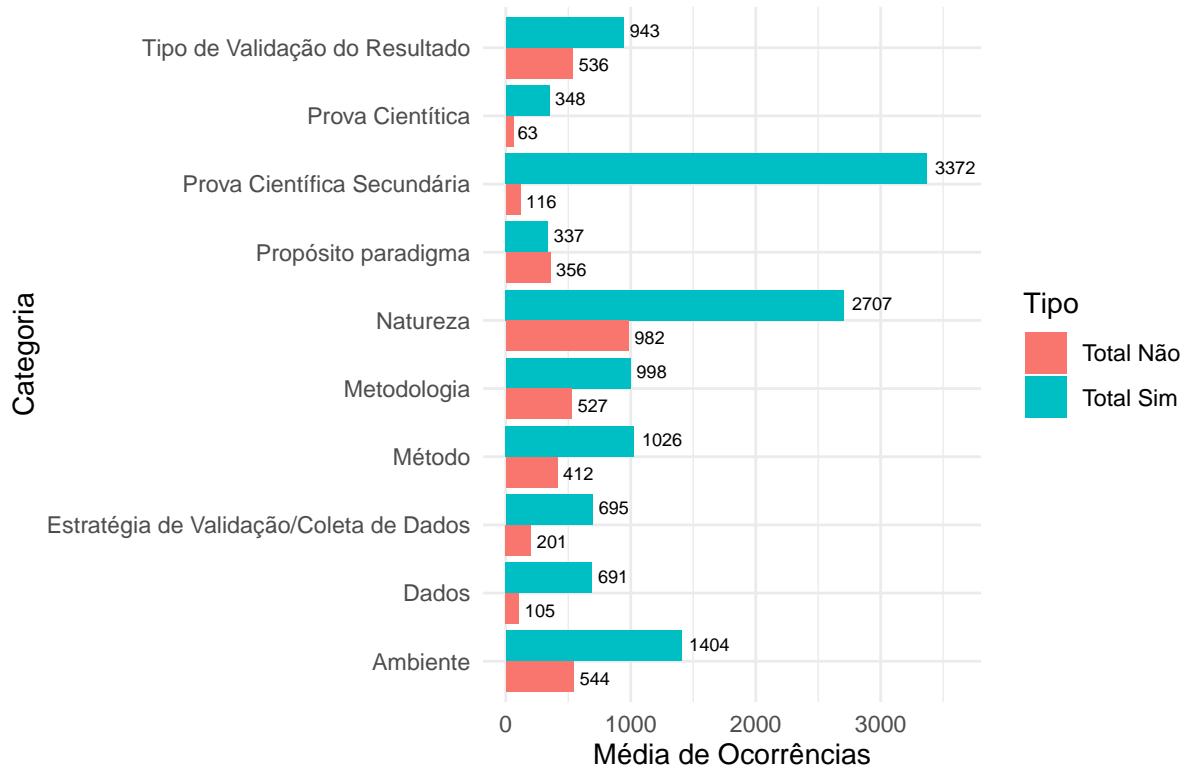


Figura 10 – Médias de Ocorrências por Categoria

4.6.4 Resultados por descritores da ontologia

Do total de 96 descritores, 89 deles obtiveram resultados na busca textual nos artigos utilizados para tal. Alguns descritores, sete no total, não obtiveram resultado algum na busca, conforme mostra a Tabela 9, e, por consequência, não estão na Tabela 10. Destes, a maioria corresponde ao segundo descritor, tendo o primeiro

Tabela 9 – Descritores sem ocorrência

Categoria	descritor	Número do descritor
Método	formulative	1
	illustrative	2
	problem oriented	2
	theoretical	2
Dados	horizontals	1
Estratégia de Validação/Coleta de Dados	project monitoring	1
	RSL	2

encontrado resultados durante o processo de validação. Ainda, especificamente o descritor “RSL” acabou não sendo traduzido para o respectivo acrônimo na língua inglesa.

Os resultados expressos na Tabela 10 representam o número de ocorrências obtidas para cada descritor da ontologia de maneira individual, sem qualquer tipo de associação entre eles, tão pouco sua categoria. Os números também correspondem ao total de aceitos (ordenação da tabela juntamente com a categoria) e não aceitos na classificação pelo algoritmo k-NN. Os números são puros e os percentuais foram calculados com base no total do descritor (%Desc) e ao total geral de ocorrências (%Tot). Novamente, note que alguns percentuais não puderam ser representados matematicamente com o número de casas decimais utilizadas nesta análise. Reitera-se que tais resultados não foram descartados por comporem os resultados totais por categoria do descritor.

O destaque da tabela cabe à última coluna - Pred. - Predominância do resultado. Ela expressa a comparação dos resultados aceitos e não aceitos. Se um é maior que o outro. Se o “n” de aceitos é maior que o de não aceitos, na coluna será mostrado o conteúdo “S”; caso contrário, “N” (predominância de “Sim” ou “Não”). Analisando esta coluna, pode-se perceber que os descritores “verticals”, “controled model” e “doctrinal”, os quais, além da baixa ocorrência, não tiveram nenhum caso de classificação positiva, indicando sua não utilização dentro da pesquisa em Computação para a população estudada.

Ainda analisando a coluna predominância, pode-se perceber que quase a totalidade daqueles descritores onde predominam os não aceitos na classificação os percentuais em relação ao total de ocorrências são inferiores a 1% (0,6%, para ser exato), excetuando os descritores “modeling” (4,7%), “concept” (2,1%) e “effectiveness” (2,07%). Analisando individualmente estes casos, nota-se que as diferenças entre aceitos e não aceitos são relativamente pequenas. Quase inexistentes e não representam mais que 1,5% em relação ao total do descritor.

A seguir são feitos os destaques referentes às Tabelas 6 a 15 do Apêndice A. Como já mencionado alguns resultados em função do número de casas decimais utilizado nas tabelas não possuem significativa representação matemática e/ou tiveram seus

Tabela 10 – Aceite por descritores da ontologia

Categoría	Descriptor	n	%	Total Sim	%Desc.	%Tot.	Total Não	%Desc.	%Tot.	Pred.
Ambiente	real	6799	5,46	4718	69	3,79	2081	30,6	1,67	S
Ambiente	real-world	3028	2,43	2338	77	1,88	690	22,8	0,55	S
Ambiente	simulated	1496	1,20	1167	78	0,94	329	22,0	0,26	S
Ambiente	real world	304	0,24	161	53	0,13	143	47,0	0,11	S
Ambiente	in vivo	47	0,04	32	68	0,03	15	31,9	0,01	S
Ambiente	in vitro	19	0,02	8	42	0,01	11	57,9	0,01	N
Dados	qualitative	3153	2,53	2995	95	2,40	158	5,0	0,13	S
Dados	quantitative	1762	1,41	1403	80	1,13	359	20,4	0,29	S
Dados	longitudinal	536	0,43	365	68	0,29	171	31,9	0,14	S
Dados	cross-sectional	75	0,06	59	79	0,05	16	21,3	0,01	S
Dados	cross-section	46	0,04	19	41	0,02	27	58,7	0,02	N
Dados	verticals	4	0,00	0	0	0,00	4	100,0	0,00	N
Estratégia de Validação/Coleta de Dados	survey	6692	5,37	6082	91	4,88	610	9,1	0,49	S
Estratégia de Validação/Coleta de Dados	experiment	3978	3,19	2463	62	1,98	1515	38,1	1,22	S
Estratégia de Validação/Coleta de Dados	workload	2616	2,10	2080	80	1,67	536	20,5	0,43	S
Estratégia de Validação/Coleta de Dados	benchmark	2957	2,37	2006	68	1,61	951	32,2	0,76	S
Estratégia de Validação/Coleta de Dados	interviews	2306	1,85	1876	81	1,51	430	18,6	0,35	S
Estratégia de Validação/Coleta de Dados	interview	2073	1,66	1673	81	1,34	400	19,3	0,32	S
Estratégia de Validação/Coleta de Dados	case study	1395	1,12	1236	89	0,99	159	11,4	0,13	S
Estratégia de Validação/Coleta de Dados	pilot	870	0,70	773	89	0,62	97	11,2	0,08	S
Estratégia de Validação/Coleta de Dados	comparative	703	0,56	565	80	0,45	138	19,6	0,11	S
Estratégia de Validação/Coleta de Dados	static analysis	650	0,52	389	60	0,31	261	40,1	0,21	S
Estratégia de Validação/Coleta de Dados	replication	632	0,51	253	40	0,20	379	60,0	0,30	N
Estratégia de Validação/Coleta de Dados	grounded theory	237	0,19	184	78	0,15	53	22,4	0,04	S
Estratégia de Validação/Coleta de Dados	dynamic analysis	273	0,22	170	62	0,14	103	37,7	0,08	S
Estratégia de Validação/Coleta de Dados	pilot study	179	0,14	145	81	0,12	34	19,0	0,03	S
Estratégia de Validação/Coleta de Dados	meta-analysis	150	0,12	132	88	0,11	18	12,0	0,01	S
Estratégia de Validação/Coleta de Dados	action research	72	0,06	51	71	0,04	21	29,2	0,02	S
Estratégia de Validação/Coleta de Dados	phenomenology	29	0,02	22	76	0,02	7	24,1	0,01	S
Estratégia de Validação/Coleta de Dados	secondary data	18	0,01	17	94	0,01	1	5,6	0,00	S
Estratégia de Validação/Coleta de Dados	systematic literature review	135	0,11	15	11	0,01	120	88,9	0,10	N
Estratégia de Validação/Coleta de Dados	math test	17	0,01	7	41	0,01	10	58,8	0,01	N
Estratégia de Validação/Coleta de Dados	case-control	6	0,00	6	100	0,00	0	0,0	0,00	S
Estratégia de Validação/Coleta de Dados	conceptual analysis	6	0,00	6	100	0,00	0	0,0	0,00	S
Estratégia de Validação/Coleta de Dados	community research	10	0,01	4	40	0,00	6	60,0	0,00	N
Estratégia de Validação/Coleta de Dados	protocol analysis	3	0,00	3	100	0,00	0	0,0	0,00	S
Estratégia de Validação/Coleta de Dados	speech analysis	4	0,00	3	75	0,00	1	25,0	0,00	S
Estratégia de Validação/Coleta de Dados	hermeneutics	2	0,00	1	50	0,00	1	50,0	0,00	S
Estratégia de Validação/Coleta de Dados	instrument application	1	0,00	1	100	0,00	0	0,0	0,00	S
Metodología	simulation	6111	4,90	4182	68	3,36	1929	31,6	1,55	S
Metodología	modeling	5681	4,56	2681	47	2,15	3000	52,8	2,41	N
Metodología	prototype	1452	1,17	1388	96	1,11	64	4,4	0,05	S
Metodología	historical	1365	1,10	1270	93	1,02	95	7,0	0,08	S
Metodología	observational	483	0,39	422	87	0,34	61	12,6	0,05	S
Metodología	control model	30	0,02	25	83	0,02	5	16,7	0,00	S
Metodología	historic	131	0,11	16	12	0,01	115	87,8	0,09	N
Metodología	process analysis	4	0,00	4	100	0,00	0	0,0	0,00	S
Metodología	processual	3	0,00	1	33	0,00	2	66,7	0,00	N
Metodología	controlled model	1	0,00	0	0	0,00	1	100,0	0,00	N
Método	evaluation	9298	7,46	8409	90	6,75	889	9,6	0,71	S
Método	experimental	4924	3,95	3919	80	3,15	1005	20,4	0,81	S
Método	engineering	5658	4,54	3550	63	2,85	2108	37,3	1,69	S
Método	investigate	2397	1,92	2247	94	1,80	150	6,3	0,12	S
Método	formal	3383	2,72	2022	60	1,62	1361	40,2	1,09	S
Método	theoretical	2502	2,01	1515	61	1,22	987	39,5	0,79	S
Método	causal	1589	1,28	1369	86	1,10	220	13,8	0,18	S
Método	scientific	1916	1,54	1281	67	1,03	635	33,1	0,51	S
Método	conceptual	1171	0,94	1095	94	0,88	76	6,5	0,06	S
Método	concept	2531	2,03	775	31	0,62	1756	69,4	1,41	N
Método	analytical	832	0,67	667	80	0,54	165	19,8	0,13	S
Método	exploratory	715	0,57	536	75	0,43	179	25,0	0,14	S
Método	descriptive	765	0,61	482	63	0,39	283	37,0	0,23	S
Método	normative	432	0,35	240	56	0,19	192	44,4	0,15	S
Método	math	497	0,40	164	33	0,13	333	67,0	0,27	N
Método	problem-solving	234	0,19	134	57	0,11	100	42,7	0,08	S
Método	standards	863	0,69	130	15	0,10	733	84,9	0,59	N
Método	constructive	401	0,32	91	23	0,07	310	77,3	0,25	N
Método	investigative	98	0,08	51	52	0,04	47	48,0	0,04	S
Método	evaluative	62	0,05	37	60	0,03	25	40,3	0,02	S
Método	empiric	14	0,01	11	79	0,01	3	21,4	0,00	S
Método	cause-and-effect	8	0,01	8	100	0,01	0	0,0	0,00	S
Método	problem-oriented	1	0,00	1	100	0,00	0	0,0	0,00	S
Método	doctrinal	1	0,00	0	0	0,00	1	100,0	0,00	N
Natureza	applied	6416	5,15	4807	75	3,86	1609	25,1	1,29	S
Natureza	pure	964	0,77	608	63	0,49	356	36,9	0,29	S
Propósito paradigma	behavioral	2202	1,77	1145	52	0,92	1057	48,0	0,85	S
Propósito paradigma	interpretive	137	0,11	100	73	0,08	37	27,0	0,03	S
Propósito paradigma	critique	400	0,32	77	19	0,06	323	80,8	0,26	S
Propósito paradigma	design science	37	0,03	27	73	0,02	10	27,0	0,01	S
Prova Científica Secundária	proof	6969	5,59	6739	97	5,41	230	3,3	0,18	S
Prova Científica Secundária	rebuttal	8	0,01	6	75	0,00	2	25,0	0,00	S
Prova Científica	induction	1131	0,91	979	87	0,79	152	13,4	0,12	S
Prova Científica	deduction	64	0,05	34	53	0,03	30	46,9	0,02	S
Prova Científica	abduction	39	0,03	31	79	0,02	8	20,5	0,01	S
Tipo de Validação do Resultado	efficiency	3194	2,56	2586	81	2,08	608	19,0	0,49	S
Tipo de Validação do Resultado	usability	1546	1,24	1345	87	1,08	201	13,0	0,16	S
Tipo de Validação do Resultado	effectiveness	2499	2,01	640	26	0,51	1859	74,4	1,49	N
Tipo de Validação do Resultado	cost-effectiveness	86	0,07	79	92	0,06	7	8,1	0,01	S
Tipo de Validação do Resultado	operationalization	73	0,06	65	89	0,05	8	11,0	0,01	S
NA	Total	124601	100,00	91419	NA	73,37	33182	NA	26,63	NA

valores arredondados para cima. Note que os destaques feitos aqui, refletem o maior e o menor número de ocorrências e percentuais de aceitação. Em alguns casos se referem ao mesmo descritor. Em outros casos a categoria tem apenas dois descritores, logicamente um deles se enquadrará em algum dos casos destacados. Todavia, não há demérito neste resultado.

Categoria - Ambiente. Maior número de ocorrências descritor: *real*, com média 1133. Com percentual de aceitação 3,8% e não aceitação de 1,7%. Menor número de ocorrências descritor: *in vitro*, com média 3,2. Com percentual de aceitação 0,0064% e não aceitação de 0,0088%. Maior percentual de aceitação sobre o total, descritor: *real*, com média 1133. Com percentual de aceitação 3,8% e não aceitação de 1,7%. Menor percentual de aceitação sobre o total, descritor: *in vitro*, com média 3,2. Com percentual de aceitação 0,0064% e não aceitação de 0,0088%.

Categoria - Dados. Maior número de ocorrências descritor: *qualitative*, com média 450. Com percentual de aceitação 2,4% e não aceitação de 0,13%. Menor número de ocorrências descritor: *verticals*, com média 0,57. Com percentual de aceitação 0% e não aceitação de 0,0032%. Maior percentual de aceitação sobre o total, descritor: *qualitative*, com média 450. Com percentual de aceitação 2,4% e não aceitação de 0,13%. Menor percentual de aceitação sobre o total, descritor: *verticals*, com média 0,57. Com percentual de aceitação 0% e não aceitação de 0,0032%.

Categoria - Estratégia de Validação/Coleta de Dados. Maior número de ocorrências descritor: *survey*, com média 231. Com percentual de aceitação 4,9% e não aceitação de 0,49%. Menor número de ocorrências descritor: *instrument application*, com média 0,034. Com percentual de aceitação 0,00083% e não aceitação de 0%. Maior percentual de aceitação sobre o total, descritor: *survey*, com média 231. Com percentual de aceitação 4,9% e não aceitação de 0,51%. Menor percentual de aceitação sobre o total, descritor: *case-control*, com média 0,21. Com percentual de aceitação 0,0048% e não aceitação de 0%.

Categoria - Metodologia. Maior número de ocorrências descritor: *simulation*, com média 611. Com percentual de aceitação 3,4% e não aceitação de 1,5%. Menor número de ocorrências descritor: *controlled model*, com média 0,1. Com percentual de aceitação 0% e não aceitação de 0,00083%. Maior percentual de aceitação sobre o total, descritor: *simulation*, com média 611. Com percentual de aceitação 3,4% e não aceitação de 1,5%. Menor percentual de aceitação sobre o total, descritor: *process analysis*, com média 0,4. Com percentual de aceitação 0,0032% e não aceitação de 0%.

Categoria - Método. Maior número de ocorrências descritor: *evaluation*, com média 332. Com percentual de aceitação 6,7% e não aceitação de 0,71%. Menor número de ocorrências descritor: *doctrinal*, com média 0,036. Com percentual de aceitação 0% e não aceitação de 0,00083%. Maior percentual de aceitação sobre o total, descri-

tor: *evaluation*, com média 332. Com percentual de aceitação 6,7% e não aceitação de 0,71%. Menor percentual de aceitação sobre o total, desritor: *cause-and-effect*, com média 0,29. Com percentual de aceitação 0,0064% e não aceitação de 0%.

Categoria - Natureza. Maior número de ocorrências desritor: *applied*, com média 3208. Com percentual de aceitação 3,9% e não aceitação de 1,3%. Menor número de ocorrências desritor: *pure*, com média 482. Com percentual de aceitação 0,49% e não aceitação de 0,29%. Maior percentual de aceitação sobre o total, desritor: *applied*, com média 3208. Com percentual de aceitação 3,9% e não aceitação de 1,3%. Menor percentual de aceitação sobre o total, desritor: *pure*, com média 482. Com percentual de aceitação 0,49% e não aceitação de 0,29%.

Categoria - Propósito Paradigma. Maior número de ocorrências desritor: *behavioral*, com média 550. Com percentual de aceitação 0,92% e não aceitação de 0,85%. Menor número de ocorrências desritor: *design science*, com média 9,2. Com percentual de aceitação 0,022% e não aceitação de 0,008%. Maior percentual de aceitação sobre o total, desritor: *behavioral*, com média 550. Com percentual de aceitação 0,92% e não aceitação de 0,85%. Menor percentual de aceitação sobre o total, desritor: *design science*, com média 9,2. Com percentual de aceitação 0,022% e não aceitação de 0,008%.

Categoria - Prova Científica Secundária. Maior número de ocorrências desritor: *proof*, com média 3484. Com percentual de aceitação 5,4% e não aceitação de 0,18%. Menor número de ocorrências desritor: *rebuttal*, com média 4. Com percentual de aceitação 0,0048% e não aceitação de 0,0016%. Maior percentual de aceitação sobre o total, desritor: *proof*, com média 3484. Com percentual de aceitação 5,4% e não aceitação de 0,18%. Menor percentual de aceitação sobre o total, desritor: *rebuttal*, com média 4. Com percentual de aceitação 0,0048% e não aceitação de 0,0016%.

Categoria - Prova Científica. Maior número de ocorrências desritor: *induction*, com média 377. Com percentual de aceitação 0,79% e não aceitação de 0,12%. Menor número de ocorrências desritor: *abduction*, com média 13. Com percentual de aceitação 0,025% e não aceitação de 0,0064%. Maior percentual de aceitação sobre o total, desritor: *induction*, com média 377. Com percentual de aceitação 0,79% e não aceitação de 0,12%. Menor percentual de aceitação sobre o total, desritor: *abduction*, com média 13. Com percentual de aceitação 0,025% e não aceitação de 0,0064%.

Categoria - Tipo de Validação do Resultado. Maior número de ocorrências desritor: *efficiency*, com média 639. Com percentual de aceitação 2,1% e não aceitação de 0,49%. Menor número de ocorrências desritor: *operationalization*, com média 15. Com percentual de aceitação 0,052% e não aceitação de 0,0064%. Maior percentual de aceitação sobre o total, desritor: *efficiency*, com média 639. Com percentual de aceitação 2,1% e não aceitação de 0,49%. Menor percentual de aceitação sobre o to-

tal, descriptor: *cost-effectiveness*, com média 17. Com percentual de aceitação 0,063% e não aceitação de 0,0056 %.

4.6.5 Análise da cobertura das áreas do conhecimento

No total foram pesquisados 2530 em 68 periódicos da ACM. A Tabela 16 do Apêndice A, mostra os totais individualizados por periódico ordenados pelo total de ocorrências. Note que a média de ocorrências não está ligada ao número de artigos do periódico. A maior média é do Periódico *ACM Trans. Softw. Eng. Methodol.* com 148,5 ocorrências em 19 artigos. Em contrapartida do maior número de ocorrências (522 artigos), do periódico *Proc. ACM Hum.-Comput. Interact.* com média de 53,4 ocorrências. Percentualmente o periódico com maior número de aceites na classificação é o *ACM Trans. Comput. Logic* com 89% em 3 artigos.

Cada observação na base de dados contém informações a respeito do artigo a que se refere. Dentre tais informações está a variável “areas”, a qual contém as áreas as quais fazem parte do escopo de publicação do periódico. O conteúdo desta variável se refere ao periódico e não ao artigo conforme mostra a Tabela 17 do Apêndice A.

Na mesma tabela é possível ver que mais de uma área está agrupada por linha, e uma mesma área está presente em mais de uma linha. Isto se deve ao fato de, como já descrito, a variável “areas” se refere ao periódico e não ao artigo. Tal informação é categorizada segundo a classificação All Science Journal Classification Codes (ASJC)⁴ que é organizada e mantida pela SCOPUS/Elsevier e compreende uma lista das grandes áreas da ciência (da: vida, física, saúde, social e humanas) divididas em 333 sub áreas das quais apenas as listadas foram nesta análise.

Na Tabela 18 do Apêndice A, o conteúdo da variável “areas” foi desmembrado do seu agrupamento. Cada linha na tabela representa a contagem de artigos na respectiva área. Note que o total não é compatível com o número de artigos pesquisados, isto se deve ao fato de que os periódicos são categorizados em mais de uma área, como já mencionado, e há uma sobreposição na contagem.

Como os artigos pesquisados são publicações de periódicos da ACM e a classificação de área fornecida pela API *Crossref* diz respeito à classificação da SCOPUS, foi realizado um exercício para relacionar as classificações de área ASJC e ACM-CCS⁵. Esta última compreende uma ontologia elaborada em 2012 a qual representa um padrão de classificação para a área da Computação. Levando em consideração que no Brasil há iniciativa semelhante elaborada pela SBC, as Comissões Especiais⁶ (CE), as quais são divididas em Grandes Áreas (GA), esta também foi levada em consideração na realização desta análise. Como pode ser visto na Tabela 19 do Apêndice A (asso-

⁴<https://scientificresearch.in/asjc-all-science-journal-classification-codes/>

⁵<https://dl.acm.org/ccs>

⁶<https://www.sbc.org.br/403-comissoes-especiais>

ciação de equivalência das classificações), há uma equivalência entre elas. O que já era esperado uma vez que as três versam sobre a mesma área da ciência.

Com base nesta tabela de equivalência de áreas (ASJC, ACM-CCS e SBC), foi feito o cruzamento dos seus dados com as áreas desmembradas obtidas nas observações dos dados do presente trabalho (Tabela 18 do Apêndice A), uma vez que estes estão distribuídos segundo a classificação da SCOPUS. Note que no que tange a Computação, todas as áreas foram contempladas nos resultados. Ainda é possível notar que o escopo dos periódicos pesquisados não diz respeito apenas a Computação, mas também a outras áreas do conhecimento, conforme mostra a Tabela 11.

4.7 A tabela de aprendizagem do k-NN

A tabela de aprendizagem do algoritmo k-NN apresenta informações sobre as regras de aceite das observações. Não foram consideradas as regras de descarte. No total a tabela de aprendizagem possui 6.949 entradas ao final da execução do algoritmo de classificação. O que corresponde a 5,58% do total de 124.601 observações. A taxa de classificação média para cada regra da tabela de aprendizagem é de 21,8 observações com uma relação de aproximadamente 4,7 aceites para 1 descarte, conforme mostra a Tabela 12.

4.8 Ameaças à Validade

O primeiro e evidente risco aos resultados é a interpretação dada ao conceito por trás do descritor da taxonomia e, consequentemente o treinamento dado ao algoritmo de classificação. Para mitigar este possível viés os conceitos dos descritores foram simplificados ao máximo para facilitar sua compreensão e evitar também ambiguidades. Durante o processo de treinamento e seleção, quando da dúvida se no fragmento de texto onde aconteceu a ocorrência do descritor o conceito estava implícito ou não, um fragmento maior era considerado lendo-se diretamente no artigo analisado. Todavia, ainda podem ocorrer casos de má classificação derivados da interpretação dos conceitos dos descritores.

A escolha da técnica de classificação e seus parâmetros também pode representar um risco aos resultados. A utilização de um “k” diferente do utilizado ($k=1$) certamente implicará em um resultado diferente. Contudo, o k-NN mostra-se bastante robusto e eficaz (BOZKURT et al., 2015; SR et al., 2017; GHOSH et al., 2021). Os resultados obtidos na classificação, percentualmente, são similares à bibliografia.

Tabela 11 – Contagem de artigos por área e associação de áreas segundo SCOPUS, ACM e SBC

Área (SCOPUS)	n	Classif. ACM	Classif. SBC
Applied Mathematics	25		
Artificial Intelligence	147	Computing methodologies	GA 3 Técnicas e Tecnologias de Computação
Biomedical Engineering	4		
Computational Mathematics	20		
Computational Theory and Mathematics	26	Theory of computation	GA 5 Gestão de Dados e de Informações
Computational Theory and Mathematics	26	Mathematics of computing	GA 2 Sistemas de Software
Computer Graphics and Computer-Aided Design	292	Computing methodologies	GA 4 Aplicações da Computação
Computer Graphics and Computer-Aided Design	292	Applied computing	GA 4 Aplicações da Computação
Computer Networks and Communications	852	Networks	GA 1 Sistemas Computacionais
Computer Networks and Communications	852	Security and privacy	GA 1 Sistemas Computacionais
Computer Science (miscellaneous)	86	Social and professional topics	GA 4 Aplicações da Computação
Computer Science Applications	181	Applied computing	GA 4 Aplicações da Computação
Computer Vision and Pattern Recognition		Computing methodologies	GA 4 Aplicações da Computação
Conservation	4		
Control and Optimization	35		
Control and Systems Engineering	35		
Discrete Mathematics and Combinatorics	16		
Economics and Econometrics	17		
Education	50		
Electrical and Electronic Engineering	70		
Experimental and Cognitive Psychology	10		
General Business	10		
General Computer Science	324	General and reference	GA 4 Aplicações da Computação
General Earth and Planetary Sciences	2		
General Environmental Science	2		
General Medicine	1		
Geology	12		
Geometry and Topology	16		
Hardware and Architecture	498	Hardware	GA 1 Sistemas Computacionais
Health Informatics	4		
Health Information Management	4		
Human-Computer Interaction	855	Human-centered computing	GA 3 Técnicas e Tecnologias de Computação
Information Systems	174	Information systems	GA 2 Sistemas de Software
Information Systems and Management			
Logic	7		
Management and Accounting	3		
Management Information Systems	10		
Marketing	17		
Mathematics (miscellaneous)	17		
Media Technology	24		
Medicine (miscellaneous)	10		
Microbiology	4		
Modeling and Simulation	13		
Ocean Engineering	71		
Reliability and Quality	14		
Risk	305		
Safety	305		
Safety Research	1		
Signal Processing	16	Hardware	GA 1 Sistemas Computacionais
Social Sciences (miscellaneous)	522		
Software	623	Software and its engineering	GA 2 Sistemas de Software
Software	623	Hardware	GA 2 Sistemas de Software
Software	623	Computer systems organization	GA 2 Sistemas de Software
Statistics and Probability	17		
Theoretical Computer Science	128		
Total	6196		
Water Science and Technology	12		

Tabela 12 – Distribuição dos valores da Tabela de Aprendizagem do k-NN.

Categoria	n	Sim	Não	Taxa	Média
Ambiente	553	10992	3143	3,5	19,9
Dados	421	7894	709	11,1	18,8
Estratégia de Validação/Coleta de Dados	1504	33381	5641	5,9	22,2
Metodologia	1068	14270	5029	2,8	13,4
Método	2177	43533	11109	3,9	20,0
Natureza	189	13550	1894	7,2	71,7
Propósito paradigma	242	1888	1392	1,4	7,8
Prova Científica Secundária	202	15926	229	69,5	78,8
Prova Científica	77	1307	189	6,9	17,0
Tipo de Validação do Resultado	385	6177	2498	2,5	16,0
Total	6818	148918	31833	4,7	21,8

4.9 Considerações Finais do Capítulo

Neste capítulo foi apresentado o objeto de estudo do presente trabalho: a Taxonomia de Pesquisa em Computação. Como consequência foi apresentado o seu processo de validação de maneira detalhada em relação à aplicabilidade da taxonomia proposta em publicações científicas da Computação, bem como a sua completude no que se refere à Computação como área do conhecimento.

No próximo capítulo são apresentados conceitos e detalhamentos sobre a pesquisa científica, os quais complementam a taxonomia de pesquisa e de maneira conjunta, agregam um importante conjunto de materiais de apoio ao ensino, bem como, à condução da pesquisa científica.

5 COMPLEMENTOS À TAXONOMIA

Embora a taxonomia seja autoexplicativa, para uma melhor compreensão da pesquisa científica bem como contextualização, algumas ideias e conceitos complementares são necessários. Neste capítulo estão destacados temas complementares à taxonomia. As etapas da pesquisa científica são compiladas da bibliografia, descritas e representadas graficamente para melhor visualização e entendimento. O protocolo RODA, embora não seja um resultado direto, é um importante documento de apoio e complementação da taxonomia proposta.

Como já mencionado, uma das causas de problemas nos resultados de publicações científicas em Computação é a falta de conhecimento sobre a pesquisa científica em si. Sob a ótica pedagógica, é necessário contemplar todos os conteúdos e temas no que se refere ao ensino da metodologia científica. Desde o pensamento crítico sobre um determinado problema a ser investigado até a redação do relatório final da pesquisa realizada, passando é claro, pela escolha dos critérios e técnicas científicas a serem adotados na condução da pesquisa.

5.1 Etapas da Pesquisa Científica

A representação gráfica do processo da pesquisa científica normalmente é vista como um fluxograma (KÖCHE, 2016) (PRAJAPATI et al., 2015), representando o fluxo do processo em suas etapas, facultando a existência de laços dentro de algumas etapas. Dando sempre a ideia principal de linearidade.

Uma das mais importantes qualidades da ciência é a contínua autocorreção. Dentro de uma lógica recursiva (DODIG-CRNKOVIC, 2002), onde novas teorias têm origem em problemas já pré-existentes. Desse modo o processo de pesquisa é circular. Começa em um problema e termina em um problema novamente, independente do resultado. Quer ele seja positivo, quer seja negativo, ao final sempre existirão novos questionamentos, novos problemas a serem solucionados e o reinicio do ciclo (MC-GRATH, 1981).

As etapas de uma pesquisa podem compreender segundo (PRAJAPATI et al.,

2015), a definição do problema, revisão de literatura sobre o tema, formulação de hipótese, desenvolvimento, coleta de dados, análise de dados, interpretação dos resultados e escrita do relatório. Em (VAISHNAVI; KUECHLER, 2004), há o aprimoramento das ideias de (MCGRATH, 1981) e a proposição de um ciclo com etapas estabelecidas e o resultado esperado no processo de pesquisa, bem como sua contribuição à produção do conhecimento como um todo. Para (KÖCHE, 2016), há a divisão dentro de um fluxograma em quatro etapas: preparação e delimitação do problema, construção do plano, execução do plano e construção e apresentação do relatório.

Na Figura 11 busca-se uma associação do conceito de representação gráfica adequada. Lógica e circular-recursiva para as etapas da pesquisa científica, bem como contemplar a conceituação a seguir a qual ela sintetiza. Cabe ressaltar que tais etapas compreendem uma especialização à apresentação das etapas do método científico, considerando as necessidades específicas à Computação.

Essas etapas fazem parte de um processo de construção do conhecimento. Olhando sob a ótica de um processo construtivo/evolutivo, as etapas são subsequentes. Todavia não há impedimento de retroceder etapas. Seja para acomodar possíveis falhas e refinar sucessivamente a pesquisa até que ela esteja realmente adequada. Ou para reproduzir um trabalho já realizado ou refinar o tema segundo um aspecto diferente.

A seguir, são descritas as ideias e conceitos percebidos em cada uma das etapas, bem como a fundamentação bibliográfica de maneira individualizada.

1. *Escolha do Tema.* A pergunta inicial determina o que será estudado. Qual aspecto sobre um tema é de interesse. Necessariamente não precisa ser um tema original; pode-se mudar o enfoque sobre um mesmo tema já estudado anteriormente. Possíveis fontes de temas são a vivência, o cotidiano, temas polêmicos, leituras, debates ou discussões. A escolha do tema é a parte mais importante da pesquisa (PRAJAPATI et al., 2015) e deve ser de interesse e intelectualmente compatível com o autor para estudar tal problema (KÖCHE, 2016). Outro fator é a existência ou não de bibliografia sobre o tema escolhido. De acordo com esse fato e a “novidade do tema”, um tipo de pesquisa é mais recomendado que outro.

2. *Revisão de Literatura.* A revisão da literatura procura por aquilo que já foi feito de maneira semelhante, idêntica ou afim. Para realizar uma análise crítica e acomodar o escopo e enfoque que será dado ao trabalho científico (PRAJAPATI et al., 2015). O objetivo principal da revisão de literatura é dar embasamento teórico ao pesquisador (KÖCHE, 2016) e por consequência ao seu trabalho. Também mostra ao pesquisador o contexto científico e nesse sentido evidencia as soluções já desenvolvidas e as lacunas a serem preenchidas (PRAJAPATI et al., 2015). Dentro da revisão de literatura uma boa fonte de temas para pesquisa são os trabalhos futuros ou continuidades propostos nas publicações que compõem a revisão.



Figura 11 – Etapas da Pesquisa Científica

3. *Justificativa*. A justificativa aponta os porquês de estudar o que vai ser estudado. Quais as vantagens, desvantagens, importância dentro do segmento ou área de conhecimento. É onde é mostrada a relevância de investigar o tema proposto dentro do contexto de pesquisa que se pretende estudar (KÖCHE, 2016). Pode apontar para lacunas evidenciadas no processo de revisão de literatura a qual se pretende complementar.

4. *Formulação do Problema*. Quais as perguntas que se pretende responder com a pesquisa baseando-se no contexto das lacunas apontadas pela literatura (PRAJAPATI et al., 2015). Compreendendo sua definição e formulação do enunciado (KÖCHE, 2016). O problema pode ser sucessivamente refinado ou reformulado e deve apontar para as possíveis hipóteses de pesquisa (DODIG-CRNKOVIC, 2002) (PRAJAPATI et al., 2015).

5. *Determinação de Objetivos*. O que se busca alcançar com a pesquisa e quais as etapas (objetivos) se pretende cumprir para testar as hipóteses. É o refinamento da formulação do problema e a derivação das suas hipóteses (PRAJAPATI et al., 2015) em linhas gerais e específicas.

6. *Metodologia*. É o passo a passo para mostrar como a pesquisa será feita para alcançar os objetivos. Todos os processos, etapas e materiais utilizados na obtenção dos resultados (PRAJAPATI et al., 2015). Nesta etapa, aparecem o tipo, o universo e escopo da pesquisa. Como será desenvolvida a técnica proposta. Se será feita a coleta de dados, qual o tamanho amostral e a análise que se projeta realizar sob esses dados. Quais serão os instrumentos de coleta ou qual a origem destes dados, assim como estes serão construídos (KÖCHE, 2016).

7. *Desenvolvimento*. É a etapa de aplicação daquilo que foi desenvolvido. Onde se realiza a coleta de dados e se explica como foi esse processo, através de que meios, por quem, quando e onde (PRAJAPATI et al., 2015). Muitas vezes, é necessária a testagem de toda a metodologia envolvida, bem como as saídas de dados por meio de um estudo piloto (KÖCHE, 2016; PRAJAPATI et al., 2015).

8. *Tabulação dos Dados*. Como foram organizados os dados obtidos. Recursos como índices, cálculos estatísticos, tabelas, quadros e gráficos. Compreende a organização dos dados coletados e toda a sua manipulação até a obtenção do conjunto final para análise (KÖCHE, 2016).

9. *Análise e Discussão dos Resultados*. Como foram analisados os dados coletados e quais os resultados obtidos. Estes confirmam ou refutam as hipóteses anunciamadas. Os testes aplicados, a relação entre as variáveis, a consonância dos tipos de variáveis como a relação que se deseja testar entre estas variáveis (KÖCHE, 2016).

10. *Conclusão da Análise dos Resultados*. A que conclusões se chega por meio dos resultados? Eram os esperados? Qual a aplicabilidade dos resultados obtidos.

11. *Redação e Apresentação do Trabalho Científico*. Por último, a escrita do relatório da pesquisa científica. É a parte da pesquisa voltada para o leitor e tem o objetivo não apenas de comunicar sobre os processos e resultados obtidos (KÖCHE, 2016), mas acima de tudo garantir o preceito científico da reproduzibilidade (MUNAFÒ et al., 2017; TEDRE, 2011b; GHEZZI, 2020; AYASH, 2014). Esse relatório é normalmente escrito segundo o corpo editorial do periódico onde vai ser publicado. Porém, algumas informações referentes à pesquisa não mudam independentemente do lugar onde ela será publicada, e nessa etapa, o protocolo RODA pode ser utilizado como ferramenta de apoio.

5.2 O Protocolo RODA

O protocolo RODA (OSORIO, 2020) é uma iniciativa desenvolvida para auxiliar na concepção, condução e redação de pesquisas científicas em Computação. Essa iniciativa é composta por 26 orientações distribuídas em seis seções estruturais lógicas de uma publicação científica.

A motivação principal do RODA reside nos problemas de comprovação de resulta-

dos científicos em Computação. Estudos evidenciam esses problemas, como mencionado nas Seções 1.1 e 1.2. No entanto, é possível observar que, nos últimos anos, a percepção dos resultados tem mudado. Já não se vê o resultado da mesma maneira que antes; mais confiabilidade e robustez são esperadas. A simples apresentação do resultado não é mais suficiente. É preciso mostrar como foi obtido do início ao fim, de maneira objetiva e transparente, garantindo inovação, reproduzibilidade e disseminação do conhecimento, seguindo os preceitos básicos da ciência.

Um movimento semelhante ocorreu em outras áreas, e a utilização de guias de pesquisa tem se mostrado de grande ajuda nesse panorama de mudança. Há claramente a necessidade de atenção nestes destaques, e a promoção da reflexão sobre a necessidade da inclusão do correto uso do estudo metodológico na comprovação de resultados científicos aplicada à Computação nas disciplinas básicas de formação acadêmica.

Outro aspecto importante a ressaltar é o caráter complementar do RODA em relação à taxonomia. A taxonomia, por si só, é capaz de orientar o projeto de pesquisa e as necessidades na condução. No entanto, ao analisarmos as etapas da pesquisa científica (Figura 11, Seção 5.1), o RODA complementa a taxonomia, pois sua concepção é orientada à última etapa, a Redação e Apresentação.

5.3 Protótipo de Extensão para Navegador

Entendendo o papel que a taxonomia proposta poderá a vir a desempenhar no ensino sobre pesquisa científica em Computação e na identificação de trabalhos relevantes na literatura, foi concebida uma ferramenta para facilitar a categorização de artigos científicos. O protótipo desta ferramenta foi desenvolvido como uma extensão de navegador. Disponível na forma de um plugin com o objetivo de tornar mais acessível e eficiente o processo de classificação de trabalhos acadêmicos.

Para composição do protótipo deste plugin, as regras de aprendizagem obtidas no processo de validação para os dois descritores da categoria *Natureza* foram transpostas, de forma programática, para a linguagem JavaScript. Durante a execução deste plugin, tais regras são avaliadas, verificando sua presença, ou não, no artigo aberto no navegador.

O protótipo de plugin encontra-se inteiramente funcional para o navegador Chrome. Nesta versão protótipo, apenas 194 regras de aprendizagem dos descritores *Pure* e *Applied* da categoria *Nature* estão contemplados. O pacote com esta extensão pode ser recuperado em um repositório público¹.

¹<https://github.com/metasciencestudies/taxonomy>

6 DISCUSSÃO SOBRE OS RESULTADOS

A pesquisa científica está centrada em encontrar e desenvolver o desconhecido, sendo os maiores objetivos da pesquisa moderna a descoberta e a disseminação do conhecimento (SUDHAKAR et al., 2021). Em uma segunda interpretação, o conhecimento científico gerado pela pesquisa não se limita à descoberta de fatos e coisas, mas capacita as pessoas e as permite progredir enquanto sociedade (GHEZZI, 2020). Dentro deste trabalho, esses foram assumidos como os objetivos fundamentais da pesquisa científica.

Na pesquisa científica, a validação de resultados possui uma importante propriedade que deve ser levada em consideração: a reproduzibilidade (GHEZZI, 2020). Qualquer pessoa deve poder, se assim o desejar, reproduzir uma pesquisa nas mesmas condições e obter os mesmos resultados. A repetição é uma premissa do método científico (KOZIOLEK, 2005) e pode ser utilizada para aferir os resultados, outro ponto extremamente importante no contexto da pesquisa científica.

Recentemente, os resultados da pesquisa científica em Computação e seus possíveis problemas de apresentação e validação, estão sendo investigados de maneira mais aprofundada, como discutido em (OSORIO; DIAS; CAVALHEIRO, 2018). Tais problemas passaram a compor a lista de objetos de pesquisa em Computação. Embora não exista até o momento uma linha de pesquisa específica que trate desse tema como objeto de estudo. Essa situação é compreensível, uma vez que a Computação é uma área relativamente nova em comparação com outras disciplinas científicas (GHEZZI, 2020; HASSANI, 2017). Como uma área emergente, a Computação precisa ser estudada epistemologicamente, assim como sua contribuição como disciplina interdisciplinar em outras áreas também precisa ser continuamente examinada.

Ao longo do tempo, tem havido discussões significativas sobre a natureza da Computação e seus objetos de estudo. A questão de que tipo de ciência é realizada em Computação é frequentemente levantada (DODIG-CRNKOVIC, 2023). Ao contrário das ciências naturais, a Computação é considerada uma ciência não natural, artificial, experimental e até mesmo um novo e completo paradigma científico (TEDRE, 2011a; DODIG-CRNKOVIC, 2023). Mais recentemente, o termo “sintético” tem sido utilizado

para tentar definir a área (TEDRE, 2011a).

A diversidade de técnicas de pesquisa em Computação pode gerar problemas de ordem epistemológica e metodológica (DODIG-CRNKOVIC, 2023). Cada técnica possui uma visão epistemológica distinta de seus objetos de estudo, empregando metodologias diferentes e não necessariamente interaplicáveis, resultando em conclusões de naturezas diversas (TEDRE, 2007; DODIG-CRNKOVIC, 2023). Os pontos de vista dos pesquisadores também podem variar significativamente de acordo com sua área ou subárea de atuação (HASSANI, 2017). Portanto, é crucial ter um profundo conhecimento da teoria das técnicas de pesquisa científica, suas divisões e conceitos. Caso contrário, os resultados tendem a ser imperfeitos do ponto de vista metodológico (TEDRE, 2007).

Dentro desse contexto, é evidente a contribuição da iniciativa apresentada neste trabalho. Os resultados indicam a plena aplicabilidade da taxonomia à Computação como área do conhecimento. Os conceitos nela contidos mostram-se representativos da pesquisa realizada na área e estão consistentes com o universo de publicações estudadas.

Além da aplicabilidade os resultados também demonstram a abrangência da taxonomia em relação às áreas de pesquisa em Computação. Inicialmente de acordo com a classificação da SCOPUS todas as áreas relacionadas à Computação estão inseridas nos resultados. Ao comparar essa classificação com as divisões de grandes áreas da SBC e a classificação de área da ACM, percebe-se que todas essas áreas estão contempladas nos resultados. Isso destaca a amplitude e a completude da taxonomia em face das diferentes nuances e temas de pesquisa em Computação.

Além disso, divisões de áreas de outras ciências segundo a classificação da SCOPUS também estão refletidas nos resultados. Isso evidencia a multidisciplinaridade da Computação e reforça a argumentação sobre a inserção da Computação na ciência de modo geral, tanto como meio, sendo utilizada como ferramenta para alcançar resultados em outras áreas do conhecimento, quanto como fim investigando a técnica computacional em si.

É importante ressaltar que dentro dos contextos de inserção da Computação na ciência, não foram encontradas evidências de que a Computação seja um objeto de estudo em si mesma (*meta*). Nenhuma das classificações de área (ACM, SCOPUS e SBC) apresenta uma divisão dentro da Computação onde as técnicas de pesquisa sejam o objeto de estudo. Esse resultado está alinhado com a proposta deste trabalho.

Considerando a multidisciplinaridade da Computação tanto no sentido de incorporar áreas de conhecimento externas quanto no sentido de ser incorporada por outras áreas, é necessário um melhor entendimento da pesquisa em Computação para facilitar e fortalecer o diálogo com outras áreas do conhecimento.

Enquanto há discussões sobre a natureza da pesquisa em Computação, seus re-

sultados, suas técnicas e o rigor metodológico adotado, também é imperativo educar melhor os futuros pesquisadores. Dentro do processo formativo é essencial ensinar aquilo que é evidenciado nas discussões realizadas. Essa necessidade fica evidente quando se percebe que poucos estudos utilizam efetivamente a metodologia adequada e que são raras as pesquisas em Computação que formulam e validam suas hipóteses de maneira iterativa dentro de um ciclo de refinamento (KOZIOLEK, 2005; WAZLAWICK, 2010). Assim, além da preocupação com a formação de novos pesquisadores, existe espaço para a reciclagem das técnicas de pesquisa empregadas por pesquisadores já formados.

As taxonomias desempenham um papel crucial no apoio à compreensão dos educadores sobre o processo de aprendizagem, suas dimensões e o processo formativo (IRVINE, 2021). Elas servem como referencial para a correta classificação e comparação com trabalhos semelhantes (MOTGER; FRANCH; MARCO, 2022), além de serem ponto de partida para estudos epistemológicos sobre a própria ciência em Computação, visando aprimorar e diversificar as pesquisas realizadas na área. Qualificando ainda mais a pesquisa e o ensino em Computação.

7 CONCLUSÃO

Este trabalho apresentou uma taxonomia de pesquisa em Computação composta por uma lista de 96 descritores de técnicas de pesquisa aplicáveis à área divididos em 10 categorias. A proposição foi validada com resultados positivos segundo critérios de aplicabilidade e completude.

A motivação para este trabalho surge da constatação de que várias pesquisas indicam problemas na apresentação de resultados em publicações científicas em Computação. Seja natureza metodológica ou na redação do relatório final da pesquisa, esses problemas são muitas vezes atribuídos à falta de conhecimento, destacando a importância da formação nesse contexto.

Como uma iniciativa para abordar os problemas descritivos do relatório final da pesquisa o protocolo RODA foi desenvolvido. No entanto, os problemas de natureza metodológica não foram totalmente resolvidos. Nesse sentido, a principal contribuição deste trabalho é a taxonomia de pesquisa, que tem uma natureza formativa e servirá como uma ferramenta essencial no ensino sobre pesquisa científica.

Dentro desse contexto educacional, a recomendação da Sociedade Brasileira de Computação (SBC) para a inclusão de Metodologia Científica e Estatística nos cursos de Graduação e Pós-Graduação em Computação destaca a importância desses conhecimentos. Essas recomendações indicam as competências que os estudantes devem desenvolver, com referência à taxonomia de *Bloom* e aos eixos de formação dos cursos na área de Computação.

No entanto as recomendações são em geral, amplas e versam sobre o “uso de métodos de pesquisa apropriados”. Não há uma especificação clara dos conteúdos e uma abordagem detalhada das técnicas científicas necessárias. A taxonomia apresentada neste trabalho preenche essa lacuna, fornecendo uma estrutura detalhada e específica das técnicas de pesquisa em Computação.

A contribuição deste trabalho visa promover a reflexão sobre a necessidade da inclusão de conteúdos relacionados às técnicas científicas para melhorar a qualidade das publicações em Computação e seus resultados ao longo de todo o ciclo de formação dos estudantes de Computação. A taxonomia desempenhará um papel crucial

nesse sentido, e sua adoção como ferramenta de ensino deve ser incentivada.

Além disso, destaca-se a importância de estudar a pesquisa em Computação como uma área de pesquisa em si mesma. A comunidade científica da área deve assumir um papel proativo nesse sentido, ditando e organizando suas diretrizes de pesquisa.

Possíveis perspectivas de desdobramentos futuros incluem o refinamento da taxonomia por meio de novas aferições e a extensão da metodologia de pesquisa utilizada no presente trabalho para outras bases indexadas. Como a do Instituto de Engenheiros Eletricistas e Eletrônicos (IEEE) por exemplo. Outra frente de trabalho seria a investigação individualizada e detalhada de cada descritor em relação às possibilidades específicas de uso na pesquisa em Computação.

A relação de dependência entre os descritores é outra frente de estudo. Estabelecer quais descritores são interdependentes poderia fomentar a elaboração de um fluxograma que auxiliaria no processo de classificação.

A elaboração de um guia visual de maneira individualizada por categoria da taxonomia para utilização em sala de aula. Juntamente com a seleção de um conjunto de materiais didáticos de apoio relacionados à taxonomia será uma contribuição valiosa. Essa iniciativa, combinada com as etapas da pesquisa e o protocolo RODA, poderia oferecer uma abordagem abrangente e profunda para o processo da pesquisa científica.

De grande utilidade tanto para utilização em sala de aula quanto em pesquisas, uma ferramenta de classificação de publicações científicas *on-line* poderia ser desenvolvida. Tal ferramenta aplicaria o modelo de classificação proposto em uma publicação escolhida pelo usuário, resultando na caracterização do mesmo face a taxonomia. O resultado pode ser utilizado em sala de aula como exemplo prático de fácil visualização para o aluno de modo direto, sem subjetividades e sem a necessidade da leitura para evidenciar as técnicas utilizadas, facilitando o trabalho do educador na obtenção dos objetivos pedagógicos. De maneira semelhante a ferramenta pode ser utilizada na pesquisa científica auxiliando pesquisadores a desenvolverem seus referenciais bibliográficos. Principalmente àqueles sistemáticos onde os resultados podem ser utilizados como critérios de inclusão ou descarte. Um passo nesta direção foi dado pelo desenvolvimento de protótipo de uma extensão para o Chrome, a qual permite categorizar artigos abertos no navegador.

Finalmente, uma possível contribuição adicional seria a elaboração de um projeto para a criação junto à Sociedade Brasileira de Computação, de uma Grande Área (GA) ou de uma Comissão Especial (CE) sobre o meta-estudo da pesquisa científica em Computação. Esse seria um passo importante para fortalecer a pesquisa na área e promover o avanço contínuo do conhecimento em Computação.

REFERÊNCIAS

- ADLER, S.; SCHMITT, S.; WOLTER, K.; KYAS, M. A survey of experimental evaluation in indoor localization research. In: INDOOR POSITIONING AND INDOOR NAVIGATION (IPIN), 2015 INTERNATIONAL CONFERENCE ON, 2015, Nova Iorque. **Anais...** IEEE, 2015. p.1–10.
- AYASH, M. Research Methodologies in Computer Science and Information Systems. **Computer Science**, [S.I.], v.2014, p.1–6, 2014.
- BACON, F. **Novum organum**. [S.I.]: Clarendon press, 1878.
- BADIUS, J.; TRECHSEL, J. **Quaestiones et decisiones in quattuor libros Sententiarum Petri Lombardi**: Centilogium theologicum. [S.I.]: Johannes Trechsel, 1495.
- BEZIAU, J.-Y. What is an axiom? **A True Polymath-A Tribute to Francisco Antonio Doria**, College Publications, London, [S.I.], p.122–142, 2020.
- BOZKURT, S.; ELIBOL, G.; GUNAL, S.; YAYAN, U. A comparative study on machine learning algorithms for indoor positioning. In: INTERNATIONAL SYMPOSIUM ON INNOVATIONS IN INTELLIGENT SYSTEMS AND APPLICATIONS (INISTA), 2015., 2015. **Anais...** [S.I.: s.n.], 2015. p.1–8.
- COMPUTER MACHINERY, A. Association for. **ACM - Special Interest Groups (SIGS)**. Disponível em: <<https://www.acm.org/special-interest-groups/alphabetical-listing>>.
- CUNHA, J. H. d. C.; ARAÚJO, R. H. d. Uso da taxonomia como ferramenta de redução da subjetividade nos processos de auditoria. **Perspectivas em Ciência da Informação**, [S.I.], v.24, p.216–231, 2019.
- DESCARTES, R.; CRESS, D. A. **Discourse on method**. [S.I.]: Hackett Publishing, 1998.
- DODIG-CRNKOVIC, G. Scientific methods in computer science. In: CONFERENCE FOR THE PROMOTION OF RESEARCH IN IT AT NEW UNIVERSITIES AND AT

- UNIVERSITY COLLEGES IN SWEDEN, SKÖVDE, SUECIA, 2002. **Proceedings...** [S.I.: s.n.], 2002. p.126–130.
- DODIG-CRNKOVIC, G. On the foundations of computing. computing as the fourth great domain of science. **Global Philosophy**, [S.I.], v.33, n.1, p.16, 2023.
- FERREIRA, L. D. **Técnicas de aprendizado de máquina aplicadas à classificação de estudantes a partir de estilos de aprendizagem**. 2018. Tese (Doutorado em Ciência da Computação) — Universidade de São Paulo.
- GHEZZI, C. What Is Research and Why We Do It. In: **Being a Researcher**. [S.I.]: Springer, 2020. p.1–23.
- GHOSH, M. et al. A Comparative Analysis of Machine Learning Algorithms to Predict Liver Disease. **Intelligent Automation & Soft Computing**, [S.I.], v.30, n.3, 2021.
- HARTMANIS, J. Some observations about the nature of computer science. In: INTERNATIONAL CONFERENCE ON FOUNDATIONS OF SOFTWARE TECHNOLOGY AND THEORETICAL COMPUTER SCIENCE, 1993. **Anais...** [S.I.: s.n.], 1993. p.1–12.
- HASSANI, H. Research methods in computer science: The challenges and issues. **arXiv preprint arXiv:1703.04080**, Nova Iorque, 2017.
- HAZZAN, O. et al. Qualitative research in computer science education. **Acm Sigcse Bulletin**, [S.I.], v.38, n.1, p.408–412, 2006.
- HOLZ, H. J. et al. Research Methods in Computing: What are they, and how should we teach them? In: **Working group reports on ITiCSE on Innovation and technology in computer science education**. [S.I.: s.n.], 2006. p.96–114.
- HUSTADT, U. **Research Methods in Computer Science**. [S.I.]: Citeseer, 2012.
- IRVINE, J. Taxonomies in education: Overview, comparison, and future directions. **Journal of Education and Development**, [S.I.], v.5, n.2, p.1, 2021.
- KÖCHE, J. C. **Fundamentos de metodologia científica**. [S.I.]: Editora Vozes, 2016.
- KODRATOFF, Y. **Introduction to Machine Learning**. [S.I.]: Elsevier Science, 2014.
- KOTSIANTIS, S. B.; ZAHARAKIS, I.; PINTELAS, P. et al. Supervised machine learning: A review of classification techniques. **Emerging artificial intelligence applications in computer engineering**, [S.I.], v.160, n.1, p.3–24, 2007.
- KOZIOLEK, H. The role of experimentation in software engineering. In: SEMINAR “RESEARCH METHODS”, SUMMER TERM, 2005. **Anais...** [S.I.: s.n.], 2005.

- MALTA, M. et al. Iniciativa STROBE: subsídios para a comunicação de estudos observacionais. **Revista de Saúde Pública**, São Paulo, v.44, p.559–565, 2010.
- MCGRATH, J. E. Dilemmatics: The study of research choices and dilemmas. **American Behavioral Scientist**, [S.I.], v.25, n.2, p.179–210, 1981.
- MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. **Sistemas inteligentes-Fundamentos e aplicações**, [S.I.], v.1, n.1, p.32, 2003.
- MOTGER, Q.; FRANCH, X.; MARCO, J. Software-based dialogue systems: survey, taxonomy, and challenges. **ACM Computing Surveys**, [S.I.], v.55, n.5, p.1–42, 2022.
- MUNAFÒ, M. R. et al. A manifesto for reproducible science. **Nature Human Behaviour**, Berlim, v.1, n.1, p.0021, 2017.
- OSORIO, A. **Uma proposta de protocolo de apoio à pesquisa em computação**. 2020. Dissertação (Mestrado em Ciência da Computação) — Universidade Federal de Pelotas.
- OSORIO, A.; DIAS, M.; CAVALHEIRO, G. G. H. **Tangible assets to improve research quality**: a meta analysis case study. [S.I.: s.n.], 2018. 117–132p.
- PRAJAPATI, H. B.; DABHI, V. K.; BHENSDADIA, C.; OTHERS. Taking a Deep Breath before Jumping into Research in Computer Science and Engineering. In: FIFTH INTERNATIONAL CONFERENCE ON ADVANCED COMPUTING & COMMUNICATION TECHNOLOGIES, 2015., 2015. **Anais...** [S.I.: s.n.], 2015. p.283–291.
- PRECHELT, L. A quantitative study of experimental evaluations of neural network learning algorithms: Current research practice. **IEEE Transactions on Neural Networks**, Pensilvânia, v.6, 1994.
- PSCHEIDT, M. **Appropriate information system development**: A methodology for sustainable cross-cultural information system production and use. [S.I.]: [SI: sn], 2012.
- RAMESH, V.; GLASS, R. L.; VESSEY, I.; OTHERS. Research in computer science: an empirical study. **Journal of systems and software**, [S.I.], v.70, n.1-2, p.165–176, 2004.
- RASCHKA, S. **Machine Learning Lecture Notes**. Disponível em: <<https://sebastianraschka.com/teaching/stat479-fs2018/>>.
- ROSSI, R. G. **Classificação automática de textos por meio de aprendizado de máquina baseado em redes**. 2016. Tese (Doutorado em Ciência da Computação) — Universidade de São Paulo.

SEN, P. C.; HAJRA, M.; GHOSH, M. Supervised classification algorithms in machine learning: A survey and review. In: EMERGING TECHNOLOGY IN MODELLING AND GRAPHICS: PROCEEDINGS OF IEM GRAPH 2018, 2020. **Anais...** [S.I.: s.n.], 2020. p.99–111.

SOUSA, R. T. B. de; ARAÚJO JÚNIOR, R. H. de. A classificação e a taxonomia como instrumentos efetivos para a recuperação da informação arquivística. **Ciência da informação**, [S.I.], v.42, n.1, 2013.

SR, M. N. et al. Comparative study of machine learning approach on Malay translated hadith text classification based on Sanad. In: MATEC WEB OF CONFERENCES, 2017. **Anais...** [S.I.: s.n.], 2017.

SUDHAKAR, G. P.; TECH, M.; MIMA, A.; OTHERS. Scientific Research Methodology Vs. Social Science Research Methodology. , [S.I.], 2021.

TEDRE, M. Know your discipline: Teaching the philosophy of computer science. **Journal of Information Technology Education: Research**, [S.I.], v.6, n.1, p.105–122, 2007.

TEDRE, M. Computing as a science: A survey of competing viewpoints. **Minds and Machines**, [S.I.], v.21, n.3, p.361–387, 2011.

TEDRE, M. The Social Study of Computer Science. In: **Virtual Communities**: Concepts, Methodologies, Tools and Applications. [S.I.]: IGI Global, 2011. p.336–352.

TEDRE, M.; MOISSEINEN, N. Experiments in computing: A survey. **The Scientific World Journal**, Londres, v.2014, 2014.

TICHY, W. F.; LUKOWICZ, P.; PRECHELT, L.; HEINZ, E. A. Experimental evaluation in computer science: A quantitative study. **Journal of Systems and Software**, Amsterdā, v.28, n.1, p.9–18, 1995.

VAISHNAVI, V.; KUECHLER, W. Design science research in information systems. **January**, <http://www.desrist.org/design-research-in-information-systems/>, v.20, p.2004, 2004.

VESSEY, I.; RAMESH, V.; GLASS, R. L. A unified classification system for research in the computing disciplines. **Information and Software Technology**, [S.I.], v.47, n.4, p.245–255, 2005.

WAINER, J.; BARSOTTINI, C. G. N.; LACERDA, D.; MARCO, L. R. M. de. Empirical evaluation in Computer Science research published by ACM. **Information and Software Technology**, Amesterdā, v.51, n.6, p.1081–1085, 2009.

WAINER, J. et al. Métodos de pesquisa quantitativa e qualitativa para a Ciência da Computação. **Atualização em informática**, Rio de Janeiro, v.1, p.221–262, 2007.

WAZLAWICK, R. S. Uma reflexão sobre a pesquisa em ciência da computação à luz da classificação das ciências e do método científico. **Revista de Sistemas de Informação da FSMA**, [S.I.], v.6, p.3–10, 2010.

APÊNDICE A RELATÓRIO DA ANÁLISE DE DADOS

Análise de Dados

Alessander Osorio

07/09/2023

Sumarizações iniciais

A presente análise diz respeito aos dados oriundos do processo de validação da taxonomia proposta segundo a ontologia dela derivada. Tal ontologia, composta por 89 descritores divididos em 10 categorias, foi pesquisada no texto de 2508 artigos de 68 periódicos da ACM, totalizando 124601 observações. Estas foram submetidas ao algoritmo semisupervisionado de classificação k-nn para verificar sua adequação aos conceitos referentes a cada descritor levantados pela revisão sistemática de literatura.

Ao final, foram considerados aceitos como dentro do contexto do conceito a que se refere, 91419 observações representando 73,37% do total. Em contraposição, dentro dos mesmos padrões, 33182 observações não foram aceitas, as quais representam 26,63% do total, conforme a Tabela 1.

As Tabelas 2 e 3, mostram os totais de ocorrências por Categoria e Descritor respectivamente. Note que dado o número total de observações, alguns descritores percentualmente apresentam valores não passíveis de representação matemática com o número de casas decimais utilizadas. Todavia não é possível desconsiderar tais valores pois estes compõem o total da categoria

A variável de desfecho analisada é de natureza categorica, tendo o valor “Sim” para os aceitos, e “Não” para os não aceitos. Os valores apresentados aqui, representam a contagem de cada uma delas individualizada. As categorias e seus descritores serão analisados individualmente sem procurar estabelecer uma relação entre elas. Os histogramas por Categoria e Descritor mostram que os dados não seguem a distribuição normal, sendo o teste do Qui-Quadrado o recomendado para verificação estatística dos valores desta análise.

Table 1: Totais Gerais Aceite

Aceite	n	%
Não	33182	27
Sim	91419	73

Table 2: Totais por Categoria dos Descritores

Categoria	n	%
Ambiente	11693	9,4
Dados	5576	4,5
Estratégia de Validação/Coleta de Dados	26014	20,9
Método	40292	32,3
Metodologia	15261	12,2
Natureza	7380	5,9
Propósito paradigma	2776	2,2
Prova Científica Secundária	6977	5,6
Prova Científica	1234	1,0

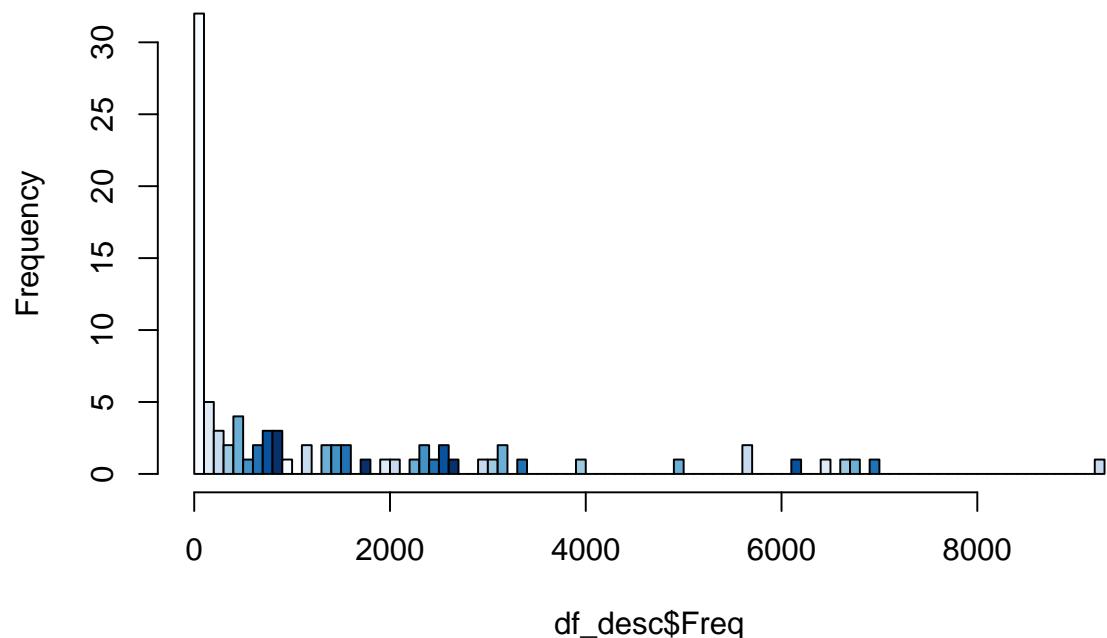
Categoría	n	%
Tipo de Validação do Resultado	7398	5,9

Table 3: Totais por Descritores

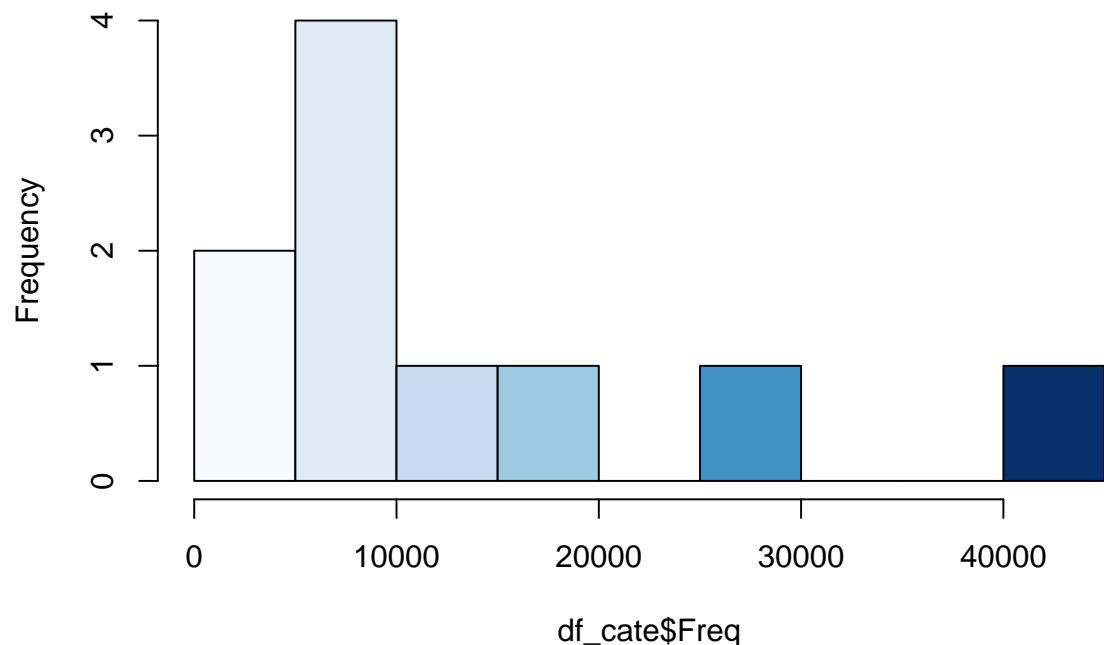
Descriptor	n	%
abduction	39	0,0
action research	72	0,1
analytical	832	0,7
applied	6416	5,1
behavioral	2202	1,8
benchmark	2957	2,4
case-control	6	0,0
case study	1395	1,1
causal	1589	1,3
cause-and-effect	8	0,0
community research	10	0,0
comparative	703	0,6
concept	2531	2,0
conceptual	1171	0,9
conceptual analysis	6	0,0
constructive	401	0,3
control model	30	0,0
controlled model	1	0,0
cost-effectiveness	86	0,1
critique	400	0,3
cross-section	46	0,0
cross-sectional	75	0,1
deduction	64	0,1
descriptive	765	0,6
design science	37	0,0
doctrinal	1	0,0
dynamic analysis	273	0,2
effectiveness	2499	2,0
efficiency	3194	2,6
empiric	14	0,0
engineering	5658	4,5
evaluation	9298	7,5
evaluative	62	0,0
experiment	3978	3,2
experimental	4924	4,0
exploratory	715	0,6
formal	3383	2,7
grounded theory	237	0,2
hermeneutics	2	0,0
historic	131	0,1
historical	1365	1,1
in vitro	19	0,0
in vivo	47	0,0
induction	1131	0,9
instrument application	1	0,0

Descriptor	n	%
interpretive	137	0,1
interview	2073	1,7
interviews	2306	1,9
investigate	2397	1,9
investigative	98	0,1
longitudinal	536	0,4
math	497	0,4
math test	17	0,0
meta-analysis	150	0,1
modeling	5681	4,6
normative	432	0,3
observational	483	0,4
operationalization	73	0,1
phenomenology	29	0,0
pilot	870	0,7
pilot study	179	0,1
problem-oriented	1	0,0
problem-solving	234	0,2
process analysis	4	0,0
processual	3	0,0
proof	6969	5,6
protocol analysis	3	0,0
prototype	1452	1,2
pure	964	0,8
qualitative	3153	2,5
quantitative	1762	1,4
real	6799	5,5
real-world	3028	2,4
real world	304	0,2
rebuttal	8	0,0
replication	632	0,5
scientific	1916	1,5
secondary data	18	0,0
simulated	1496	1,2
simulation	6111	4,9
speech analysis	4	0,0
standards	863	0,7
static analysis	650	0,5
survey	6692	5,4
systematic literature review	135	0,1
theoretical	2502	2,0
usability	1546	1,2
verticals	4	0,0
workload	2616	2,1

Histograma segundo ocorrências por Descritor



Histograma segundo ocorrências por Categoria



Teste Global dos aceitos/não aceitos

Esta seção mostra a aplicação dos testes estatísticos Qui-Quadrado e Teste de Proporcionalidade para toda a população das observações. O p-valor encontrado em ambos (0.0000000000000001) é inferior a 0,01 ($p<0,01$), o que rejeita a Hipótese Nula e aceita a Hipótese Alternativa para os dados analisados. Individualmente foi aplicado para cada categoria obtendo-se resultado semelhante.

Chi-squared test for given probabilities

```
data: tabela_contingencia
X-squared = 27219, df = 1, p-value <2e-16
```

Não	Sim
33182	91419

```
[1] "Categoria: Ambiente qui-quadrado p-valor: 0"
[1] "Categoria: Dados qui-quadrado p-valor: 0"
[1] "Categoria: Estratégia de Validação/Coleta de Dados qui-quadrado p-valor: 0"
[1] "Categoria: Método qui-quadrado p-valor: 0"
[1] "Categoria: Metodologia qui-quadrado p-valor: 0"
[1] "Categoria: Natureza qui-quadrado p-valor: 0"
[1] "Categoria: Propósito paradigma qui-quadrado p-valor: 0,14"
```

```
[1] "Categoria: Prova Científica Secundária qui-quadrado p-valor: 0"
[1] "Categoria: Prova Científica qui-quadrado p-valor: 0"
[1] "Categoria: Tipo de Validação do Resultado qui-quadrado p-valor: 0"
```

Teste Individual dos aceitos/não aceitos e descriptor individualmente

```
[1] "Descriptor abduction qui-quadrado p-valor: 0,00023"
[1] "Descriptor action research qui-quadrado p-valor: 0,00041"
[1] "Descriptor analytical qui-quadrado p-valor: 0"
[1] "Descriptor applied qui-quadrado p-valor: 0"
[1] "Descriptor behavioral qui-quadrado p-valor: 0,061"
[1] "Descriptor benchmark qui-quadrado p-valor: 0"
[1] "Descriptor case-control qui-quadrado p-valor: Sem ocorrência de uma das categorias"
[1] "Descriptor case study qui-quadrado p-valor: 0"
[1] "Descriptor causal qui-quadrado p-valor: 0"
[1] "Descriptor cause-and-effect qui-quadrado p-valor: Sem ocorrência de uma das categorias"
[1] "Descriptor community research qui-quadrado p-valor: 0,53"
[1] "Descriptor comparative qui-quadrado p-valor: 0"
[1] "Descriptor concept qui-quadrado p-valor: 0"
[1] "Descriptor conceptual qui-quadrado p-valor: 0"
[1] "Descriptor conceptual analysis qui-quadrado p-valor: Sem ocorrência de uma das categorias"
[1] "Descriptor constructive qui-quadrado p-valor: 0"
[1] "Descriptor control model qui-quadrado p-valor: 0,00026"
[1] "Descriptor controlled model qui-quadrado p-valor: Sem ocorrência de uma das categorias"
[1] "Descriptor cost-effectiveness qui-quadrado p-valor: 0"
[1] "Descriptor critique qui-quadrado p-valor: 0"
[1] "Descriptor cross-section qui-quadrado p-valor: 0,24"
[1] "Descriptor cross-sectional qui-quadrado p-valor: 0"
[1] "Descriptor deduction qui-quadrado p-valor: 0,62"
[1] "Descriptor descriptive qui-quadrado p-valor: 0"
[1] "Descriptor design science qui-quadrado p-valor: 0,0052"
[1] "Descriptor doctrinal qui-quadrado p-valor: Sem ocorrência de uma das categorias"
[1] "Descriptor dynamic analysis qui-quadrado p-valor: 0,00005"
[1] "Descriptor effectiveness qui-quadrado p-valor: 0"
[1] "Descriptor efficiency qui-quadrado p-valor: 0"
[1] "Descriptor empiric qui-quadrado p-valor: 0,033"
[1] "Descriptor engineering qui-quadrado p-valor: 0"
[1] "Descriptor evaluation qui-quadrado p-valor: 0"
[1] "Descriptor evaluative qui-quadrado p-valor: 0,13"
[1] "Descriptor experiment qui-quadrado p-valor: 0"
[1] "Descriptor experimental qui-quadrado p-valor: 0"
[1] "Descriptor exploratory qui-quadrado p-valor: 0"
[1] "Descriptor formal qui-quadrado p-valor: 0"
[1] "Descriptor grounded theory qui-quadrado p-valor: 0"
[1] "Descriptor hermeneutics qui-quadrado p-valor: 1"
[1] "Descriptor historic qui-quadrado p-valor: 0"
[1] "Descriptor historical qui-quadrado p-valor: 0"
[1] "Descriptor in vitro qui-quadrado p-valor: 0,49"
[1] "Descriptor in vivo qui-quadrado p-valor: 0,013"
[1] "Descriptor induction qui-quadrado p-valor: 0"
[1] "Descriptor instrument application qui-quadrado p-valor: Sem ocorrência de uma das categorias"
[1] "Descriptor interpretive qui-quadrado p-valor: 0"
[1] "Descriptor interview qui-quadrado p-valor: 0"
```

```

[1] "Descriptor interviews qui-quadrado p-valor: 0"
[1] "Descriptor investigate qui-quadrado p-valor: 0"
[1] "Descriptor investigative qui-quadrado p-valor: 0,69"
[1] "Descriptor longitudinal qui-quadrado p-valor: 0"
[1] "Descriptor math qui-quadrado p-valor: 0"
[1] "Descriptor math test qui-quadrado p-valor: 0,47"
[1] "Descriptor meta-analysis qui-quadrado p-valor: 0"
[1] "Descriptor modeling qui-quadrado p-valor: 0,00002"
[1] "Descriptor normative qui-quadrado p-valor: 0,021"
[1] "Descriptor observational qui-quadrado p-valor: 0"
[1] "Descriptor operationalization qui-quadrado p-valor: 0"
[1] "Descriptor phenomenology qui-quadrado p-valor: 0,0053"
[1] "Descriptor pilot qui-quadrado p-valor: 0"
[1] "Descriptor pilot study qui-quadrado p-valor: 0"
[1] "Descriptor problem-oriented qui-quadrado p-valor: Sem ocorrência de uma das categorias"
[1] "Descriptor problem-solving qui-quadrado p-valor: 0,026"
[1] "Descriptor process analysis qui-quadrado p-valor: Sem ocorrência de uma das categorias"
[1] "Descriptor processual qui-quadrado p-valor: 0,56"
[1] "Descriptor proof qui-quadrado p-valor: 0"
[1] "Descriptor protocol analysis qui-quadrado p-valor: Sem ocorrência de uma das categorias"
[1] "Descriptor prototype qui-quadrado p-valor: 0"
[1] "Descriptor pure qui-quadrado p-valor: 0"
[1] "Descriptor qualitative qui-quadrado p-valor: 0"
[1] "Descriptor quantitative qui-quadrado p-valor: 0"
[1] "Descriptor real qui-quadrado p-valor: 0"
[1] "Descriptor real-world qui-quadrado p-valor: 0"
[1] "Descriptor real world qui-quadrado p-valor: 0,3"
[1] "Descriptor rebuttal qui-quadrado p-valor: 0,16"
[1] "Descriptor replication qui-quadrado p-valor: 0"
[1] "Descriptor scientific qui-quadrado p-valor: 0"
[1] "Descriptor secondary data qui-quadrado p-valor: 0,00016"
[1] "Descriptor simulated qui-quadrado p-valor: 0"
[1] "Descriptor simulation qui-quadrado p-valor: 0"
[1] "Descriptor speech analysis qui-quadrado p-valor: 0,32"
[1] "Descriptor standards qui-quadrado p-valor: 0"
[1] "Descriptor static analysis qui-quadrado p-valor: 0"
[1] "Descriptor survey qui-quadrado p-valor: 0"
[1] "Descriptor systematic literature review qui-quadrado p-valor: 0"
[1] "Descriptor theoretical qui-quadrado p-valor: 0"
[1] "Descriptor usability qui-quadrado p-valor: 0"
[1] "Descriptor verticals qui-quadrado p-valor: Sem ocorrência de uma das categorias"
[1] "Descriptor workload qui-quadrado p-valor: 0"

```

Resultados por descritores da ontologia

Os resultados expressos na tabela abaixo representam o número de ocorrências obtidas para cada descritor da ontologia de maneira individual, sem qualquer tipo de associação entre eles. Os números também correspondem ao total de aceitos (ordenação da tabela juntamente com a categoria) e não aceitos na classificação pelo algoritmo k-nn. Os números são puros e os percentuais foram calculados com base no total do descritor (%Desc) e ao total geral de ocorrências (%Tot). Novamente note que alguns percentuais não puderam ser representados matematicamente com o número de casas decimais utilizados nesta análise. Reitera-se que tais resultados não foram descartados por comporem os resultados totais por categoria do descritor.

O destaque da tabela cabe à última coluna - Pred. - Predominância do resultado. Ela expressa a comparação

dos resultados aceitos e não aceitos. Se um é maior que o outro. Se o 'n' de aceitos é maior que o de não aceitos na coluna será mostrado o conteúdo 'S' caso contrário 'N'. Analisando esta coluna pode-se ver que os descritores 'verticals', 'controled model' e 'doctrinal' os quais, além da baixa ocorrência, não tiveram nenhum caso de classificação positiva. Indicando sua não utilização dentro da pesquisa em computação para a população estudada.

Ainda analisando a coluna predominância, pode-se perceber que quase a totalidade, daqueles descritores onde predominam os não aceitos na classificação, os percentuais em relação ao total de ocorrências são inferiores a 1% (0,6% para ser exato). Excetuando os descritores 'modeling' (4,7%), 'concept' (2,1%) e 'effectiveness' (2,07%). Analisando individualmente estes casos nota-se que as diferenças entre aceitos e não aceitos, são relativamente pequenas, quase inexistentes não representando mais que 1,5% em relação ao total do descritor.

Table 4: Aceite por descritores da ontologia

Categoría	Descriptor	n	%	Total	Sim	%Desc.	%Tot.	Total	Não	%Desc.	%Tot.	Pred.
Ambiente	real	6799	5,46	4718	69	3,79		2081	30,6	1,67	S	
Ambiente	real-world	3028	2,43	2338	77	1,88		690	22,8	0,55	S	
Ambiente	simulated	1496	1,20	1167	78	0,94		329	22,0	0,26	S	
Ambiente	real world	304	0,24	161	53	0,13		143	47,0	0,11	S	
Ambiente	in vivo	47	0,04	32	68	0,03		15	31,9	0,01	S	
Ambiente	in vitro	19	0,02	8	42	0,01		11	57,9	0,01	N	
Dados	qualitative	3153	2,53	2995	95	2,40		158	5,0	0,13	S	
Dados	quantitative	1762	1,41	1403	80	1,13		359	20,4	0,29	S	
Dados	longitudinal	536	0,43	365	68	0,29		171	31,9	0,14	S	
Dados	cross-sectional	75	0,06	59	79	0,05		16	21,3	0,01	S	
Dados	cross-section	46	0,04	19	41	0,02		27	58,7	0,02	N	
Dados	verticals	4	0,00	0	0	0,00		4	100,0	0,00	N	
Estratégia de Validação/Coleta de Dados	survey	6692	5,37	6082	91	4,88		610	9,1	0,49	S	
Estratégia de Validação/Coleta de Dados	experiment	3978	3,19	2463	62	1,98		1515	38,1	1,22	S	
Estratégia de Validação/Coleta de Dados	workload	2616	2,10	2080	80	1,67		536	20,5	0,43	S	
Estratégia de Validação/Coleta de Dados	benchmark	2957	2,37	2006	68	1,61		951	32,2	0,76	S	
Estratégia de Validação/Coleta de Dados	interviews	2306	1,85	1876	81	1,51		430	18,6	0,35	S	
Estratégia de Validação/Coleta de Dados	interview	2073	1,66	1673	81	1,34		400	19,3	0,32	S	
Estratégia de Validação/Coleta de Dados	case study	1395	1,12	1236	89	0,99		159	11,4	0,13	S	
Estratégia de Validação/Coleta de Dados	pilot	870	0,70	773	89	0,62		97	11,2	0,08	S	
Estratégia de Validação/Coleta de Dados	comparative	703	0,56	565	80	0,45		138	19,6	0,11	S	
Estratégia de Validação/Coleta de Dados	static analysis	650	0,52	389	60	0,31		261	40,1	0,21	S	
Estratégia de Validação/Coleta de Dados	replication	632	0,51	253	40	0,20		379	60,0	0,30	N	
Estratégia de Validação/Coleta de Dados	grounded theory	237	0,19	184	78	0,15		53	22,4	0,04	S	
Estratégia de Validação/Coleta de Dados	dynamic analysis	273	0,22	170	62	0,14		103	37,7	0,08	S	
Estratégia de Validação/Coleta de Dados	pilot study	179	0,14	145	81	0,12		34	19,0	0,03	S	
Estratégia de Validação/Coleta de Dados	meta-analysis	150	0,12	132	88	0,11		18	12,0	0,01	S	
Estratégia de Validação/Coleta de Dados	action research	72	0,06	51	71	0,04		21	29,2	0,02	S	
Estratégia de Validação/Coleta de Dados	phenomenology	29	0,02	22	76	0,02		7	24,1	0,01	S	
Estratégia de Validação/Coleta de Dados	secondary data	18	0,01	17	94	0,01		1	5,6	0,00	S	
Estratégia de Validação/Coleta de Dados	systematic literature review	135	0,11	15	11	0,01		120	88,9	0,10	N	
Estratégia de Validação/Coleta de Dados	math test	17	0,01	7	41	0,01		10	58,8	0,01	N	
Estratégia de Validação/Coleta de Dados	case-control	6	0,00	6	100	0,00		0	0,0	0,00	S	
Estratégia de Validação/Coleta de Dados	conceptual analysis	6	0,00	6	100	0,00		0	0,0	0,00	S	
Estratégia de Validação/Coleta de Dados	community research	10	0,01	4	40	0,00		6	60,0	0,00	N	
Estratégia de Validação/Coleta de Dados	protocol analysis	3	0,00	3	100	0,00		0	0,0	0,00	S	
Estratégia de Validação/Coleta de Dados	speech analysis	4	0,00	3	75	0,00		1	25,0	0,00	S	
Estratégia de Validação/Coleta de Dados	hermeneutics	2	0,00	1	50	0,00		1	50,0	0,00	S	
Estratégia de Validação/Coleta de Dados	instrument application	1	0,00	1	100	0,00		0	0,0	0,00	S	
Metodología	simulation	6111	4,90	4182	68	3,36		1929	31,6	1,55	S	
Metodología	modeling	5681	4,56	2681	47	2,15		3000	52,8	2,41	N	
Metodología	prototype	1452	1,17	1388	96	1,11		64	4,4	0,05	S	
Metodología	historical	1365	1,10	1270	93	1,02		95	7,0	0,08	S	
Metodología	observational	483	0,39	422	87	0,34		61	12,6	0,05	S	
Metodología	control model	30	0,02	25	83	0,02		5	16,7	0,00	S	
Metodología	historic	131	0,11	16	12	0,01		115	87,8	0,09	N	
Metodología	process analysis	4	0,00	4	100	0,00		0	0,0	0,00	S	
Metodología	processual	3	0,00	1	33	0,00		2	66,7	0,00	N	
Metodología	controlled model	1	0,00	0	0	0,00		1	100,0	0,00	N	
Método	evaluation	9298	7,46	8409	90	6,75		889	9,6	0,71	S	
Método	experimental	4924	3,95	3919	80	3,15		1005	20,4	0,81	S	
Método	engineering	5658	4,54	3550	63	2,85		2108	37,3	1,69	S	
Método	investigate	2397	1,92	2247	94	1,80		150	6,3	0,12	S	

Categoria	Descriptor	n	%	Total Sim			Total Não			%Desc.	%Tot.	Pred.
				%Desc.	%Tot.	%Desc.	%Tot.	%Desc.	%Tot.			
Método	formal	3383	2,72	2022	60	1,62	1361	40,2	1,09	S		
Método	theoretical	2502	2,01	1515	61	1,22	987	39,5	0,79	S		
Método	causal	1589	1,28	1369	86	1,10	220	13,8	0,18	S		
Método	scientific	1916	1,54	1281	67	1,03	635	33,1	0,51	S		
Método	conceptual	1171	0,94	1095	94	0,88	76	6,5	0,06	S		
Método	concept	2531	2,03	775	31	0,62	1756	69,4	1,41	N		
Método	analytical	832	0,67	667	80	0,54	165	19,8	0,13	S		
Método	exploratory	715	0,57	536	75	0,43	179	25,0	0,14	S		
Método	descriptive	765	0,61	482	63	0,39	283	37,0	0,23	S		
Método	normative	432	0,35	240	56	0,19	192	44,4	0,15	S		
Método	math	497	0,40	164	33	0,13	333	67,0	0,27	N		
Método	problem-solving	234	0,19	134	57	0,11	100	42,7	0,08	S		
Método	standards	863	0,69	130	15	0,10	733	84,9	0,59	N		
Método	constructive	401	0,32	91	23	0,07	310	77,3	0,25	N		
Método	investigative	98	0,08	51	52	0,04	47	48,0	0,04	S		
Método	evaluative	62	0,05	37	60	0,03	25	40,3	0,02	S		
Método	empiric	14	0,01	11	79	0,01	3	21,4	0,00	S		
Método	cause-and-effect	8	0,01	8	100	0,01	0	0,0	0,00	S		
Método	problem-oriented	1	0,00	1	100	0,00	0	0,0	0,00	S		
Método	doctrinal	1	0,00	0	0	0,00	1	100,0	0,00	N		
Natureza	applied	6416	5,15	4807	75	3,86	1609	25,1	1,29	S		
Natureza	pure	964	0,77	608	63	0,49	356	36,9	0,29	S		
Propósito paradigma	behavioral	2202	1,77	1145	52	0,92	1057	48,0	0,85	S		
Propósito paradigma	interpretive	137	0,11	100	73	0,08	37	27,0	0,03	S		
Propósito paradigma	critique	400	0,32	77	19	0,06	323	80,8	0,26	N		
Propósito paradigma	design science	37	0,03	27	73	0,02	10	27,0	0,01	S		
Prova Científica Secundária	proof	6969	5,59	6739	97	5,41	230	3,3	0,18	S		
Prova Científica Secundária	rebuttal	8	0,01	6	75	0,00	2	25,0	0,00	S		
Prova Científica	induction	1131	0,91	979	87	0,79	152	13,4	0,12	S		
Prova Científica	deduction	64	0,05	34	53	0,03	30	46,9	0,02	S		
Prova Científica	abduction	39	0,03	31	79	0,02	8	20,5	0,01	S		
Tipo de Validação do Resultado	efficiency	3194	2,56	2586	81	2,08	608	19,0	0,49	S		
Tipo de Validação do Resultado	usability	1546	1,24	1345	87	1,08	201	13,0	0,16	S		
Tipo de Validação do Resultado	effectiveness	2499	2,01	640	26	0,51	1859	74,4	1,49	N		
Tipo de Validação do Resultado	cost-effectiveness	86	0,07	79	92	0,06	7	8,1	0,01	S		
Tipo de Validação do Resultado	operationalization	73	0,06	65	89	0,05	8	11,0	0,01	S		
NA	Total	124601	100,00	91419	NA	73,37	33182	NA	26,63	NA		

Resultados por categoria da ontologia

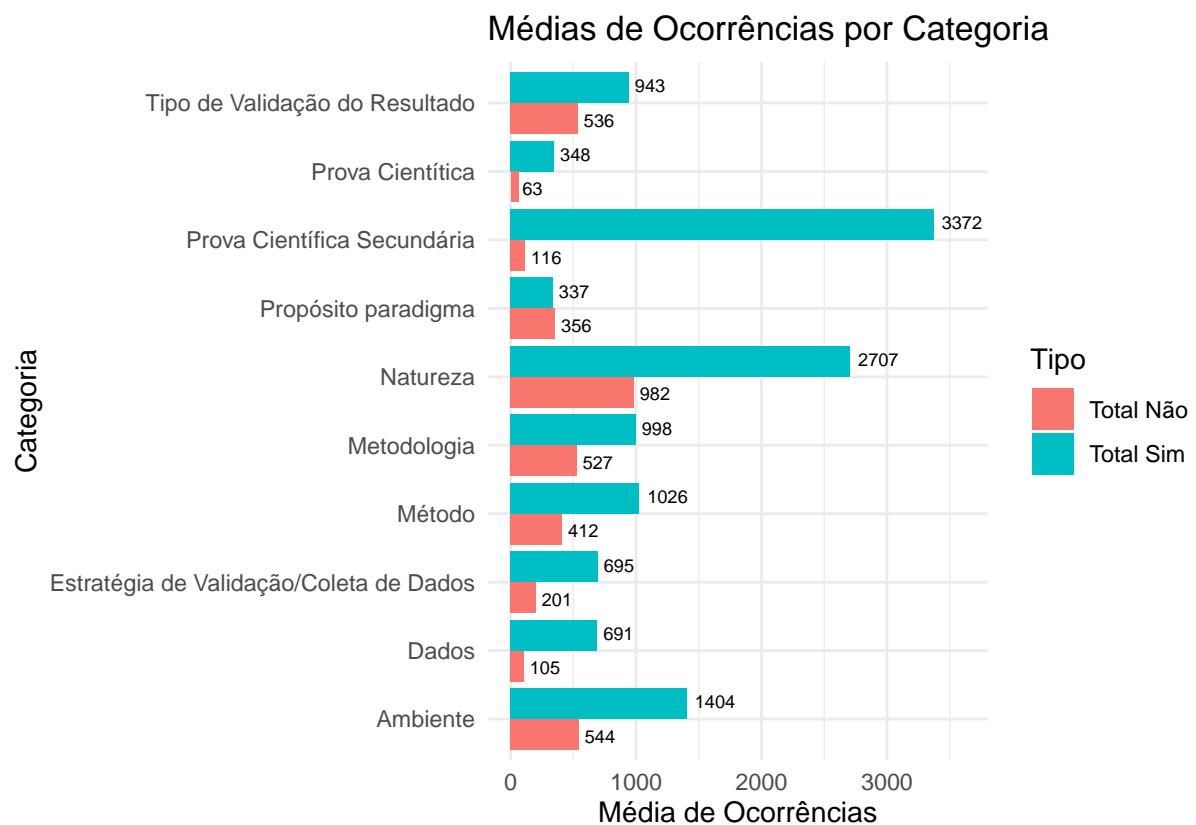
Para os resultados por categoria da ontologia não considerou-se o ‘n’ puro, uma vez que as categorias possuem, internamente, números distintos de descritores. Isto faria com que uma categoria com maior número de descritores, logicamente, tivesse maior número de ocorrências e por consequencia valores enviezados em suas contagens. Nesta parte da análise foram considerados os número de descritores por categoria, sendo o número de ocorrências dividido por tais valores componto assim um valor para ‘n’ normalizado ou a média interna por categoria.

Exemplificando, percebe-se que a categoria com maior ocorrências brutas é ‘Método’ com um total de 39.429 ocorrências e 28 descritores. Ao se normalizar este valor obtém-se um ‘n’ normalizado de 1408. Comparando-se com a categoria ‘Natureza’, esta possuir apenas 2 descritores resultando em um ‘n’ de 3690. Note que o objetivo desta análise não é estabelecer uma comparação entre as categorias, tal ação é apenas para exemplificar e justificar a forma como os números foram analisados.

Dito isto, os demais dados da tabela foram calculados levando-se em consideração o ‘n’ normalizado ou média por categoria. Nesta tabela não foi calculada a predominância dos valores, uma vez que, exceto a categoria ‘Propósito paradigma’, todas as demais possuem predominância positiva. A categoria em questão possui uma diferença entre os valores para aceito e não aceitos percentualmente nula inferior a 0,01% ou 19 ocorrências.

Table 5: Aceito por categorias e descritores da ontologia

Categoria	N.Desc.	n	%	n.Norm.	Total Sim			Total Não			%NCat.	%NTot.
					%SCat.	%STot.	Total	%N	%T			
Ambiente	6	11693	9,38	1949	1404	12,0	1,13	544	4,65	0,44		
Dados	7	5576	4,48	797	691	12,4	0,55	105	1,88	0,08		
Estratégia de Validação/Coleta de Dados	29	26014	20,88	897	695	2,7	0,56	201	0,77	0,16		
Metodologia	10	15261	12,25	1526	998	6,5	0,80	527	3,45	0,42		
Método	28	40292	32,34	1439	1026	2,5	0,82	412	1,02	0,33		
Natureza	2	7380	5,92	3690	2707	36,7	2,17	982	13,31	0,79		
Propósito paradigma	4	2776	2,23	694	337	12,1	0,27	356	12,82	0,29		
Prova Científica Secundária	2	6977	5,60	3488	3372	48,3	2,71	116	1,66	0,09		
Prova Científica	3	1234	0,99	411	348	28,2	0,28	63	5,11	0,05		
Tipo de Validação do Resultado	5	7398	5,94	1480	943	12,8	0,76	536	7,25	0,43		
Total	96	124601	100,00	1298	12521	NA	10,05	3842	NA	3,08		



Resultados individuais por categoria

A seguir, as categorias são sumarizadas individualmente e seus destaques apresentados.

Categoria - Ambiente

Maior número de ocorrências descritor: real , com média 1133 . Com percentual de aceitação 3,8 % e não aceitação de 1,7 %.

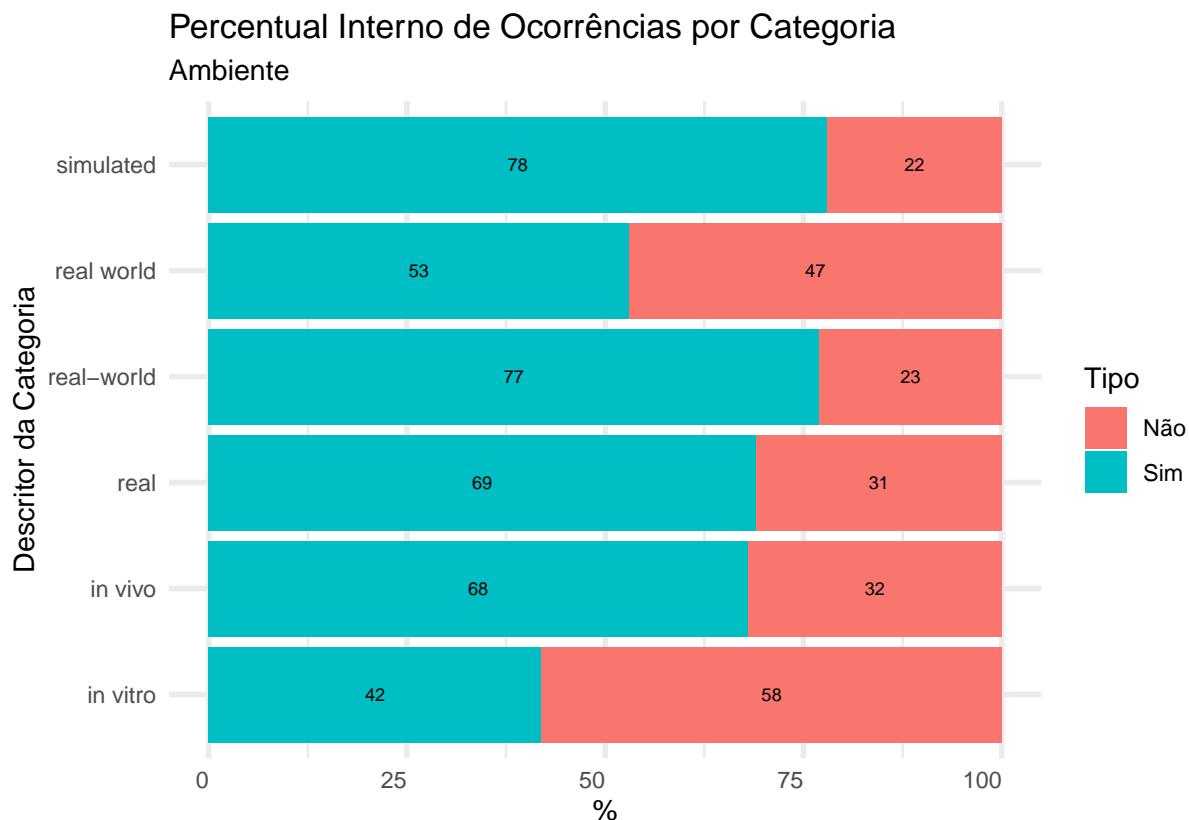
Menor número de ocorrências descritor: in vitro , com média 3,2 . Com percentual de aceitação 0,0064 % e não aceitação de 0,0088 %.

Maior percentual de aceitação sobre o total, descritor: real , com média 1133 . Com percentual de aceitação 3,8 % e não aceitação de 1,7 %.

Menor percentual de aceitação sobre o total, descritor: in vitro , com média 3,2 . Com percentual de aceitação 0,0064 % e não aceitação de 0,0088 %.

Table 6: Distribuição dos resultados individuais da categoria: Ambiente

Descriptor	n	%	n.Norm.	Total Sim	Sim	%STot.	Total Não	Não	%NTot.
in vitro	19	0,02	3,2	8	42	0,01	11	58	0,01
in vivo	47	0,04	7,8	32	68	0,03	15	32	0,01
real	6799	5,46	1133,2	4718	69	3,79	2081	31	1,67
real world	304	0,24	50,7	161	53	0,13	143	47	0,11
real-world	3028	2,43	504,7	2338	77	1,88	690	23	0,55
simulated	1496	1,20	249,3	1167	78	0,94	329	22	0,26



Maior número de ocorrências descriptor: qualitative , com média 450 . Com percentual de aceitação 2,4 % e não aceitação de 0,13 %.

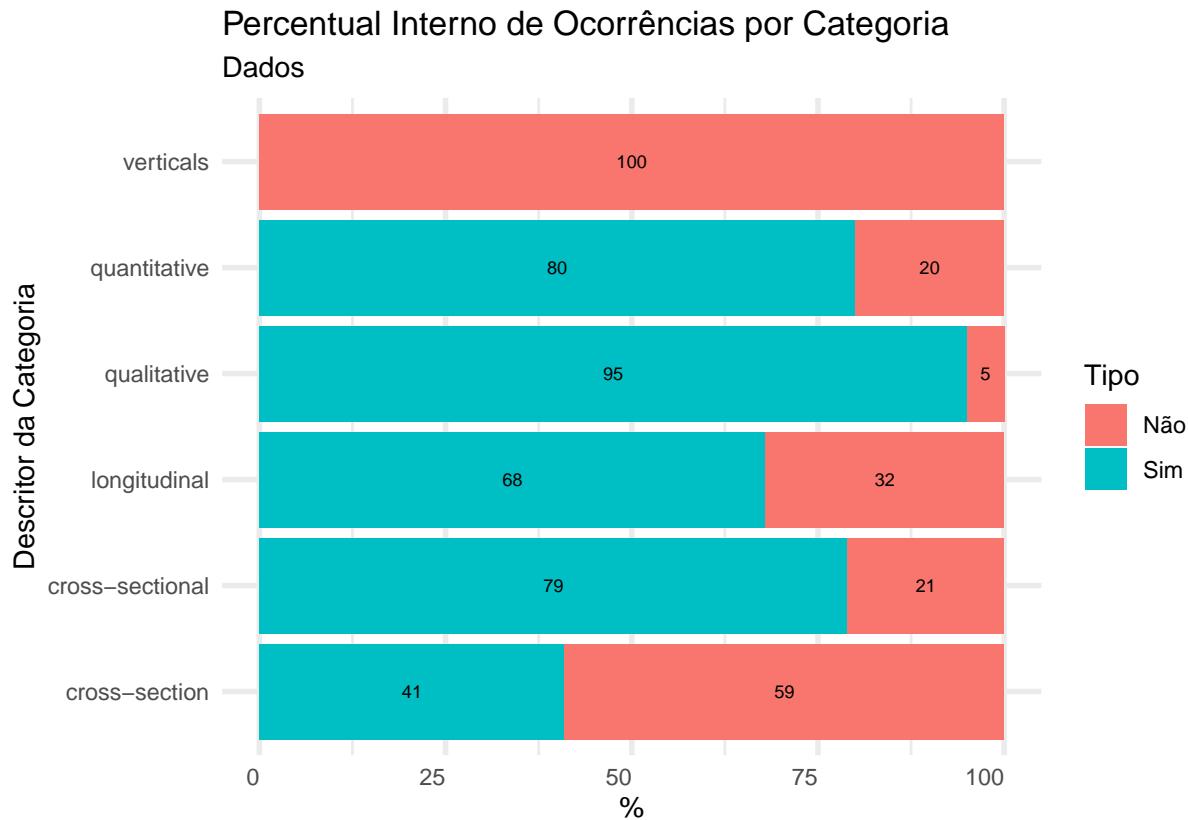
Menor número de ocorrências descriptor: verticals , com média 0,57 . Com percentual de aceitação 0 % e não aceitação de 0,0032 %.

Maior percentual de aceitação sobre o total, descriptor: qualitative , com média 450 . Com percentual de aceitação 2,4 % e não aceitação de 0,13 %.

Menor percentual de aceitação sobre o total, descriptor: verticals , com média 0,57 . Com percentual de aceitação 0 % e não aceitação de 0,0032 %.

Table 7: Distribuição dos resultados individuais da categoria: Dados

Descriptor	n	%	n.Norm.	Total Sim	Sim	%STot.	Total Não	Não	%NTot.
cross-section	46	0,04	6,57	19	41	0,02	27	59	0,02
cross-sectional	75	0,06	10,71	59	79	0,05	16	21	0,01
longitudinal	536	0,43	76,57	365	68	0,29	171	32	0,14
qualitative	3153	2,53	450,43	2995	95	2,40	158	5	0,13
quantitative	1762	1,41	251,71	1403	80	1,13	359	20	0,29
verticals	4	0,00	0,57	0	0	0,00	4	100	0,00



Categoria - Estratégia de Validação/Coleta de Dados

Maior número de ocorrências descriptor: survey , com média 231 . Com percentual de aceitação 4,9 % e não aceitação de 0,49 %.

Menor número de ocorrências descriptor: instrument application , com média 0,034 . Com percentual de aceitação 8e-04 % e não aceitação de 0 %.

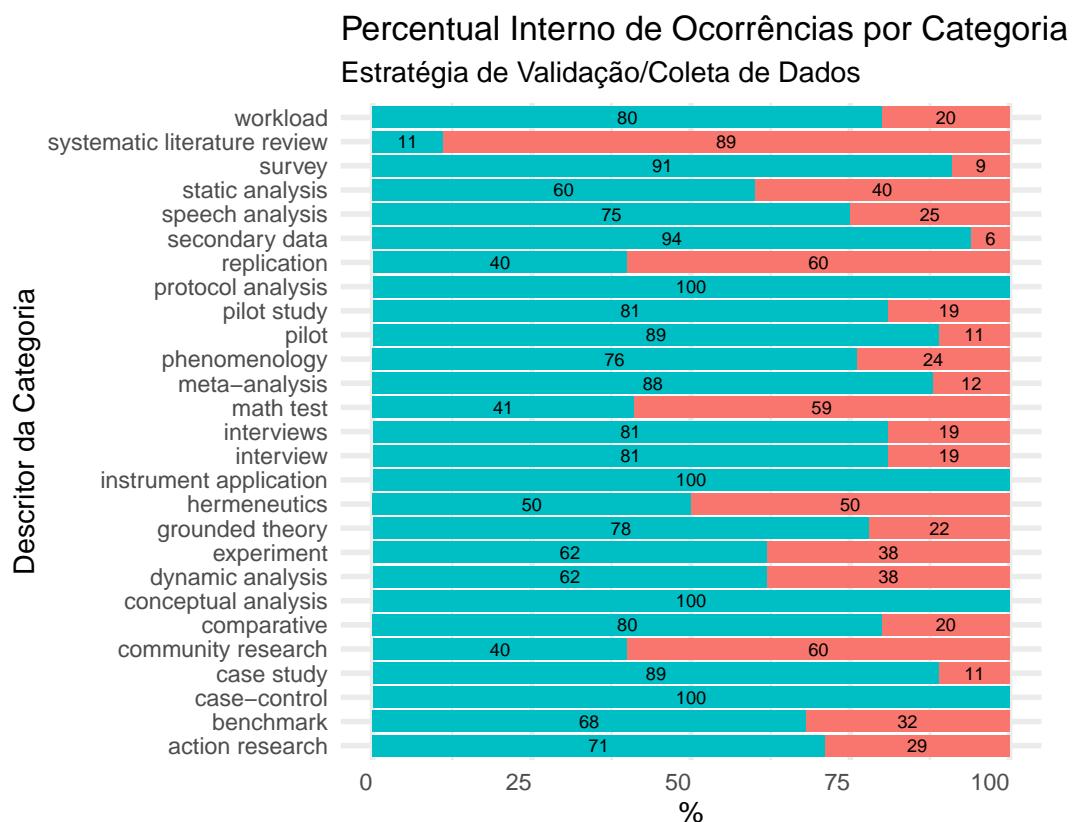
Maior percentual de aceitação sobre o total, descriptor: survey , com média 231 . Com percentual de aceitação 4,9 % e não aceitação de 0,49 %.

Menor percentual de aceitação sobre o total, descriptor: case-control , com média 0,21 . Com percentual de aceitação 0,0048 % e não aceitação de 0 %.

Table 8: Distribuição dos resultados individuais da categoria: Estratégia de Validação/Coleta de Dados

Descriptor	n	%	n.Norm.	Total Sim	Sim	%STot.	Total Não	Não	%NTot.
action research	72	0,06	2,48	51	71	0,04	21	29	0,02
benchmark	2957	2,37	101,97	2006	68	1,61	951	32	0,76
case study	1395	1,12	48,10	1236	89	0,99	159	11	0,13
case-control	6	0,00	0,21	6	100	0,00	0	0	0,00
community research	10	0,01	0,34	4	40	0,00	6	60	0,00
comparative	703	0,56	24,24	565	80	0,45	138	20	0,11
conceptual analysis	6	0,00	0,21	6	100	0,00	0	0	0,00
dynamic analysis	273	0,22	9,41	170	62	0,14	103	38	0,08
experiment	3978	3,19	137,17	2463	62	1,98	1515	38	1,22
grounded theory	237	0,19	8,17	184	78	0,15	53	22	0,04
hermeneutics	2	0,00	0,07	1	50	0,00	1	50	0,00
instrument application	1	0,00	0,03	1	100	0,00	0	0	0,00
interview	2073	1,66	71,48	1673	81	1,34	400	19	0,32
interviews	2306	1,85	79,52	1876	81	1,51	430	19	0,35
math test	17	0,01	0,59	7	41	0,01	10	59	0,01
meta-analysis	150	0,12	5,17	132	88	0,11	18	12	0,01
phenomenology	29	0,02	1,00	22	76	0,02	7	24	0,01
pilot	870	0,70	30,00	773	89	0,62	97	11	0,08
pilot study	179	0,14	6,17	145	81	0,12	34	19	0,03
protocol analysis	3	0,00	0,10	3	100	0,00	0	0	0,00
replication	632	0,51	21,79	253	40	0,20	379	60	0,30

Descriptor	n	%	n.Norm.	Total	Sim	Sim	%STot.	Total	Não	Não	%NTot.
secondary data	18	0,01	0,62	17	94	0,01		1	6	0,00	
speech analysis	4	0,00	0,14	3	75	0,00		1	25	0,00	
static analysis	650	0,52	22,41	389	60	0,31		261	40	0,21	
survey	6692	5,37	230,76	6082	91	4,88		610	9	0,49	
systematic literature review	135	0,11	4,66	15	11	0,01		120	89	0,10	
workload	2616	2,10	90,21	2080	80	1,67		536	20	0,43	



Categoria - Metodologia

Maior número de ocorrências descriptor: simulation , com média 611 . Com percentual de aceitação 3,4 % e não aceitação de 1,5 %.

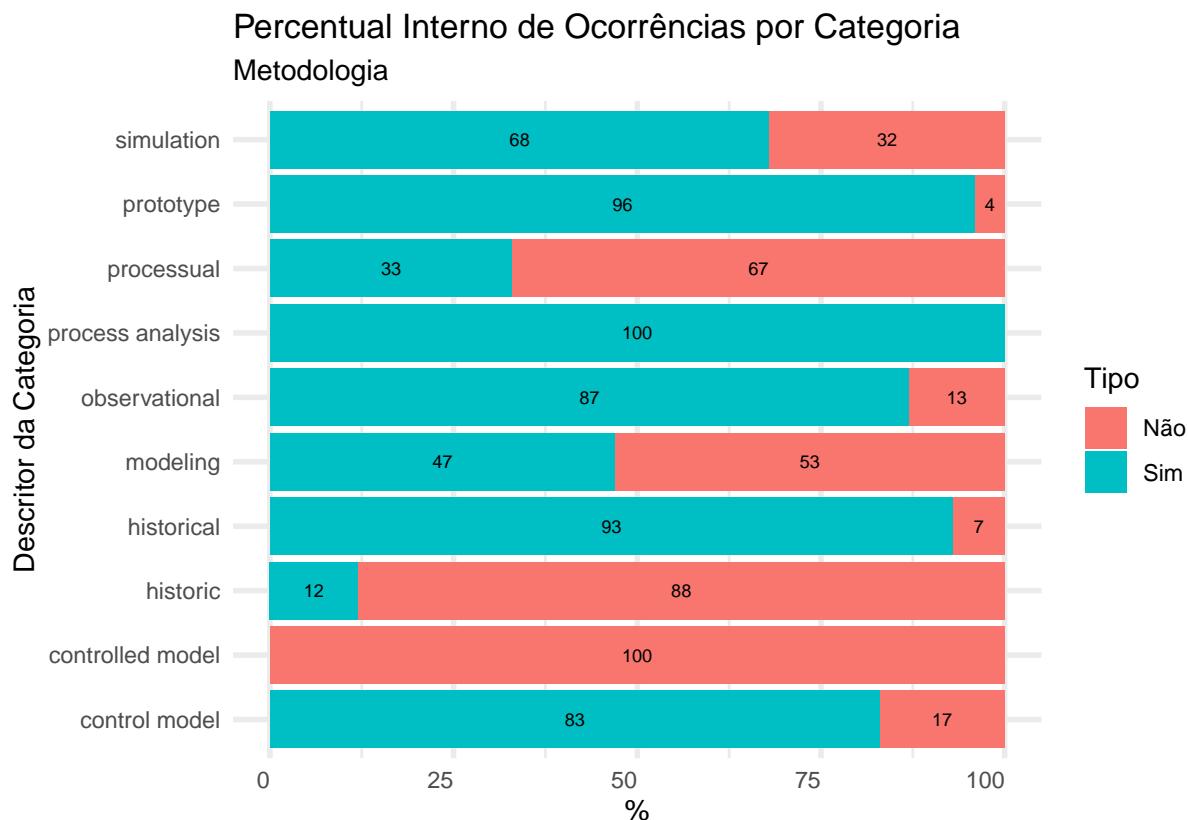
Menor número de ocorrências descriptor: controlled model , com média 0,1 . Com percentual de aceitação 0 % e não aceitação de 8e-04 %.

Maior percentual de aceitação sobre o total, descriptor: simulation , com média 611 . Com percentual de aceitação 3,4 % e não aceitação de 1,5 %.

Menor percentual de aceitação sobre o total, descriptor: process analysis , com média 0,4 . Com percentual de aceitação 0,0032 % e não aceitação de 0 %.

Table 9: Distribuição dos resultados individuais da categoria: Metodologia

Descriptor	n	%	n.Norm.	Total Sim	Sim	%STot.	Total Não	Não	%NTot.
control model	30	0,02	3,0	25	83	0,02	5	17	0,00
controlled model	1	0,00	0,1	0	0	0,00	1	100	0,00
historic	131	0,11	13,1	16	12	0,01	115	88	0,09
historical	1365	1,10	136,5	1270	93	1,02	95	7	0,08
modeling	5681	4,56	568,1	2681	47	2,15	3000	53	2,41
observational	483	0,39	48,3	422	87	0,34	61	13	0,05
process analysis	4	0,00	0,4	4	100	0,00	0	0	0,00
processual	3	0,00	0,3	1	33	0,00	2	67	0,00
prototype	1452	1,17	145,2	1388	96	1,11	64	4	0,05
simulation	6111	4,90	611,1	4182	68	3,36	1929	32	1,55



Categoria - Método

Maior número de ocorrências descriptor: evaluation , com média 332 . Com percentual de aceitação 6,7 % e não aceitação de 0,71 %.

Menor número de ocorrências descriptor: doctrinal , com média 0,036 . Com percentual de aceitação 0 % e não aceitação de 8e-04 %.

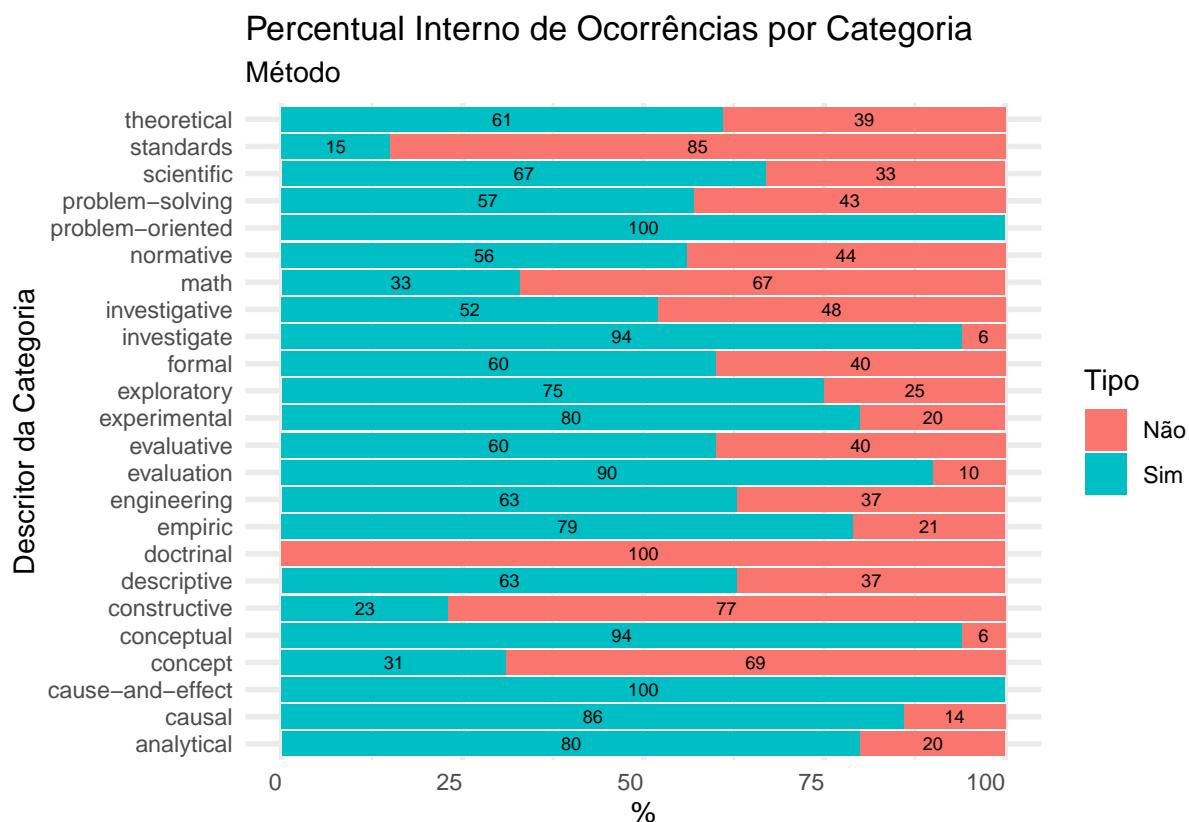
Maior percentual de aceitação sobre o total, descriptor: evaluation , com média 332 . Com percentual de aceitação 6,7 % e não aceitação de 0,71 %.

Menor percentual de aceitação sobre o total, descriptor: cause-and-effect , com média 0,29 . Com percentual de aceitação 0,0064 % e não aceitação de 0 %.

Table 10: Distribuição dos resultados individuais da categoria: Método

Descriptor	n	%	n.Norm.	Total Sim	Sim	%STot.	Total Não	Não	%NTot.
analytical	832	0,67	29,71	667	80	0,54	165	20	0,13
causal	1589	1,28	56,75	1369	86	1,10	220	14	0,18
cause-and-effect	8	0,01	0,29	8	100	0,01	0	0	0,00
concept	2531	2,03	90,39	775	31	0,62	1756	69	1,41
conceptual	1171	0,94	41,82	1095	94	0,88	76	6	0,06
constructive	401	0,32	14,32	91	23	0,07	310	77	0,25
descriptive	765	0,61	27,32	482	63	0,39	283	37	0,23
doctrinal	1	0,00	0,04	0	0	0,00	1	100	0,00
empiric	14	0,01	0,50	11	79	0,01	3	21	0,00
engineering	5658	4,54	202,07	3550	63	2,85	2108	37	1,69
evaluation	9298	7,46	332,07	8409	90	6,75	889	10	0,71
evaluative	62	0,05	2,21	37	60	0,03	25	40	0,02
experimental	4924	3,95	175,86	3919	80	3,15	1005	20	0,81
exploratory	715	0,57	25,54	536	75	0,43	179	25	0,14
formal	3383	2,72	120,82	2022	60	1,62	1361	40	1,09
investigate	2397	1,92	85,61	2247	94	1,80	150	6	0,12
investigative	98	0,08	3,50	51	52	0,04	47	48	0,04
math	497	0,40	17,75	164	33	0,13	333	67	0,27
normative	432	0,35	15,43	240	56	0,19	192	44	0,15
problem-oriented	1	0,00	0,04	1	100	0,00	0	0	0,00
problem-solving	234	0,19	8,36	134	57	0,11	100	43	0,08
scientific	1916	1,54	68,43	1281	67	1,03	635	33	0,51

Descriptor	n	%	n.Norm.	Total	Sim	Sim	%STot.	Total	Não	Não	%NTot.
standards	863	0,69	30,82	130	15		0,10	733	85		0,59
theoretical	2502	2,01	89,36	1515	61		1,22	987	39		0,79



Categoria - Natureza

Maior número de ocorrências descriptor: applied , com média 3208 . Com percentual de aceitação 3,9 % e não aceitação de 1,3 %.

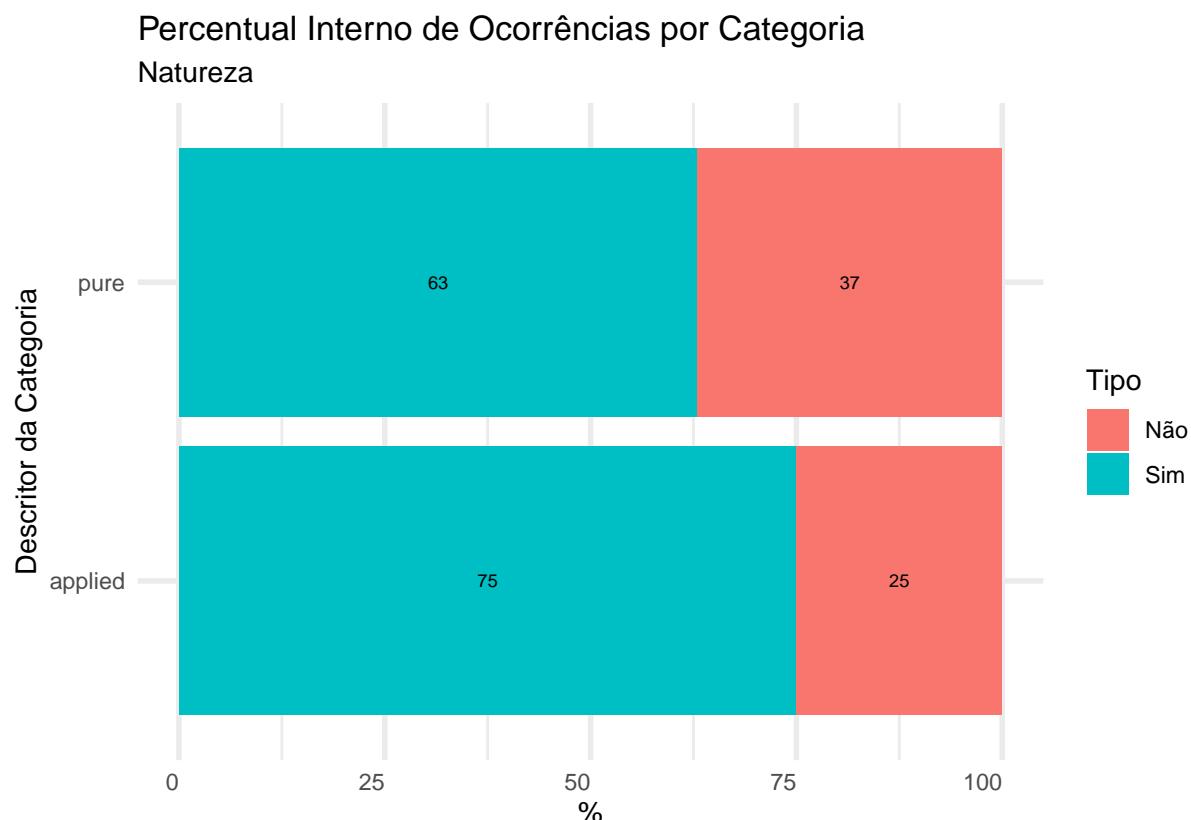
Menor número de ocorrências descriptor: pure , com média 482 . Com percentual de aceitação 0,49 % e não aceitação de 0,29 %.

Maior percentual de aceitação sobre o total, descriptor: applied , com média 3208 . Com percentual de aceitação 3,9 % e não aceitação de 1,3 %.

Menor percentual de aceitação sobre o total, descriptor: pure , com média 482 . Com percentual de aceitação 0,49 % e não aceitação de 0,29 %.

Table 11: Distribuição dos resultados individuais da categoria: Natureza

Descriptor	n	%	n.Norm.	Total Sim	Sim	%STot.	Total Não	Não	%NTot.
applied	6416	5,15	3208	4807	75	3,86	1609	25	1,29
pure	964	0,77	482	608	63	0,49	356	37	0,29



Categoria - Propósito paradigma

Maior número de ocorrências descriptor: behavioral , com média 550 . Com percentual de aceitação 0,92 % e não aceitação de 0,85 %.

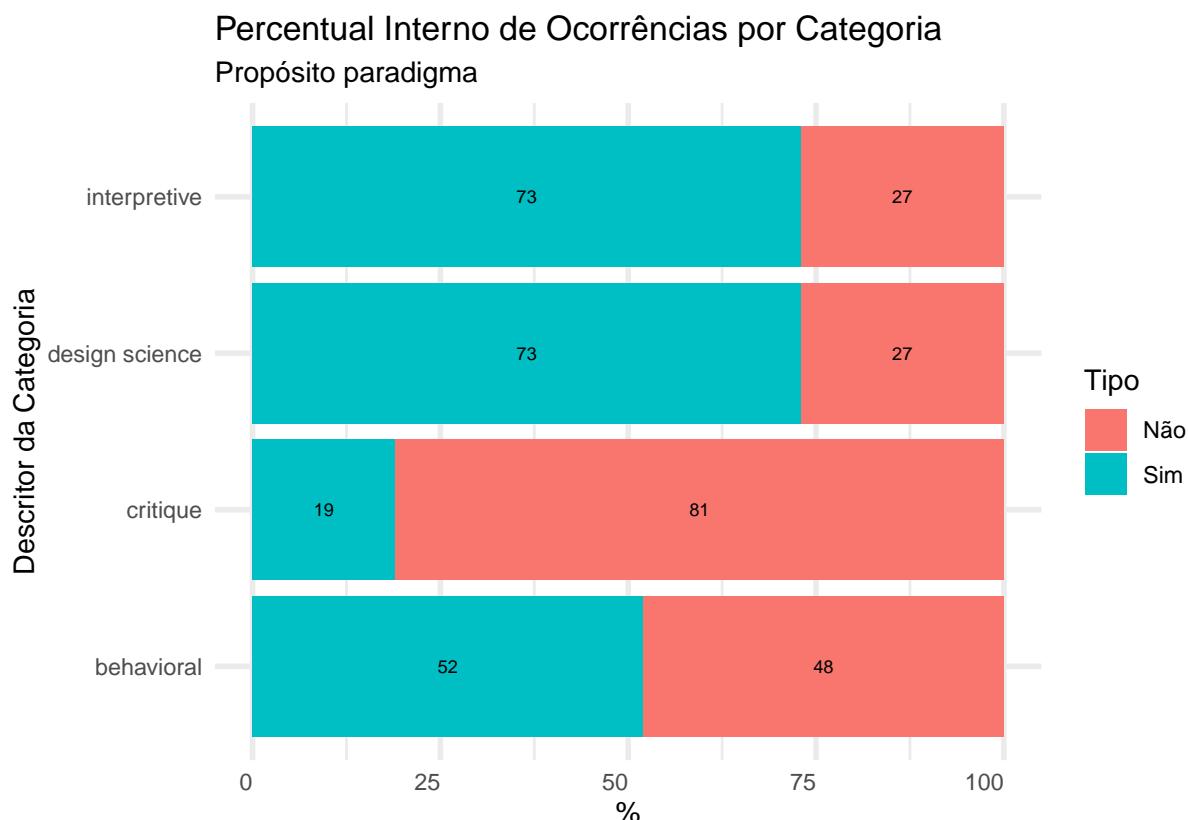
Menor número de ocorrências descriptor: design science , com média 9,2 . Com percentual de aceitação 0,022 % e não aceitação de 0,008 %.

Maior percentual de aceitação sobre o total, descriptor: behavioral , com média 550 . Com percentual de aceitação 0,92 % e não aceitação de 0,85 %.

Menor percentual de aceitação sobre o total, descriptor: design science , com média 9,2 . Com percentual de aceitação 0,022 % e não aceitação de 0,008 %.

Table 12: Distribuição dos resultados individuais da categoria: Propósito paradigma

Descriptor	n	%	n.Norm.	Total Sim	Sim	%STot.	Total Não	Não	%NTot.
behavioral	2202	1,77	550,5	1145	52	0,92	1057	48	0,85
critique	400	0,32	100,0	77	19	0,06	323	81	0,26
design science	37	0,03	9,2	27	73	0,02	10	27	0,01
interpretive	137	0,11	34,2	100	73	0,08	37	27	0,03



Maior número de ocorrências descriptor: proof , com média 3484 . Com percentual de aceitação 5,4 % e não aceitação de 0,18 %.

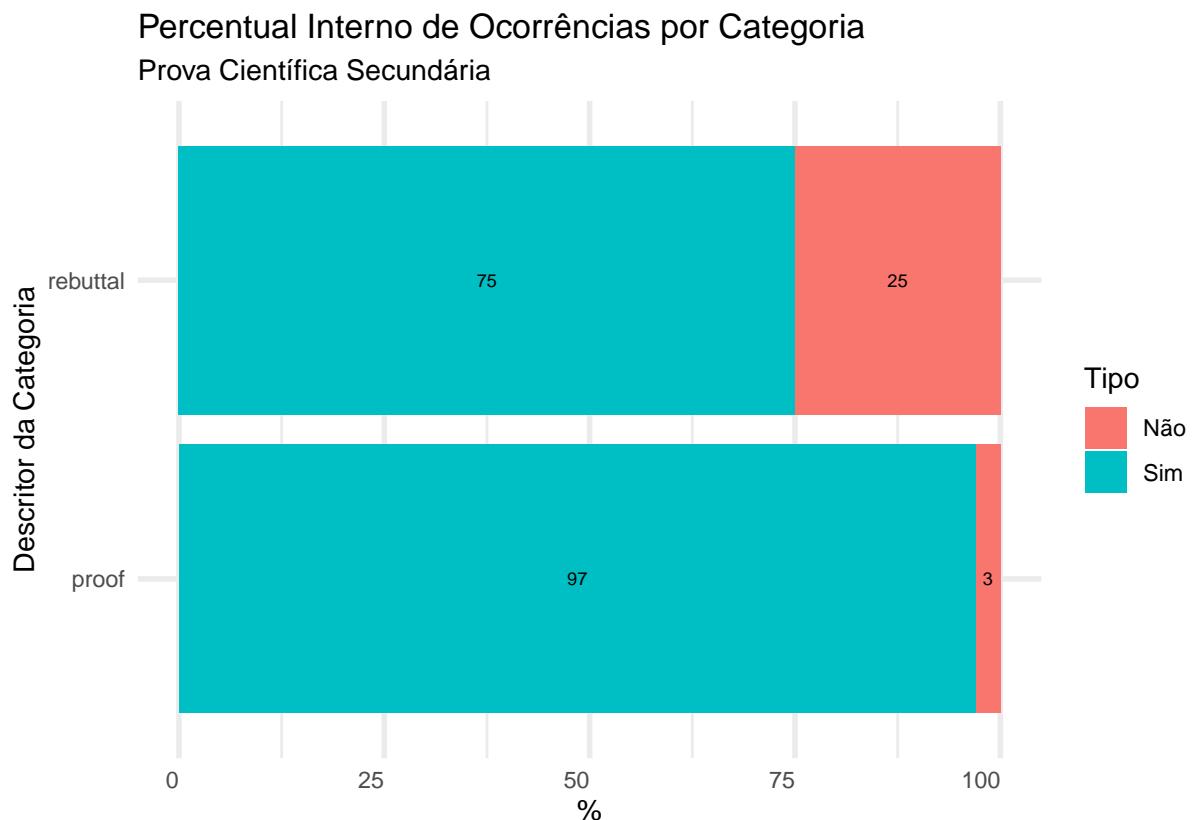
Menor número de ocorrências descriptor: rebuttal , com média 4 . Com percentual de aceitação 0,0048 % e não aceitação de 0,0016 %.

Maior percentual de aceitação sobre o total, descriptor: proof , com média 3484 . Com percentual de aceitação 5,4 % e não aceitação de 0,18 %.

Menor percentual de aceitação sobre o total, descriptor: rebuttal , com média 4 . Com percentual de aceitação 0,0048 % e não aceitação de 0,0016 %.

Table 13: Distribuição dos resultados individuais da categoria: Prova Científica Secundária

Descriptor	n	%	n.Norm.	Total Sim	Sim	%STot.	Total Não	Não	%NTot.
proof	6969	5,59	3484	6739	97	5,4	230	3	0,18
rebuttal	8	0,01	4	6	75	0,0	2	25	0,00



Maior número de ocorrências descriptor: induction , com média 377 . Com percentual de aceitação 0,79 % e não aceitação de 0,12 %.

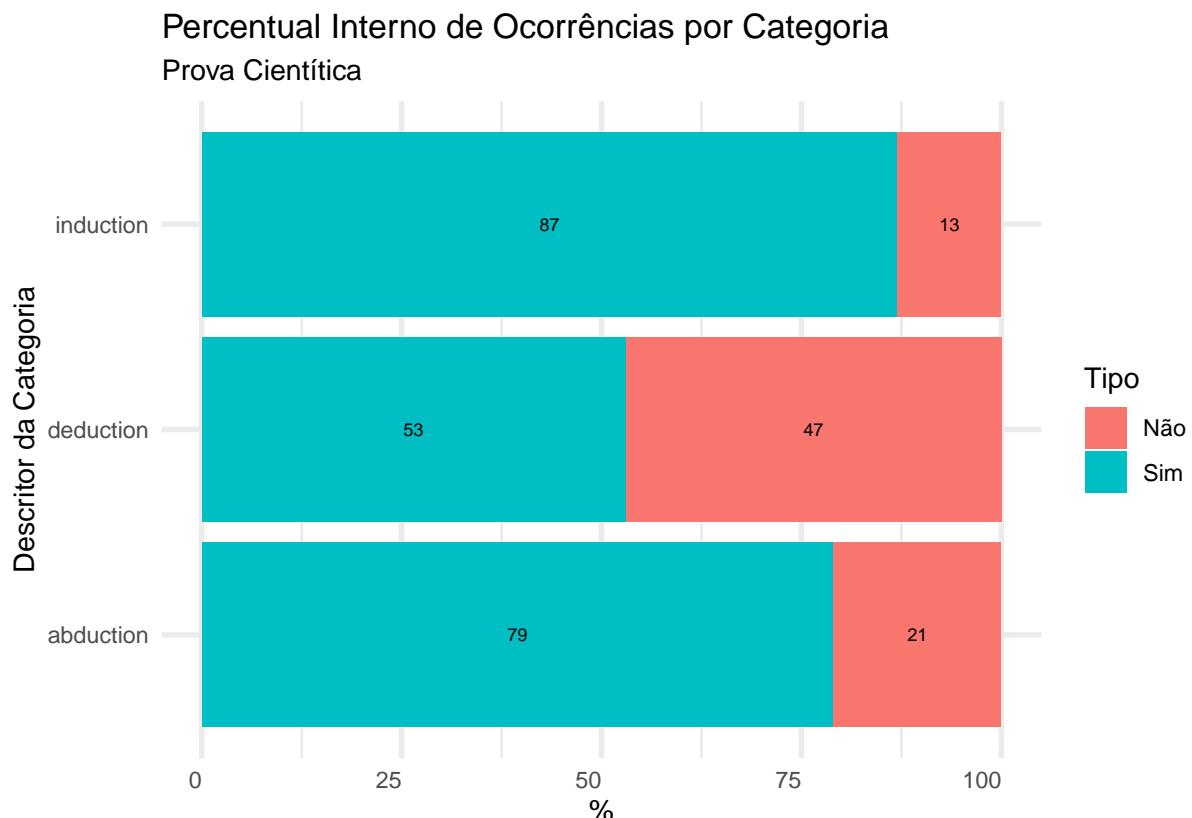
Menor número de ocorrências descriptor: abduction , com média 13 . Com percentual de aceitação 0,025 % e não aceitação de 0,0064 %.

Maior percentual de aceitação sobre o total, descriptor: induction , com média 377 . Com percentual de aceitação 0,79 % e não aceitação de 0,12 %.

Menor percentual de aceitação sobre o total, descriptor: abduction , com média 13 . Com percentual de aceitação 0,025 % e não aceitação de 0,0064 %.

Table 14: Distribuição dos resultados individuais da categoria: Prova Científica

Descriptor	n	%	n.Norm.	Total Sim	Sim	%STot.	Total Não	Não	%NTot.
abduction	39	0,03	13	31	79	0,02	8	21	0,01
deduction	64	0,05	21	34	53	0,03	30	47	0,02
induction	1131	0,91	377	979	87	0,79	152	13	0,12



Categoria - Tipo de Validação do Resultado

Maior número de ocorrências descriptor: efficiency , com média 639 . Com percentual de aceitação 2,1 % e não aceitação de 0,49 %.

Menor número de ocorrências descriptor: operationalization , com média 15 . Com percentual de aceitação 0,052 % e não aceitação de 0,0064 %.

Maior percentual de aceitação sobre o total, descriptor: efficiency , com média 639 . Com percentual de aceitação 2,1 % e não aceitação de 0,49 %.

Menor percentual de aceitação sobre o total, descriptor: cost-effectiveness , com média 17 . Com percentual de aceitação 0,063 % e não aceitação de 0,0056 %.

Table 15: Distribuição dos resultados individuais da categoria: Tipo de Validação do Resultado

Descriptor	n	%	n.Norm.	Total Sim	Sim	%STot.	Total Não	Não	%NTot.
cost-effectiveness	86	0,07	17	79	92	0,06	7	8	0,01
effectiveness	2499	2,01	500	640	26	0,51	1859	74	1,49
efficiency	3194	2,56	639	2586	81	2,08	608	19	0,49
operationalization	73	0,06	15	65	89	0,05	8	11	0,01
usability	1546	1,24	309	1345	87	1,08	201	13	0,16

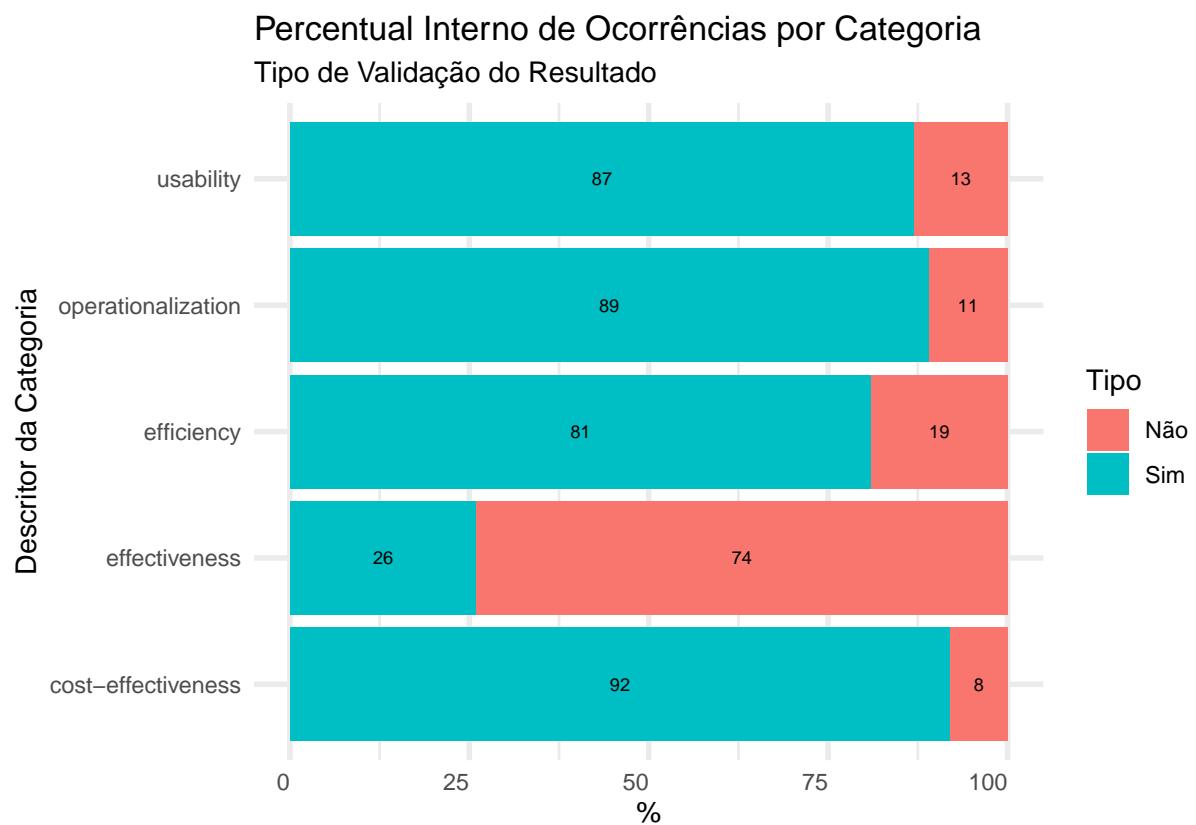


Tabela sumarizada por Periódico das publicações

No total foram pesquisados 2508 em 68 periódicos da ACM. A tabela abaixo mostra os totais individualizados por periódico ordenados pelo total de ocorrências. Note que a média de ocorrências não está ligada ao número de artigos. A maior média é do Periódico ACM Trans. Softw. Eng. Methodol. com 146,7 ocorrências em 19 artigos!

Table 16: Distribuição dos resultados por periódico da ACM.

Periódico	Artigos	n	Média	%	Total			Total		
					Sim	%Per.	%Tot.	Não	%Per.	%Tot.
Proc. ACM Hum.-Comput. Interact.	522	28263	54,1	22,68	20367	72	16,35	7896	28	6,34
Proc. ACM Program. Lang.	217	12214	56,3	9,80	9552	78	7,67	2662	22	2,14
ACM Trans. Comput.-Hum. Interact.	105	8464	80,6	6,79	5937	70	4,76	2527	30	2,03
ACM Comput. Surv.	65	6387	98,3	5,13	4929	77	3,96	1458	23	1,17
ACM Trans. Graph.	159	5970	37,5	4,79	3890	65	3,12	2080	35	1,67
Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.	113	5447	48,2	4,37	4097	75	3,29	1350	25	1,08
ACM Trans. Model. Comput. Simul.	40	4172	104,3	3,35	2526	61	2,03	1646	39	1,32
ACM Trans. Archit. Code Optim.	79	3254	41,2	2,61	2512	77	2,02	742	23	0,60
ACM/IMS Trans. Data Sci.	73	3253	44,6	2,61	2452	75	1,97	801	25	0,64
Proc. ACM Meas. Anal. Comput. Syst.	57	2856	50,1	2,29	2219	78	1,78	637	22	0,51
ACM Trans. Softw. Eng. Methodol.	19	2821	148,5	2,26	2245	80	1,80	576	20	0,46
ACM Trans. Comput. Educ.	40	2339	58,5	1,88	1680	72	1,35	659	28	0,53
ACM Trans. Program. Lang. Syst.	26	2223	85,5	1,78	1695	76	1,36	528	24	0,42
ACM Trans. Knowl. Discov. Data	41	2143	52,3	1,72	1671	78	1,34	472	22	0,38
J. Hum.-Robot Interact.	31	2099	67,7	1,68	1559	74	1,25	540	26	0,43
J. ACM	33	1941	58,8	1,56	1666	86	1,34	275	14	0,22
SIGPLAN Not.	53	1611	30,4	1,29	1127	70	0,90	484	30	0,39
ACM Trans. Cyber-Phys. Syst.	35	1579	45,1	1,27	1099	70	0,88	480	30	0,39
ACM Trans. Embed. Comput. Syst.	41	1402	34,2	1,13	1020	73	0,82	382	27	0,31
Commun. ACM	72	1377	19,1	1,11	897	65	0,72	480	35	0,39
ACM Trans. Intell. Syst. Technol.	29	1267	43,7	1,02	944	75	0,76	323	25	0,26
J. Emerg. Technol. Comput. Syst.	38	1252	33,0	1,00	906	72	0,73	346	28	0,28
ACM Trans. Interact. Intell. Syst.	19	1248	65,7	1,00	935	75	0,75	313	25	0,25
ACM Trans. Storage	26	1177	45,3	0,94	852	72	0,68	325	28	0,26
ACM Trans. Des. Autom. Electron. Syst.	32	1107	34,6	0,89	824	74	0,66	283	26	0,23
ACM Trans. Priv. Secur.	21	1036	49,3	0,83	792	76	0,64	244	24	0,20
ACM Trans. Sen. Netw.	22	964	43,8	0,77	708	73	0,57	256	27	0,21
Proc. ACM Comput. Graph. Interact. Tech.	44	946	21,5	0,76	668	71	0,54	278	29	0,22
ACM Trans. Comput. Syst.	21	914	43,5	0,73	718	79	0,58	196	21	0,16
ACM Trans. Appl. Percept.	10	860	86,0	0,69	681	79	0,55	179	21	0,14
ACM Trans. Access. Comput.	15	824	54,9	0,66	626	76	0,50	198	24	0,16
ACM Trans. Math. Softw.	25	736	29,4	0,59	498	68	0,40	238	32	0,19
ACM Trans. Multimedia Comput. Commun.	19	730	38,4	0,59	565	77	0,45	165	23	0,13
Appl.										
ACM Trans. Manage. Inf. Syst.	17	725	42,6	0,58	522	72	0,42	203	28	0,16
ACM Trans. Parallel Comput.	15	697	46,5	0,56	511	73	0,41	186	27	0,15
ACM Trans. Spatial Algorithms Syst.	16	681	42,6	0,55	526	77	0,42	155	23	0,12
ACM Trans. Database Syst.	10	638	63,8	0,51	516	81	0,41	122	19	0,10
ACM Trans. Internet Things	8	541	67,6	0,43	366	68	0,29	175	32	0,14
ACM Trans. Model. Perform. Eval. Comput. Syst.	10	513	51,3	0,41	384	75	0,31	129	25	0,10
Digital Threats	8	507	63,4	0,41	304	60	0,24	203	40	0,16
ACM Trans. Algorithms	24	499	20,8	0,40	421	84	0,34	78	16	0,06
ACM Trans. Inf. Syst.	10	499	49,9	0,40	356	71	0,29	143	29	0,11
ACM Trans. Econ. Comput.	17	497	29,2	0,40	433	87	0,35	64	13	0,05
ACM Trans. Comput. Healthcare	8	460	57,5	0,37	324	70	0,26	136	30	0,11
ACM Trans. Comput. Theory	11	460	41,8	0,37	380	83	0,30	80	17	0,06
ACM Trans. Internet Technol.	11	414	37,6	0,33	307	74	0,25	107	26	0,09
Ubiquity	14	339	24,2	0,27	203	60	0,16	136	40	0,11
SIGCOMM Comput. Commun. Rev.	26	317	12,2	0,25	215	68	0,17	102	32	0,08
Form. Asp. Comput.	3	297	99,0	0,24	231	78	0,19	66	22	0,05
Digit. Gov.: Res. Pract.	13	293	22,5	0,24	219	75	0,18	74	25	0,06
ACM J. Exp. Algorithms	7	289	41,3	0,23	218	75	0,17	71	25	0,06
ACM Trans. Asian Low-Resour. Lang. Inf. Process.	11	288	26,2	0,23	208	72	0,17	80	28	0,06
J. Data and Information Quality	7	272	38,9	0,22	215	79	0,17	57	21	0,05
SIGMETRICS Perform. Eval. Rev.	31	268	8,6	0,22	166	62	0,13	102	38	0,08
ACM Trans. Evol. Learn. Optim.	4	265	66,2	0,21	171	65	0,14	94	35	0,08
Digital Threats: Research and Practice	5	259	51,8	0,21	169	65	0,14	90	35	0,07
Trans. Soc. Comput.	5	254	50,8	0,20	171	67	0,14	83	33	0,07
ACM Trans. Reconfigurable Technol. Syst.	9	247	27,4	0,20	196	79	0,16	51	21	0,04
J. Comput. Cult. Herit.	4	246	61,5	0,20	190	77	0,15	56	23	0,04
ACM Inroads	10	227	22,7	0,18	143	63	0,11	84	37	0,07
ACM Trans. Web	5	148	29,6	0,12	91	61	0,07	57	39	0,05
ACM Trans. Comput. Logic	3	133	44,3	0,11	118	89	0,09	15	11	0,01
ACM Trans. Auton. Adapt. Syst.	2	107	53,5	0,09	88	82	0,07	19	18	0,02
interactions	15	103	6,9	0,08	53	51	0,04	50	49	0,04
XRDS	20	97	4,8	0,08	50	52	0,04	47	48	0,04
ACM Transactions on Quantum Computing	2	84	42,0	0,07	54	64	0,04	30	36	0,02
Queue	4	53	13,2	0,04	36	68	0,03	17	32	0,01
SIGACCESS Access. Comput.	1	8	8,0	0,01	7	88	0,01	1	12	0,00
Total	2508	124601	NA	100,00	91419	NA	73,37	33182	NA	26,63

Tabela Sumarizada por área dos periódicos segundo classificação da SCOPUS para a Ciência

Cada observação na base de dados contém informações a respeito do artigo a que se refere. Dentre tais informações está a variável área, a qual contém as áreas as quais fazem parte do escopo de publicação do periódico. O conteúdo se refere ao periódico e não ao artigo conforme mostra a tabela abaixo.

Nesta tabela é possível ver que mais de uma área estão agrupadas por linha, e, uma mesma área está presente em mais de uma linha. Isto se deve ao fato de, como já descrito, a variável *areas* se referir ao periódico e não ao artigo.

Tal informação é categorizada segundo a classificação ASJC (All Science Journal Classification Codes - <https://scientificresearch.in/asjc-all-science-journal-classification-codes/>) que é organizada e mantida pela SCOPUS/Elsevier (<https://www.scopus.com/home.uri>) e compreende uma lista das grandes áreas da ciência (da: vida, física, saúde, social e humanas) divididas em 333 sub áreas não consideradas nesta análise.

Table 17: Distribuição segundo o agrupamento de áreas dos periódicos

Areas	n	%	Total			Total		
			Sim	%Cat.	%Tot.	Não	%Cat.	%Tot.
Social Sciences (miscellaneous), Human-Computer Interaction, Computer Networks and Communications,	28263	22,68	20367	72	16,35	7896	28	6,34
Software, Safety, Risk, Reliability and Quality,	12214	9,80	9552	78	7,67	2662	22	2,14
Human-Computer Interaction,	8567	6,88	5990	70	4,81	2577	30	2,07
Theoretical Computer Science, General Computer Science,	6387	5,13	4929	77	3,96	1458	23	1,17
Computer Graphics and Computer-Aided Design,	5970	4,79	3890	65	3,12	2080	35	1,67
Human-Computer Interaction, Hardware and Architecture, Computer Networks and Communications,	5447	4,37	4097	75	3,29	1350	25	1,08
Software,	5044	4,05	3940	78	3,16	1104	22	0,89
General Computer Science,	5022	4,03	3726	74	2,99	1296	26	1,04
4810	3,86	3461	72	2,78	1349	28	1,08	
Modeling and Simulation, Computer Science Applications,	4172	3,35	2526	61	2,03	1646	39	1,32
Human-Computer Interaction, Artificial Intelligence,	3347	2,69	2494	75	2,00	853	25	0,68
Software, Information Systems, Hardware and Architecture,	3254	2,61	2512	77	2,02	742	23	0,60
Computer Science (miscellaneous), Safety, Risk, Reliability and Quality, Hardware and Architecture, Computer Networks and Communications,	2856	2,29	2219	78	1,78	637	22	0,51
General Computer Science, Education,	2566	2,06	1823	71	1,46	743	29	0,60
Software, Control and Systems Engineering, Information Systems, Hardware and Architecture, Artificial Intelligence,	1941	1,56	1666	86	1,34	275	14	0,22
Software, Computer Graphics and Computer-Aided Design,	1611	1,29	1127	70	0,90	484	30	0,39
Human-Computer Interaction, Hardware and Architecture, Computer Networks and Communications, Control and Optimization, Artificial Intelligence,	1579	1,27	1099	70	0,88	480	30	0,39
Computer Networks and Communications,	1526	1,22	1106	72	0,89	420	28	0,34
Software, Hardware and Architecture,	1402	1,13	1020	73	0,82	382	27	0,31
Theoretical Computer Science, Artificial Intelligence,	1267	1,02	944	75	0,76	323	25	0,26
Software, Hardware and Architecture, Electrical and Electronic Engineering,	1252	1,00	906	72	0,73	346	28	0,28
Hardware and Architecture,	1177	0,94	852	72	0,68	325	28	0,26
Computer Science Applications, Computer Graphics and Computer-Aided Design, Electrical and Electronic Engineering,	1107	0,89	824	74	0,66	283	26	0,23
General Computer Science, Safety, Risk, Reliability and Quality,	1036	0,83	792	76	0,64	244	24	0,20
Computer Science Applications, Computer Graphics and Computer-Aided Design,	946	0,76	668	71	0,54	278	29	0,22
Theoretical Computer Science, General Computer Science, Experimental and Cognitive Psychology,	860	0,69	681	79	0,55	179	21	0,14
Human-Computer Interaction, Computer Science Applications,	824	0,66	626	76	0,50	198	24	0,16
Software, Applied Mathematics,	736	0,59	498	68	0,40	238	32	0,19
Hardware and Architecture, Computer Networks and Communications,	730	0,59	565	77	0,45	165	23	0,13
Water Science and Technology, Ocean Engineering, Geology,	726	0,58	445	61	0,36	281	39	0,23
Management Information Systems, General Computer Science,	725	0,58	522	72	0,42	203	28	0,16
Software, Modeling and Simulation, Hardware and Architecture, Computer Science Applications, Computational Theory and Mathematics,	697	0,56	511	73	0,41	186	27	0,15
Signal Processing, Information Systems, Modeling and Simulation, Computer Science Applications, Geometry and Topology, Discrete Mathematics and Combinatorics,	681	0,55	526	77	0,42	155	23	0,12
Information Systems,	638	0,51	516	81	0,41	122	19	0,10
Computer Science (miscellaneous), Software, Information Systems, Media Technology, Safety, Risk, Reliability and Quality, Hardware and Architecture, Computer Networks and Communications,	513	0,41	384	75	0,31	129	25	0,10
Information Systems, General Business, Management and Accounting, Computer Science Applications,	499	0,40	356	71	0,29	143	29	0,11
Mathematics (miscellaneous), Computer Science (miscellaneous), Statistics and Probability, Economics and Econometrics,	499	0,40	421	84	0,34	78	16	0,06
Marketing, Computational Mathematics,	497	0,40	433	87	0,35	64	13	0,05
Theoretical Computer Science, Computational Theory and Mathematics,	460	0,37	380	83	0,30	80	17	0,06
Software, Computer Networks and Communications,	317	0,25	215	68	0,17	102	32	0,08
Software, Medicine (miscellaneous), Information Systems, Biomedical Engineering, Computer Science Applications, Health Informatics, Health Information Management,	302	0,24	210	70	0,17	92	30	0,07
Software, Theoretical Computer Science,	297	0,24	231	78	0,19	66	22	0,05
Microbiology,	293	0,24	219	75	0,18	74	25	0,06
Theoretical Computer Science,	289	0,23	218	75	0,17	71	25	0,06
Information Systems, Information Systems and Management,	272	0,22	215	79	0,17	57	21	0,05
Software, Hardware and Architecture, Computer Networks and Communications,	268	0,22	166	62	0,13	102	38	0,08
Conservation, Information Systems, Computer Science Applications, Computer Graphics and Computer-Aided Design,	246	0,20	190	77	0,15	56	23	0,04
Theoretical Computer Science, General Computer Science, Logic, Computational Mathematics,	133	0,11	118	89	0,09	15	11	0,01
General Environmental Science, General Earth and Planetary Sciences,	121	0,10	80	66	0,06	41	34	0,03
Control and Systems Engineering, Computer Science (miscellaneous), Software,	107	0,09	88	82	0,07	19	18	0,02
Software, Information Systems, Safety Research, Hardware and Architecture, Computer Science Applications, Computer Networks and Communications,	40	0,03	28	70	0,02	12	30	0,01
General Medicine,	34	0,03	23	68	0,02	11	32	0,01
Ocean Engineering,	34	0,03	24	71	0,02	10	29	0,01
Total	124601	100,00	91419	NA	73,37	33182	NA	26,63

Na tabela abaixo, o conteúdo da variável *areas* foi desmembrado do agrupamento. Cada linha representa a contagem de artigos em cada área respectivamente. Note que o total não é compatível com o número de artigos pesquisados, isto se deve de que os periódicos são categorizados em mais de uma área e há uma sobreposição na contagem.

Table 18: Distribuição dos resultados por áreas segundo SCOPUS.

Areas	n
Applied Mathematics	25
Artificial Intelligence	147
Biomedical Engineering	4
Computational Mathematics	20
Computational Theory and Mathematics	26
Computer Graphics and Computer-Aided Design	292
Computer Networks and Communications	852
Computer Science (miscellaneous)	86
Computer Science Applications	181
Conservation	4
Control and Optimization	35
Control and Systems Engineering	35
Discrete Mathematics and Combinatorics	16
Economics and Econometrics	17
Education	50
Electrical and Electronic Engineering	70
Experimental and Cognitive Psychology	10
General Business	10
General Computer Science	324
General Earth and Planetary Sciences	2
General Environmental Science	2
General Medicine	1
Geology	12
Geometry and Topology	16
Hardware and Architecture	498
Health Informatics	4
Health Information Management	4
Human-Computer Interaction	855
Information Systems	174
Information Systems and Management	7
Logic	3
Management Information Systems	17
Management and Accounting	10
Marketing	17
Mathematics (miscellaneous)	24
Media Technology	10
Medicine (miscellaneous)	4
Microbiology	13
Modeling and Simulation	71
Ocean Engineering	14
Reliability and Quality	305
Risk	305
Safety	305
Safety Research	1
Signal Processing	16
Social Sciences (miscellaneous)	522
Software	623
Statistics and Probability	17
Theoretical Computer Science	128
Water Science and Technology	12

Areas	n
Total	6196

Na tabela abaixo foi realizado um exercício para relacionar as classificações de área ASJC, ACM-CCS e Grandes áreas da computação segundo a Sociedade Brasileira de Computação (SBC). Como pode ser visto há uma equivalência entre elas, o que já era esperado uma vez que as três versam sobre a mesma área da ciência.

Table 19: Associação de equivalência das classificações da Ciência da Computação segundo ACM, SCOPUS e SBC.

Classif. ACM	Classif. SCOPUS	Classif. SBC
General and reference	General Computer Science	GA 4 Aplicações da Computação
Hardware	Hardware and Architecture	GA 1 Sistemas Computacionais
Hardware	Signal Processing	GA 1 Sistemas Computacionais
Hardware	Software	GA 2 Sistemas de Software
Computer systems organization	Software	GA 2 Sistemas de Software
Networks	Computer Networks and Communications	GA 1 Sistemas Computacionais
Software and its engineering	Software	GA 2 Sistemas de Software
Theory of computation	Computational Theory and Mathematics	GA 5 Gestão de Dados e de Informações
Mathematics of computing	Computational Theory and Mathematics	GA 2 Sistemas de Software
Information systems	Information Systems	GA 2 Sistemas de Software
Security and privacy	Computer Networks and Communications	GA 1 Sistemas Computacionais
Human-centered computing	Human-Computer Interaction	GA 3 Técnicas e Tecnologias de Computação
Computing methodologies	Artificial Intelligence	GA 3 Técnicas e Tecnologias de Computação
Computing methodologies	Computer Vision and Pattern Recognition	GA 4 Aplicações da Computação
Computing methodologies	Computer Graphics and Computer-Aided Design	GA 4 Aplicações da Computação
Applied computing	Computer Science Applications	GA 4 Aplicações da Computação
Applied computing	Computer Graphics and Computer-Aided Design	GA 4 Aplicações da Computação
Social and professional topics	Computer Science (miscellaneous)	GA 4 Aplicações da Computação

Com base na tabela de equivalência de áreas (ASJC, ACM-CCS e SBC), foi feita a relação entre tal tabela e as áreas desmembradas obtidas nas observações dos dados do presente trabalho, uma vez que estes estão distribuídos segundo a classificação da SCOPUS. Note que no que tange a computação, todas as áreas foram contempladas nos resultados. Ainda é possível notar que o escopo dos periódicos pesquisados não dizem respeito apenas a computação, mas também a outras áreas do conhecimento. Conforme mostra a tabela abaixo.

Table 20: Contagem de artigos por área e associação de áreas segundo SCOPUS, ACM e SBC

Área (SCOPUS)	n	Classif. ACM	Classif. SBC
Applied Mathematics	25	NA	NA
Artificial Intelligence	147	Computing methodologies	GA 3 Técnicas e Tecnologias de Computação
Biomedical Engineering	4	NA	NA
Computational Mathematics	20	NA	NA
Computational Theory and Mathematics	26	Theory of computation	GA 5 Gestão de Dados e de Informações
Computational Theory and Mathematics	26	Mathematics of computing	GA 2 Sistemas de Software
Computer Graphics and Computer-Aided Design	292	Computing methodologies	GA 4 Aplicações da Computação
Computer Graphics and Computer-Aided Design	292	Applied computing	GA 4 Aplicações da Computação
Computer Networks and Communications	852	Networks	GA 1 Sistemas Computacionais
Computer Networks and Communications	852	Security and privacy	GA 1 Sistemas Computacionais
Computer Science (miscellaneous)	86	Social and professional topics	GA 4 Aplicações da Computação
Computer Science Applications	181	Applied computing	GA 4 Aplicações da Computação
Computer Vision and Pattern Recognition	NA	Computing methodologies	GA 4 Aplicações da Computação
Conservation	4	NA	NA
Control and Optimization	35	NA	NA
Control and Systems Engineering	35	NA	NA
Discrete Mathematics and Combinatorics	16	NA	NA
Economics and Econometrics	17	NA	NA
Education	50	NA	NA
Electrical and Electronic Engineering	70	NA	NA
Experimental and Cognitive Psychology	10	NA	NA
General Business	10	NA	NA
General Computer Science	324	General and reference	GA 4 Aplicações da Computação
General Earth and Planetary Sciences	2	NA	NA
General Environmental Science	2	NA	NA
General Medicine	1	NA	NA
Geology	12	NA	NA
Geometry and Topology	16	NA	NA
Hardware and Architecture	498	Hardware	GA 1 Sistemas Computacionais
Health Informatics	4	NA	NA
Health Information Management	4	NA	NA
Human-Computer Interaction	855	Human-centered computing	GA 3 Técnicas e Tecnologias de Computação
Information Systems	174	Information systems	GA 2 Sistemas de Software
Information Systems and Management	7	NA	NA
Logic	3	NA	NA
Management and Accounting	10	NA	NA
Management Information Systems	17	NA	NA
Marketing	17	NA	NA
Mathematics (miscellaneous)	24	NA	NA
Media Technology	10	NA	NA
Medicine (miscellaneous)	4	NA	NA
Microbiology	13	NA	NA
Modeling and Simulation	71	NA	NA
Ocean Engineering	14	NA	NA

Área (SCOPUS)	n	Classif. ACM	Classif. SBC
Reliability and Quality	305	NA	NA
Risk	305	NA	NA
Safety	305	NA	NA
Safety Research	1	NA	NA
Signal Processing	16	Hardware	GA 1 Sistemas Computacionais
Social Sciences (miscellaneous)	522	NA	NA
Software	623	Software and its engineering	GA 2 Sistemas de Software
Software	623	Hardware	GA 2 Sistemas de Software
Software	623	Computer systems organization	GA 2 Sistemas de Software
Statistics and Probability	17	NA	NA
Theoretical Computer Science	128	NA	NA
Total	6196	NA	NA
Water Science and Technology	12	NA	NA

A tabela de aprendizagem do K-nn

No total, a tabela de aprendizagem possui 6809 entradas ao final da execução do algorítimo de classificação. O que corresponde a 5,46% do total de 124601 observações. A taxa de classificação média para cada regra da tabela de aprendizagem é de 22 observações com uma relação de aproximadamente 5 aceites para 1 descarte.

Table 21: Distribuição dos valores da Tabela de Aprendizagem do K-nn.

Categoria	n	Sim	Não	Taxa	Média
Ambiente	553	10992	3143	3,5	19,9
Dados	421	7894	709	11,1	18,8
Estratégia de Validação/Coleta de Dados	1504	33381	5641	5,9	22,2
Metodologia	1068	14270	5029	2,8	13,4
Método	2177	43533	11109	3,9	20,0
Natureza	189	13550	1894	7,2	71,7
Propósito paradigma	242	1888	1392	1,4	7,8
Prova Científica Secundária	202	15926	229	69,5	78,8
Prova Científica	77	1307	189	6,9	17,0
Tipo de Validação do Resultado	385	6177	2498	2,5	16,0
Total	6818	148918	31833	4,7	21,8

Fim