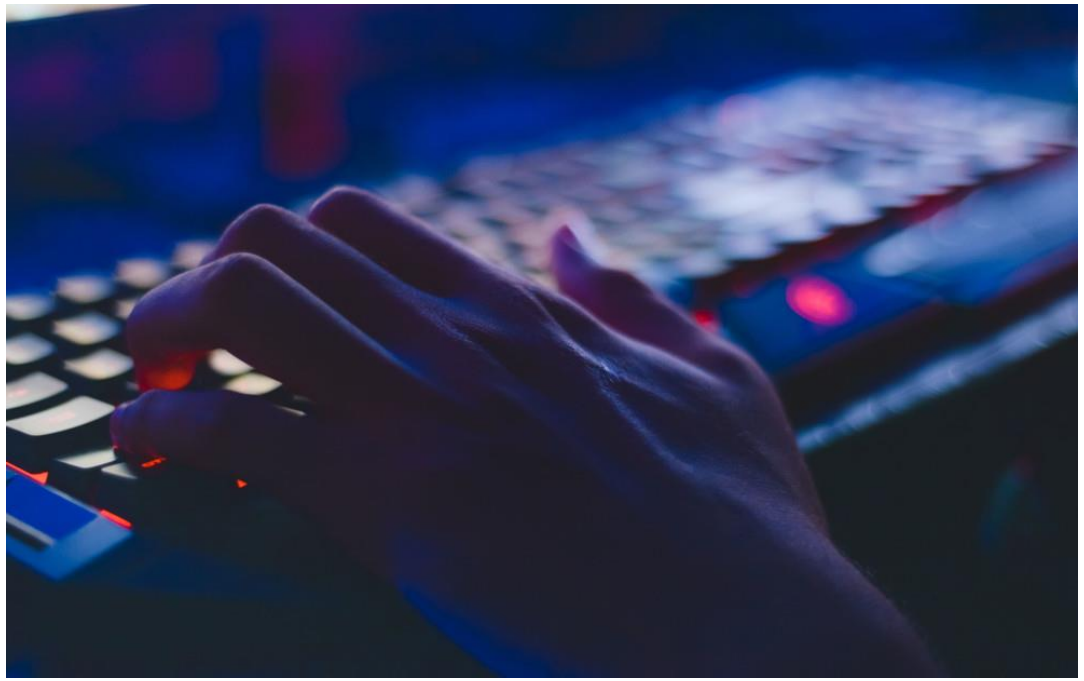
5-2024

Getting published and raising research visibility

Pin Pin Yeo

Getting Published & Raising Research Visibility



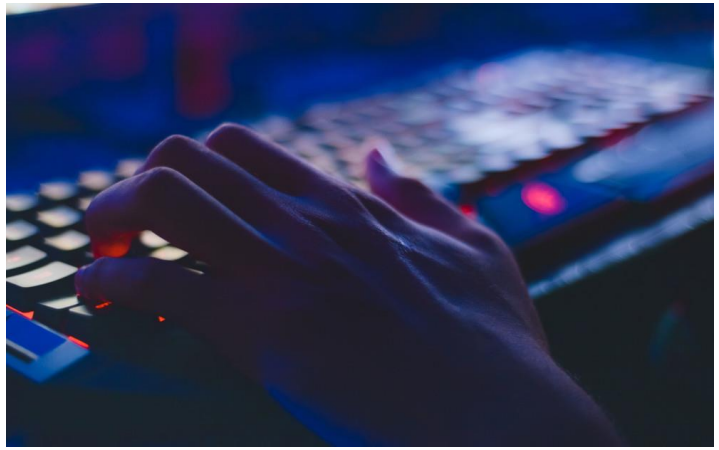
Yeo Pin Pin

Head, Research Services
SMU Libraries

17 May 2024

Presented at Rising Scholars Conference, Singapore, 16-17 May 2024

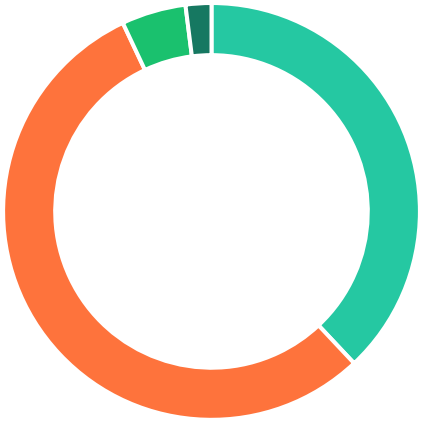
Agenda: Publishing & Research Visibility



- Publishing process
  - Determine if you have original research to write up
  - Writing your paper
  - Peer review process
  - Make your work open access – green open access
- Research visibility
  - Share your work on online platforms
  - Author profiles

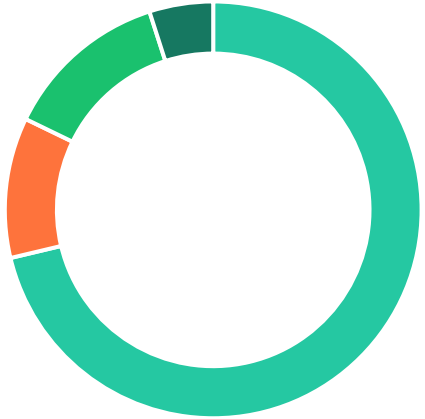
Research Outputs

COMPUTER SCIENCE



Journal article	38%
Conference proceedings article	55%
Book chapter	5%
Book	2%

BUSINESS

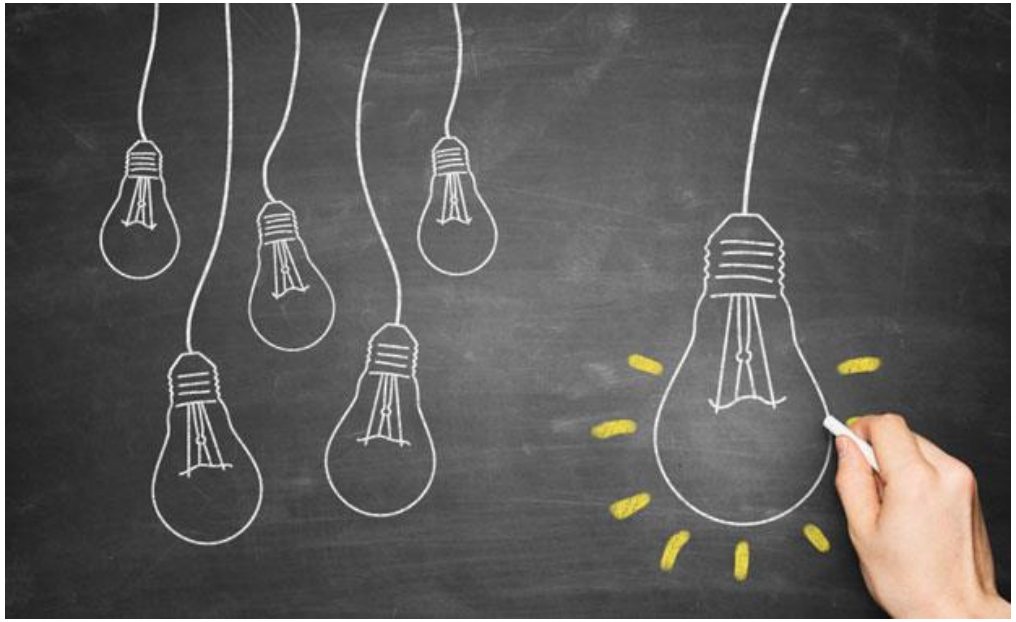


Journal article	71%
Conference proceedings article	11%
Book chapter	13%
Book	5%

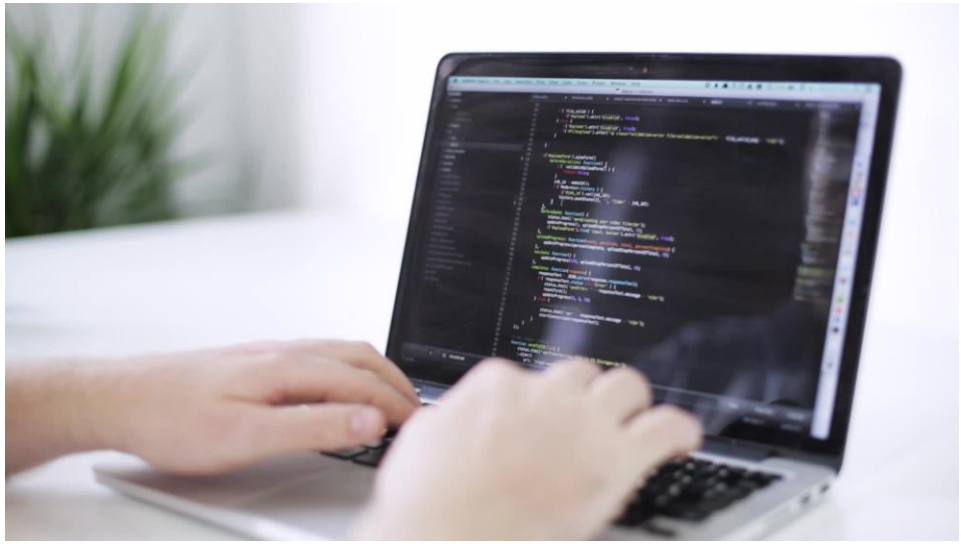
Source: The rhythms of scholarly publication (2021). https://osf.io/cwbuf/

Original research

- Think about your current project / work
- What is new and novel
  - originality of thought
  - unique hypothesis
  - counter-intuitive outcomes / results
- How would your research add value
  - new knowledge to add to the field
- What problem does it solve
- Who would be interested
  - researchers versus practitioner



Writing your paper



- Keep it simple
- Express complicated ideas plainly
- Tell a story that is easy to follow
- Make sure your argument / key ideas are in the first paragraph and abstract
- Extensive literature review
- Figures / tables / charts to illustrate your ideas better

Paper structure (IMRad)

- Introduction
  - Question / problem
  - Thesis / hypothesis / argument
  - Roadmap / outline
- Methods / Theory
- Results
- Discussion
- Conclusion
- *Check if your target journal has a specific format and follow it*

Title
Authors
Abstract
Keywords
Introduction
Methodology
Equations
Results
Discussion
Conclusion
References
Acknowledgments

- https://journals.ieeeauthorcenter.ieee.org/create-your-ieee-journal-article/create-the-text-of-your-article/structure-your-article/

Publishing process

- Have original research work of current interest.

- Identify a journal with aims & scope close to your research work.
- Format manuscript according to the author guideline of targeted journal.

- Submit your manuscript online by providing complete detail in online submission system.
- Do not submit manuscript to more than one journal at the same time.

- Editor will initiate peer-review process for well formatted manuscript within the scope of journal
- Keep patience and wait for a reply from editor during review process.

Based on reviewer's recommendation, editor will send you one of the following decision letters-

- **Revision Required:** Revise manuscript as per comment and resubmit.
- **Decline:** Look at your manuscript and decline reason. Spend plenty of time to improve your overall manuscript and resubmit as a fresh manuscript.

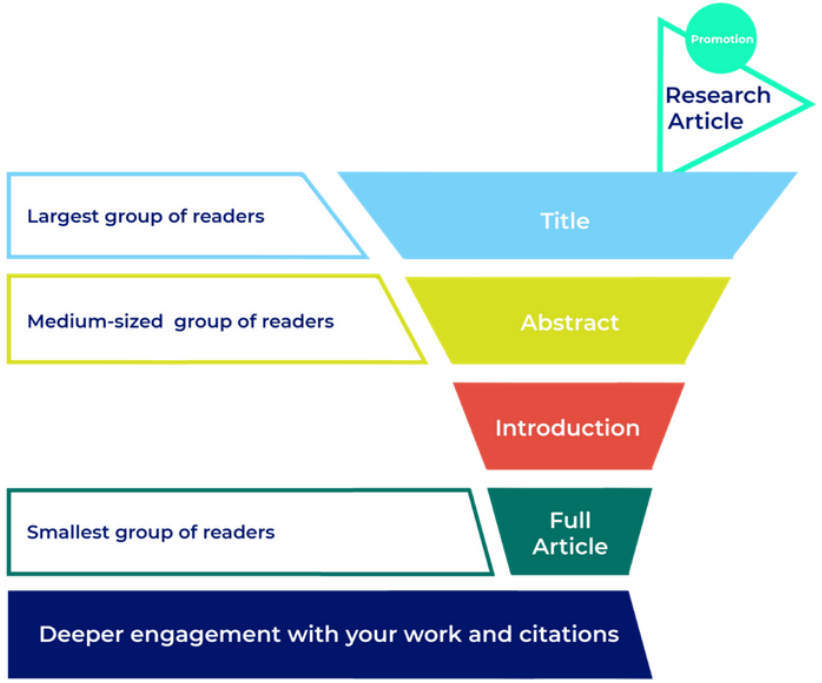
- **Accepted:** Its time to celebrate, your manuscript will be copyedited by the journal for final publication.

- You will receive galley proof version for minor proofreading corrections and your paper will be published.
- It's time to share your published work and cite in other related papers.

Source: https://aijr.org/paper-publishing-process/

Title

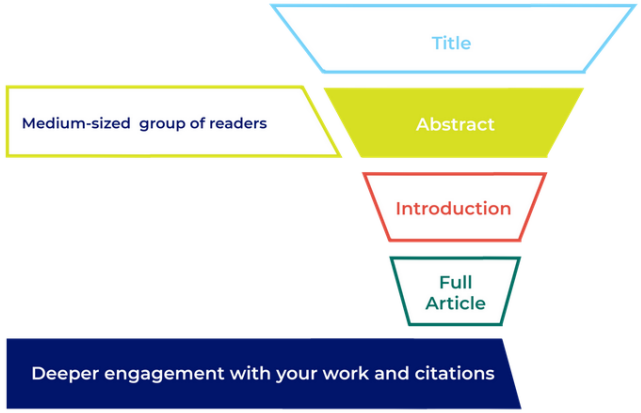
- Keep it short and to the point (< 16 words)
- Include key information about your paper
- What you discovered
- Find a way to pique your readers' interest, give them enough information to keep them reading
- Avoid cliché phrases, if using common phrase, have a more informative subtitle



Example:
Out of sight, out of mind? How vulnerable dependencies affect open-source projects

https://plos.org/resource/how-to-write-a-great-title/

Abstract



- Intended audience and frame it for them
  - Industry versus academics/researchers
- Abstracts should include the key points of research
- Include background, objective/aim, methods, results, conclusion
- Focus on important information – key results, conclusions
- Is it clear
- Is it easy to understand

General audience versus researcher

big trends

DOI:10.1145/3378422

BY MEEYOUNG CHA /KAIST, WEI GAO /SINGAPORE MANAGEMENT UNIV., AND CHENG-TE LI /NATIONAL CHENG KUNG UNIV.

Detecting Fake News in Social Media: An Asia-Pacific Perspective

IN MARCH 2011, the catastrophic accident known as "The Fukushima Daiichi nuclear disaster" took place, initiated by the Tohoku earthquake and tsunami in Japan. The only nuclear accident to receive a Level-7 classification on the International Nuclear Event Scale since the Chernobyl nuclear power plant disaster in 1986, the Fukushima event triggered global concerns and rumors regarding radiation leaks. Among the false rumors was an image, which had been described as a map of radioactive discharge emanating into the Pacific Ocean, as illustrated in the accompanying figure. In fact, this figure, depicting the wave height of the tsunami that followed, still to this date circulates on social media with the inaccurate description.

Social media is ideal for spreading rumors, because it lacks censorship. Confirmation bias and filter-bubble effects further amplify the spread of unconfirmed information. Upon public outcry, independent fact-checking organizations have emerged globally, and many platforms are making efforts to fight against fake news. For example, the state-run *Factually* website in Singapore has been known to clarify falsehoods since its inception in May 2012, which was followed recently by the implementation of the Protection from Online Falsehoods and Manipulation Act (POFMA) in October 2019. In Taiwan, the government officially created a feature on the website of the Executive Yuan (the executive branch of Taiwan's government) to identify erroneous reporting and combat the spread of fake news. Taiwan's Open Culture Foundation has also developed and introduced the well-known anti-fake fact-checking chatbot *Cofacts* in May 2018. The Indonesia government since 2018 has held weekly briefings on hoax news; that same year, the country revised its Criminal Code to permit the imprisonment for up to six years of anyone spreading fake news. Governments in the Asia and Oceania region, including South Korea, Singapore, Japan, Taiwan, Philippines, Cambodia, Malaysia, have enacted relevant laws to prevent fake news from spreading.

Nonetheless, fact-checking of fake news remains daunting, and requires tremendous time and effort in terms of human investigation. Moreover, it is prone to low efficiency and inadequate coverage due to the complexity of the topics being checked, and is incapable of keeping up with the fast

2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining

Evaluating Vulnerability to Fake News in Social Networks: A Community Health Assessment Model

Bhavtosh Rath[1], Wei Gao[2], and Jaideep Srivastava[1]
[1]Dept. of Computer Science & Engineering, University of Minnesota, Twin Cities, MN, USA
[2]School of Information Management, Victoria University of Wellington, Wellington, New Zealand
rathx082@umn.edu, wei.gao@vuw.ac.nz, srivasta@umn.edu

***Abstract*—Understanding the spread of false information in social networks has gained a lot of recent attention. In this paper, we explore the role community structures play in determining how people get exposed to fake news. Inspired by approaches in epidemiology, we propose a novel *Community Health Assessment* model, whose goal is to understand the vulnerability of communities to fake news spread. We define the concepts of neighbor, boundary and core nodes of a community and propose appropriate metrics to quantify the vulnerability of nodes (individual-level) and communities (group-level) to spreading fake news. We evaluate our model on communities identified using three popular community detection algorithms for twelve real-world news spreading networks collected from Twitter. Experimental results show that the proposed metrics perform significantly better on the fake news spreading networks than on the true news, indicating that our community health assessment model is effective.**

I. Introduction

Use of social media platforms like Facebook and Twitter is ubiquitous in modern times, making them powerful platforms for news propagation and consumption. However, the good inevitably is accompanied by the bad, which can be witnessed by the problem of *fake news spreading* [1]. It spreads when someone propagates it via endorsements such as replying, sharing or re-posting, without validating its authenticity. There is significant interest in understanding the nature of fake news spreading. Our focus is on *assessing the vulnerability of social networks to fake news spreading*. Specifically, we focus on people and the communities they create, with the goal of identifying how vulnerable individuals and communities are to believing and propagating fake news. We propose the Community Health Assessment model that distinguishes between neighbor, boundary and core nodes of a community, and propose novel metrics to quantify the vulnerability of an individual node, as well as the community, to external exposure. We propose methods to estimate the likelihood of a boundary node of a community to believe fake news sent from its immediate neighbors; and also estimate the likelihood of a community's entire boundary node set to believe fake news coming from its neighborhood. It is important to note that the method used to quantify vulnerability of a boundary node can be generalized to any node. Intuitively, if an external node infects a member of a community, the likelihood of the entire community getting infected increases due to high connectivity and trust among its members. Thus, while assessing vulnerability of community, we focus on examining the influence of news propagated from external nodes into the community rather than considering the propagation of the news within the community. We evaluate our model on the propagation networks of twelve real-world news collected from snopes.com[1].

Our contributions are summarized as follows:

- We propose the Community Health Assessment model that introduces the ideas of neighbor, boundary and core nodes for a community structure in a social network.
- We propose metrics to quantify the vulnerability of a node and a community to fake news exposure from outside.
- Using Twitter news item spreading network (a subgraph of Twitter network induced by the news item, *news network* in short) we demonstrate that our proposed metrics can assess the vulnerability of social networks to fake news better than for true news.

II. Related Work

There has been a recent surge in interest among researchers and practitioners to develop approaches to prevent fake news spread. Most approaches in the literature use content-based [2], [3] and propagation-based characteristics [4], [5]. Approaches using neural networks [6], [7] have also shown promising results. Infection spread models inspired from epidemiology [8], [9] have also been used to model rumor spreading. Other models have tried to identify the rumor spreading source [10], [11]. A community perspective to rumor spread has also been attempted. Fan et al. [12] proposed an approach to identify a minimal set of boundary nodes that would prevent spread of rumors from neighboring communities. Nguyen et al. [13] proposed a community-based heuristic method to find the smallest set of highly influential nodes whose decontamination with good information would contain rumor spreading. Vosoughi

[1]https://www.snopes.com/fact-check-ratings/

ASONAM '19, August 27-30, 2019, Vancouver, Canada

ACM ISBN 978-1-4503-6868-1/19/08?$15.00
http://dx.doi.org/10.1145/3341161.3342920

General audience versus researcher - 2

Detecting fake news in social media - Abstract: In March 2011, the catastrophic accident known as "The Fukushima Daiichi nuclear disaster" took place, initiated by the Tohoku earthquake and tsunami in Japan. The only nuclear accident to receive a Level-7 classification on the International Nuclear Event Scale since the Chernobyl nuclear power plant disaster in 1986, the Fukushima event triggered global concerns and rumors regarding radiation leaks. Among the false rumors was an image, which had been described as a map of radioactive discharge emanating into the Pacific Ocean, as illustrated in the accompanying figure. In fact, this figure, depicting the wave height of the tsunami that followed, still to this date circulates on social media with the inaccurate description. Social media is ideal for spreading rumors, because it lacks censorship. Confirmation bias and filter-bubble effects further amplify the spread of unconfirmed information. Upon public outcry, independent fact-checking organizations have emerged globally, and many platforms are making efforts to fight against fake news.

Evaluating vulnerability to fake news in social networks - Abstract:
Understanding the spread of false information in social networks has gained a lot of recent attention. In this paper, we explore the role community structures play in determining how people get exposed to fake news. Inspired by approaches in epidemiology, we propose a novel *Community Health Assessment* model, whose goal is to understand the vulnerability of communities to fake news spread. We define the concepts of neighbor, boundary and core nodes of a community and propose appropriate metrics to quantify the vulnerability of nodes (individual-level) and communities (group-level) to spreading fake news. We evaluate our model on communities identified using three popular community detection algorithms for twelve real-world news spreading networks collected from Twitter. Experimental results show that the proposed metrics perform significantly better on the fake news spreading networks than on the true news, indicating that our community health assessment model is effective.

General audience versus researcher – 3

Detecting fake news in social media – Conclusion: The fake news phenomenon is taking new turns. YouTube and instant messaging (IM) services (for example, Whatsapp, Kakaotalk) are emerging as hotbeds of fake news. According to a survey conducted by the Korea Press Foundation, 34% of Korean YouTube viewers report having watched or received videos containing fake news. Taiwan's Open Government Foundation g0v reported that in 2017 only 46% of chatbot responses on that nation's most popular IM app LINE is correct. Fake news on streaming platforms and IM services is particularly concerning because it contains visual content, which is more persuasive than mere text posts. Also, IM may reinforce the credibility of fake claims because people are more likely to follow trusted social contacts blindly. … Consequently, the task of detecting fake news in the era of big data, social media, and artificial intelligence calls for greater attention from the research community.

Evaluating vulnerability to fake news in social networks – Conclusion: We propose novel metrics based on the concept of believability derived from computational trust measures to compute vulnerability of nodes and communities to news spread and show that the metrics are much more sensitive to fake news than true news. This confirms our hypothesis that fake news have to rely on strong trust among spreaders to propagate while true news do not. Through experiments on large news spread networks on Twitter we show that our proposed metrics can identify the vulnerable nodes for false news networks with higher precision.

Keywords

- Flags in searches by end users
- Think about keywords and phrases that describe your work
- Add some synonyms
- Use phrases when appropriate
- Test in Google / Google Scholar – does it bring back the results you expect

Keywords Example 1

- Higher well-being individuals are more receptive to cultivated meat: An investigation of their reasoning for consuming cultivated meat
- **Keywords:** Cultivated meat, Psychological well-being, Consumer acceptance, Climate change
- Abstract: It is evident that over-consumption of meat can contribute to the emission of hazardous greenhouse gases. One viable way to address such climate impact is to make people become more aware of more sustainable diet options, such as cultivated meat. However, it is challenging to instigate change in people's meat-eating habit, and empirical works have been examining the psychological factors that are related to consumers' willingness to consume cultivated meat. Research has suggested that psychological well-being can play a role in the meaning-making of food consumption, with higher well-being individuals showing more recognition of other sociocultural benefits of consuming food beyond just fulfilling their sustenance needs. As existing works have yet to understand the link between well-being and consumption of novel foods, the current research set out to fill this gap by examining the relationship between people's psychological well-being and their willingness to consume cultivated meat via different reasons (mediators) for consuming cultivated meat ...

Keywords Example 2

Information trust and COVID-19 vaccine hesitancy amongst middle-aged and older adults in Singapore: A latent class analysis approach

Keywords: Social media use, health literacy, support, communications, attitudes, barriers

Abstract: Rationale: COVID-19 vaccine hesitancy presents significant challenges for public health. Objective: Vaccine hesitancy among middle-aged and older adults has been a significant barrier in Singapore's battle against COVID-19. We hypothesize that the trust middle-aged and older adults place in various sources of information influences vaccine hesitancy, and that distinct typologies of trust can be identified to better inform targeted health communication efforts. Method: Data from a nationally representative panel survey of Singaporeans aged 56-75 (N =6094) was utilized. Modules fielded in August and November 2020, and June 2021 were analyzed, assessing social networks, trust in sources of information, and vaccination status respectively. Predictors of vaccination status were first examined. Latent class analysis was then used to identify typologies of trust in various sources of information. Results: Trust in formal sources of information (e.g government sources) is found to predict vaccination status among respondents. Contrary to expectations, trust in social media and informal sources (family and friends), and perceived social support did not predict vaccination status. Latent class analysis identified 4 typologies of respondents based on their patterns of trust in these sources. Significantly, it is found that a portion of respondents with low trust in formal sources of information have high trust in informal sources...

Keywords Example 3

Security and privacy in smart health: Efficient policy-hiding attribute-based access control

Keywords: Attribute-based encryption (ABE), decryption test, full security, large universe; privacy protection, smart health (s-health)

Abstract: With the rapid development of the Internet of Things and cloud computing technologies, smart health (s-health) is expected to significantly improve the quality of health care. However, data security and user privacy concerns in s-health have not been adequately addressed. As a well-received solution to realize fine-grained access control, ciphertext-policy attributebased encryption (CP-ABE) has the potential to ensure data security in s-health. Nevertheless, direct adoption of the traditional CP-ABE in s-health suffers two flaws. For one thing, access policies are in cleartext form and reveal sensitive health-related information in the encrypted s-health records (SHRs). For another, it usually supports small attribute universe, which places an undesirable limitation on practical deployments of CP-ABE because the size of its public parameters grows linearly with the size of the universe. To address these problems, we introduce PASH, a privacy-aware s-health access control system, in which the key ingredient is a large universe CP-ABE with access policies partially hidden. ...

Which journal / conference ?

- Look at your References, see if there are possible venues
- If you are citing them in your paper, they are likely to focus on same areas
- Check Journal Citation Report for impact factors, CiteScore in Scopus, as they identify journals with high average citations
- Ask your supervisor for advice

Conference recommender

- Scimago
  https://www.scimagojr.com/journalrank.php?type=p&order=sjr&ord=desc
- Guide2Research
  https://www.guide2research.com/conferences/
- CORE Conference Portal
  http://portal.core.edu.au/conf-ranks/

Think. Check. Attend https://thinkcheckattend.org/

Journal recommenders

- IEEE Publication Recommender https://publication-recommender.ieee.org/periodicals
- Elsevier Journal Finder https://journalfinder.elsevier.com/
- Springer Journal Suggestor https://journalsuggester.springer.com/
- Master Journal List (Clarivate) https://mjl.clarivate.com/home
- JANE: Journal/Author Name Estimator https://jane.biosemantics.org/
- JournalGuide https://www.journalguide.com/

Journal requirements

- Is there a **template**
- Are there page limits
- Which style to use for references
- Check the journal website
- Look at the articles published in the journal

What is peer review?

- Peer review is designed to assess the validity, quality and often the originality of articles for publication
- Purpose is to maintain the integrity of science by filtering out invalid or poor quality articles
- Running articles through the process of peer review adds value to them and improves the article
- Not perfect – some articles still contain inaccuracies – after all peer review is a human activity
- Video: https://www.wiley.com/network/researchers/peer-review-week-2020/wiley-peer-review-part-1

Peer review process

https://authorservices.wiley.com/Reviewers/journal-reviewers/what-is-peer-review/the-peer-review-process.html

Types of peer review

Single blind	Author does not know the identity of the reviewer
Double blind	Reviewer does not know the identity of the author, and vice-versa
Open peer review	The identity of the author and the reviewer is known by all participants, during or after the review process – some reviewers even make their comments open to others

Outcomes of peer review

- Accept
- Accept after minor revisions
- Accept after major revisions
- Reject

Questions reviewers ask

- Does the paper fit the standards and scope of the journal it is being considered for?
- Is the research question clear?
- Was the approach appropriate?
- Is the study design, methods and analysis appropriate to the question being studied?
- Is the study innovative or original?
- Does the study challenge existing paradigms or add to existing knowledge? Does it develop novel concepts?
- Are the methods described clearly enough for other researchers to replicate?
- Are the methods of statistical analysis and level of significance appropriate?
- Could presentation of the results be improved and do they answer the question?
- Are the conclusions appropriate?
- If humans, human tissues or animals are involved, was ethics approval gained and was the study ethical?

Source: Peer review: The nuts and bolts. https://senseaboutscience.org/wp-content/uploads/2016/09/peer-review-the-nuts-and-bolts.pdf

If your paper is rejected

- View the rejection as a form of feedback
- Take the time and the space to improve your work
- Check if feedback from the reviewers were included and learn from it
- Pull apart your paper and re-write your paper
- Consider submitting to another journal

- Source: Leglu, C. (2022, May). What to do when an academic journal rejects your article. *The Times Higher Education*. URL: https://www.timeshighereducation.com/campus/what-do-when-academic-journal-rejects-your-article

Responding to reviewer comments

- Read the comments carefully
- Make a list of all the comments with a column to indicate how it was corrected
- Tackle the minor ones first (e.g. spelling mistakes, grammatical errors, inconsistencies)
- Next tackle the major ones – answer politely, completely and with evidence

Open peer review

Found on WoS for Sensors https://www.webofscience.com/wos/woscc/full-record/WOS:000603275500001 click on Open Peer Reviews

PRE-PUBLICATION REVIEW **NOV 2020**

The authors propose a model-based indoor positioning system (IPS), that employs a small amount of data-sampling from a large-scale scenario, using dynamic parameters (obtained from the data-sampling), and an automatic selection of access points.

The article is clearly written, with some minor issues that need to be acknowledged:

- The error of the proposed system is around 3 m., however, the authors mentioned in Section 2 (Related Work) that the classic RADAR system "resulted in an accuracy of 2 to 3m." The authors need to clarify why is this system not evaluated, and to contextualize the reader of the error of the proposed system, which seemingly is not outperforming RADAR.
- The paragraph in lines 308-314 should be located at the end of Section 3.1, since it provides important details of data gathering.
- Practicality seems to be an important contribution of the proposed system, since, as the authors state that they "only needed to collect some samples from 15 different locations to generate our model. This can be done in fewer than 15 minutes". While this reviewer does not doubt that this process only took 15 minutes, it seems misleading to state that samples were collected from only from 15 different locations, since in Section 4.1 (lines 299-300) the authors state that they "gathered 100 packet samples from evenly-spaced, 2m apart locations, to a total of 150 different test points". This should be clarified
- Small typo in line 362: "we decided to use the Δ value has a filter" -> "we decided to use the Δ value as a filter"

INCREMENTAL

8.5

DETAILS

Tweet

PUBLISHED IN

Sensors

REVIEWED BY

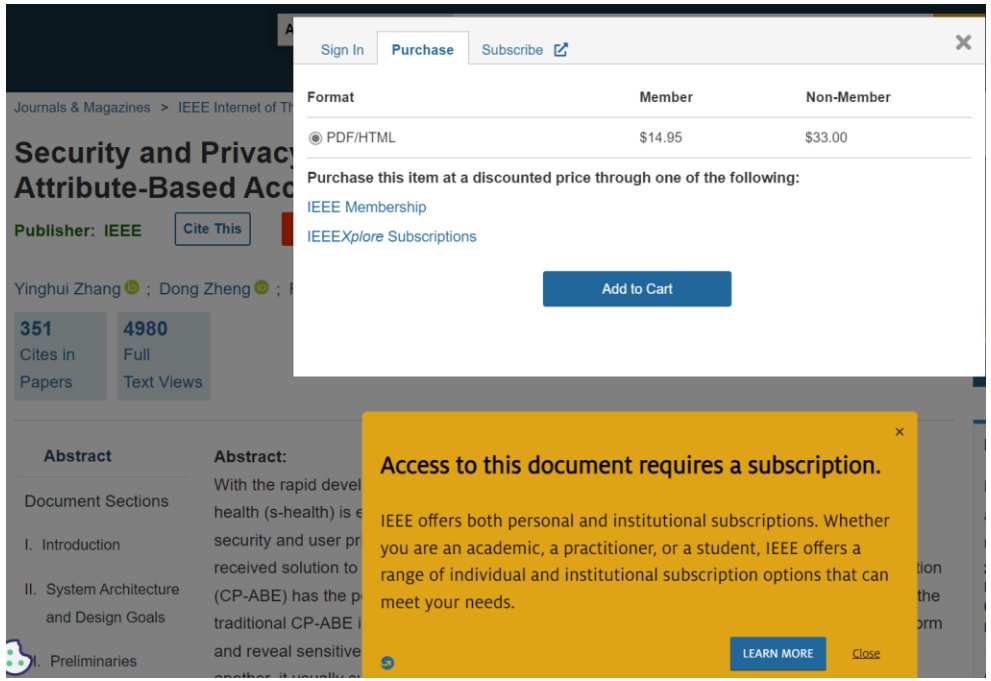
Caleb Rascon

After your paper is accepted

- Review the proofs of your paper in the journal format
- Read carefully before signing copyright license
  - Aim to transfer copyright only for the published version
  - Aim to retain rights to pre-print and author-final version

Your paper on the publisher website

- Most journals must be subscribed
- Without a subscription, you have no access to the paper



Benefits of Open Access

- Open access (OA) makes scholarly research literature freely available to all, to read, to reuse with open license for copyright with acknowledgement
- OA papers are viewed, downloaded and read by more people
- OA papers are built upon by others, and cited more by other researchers

Open access publishing

- **Green open access** – free by making an earlier version of the article available on an open access repository e.g. InK, arXIV.org, SSRN – author-final version (post-print), submitted version (pre-print) – some publishers impose embargo period 12-24 months

Privacy preserving search services against online attack

Yi Zhao[a,e], Jianting Ning[b,c], Kaitai Liang[f], Yanqi Zhao[d,e], Liqun Chen[g], Bo Yang[a,e,*]

a School of Information Engineering, Chang'an University, Xi'an 710064, China
b School of Computer Science, Shaanxi Normal University, Xi'an 710062, China
c State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 10 0093, China
d Department of Computer Science, Fujian Normal University, China
e Department of Computer Science, Singapore Management University, Singapore
f Department of Computer Science, University of Surrey, UK

Published in Computers and Security, August 2020, 95, 101836, pp. 1-10. DOI: 10.1016/j.cose.2020.101836

Abstract: Searchable functionality is provided in many online services such as mail services or outsourced data storage. To protect users privacy, data in these services is usually stored after being encrypted using searchable encryption. This enables the data user to securely search over encrypted data from a remote server without leaking data and query information. Public key encryption with keyword search is one of the research branches of searchable encryption; this provides privacy-preserving searchable functionality for applications such as encrypted email systems. However, it has an inherent vulnerability in that the information of a query may be leaked using a keyword guessing attack. Most of existing works aim to make the system resistant to offline keyword guessing, but this does not protect against online attacks on real world services. In this paper, we move a step forward to present a generic framework able to resist online keyword guessing attack using a server-assisted model. Specifically, we design a novel primitive C statistical all-but-one lossy encryption, which can prevent a specific user from generating valid encryptions. This primitive can be seen as an access control on encryption ability. Combining searchable encryption technique with the new primitive makes online keyword guessing attack impossible for the specified user, even if the attack is launched online. We further give formal security analysis for the generic framework, and a concrete implementation with efficiency analysis to show that our design is practical.

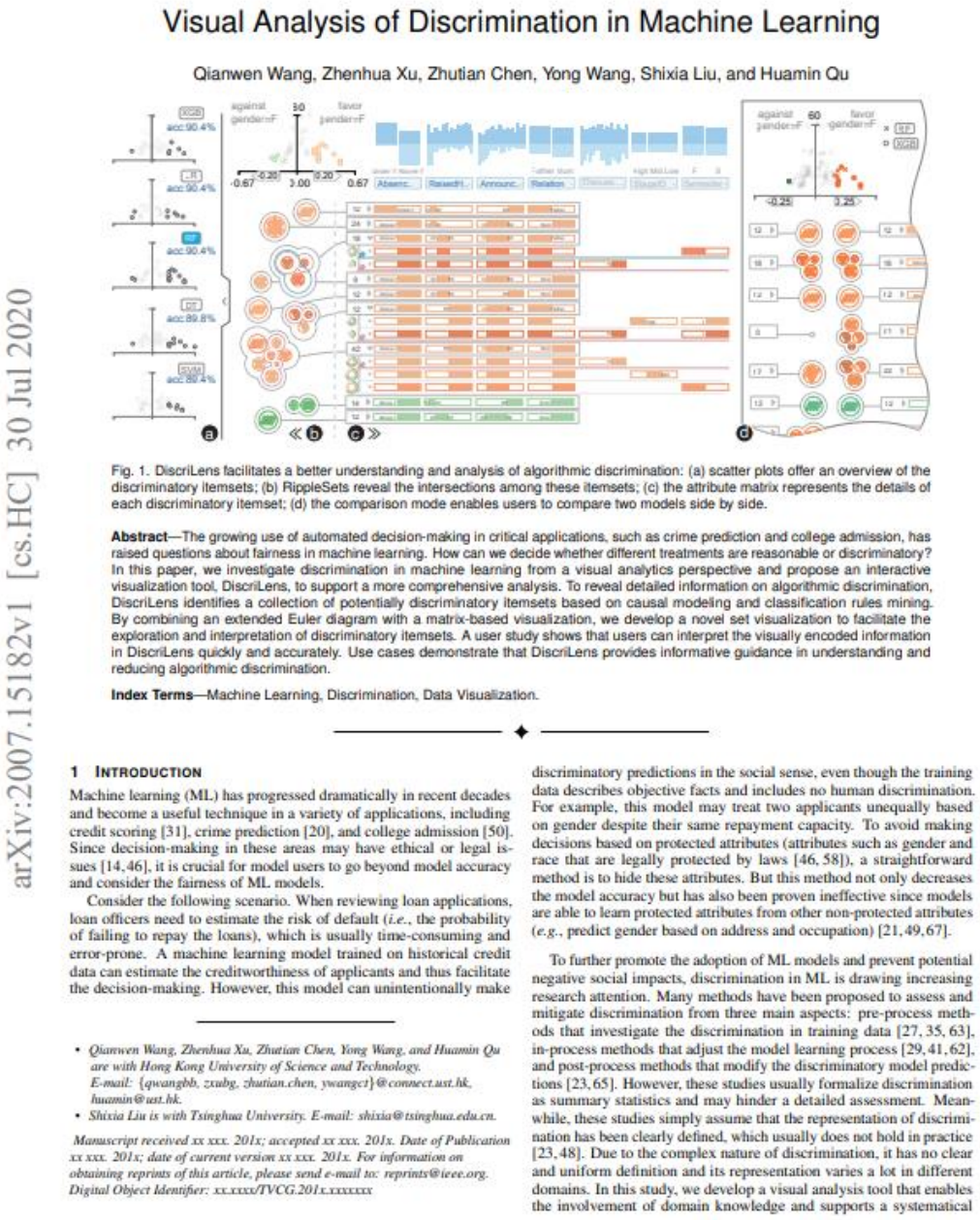
Keywords: Keyword search, Encrypted data, Security, Online keyword guessing attack

1. Introduction

Searchable functionality is popular in many online services like email system and data storage. One can search his email box with date to find the emails he received at the date. Or one can also use keywords to find out those goods with names which contain the keywords in online shopping. However, searching by keywords makes a threat to user's privacy. For example, attackers can use data mining technology to analyze the search history of an online merchant user to determine his hobby or income range. So in practice, some cryptographic tools are used in the protocol to protect privacy, like encrypted email systems.

Encryption technologies (e.g., AES(Advanced Encryption Standard), DES (Data Encryption Standard), RSA (Rivest et al., 1978) have been widely employed in real-world applications, such as ProtonMail[1] and SpiderOak[2], to protect the confidentiality of client data. However, the encryption may prevent further operations over encrypted data. For instance, making use of RSA to encrypt emails and further store the encryptions to a remote server that makes search over data somewhat impossible on server side. This is because server may not be shared decryption rights by clients. There are some naive approaches to enable server to search over encrypted emails. One solution is to persuade clients to share secret keys with server, so that server can first decrypt and further search over plain data. This, however, may infringe the privacy of clients and meanwhile, it may not be necessary for clients to share fully trust with server. Another solution could be that clients downloads all encrypted emails to local and further fulfil throughout search over decryption. Although leaking no email information to server, this solution may yield heavy bandwidth and computational cost

[1] https://protonmail.com/
[2] https://spideroak.com



Visual Analysis of Discrimination in Machine Learning

Qianwen Wang, Zhenhua Xu, Zhutian Chen, Yong Wang, Shixia Liu, and Huamin Qu

Fig. 1. DiscriLens facilitates a better understanding and analysis of algorithmic discrimination: (a) scatter plots offer an overview of the discriminatory itemsets; (b) RippleSets reveal the intersections among these itemsets; (c) the attribute matrix represents the details of each discriminatory itemset; (d) the comparison mode enables users to compare two models side by side.

Abstract—The growing use of automated decision-making in critical applications, such as crime prediction and college admission, has raised questions about fairness in machine learning. How can we decide whether different treatments are reasonable or discriminatory? In this paper, we investigate discrimination in machine learning from a visual analytics perspective and propose an interactive visualization tool, DiscriLens, to support a more comprehensive analysis. To reveal detailed information on algorithmic discrimination, DiscriLens identifies a collection of potentially discriminatory itemsets based on causal modeling and classification rules mining. By combining an extended Euler diagram with a matrix-based visualization, we develop a novel set visualization to facilitate the exploration and interpretation of discriminatory itemsets. A user study shows that users can interpret the visually encoded information in DiscriLens quickly and accurately. Use cases demonstrate that DiscriLens provides informative guidance in understanding and reducing algorithmic discrimination.

Index Terms—Machine Learning, Discrimination, Data Visualization.

1 Introduction

Machine learning (ML) has progressed dramatically in recent decades and become a useful technique in a variety of applications, including credit scoring [31], crime prediction [20], and college admission [50]. Since decision-making in these areas may have ethical or legal issues [14, 46], it is crucial for model users to go beyond model accuracy and consider the fairness of ML models.

Consider the following scenario. When reviewing loan applications, loan officers need to estimate the risk of default (*i.e.*, the probability of failing to repay the loans), which is usually time-consuming and error-prone. A machine learning model trained on historical credit data can estimate the creditworthiness of applicants and thus facilitate the decision-making. However, this model can unintentionally make discriminatory predictions in the social sense, even though the training data describes objective facts and includes no human discrimination. For example, this model may treat two applicants unequally based on gender despite their same repayment capacity. To avoid making decisions based on protected attributes (attributes such as gender and race that are legally protected by laws [46, 58]), a straightforward method is to hide these attributes. But this method not only decreases the model accuracy but has also been proven ineffective since models are able to learn protected attributes from other non-protected attributes (*e.g.*, predict gender based on address and occupation) [21, 49, 67].

To further promote the adoption of ML models and prevent potential negative social impacts, discrimination in ML is drawing increasing research attention. Many methods have been proposed to assess and mitigate discrimination from three main aspects: pre-process methods that investigate the discrimination in training data [27, 35, 63], in-process methods that adjust the model learning process [29, 41, 62], and post-process methods that modify the discriminatory model predictions [23, 65]. However, these studies usually formalize discrimination as summary statistics and may hinder a detailed assessment. Meanwhile, these studies simply assume that the representation of discrimination has been clearly defined, which usually does not hold in practice [23, 48]. Due to the complex nature of discrimination, it has no clear and uniform definition and its representation varies a lot in different domains. In this study, we develop a visual analysis tool that enables the involvement of domain knowledge and supports a systematical

- *Qianwen Wang, Zhenhua Xu, Zhutian Chen, Yong Wang, and Huamin Qu are with Hong Kong University of Science and Technology. E-mail: {qwangbb, zxubg, zhutian.chen, ywangct}@connect.ust.hk, huamin@ust.hk.*
- *Shixia Liu is with Tsinghua University. E-mail: shixia@tsinghua.edu.cn.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

Gold open access

- Gold open access – payment of **Article Processing Charge** to make the article open access on the publisher website – published version

Data Science and Engineering (2019) 4:93–108
https://doi.org/10.1007/s41019-019-0095-7

Distributed Similarity Queries in Metric Spaces

Keyu Yang[1] · Xin Ding[1] · Yuanliang Zhang[1] · Lu Chen[2] · Baihua Zheng[3] · Yunjun Gao[1]

Received: 3 June 2019 / Accepted: 12 June 2019 / Published online: 28 June 2019
© The Author(s) 2019

Abstract
Similarity queries, including range queries and k nearest neighbor (kNN) queries, in metric spaces have applications in many areas such as multimedia retrieval, computational biology and location-based services. With the growing volumes of data, a distributed method is required. In this paper, we propose an Asynchronous Metric Distributed System (AMDS), to support efficient metric similarity queries in the distributed environment. AMDS uniformly partitions the data with the pivot-mapping technique to ensure the load balancing, and employs publish/subscribe communication model to asynchronous process large scale of queries. The employment of asynchronous processing model also improves robustness and efficiency of AMDS. In addition, we develop efficient similarity search algorithms using AMDS. Extensive experiments using real and synthetic data demonstrate the performance of metric similarity queries using AMDS. Moreover, the AMDS scales sublinearly with the growing data size.

Keywords Similarity query · Range query · kNN query · Metric space · Distributed processing · Algorithm

1 Introduction

Similarity queries in metric spaces find objects similar to a given query object under a certain criterion. This functionality has been widely used in real-life applications. This is because metric spaces can support various data types (e.g., images, words, DNA sequences) and flexible distance metrics (e.g., L_p-norm distance, edit distance). Here, we give two representative examples below.

Application 1 (Multimedia Retrieval) In an image retrieval system, the similarity between images can be measured using L_p-norm metric, earth mover's distance or other distance metrics between their corresponding feature vectors. Here, similarity queries in metric space can help users find similar figures for a given one.

Application 2 (Nature Language Processing) In the *WordNet*, a knowledge graph for better nature language understanding, the similarity between two words could be measured by the shortest path, maximum flow or other distance

Extended from their APWEB-WAIM 2018 papers

This journal paper is an extended version of the conference paper, titled "*Distributed k- Nearest Neighbour Queries in Metric Spaces*," which has been published in APWeb-WAIM 2018. Specifically, the paper extends the conference paper by including (i) one additional interesting query, i.e., metric range queries (Sect. 5.1); (ii) an enhanced experimental evaluation that incorporates the new class of queries and the flexibility of the system (Sect. 6); and (iii) more comprehensive coverage of related work (Sect. 2). Also, the introduction (Sect. 1) has been revised and additional theoretical analysis is included.

Yunjun Gao
gaoyj@zju.edu.cn; raygaoyj@gmail.com
Keyu Yang
kyyang@zju.edu.cn
Xin Ding
dingxin@zju.edu.cn
Yuanliang Zhang
yuanlz@zju.edu.cn
Lu Chen
luchen@cs.aau.dk
Baihua Zheng
bhzheng@smu.edu.sg
1 College of Computer Science, Zhejiang University, 38 Zheda Road, Hangzhou 310027, China
2 Department of Computer Science, Aalborg University, Aalborg, Denmark
3 School of Information Systems, Singapore Management University, 80 Stamford Road, Singapore 178902, Singapore

Springer

Computers & Security 95 (2020) 101836

Contents lists available at ScienceDirect

Computers & Security

journal homepage: www.elsevier.com/locate/cose

Privacy preserving search services against online attack

Yi Zhao[a,b,c], Jianting Ning[d,e], Kaitai Liang[f], Yanqi Zhao[b,c], Liqun Chen[f], Bo Yang[b,c,*]

[a] School of Information Engineering, Chang'an University, Xi'an 710064, China
[b] School of Computer Science, Shaanxi Normal University, Xi'an 710062, China
[c] State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China
[d] Department of Computer Science, Fujian Normal University, China
[e] Department of Computer Science, Singapore Management University, Singapore
[f] Department of Computer Science, University of Surrey, UK

ARTICLE INFO

Article history:
Received 5 June 2019
Revised 6 February 2020
Accepted 11 April 2020
Available online 29 April 2020

Keywords:
Keyword search
Encrypted data
Security
Online keyword guessing attack

ABSTRACT

Searchable functionality is provided in many online services such as mail services or outsourced data storage. To protect users privacy, data in these services is usually stored after being encrypted using searchable encryption. This enables the data user to securely search encrypted data from a remote server without leaking data and query information. Public key encryption with keyword search is one of the research branches of searchable encryption; this provides privacy-preserving searchable functionality for applications such as encrypted email systems. However, it has an inherent vulnerability in that the information of a query may be leaked using a keyword guessing attack. Most of existing works aim to make the system resistant to offline keyword guessing, but this does not protect against online attacks on real world services. In this paper, we move a step forward to present a generic framework able to resist online keyword guessing attack using a server-assisted model. Specifically, we design a novel primitive C mirrored all-but-one lossy encryption, which can prevent a specific user from generating valid encryptions. This primitive can be seen as an access control on encryption ability. Combining searchable encryption technique with the new primitive makes online keyword guessing attack impossible for the specified user, even if the attack is launched online. We further give formal security analysis for the generic framework, and a concrete implementation with efficiency analysis to show that our design is practical.

1. Introduction

Searchable functionality is popular in many online services like email system and data storage. One can search his email box with date to find the emails he received at the date. Or one can also use keywords to find out those goods with names which contain the keywords in online shopping. However, searching by keywords makes a threat to user's privacy. For example, attackers can use data mining technology to analyse the search history of an online merchant user to determine his hobby or income range. So in practice, some cryptographic tools are used in the protocol to protect privacy, like encrypted email systems.

Encryption technologies (e.g., AES(Advanced Encryption Standard), DES (Data Encryption Standard), RSA (Rivest et al., 1978) have been widely employed in real-world applications, such as ProtonMail[1] and SpiderOak[2], to protect the confidentiality of client data. However, the encryption may prevent further operations over encrypted data. For instance, making use of RSA to encrypt emails and further store the encryptions to a remote server that makes search over data somewhat impossible on server side. This is because server may not be shared decryption rights by clients. There are some naive approaches to enable server to search over encrypted emails. One solution is to persuade clients to share secret keys with server, so that server can first decrypt and further search over plain data. This, however, may infringe the privacy of clients and meanwhile, it may not be necessary for clients to share fully trust with server. Another solution could be that clients downloads all encrypted emails to local and further fulfil throughout search over decryption. Although leaking no email information to server, this solution may yield heavy bandwidth and computational cost

* Corresponding author at: School of Computer Science, Shaanxi Normal University, Xi'an 710062, China.
E-mail address: byang@snnu.edu.cn (B. Yang).

https://doi.org/10.1016/j.cose.2020.101836
0167-4048/© 2020 Elsevier Ltd. All rights reserved.

[1] https://protonmail.com/.
[2] https://spideroak.com/.

Commanding and Re-Dictation: Developing Eyes-Free Voice-Based Interaction for Editing Dictated Text

DEBJYOTI GHOSH, NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore
CAN LIU, School of Creative Media, City University of Hong Kong
SHENGDONG ZHAO, School of Computing, National University of Singapore
KOTARO HARA, School of Information Systems, Singapore Management University

Existing voice-based interfaces have limited support for text editing, especially when *seeing* the text is difficult, e.g., while walking or cooking. This research develops voice interaction techniques for eyes-free text editing. First, with a Wizard-of-Oz study, we identified two primary user strategies: using commands, e.g., "REPLACE *go* with *goes*" and re-dictating over an erroneous portion, e.g., correcting "he go there" by saying "he goes there." To support these user strategies with an actual system implementation, we developed two eyes-free voice interaction techniques, *Commanding* and *Re-dictation*, and evaluated them with a controlled experiment. Results showed that while Re-dictation performs significantly better for more semantically complex edits, Commanding is more suitable for making one-word edits, especially deletions. We developed *VoiceRev* to combine both the techniques in the same interface and evaluated it with realistic tasks. Results showed improved usability of the combined techniques over either of the two techniques used individually.

CCS Concepts: • **Human-centered computing** → **Interaction techniques**; **Text input**;

Additional Key Words and Phrases: Text editing, commanding, re-dictation, eyes-free, voice-based text editing, voice interaction, voice user interfaces

ACM Reference format:
Debjyoti Ghosh, Can Liu, Shengdong Zhao, and Kotaro Hara. 2020. Commanding and Re-Dictation: Developing Eyes-Free Voice-Based Interaction for Editing Dictated Text. *ACM Trans. Comput.-Hum. Interact.* 27, 4, Article 28 (August 2020), 31 pages.
https://doi.org/10.1145/3390889

1 INTRODUCTION

In Human–Computer Interaction (HCI), text editing typically involves visual engagement: Although text *input* can be performed using various techniques, e.g., physical keyboard, touch-based

This work was supported in part by the NUS Advanced Robotics Centre.

Authors' addresses: D. Ghosh, NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore, Singapore 119077; email: debjyoti@u.nus.edu; C. Liu, School of Creative Media, City University of Hong Kong, Hong Kong; email: canliu@cityu.edu.hk; S. Zhao, School of Computing, National University of Singapore, Singapore 119077; email: zhaosd@comp.nus.edu.sg; K. Hara, School of Information Systems, Singapore Management University, Singapore 178902; email: kotarohara@smu.edu.sg.

1073-0516/2020/08-ART28 $15.00
https://doi.org/10.1145/3390889

Article Processing Charges

- IEEE – journals US$1,995 – US$2,995
- Elsevier – Computers & Security US$3,190; Journal of Econometrics US$ 3,850; Leadership Quarterly US$ 4,470
- Wiley – Strategic Management Journal US$ 4,050
- Springer – Nature US$ 12,290
- Springer Open – Data Science and Engineering – publication costs covered by the sponsor Nanjing Sinovatio Technology Co. Ltd, authors do not pay APC
- Check if your institution has a deal with the publisher to allow you to publish Gold Open Access without additional charge, check terms, e.g. corresponding author, any quota?

Publishing process

- Have original research work of current interest.

- Identify a journal with aims & scope close to your research work.
- Format manuscript according to the author guideline of targeted journal.

- Submit your manuscript online by providing complete detail in online submission system.
- Do not submit manuscript to more than one journal at the same time.

- Editor will initiate peer-review process for well formatted manuscript within the scope of journal
- Keep patience and wait for a reply from editor during review process.

Based on reviewer's recommendation, editor will send you one of the following decision letters-

- **Revision Required:** Revise manuscript as per comment and resubmit.
- **Decline:** Look at your manuscript and decline reason. Spend plenty of time to improve your overall manuscript and resubmit as a fresh manuscript.

- **Accepted:** Its time to celebrate, your manuscript will be copyedited by the journal for final publication.

- You will receive galley proof version for minor proofreading corrections and your paper will be published.
- It's time to share your published work and cite in other related papers.

Source: https://aijr.org/paper-publishing-process/

Criteria for Authorship (ACM)

Anyone listed as Author on an ACM manuscript submission must meet all the following criteria:

- they must be the "creator or originator of an idea"
- they must have made a substantial intellectual contribution to the work
- they take full responsibility for all content in the work
- Other contributors may be acknowledged at the end of the paper, before the bibliography, with explicitly described roles, preferably using the roles found in the NISO Contributor Roles Taxonomy at https://credit.niso.org/
- ACM requires that all published papers include the names, affiliations where research took place, and email address' of all authors listed on the paper, as well as provide accurate contact information to ACM as required in the ACM rights contract. ACM does not allow anonymous authors.

Source: https://www.acm.org/publications/policies/new-acm-policy-on-authorship; https://publicationethics.org/resources/discussion-documents/authorship

CRediT

CRediT (Contributor Roles Taxonomy) is a high-level taxonomy with 14 roles, to represent roles typically played by contributors to research outputs.

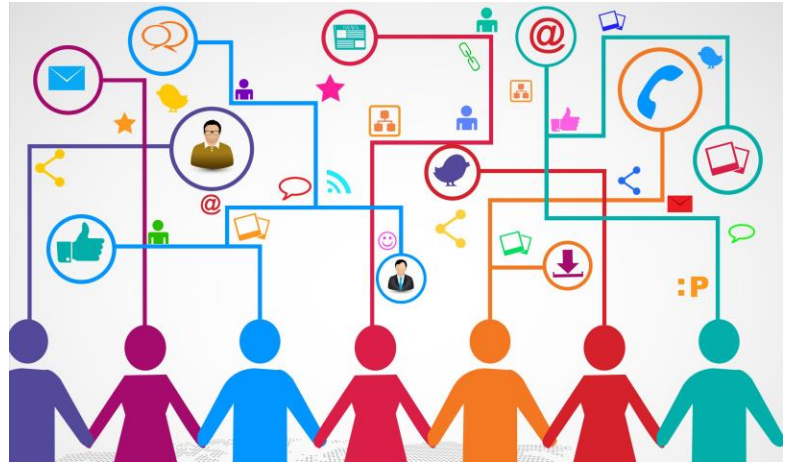
- Conceptualization
- Data Curation
- Formal Analysis
- Funding acquisition
- Investigation
- Methodology
- Project administration
- Resources
- Software
- Supervision
- Validation
- Visualization
- Writing – original draft
- Writing – review and editing

Source: https://credit.niso.org/

Research visibility

- In an age of abundance of scholarly research, it is important to increase the visibility of your research and make meaningful impact
- Share your research in meaningful ways
- Present in your institution, or at relevant conferences
- Use online platforms and social media to amplify your research and its findings

Sharing your work



- Include in your email footer
- Include in your conference presentation, or school presentation
- Announce on LinkedIn, X, Facebook
- Add appropriate version to your institution research information system, institutional repository, pre-print servers and school website – for SMU: IRIS → InK
- Add to your author profile, e.g. ORCID, Google Scholar profile, ResearchGate
- Ensure it is indexed in major databases (Scopus & Web of Science) and your subject discipline databases
- Search in various search engines to confirm it can be found

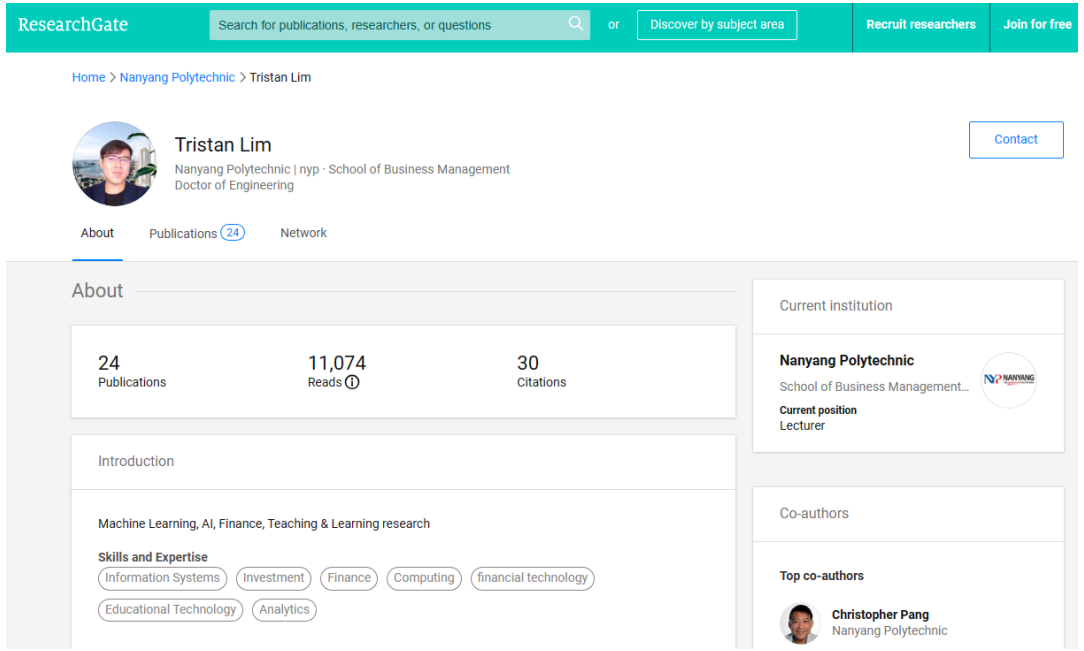
Presenting your work

- Conferences (select reputable and relevant ones)
- School presentations
- Competitions

Online platforms

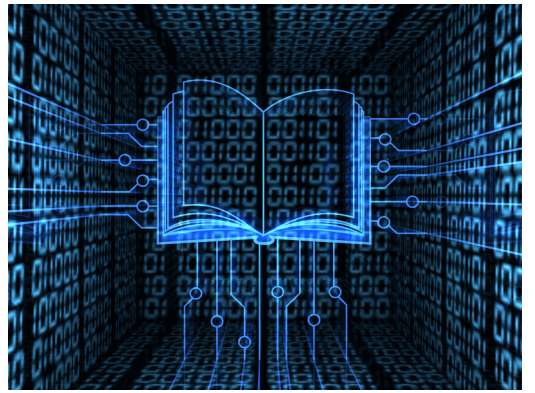
- Reasons to use online platforms:
  - Share information quickly
  - Get recognition and exposure
  - Engage with other researchers and potential collaborators

ResearchGate



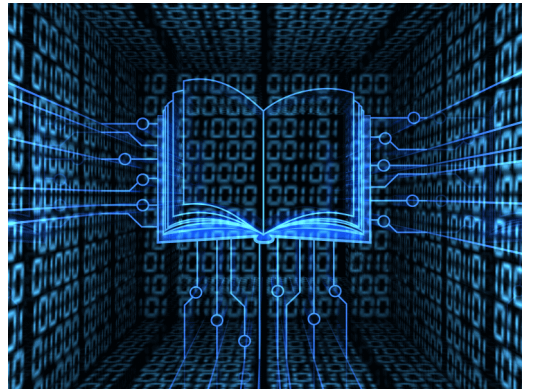
- Platform to share publications
- Network with other researchers

Pre-print Repositories



- Pre-print servers are a way to share your work before it gets accepted in a journal
- Check the policies of the journals you submit your papers on how they view pre-prints
- Look for pre-print servers in your discipline:
  - Social Science Research Network (SSRN) https://www.ssrn.com/
  - arXiv https://arxiv.org/
  - SocArXiv https://osf.io/preprints/socarxiv

Institutional Repositories



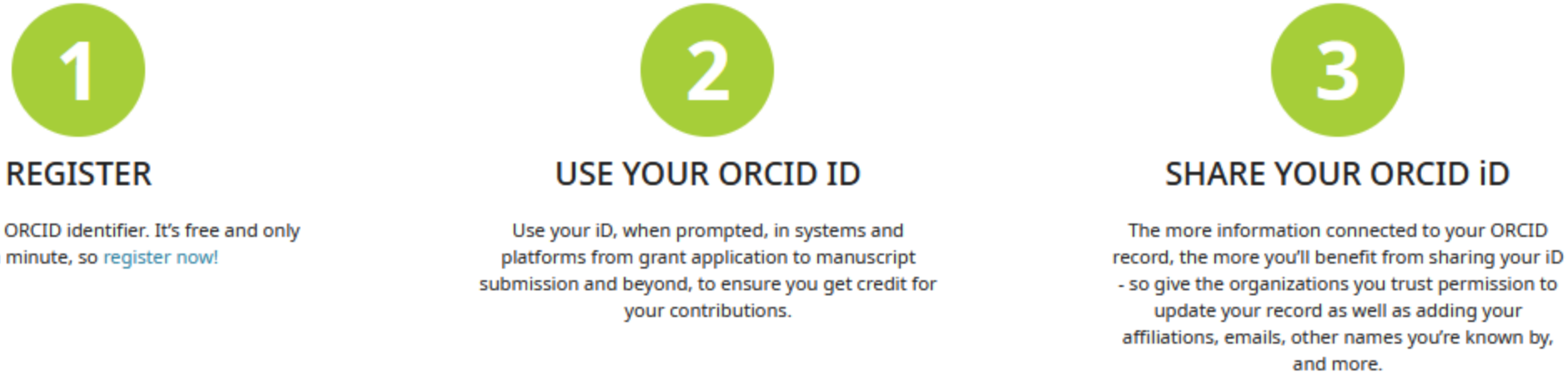
- Leverage on your institutional repository and use it to share your papers green open access, with good discoverability in search engines
  - NTU Digital Repository https://dr.ntu.edu.sg/
  - NUS ScholarBank https://scholarbank.nus.edu.sg/
  - SMU Institutional Knowledge https://ink.library.smu.edu.sg

Author profiles

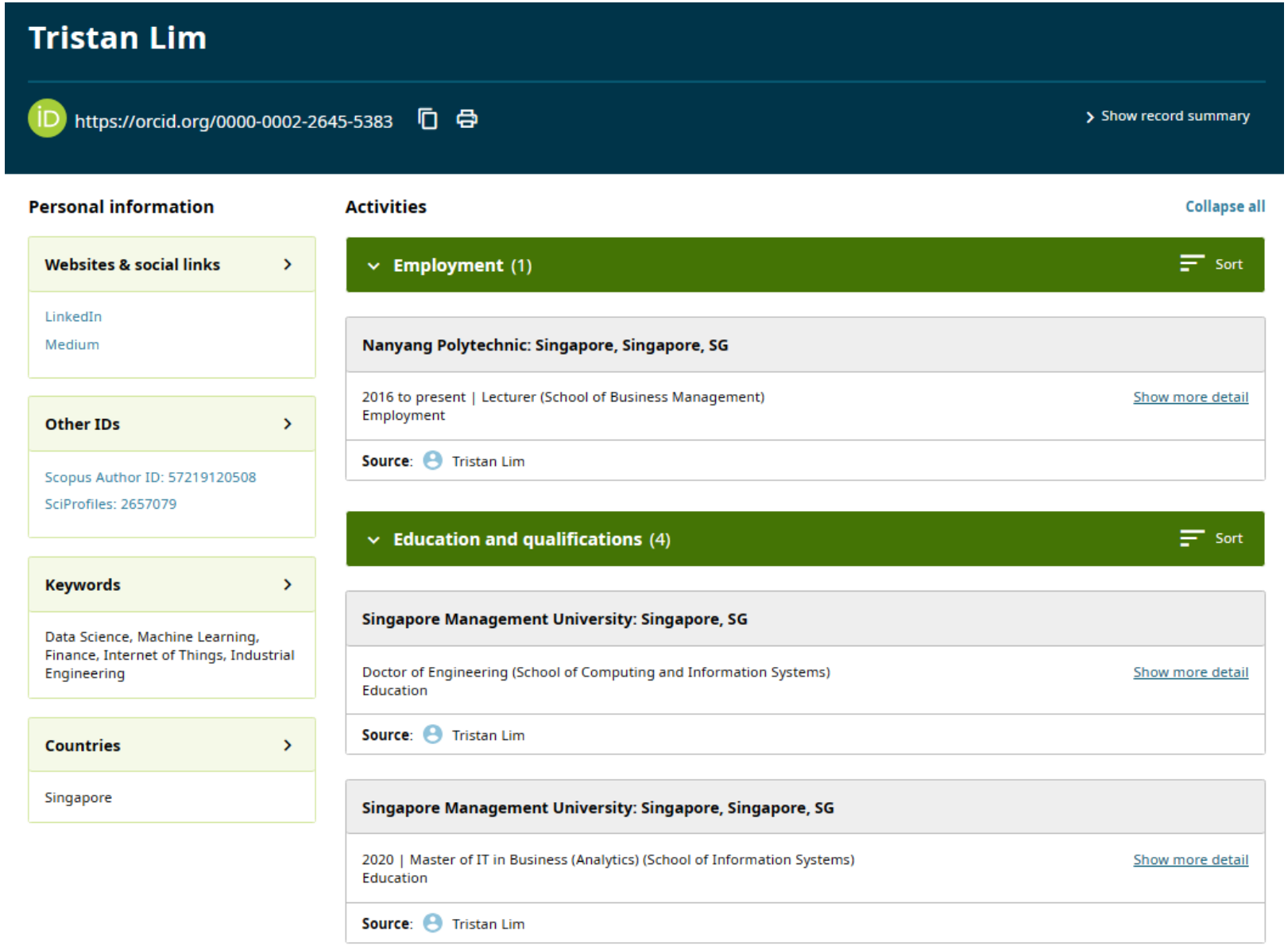
- Author profiles can be used as a tool to:
  - associate all your publications to your profile
  - promote publications and expertise
  - facilitate scholarly networking
  - share publications
  - make an impact
  - increase visibility and scholarly reputation
- Different platforms offer variations in their range of services
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5639053/

ORCID

- When you submit to a journal, you may be asked to provide ORCID
- What is ORCID? https://orcid.org/
- ORCID stands for Open Researcher and Contributor ID
- Provide a unique persistent identifier for individuals to use
- Open and non-proprietary



ORCID example



- Created by user
- Non-proprietary
- Not tied to institution
- User-controlled
- User has to grant permission for other systems to update ID
- Connects to various systems used by publishers, databases, institutions

https://orcid.org/0000-0002-2645-5383

Google Scholar profile – PG student

Google Scholar SIGN IN

Tristan LIM

FOLLOW

Nanyang Polytechnic
Verified email at nyp.edu.sg - Homepage

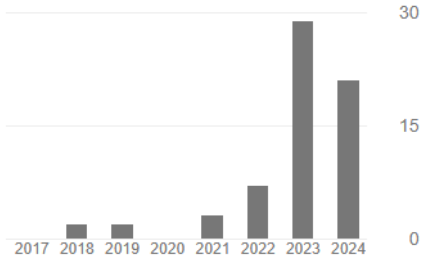
Fintech Finance AI Machine Learning Education Technology

GET MY OWN PROFILE

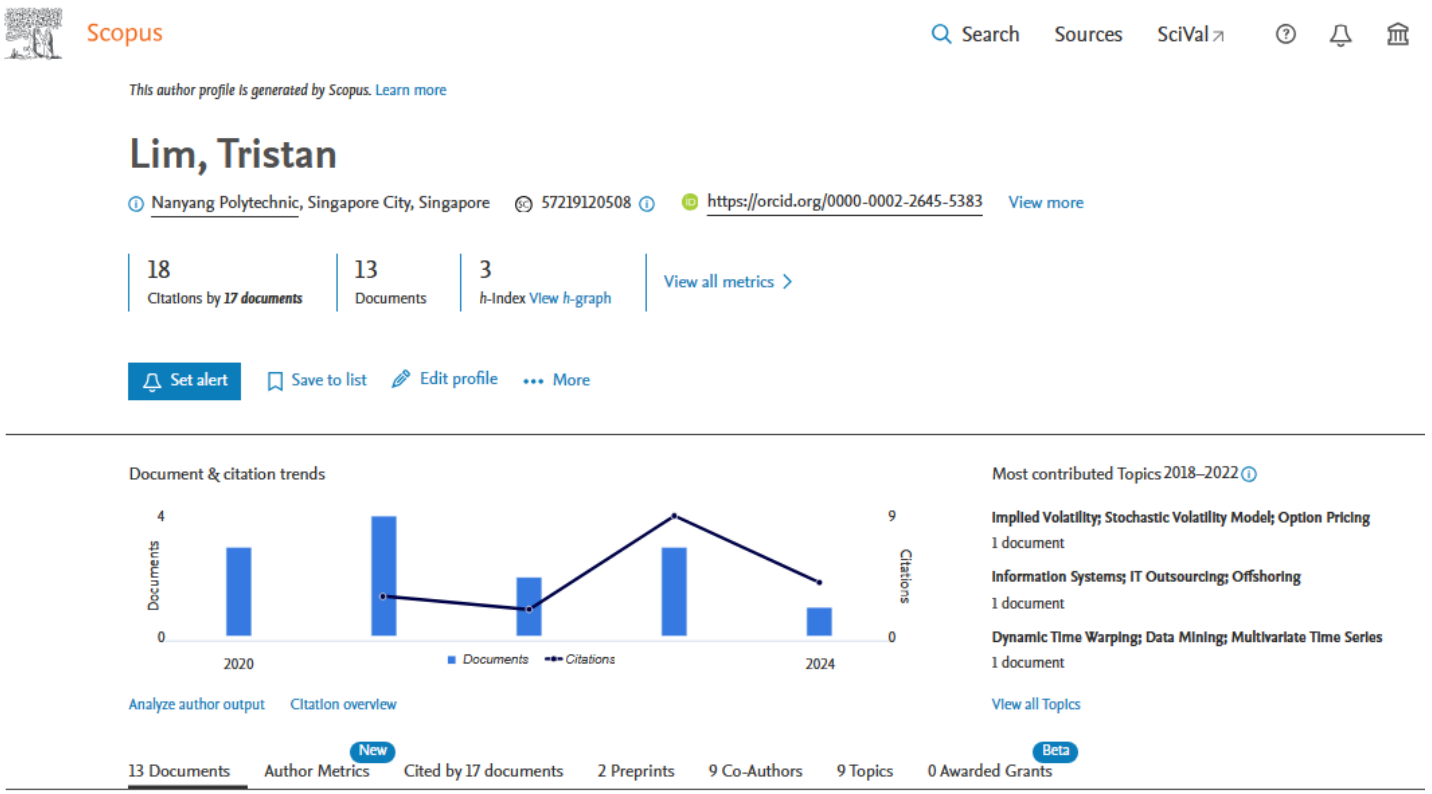
TITLE	CITED BY	YEAR
Environmental, social, and governance (ESG) and artificial intelligence in finance: State-of-the-art and research takeaways T Lim Artificial Intelligence Review 57 (4), 1-64	1 *	2024
Educational Technologies and Assessment Practices: Evolution and Emerging Research Gaps T Lim, S Gottipati, MLF Cheong Reshaping Learning with Next Generation Educational Technologies, 136-172		2024
Theory-Guided Analytics Process: Using Theories to Underpin an Analytics Process for New Banking Product Development Using Segmentation-Based Marketing Analytics Leveraging on … T Lim, T Pan, CS Ong, S Chen, JJJ Chia Analytics 2 (1), 105-131	1	2023

Cited by VIEW ALL

	All	Since 2019
Citations	70	64
h-index	5	4
i10-index	2	2



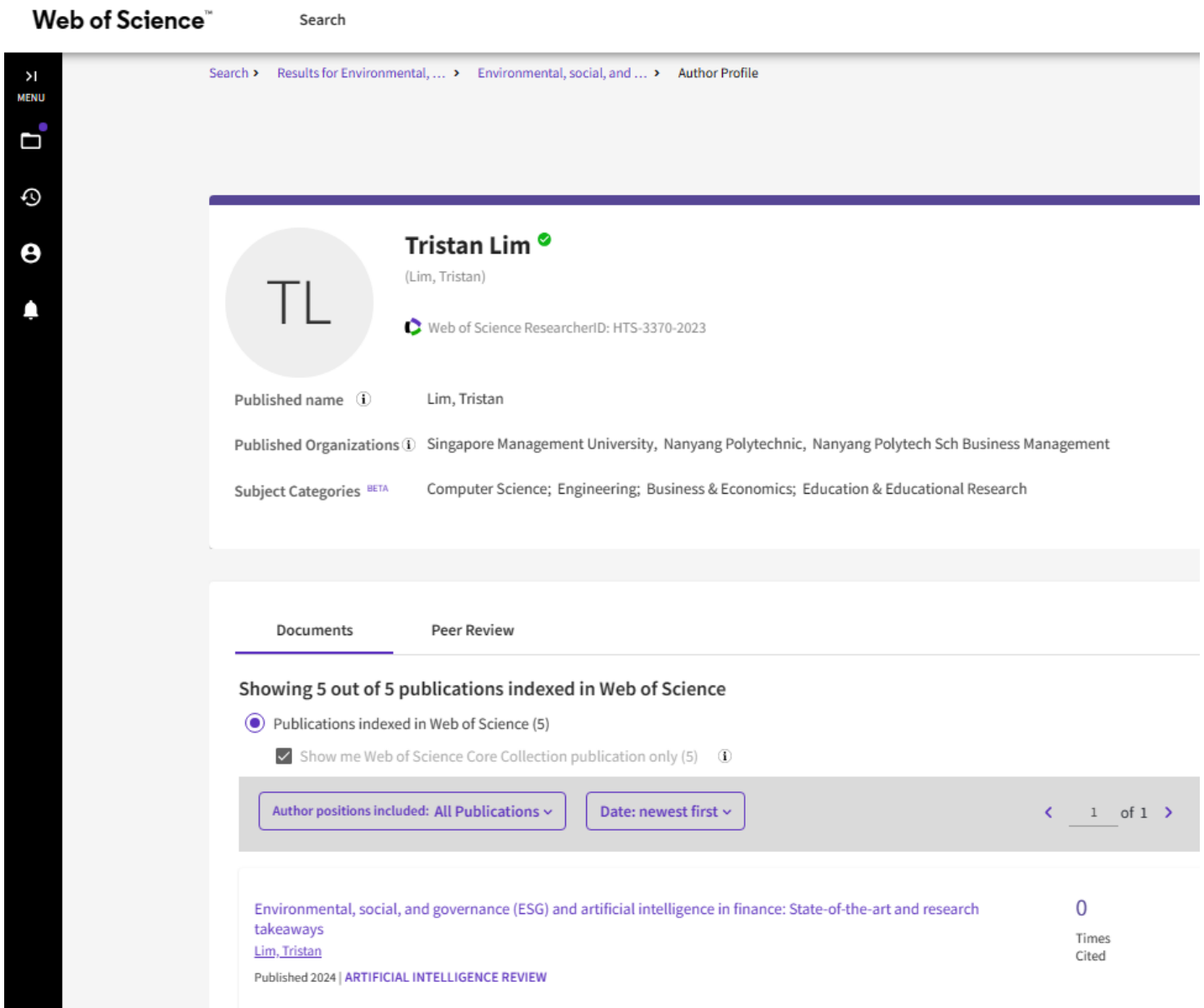
Scopus Author Profile



- Created by Scopus
- Uses data in Scopus for citations, h-index
- Proprietary

https://www.scopus.com/authid/detail.uri?authorId=57219120508

Web of Science ResearcherID

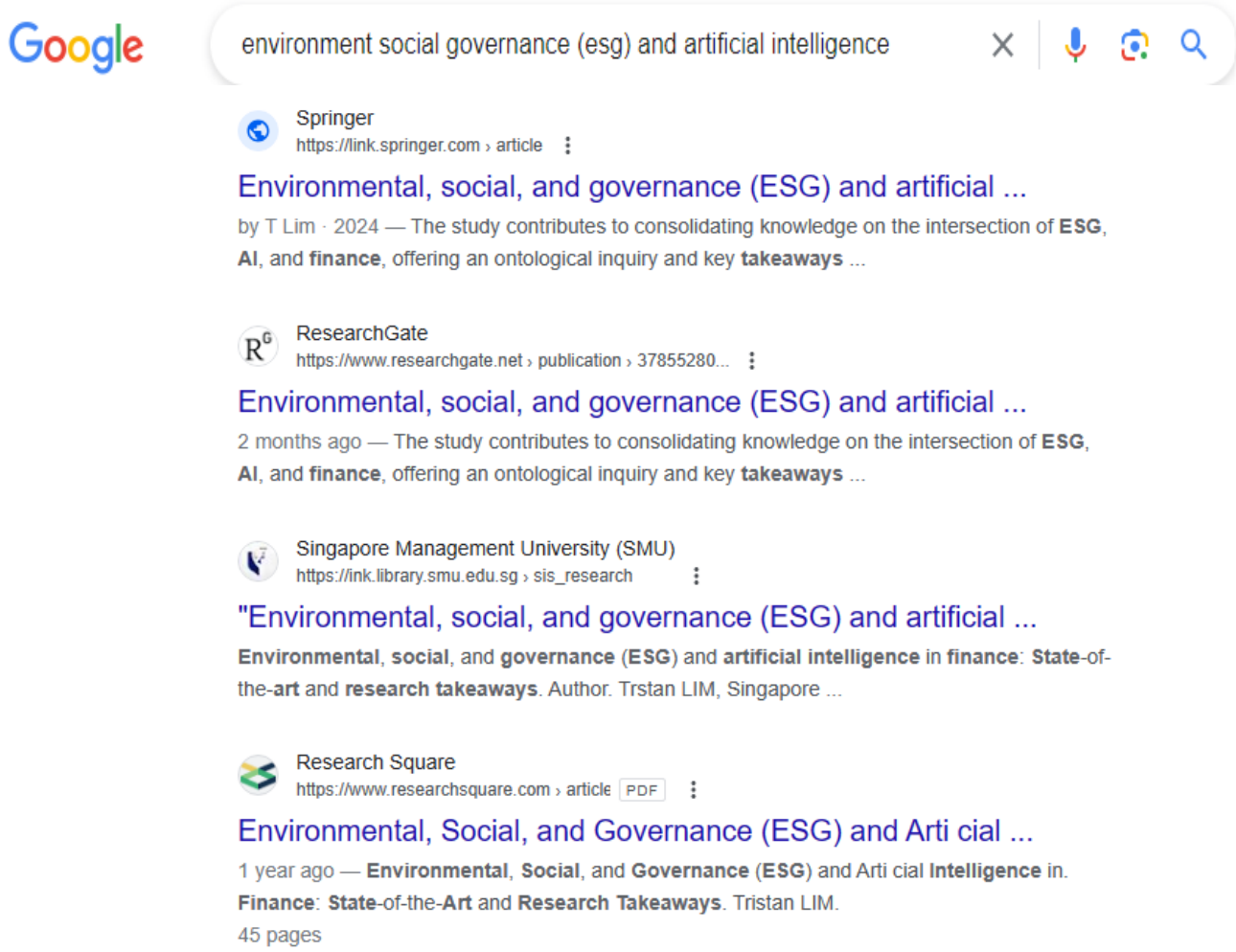


- Created by user/ Web of Science
- Uses data in Web of Science for citations, h-index
- Proprietary

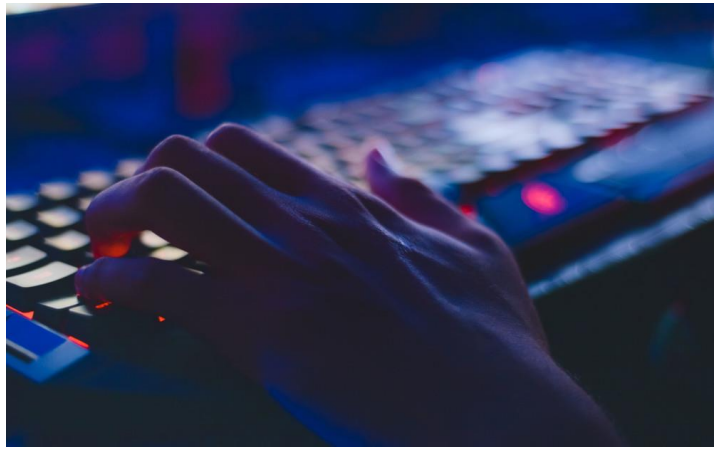
https://www.webofscience.com/wos/author/record/40043553

Search discovery

For better discoverability, it is important to enable your research to be found in search engines



Summary: Publishing & Research Visibility



- Publishing process
  - Determine if you have original research to write up
  - Understand the peer review process
  - Make your work open access – green open access is free
- Research visibility
  - Share your work on online platforms
  - Keep your author profiles up-to-date
- Questions?

Resources

- ACM Information for Authors https://www.acm.org/publications/authors/information-for-authors
- Cambridge University Press Author Resources https://www.cambridge.org/authorhub/resources
- SAGE Journal Author Gateway https://us.sagepub.com/en-us/nam/journal-author-gateway
- Springer Information for Journal Article Authors https://www.springer.com/gp/authors-editors/journal-author
- Wiley Author Resources. https://authorservices.wiley.com/author-resources/index.html