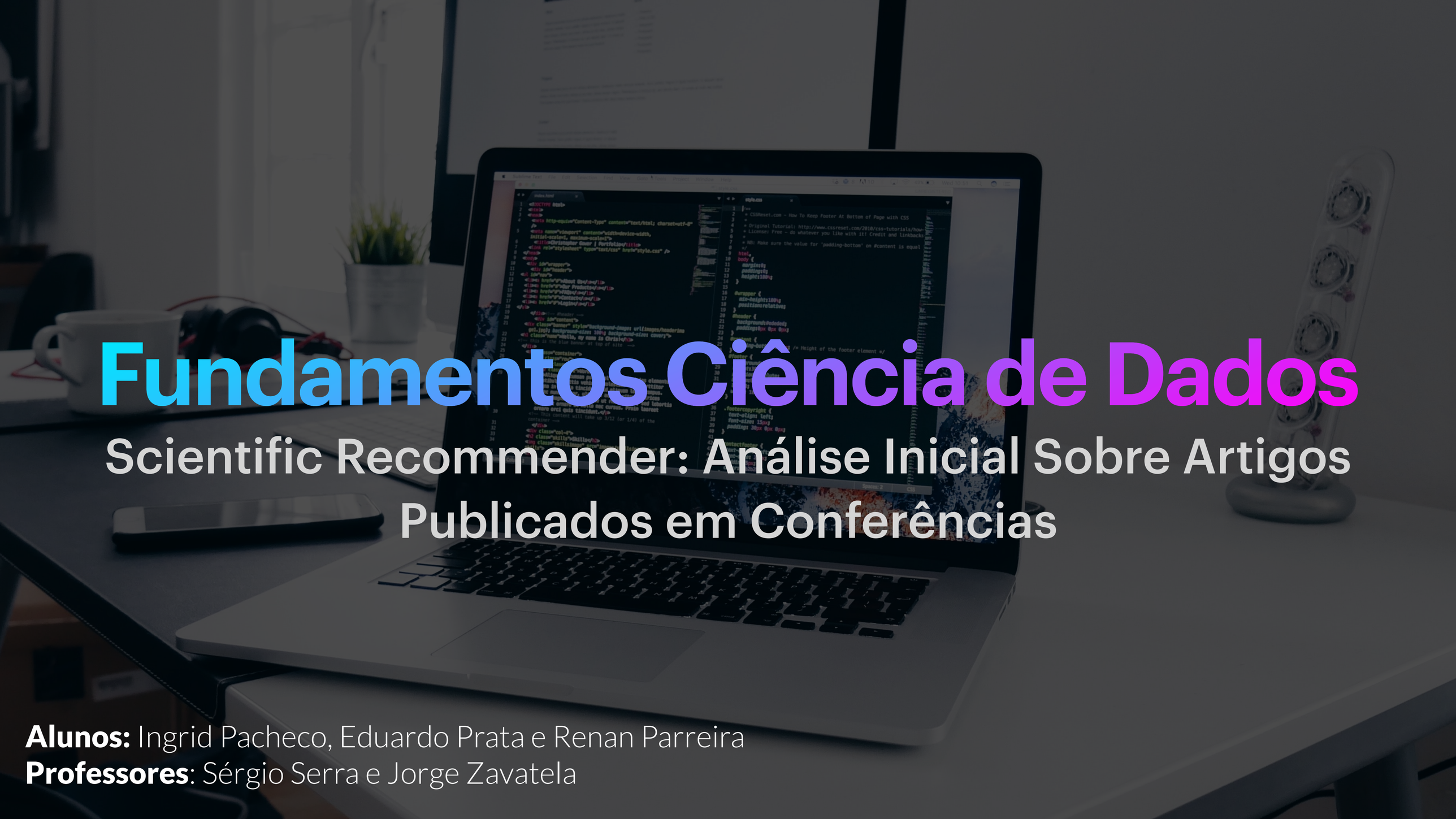


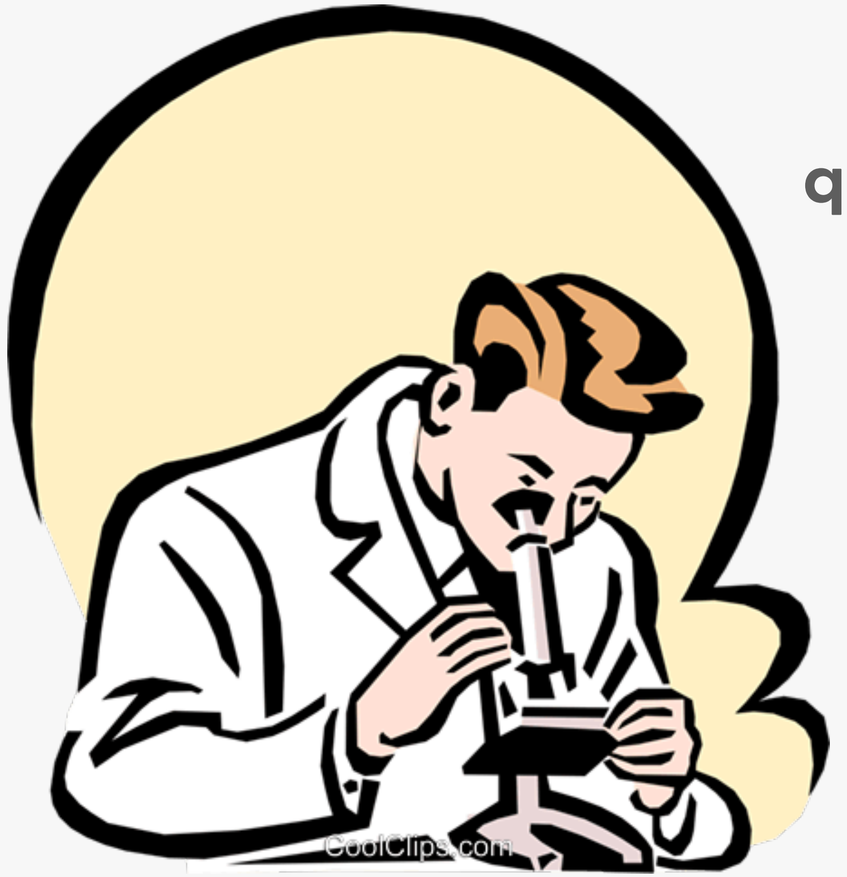
Fundamentos Ciência de Dados

Scientific Recommender: Análise Inicial Sobre Artigos Publicados em Conferências

Alunos: Ingrid Pacheco, Eduardo Prata e Renan Parreira
Professores: Sérgio Serra e Jorge Zavatela

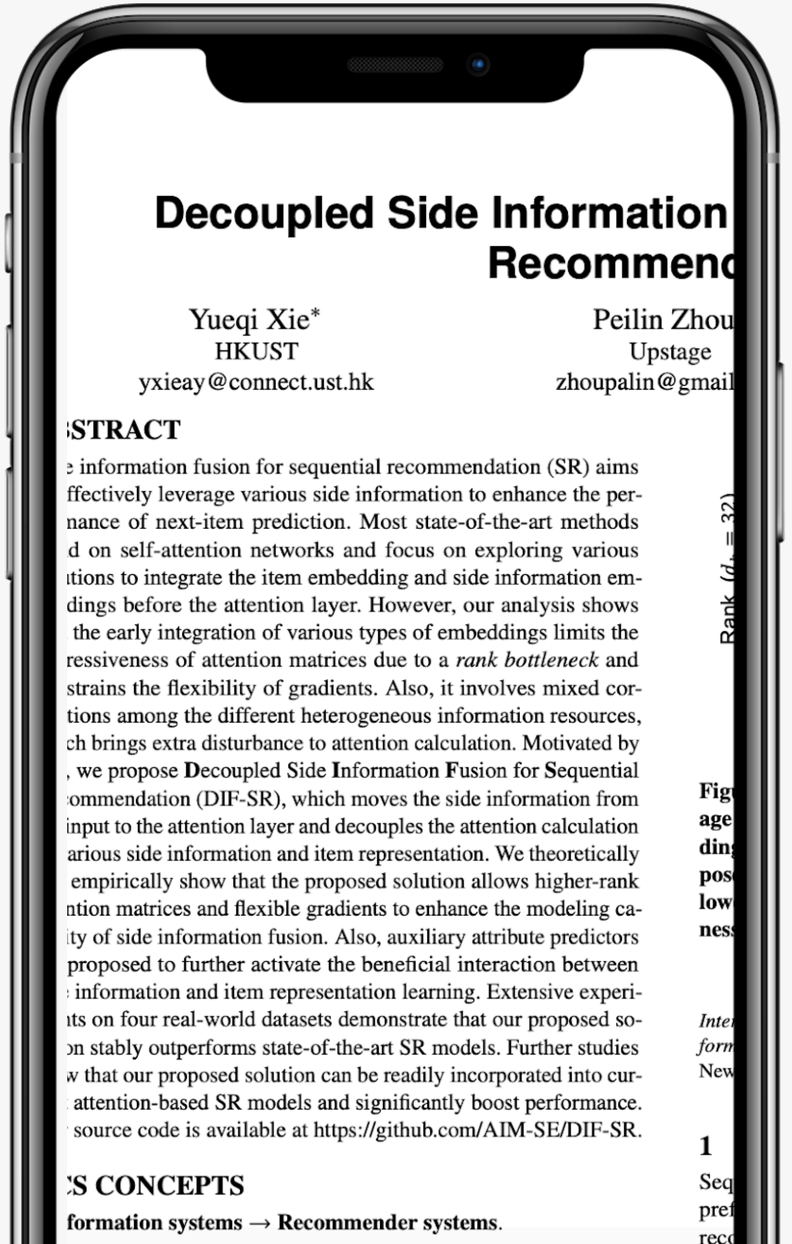
PROBLEMA INICIAL

Problema Inicial



Maior quantidade de pessoas

Problema Inicial



Decoupled Side Information
Recommend

Yueqi Xie*
HKUST
yxieay@connect.ust.hk

Peilin Zhou
Upstage
zhoupalin@gmail

STRACT

information fusion for sequential recommendation (SR) aims
ffectively leverage various side information to enhance the per-
nance of next-item prediction. Most state-of-the-art methods
d on self-attention networks and focus on exploring various
itions to integrate the item embedding and side information em-
dings before the attention layer. However, our analysis shows
the early integration of various types of embeddings limits the
ressiveness of attention matrices due to a *rank bottleneck* and
strains the flexibility of gradients. Also, it involves mixed cor-
tions among the different heterogeneous information resources,
ch brings extra disturbance to attention calculation. Motivated by
, we propose **D**ecoupled Side **I**nformation **F**usion for **S**equential
ommendation (DIF-SR), which moves the side information from
input to the attention layer and decouples the attention calculation
arious side information and item representation. We theoretically
empirically show that the proposed solution allows higher-rank
ntion matrices and flexible gradients to enhance the modeling ca-
ty of side information fusion. Also, auxiliary attribute predictors
proposed to further activate the beneficial interaction between
information and item representation learning. Extensive experi-
ts on four real-world datasets demonstrate that our proposed so-
n stably outperforms state-of-the-art SR models. Further studies
w that our proposed solution can be readily incorporated into cur-
attention-based SR models and significantly boost performance.
source code is available at https://github.com/AIM-SE/DIF-SR.

S CONCEPTS

formation systems → Recommender systems.

Rank ($d_$ = 32)

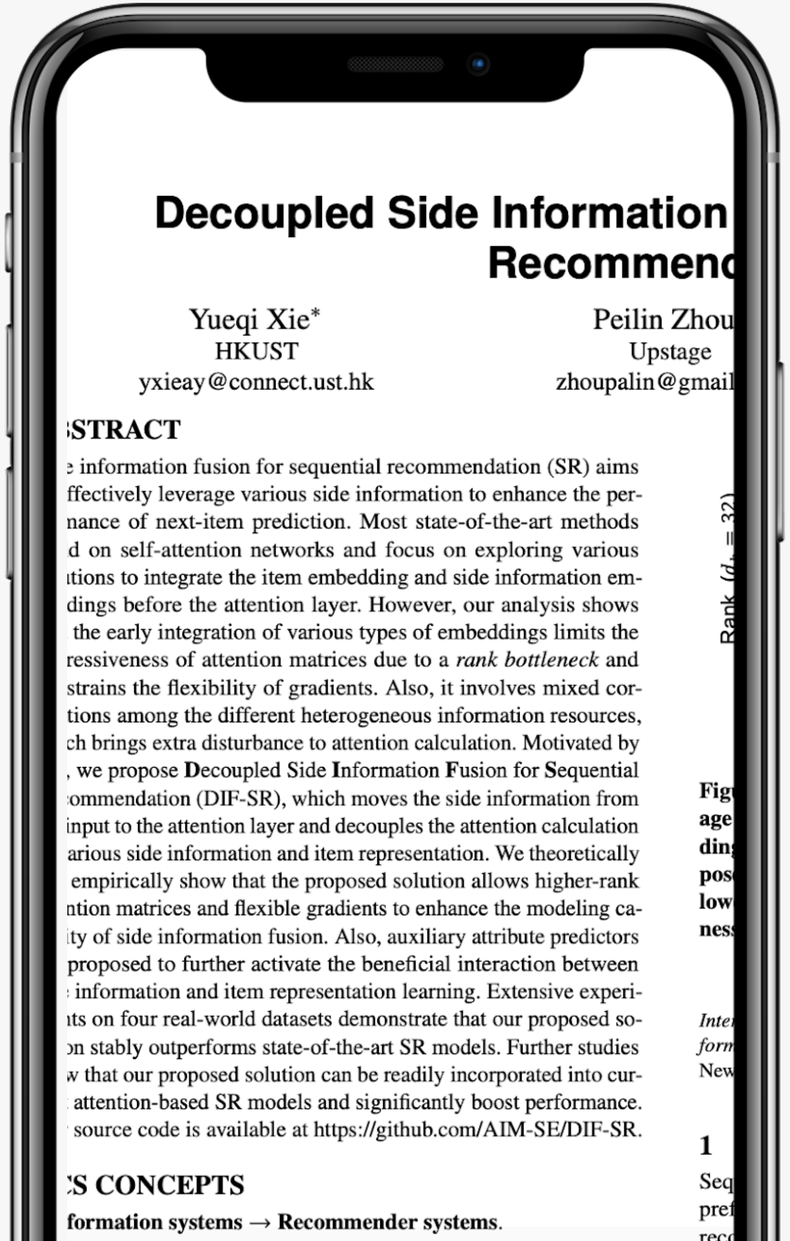
Fig
age
din
pos
low
nes

Inte
form
New

1

Seq
pref
reco

Problema Inicial



Decoupled Side Information
Recommenc

Yueqi Xie*
HKUST
yxieay@connect.ust.hk

Peilin Zhou
Upstage
zhoupalin@gmail

STRACT

e information fusion for sequential recommendation (SR) aims
ffectively leverage various side information to enhance the per-
nance of next-item prediction. Most state-of-the-art methods
d on self-attention networks and focus on exploring various
tions to integrate the item embedding and side information em-
dings before the attention layer. However, our analysis shows
the early integration of various types of embeddings limits the
ressiveness of attention matrices due to a *rank bottleneck* and
strains the flexibility of gradients. Also, it involves mixed cor-
tions among the different heterogeneous information resources,
ch brings extra disturbance to attention calculation. Motivated by
, we propose **D**ecoupled Side **I**nformation **F**usion for **S**equential
ommendation (DIF-SR), which moves the side information from
input to the attention layer and decouples the attention calculation
arious side information and item representation. We theoretically
empirically show that the proposed solution allows higher-rank
ntion matrices and flexible gradients to enhance the modeling ca-
ty of side information fusion. Also, auxiliary attribute predictors
proposed to further activate the beneficial interaction between
information and item representation learning. Extensive experi-
ts on four real-world datasets demonstrate that our proposed so-
n stably outperforms state-of-the-art SR models. Further studies
w that our proposed solution can be readily incorporated into cur-
attention-based SR models and significantly boost performance.
source code is available at https://github.com/AIM-SE/DIF-SR.

S CONCEPTS

formation systems → Recommender systems.

Rank ($d_k = 32$)

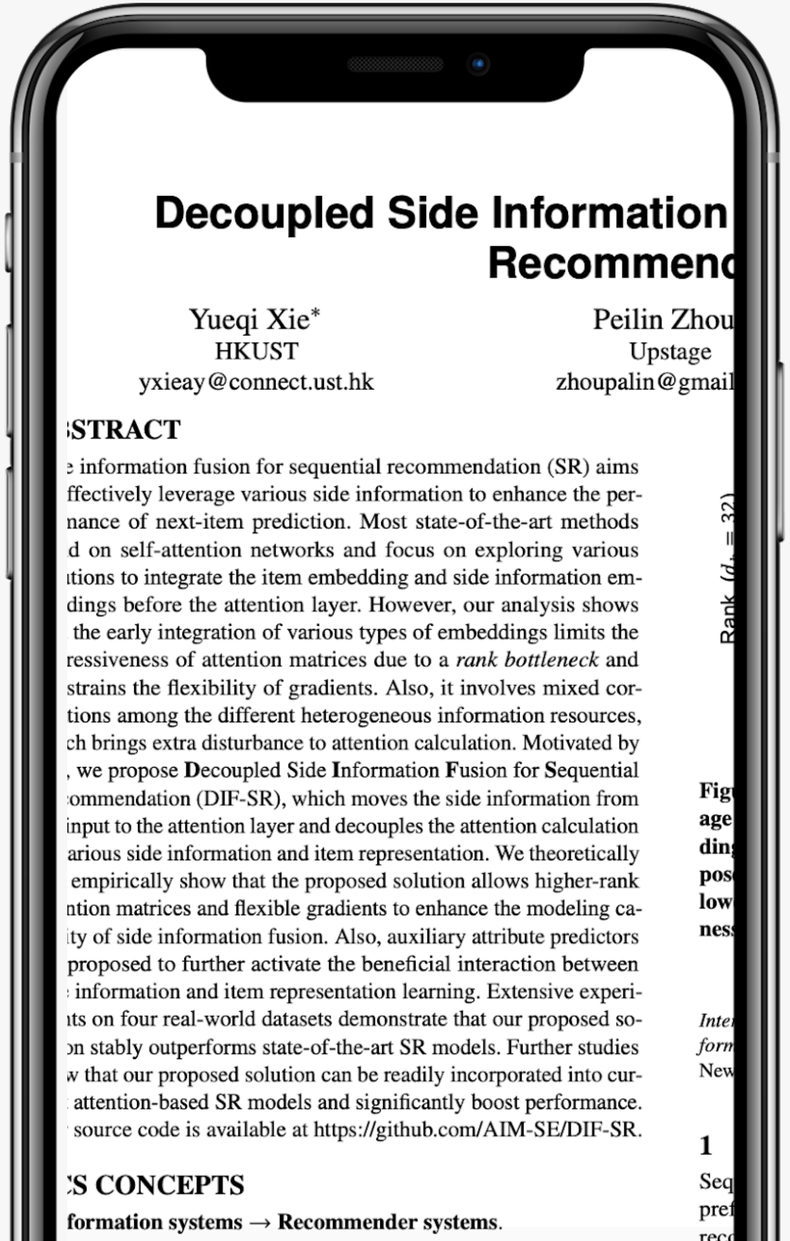
Fig
age
din
pos
low
ness

Inte
form
New

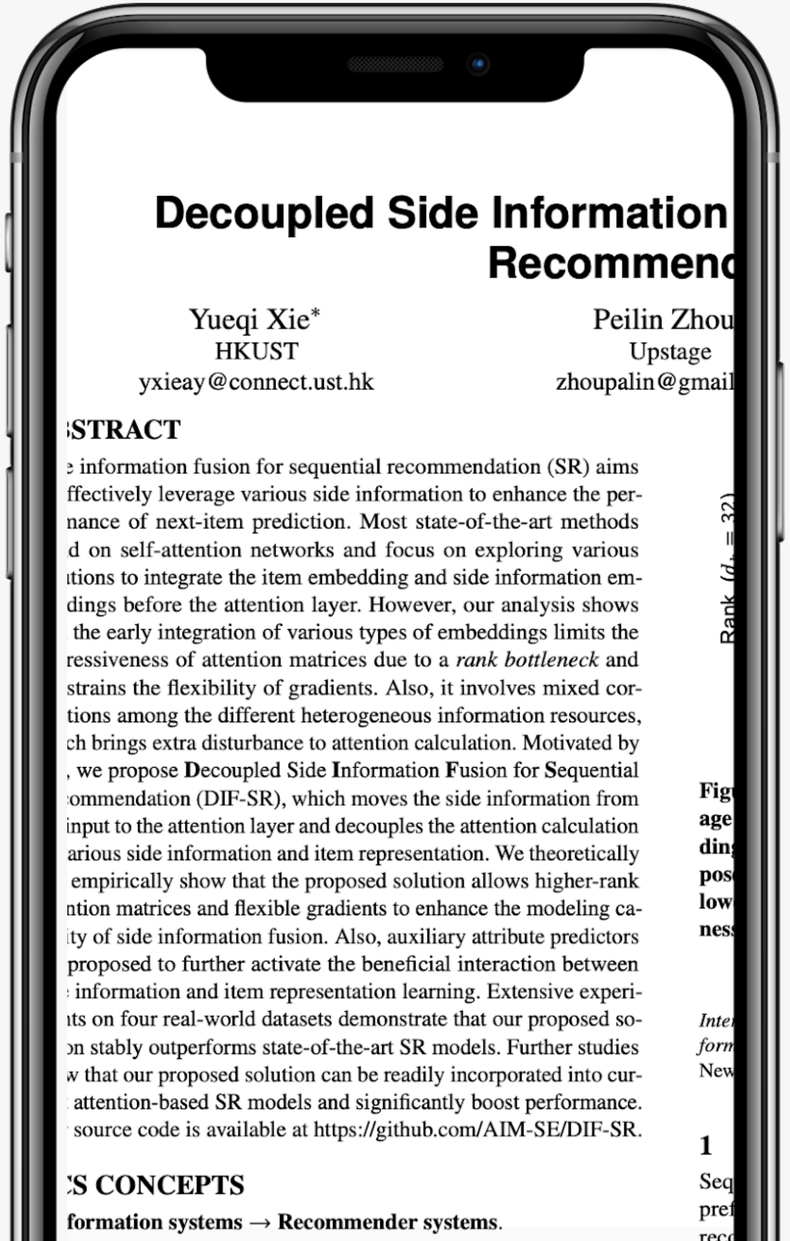
1

Seq
pref
reco

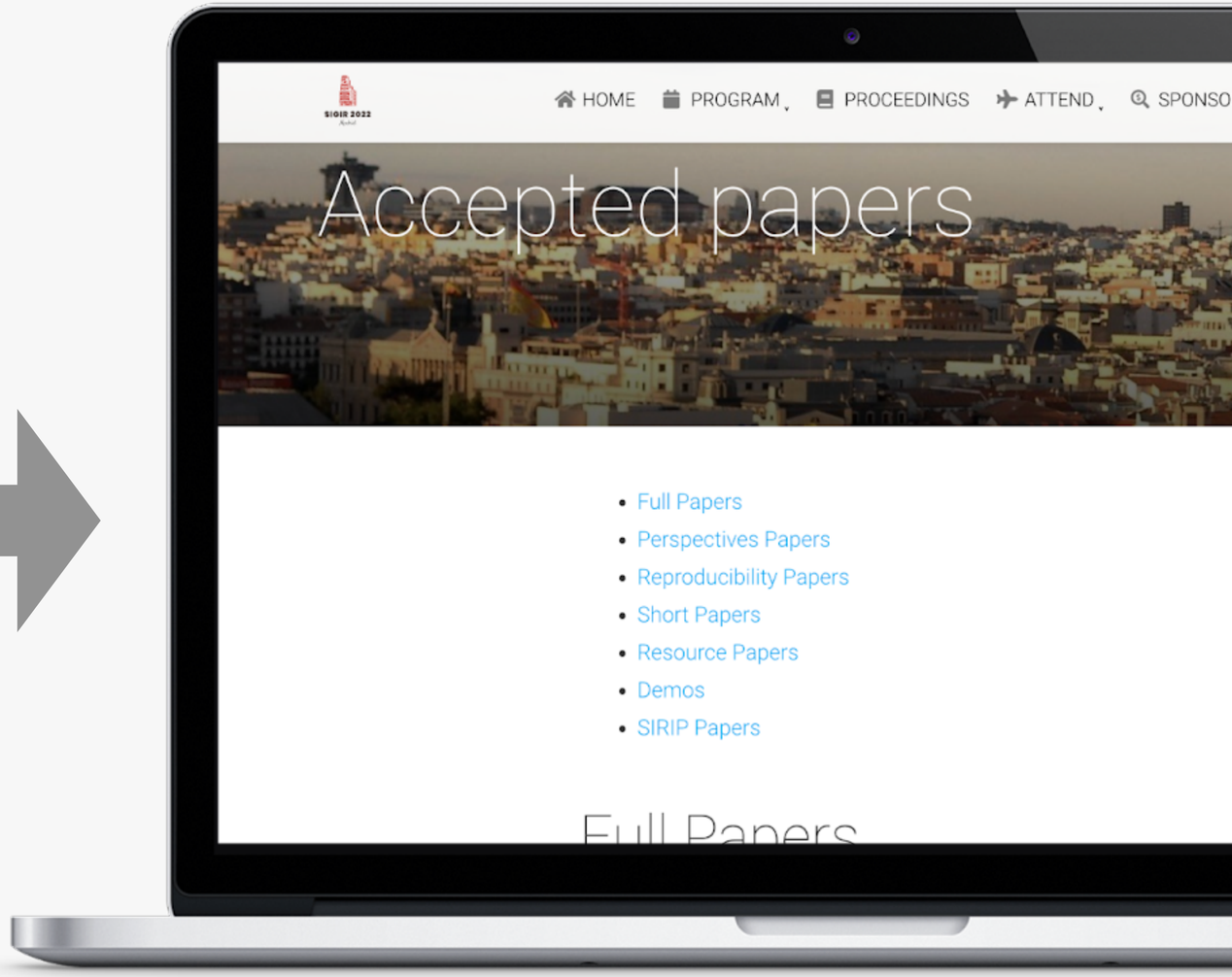
Problema Inicial



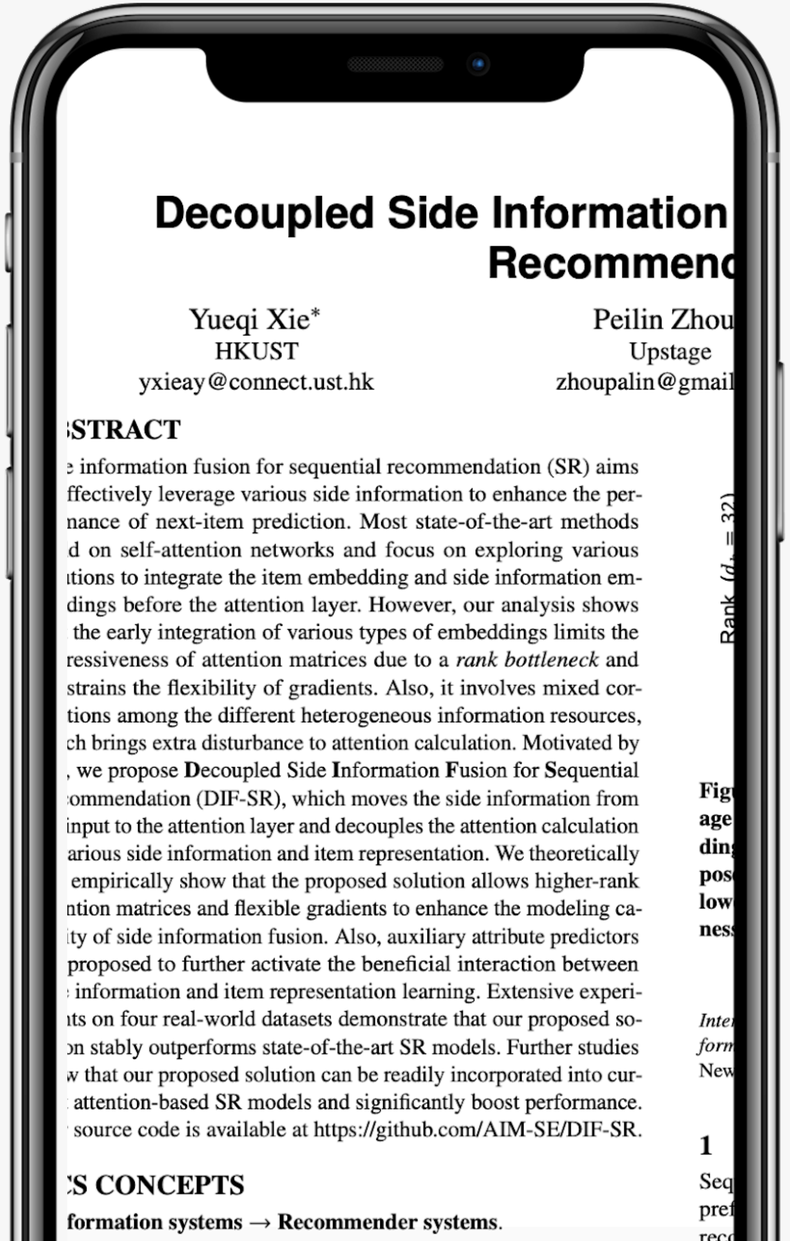
Problema Inicial



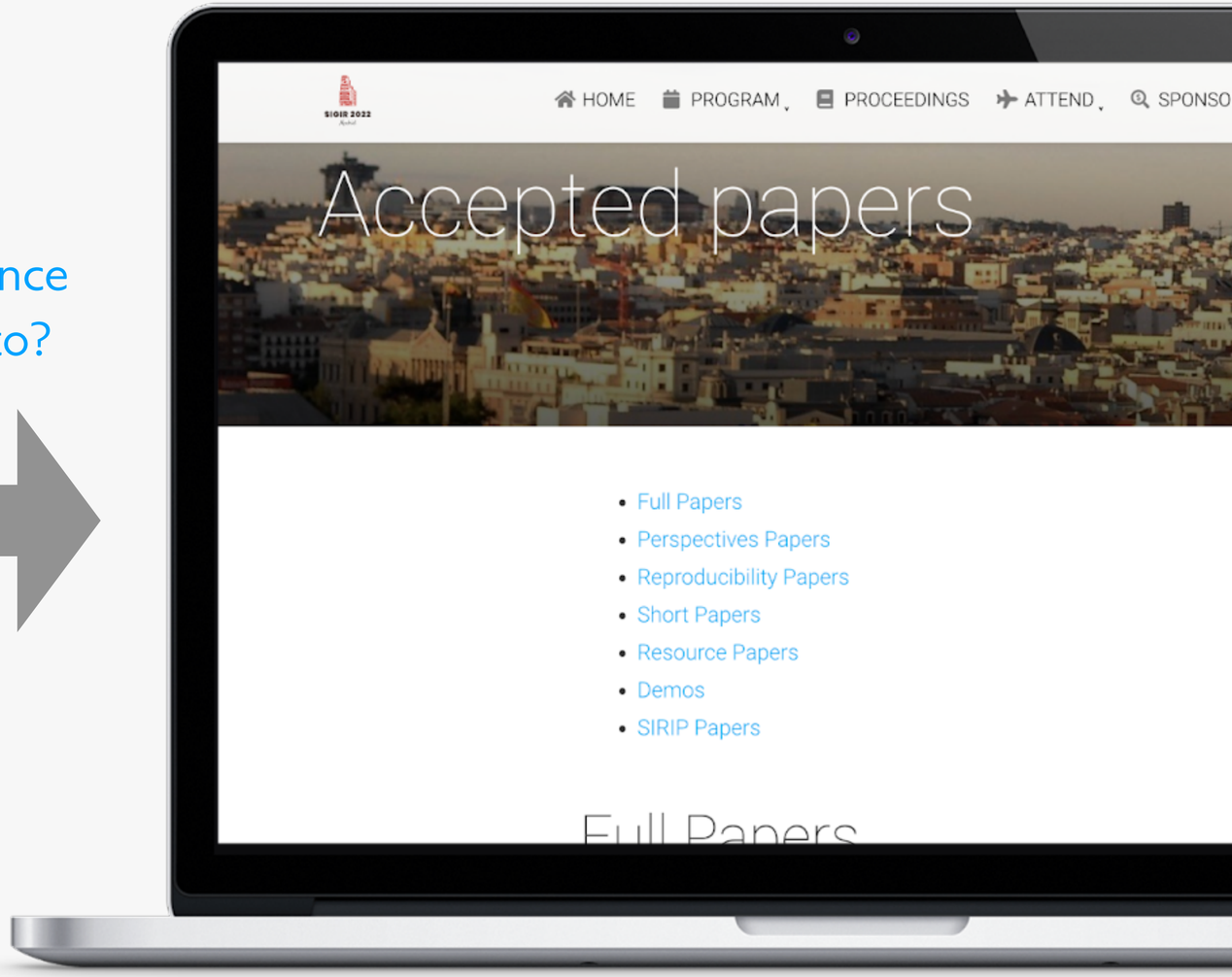
Ser aceito



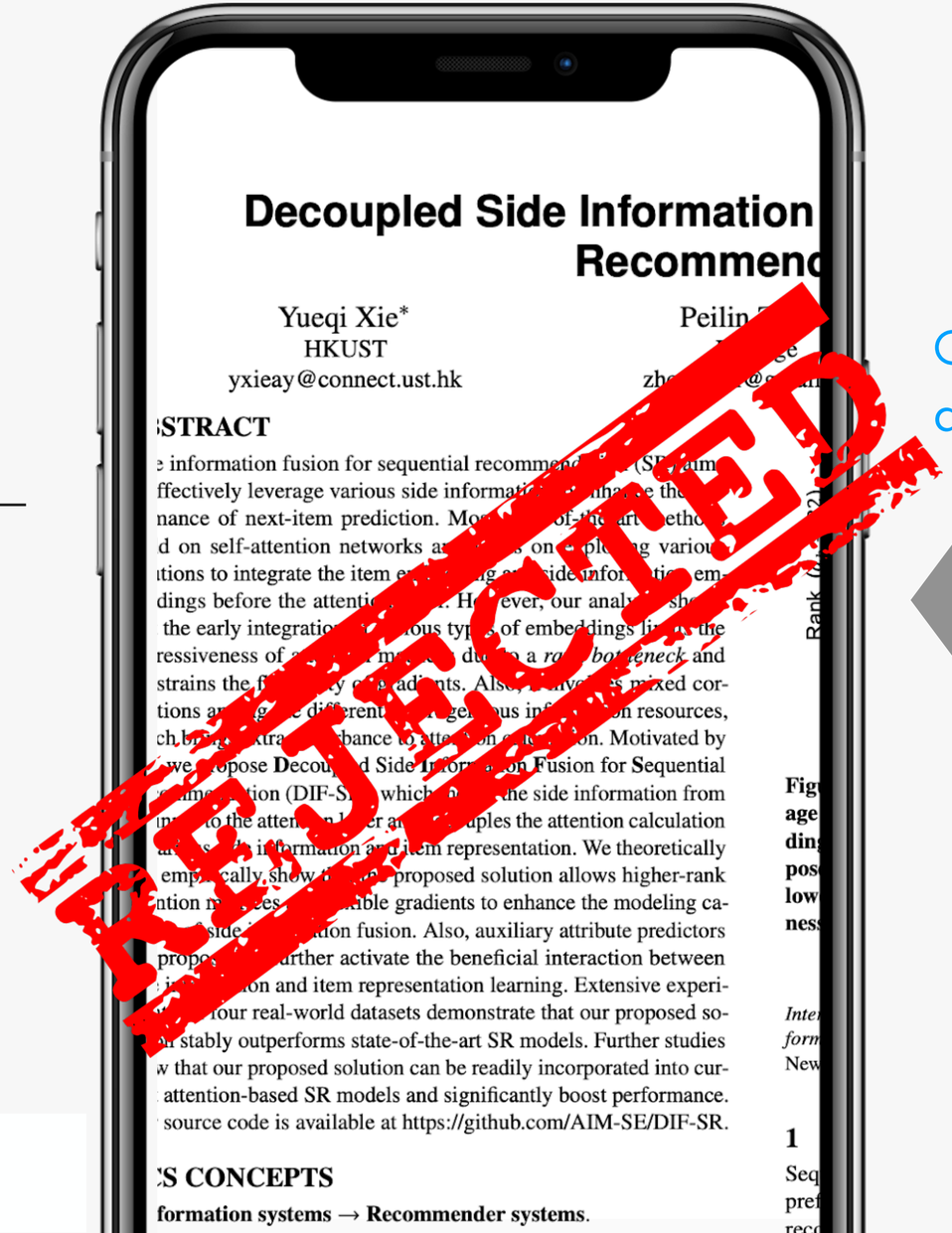
Problema Inicial



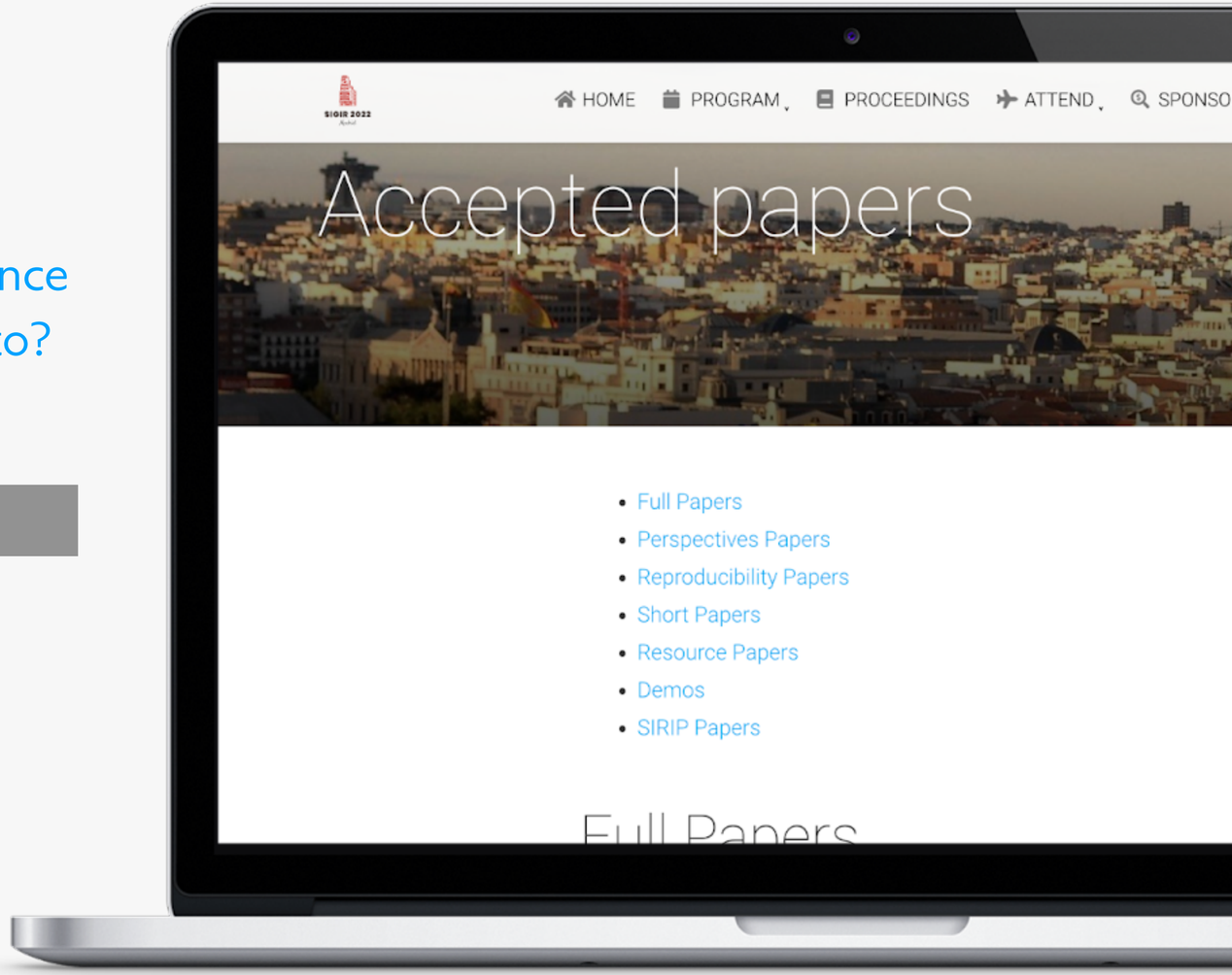
Qual a chance de ser aceito?



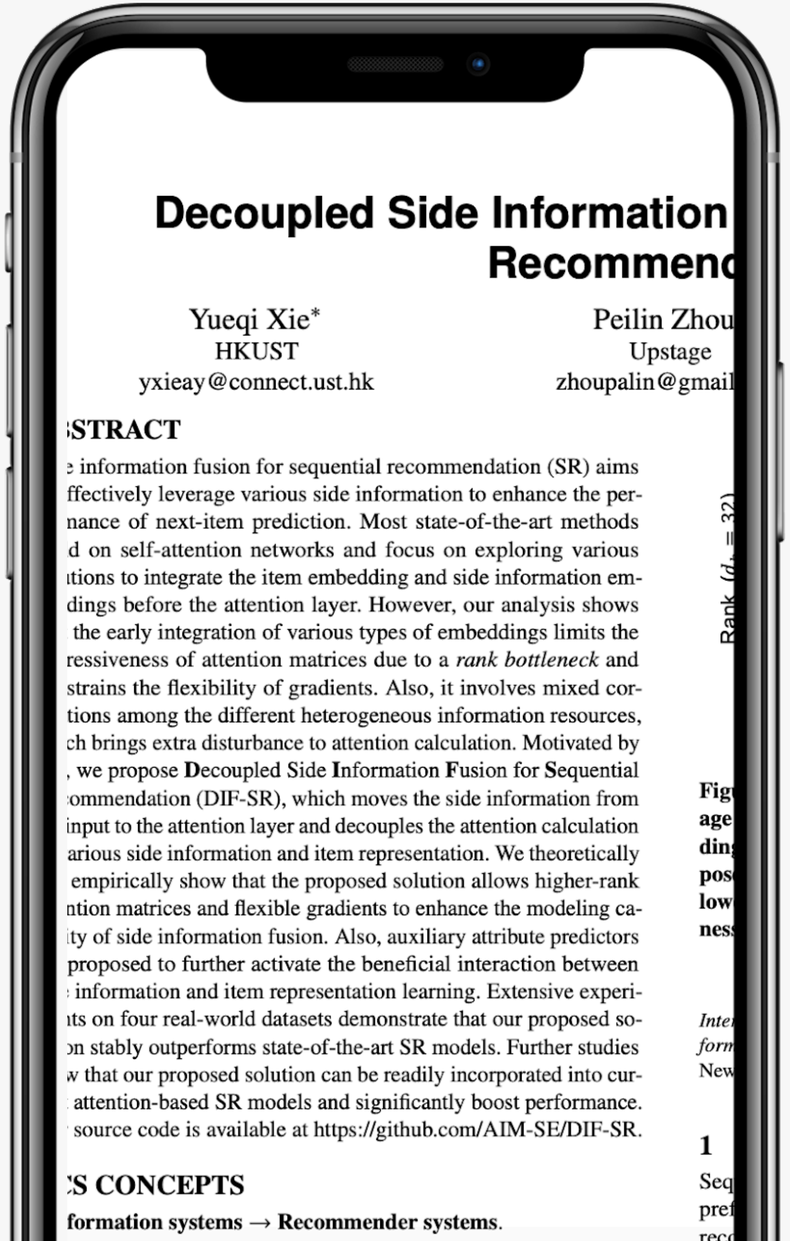
Problema Inicial



Qual a chance de ser aceito?



Problema Inicial

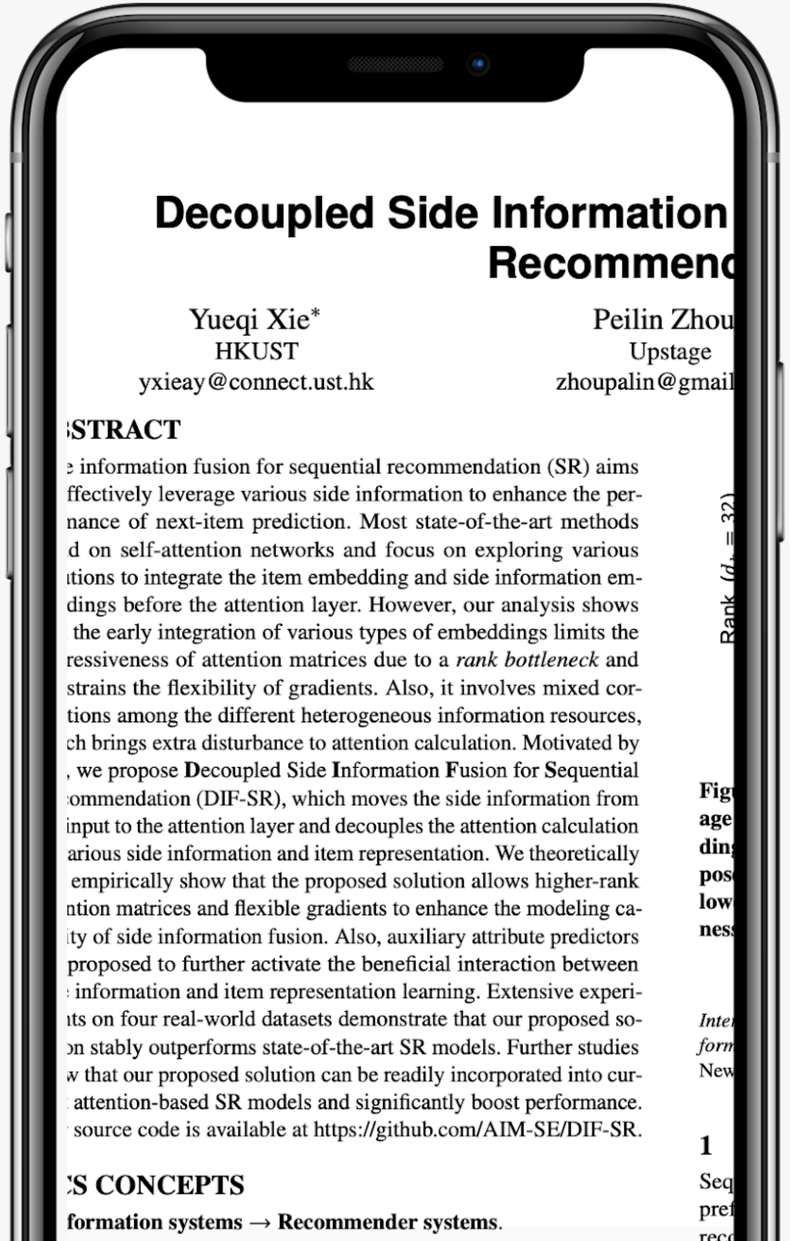


41%

87%

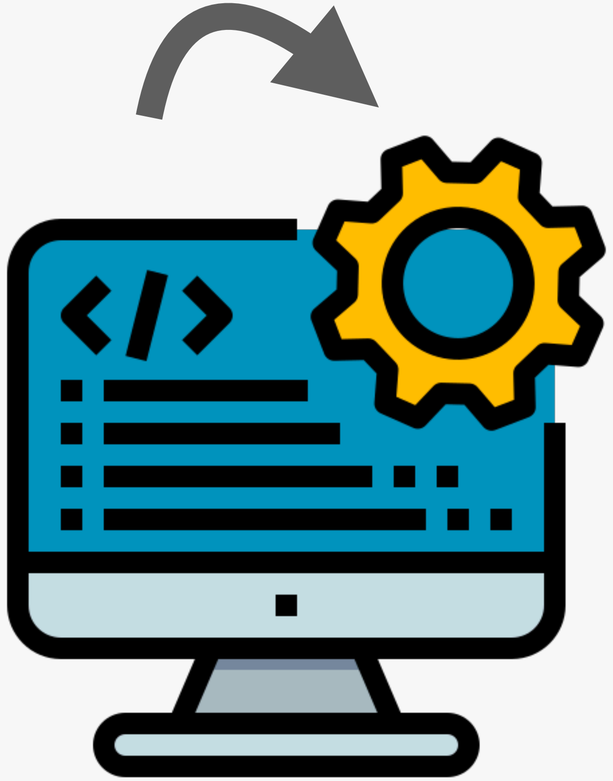
63%

Problema Inicial



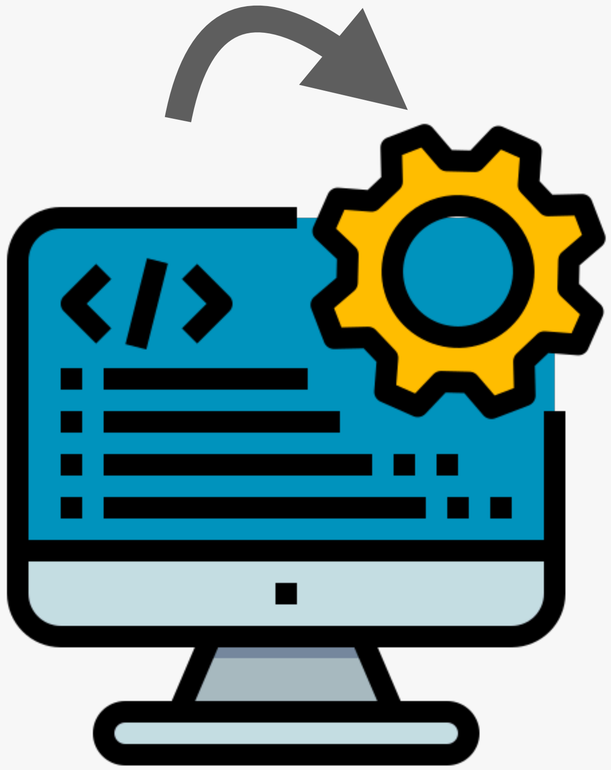
Recebe dados
do artigo escrito

Faz o processamento utilizando
dados de artigos já publicados
em conferências



Problema Inicial

Faz o processamento utilizando dados de artigos já publicados em conferências



Entrega a lista das melhores conferências para publicação do artigo

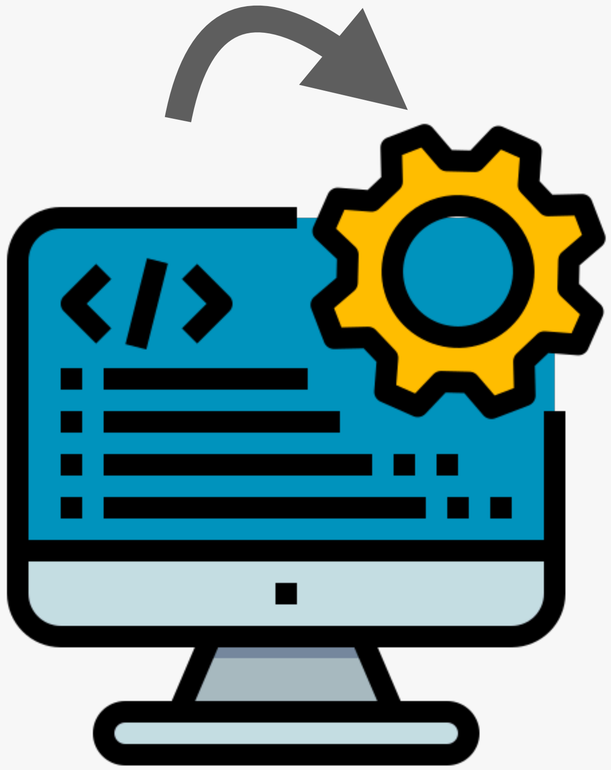
41%

87%

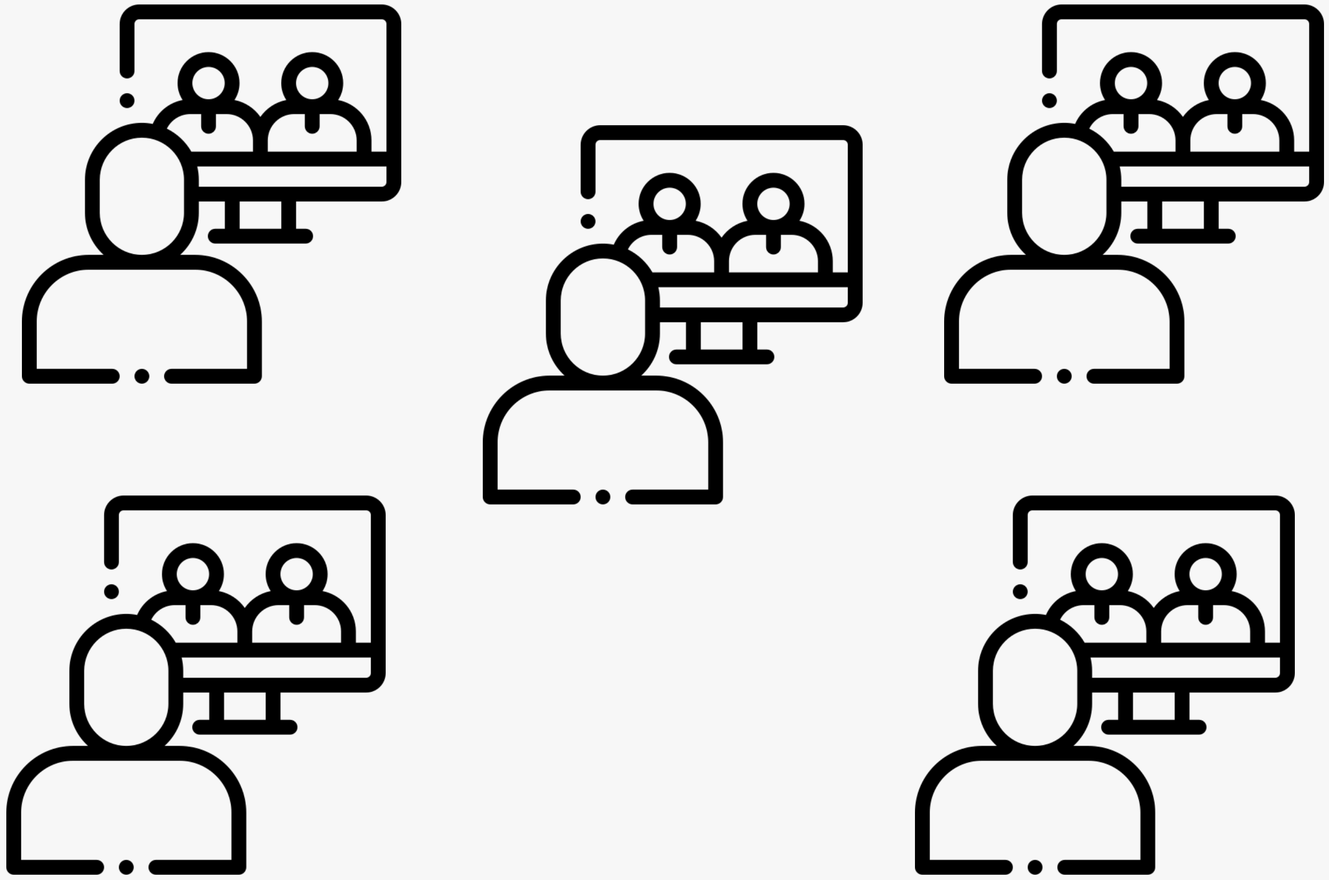
63%

Problema Inicial

Faz o processamento utilizando dados de artigos já publicados em conferências



Entender como estão as conferências em relação aos temas publicados nelas e um histórico

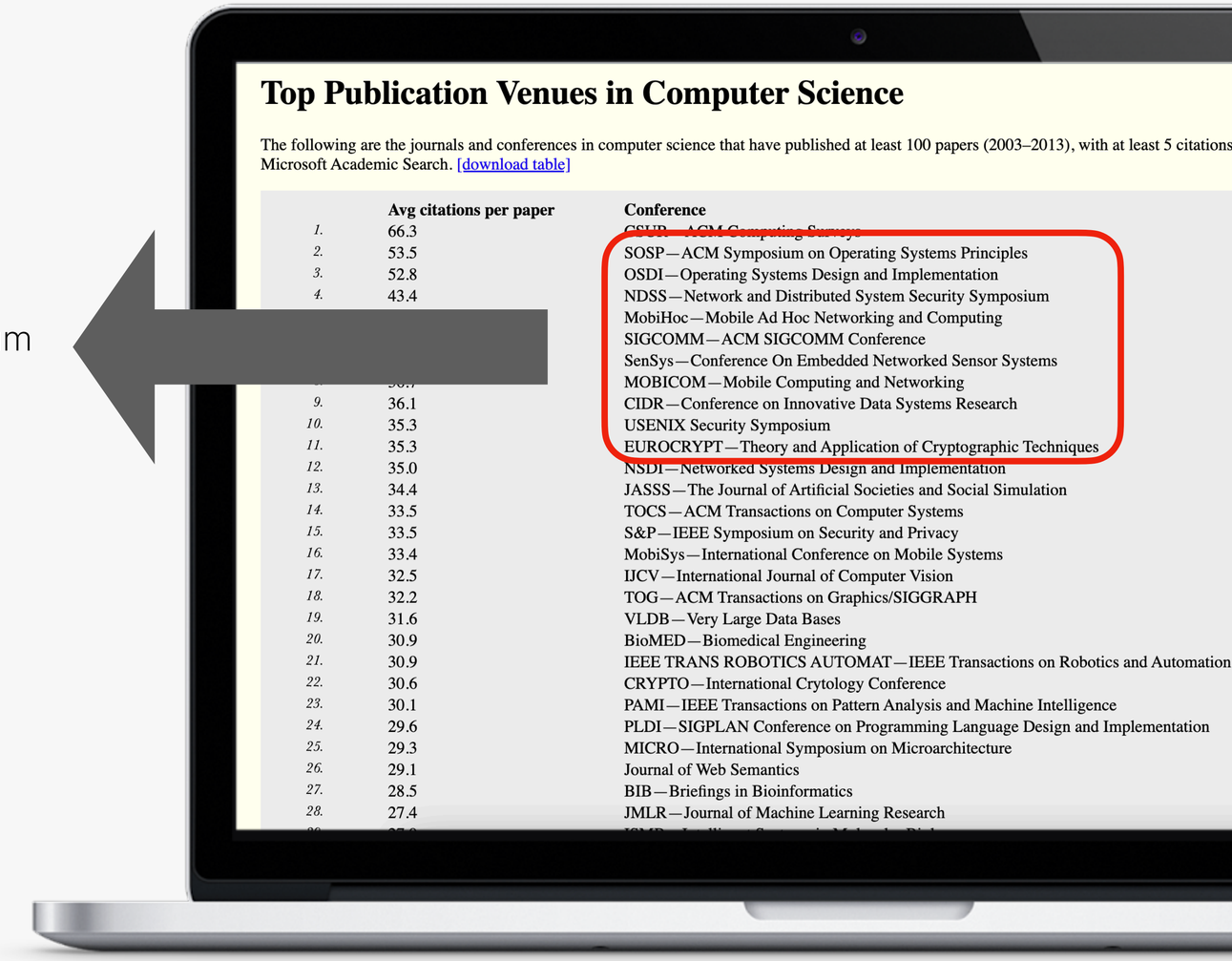


Geração do Dataset

Geração do dataset

10 conferências:

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



Geração do dataset

10 conferências:

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

2018 a 2022

Geração do dataset

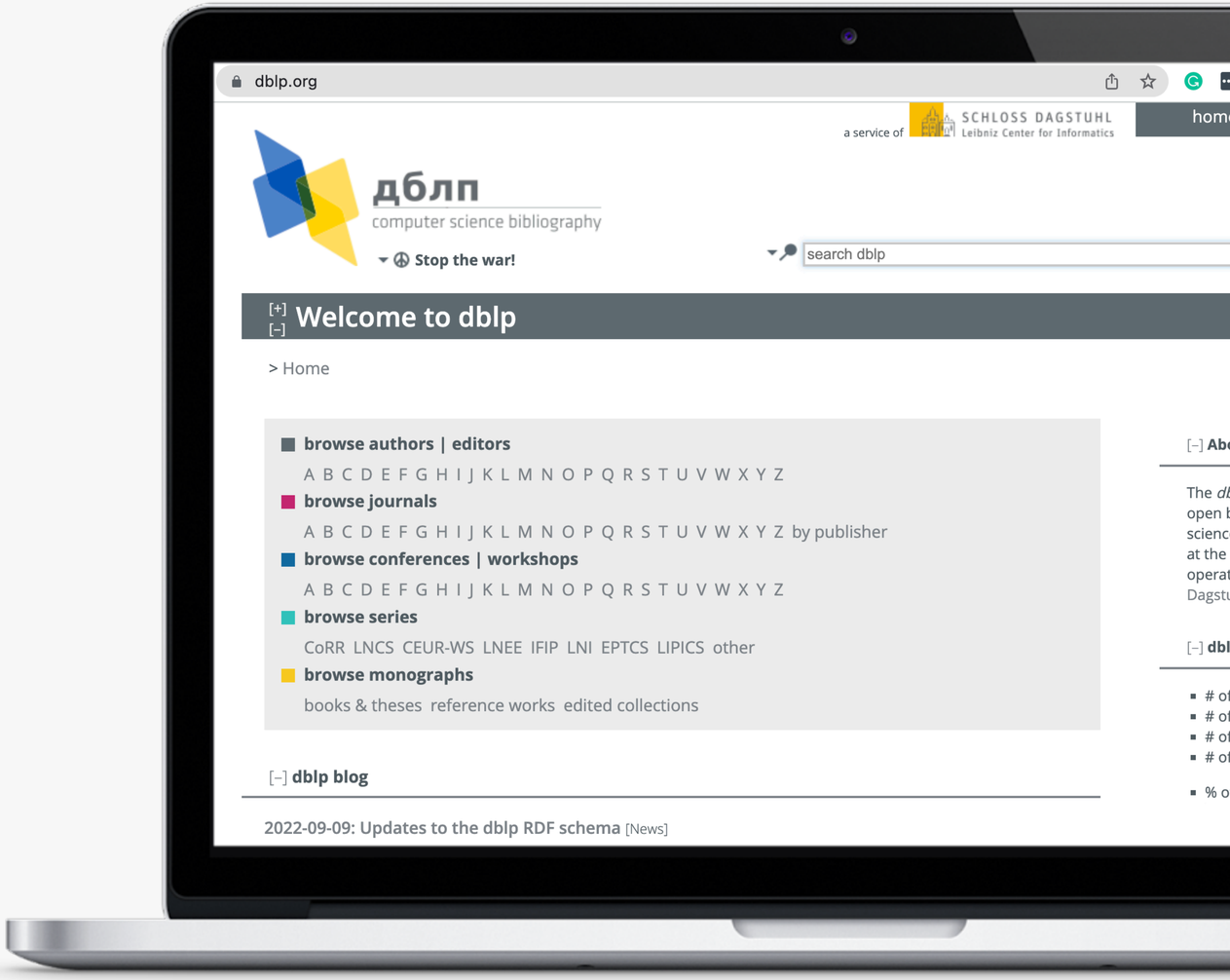
10 conferências:

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

Geração do dataset

Artigos são extraídos do site da **Dblp**

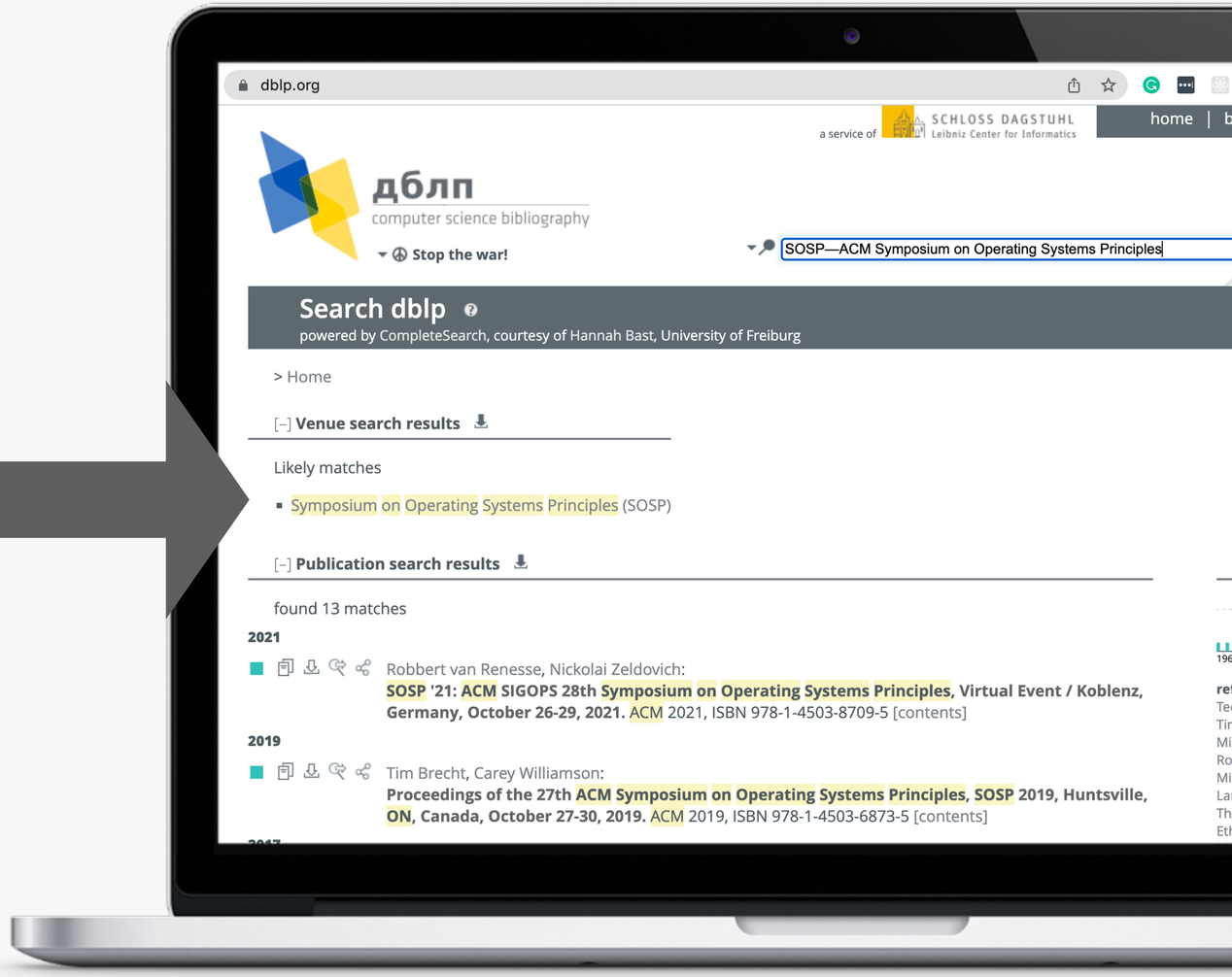
publications: 6,303,209
authors: 3,085,450
conferences: 5,978
journals: 1,815



Geração do dataset

Para cada conferência, ela é buscada na rota de busca do site

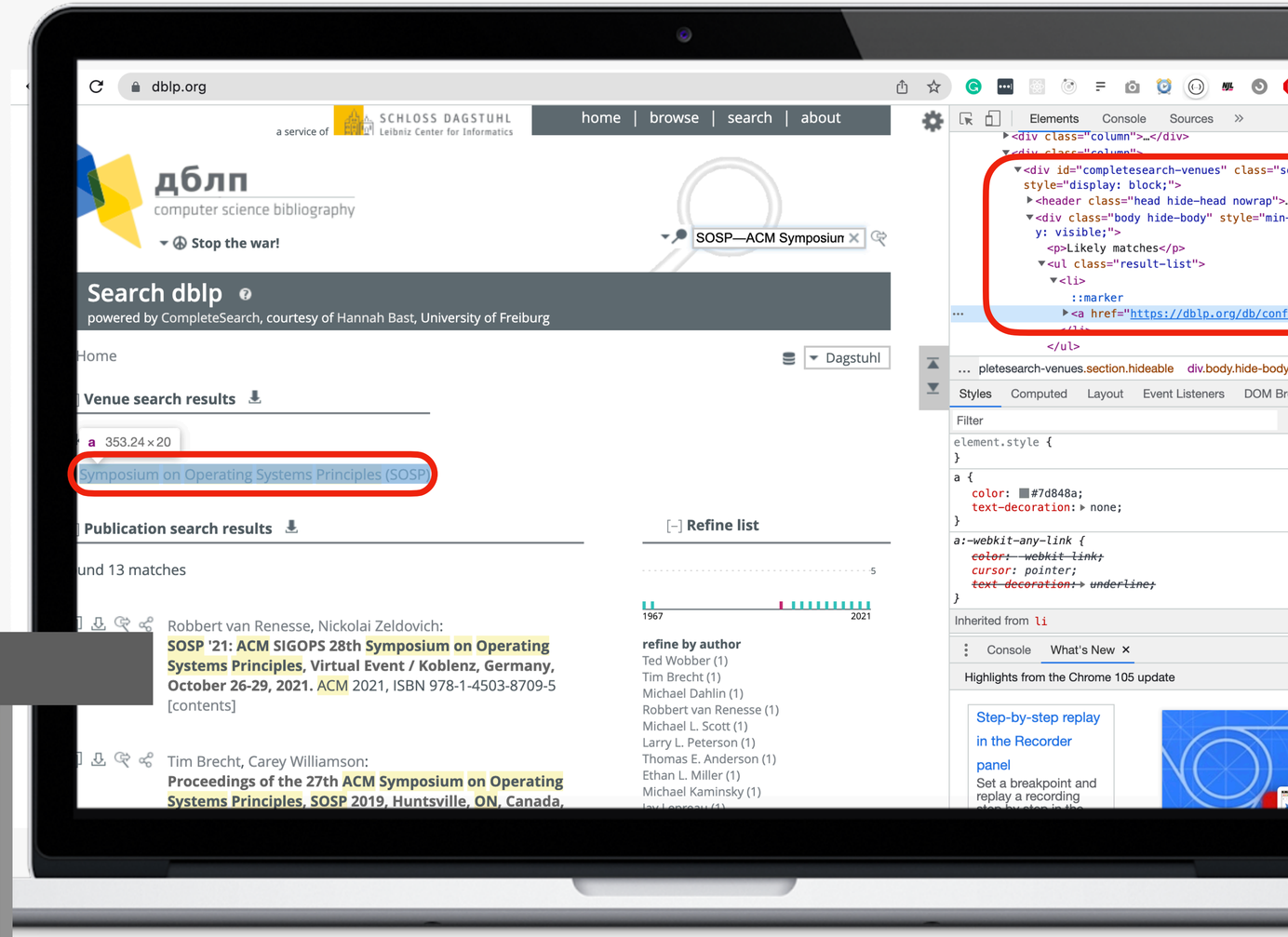
(**https://dblp.org/search?q=**)



Geração do dataset

Utiliza Web Scraper para entrar no link de listagem da conferência

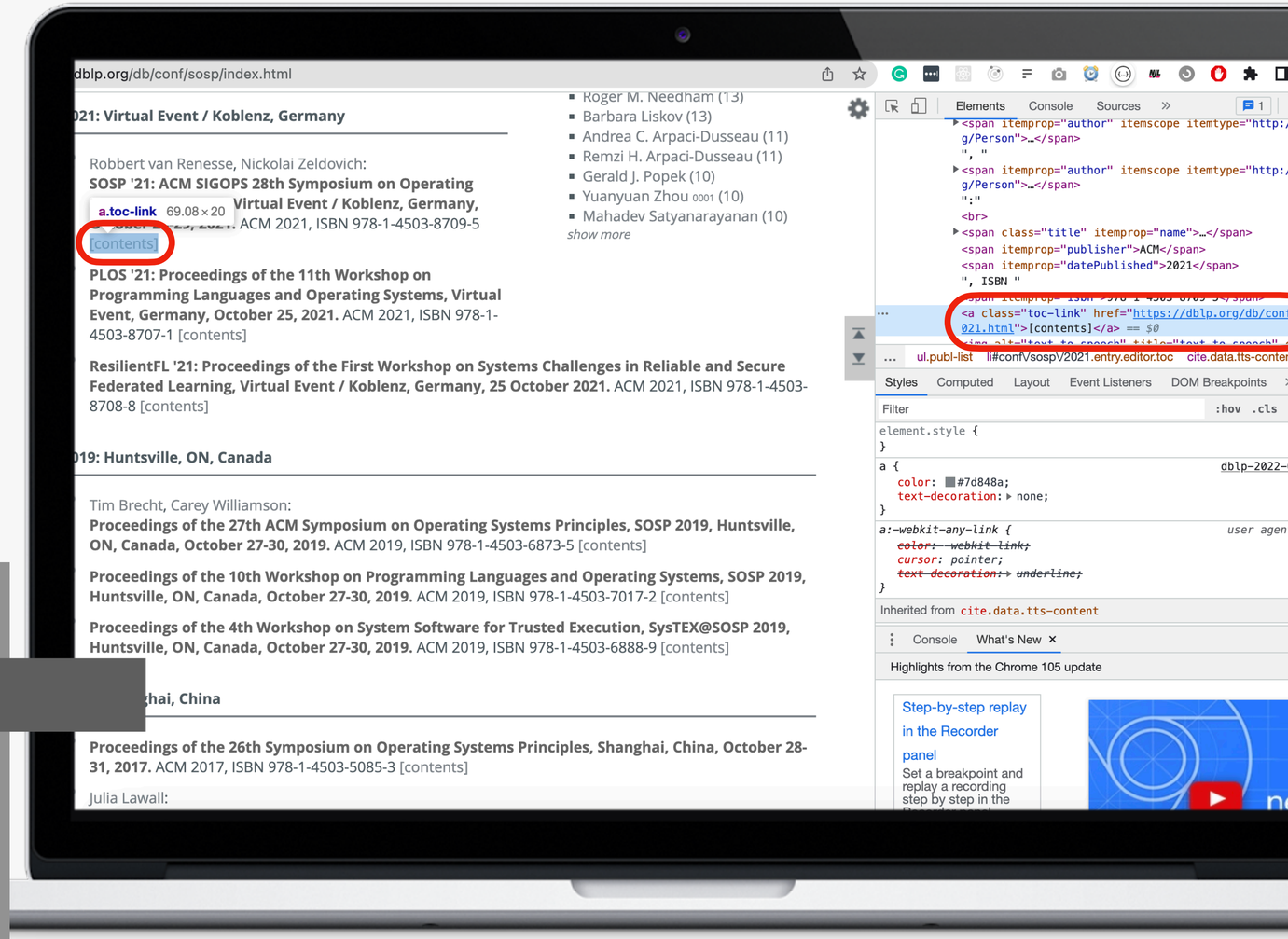
```
# Formatando o retorno
sopa = BeautifulSoup(resposta.text, 'html.parser')

infos = sopa.find('div', {'id':'completesearch-venues'})

venue = infos.find('ul', {'class':'result-list'}).find('a')

print('Href: ' + venue.get('href'))
venue_href = venue.get('href')
resposta = requests.get(venue_href, headers = cabecalho)
```



Geração do dataset

A partir da listagem de edições, entra no link das que são entre 2018 e 2022

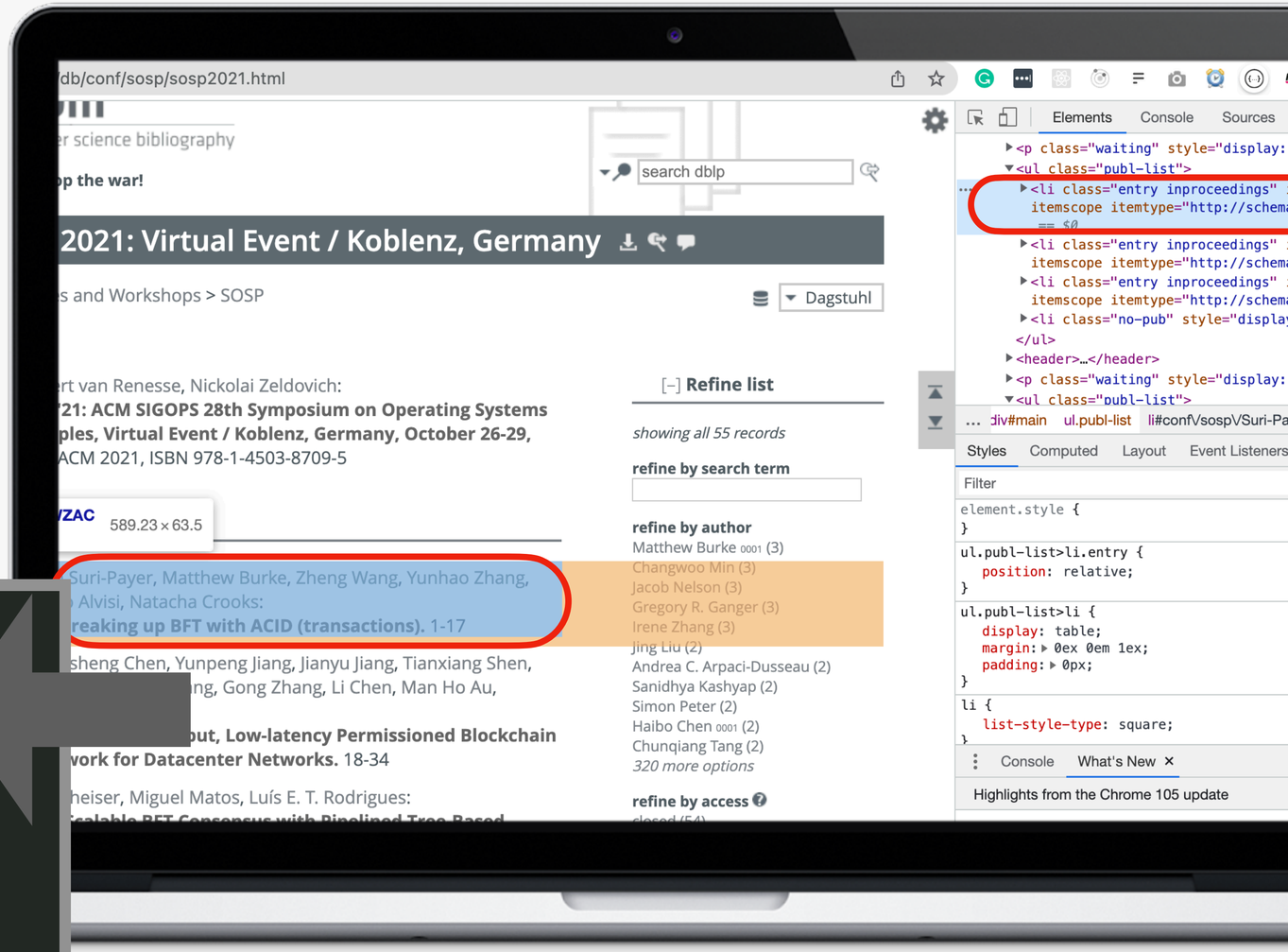
```
for conf_edition in sopa.find_all('ul', {'class':'publ-list'}):
  conf_title = conf_edition.find('span',{'itemprop':'name'}).get_text().lower()

  # Desconsidera os workshops
  if 'workshop' in conf_title:
    continue

  published_year = conf_edition.find('span',{'itemprop':'datePublished'}).get_text()
  print('Ano: ' + published_year)

  # So pega os artigos ate 2018
  if int(published_year) < 2018:
    break
  content = conf_edition.find('a',{'class':'toc-link'}).get('href')
  print(content)
  resposta = requests.get(content, headers = cabecalho)
```



Geração do dataset

Entra em cada artigo, pegando os dados dele:
título, ano, gênero (CC), doi, conferência

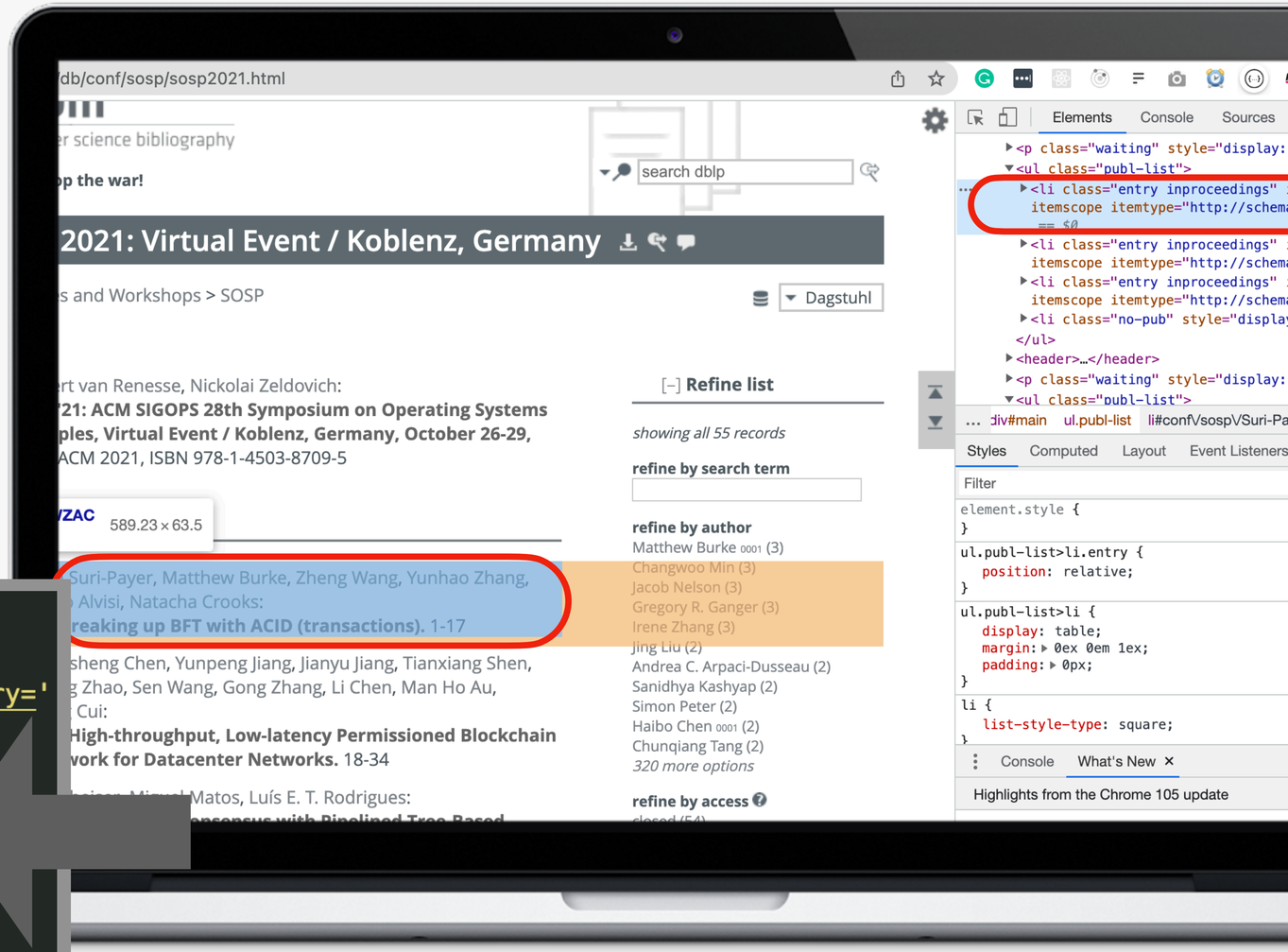
```
for article in sopa.find_all('li', {'class':'entry inproceedings'}):
  header = article.find_all_previous('h2')[0].get('id').lower()

  # Desconsidera os keynotes
  if 'keynote' in header:
    continue
  title = article.find('span', {'class':'title'}).get_text()
  print(title)
  year = article.find('meta', {'itemprop':'datePublished'})['content']
  print(year)
  genre = article.find('meta', {'property':'genre'})['content']
  print(genre)
  doi_url = article.find('li', {'class': 'drop-down'}).find('a').get('href')
```



Geração do dataset

Quando não tem doi, faz uma busca na api do **Semantic Scholar**

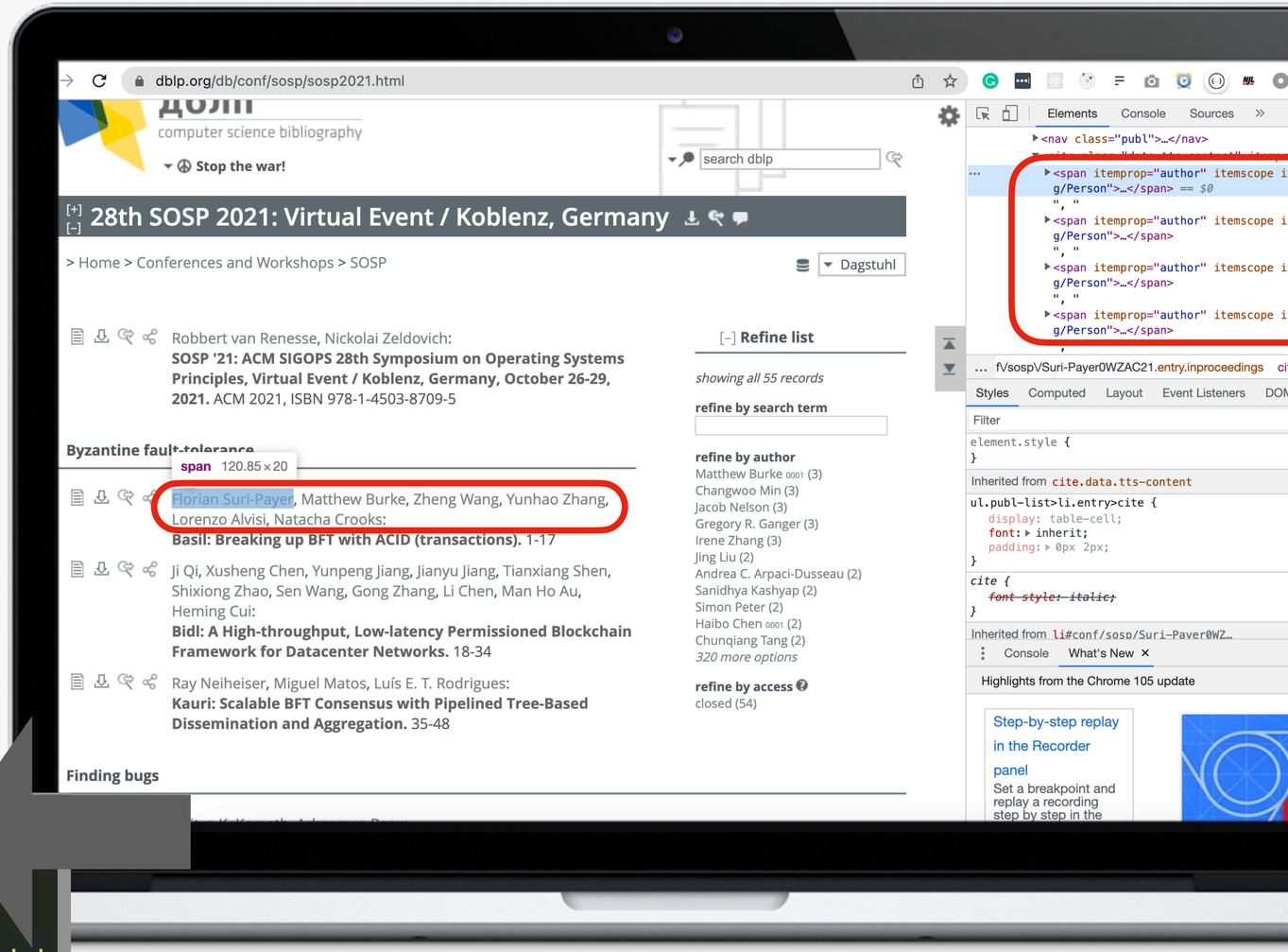
```python
def findArticle(title,cabecalho):
  filtro = title.replace(' ','+')
  link = 'https://api.semanticscholar.org/graph/v1/paper/search?query='
  print('Não tem DOI')
  retorno = requests.get( link + filtro, headers = cabecalho)
  if retorno.status_code == 429:
      time.sleep(300)
      print('Esperando um pouco')
      retorno = requests.get( link + filtro, headers = cabecalho)
  sp_artigo = BeautifulSoup(retorno.text, 'html.parser')
  artigo_json = json.loads(sp_artigo.text)
  if len(artigo_json["data"]) > 0:
    str_paperId = artigo_json["data"][0]["paperId"]
    return str_paperId
  return None
```



Geração do dataset

Se não tem **paperId**, pega os autores do próprio artigo e cria uma estrutura básica

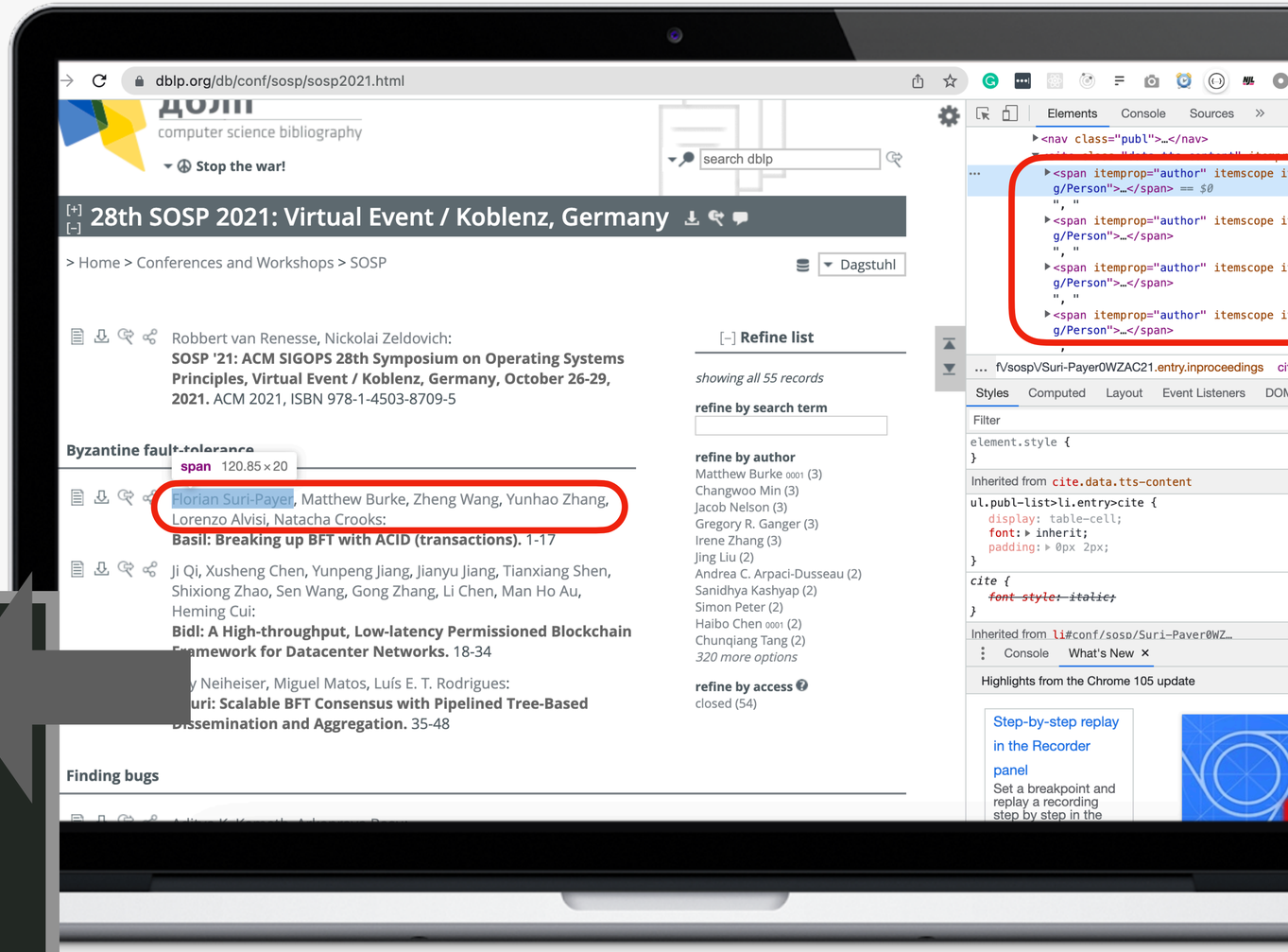
```
if paperId is None:
  print('paperId is None')
  authors = ''
  for author in article.find_all('span',{'itemprop':'author'}):
    authors = authors + author.find('span',{'itemprop':'name'}).get('title') + ','
  authors = authors[0:len(authors)-1]
  article_data = ['','',authors,conf,'',genre]
```



Geração do dataset

Caso contrário, complementa com dados da **API do semantic scholar** (lib do Python): **fieldsOfStudy, authors, topics...**

```
  else:
    print('paperId: ' + paperId)
    article_data = get_complementary_data(qty_article, pape
    print('articledata1: ', article_data)
    doi = article_data[1]
    print('doi: ', doi)

else:
  doi = doi_url.split('https://doi.org/')[1]
  print(doi)
  article_data = get_complementary_data(qty_article, doi)
  print('articledata2: ', article_data)
```

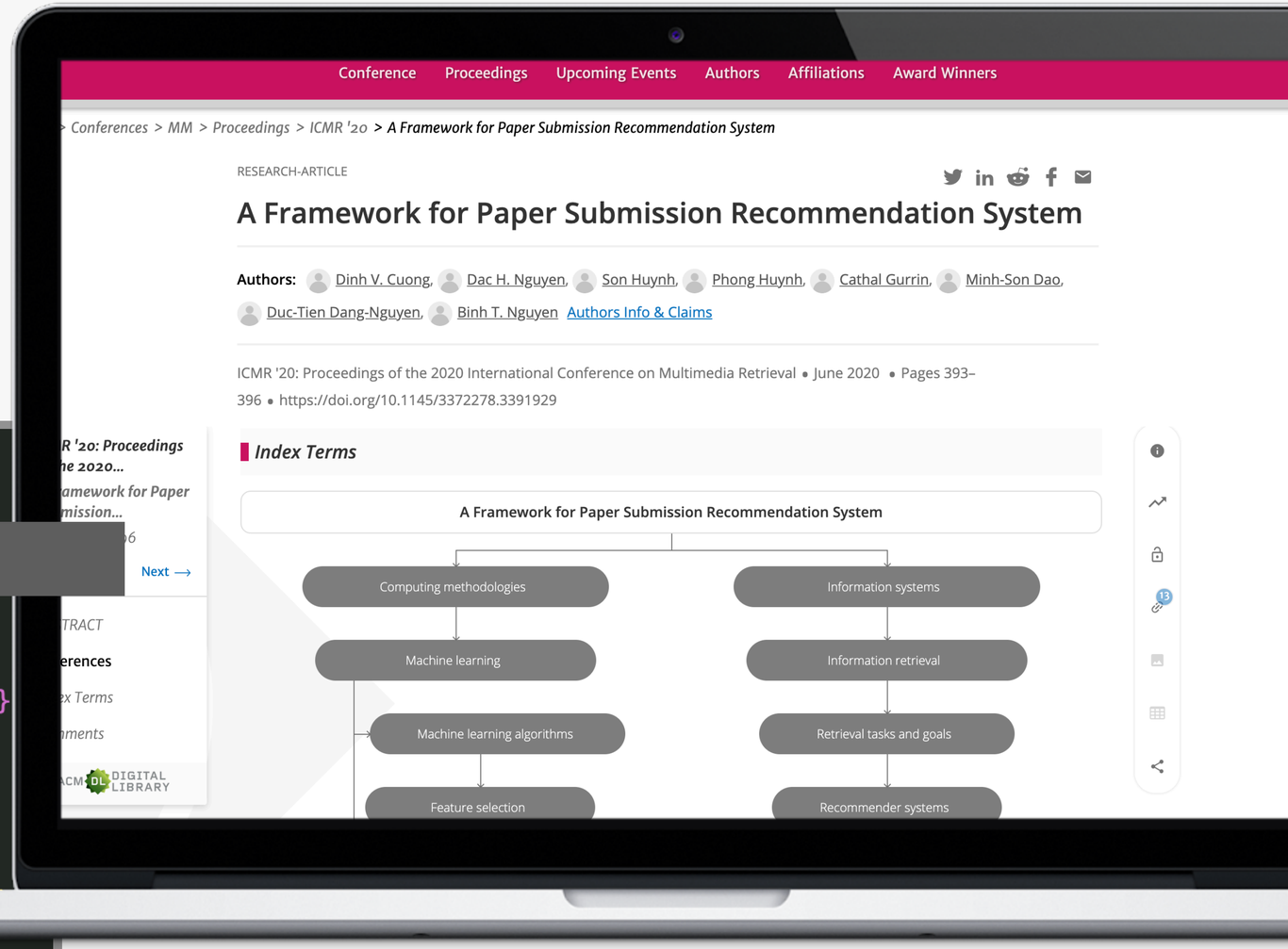


*A lib traz mais informações que a rota direta não traz, mas só pode ser acessada por doi ou paperId

Geração do dataset

Se tem doi pega **keywords** e **autores** do site do doi (vem do próprio dblp)

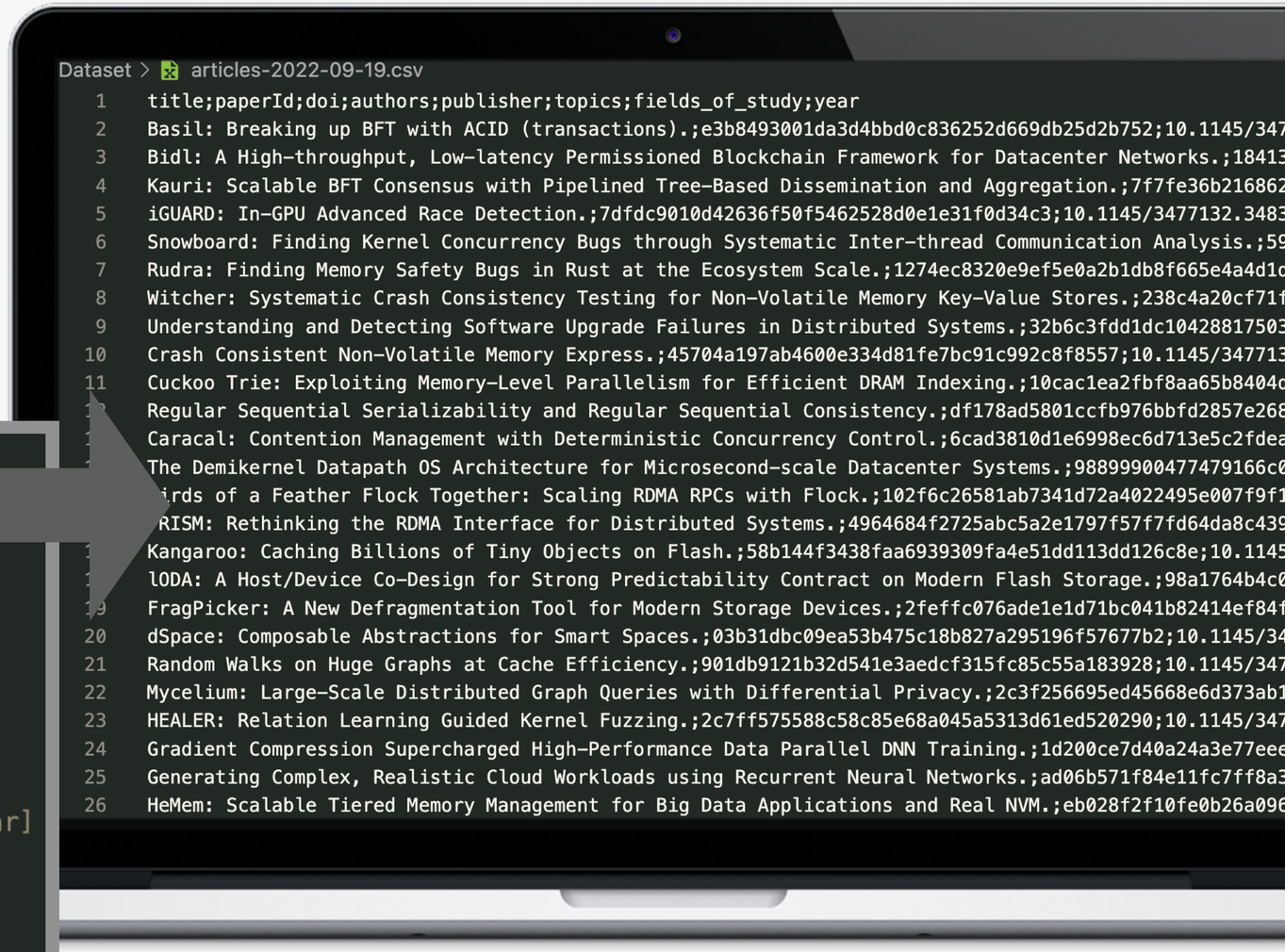
```
if doi != '':
  resposta = requests.get(doi_url, headers = cabecalho)
  sopa = BeautifulSoup(resposta.text, 'html.parser')
  # print(sopa.prettify())
  author_list = sopa.find('ul',{'ariaa-label':'authors'})
  if author_list is not None:
    for author in author_list.find_all('a',{'class':'author-name'}
      authors = authors + author.get('title') + ','
    authors = authors[0:len(authors)-1]
  print(authors)

  kw_list = sopa.find('ol',{'class':'rlist organizational
  if kw_list is not None:
    for kw in kw_list.find_all('a'):
      keywords = keywords + kw.get_text() + ','
    keywords = keywords[0:len(keywords)-1]
  print(keywords)
```



Geração do dataset

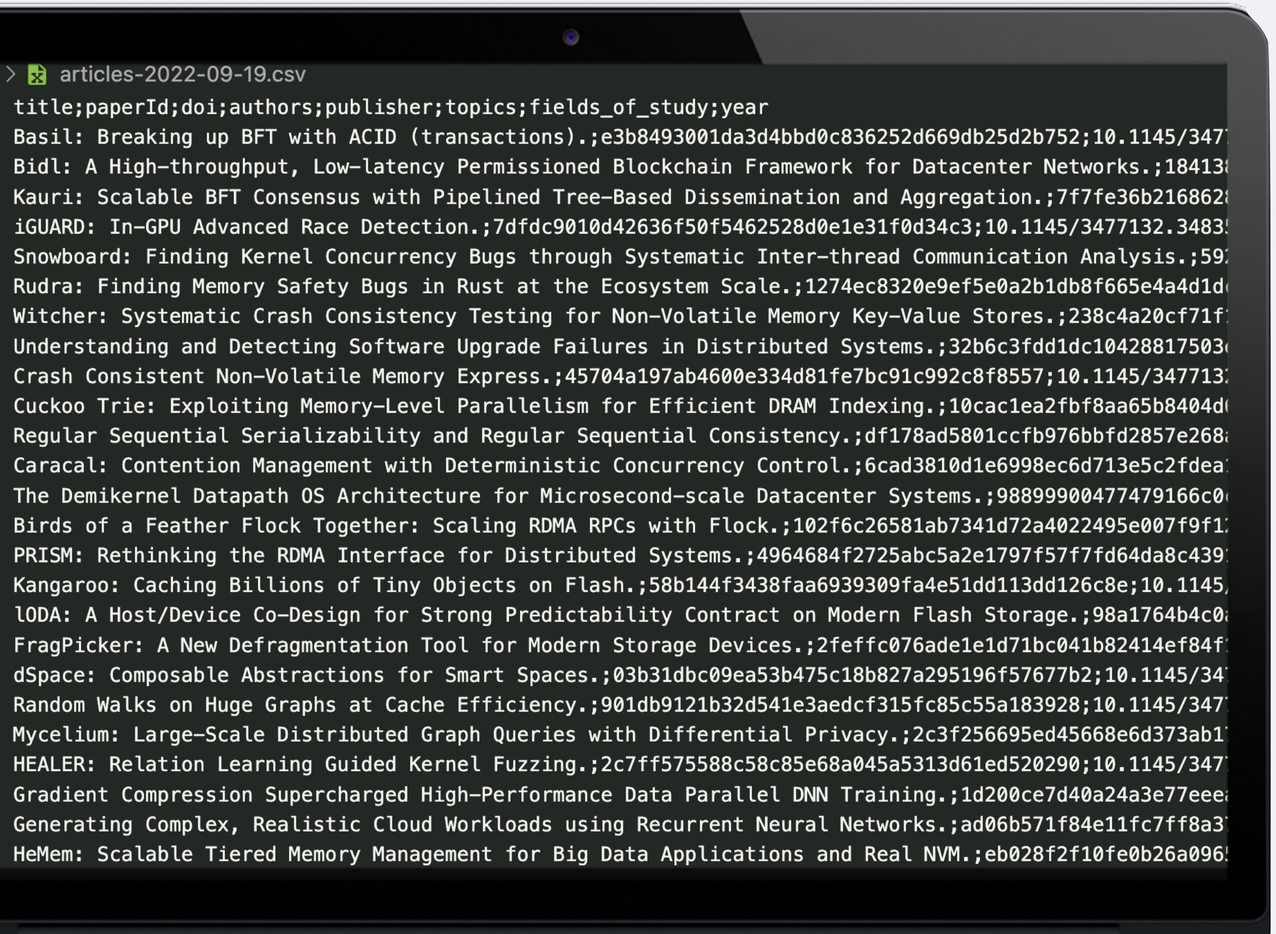
Adequa os dados e escreve em um arquivo **CSV**

```python
if article_data is not None:
  article_data.insert(0,title)
  article_data.append(year)
  article_data[3] = authors if authors != '' else article_data[3]
  article_data[4] = conf
  if article_data[5] == '' and keywords != '':
    article_data[5] = keywords
  elif article_data[5] != '' and keywords != '':
    article_data[5] = article_data[5] + ', ' + keywords
  # [title,paperId,doi,authors,publisher,topics,fields_of_study,year]
  if article_data[6] == '':
    article_data[6] = genre
  elif genre.lower() not in article_data[6].lower():
    article_data[6] = article_data[6] + ', ' + genre
else:
  article_data = [title,'',doi,authors,conf,keywords,genre,year]

writer.writerow(article_data)
qty_article += 1
```



Geração do dataset

Adequa os dados e escreve em um arquivo **CSV**

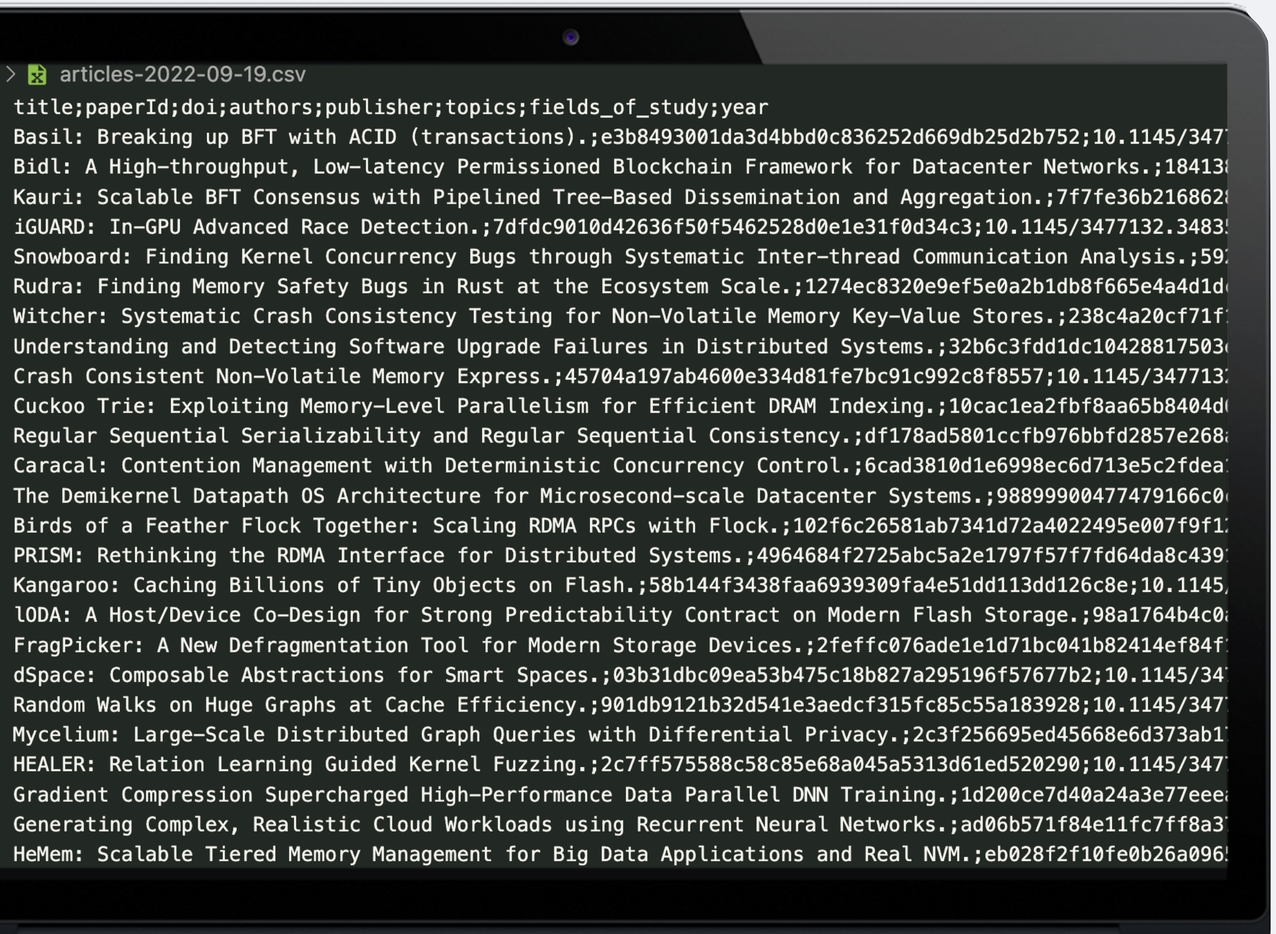


python™

title,
paperId,doi,authors,
publisher,topics,field
s_of_study,year

Geração do dataset

Adequa os dados e escreve em um arquivo **CSV**



python™

2783 artigos

3597 tópicos

Análise e Resultados

Análise e Resultados

2783 artigos

10 conferências

Entre 2018 e 2022

3597 tópicos

Análise e Resultados

Quantidade de tópicos por conferência:

Tabela 2. Quantidade de tópicos por conferência

Conferência	Quantidade de Tópicos
SOSP	383
OSDI	323
NDSS	1253
Mobihoc	628
SIGCOMM	485
SenSys	444
MOBICOM	702
CIDR	421
USENIX Security Symposium	1343
EUROCRYPT	395

Análise e Resultados

Criado um
novo Script
para a análise

Análises e resultados

Quais são os 6 tópicos mais publicados nos eventos?

Dicionário com a quantidade de vezes que um tópico aparece

```
{
  'COMPUTER SYSTEMS ORGANIZATION': 29,
  'DEPENDABLE AND FAULT-TOLERANT SYSTEMS AND NETWORKS': 18,
  'SECURITY AND PRIVACY': 25,
  ...
}
```

Análises e resultados

Quais são os 6 tópicos mais publicados nos eventos?

Dicionário com a quantidade de vezes que um tópico aparece

Ordenado de forma decrescente

Pegou os 6 primeiros

Análises e resultados

Quais são os 6 tópicos mais publicados nos eventos?

Tabela 3. Quantidade de artigos de cada um dos 6 maiores tópicos

Tópico	Quantidade de artigos
'EXPERIMENT'	154
'ALGORITHM'	150
'COMPUTER SCIENCE'	131
'SENSOR'	114
'PROTOTYPE'	113
'MACHINE LEARNING'	107

Análises e resultados

Quais são os 6 tópicos mais publicados por ano (de 2018 a 2022) nestes eventos?

Dicionário de anos com a quantidade de vezes que um tópico foi publicado naquele ano

```
{
  2022: {
    'NETWORK CONGESTION': 1,
    'ELASTICITY (DATA STORE)': 1,
    'THROUGHPUT': 1,
    'FREQUENCY RESPONSE': 1,
    ...
  }
}
```

Análises e resultados

Quais são os 6 tópicos mais publicados por ano (de 2018 a 2022) nestes eventos?

Dicionário de anos com a quantidade de vezes que um tópico foi publicado naquele ano

Ordenado de forma decrescente

Pegou os 6 primeiros de cada ano

Análises e resultados

Quais são os 6 tópicos mais publicados por ano (de 2018 a 2022) nestes eventos?

Tabela 4. Top 6 Tópicos Publicados Em Cada Ano

Ano	Top 6 Tópicos
’2018’	’EXPERIMENT’, ’SENSOR’, ’PROTOTYPE’, ’ALGORITHM’, ’OVERHEAD (COMPUTING)’, ’THROUGHPUT’
’2019’	’ALGORITHM’, ’EXPERIMENT’, ’PROTOTYPE’, ’MACHINE LEARNING’, ’SENSOR’, ’SCALABILITY’
’2020’	’COMPUTER SCIENCE’, ’ALGORITHM’, ’EXPERIMENT’, ’NETWORKS’, ’MACHINE LEARNING’, ’ADVERSARY (CRYPTOGRAPHY)’
’2021’	’COMPUTER SCIENCE’, ’NETWORKS’, ’SOFTWARE AND ITS ENGINEERING’, ’COMPUTER SYSTEMS ORGANIZATION’, ’SOFTWARE ORGANIZATION AND PROPERTIES’, ’INFORMATION SYSTEMS’
’2022’	’COMPUTER SCIENCE’, ’NETWORK CONGESTION’, ’ELASTICITY (DATA STORE)’, ’THROUGHPUT’, ’FREQUENCY RESPONSE’, ’FLOW’

Análises e resultados

Para cada tópico, quais as conferências que mais os publicam?

```
{
 'NP-HARDNESS': {
    'MobiHoc ': {
        2020: 2,
        2019: 2,
        2018: 2
    },
    'MOBICOM ': {
        2018: 1
    }
 },
 'ONLINE LEARNING SETTINGS': {
    'MobiHoc ': {
        2021: 1
    }
 }
  ...
}
```

```
{
 'TROJAN HORSE (COMPUTING)': {
    2018: {
        'NDSS ': 1,
        'USENIX Security Symposium': 1
    },
    2021: {
        'USENIX Security Symposium': 1
    },
    2020: {
        'USENIX Security Symposium': 1
    },
    2019: {
        'USENIX Security Symposium': 3
    }
 },
  ...
```

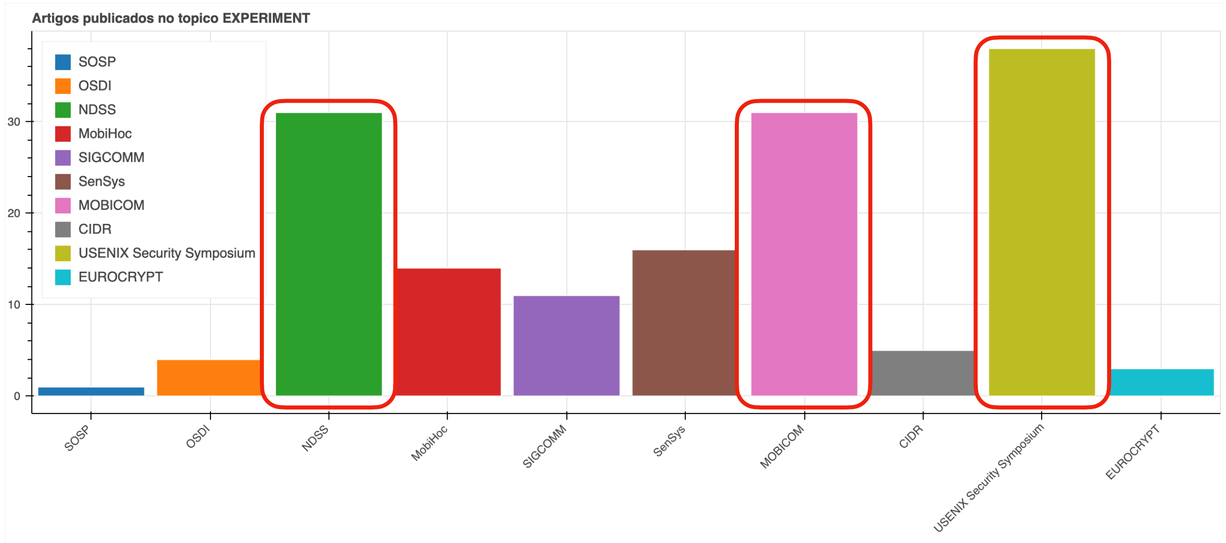
Análises e resultados

Para cada tópico, quais as conferências que mais os publicam?

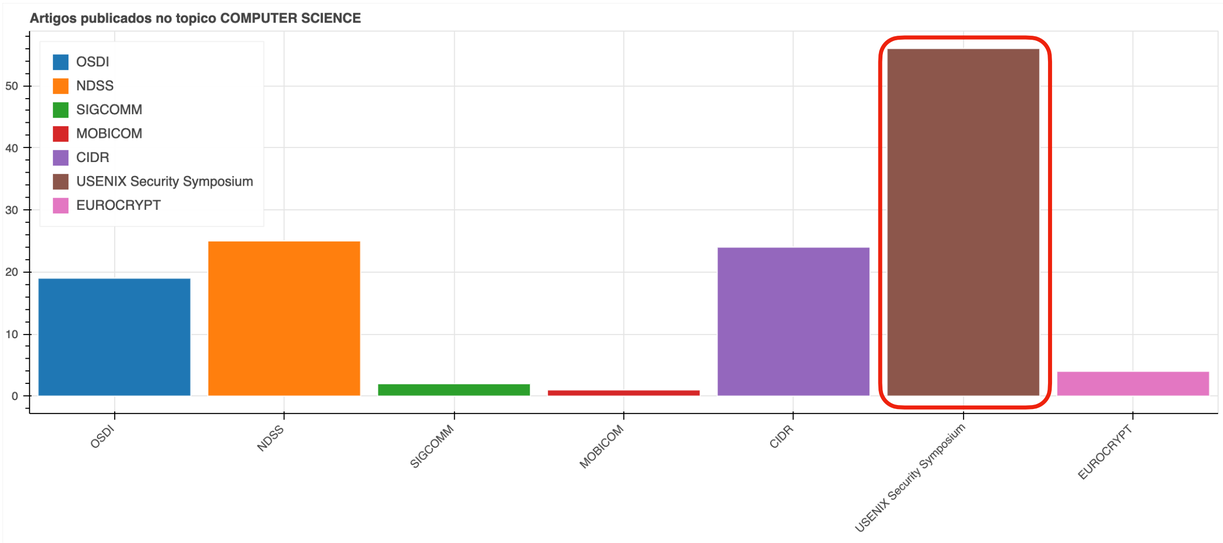
Para cada tópico, faz a soma da quantidade de publicações por conferência

Análises e resultados

Para cada tópico, quais as conferências que mais os publicam?



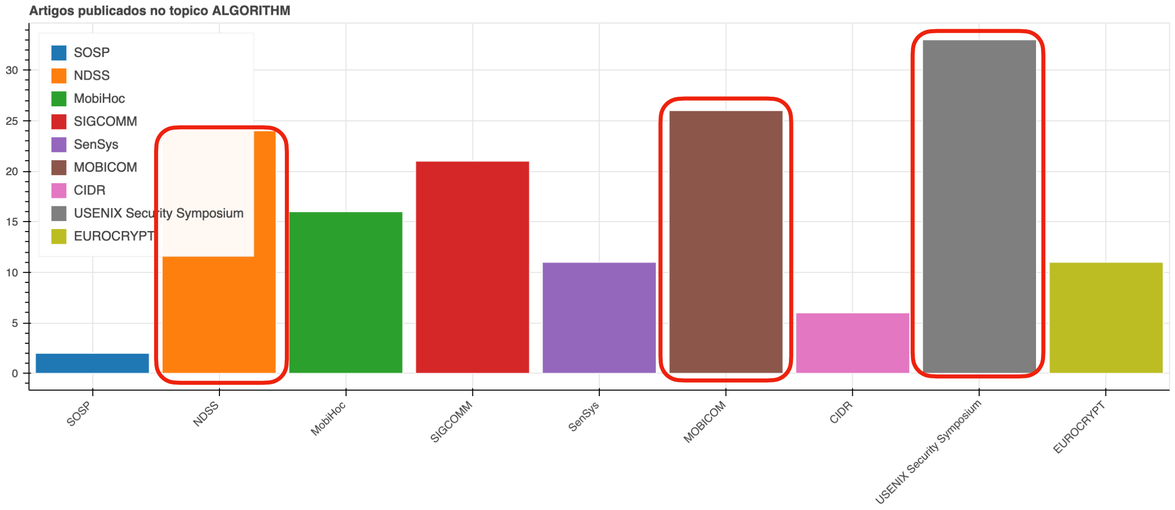
EXPERIMENT



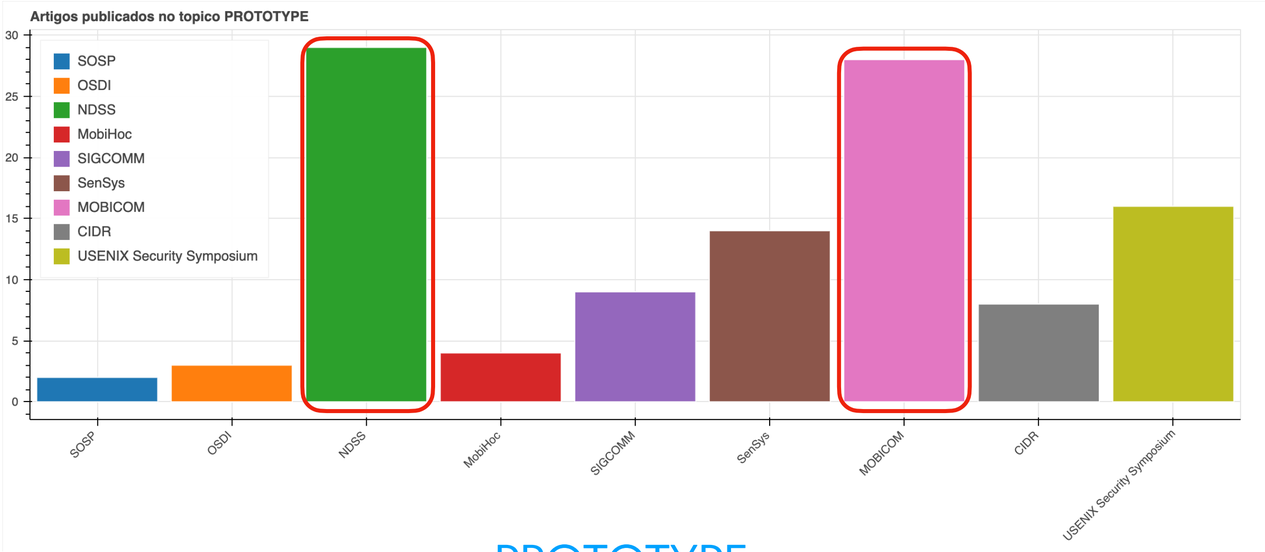
COMPUTER SCIENCE

Análises e resultados

Para cada tópico, quais as conferências que mais os publicam?



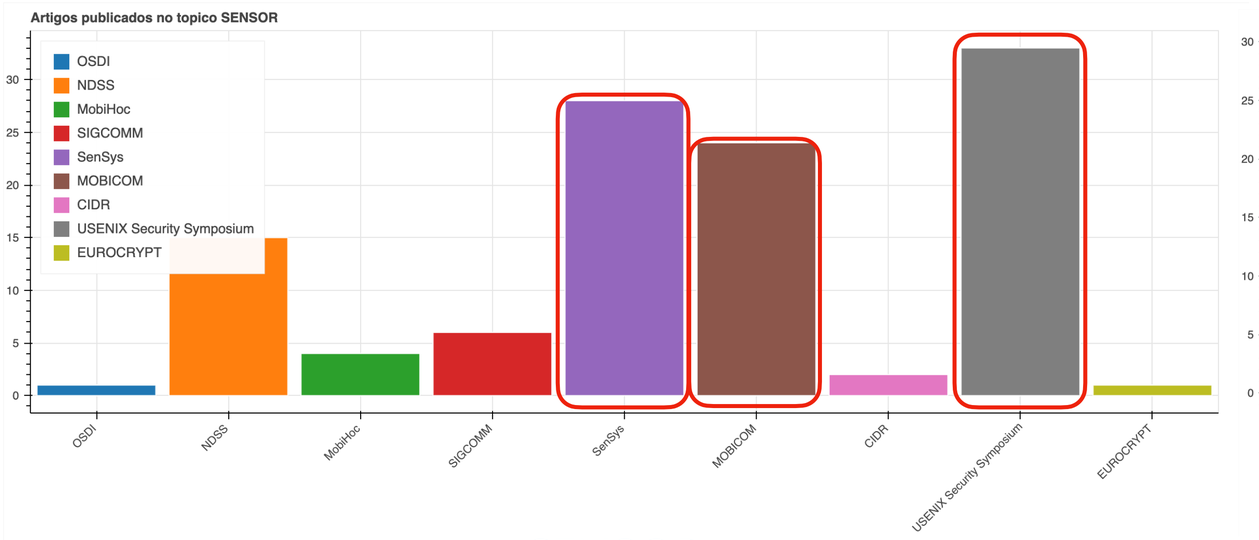
ALGORITHM



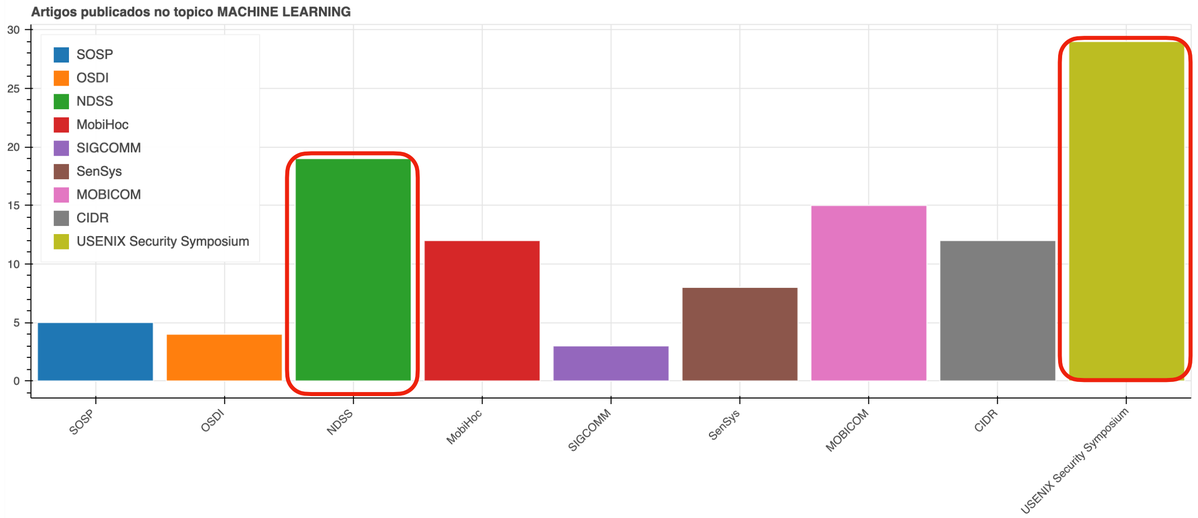
PROTOTYPE

Análises e resultados

Para cada tópico, quais as conferências que mais os publicam?



SENSOR



MACHINE LEARNING

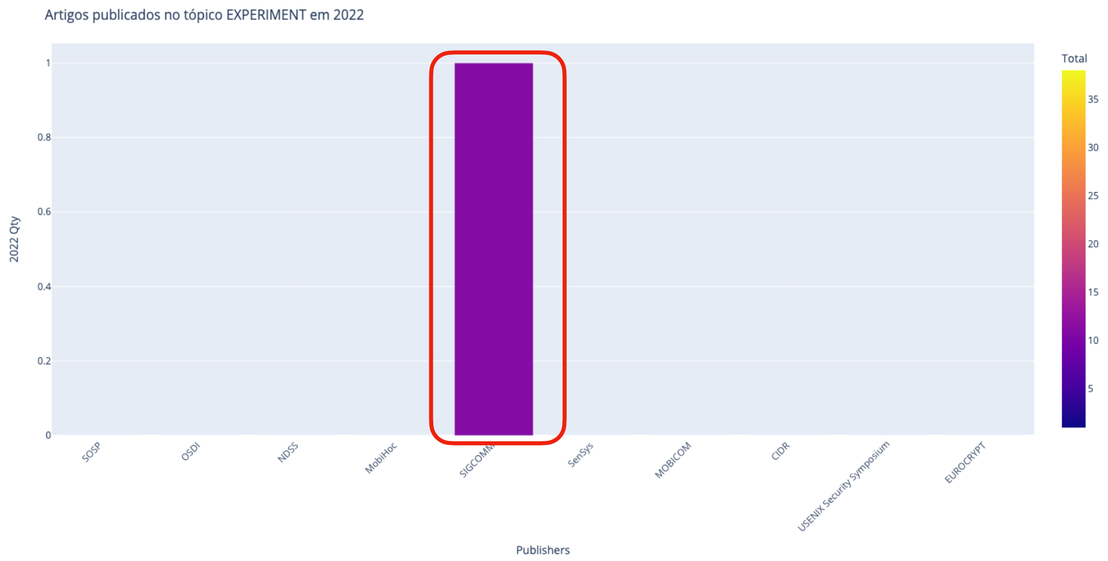
Análises e resultados

Para cada tópico, quais as conferências que mais os publicam?

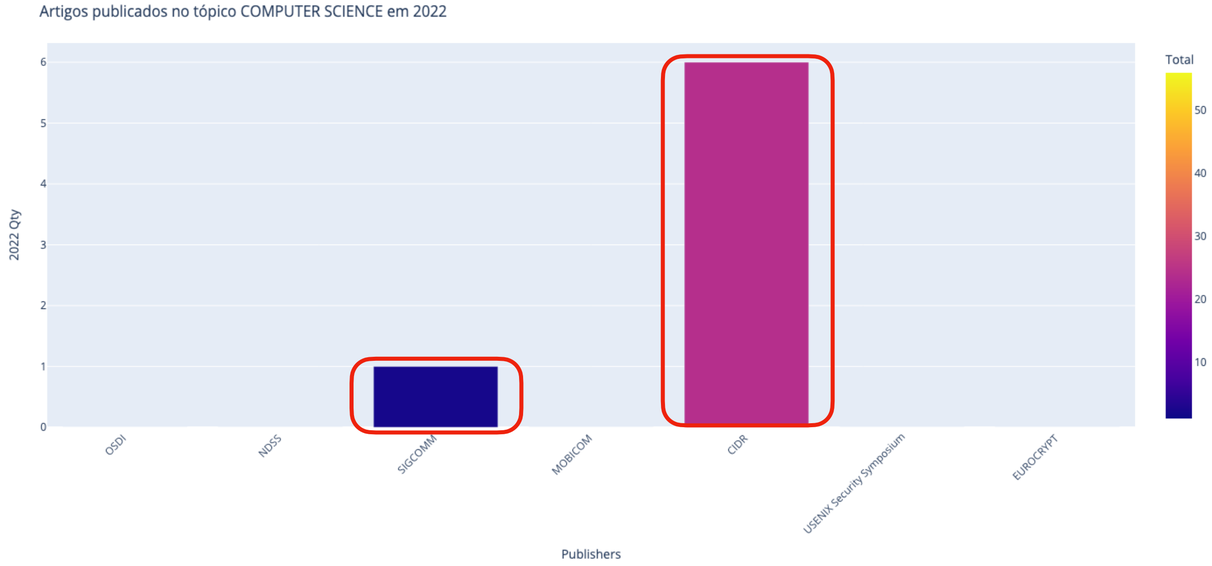
Publicações do último ano: Ordenar o ano de forma crescente e pegar a última linha para todas as conferências (de cada tópico)

Análises e resultados

Para cada tópico, quais as conferências que mais os publicam?



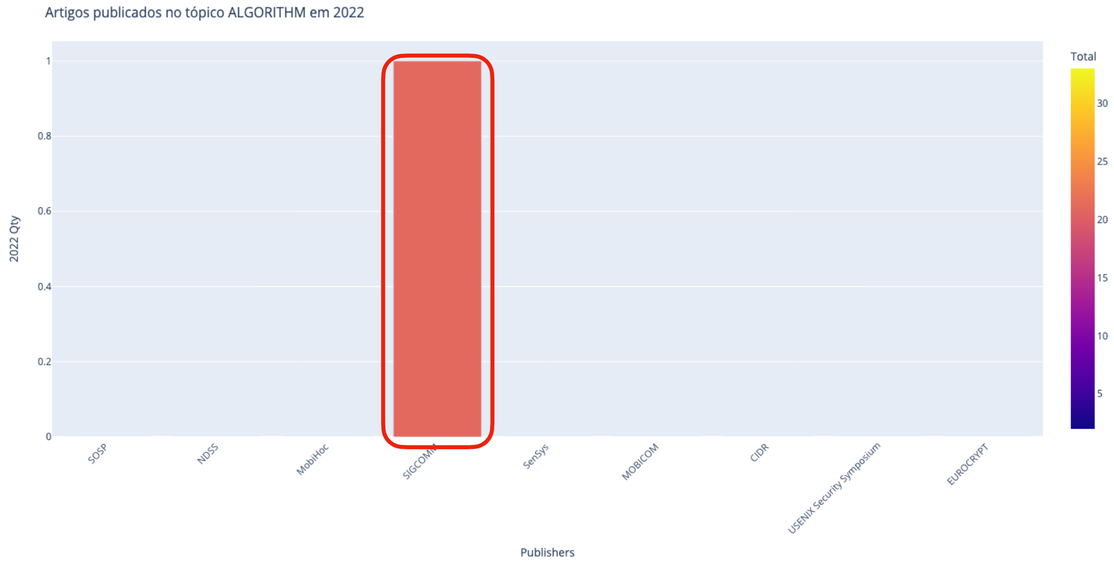
EXPERIMENT



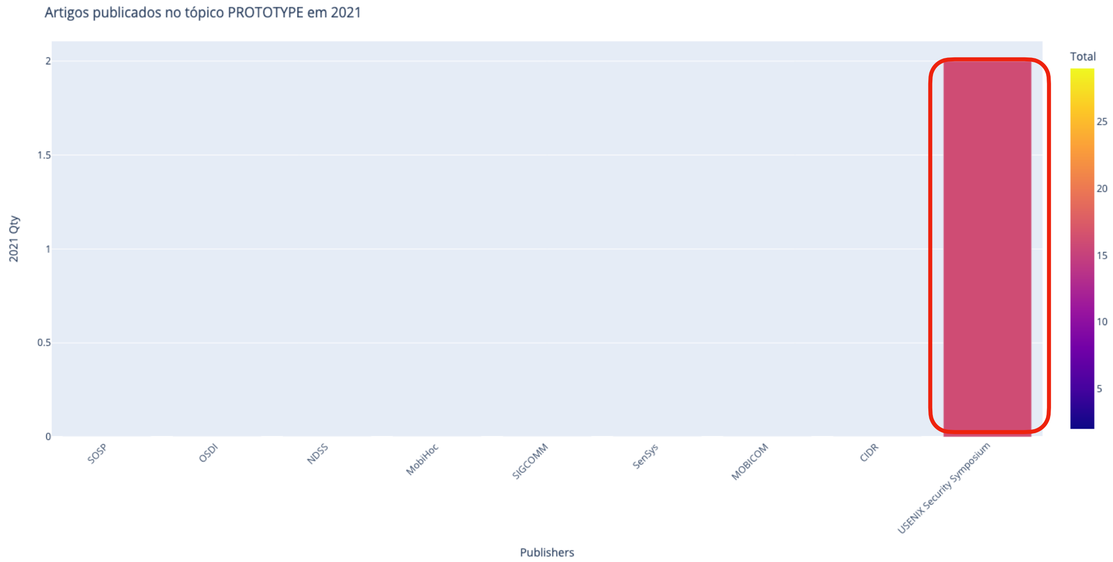
COMPUTER SCIENCE

Análises e resultados

Para cada tópico, quais as conferências que mais os publicam?



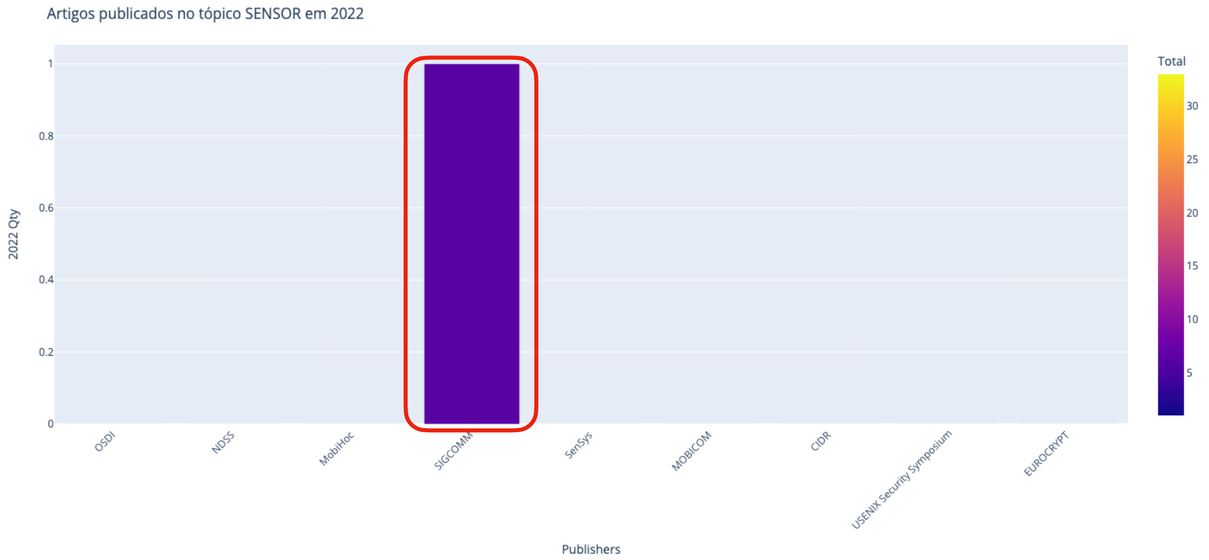
ALGORITHM



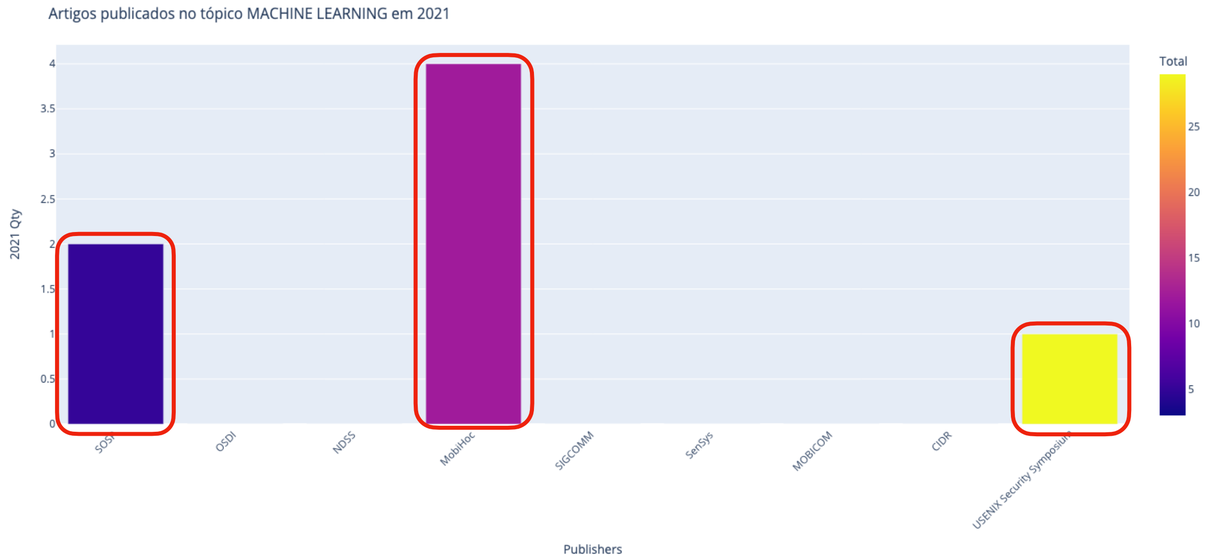
PROTOTYPE

Análises e resultados

Para cada tópico, quais as conferências que mais os publicam?



SENSOR



MACHINE LEARNING

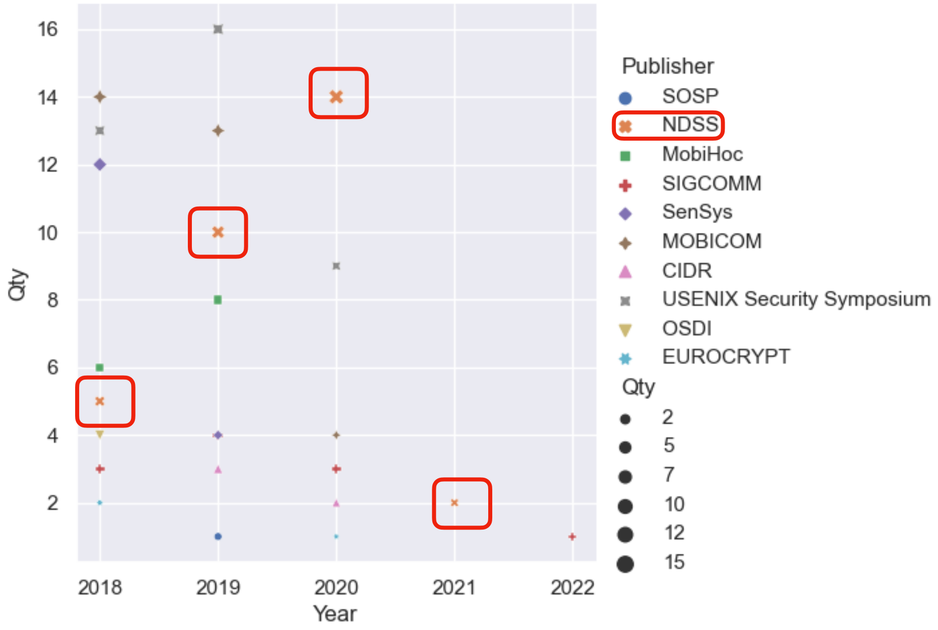
Análises e resultados

Como foram as evoluções para cada tópico das publicações das conferências ao longo dos anos?

Para cada tópico pega o quanto cada conferência publicou em cada ano

Análises e resultados

Como foram as evoluções ao longo dos anos de publicações das conferências para cada tópico?



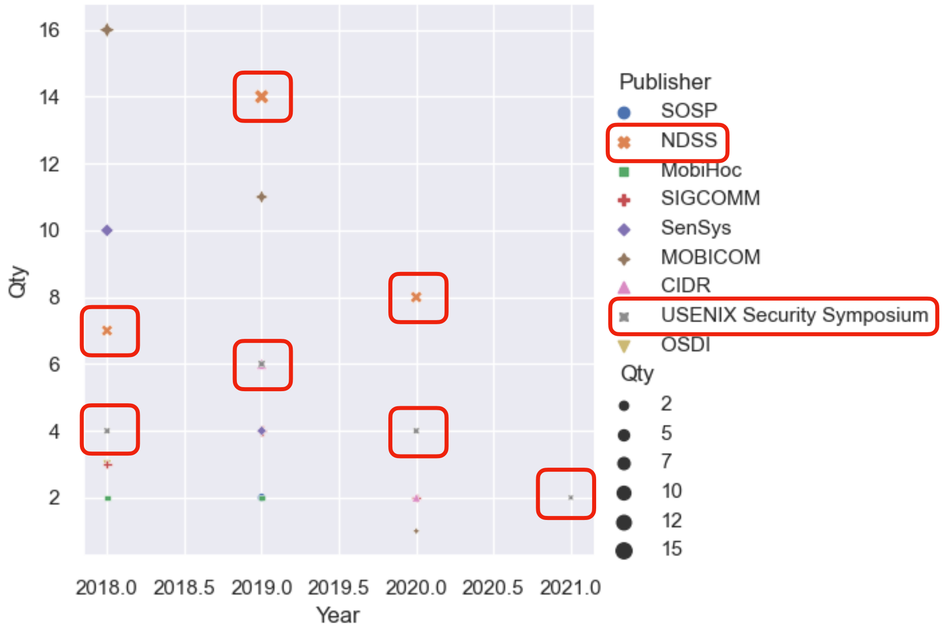
EXPERIMENT

COMPUTER SCIENCE

Análises e resultados

Como foram as evoluções ao longo dos anos de publicações das conferências para cada tópico?

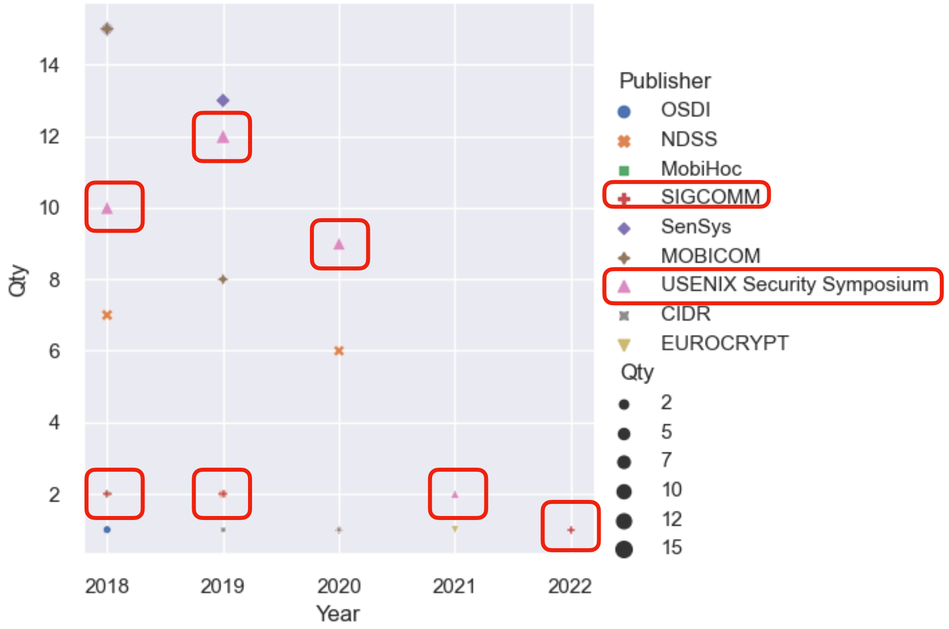
ALGORITHM



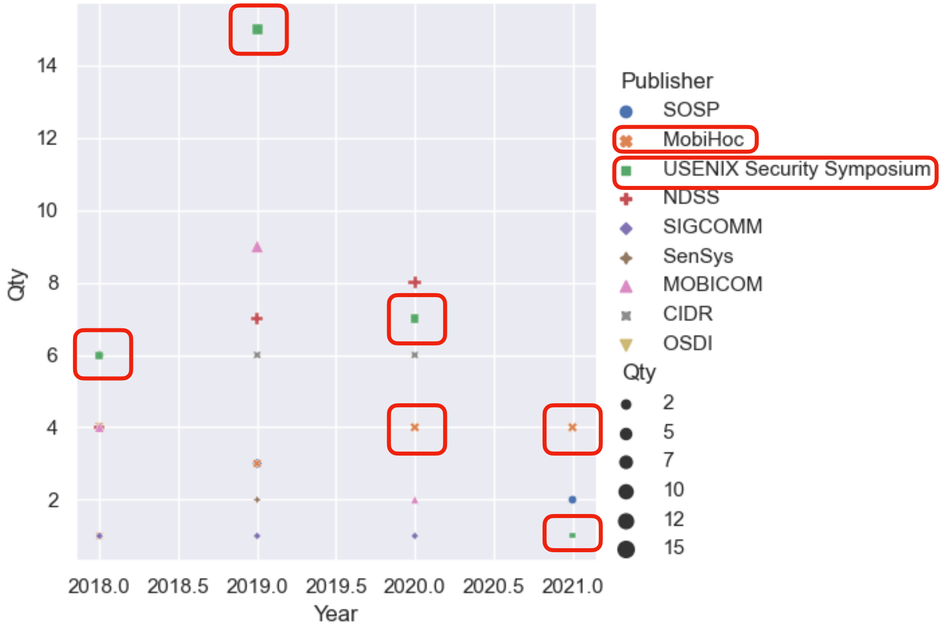
PROTOTYPE

Análises e resultados

Como foram as evoluções ao longo dos anos de publicações das conferências para cada tópico?



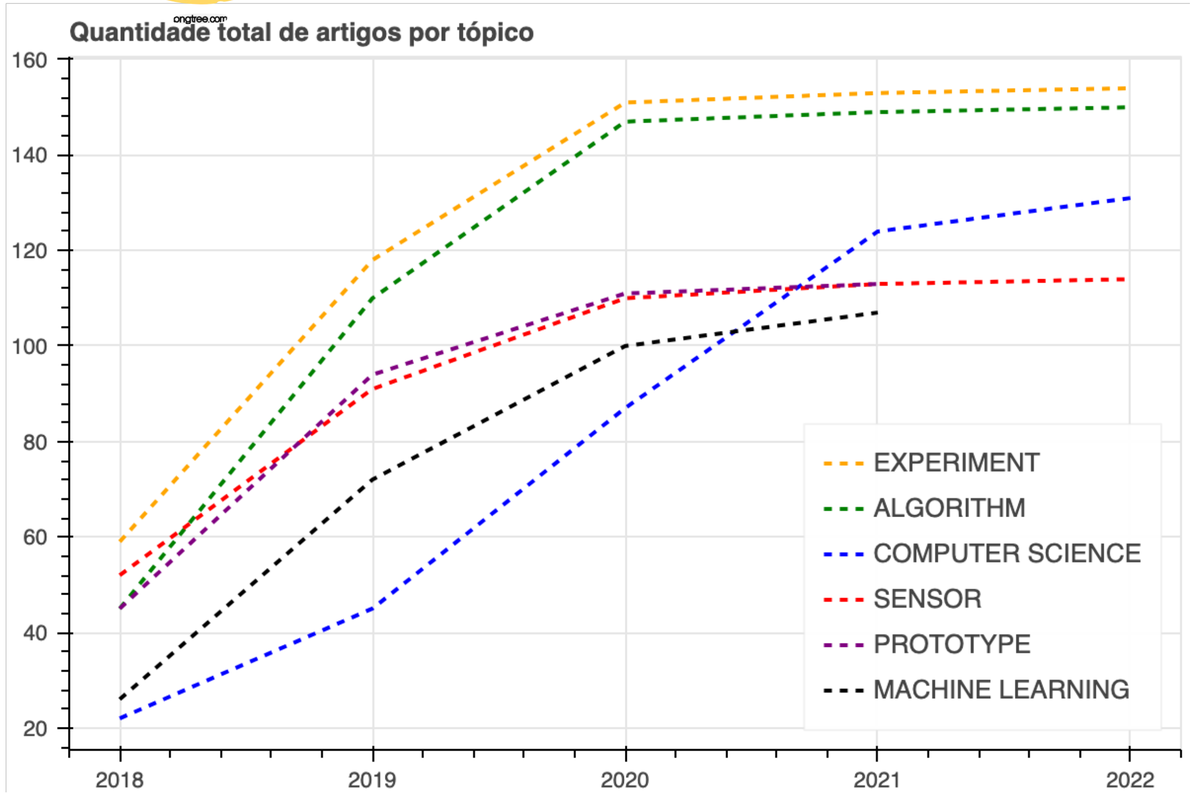
SENSOR

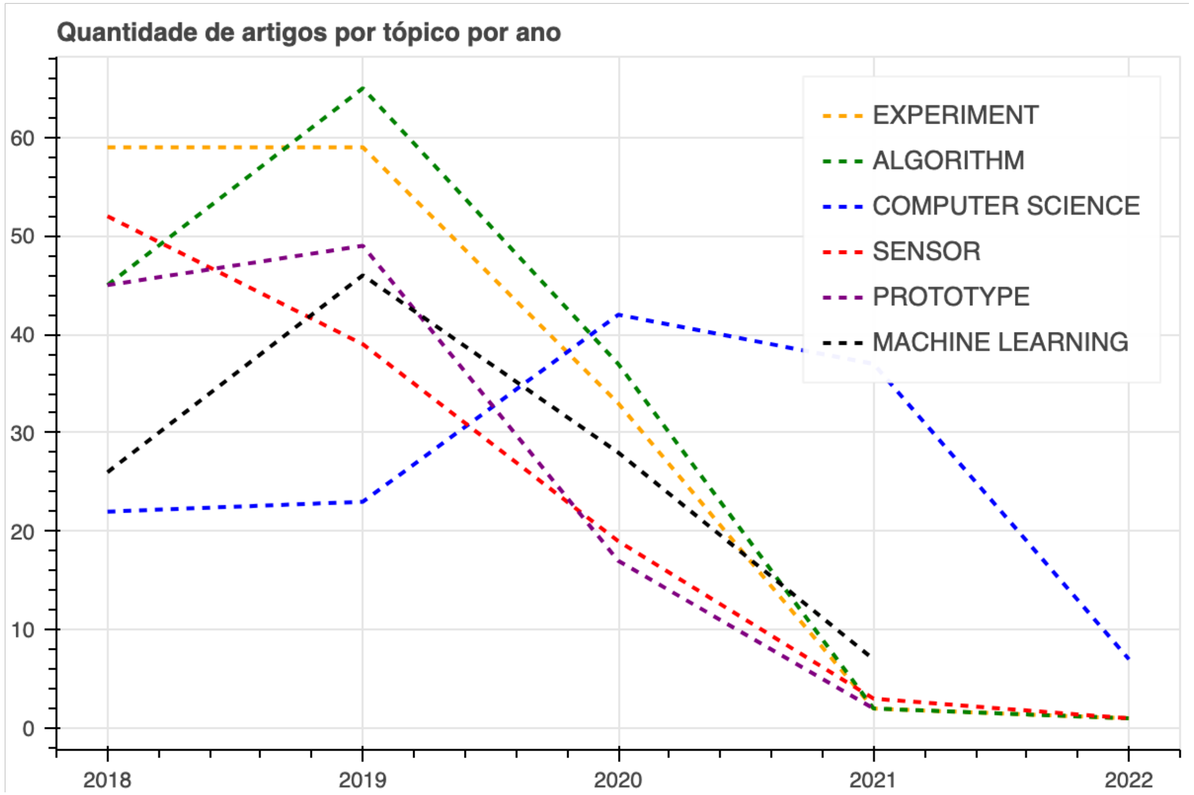


MACHINE LEARNING

Análises e resultados

Como foram as evoluções ao longo dos anos de publicações das conferências para cada tópico?





Análises e resultados

Para cada tópico, em que categoria as conferências entram em relação a publicação dos mesmos?

Dicionário de publicações de conferências por tópico

```
{
    'SOSP ': {
        'COMPUTER SYSTEMS ORGANIZATION': 23,
        'DEPENDABLE AND FAULT-TOLERANT SYSTEMS AND NETWORKS
            ': 18,
        'SECURITY AND PRIVACY': 12,
        'DATABASE AND STORAGE SECURITY': 2
        ...
    }
}
```

Análises e resultados

Para cada tópico, em que categoria as conferências entram em relação a publicação dos mesmos?

Dicionário de publicações de conferências por ano

```
{
    'SOSP ': {
        2021: 54,
        2019: 38
    },
    'OSDI ': {
        2021: 31,
        2020: 70,
        2018: 47
    }
}
```

Análises e resultados

Para cada tópico, em que categoria as conferências entram em relação a publicação dos mesmos?

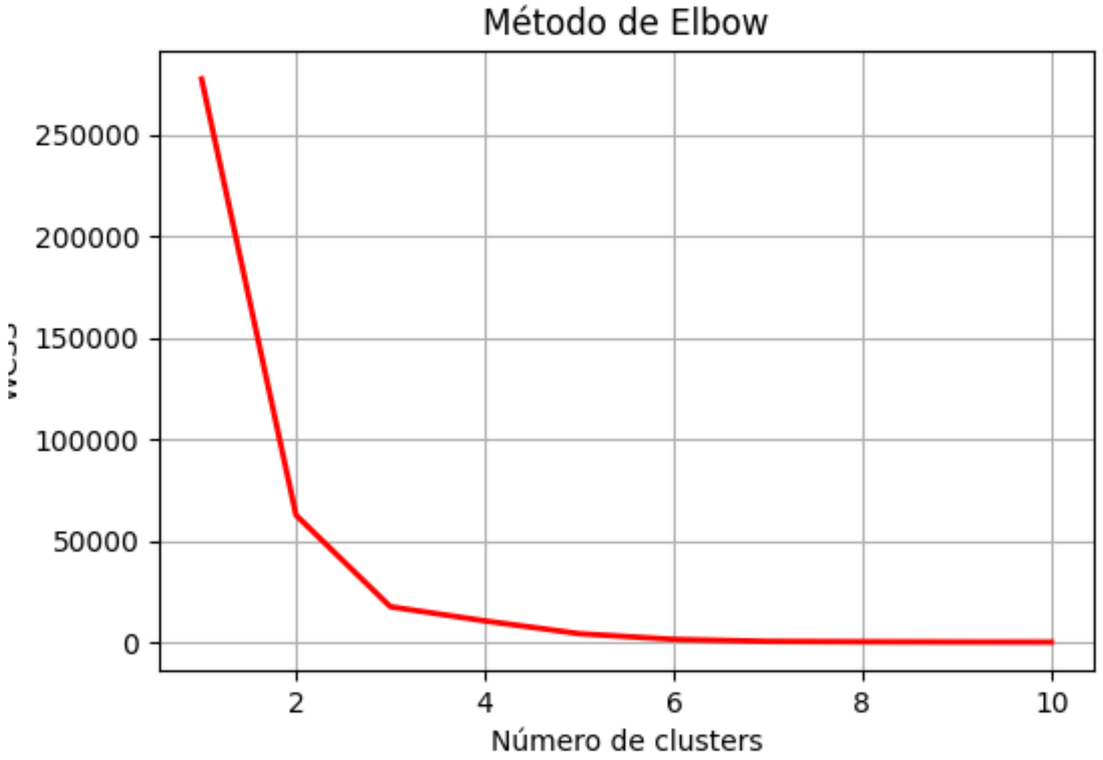
Para cada um dos 6 tópicos: Pega a quantidade total de publicações por conferência e compara com a quantidade de publicações do tópico na conferência

Algoritmo de clusterização K-Means por tópico

(Quantidade total do evento vs quantidade do evento pro tópico)

Análises e resultados

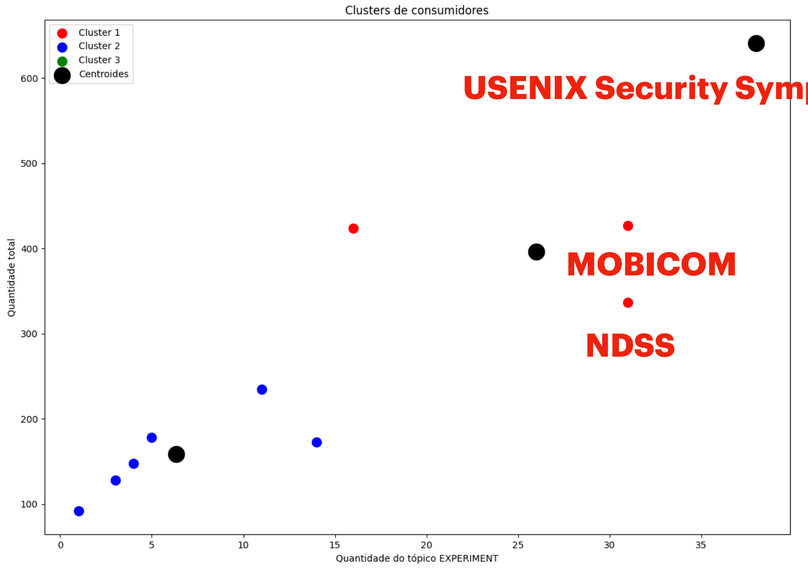
Para cada tópico, em que categoria as conferências entram em relação a publicação dos mesmos?

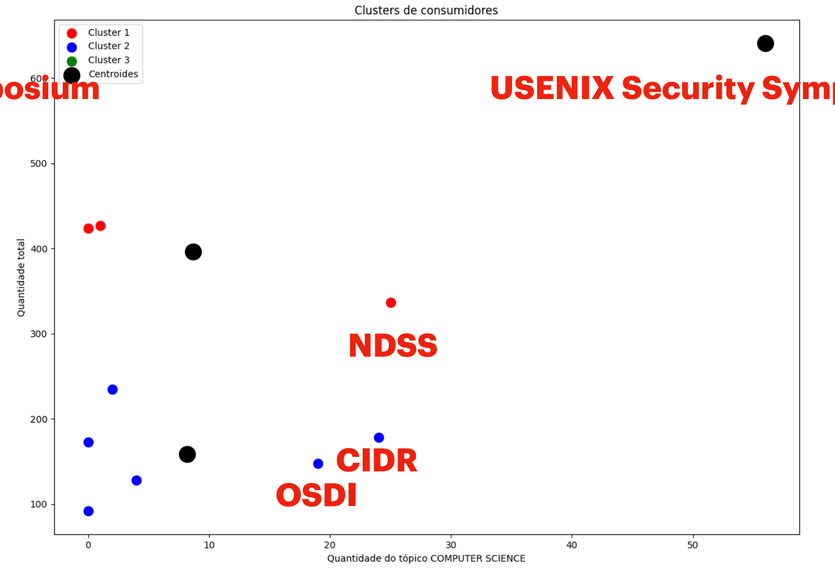


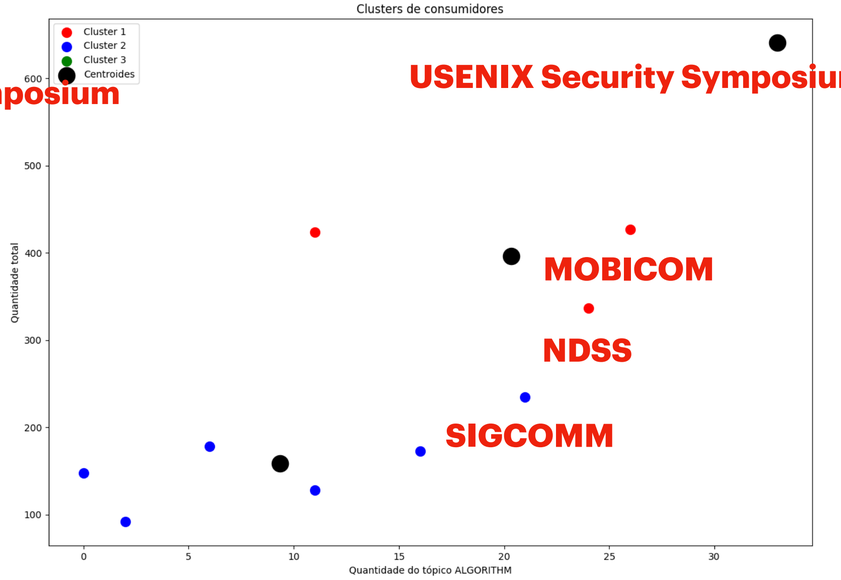
Quantidade de Clusters: 3

Análises e resultados

Para cada tópico, em que categoria as conferências entram em relação a publicação dos mesmos?

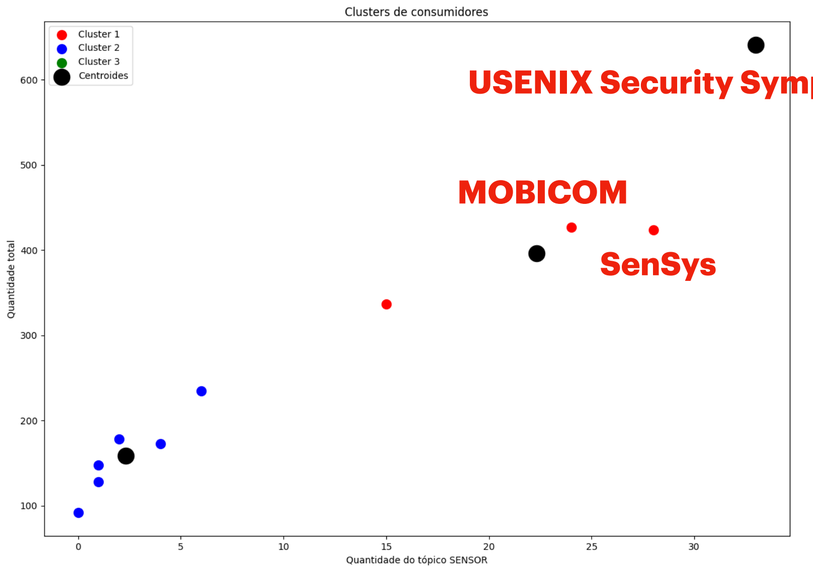


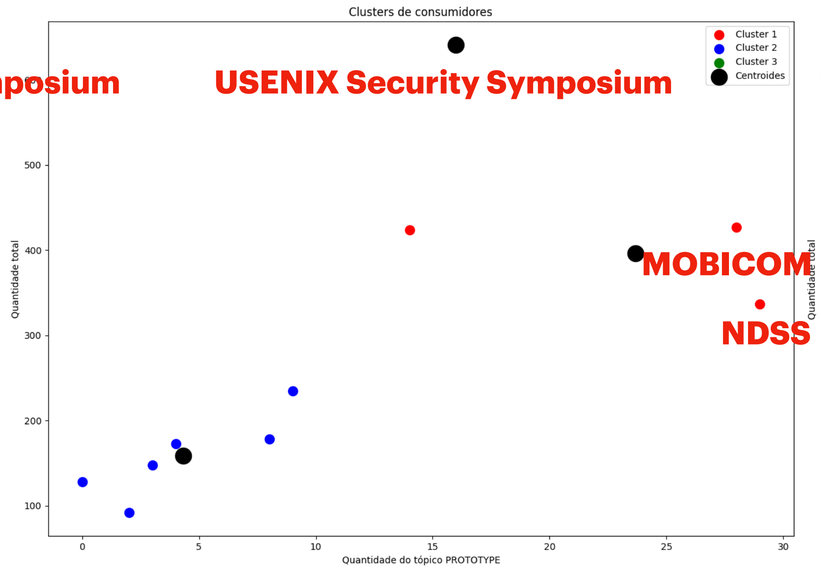


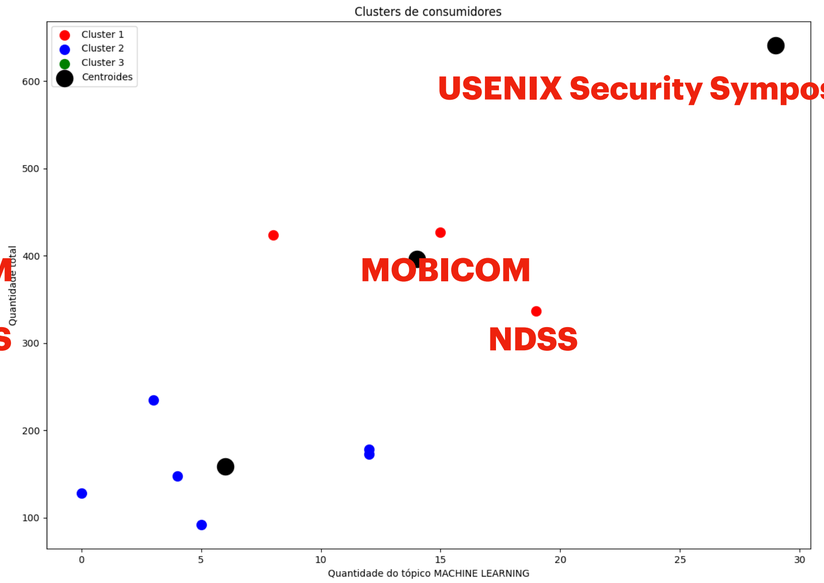


Análises e resultados

Para cada tópico, em que categoria as conferências entram em relação a publicação dos mesmos?







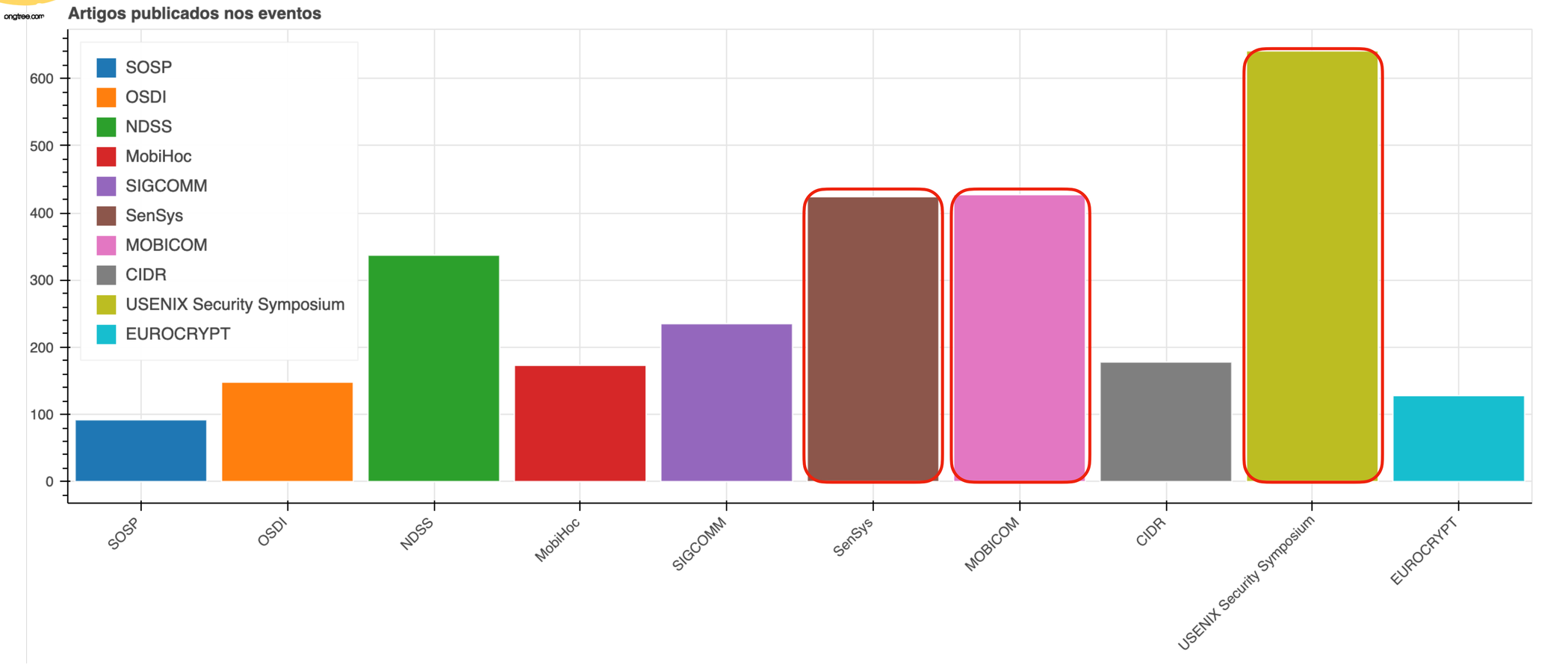
Análises e resultados

Quais as conferências que mais publicaram?

Utiliza o dicionário de publicações de conferências por ano e faz a soma de publicações por conferência - Total de publicações

Análises e resultados

Quais as conferências que mais publicaram?



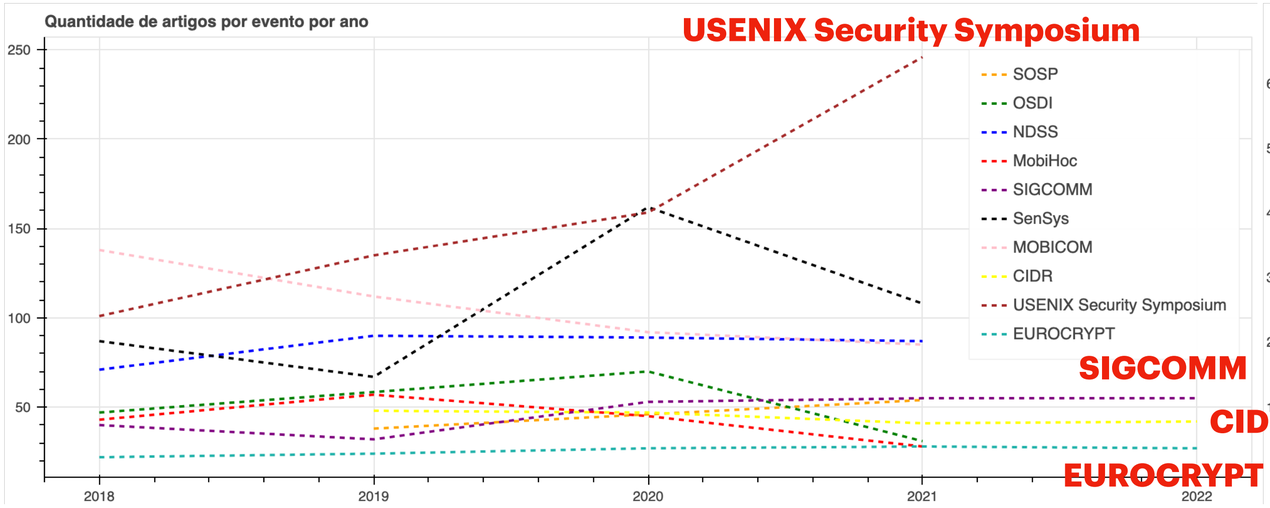
Análises e resultados

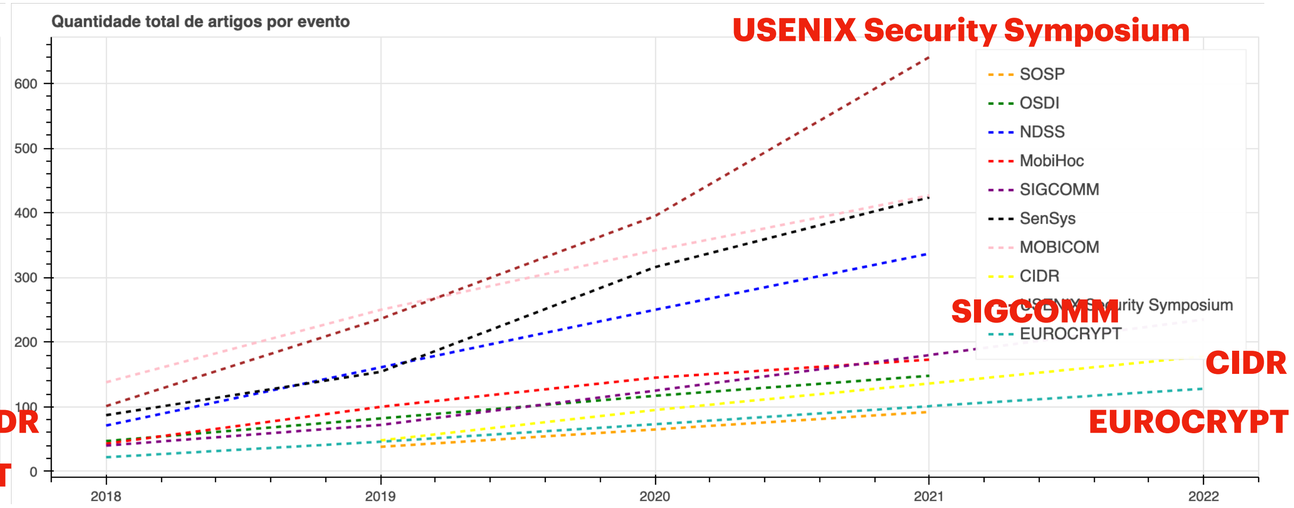
Como foram as evoluções ao longo dos anos para as conferências?

Utiliza o dicionário de publicações de conferências por ano e vai somando a quantidade de publicações por conferência ao longo dos anos

Análises e resultados

Como foram as evoluções ao longo dos anos para as conferências?





Análises e resultados

Para cada conferência, qual o tópico que mais aparece?

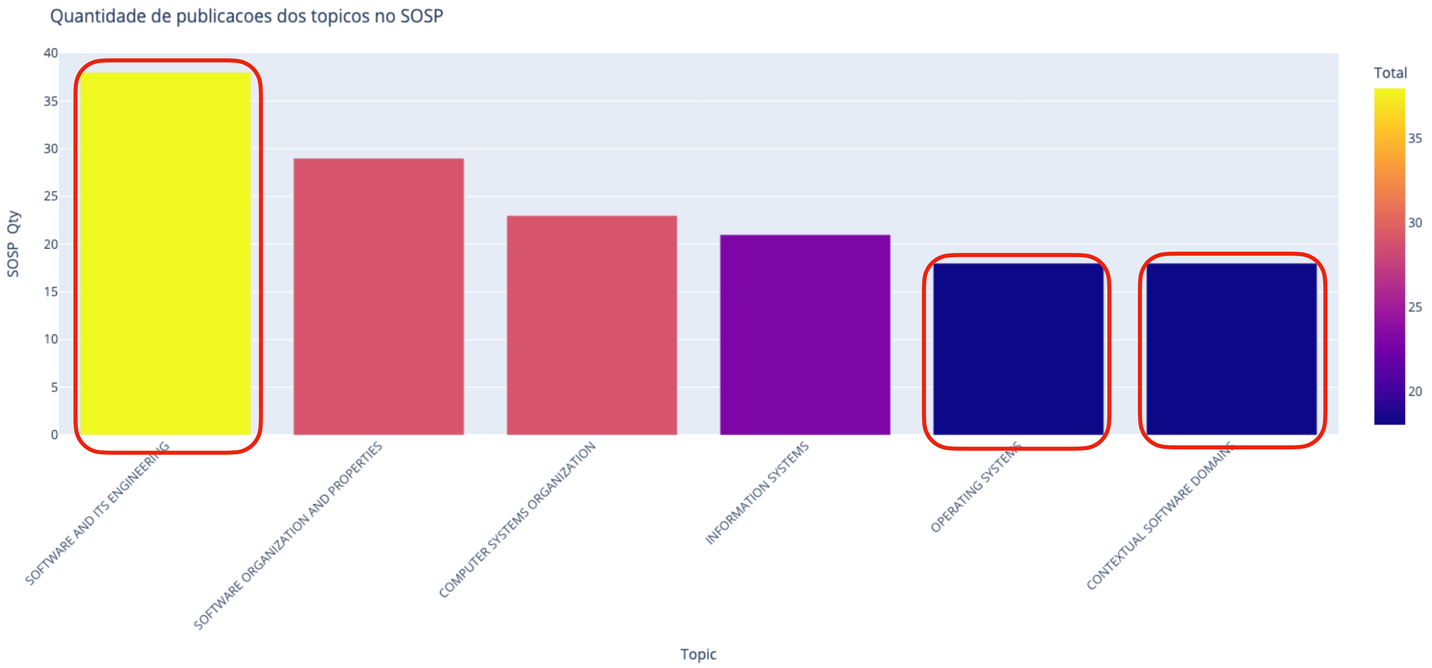
Para cada conferência, pega os 6 tópicos que mais tem publicações

Compara a quantidade total de publicações do tópico com a quantidade de publicações dele na conferência

Análises e resultados

Para cada conferência, qual o tópico que mais aparece?

SOSP

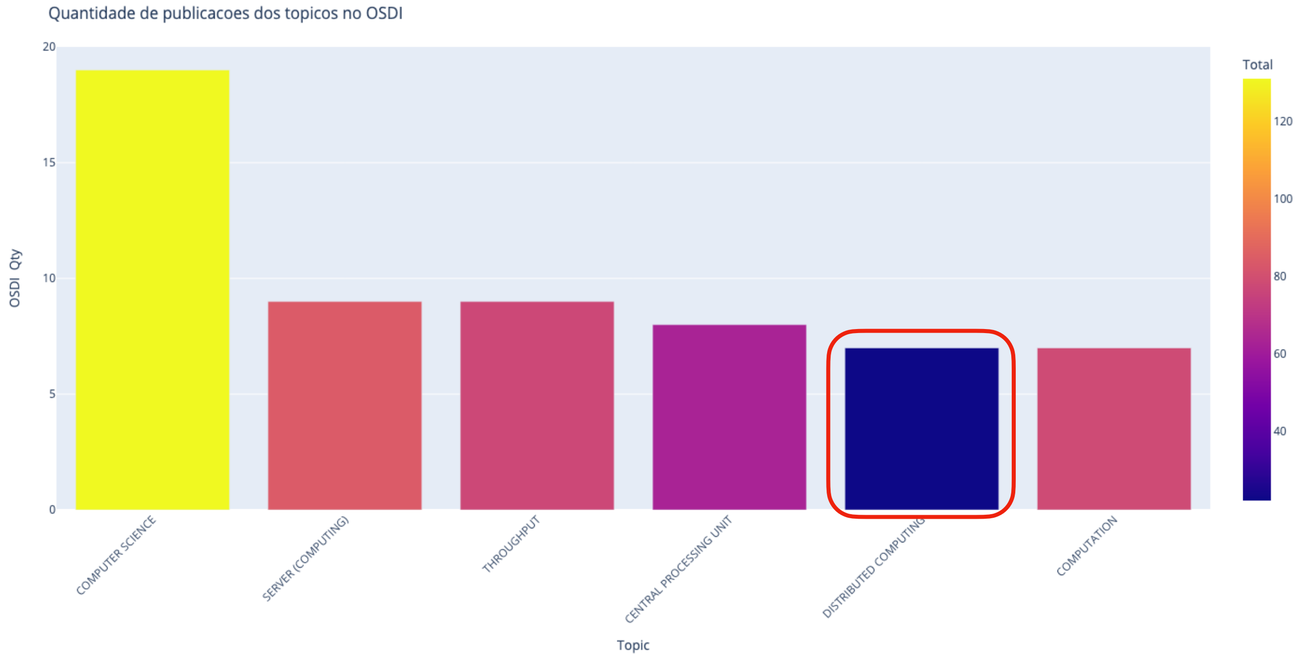


	Topic	Total	SOSP	Qty
0	SOFTWARE AND ITS ENGINEERING	38		38.0
1	SOFTWARE ORGANIZATION AND PROPERTIES	29		29.0
2	COMPUTER SYSTEMS ORGANIZATION	29		23.0
3	INFORMATION SYSTEMS	23		21.0
4	OPERATING SYSTEMS	18		18.0
5	CONTEXTUAL SOFTWARE DOMAINS	18		18.0

Análises e resultados

Para cada conferência, qual o tópico que mais aparece?

OSDI

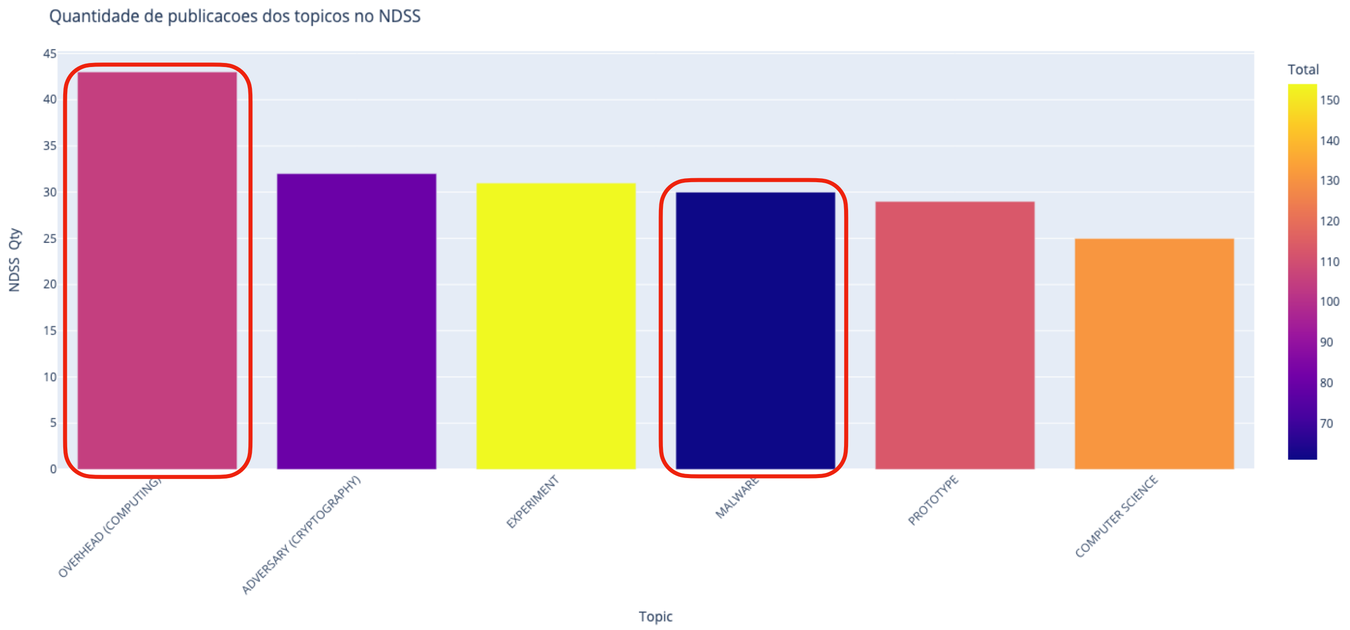


	Topic	Total	OSDI	Qty
0	COMPUTER SCIENCE	131		19.0
1	SERVER (COMPUTING)	84		9.0
2	THROUGHPUT	77		9.0
3	CENTRAL PROCESSING UNIT	63		8.0
4	DISTRIBUTED COMPUTING	22		7.0
5	COMPUTATION	78		7.0

Análises e resultados

Para cada conferência, qual o tópico que mais aparece?

NDSS

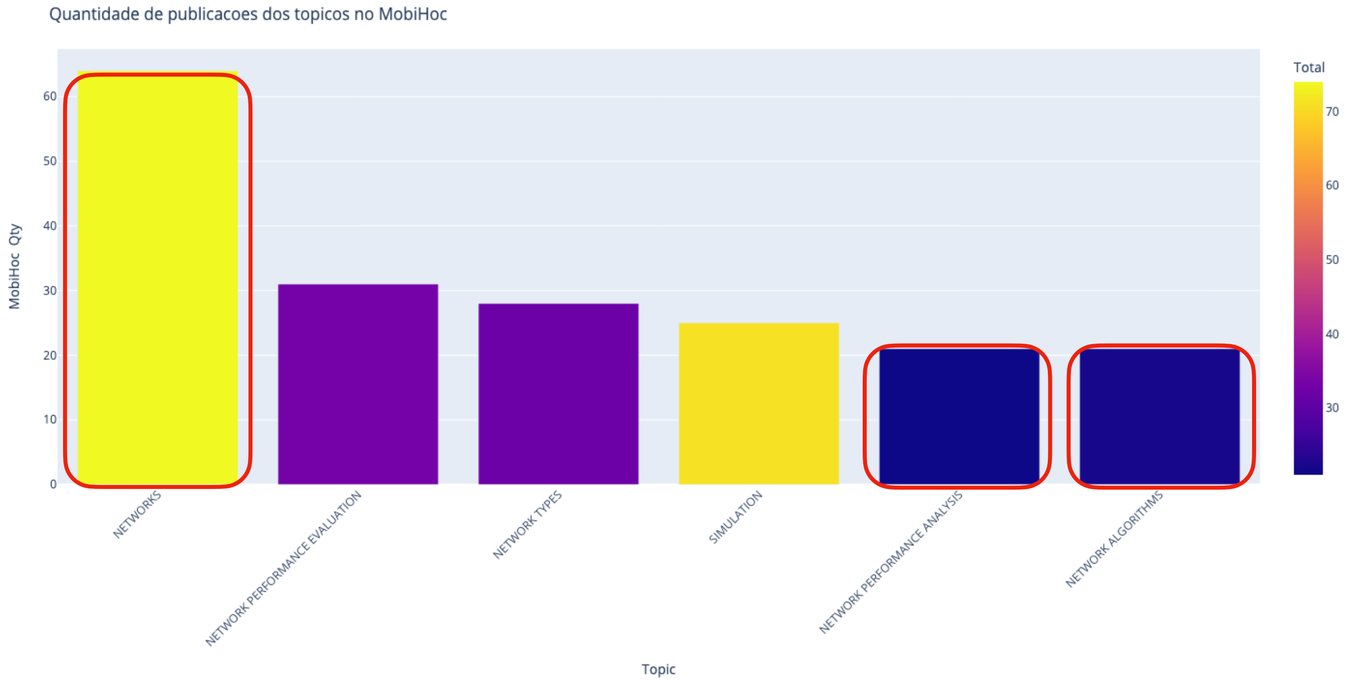


	Topic	Total	NDSS Qty
0	OVERHEAD (COMPUTING)	105	43.0
1	ADVERSARY (CRYPTOGRAPHY)	80	32.0
2	EXPERIMENT	154	31.0
3	MALWARE	61	30.0
4	PROTOTYPE	113	29.0
5	COMPUTER SCIENCE	131	25.0

Análises e resultados

Para cada conferência, qual o tópico que mais aparece?

MobiHoc

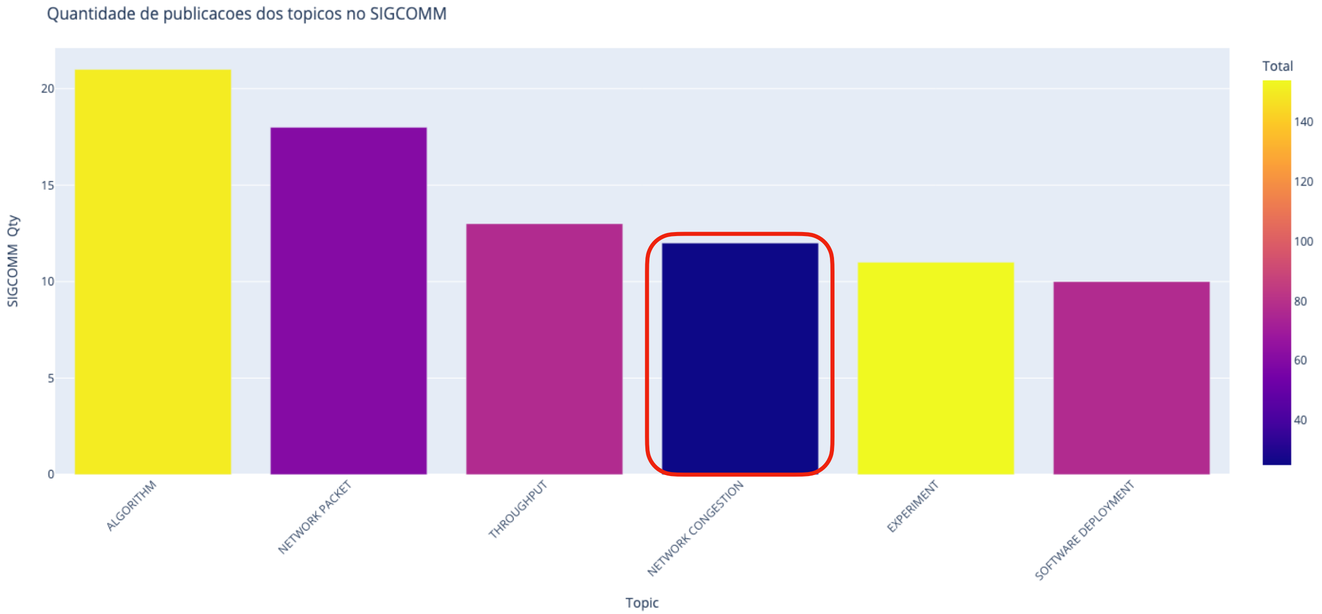


	Topic	Total	MobiHoc	Qty
0	NETWORKS	74		64.0
1	NETWORK PERFORMANCE EVALUATION	33		31.0
2	NETWORK TYPES	32		28.0
3	SIMULATION	71		25.0
4	NETWORK PERFORMANCE ANALYSIS	21		21.0
5	NETWORK ALGORITHMS	22		21.0

Análises e resultados

Para cada conferência, qual o tópico que mais aparece?

SIGCOMM

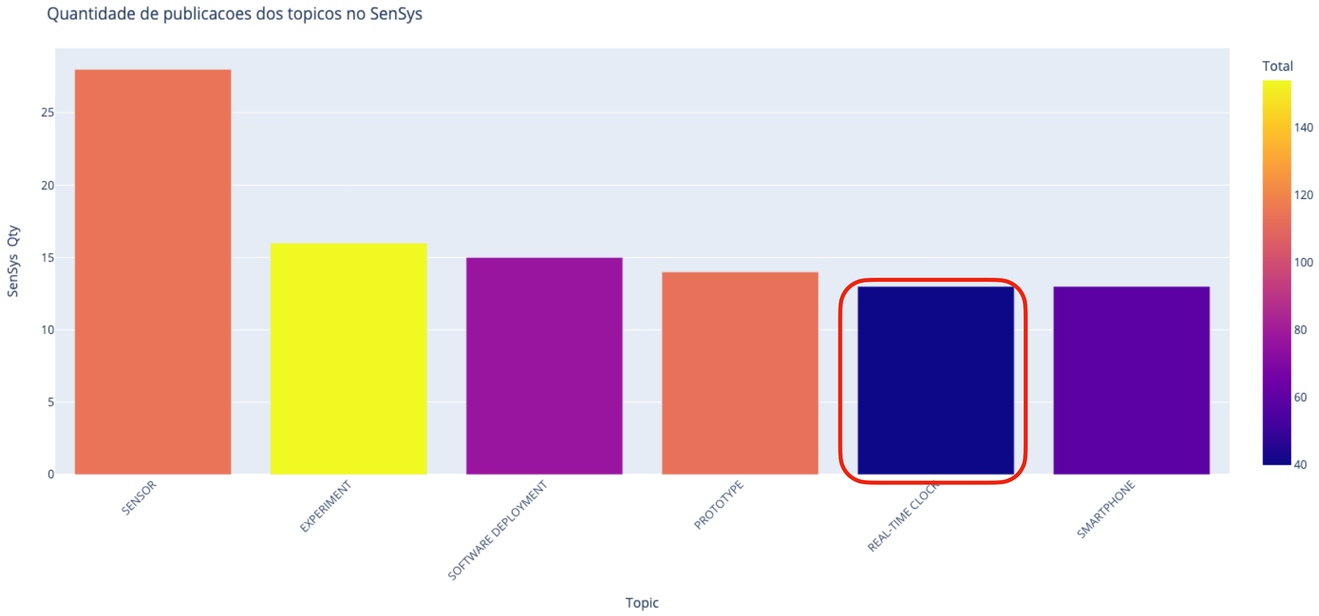


	Topic	Total	SIGCOMM	Qty
0	ALGORITHM	150		21.0
1	NETWORK PACKET	60		18.0
2	THROUGHPUT	77		13.0
3	NETWORK CONGESTION	25		12.0
4	EXPERIMENT	154		11.0
5	SOFTWARE DEPLOYMENT	77		10.0

Análises e resultados

Para cada conferência, qual o tópico que mais aparece?

SenSys

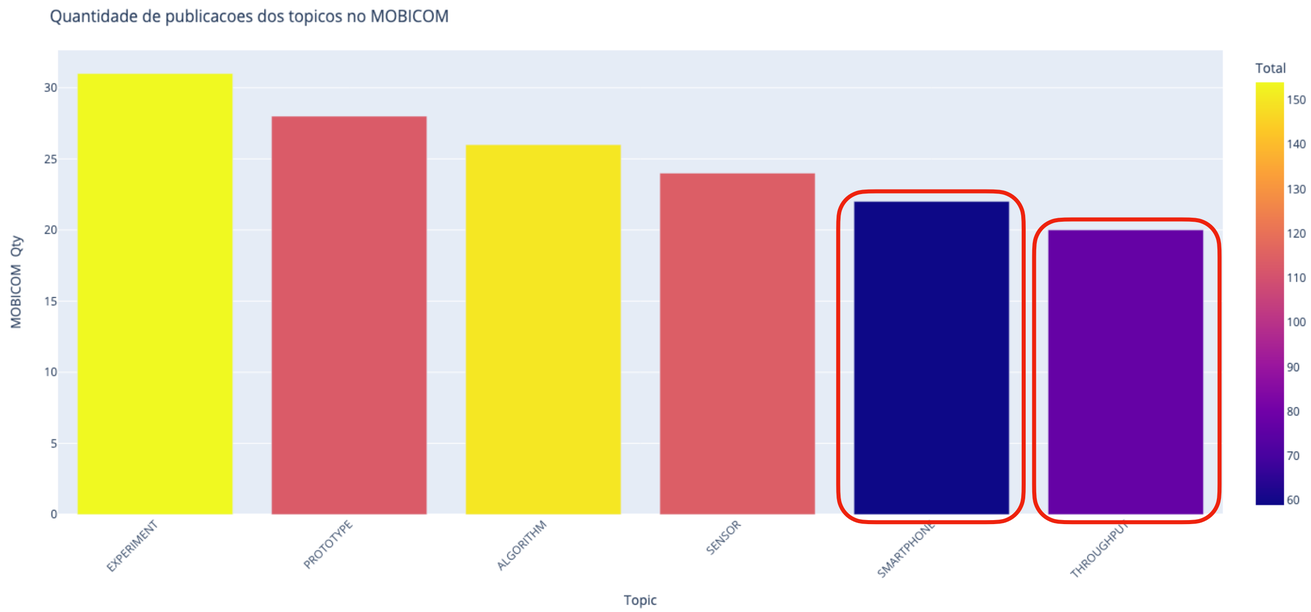


	Topic	Total	SenSys Qty
0	SENSOR	114	28.0
1	EXPERIMENT	154	16.0
2	SOFTWARE DEPLOYMENT	77	15.0
3	PROTOTYPE	113	14.0
4	REAL-TIME CLOCK	40	13.0
5	SMARTPHONE	59	13.0

Análises e resultados

Para cada conferência, qual o tópico que mais aparece?

MOBICOM

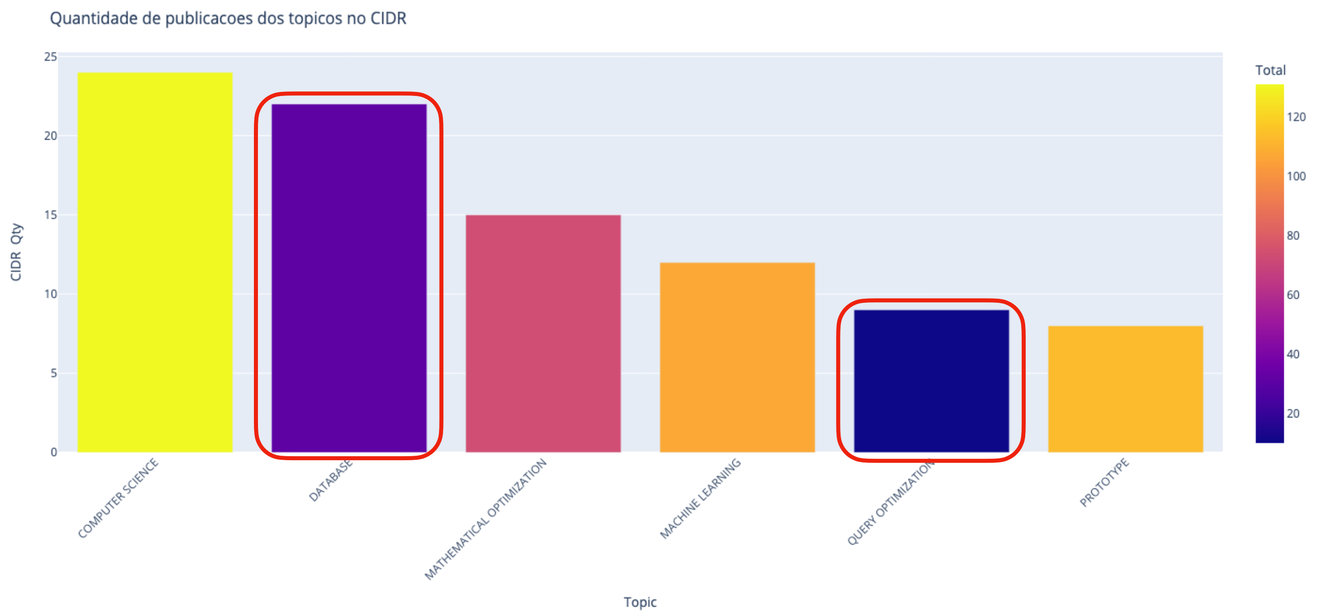


	Topic	Total	MOBICOM	Qty
0	EXPERIMENT	154		31.0
1	PROTOTYPE	113		28.0
2	ALGORITHM	150		26.0
3	SENSOR	114		24.0
4	SMARTPHONE	59		22.0
5	THROUGHPUT	77		20.0

Análises e resultados

Para cada conferência, qual o tópico que mais aparece?

CIDR

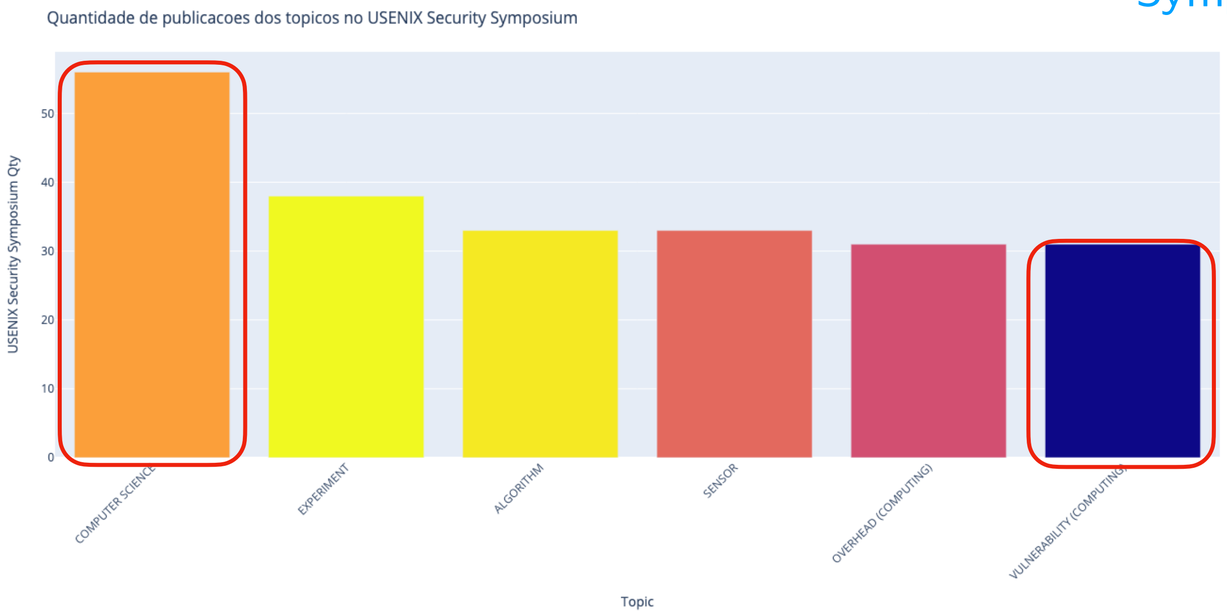


	Topic	Total	CIDR	Qty
0	COMPUTER SCIENCE	131		24.0
1	DATABASE	31		22.0
2	MATHEMATICAL OPTIMIZATION	73		15.0
3	MACHINE LEARNING	107		12.0
4	QUERY OPTIMIZATION	10		9.0
5	PROTOTYPE	113		8.0

Análises e resultados

Para cada conferência, qual o tópico que mais aparece?

USENIX Security Symposium

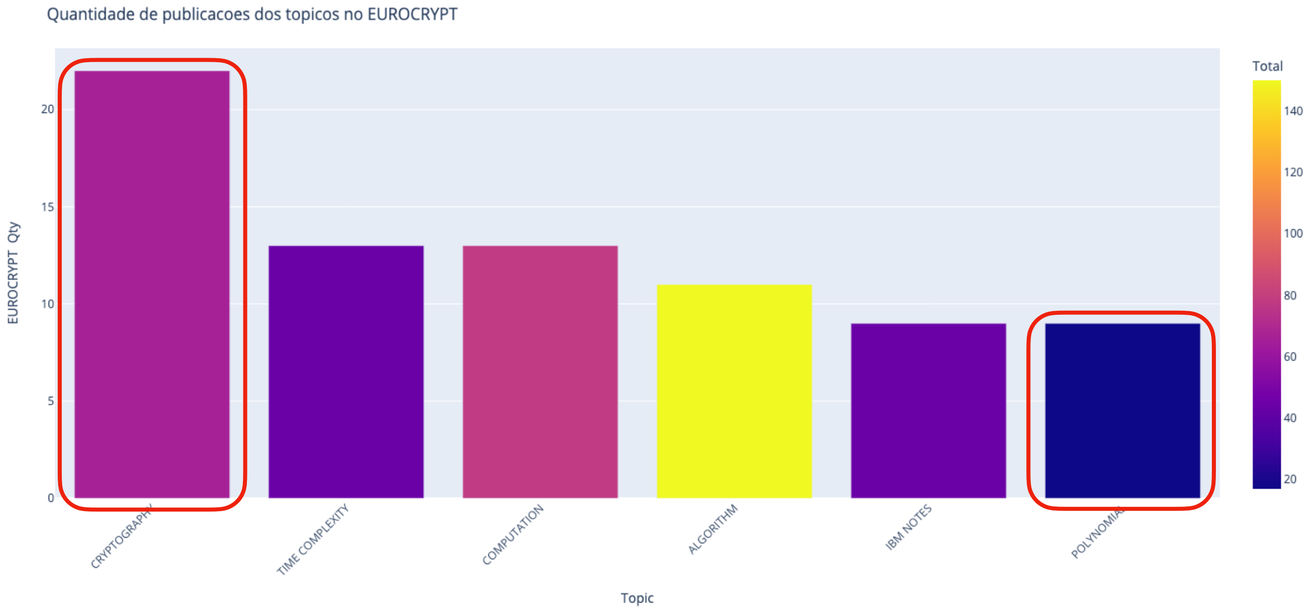


	Topic	Total	USENIX Security Symposium Qty
0	COMPUTER SCIENCE	131	56.0
1	EXPERIMENT	154	38.0
2	ALGORITHM	150	33.0
3	SENSOR	114	33.0
4	OVERHEAD (COMPUTING)	105	31.0
5	VULNERABILITY (COMPUTING)	50	31.0

Análises e resultados

Para cada conferência, qual o tópico que mais aparece?

EUROCRYPT



	Topic	Total	EUROCRYPT	Qty
0	CRYPTOGRAPHY	66		22.0
1	TIME COMPLEXITY	44		13.0
2	COMPUTATION	78		13.0
3	ALGORITHM	150		11.0
4	IBM NOTES	44		9.0
5	POLYNOMIAL	17		9.0

Análises e resultados

Como foram as evoluções dos tópicos publicados por cada conferência ao longo dos anos?

Dicionário de publicações de conferências por ano e tópico

```
{
    'MOBICOM ': {
        2020: {
            'LOCATION (GEOGRAPHY)': 1,
            'DUPLEX (TELECOMMUNICATIONS)': 1,
            'DEFINITION': 1,
            'PHASED ARRAY': 2,
            'EXPERIMENT': 4,
            'FEEDBACK': 1,
            'NETWORK PERFORMANCE': 1,
            'APPROXIMATION ALGORITHM': 1,
            'PROGRAMMING PARADIGM': 1,
            'DOWNTIME': 2,
            'IBM NOTES': 1,
            'BACKSCATTER (EMAIL)': 1,
            ...
        }
    }
}
```

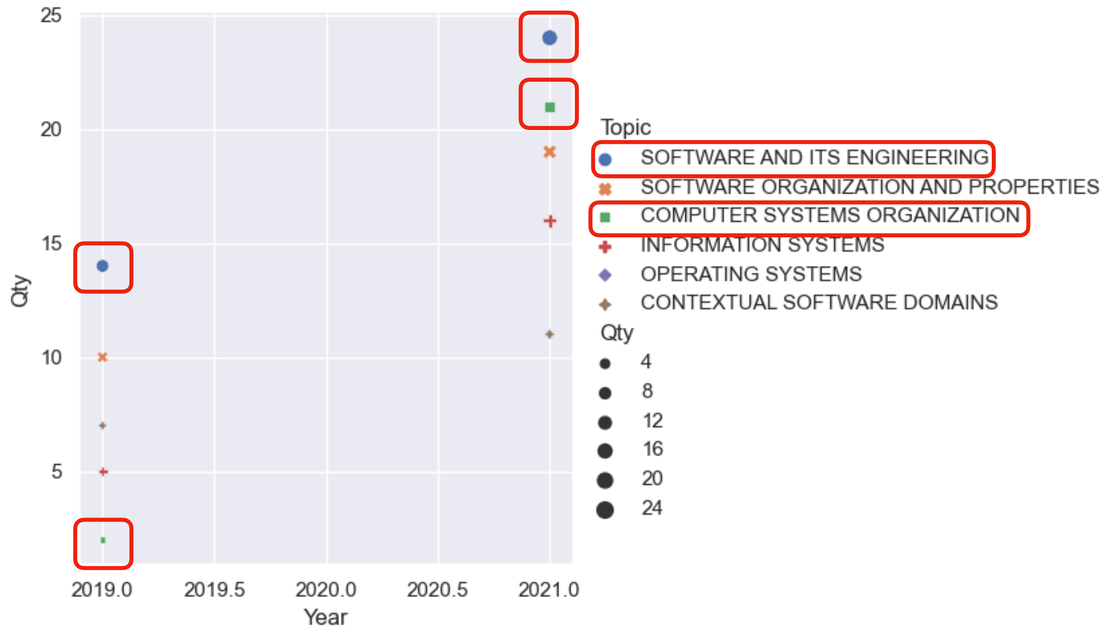
Análises e resultados

Como foram as evoluções dos tópicos publicados por cada conferência ao longo dos anos?

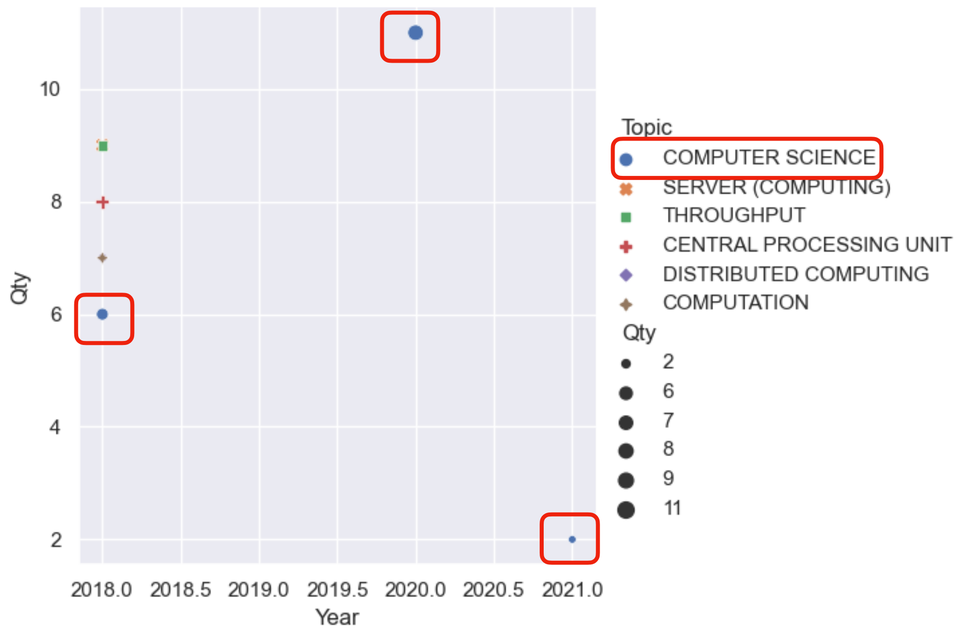
Pega os 6 tópicos com mais publicações para cada conferência e suas quantidades por ano

Análises e resultados

Como foram as evoluções dos tópicos publicados por cada conferência ao longo dos anos?



SOSP

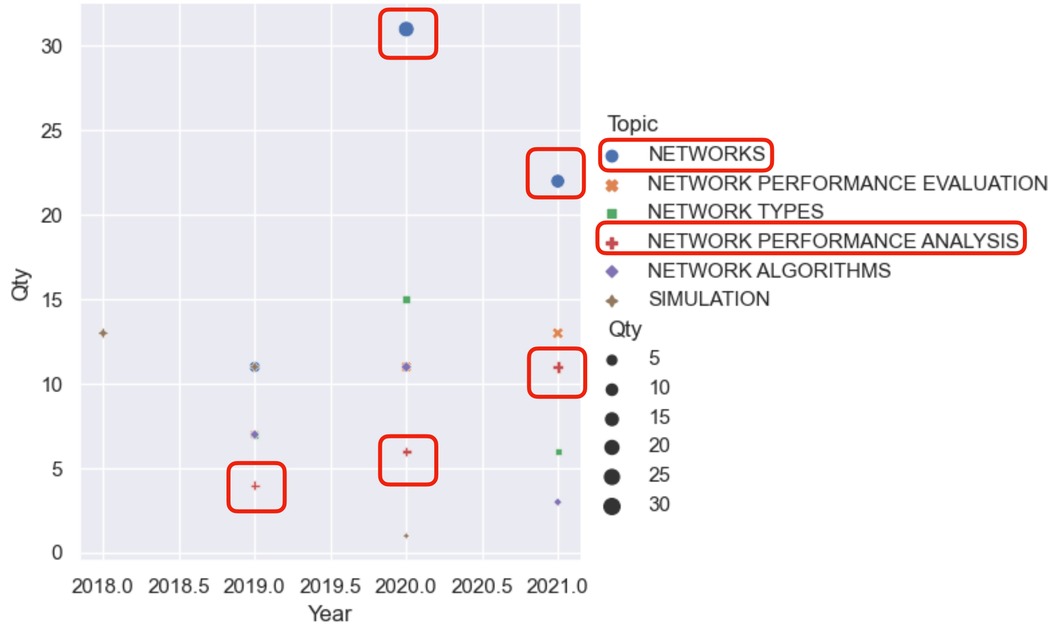


OSDI

Análises e resultados

Como foram as evoluções dos tópicos publicados por cada conferência ao longo dos anos?

NDSS

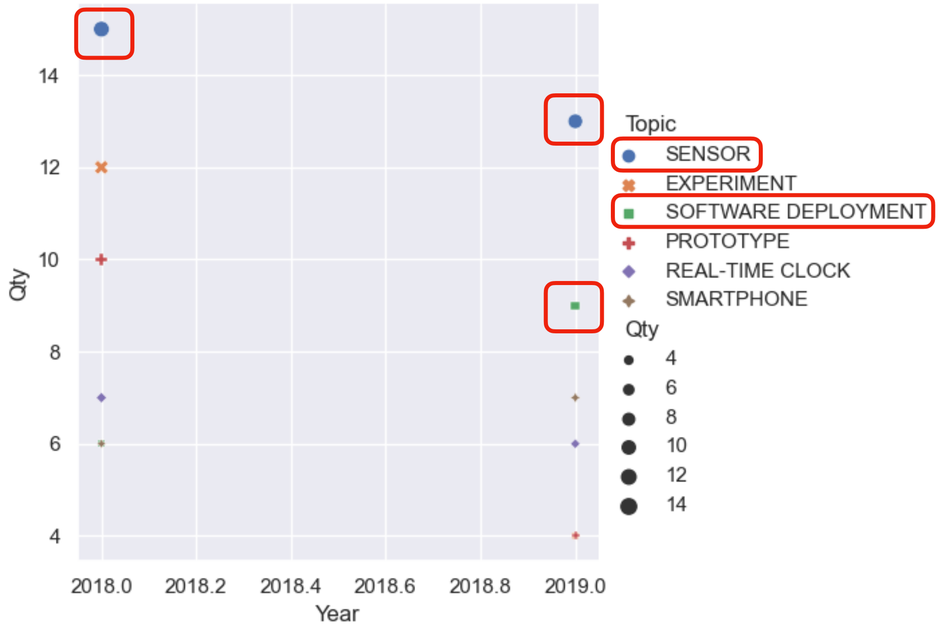


MobiHoc

Análises e resultados

Como foram as evoluções dos tópicos publicados por cada conferência ao longo dos anos?

SIGCOMM

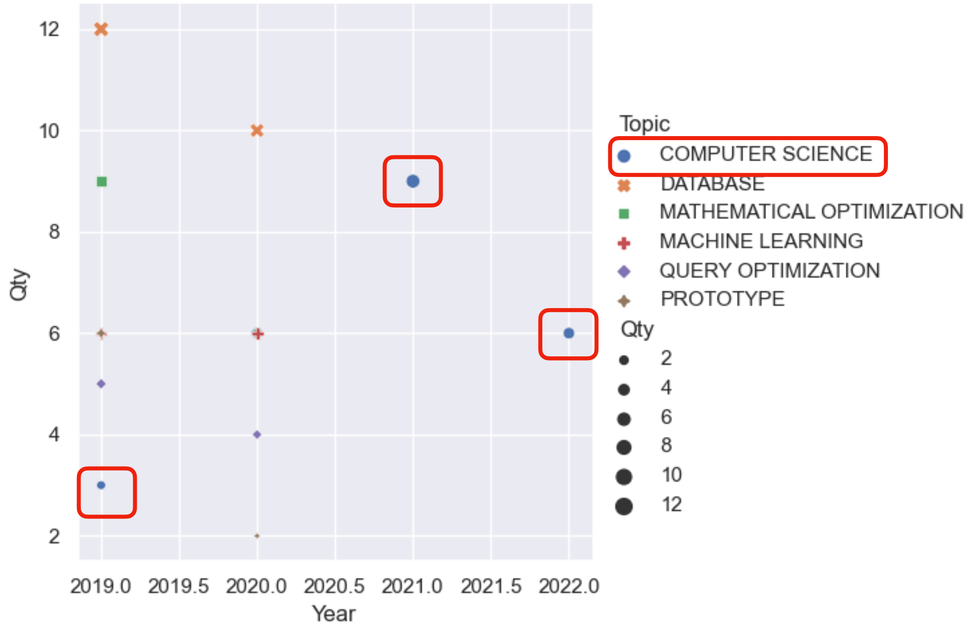


SenSys

Análises e resultados

Como foram as evoluções dos tópicos publicados por cada conferência ao longo dos anos?

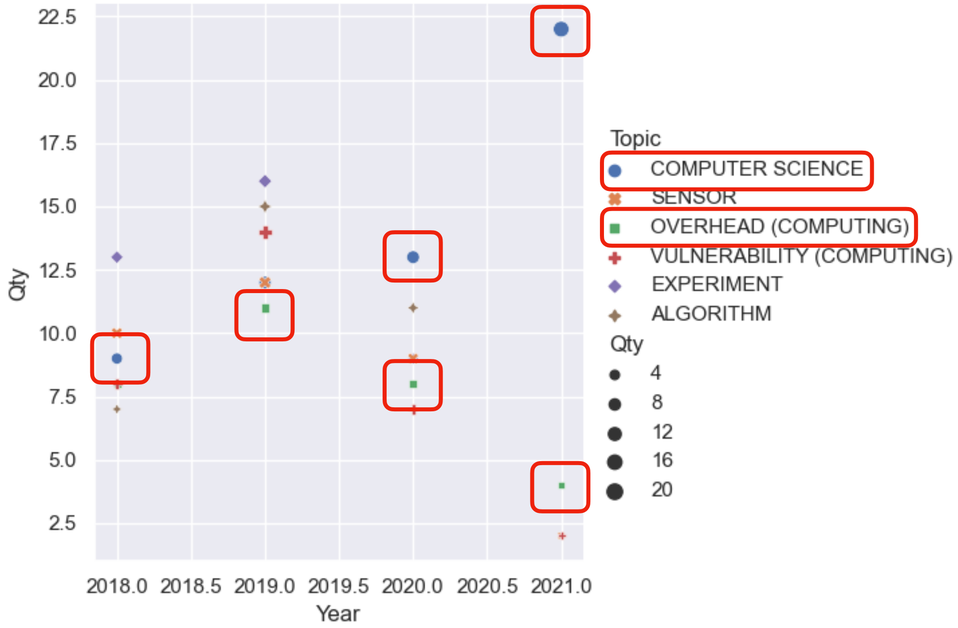
MOBICOM



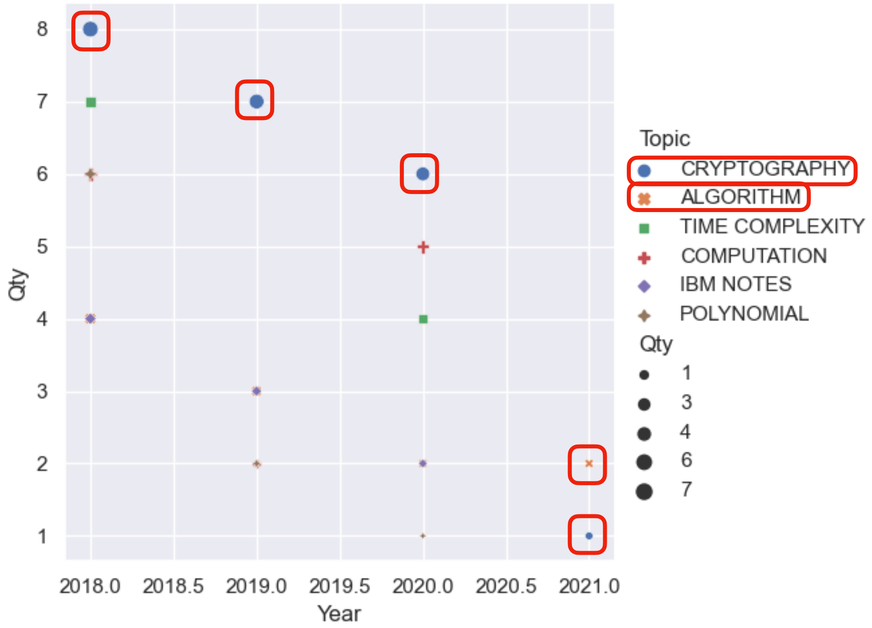
CIDR

Análises e resultados

Como foram as evoluções dos tópicos publicados por cada conferência ao longo dos anos?



USENIX Security Symposium



EUROCRYPT

Problemas

Problemas

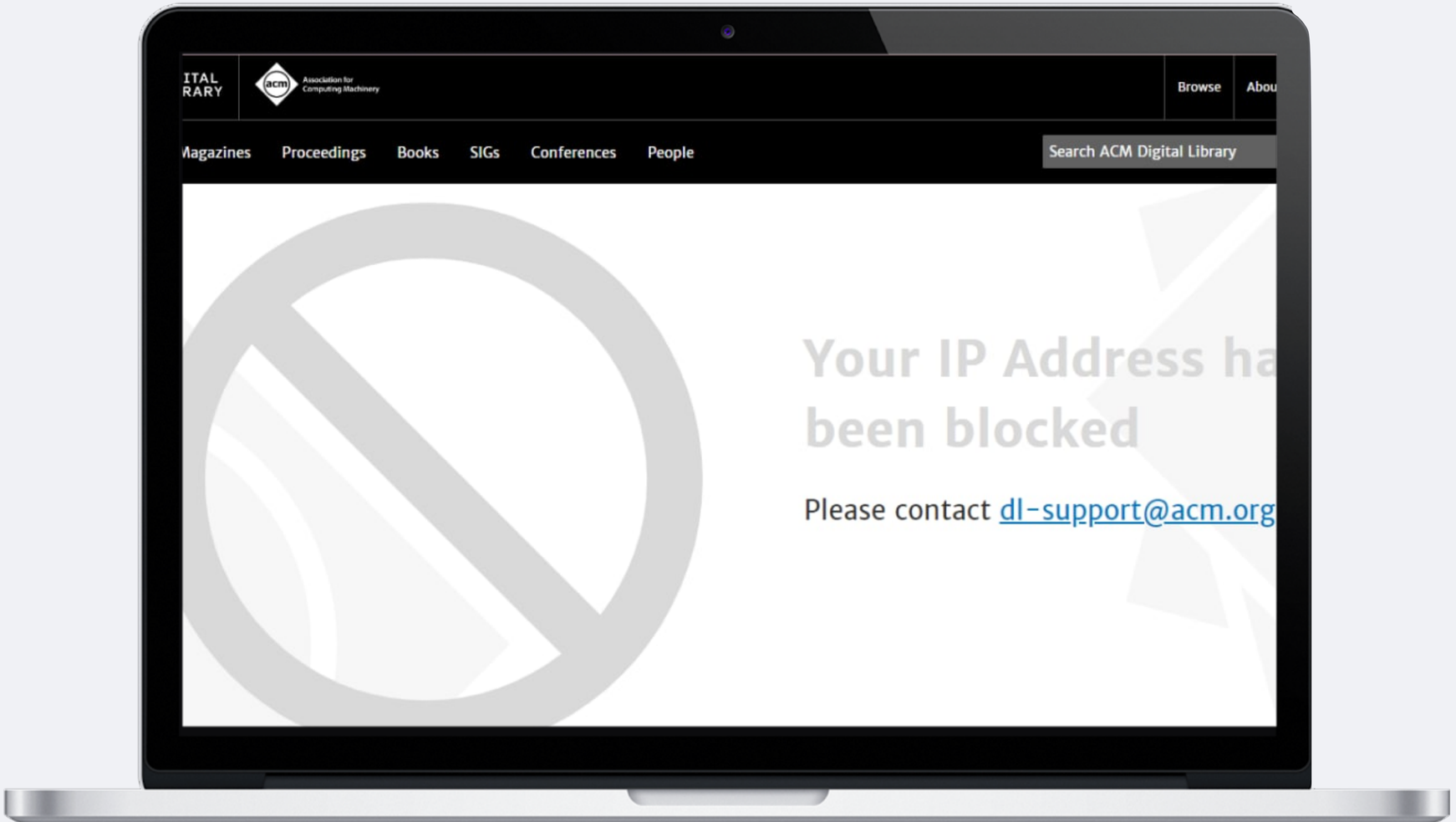
Falta de um único dataset com todos os dados necessários

Limite de requisições na API do Semantic Scholar de **100 requests por 5 minutos**

IP Bloqueado por bastante acesso ao site do DL ACM

Problemas

IP Bloqueado por bastante acesso ao site do DL ACM

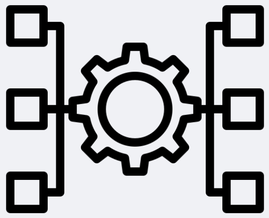


Problemas

Falta de um único dataset com todos os dados necessários

Limite de requisições na API do Semantic Scholar de **100 requests por 5 minutos**

IP Bloqueado por bastante acesso ao site do DL ACM



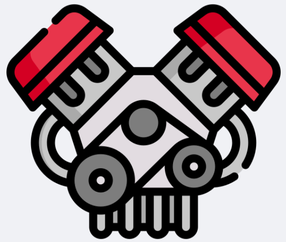
Estruturas distintas dos artigos no site da dblp

Trabalhos Futuros

Trabalhos futuros

Expandir a análise para mais conferências e maiores intervalos de tempo

Expandir a análise para países dos artigos publicados nas conferências e se existe relação entre os artigos publicados em uma conferência



Implementar o sistema de recomendação que utilize a base de dados criada

Encontrar forma de trazer dados completos de todos os artigos de forma automática (uma fonte única)

Referências:

- Cuong, D. V., Nguyen, D. H., Huynh, S., Huynh, P., Gurrin, C., Dao, M.-S., Dang-Nguyen, D.-T., e Nguyen, B. T. (2020). A framework for paper submission recommen-dation system. In Proceedings of the 2020 International Conference on Multimedia Retrieval, ICMR ’20, page 393–396, New York, NY, USA. Association for Computing Machinery.

- GOLDBERG, D. e. a. (1992). Using collaborative filtering to weave an information ta-pestry. 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.

- Wang, D., Liang, Y., Xu, D., Feng, X., e Guan, R. (2018). A content-based recommender system for computer science publications. Knowledge-Based Systems, 157:1–9.

Fundamentos Ciência de Dados

Scientific Recommender: Análise Inicial Sobre Artigos Publicados em Conferências