

Scientific Recommender: Análise Inicial Sobre Artigos Publicados em Eventos

Scientific Recommender: Initial Analysis On Articles Published in Events

Ingrid Pacheco ¹, Eduardo Prata, Renan Parreira

¹Programa de Pós-Graduação em Informática – Instituto de Computação
Universidade Federal do Rio de Janeiro (UFRJ)
Rio de Janeiro – RJ – Brasil

{ingridpacheco}@dcc.ufrj.br

Abstract. *As a daily task, plenty of researchers submit articles to events and magazines hoping that their studies will be shared with more people. However, the frustration of being rejected can discourage them and overshadow their research. In the present work, analyses are performed on published articles over a lot of conferences in order to recognize patterns through their histories and provide a foundation for the creation of a recommendation system that will assist them in the search of the perfect event.*

Keywords. *Recommendation Systems; Clusterization; Published Articles; Data Science ; Data Analysis.*

Resumo. *Como uma tarefa rotineira, diversos pesquisadores submetem artigos a eventos e revistas com a esperança de terem seus estudos compartilhados com mais pessoas. Entretanto, a frustração ao ser rejeitado pode acabar o desestimulando e ofuscando uma brilhante pesquisa. Neste trabalho, são feitas análises em cima de artigos publicados em diversas conferências a fim de reconhecer padrões através de seus históricos e prover uma base para a criação de um sistema de recomendação que auxilie na busca pelo evento perfeito.*

Palavras-Chave. *Sistemas de Recomendação; Clusterização; Artigos Publicados; Ciência de Dados; Análise de Dados.*

1. Introdução

Um dos passos mais importantes na vida de um cientista, acadêmico, aluno ou pesquisador é a publicação de um artigo científico. O artigo dá mais importância e faz os trabalhos criados ganharem mais sentido pois o mesmo pode beneficiar outras pessoas agregando mais conhecimento em seus trabalhos e/ou pesquisas. Se ninguém sabe ou não pode

ser beneficiado por trabalhos anteriormente elaborados, os mesmos têm suas relevâncias reduzidas.

Quando o pesquisador entende que seu trabalho já tem argumentos suficiente para responder as questões as quais ele se propõe a caracterizar e os problemas identificados possuem embasamento suficiente para serem resolvidos junto a suas hipóteses testadas, seria o momento perfeito para o mesmo partir para a publicação de seu artigo.

Um artigo científico deve ser relevante para os seus leitores e contribuir para novos conhecimentos. No entanto, existe uma grande preocupação por parte dos pesquisadores ao tentar a publicação de seus artigos pois alguns deles são negados.

Algumas situações podem ser citadas como exemplos no que podem influenciar na rejeição de artigos e não se restringem a somente estes, como “o artigo não contribuir para novos conhecimentos na disciplina ou na aplicação de conhecimento” ou “o artigo não atender aos padrões éticos” ou até mesmo “o artigo estar mal escrito”.

Além dos pontos citados, um outro grande fator que pode contribuir para a rejeição de um artigo seria por parte da relevância oferecida para os seus leitores.

Para ajudar neste ponto, uma ferramenta de sistema de recomendação se faz fundamental para entender as necessidades dos pesquisadores e trazer para eles as melhores indicações de eventos nos quais os seus artigos teriam uma maior probabilidade de serem aceitos. Entretanto, a fim de alcançar tal como um objetivo final, é preciso retroceder um pouco e entender como cada tópico se apresenta dentro de uma conferência, para entender os padrões delas, suas previsões e conseguir encontrar as melhores opções como resultado.

Neste contexto, procurando analisar os eventos e seus tópicos, são formuladas as seguintes perguntas de pesquisa a serem respondidas pelo presente trabalho:

- **P1** Quais são os 6 tópicos mais publicados nos eventos?
- **P2** Para cada tópico, quais os eventos que mais os publicam?
- **P3** Como foram as evoluções ao longo dos anos de publicações dos eventos para cada tópico?

Visando responder as questões de pesquisa elencadas e considerando a falta de um conjunto de dados único com as informações necessárias, o presente trabalho construiu um *dataset* com dados de artigos publicados em diversas conferências e realizou análises comparativas sobre eles para visualizar os atuais padrões.

O restante deste artigo está organizado da seguinte forma. Na seção 2, é feita uma fundamentação teórica dos principais conceitos relacionados ao trabalho. Na seção 2.1 é apresentado o conceito principal de um sistema de recomendação. Na seção 2.2, por sua vez, é explicado o que é o método de clusterização, relevante para a análise posterior. Na seção 3, são apresentadas as informações sobre a criação dos conjuntos de dados. Na seção 4, são apresentadas as análises realizadas e os resultados obtidos. Por fim, na seção 5, são apresentadas as conclusões e os trabalhos futuros.

2. Fundamentação Teórica

Nesta seção, são apresentados alguns conceitos fundamentais que serão abordados ao longo do trabalho, incluindo o que é um sistema de recomendação e clusterização.

2.1. Sistema de Recomendação

Com o problema de sobrecarga de informação, os Sistemas de Recomendações se tornaram uma importante área de pesquisa ao longo dos anos, inicialmente chamados de Sistemas de Filtragem Colaborativa devido à utilização no primeiro sistema de recomendação Tapestry [GOLDBERG 1992], onde usuários indicavam documentos entre si, desenvolvido pela Xerox.

Os Sistemas de Recomendação são nada mais que classes de algoritmos com inteligência artificial que recomendam itens relevantes aos utilizadores, sendo utilizadas diferentes técnicas para a filtragem e ranqueamento da informação que observam vários fatores, como por exemplo, o histórico das interações dos utilizadores[]. Investigações associando os Sistema de Recomendação com novos recursos tecnológicos e o desenvolvimento de projetos e de produtos em diferentes tecnologias estão em crescimento nos últimos anos [Shao 2021]

A utilização mais comum nos dias de hoje está presente em sites de e-commerce ou canais de streaming, baseando suas recomendação pelos dados obtidos através do navegador, buscas e avaliações de produtos e serviços pela internet, de forma implícita.

Mas também podendo ser aplicada, de forma explícita, em outras áreas como a descrita neste artigo, onde será utilizado como ferramenta de recomendação para onde submeter um Artigo Científico buscando as melhores oportunidades, usando as palavras chaves do artigo como entrada para o Sistema de Recomendação será avaliado e analisado em um base de artigos.

2.2. Clusterização

Tendo em vista que todos os dados contáveis e quantificáveis que possam ser agrupados, podem passar pelo processo de Clusterização, que nada mais é do que um agrupamento desses dados buscando alguma similaridade. Sendo assim, identificamos sua primeira e mais simples utilização que é no agrupamento de perfis ou de indivíduos com características em comum, visando representar padrões desses grupos de indivíduos, por exemplo, na identificação de um público alvo para uma campanha publicitária ou até mesmo para uma pesquisa de mercado, ou seja, a clusterização de clientes, clusterização de produtos, servidores, de dados.

Por sua vez, os clusters precisam de algoritmos com objetivo de expressarem tais critérios em forma de gráfico, como se os objetos a serem clusterizados se agrupem conforme uma maior ou menor distância em relação uns aos outros, de acordo com as características em comum.

Os algoritmos mais usados são:

- K-means Clusters: Os clusters podem ser dispostos em gráficos hipotéticos, onde o centro de cada cluster, chamado de centroide, é calculada a média dos valores

para o mesmo e desse modo o algoritmo buscará o centroide mais perto, usando métricas de distância e o atribuindo um ponto encontrado ao cluster mais próximo.

- Clusterização hierárquica: No Clustering hierárquico o algoritmo, constrói uma hierarquia disposta em forma de dendrograma.

3. Construção do Conjunto de Dados

Nesta seção, será descrito o conjunto de dados gerado para realização das análises. Ele contém dados de diversos artigos publicados em 10 dos 11 principais eventos de Ciência da Computação em um intervalo de 5 anos (2018 até 2022). Posteriormente, o mesmo é utilizado para compreender a presença de cada temática nestes eventos e seu comparativo histórico ao longo dos anos.

Em relação a análise proposta, era necessária a construção de um conjunto de dados que contivesse todas as informações requeridas, como conferência, artigo, ano, tópicos, dentre outros. Para tal, foram extraídos 10 dos 11 principais eventos que tiveram publicações no tema Ciência da Computação. Estes vieram de uma *lista do departamento de Ciência da Computação da Universidade Cornell*¹, e utilizaram como parâmetro a publicação de pelo menos 100 *papers* entre 2003 e 2013, com 5 citações cada durante os últimos 10 anos.

Os eventos utilizados para análise foram:

- *SOSP—ACM Symposium on Operating Systems Principles*
- *OSDI—Operating Systems Design and Implementation*
- *NDSS—Network and Distributed System Security Symposium*
- *MobiHoc—Mobile Ad Hoc Networking and Computing*
- *SIGCOMM—ACM SIGCOMM Conference*
- *SenSys—Conference On Embedded Networked Sensor Systems*
- *MOBICOM—Mobile Computing and Networking*
- *CIDR—Conference on Innovative Data Systems Research*
- *USENIX Security Symposium*
- *EUROCRYPT—Theory and Application of Cryptographic Techniques*

A partir das conferências selecionadas, o próximo passo foi coletar os artigos publicados em cada entre os anos de 2018 e 2022. Portanto, foi utilizada a fonte *dblp computer science bibliography*, que contém mais de 6 milhões de publicações e quase 6000 conferências mapeadas, com suas respectivas edições e conteúdos.

Entretanto, pelo fato de tal biblioteca digital não possuir uma *API* para facilitar a busca, foi necessário criar um *web scrapper* que passando por algumas propriedades da página buscava cada uma das conferências mencionadas, depois entrava nas páginas de cada edição dentro do intervalo de tempo estabelecido e por fim, coletava as informações pertinentes aos seus artigos publicados. Vale ressaltar que a cada interação com a página alguns dados eram salvos, como o nome da conferência ao buscar sobre ela, o ano da edição na listagem de edições e dados sobre o artigo na listagem dos mesmos.

As rotas utilizadas e suas respectivas responsabilidades foram estabelecidas da seguinte forma:

¹<https://www.cs.cornell.edu/andru/csconf.html>

- "<https://dblp.org/search?q=>": Consulta a biblioteca pelo nome da conferência para encontrar a listagem de edições dela.
- "<https://dblp.org/db/conf/confname>": Página com a listagem de edições da conferência. Ex: <https://dblp.org/db/conf/sosp/index.html>.
- "<https://dblp.org/db/conf/confname/confedition>": Página com a listagem de artigos publicados em determinada edição. Ex: <https://dblp.org/db/conf/sosp/sosp2021.html>.
- "<https://dl.acm.org/doi/articledoi>": Página com informações complementares sobre os artigos, como autores e *keywords*. Ex: <https://dl.acm.org/doi/10.1145/3477132.3483586>.

No fluxo em que o DOI não esteja atrelado ao artigo na listagem da edição, é feita uma busca auxiliar na *API do Semantic Scholar*² utilizando o título do artigo, de forma a encontrar o *paperId* do mesmo, para posteriormente ser possível complementar as informações.

Caso ainda assim não seja possível encontrar nem o *paperId* nem o DOI do artigo, os autores são coletados através de propriedades na própria listagem e o mesmo é salvo no dataset com as informações já existentes. Entretanto, no caso de ter sido encontrado o *paperId* ou DOI, uma busca complementar é realizada na *biblioteca do Semantic Scholar*³ de forma a pegar os autores, tópicos e campo de estudo.

Por fim, se o DOI tiver sido encontrado na própria página ou pela consulta mencionada anteriormente, a página da *DL ACM*⁴ com informações complementares do artigo é consultada, com o intuito de obter tanto os autores (de mais fácil acesso) quanto mais *keywords* para fazer com que a análise seja mais completa possível. Todas as informações são posteriormente anexadas e salvas em um arquivo CSV, de forma a representar um artigo completo.

O conjunto de dados foi criado em 19 de Setembro de 2022 no formato CSV, contendo 2783 artigos publicadas em 10 eventos (SOSP, OSDI, NDSS, MobiHoc, SIGCOMM, SenSys, MOBICOM, CIDR, USENIX Security Symposium, EUROCRYPT) durante os anos de 2018 a 2022. Além disso, as informações foram armazenadas da forma apresentada na Tabela 1. Por fim, o conjunto de dados e o *script* em *Python* utilizado para o processo de geração do arquivo CSV foi disponibilizado para consulta e utilização em futuras pesquisas no *GitHub*⁵.

²<https://api.semanticscholar.org/graph/v1/paper/search?query=>

³<https://pypi.org/project/semanticscholar/>

⁴<https://dl.acm.org/doi/>

⁵<https://github.com/ingridpacheco/Scientific-Recommender>

Tabela 1. Informações os artigos no dataset

Atributo	Descrição
title	Título do artigo
paperId	PaperId do artigo disponibilizado no Semantic Scholar
doi	DOI do artigo, ou identificador único
authors	Autores que escreveram o artigo
publisher	Conferência na qual o artigo foi publicado
topics	Keywords dos artigos
fields_of_study	Campo geral de estudo do artigo
year	Ano que o artigo foi publicado

4. Análises e Resultados

Nesta seção, são detalhadas as análises e resultados encontrados ao longo do trabalho, tanto com relação as conferências quanto aos tópicos e seus comparativos históricos.

Baseado no conjunto de dados gerado, foi criado um novo *script*⁶ para realizar as análises referentes as perguntas (especificamente de **P1** a **P6**) previamente estabelecidas (ver seção 1).

4.1. Caracterização do Dataset

Antes de efetivamente serem expostas as análises realizadas no *dataset*, algumas observações sobre ele merecem ser elucidadas, de forma que fique mais compreensível alguns resultados posteriores.

Apesar de o *web scraper* conseguir pegar uma boa parte dos dados dos artigos, nem todos eram disponibilizados nas páginas, o que acabou deixando certas brechas no *dataset* que tiveram que ser resolvidas manualmente nas análises. Outro ponto relevante é de que a análise inicial só foi feita em 10 conferências pois existiam limitações com relação a quantidade de requisições tanto na *API do Semantic Scholar* quanto ao site da DL ACM, que impossibilitavam uma raspagem maior de dados.

No total, além dos números apresentados anteriormente de 2783 artigos em 10 eventos entre 2018 e 2022.

4.2. Tópicos Que Mais Aparecem Em Cada Conferência

Considerando o objetivo final de se criar um sistema de recomendação de eventos para pesquisadores de diversas áreas, é importante compreender quais os principais tópicos abordados em cada conferência para que assim as recomendações sejam mais certas usando as temáticas de cada pesquisador e o que os eventos mais prezam.

Dessa forma, e para responder as perguntas seguintes, foi-se criado um novo *script*⁷ focado na análise do *dataset*, utilizando como auxílio as bibliotecas: *numpy*⁸,

⁶<https://github.com/ingridpacheco/Scientific-Recommender/blob/main/articleAnalysis.py>

⁷<https://github.com/ingridpacheco/Scientific-Recommender/blob/main/articleAnalysis.py>

⁸<https://numpy.org/>

pandas⁹, matplotlib¹⁰, bokeh¹¹ e plotly¹².

Para responder a **P1**, a biblioteca pandas foi utilizada na leitura do arquivo CSV criado com o dataset. Ela o transforma em um *DataFrame*¹³, estrutura de posterior fácil manipulação. Com este, um loop foi feito passando por cada linha (artigo) e um dicionário foi criado armazenando a quantidade de vezes que um tópico aparece (os tópicos tiveram que ser quebrados e armazenados um a um). Ao final da execução do mencionado trecho um dicionário da seguinte estrutura é disponibilizado:

```

1 {
2   'COMPUTER SYSTEMS ORGANIZATION' : 29,
3   'DEPENDABLE AND FAULT-TOLERANT SYSTEMS AND NETWORKS' : 18,
4   'SECURITY AND PRIVACY' : 25,
5   ...
6 }
```

A seguir, ele é ordenado de forma decrescente e os 6 tópicos com maior quantidade são recortados, gerando a listagem dos seguintes temas e respondendo assim a **P1**:

- 'EXPERIMENT': 154
- 'ALGORITHM': 150
- 'COMPUTER SCIENCE': 131
- 'SENSOR': 114
- 'PROTOTYPE': 113
- 'MACHINE LEARNING': 107

4.3. Eventos Que Mais Publicam Cada Tópico

A fim de compreender para cada um destes 6 tópicos sua distribuição nos 10 eventos outros dicionários tiveram de ser criados. O primeiro deles armazenava a correlação da quantidade de artigos publicados de cada tópico em cada evento em cada ano, como mostra a estrutura seguinte:

```

1 {
2   'NP-HARDNESS' : {
3     'MobiHoc' : {
4       2020: 2,
5       2019: 2,
6       2018: 2
7     },
8     'MOBICOM' : {
9       2018: 1
10    }
11  }
```

⁹<https://pandas.pydata.org/>

¹⁰<https://matplotlib.org/>

¹¹<http://bokeh.org/>

¹²<https://www.plotly.express/>

¹³<https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.html>

```

11 },
12 'ONLINE LEARNING SETTINGS': {
13     'MobiHoc ': {
14         2021: 1
15     }
16 }
17 ...
18 }

```

Já o segundo armazenava a correlação oposta, ou seja, a quantidade de artigos publicados de um determinado tópico em cada ano em cada evento, da forma:

```

1 {
2   'TROJAN HORSE (COMPUTING)': {
3     2018: {
4       'NDSS ': 1,
5       'USENIX Security Symposium': 1
6     },
7     2021: {
8       'USENIX Security Symposium': 1
9     },
10    2020: {
11      'USENIX Security Symposium': 1
12    },
13    2019: {
14      'USENIX Security Symposium': 3
15    }
16  },
17  ...
18 }

```

Com os dois dicionários já criados, para cada tópico do top 6 o processo foi pegar a quantidade por ano de cada evento, transformar em um *Dataframe*, agregar a soma de cada evento numa outra variável, fazer um recorte das quantidades de 2022 e enfim, criar um *Dataframe* com as colunas: 'Publisher' (evento), 'Total' e '2022 Qty' (Quantidade de publicações de determinado tópico no evento específico no ano de 2022).

A partir da estrutura criada acima foi possível tanto compreender como cada tema era abordado de uma maneira geral por cada conferência, respondendo a **P2** quanto fazer uma comparação histórica ao longo dos anos, respondendo a **P3**, para entender em quais ele mais aumentou, se manteve o mesmo ou diminuiu, como ilustrado na figura 4.3 para o caso do tópico EXPERIMENT.

Todos os gráficos produzidos para os outros 5 tópicos estarão na seção de anexo para futuras consultas. Porém, uma resumida análise sobre cada um deles é:

- Para o tópico EXPERIMENT, o evento USENIX Security Symposium foi o que mais publicou artigos, com uma quantidade total de 38 artigos, seguido por NDSS

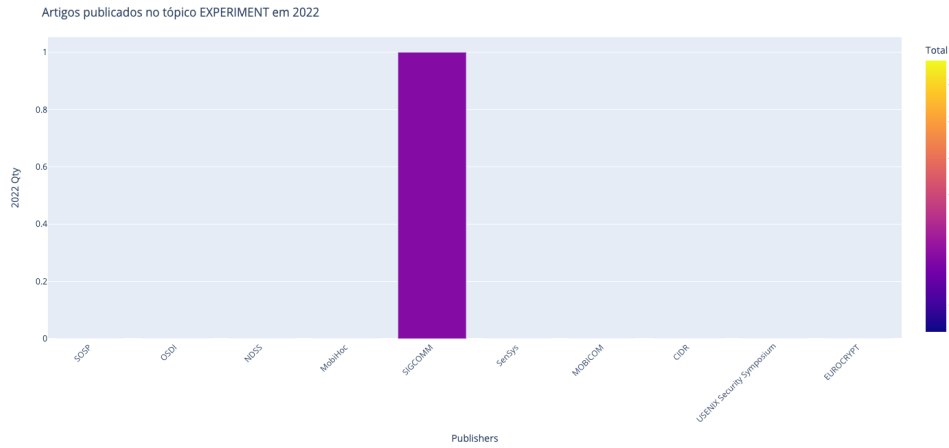


Figura 1. Artigos publicados sobre EXPERIMENT em 2022

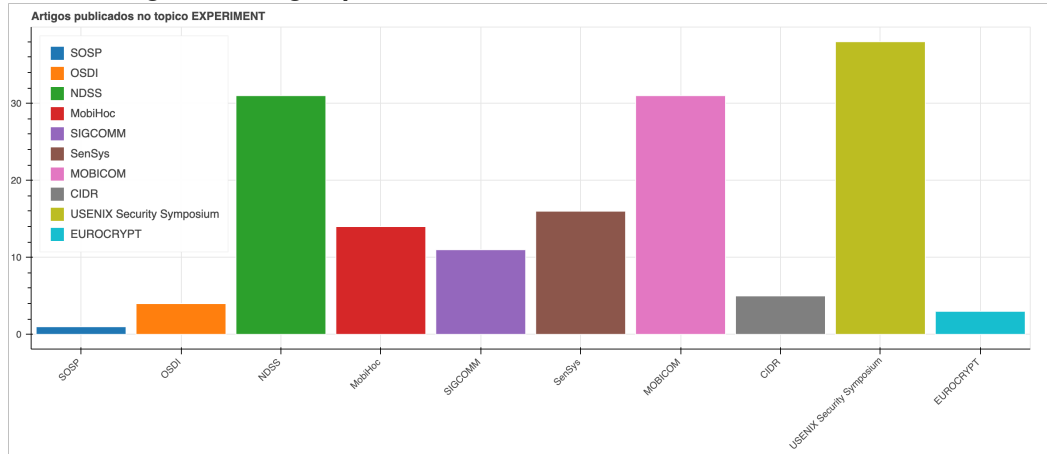


Figura 2. Artigos publicados sobre EXPERIMENT

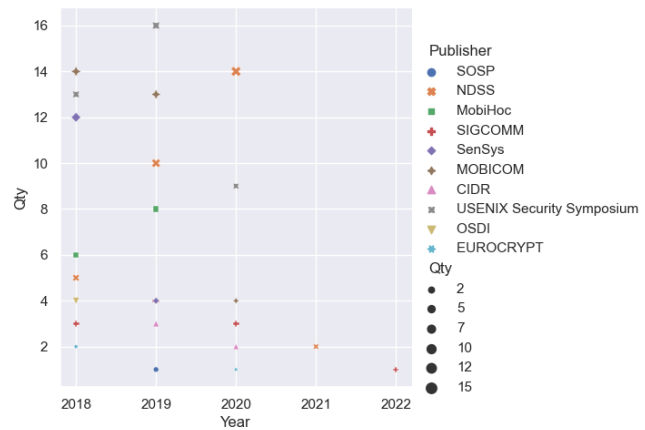


Figura 3. Comparativo hist rico dos eventos para artigos sobre EXPERIMENT

e MOBICOM, ambos com 31. Apesar de eles serem o que mais publicaram sobre o tema de uma maneira geral, no ano de 2022 apenas o SIGCOMM já teve alguma publicação com o tema. Por fim, o único evento que apresentou um aumento constante entre 2018, 2019 e 2020 foi o NDSS, mesmo a quantidade tendo diminuído em 2021, sendo o único a ter publicado artigo sobre o tema neste ano.

- Para o tópico ALGORITHM, o evento USENIX Security Symposium foi o que mais publicou com 33 artigos, seguido de MOBICOM com 26 e NDSS com 24. Até então o padrão é bem parecido com o de EXPERIMENT, mas nesse caso a única conferência que publicou artigo sobre o tema em 2022 foi EUROCRYPT.
- Para o tópico COMPUTER SCIENCE, o evento USENIX Security Symposium foi o que mais publicou com 56 artigos do tema, seguido do NDSS com 25 e para quebrar o padrão, o CIDR com 24. Além do quebrar o padrão do top 3, o CIDR também foi a conferência que mais publicou sobre o tema em 2022 com 6 artigos, seguido pelo SIGCOMM com 1 artigo, sendo os 2 únicos. Ao observar o histograma é possível perceber que apesar de não ter nenhuma publicação do tópico em 2022, a quantidade de artigos desse tipo no USENIX Security Symposium tem aumentado bastante até 2021, enquanto os outros eventos se mantêm mais constantes, até mesmo o CIDR, que diminuiu a quantidade de 2021 para 2022 (lembrando que o ano ainda está em andamento).
- Para o tópico SENSOR, o padrão dos dois primeiros se mantém quase intacto, com os três eventos no top 4, apenas com a adição da conferência SenSys em segundo lugar com 28 artigos, enquanto o USENIX Security Symposium tem 33, o MOBICOM tem 24 e o NDSS tem 15. Com relação a 2022 só os eventos USENIX Security Symposium e EUROCRYPT possuem artigos publicados, respectivamente com 2 e 1.
- Para o tópico PROTOTYPE, o primeiro lugar vai para o evento NDSS com 29 artigos, seguido do MOBICOM com 28 e o USENIX Security Symposium com 16, ficando em terceiro lugar. Mesmo assim, em 2022 somente o último teve publicações da área, com 2 artigos, inclusive se mantendo relativamente constante na análise histórica, enquanto MOBICOM apresentou uma certa decaída na quantidade de artigos publicados sobre o tema.
- Para o tópico MACHINE LEARNING, o evento que mais teve publicações foi o USENIX Security Symposium com 29, seguido do NDSS com 19 e o MOBICOM com 15. Já em 2022 é um dos temas que mais teve artigos publicados, com 28 no total, sendo que NDSS fica na frente com 8, USENIX Security Symposium em segundo com 7 e CIDR em terceiro com 6.

5. Conclusões

Como uma das atividades mais comuns na vida dos pesquisadores, submeter artigos a eventos e revistas pode ser cansativo se o mesmo não souber as melhores plataformas e as que vão valorizar as suas principais descobertas. Além de longas descrições e formulários complexos, muito se gasta na expectativa de ser aceito e poder ter uma publicação compartilhada com diversas outras pessoas. Infelizmente, devido a grande diversidade de conferências e eventos muito potencial é descartado no meio do caminho, e por isso, uma ferramenta que fosse capaz de recomendar às pessoas o local com maior chance de

aceitação diminuiria a fricção e busca, a fim de ajudar mais cientistas a terem seus estudos abertos com toda a comunidade.

Mediante o exposto, e considerando este trabalho como uma etapa inicial essencial para o objetivo final, é possível afirmar que ele, através das análises propostas, trouxe luz a forma como algumas conferências estão se portando a certa de temas centras, como protótipos, algoritmos e *machine learning*. Ademais, foi possível identificar padrões históricos que podem servir para futuramente uma previsão de quantidade de artigos publicados de um tema para cada evento, considerando como eles se portaram no passado. Além disso, este artigo pode ser útil para fomentar a discussão sobre quais temas deveriam ser mais publicados ou quais não estão recebendo atenção suficiente em prol de outros, elucidando buracos no conhecimento que por vezes podem estar obscuros.

No entanto, durante o desenvolvimento deste trabalho, alguns desafios foram encontrados. O primeiro foi a falta de um conjunto de dados único com todas as informações, sendo necessária a criação de um *script* em *Python* para, coletando os dados de diferentes fontes, fazê-lo. Além, este trabalho teve algumas limitações, como a quantidade de 100 requisições para a API do Semantic Scholar a cada 5 minutos e o fato do site do DL ACM bloquear o IP do computador após alguns *scrapings* por achar que seria um ataque *hacker*. Devido a estes fatos a pesquisa se limitou a 10 conferências, e para tentar driblar essas questões teve que usada uma função de delay no código e uma VPN externa, respectivamente. Outro ponto que cabe ressaltar é que cada evento na página do *dblp* tinha uma estrutura distinta, então alguns artigos ficaram com dados faltantes, pois o *scraping* procurava por certos padrões que não existiam.

Como trabalhos futuros, pretende-se ampliar essas análises a uma maior quantidade de conferências e a um intervalo maior de tempo para ter um contexto histórico mais completo. Outro ponto interessante é o de expandir as análises para questões relacionadas ao países dos artigos que os eventos mais publicam e se existe algum relação entre os artigos publicados por uma mesma conferência. As análises propostas neste trabalho servem como munção inicial para diversas outras abordagens, além do próprio sistema de recomendação do qual ele é base. Finalmente, é muito importante encontrar uma forma de trazer todos os dados dos artigos automaticamente, sendo pelo uso de outra *API* ou algum site que contenha as informações completas, que atualmente se encontram indisponíveis para alguns casos.

Referências

- GOLDBERG, D. e. a. (1992). Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35:61–70.
- Shao, B., L. X. . B. G. (2021). A survey of research hotspots and frontier trends of recommendation systems from the perspective of knowledge graph. *Expert Systems with Applications*, 165.