

Um estudo sobre menções de métricas computacionais, termos e testes estatísticos em artigos apresentados no WSCAD

Rodrigo S. Nurmberg¹, Rogério L. Rizzi¹, Claudia B. Rizzi¹,
Guilherme Galante¹, Thiago S. Elias, Gustavo R. Malacarne¹¹

¹ Programa de Pós-Graduação em Ciência da Computação (PPGComp)

Universidade Estadual do Oeste do Paraná (UNIOESTE)

R. Universitária, 1619 - Universitário - Cascavel - PR - Brasil - CEP: 85819-110

{rodrigo.nurmberg, rogerio.rizzi, claudia.rizzi}@unioeste.br
{guilherme.galante, gustavo.malacarne, thiago.elias}@unioeste.br

Abstract. *This paper presents and discusses the results of a survey on mentions of computational metrics, terms and statistical tests in articles presented at WSCAD in the period from 2000 to 2022. The strategies were centered on the choice of terms through gold standard lists in the domains of interest, the use of regular expressions for data extraction and the statistical analysis of the results. Under the methodological procedures followed, the statistical techniques adopted indicated that there is no difference in the frequency of mentions of the analyzed terms in the articles published in this event.*

Resumo. *Neste trabalho são apresentados e discutidos resultados de uma pesquisa sobre menções de métricas computacionais, termos e testes estatísticos em artigos apresentados no WSCAD no período de 2000 a 2022. As estratégias centraram-se na escolha dos termos por meio de listas padrão ouro nos domínios de interesse, no emprego de expressões regulares à extração de dados e na realização da análise estatística dos resultados. Sob os procedimentos metodológicos seguidos, as técnicas estatísticas adotadas indicaram não haver diferença quanto a frequência de menções dos termos analisados nos artigos publicados neste evento.*

1. Introdução

Uma sistemática discussão dos aspectos metodológicos que embasam um trabalho acadêmico ou técnico, assim como a análise e avaliação quantitativa e qualitativa dos resultados obtidos, são indispensáveis para fornecer maior confiabilidade à pesquisa apresentada, e também para viabilizar sua reprodutibilidade por pares e interessados.

Não é do escopo deste trabalho discutir prováveis causas ou falhas na irreprodutibilidade de trabalhos técnicos e científicos. Mas se pode indicar que algumas estão indubitavelmente associadas aos aspectos do delineamento do experimento ou estudo e da análise estatística dos dados. Discussões sobre a relevância do planejamento experimental e da análise estatística de dados na Ciência da Computação são realizadas há anos, já tendo uma literatura técnica consolidada, a exemplo do livro [Wohlin et al. 2000]. Considera-se, pois, que pertinentes análises qualitativa ou quantitativa devem ser realizadas, não sendo suficiente, em geral, considerar no processo avaliativo apenas métricas específicas da Computação, pois elas podem depender do material ou técnicas empregadas no estudo, a exemplo da arquitetura ou das tecnologias, entre outros fatores. Ademais, não são

discutidos significados, conceitos ou técnicas estatísticas ou de avaliação de desempenho computacional, pois uma apresentação a respeito, no contexto do trabalho, já foi realizado por Osório e colaboradores, quando analisaram trabalhos submetidos ao WSCAD no período 2000 a 2017 [Osorio et al. 2018].

Este estudo tem como objetivo avaliar o acolhimento pelo WSCAD da indicação de [Osorio et al. 2018] de que *"...artigos apresentando análise de desempenho venham acompanhados de validação estatística de seus resultados ao mesmo tempo em que solicita aos revisores destes artigos que observem se tal estudo foi devidamente apresentado"*. Para a concretização deste trabalho foram processados textos disponíveis nos anais do WSCAD. Porém, como alguns dos aspectos metodológicos deste trabalho diferem daqueles dos autores, foram processados textos disponíveis no período de 2000 a 2022, e não apenas entre 2018 e 2022, já que desse modo pode-se processar também os artigos em inglês e fazer discussões mais abrangentes, mesmo que não se possa comparar os resultados diretamente com os obtidos por [Osorio et al. 2018].

Este trabalho está dividido em seções, sendo que a Seção 2 apresenta alguns trabalhos relacionados. A Seção 3 apresenta os materiais e métodos empregados para desenvolver esta pesquisa. Os resultados são discutidos na Seção 4. Na Seção 5 são realizadas algumas considerações gerais e indicados trabalhos em andamento.

2. Trabalhos Relacionados

A condução metodológica embasou a realização de estudo de menções de específicas métricas computacionais e de termos e testes estatísticos. Para melhor conduzir as atividades e definir o escopo deste trabalho, realizou-se um levantamento de produções relacionadas com a temática. Foram recuperados trabalhos em Ciência da Computação cujos enfoques metodológicos aderem àquele de interesse do presente estudo, e as buscas foram realizadas nas bases da Scielo Brasil e Periódicos CAPES. As palavras-chave empregadas foram meta-análise, métricas e testes estatísticos.

Em [Andujar et al. 2013] é destacada a importância da experimentação em Computação, que deve ser reconhecida em todos os ramos desta Ciência como uma metodologia fundamental, com regras e práticas bem definidas. Tais experimentos devem seguir protocolos sólidos, replicáveis e verificáveis, assim como nas Ciências e nas Engenharias. Discute-se que tal abordagem melhora a confiabilidade dos resultados obtidos, a validação de hipóteses e a tomada de decisões baseada em dados.

No estudo de [Santos 2018], é apresentada uma análise comparativa entre os processos linguístico e estatístico para extração de termos. Os autores constroem seus experimentos a partir de artigos e indicam que a abordagem estatística é uma opção relevante no contexto, pois viabiliza tratar a ambiguidade e a variabilidade presentes na linguagem natural, além de possibilitar a identificação de relações semânticas e contextuais entre os termos. Os autores indicam que a abordagem estatística deve ser explorada em aplicações na Área de Computação e afins.

Em [Osorio et al. 2018], foi realizada uma meta-análise qualitativa cujo objetivo foi extrair informações sobre métodos estatísticos e métricas utilizadas em artigos publicados no WSCAD entre 2000 e 2017. Foram utilizadas técnicas de mineração de dados e análise de textos, viabilizando a extração automática de dados de interesse. Os resultados

obtidos foram sistematizados e analisados de forma crítica, buscando a identificação de tendências e de lacunas ao emprego da análise estatística nos resultados.

O estudo apresentado em [Detoni 2020] forneceu uma análise quantitativa das técnicas frequentemente utilizadas na Engenharia de Software, buscando auxiliar na correta utilização e interpretação dos resultados dos testes estatísticos. O estudo apresentou e exemplificou técnicas estatísticas de análise de dados, buscando fornecer uma base à análise de resultados em estudos de Engenharia de Software. A análise sistemática das técnicas estatísticas apresentadas pode contribuir à identificação de métodos adequados aos diferentes contextos de pesquisa e à geração de resultados confiáveis e robustos.

3. Materiais e Métodos

Esta pesquisa tem por objetivo investigar se os trabalhos apresentados no WSCAD após 2018 realizam análises estatísticas na avaliação dos seus resultados. Busca-se verificar se existe diferença estatisticamente significativa em tal aspecto após a publicação de [Osorio et al. 2018], no qual os autores identificam baixa utilização de tais análises nos trabalhos publicados no WSCAD até 2017, e dada sua importância eles alertaram a comunidade quanto à relevância da validação estatística dos resultados, sugerindo que ela integrasse os requisitos de avaliação pelos revisores de trabalhos do evento. Assim, nesta seção são discutidas a abordagem metodológica e as estratégias procedurais empregadas da extração até a análise dos dados, para atingir o objetivo proposto.

3.1. Recuperação da Informação

A recuperação de informações (RI), é um campo de estudo que se ocupa da recuperação de documentos que satisfazem uma necessidade informacional, geralmente por conterem termos de interesse [Manning et al. 2008]. Assim sendo, técnicas de RI são empregadas quando deseja-se localizar, automaticamente, documentos contendo determinado conjunto de palavras-chave. Ao conjunto de documentos que são objeto de estudo pertencentes a um mesmo domínio, dá-se o nome de *corpus*. E entre os conceitos de RI, pode-se destacar 3 utilizados neste trabalho, sendo descritos conforme [Manning et al. 2008]:

- Frequência (absoluta) de termos ($tf_{t,d}$), que corresponde ao número de vezes que um determinado termo ocorre em um documento;
- Frequência de documentos (DF_t), que refere-se ao número de documentos do *corpus* que contém um dado termo;
- Matriz de incidência de termos \times documentos (TDM) é aquela em que as linhas correspondem aos termos e as colunas aos documentos. Cada célula t, d possui um valor inteiro associado, 0 caso o termo não ocorra no documento, 1 ou o valor $tf_{t,d}$ na ocorrência do termo. Os dados também podem ser representados de forma transposta, numa matriz de incidência documentos \times termos (DTM), em que as linhas correspondem aos documentos e as colunas aos termos.

3.2. Termos

Termos são unidades léxicas portadoras de conceitos, utilizadas para expressar conhecimento dentro de um domínio específico, tal como na Ciência da Computação, e cuja identificação e seleção são fundamentais na construção de recursos computacionais que auxiliam no processamento automatizado de textos, tais como taxonomias e ontologias

[Conrado et al. 2014]. Também são importantes em tarefas computacionais, como a recuperação de informação.

De acordo com a Sociedade Americana para Catalogação¹, dentro do campo da organização de conhecimento (*Knowledge Organization Scheme - KOS*) em Biblioteconomia e Ciências da Informação, são relevantes os conceitos de:

- Vocabulário controlado, que é uma lista autoritativa de termos utilizada para assegurar uma indexação consistente de documentos, seja ela humana ou automatizada. Engloba taxonomias e *thesauri*;
- Taxonomia é um vocabulário controlado, organizado de forma hierárquica por meio de relações do tipo pai/filho ou mais amplo (*Broader term - BT*)/mais específico (*Narrower term - NT*);
- *Thesaurus* é um vocabulário controlado, que além das relações hierárquicas *BT/NT*, possui relações associativas, de termos relacionados (*Related term - RT*), e relações de equivalência/preferência, como usado para (*Used For - UF*) e use (*U*). Editores criam *thesauri* detalhados para indexação de periódicos.

No campo da representação do conhecimento (*Knowledge Representation - KR*), em Inteligência Artificial, é usual o conceito de ontologia, que são modelos abstratos da realidade, compostos por um sistema estruturado de conceitos, compreendendo processos, objetos e atributos de um domínio, e seus inter-relacionamentos [Li et al. 2008].

Diferentemente dos *thesauri*, em que os conceitos são conectados apenas por relações independentes de domínio, as ontologias podem consistir de diversas taxonomias cujos conceitos são conectados por múltiplas relações específicas do domínio, [Li et al. 2008], semanticamente mais rigorosas e formais. Em geral, as ontologias podem ser reutilizadas entre diversas aplicações e comunidades [Uschold e Gruninger 2004]. Por serem produzidos por diversos especialistas em um domínio, com elevado nível de concordância entre eles, tanto os *thesauri* quanto as ontologias podem ser considerados conjuntos de termos padrão ouro² dentro da recuperação de informação.

3.3. Procedimentos

Considerando essas discussões, e suas implicações, elaborou-se uma metodologia cujos procedimentos executados foram agrupados em duas macro etapas, que encontram-se descritas nas Seções 3.3.1 Pré-Processamento e 3.3.2 Processamento e Análise. Todas as etapas de processamento dos dados foram realizadas utilizando-se a linguagem Python, no ambiente de desenvolvimento colaborativo Google Colab³. Adicionalmente foram utilizadas planilhas eletrônicas do Google Sheets para parametrização dos métodos, as quais foram lidas utilizando-se o pacote Python *gsread*.

3.3.1. Pré-processamento

Sendo o objeto de estudo o conjunto de artigos publicados no WSCAD, deu-se início a coleta dos dados, que foram obtidos dos anais do evento, disponibilizados na Biblioteca

¹<https://www.taxonomies-sig.org/about.htm>

²Termos que especialistas no domínio concordam serem relevantes e representativos da área.

³Repositório do projeto disponível em: <https://github.com/rsn86/WSCAD2023>

Digital da Sociedade Brasileira de Computação (SBC Open Lib - SOL)⁴.

Dentre os arquivos disponibilizados pretendeu-se analisar apenas artigos completos, sendo necessário filtrar os classificados como resumos (10 arquivos), e os conteúdos editoriais, como aberturas e apresentações (10 arquivos). Também foram filtrados os arquivos PDF cujo formato não possibilitava a conversão para TXT pelo pacote tika-python, por utilizarem uma codificação customizada⁵, interpretável apenas internamente, ao contrário das codificações ASCII e Unicode, ou por estarem em formato de imagem. Esses arquivos poderiam ser processados utilizando-se o reconhecimento óptico de caracteres (OCR), mas optou-se por excluí-los, visto que foram encontrados 5 arquivos nessa situação, distribuídos em 5 edições distintas, e processá-los não alteraria os resultados. Para mitigar o viés, o artigo [Osorio et al. 2018] também foi excluído das análises.

Por conseguinte, de um total de 545 artigos nos anais do WSCAD entre os anos de 2000 a 2022, após a aplicação dos filtros restaram 519 arquivos, dos quais 87 escritos em inglês, e os demais 432 em português. Tais artigos encontravam-se em formato PDF, sendo necessário convertê-los em texto plano TXT, cujos arquivos resultantes foram gerados na codificação Unicode.

Como a codificação Unicode permite representar um mesmo caractere de mais de uma maneira, aplicou-se a normalização NFKC realizando-se a decomposição canônica dos caracteres em sequências de códigos caractere+diacrítico, transformando-os na Forma Normal Decodificada (NFD). Em seguida combinando-se os códigos, foi gerado o caractere pré-codificado de código único na Forma Normal Codificada (NFC) e então, unificaram-se os caracteres semelhantes de acordo com uma tabela de equivalências K. Para isso foi utilizada a função `normalize`⁶ do módulo Python `unicodedata`.

Com os caracteres codificados de maneira única, pode-se utilizar expressões regulares (ER) mais concisas, sem considerar as múltiplas representações dos caracteres, e dispensando a necessidade de suporte avançado a Unicode pelo módulo de processamento de expressões regulares. Com essa abordagem, evitou-se a ocorrência de falsos negativos⁷, na qual alguns termos não podiam ser localizados nos documentos devido à divergência na representação dos caracteres das buscas e dos caracteres nos documentos.

Após a normalização, aplicou-se um conjunto expressões regulares para remover linhas contendo apenas caracteres não alfanuméricos e linhas em branco, além de unificar palavras que passaram por translineação (que é o processo no qual uma palavra tem suas sílabas divididas por um hífen com uma parte ficando no final de uma linha e a outra no início da linha subsequente), removendo-se hifens seguidos de quebras de linha para concatenar as partes da palavra. Também foram removidos resumos, *abstracts* e referências bibliográficas e textos editoriais, referentes à edição do evento ou à publicação em si. Ao final desse processo, restou apenas o texto do corpo dos artigos completos. O resultado foi salvo em arquivos TXT, gerando o *corpus* a ser analisado neste trabalho. Todas essas etapas encontram-se esquematizadas na Figura 1.

⁴<https://sol.sbc.org.br/index.php/wscad/issue/archive>

⁵[https://pdfbox.apache.org/2.0/faq.html#how-come-i-am-getting-gibberish\(g38g43g36g51g5\)-when-extracting-text%3F](https://pdfbox.apache.org/2.0/faq.html#how-come-i-am-getting-gibberish(g38g43g36g51g5)-when-extracting-text%3F)

⁶<https://docs.python.org/3/library/unicodedata.html#unicodedata.normalize>

⁷Experienciados durante a utilização da funcionalidade de busca em algumas ferramentas PDF, como os leitores de PDF do Google Chrome e o Foxit Reader.

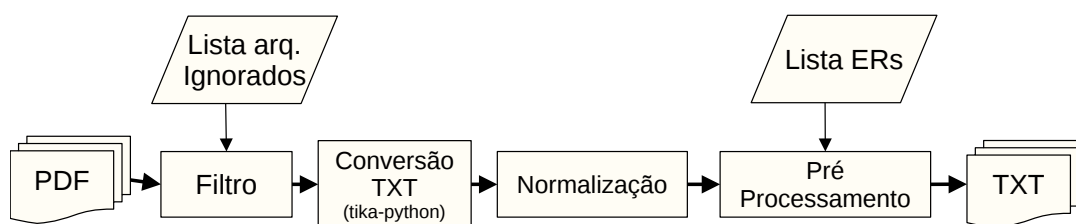


Figura 1. Etapas do pré-processamento dos dados.

3.3.2. Processamento e Análise

A seleção de termos adequados constitui etapa crítica para o sucesso do processo de recuperação da informação, de modo que passou-se a buscar fontes padrão ouro para termos relacionados com análise estatística de dados e métricas computacionais.

Na ausência de lista padrão ouro às métricas computacionais, recorreu-se à uma técnica adotada no campo da RI, na qual a lista padrão ouro é gerada com base nos elementos comuns entre os sumários de livros textos de um domínio [Conrado et al. 2014], simulando o consenso entre autoridades no assunto. Desta forma, com base no sumário de livros texto de experimentação computacional, foi gerada uma lista padrão ouro de termos para métricas, mimetizando uma eventualmente estabelecida por especialistas do domínio. Para atualizar a lista com os progressos tecnológicos, utilizaram-se folhas de especificações de equipamentos de tecnologia da informação e comunicação (TIC).

Para estabelecer uma lista padrão ouro aos termos e testes estatísticos, no âmbito da Computação e Simulações Computacionais, foi adotada a taxonomia do IEEE⁸ e a ontologia STATO⁹. A taxonomia do IEEE é derivada dos três primeiros níveis hierárquicos de cada família de termos do *thesaurus* do IEEE, que contém cerca de 11.570 termos descritivos de engenharia, técnicos e científicos, construído por especialistas no domínio e em análise de informação, com o objetivo de facilitar a busca e recuperação de informações nos campos de Engenharias e Computação. STATO é uma ontologia estatística de propósito geral, cobrindo testes estatísticos e suas condições de aplicação, assim como informações necessárias ou resultantes de métodos estatísticos. Também cobre aspectos de planejamento estatístico-experimental, e entre os objetivos e casos de uso documentados estão a associação entre os resultados e as hipóteses sob avaliação no experimento e prover um vocabulário especializado para mineração de textos.

Com o auxílio de um especialista, foram selecionados termos, testes e métricas provavelmente usuais no contexto do Simpósio, e para cada uma elaborou-se expressões regulares associadas. Essas ER foram agrupadas em três categorias, apresentadas nas Tabela 1 de métricas computacionais, Tabela 2 de termos estatísticos e Tabela 3 de testes estatísticos. Por uma limitação de espaço, foram apresentadas apenas as 3 primeiras linhas de cada tabela. As tabelas completas podem ser consultadas no repositório do artigo.

Note que apesar de as listas de termos estarem em inglês, cada ER contempla tanto o idioma inglês, quanto o português, uma vez que em Computação é usual utilizar-se de termos em língua inglesa ao redigir documentos em português.

⁸<https://www.ieee.org/publications/services/thesaurus-access-page.html>

⁹<https://stato-ontology.org/>

Tabela 1. Amostra das métricas selecionadas para coleta de dados

Descrição	Chave de Pesquisa
bandwidth (BW)	(bandwidth) OR (largura[\\s—]+de[\\s—]+banda)
bits per second (BPS)	((?:[kmgtp](?:i)?)?b(its OR ytes)?(?:/s(?:ec OR econds? OR eg OR egundos)? OR ps))
Observational error (OBERR)	(accuracy OR precision) OR (acur[áa]cia OR precis[ãa]o)
...	...

Tabela 2. Amostra dos termos estatísticos selecionados para coleta de dados

Descrição	Chave de Pesquisa
Probability (PR)	(probability) OR (probabilidades?)
Probability Distributions (DIST)	((?:(:hiper)?exponential OR (:Log[\\s—]*)?normal OR Gaussian OR Weibull OR binomial OR bernolli OR poisson OR pareto)[\\s—]+distribution) OR (distribui[çc][ãa]o[\\s—]+(?:(:hiper)?exponencial OR (:log[\\s—]*)?normal OR Gaussiana OR Weibull OR binomial OR bernolli OR poisson OR pareto))
Correlation (COR)	(correlation[\\s—]*(?:coeficientes?)) OR ((?:coeficientes?[\\s—]+de)[\\s—]*correla[çc][ãa]o)
...	...

Tabela 3. Amostra dos testes estatísticos selecionados para coleta de dados

Descrição	Chave de Pesquisa
statistical hypothesis test (HT)	((?:hypothes[ie] OR statistical)s?[\\s—]+(?:test OR evaluation)s?) OR ((?:testes?)\\s—]+(?:de)?[\\s—]+(?:hip[óo]tese OR estat[íi]tico)s?)
Student's t-test (TTST)	((?:Student[\\s—]*s[\\s—]*)?\\bt[\\s—]+test) OR (teste[\\s—]+t\\b(?:[\\s—]+de[\\s—]+Student)? OR t[\\s—]+student)
chi square test (CST)	(chi[\\s—]+square) OR ((?:chi OR qui)[\\s—]+quadrado)
...	...

A coleta de dados foi realizada empregando-se expressões regulares, ao invés de se utilizar diretamente os termos, pois elas viabilizam melhor controle sobre os padrões de correspondência, evitando, a exemplo, falsos positivos advindos de correspondências indesejadas entre palavras chaves e o texto, como no caso do termo "valor p " e textos contendo a mesma sequência de caracteres, como "valor previsto" ou "valor pequeno".

Considerando as listas categorizadas de ERs, prosseguiu-se com a extração de termos dos documentos do *corpus* que foi produzido durante o pré-processamento, contabilizando-se o número de ocorrências de cada palavra-chave por documento ($tf_{t,d}$), por meio da função *findall* do módulo *regex*¹⁰. Obteve-se, assim, uma matriz de incidência de documentos \times termos (DTM) como um dicionário Python, posteriormente convertido em um *DataFrame* Pandas para realizar as manipulações e tabulação dos dados.

Como o objetivo do trabalho é verificar se métricas e testes estatísticos estão sendo ou não mencionados, independentemente de quantas vezes estão presentes nos artigos, calculou-se a frequência de documentos (DF_t). Esse tratamento também viabilizou que fossem ignorados eventuais problemas que ocorrem na conversão do PDF em TXT, no

¹⁰<https://pypi.org/project/regex/>

tocante aos textos em gráficos e tabelas, pois o impacto restringe-se na redução do valor de $tf_{t,d}$, sem zerá-lo, uma vez que esses elementos são discutidos no corpo do texto¹¹.

Para a análise foram gerados gráficos¹² para observar as tendências, e realizados testes de hipóteses¹³ em que se considerou como hipótese nula que não existe diferenças significativas entre menções dos termos analisados "até" e "após" o ano de 2018, sendo a hipótese alternativa que existe diferença estatística significativa entre esses grupos de dados. A Figura 2 ilustra o fluxo descrito às etapas de processamento e análise dos dados.

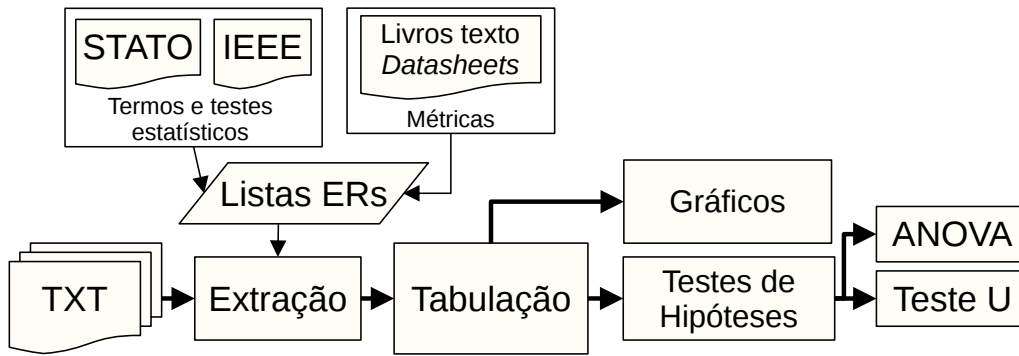


Figura 2. Etapas do processamento e análise dos dados.

Os resultados são apresentados e discutidos na Seção 4 perante uma Análise de Variância (ANOVA), que testa se dois ou mais grupos possuem iguais médias populacionais. Para sua aplicação é requerido que os erros experimentais estejam normalmente distribuídos (uma verificação pode ser realizada com o Teste de Shapiro-Wilk), e que as variâncias entre os grupos sejam iguais (sua verificação pode ser efetuada com o Teste de Levene), sendo tais igualdades no contexto estatístico, com nível de significância de $\alpha = 0,05$. E se alguma das pressuposições ao emprego da ANOVA não for satisfeita pode-se utilizar um teste não paramétrico, como o Teste U (Mann-Whitney), para investigar se dois grupos possuem mesma distribuição de dados.

4. Resultados e Conclusões

Cada lista de ERs apresenta um conjunto de termos alternativos para cada categoria, e como o objetivo não é determinar quais métricas ou testes são efetivamente utilizados, mas apenas se são mencionados ou não nos trabalhos, os dados foram agrupados por ano e por categoria, resultando na Tabela 4. As 2 primeiras colunas, ano e n , correspondem ao ano do evento e ao número de artigos analisados da Edição. São seguidas por 3 grupos de colunas referentes às categorias de ERs, em que cada grupo contém 2 colunas, designando, respectivamente, o número de artigos que apresentam termos referentes à categoria ($\sum DF_t$) e número de citações dos termos pertencentes à categoria ($\sum tf_{t,d}$).

Os gráficos das Figuras 3(a) e 3(b), apresentam, respectivamente, os percentuais dos artigos analisados, por ano que mencionam termos de cada categoria ($\sum DF_t/n$), e a razão entre o número de citações e o número de artigos analisados por ano ($\sum tf_{t,d}/n$).

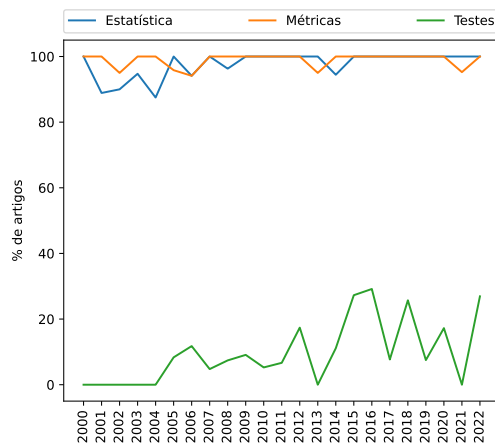
¹¹Conforme validação manual de 40 artigos sorteados aleatoriamente e distribuídos entre os autores.

¹²Com auxílio da biblioteca matplotlib.

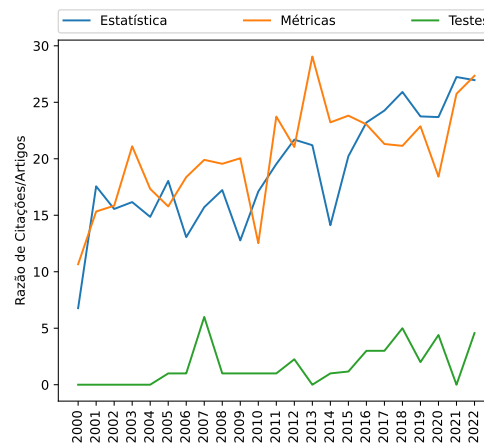
¹³Utilizando os pacotes Scipy e statsmodels.

Tabela 4. Distribuição de citações de 2000 a 2022 por tipo do termo e ano

Ano	n	Estatística		Métricas		Testes	
		$\sum DF_t$	$\sum tf_{t,d}$	$\sum DF_t$	$\sum tf_{t,d}$	$\sum DF_t$	$\sum tf_{t,d}$
2000	9	9	61	9	96	0	0
2001	18	16	281	18	276	0	0
2002	20	18	280	19	301	0	0
2003	19	18	291	19	401	0	0
2004	24	21	312	24	416	0	0
2005	24	24	433	23	363	2	2
2006	17	16	209	16	294	2	2
2007	21	21	330	21	418	1	6
2008	27	26	448	27	528	2	2
2009	22	22	281	22	441	2	2
2010	19	19	325	19	238	1	1
2011	15	15	293	15	356	1	1
2012	23	23	499	23	484	4	9
2013	20	20	424	19	552	0	0
2014	18	17	240	18	418	2	2
2015	22	22	445	22	524	6	7
2016	24	24	557	24	553	7	21
2017	26	26	631	26	554	2	6
2018	35	35	907	35	740	9	45
2019	40	40	950	40	915	3	6
2020	29	29	687	29	534	5	22
2021	21	21	572	20	515	0	0
2022	26	26	701	26	711	7	32
média	22,565	22,087	441,609	22,348	462,087	2,435	7,217
desv.pad.	6,423	6,612	221,299	6,513	177,108	2,660	11,747



(a) Proporção de artigos com citações por ano



(b) Razão de citações por artigo ano

Figura 3. Gráficos de resultados obtidos.

Como o número de artigos varia muito de edição para edição do WSCAD, a análise estatística foi conduzida com base na proporção de artigos, de cada ano, contendo correspondências às expressões regulares e se restringiu na avaliação se existe diferenças significativas às menções entre o período de 2000 a 2018, que se refere às edições até a publicação de [Osorio et al. 2018], e o período de 2019 a 2022, após tal publicação.

Observa-se que o número de amostras nos cenários "até" e "após" são muito diferentes e, provavelmente por essa condição, o teste de Shapiro-Wilk indicou pela rejeição da normalidade dos resíduos, de modo que tal critério à aplicação da ANOVA não foi satisfeito. Por este motivo optou-se pela aplicação do teste *U* exato¹⁴, que indicou pela igualdade das distribuições das amostras, podendo os resultados serem aceitos com o nível de significância adotado. Os resultados encontram-se na Tabela 5.

Tabela 5. P-valores dos testes realizados

Categoria	ANOVA	Shapiro-Wilk	Levene	Teste U
Métricas	0,910	0,000	0,910	1,000
Estatística	0,208	0,000	0,208	0,286
Testes	0,482	0,019	0,523	0,611

Pode-se afirmar, pois, que não existe diferenças estatísticas quanto nas frequências de menção dos termos estudados nos artigos deste evento no período de 2000 até 2022, sob os encaminhamentos dados neste trabalho, de modo que as análises estatísticas continuam sendo negligenciadas nos artigos do WSCAD. E, enfim, e não obstante a semelhança entre os métodos e objetivos deste trabalho e o elaborado por [Osorio et al. 2018], destaca-se que neste trabalho a escolha dos termos foi realizada por meio de listas padrão ouro nos domínios de interesse, e que foram empregados expressões regulares à mineração dos dados, processando-se também documentos em inglês. E que uma análise estatística básica foi realizada na verificação dos resultados.

5. Trabalhos em Andamento e Sugestões

Uma relevante questão apontada por [Osorio et al. 2018] é quanto "*...a necessidade da inclusão da metodologia de análise estatística de dados aplicada à computação nas disciplinas básicas de formação*". Um estudo a esse respeito foi realizado para identificar disciplinas (e conteúdos) afins ministradas por Instituições de Ensino Superior (IES) do Oeste do Paraná em cursos de graduação na Área da Computação. Por meio dele foram coletados dados de 24 IES da Mesoregião 6, sendo 10 públicas e 14 privadas. Tal coleta foi realizada por meio de um questionário semiestruturado, aplicado durante entrevistas efetuadas com os coordenadores de cursos, e ainda está sob análise. Outro estudo, em andamento, emprega a mineração automática de dissertações de mestrado de IES do Paraná. O delineamento experimental envolve a estratificação por conglomerados e o processamento dos dados será realizado empregando técnicas qualitativas e quantitativas.

Quanto aos aspectos metodológicos e estatísticos, pode-se indicar que não foram explorados os relacionamentos entre os termos da taxonomia e da ontologia, na busca por termos nos documentos. Como trabalho futuro indica-se que pode-se tirar proveito deles para melhorar a busca e recuperação de informações, como em [Li et al. 2008]. Os relacionamentos podem ser utilizados para expandir a lista de termos buscados ou para verificar se o contexto no qual são utilizados correspondem ao desejado, preservando seus conceitos. Além disso, este estudo focou a análise de menção de termos nos artigos, e não sua efetiva utilização, pois a menção de técnicas estatísticas e métricas computacionais não implica sua utilização na análise dos resultados, como mostrou uma rápida verificação em amostra aleatória de artigos e dissertações. Um refinamento adicional deve ser

¹⁴A forma exata é sugerida quando algum dos tratamentos possui menos de 8 amostras.

realizado enfocando os designados "Capítulos de Resultados". Também destaca-se que existe um forte desbalanceamento de dados intra edições e entre as edições do WSCAD, e esse aspecto requer uma reflexão para o uso de técnicas estatísticas mais elaboradas.

Enfim, ressalta-se que o padrão PDF/A (ISO 19005) na variante *u*, assegura a preservação de longo prazo de documentos eletrônicos, além de melhorar a acessibilidade dos documentos e a integração com leitores de tela, facilita o processamento automático de documentos. Sugere-se que seja adotado como requisito para publicação, a conformidade dos documentos com tal padrão. Outra sugestão é que o Simpósio WSCAD ofereça em suas próximas edições cursos específicos em análise estatística de dados, para mitigar a situação atual.

6. Agradecimentos

Agradecemos ao professor Gerson G. H. Cavalheiro e Alessandro Osorio, da UFPel, dois dos autores de [Osorio et al. 2018], pela paciência e disponibilidade em compartilharem seus conhecimentos e experiências.

Referências

- Andujar, C., Schiaffonati, V., Schreiber, F. A., Tanca, L., Tedre, M., van Hee, K., e van Leeuwen, J. (2013). The role and relevance of experimentation in informatics. *8th European Computer Science Summit of Informatics Europe*.
- Conrado, M. D. S., Felippo, A. D., Pardo, T. A. S., e Rezende, S. O. (2014). A survey of automatic term extraction for Brazilian Portuguese. *Journal of the Brazilian Computer Society*, 20(1):1–28.
- Detoni, M. N. P. (2020). Mapeamento e aplicação de testes estatísticos em engenharia de software. Master's thesis, Pontifícia Universidade Católica do Rio Grande do Sul.
- Li, Z., Raskin, V., e Ramani, K. (2008). Developing Engineering Ontology for Information Retrieval. *Journal of Computing and Information Science in Engineering*, 8(1):011003.
- Manning, C. D., Raghavan, P., e Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press, New York.
- Osorio, A., Dias, M., e Cavalheiro, G. G. (2018). Meta-analysis of scientific articles according to statistical criteria: A case study in wscad. In *Anais do XIX Simpósio de Sistemas Computacionais de Alto Desempenho*, pages 124–129, Porto Alegre, RS, Brasil. SBC.
- Santos, C. A. d. (2018). Uma análise comparativa entre as abordagens linguística e estatística para extração automática de termos relevantes de corpora. Master's thesis, Pontifícia Universidade Católica do Rio Grande do Sul.
- Uschold, M. e Gruninger, M. (2004). Ontologies and semantics for seamless connectivity. *SIGMOD Rec.*, 33(4):58–64.
- Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., e Wesslén, A. (2000). *Experimentation in software engineering: an introduction*. Number 6 in International Series in Software Engineering. Springer, New York, NY.