Full text available at: http://dx.doi.org/10.1561/1500000069

Search and Discovery in Personal Email Collections

Full text available at: http://dx.doi.org/10.1561/1500000069

Other titles in Foundations and Trends® in Information Retrieval

Extracting, Mining and Predicting Users' Interests from Social Media
Fattane Zarrinkalam, Stefano Faralli, Guangyuan Piao and Ebrahim Bagheri
ISBN: 978-1-68083-738-4

Knowledge Graphs: An Information Retrieval Perspective
Ridho Reinanda, Edgar Meij and Maarten de Rijke
ISBN: 978-1-68083-728-5

Deep Learning for Matching in Search and Recommendation
Jun Xu, Xiangnan He and Hang Li
ISBN: 978-1-68083-706-3

Explainable Recommendation: A Survey and New Perspectives
Yongfeng Zhang and Xu Chen
ISBN: 978-1-68083-658-5

Information Retrieval: The Early Years
Donna Harman
ISBN: 978-1-68083-584-7

Bandit Algorithms in Information Retrieval
Dorota Glowacka
ISBN: 978-1-68083-574-8

Neural Approaches to Conversational AI
Jianfeng Gao, Michel Galley and Lihong Li
ISBN: 978-1-68083-552-6

An Introduction to Neural Information Retrieval
Bhaskar Mitra and Nick Craswell
ISBN: 978-1-68083-532-8

Full text available at: http://dx.doi.org/10.1561/1500000069

Search and Discovery in Personal Email Collections

Michael Bendersky
Google
bemike@google.com

Xuanhui Wang
Google
xuanhui@google.com

Marc Najork
Google
najork@google.com

Donald Metzler
Google
metzler@google.com

the essence of knowledge
Boston — Delft

Full text available at: http://dx.doi.org/10.1561/1500000069

Foundations and Trends® in Information Retrieval

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

M. Bendersky, X. Wang, M. Najork, D. Metzler. *Search and Discovery in Personal Email Collections*. Foundations and Trends® in Information Retrieval, vol. 15, no. 1, pp. 1–133, 2021.

ISBN: 978-1-68083-839-8
© 2021 M. Bendersky, X. Wang, M. Najork, D. Metzler

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, mechanical, photocopying, recording or otherwise, without prior written permission of the publishers.

Photocopying. In the USA: This journal is registered at the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923. Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by now Publishers Inc for users registered with the Copyright Clearance Center (CCC). The 'services' for users can be found on the internet at: www.copyright.com

For those organizations that have been granted a photocopy license, a separate system of payment has been arranged. Authorization does not extend to other kinds of copying, such as that for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale. In the rest of the world: Permission to photocopy must be obtained from the copyright owner. Please apply to now Publishers Inc., PO Box 1024, Hanover, MA 02339, USA; Tel. +1 781 871 0245; www.nowpublishers.com; sales@nowpublishers.com

now Publishers Inc. has an exclusive license to publish this material worldwide. Permission to use this content must be obtained from the copyright license holder. Please apply to now Publishers, PO Box 179, 2600 AD Delft, The Netherlands, www.nowpublishers.com; e-mail: sales@nowpublishers.com

Full text available at: http://dx.doi.org/10.1561/1500000069

Foundations and Trends® in Information Retrieval
Volume 15, Issue 1, 2021

Editorial Board

Editors-in-Chief

Maarten de Rijke
University of Amsterdam
and Ahold Delhaize
The Netherlands

Yiqun Liu
Tsinghua University
China

Diane Kelly
University of Tennessee
USA

Editors

Barbara Poblete
University of Chile

Claudia Hauff
Delft University of Technology

Doug Oard
University of Maryland

Ellen M. Voorhees
National Institute of Standards and Technology

Fabrizio Sebastiani
Consiglio Nazionale delle Ricerche, Italy

Hang Li
Bytedance Technology

Isabelle Moulinier
Capital One

Jaap Kamps
University of Amsterdam

Jimmy Lin
University of Waterloo

Leif Azzopardi
University of Glasgow

Lynda Tamine
University of Toulouse

Mark Sanderson
RMIT University

Rodrygo Luis Teodoro Santos
Universidade Federal de Minas Gerais

Ruihua Song
Renmin University of China

Ryen White
Microsoft Research

Shane Culpepper
RMIT

Soumen Chakrabarti
Indian Institute of Technology

Xuanjing Huang
Fudan University

Zi Helen Huang
University of Queensland

Full text available at: http://dx.doi.org/10.1561/1500000069

Editorial Scope

Topics

Foundations and Trends® in Information Retrieval publishes survey and tutorial articles in the following topics:

- Applications of IR
- Architectures for IR
- Collaborative filtering and recommender systems
- Cross-lingual and multilingual IR
- Distributed IR and federated search
- Evaluation issues and test collections for IR
- Formal models and language models for IR
- IR on mobile platforms
- Indexing and retrieval of structured documents
- Information categorization and clustering
- Information extraction
- Information filtering and routing
- Metasearch, rank aggregation and data fusion
- Natural language processing for IR
- Performance issues for IR systems, including algorithms, data structures, optimization techniques, and scalability
- Question answering
- Summarization of single documents, multiple documents, and corpora
- Text mining
- Topic detection and tracking
- Usability, interactivity, and visualization issues in IR
- User modelling and user studies for IR
- Web search

Information for Librarians

Foundations and Trends® in Information Retrieval, 2021, Volume 15, 5 issues. ISSN paper version 1554-0669. ISSN online version 1554-0677. Also available as a combined paper and online subscription.

Full text available at: http://dx.doi.org/10.1561/1500000069

Contents

Full text available at: http://dx.doi.org/10.1561/1500000069

Full text available at: http://dx.doi.org/10.1561/1500000069

Search and Discovery in Personal Email Collections

Michael Bendersky[1], Xuanhui Wang[2], Marc Najork[3] and Donald Metzler[4]

[1] *Google; bemike@google.com*
[2] *Google; xuanhui@google.com*
[3] *Google; najork@google.com*
[4] *Google; metzler@google.com*

ABSTRACT

Email has been an essential communication medium for many years. As a result, the information accumulated in our mailboxes has become valuable for all of our personal and professional activities. For years, researchers have been developing interfaces, models and algorithms to facilitate search, discovery and organization of email data. In this survey, we attempt to bring together these diverse research directions, and provide both a historical background, as well as a comprehensive overview of the recent advances in the field. In particular, we lay out all the components needed in the design of a privacy-centric email search engine, including search interface, indexing, document and query understanding, retrieval, ranking and evaluation. We also go beyond search, presenting recent work on intelligent task assistance in email. Finally, we discuss some emerging trends and future directions in email search and discovery research.

Michael Bendersky, Xuanhui Wang, Marc Najork and Donald Metzler (2021), "Search and Discovery in Personal Email Collections", Foundations and Trends® in Information Retrieval: Vol. 15, No. 1, pp 1–133. DOI: 10.1561/1500000069.

Full text available at: http://dx.doi.org/10.1561/1500000069

1

Introduction

Email has thrived as an electronic communications medium for at least five decades, with the first published email standards dating back to Bhushan *et al.* (1973). While the basic email format — a header containing email metadata and a body containing the message content — remained more or less unchanged through the decades, the types of information shared through email have been continuously evolving.

While email was originally developed with organizational and enterprise communications in mind, the success of web-based services like *Hotmail* and *Yahoo! Mail* in the late 1990's made email a popular consumer communication tool. Over the years, and with the rise of the various messaging applications, there have been reports on a decline in interpersonal email communications, especially among younger users (Tsotsis, 2011). However, consumer email traffic has still consistently kept growing. This discrepancy can be attributed in large part to the rise of machine-generated messages, such as store promotions, newsletters, receipts and bills (Maarek, 2017).

Despite advances in instant communications, email also remains a vital communication tool in the enterprise setting. A recent survey of 1,000 U.S. employees by Naragon (2018) finds that users spend more

Full text available at: http://dx.doi.org/10.1561/1500000069

than 3 hours on a weekday checking their work email. Roughly 50% of survey participants check both their personal and work email at least every few hours. Naragon (2018) also reports that in a work setting, email is a more preferred communication medium than either instant messaging (11% preference), or phone (16%), and is tied in popularity with face-to-face communications (31%).

The popularity of email in both our personal lives and in the workplace is in part due to its use for collaborative task management. Collaborative task management involves reminder creation, identification of messages related to the task, synthesis of information from these messages, and interaction with others in order to complete the task. Regardless of its limitations, email is often the preferred medium for these activities (Whittaker, 2005).

As a confluence of these factors, email remains a reliable repository of information about our personal and organizational communications, social networks, activities, financial transactions, travel plans, and work commitments. As our mailboxes grow, so does the need for the development of new effective approaches to information finding in this repository. As researchers repeatedly discover, there is a substantial difference between search in public data (e.g., web search) and private email collections.

First, chronology plays an important role for both email search algorithms and interfaces (Dumais *et al.*, 2003). Second, corpus size of single mailbox is drastically smaller than that of a large web corpus. This often leads to low recall, especially for longer queries, or when there is a vocabulary mismatch between user queries and their mailboxes (Carmel *et al.*, 2015; Li *et al.*, 2019b). Finally, developing effective search algorithms while stringently preserving the privacy of user information is a difficult research challenge (Bendersky *et al.*, 2018).

Therefore, in this survey, we provide an overview of the current state-of-the art techniques that focus on these unique aspects of email management, search and discovery. Since we assume that most of our readers are more familiar with the web search counterparts of these techniques, we contrast and draw comparisons between web and email search, when appropriate.

Full text available at: http://dx.doi.org/10.1561/1500000069

1.1 Email Statistics

Before diving into describing the various email use cases, in this section we provide an overview of email usage, including general statistics, the demographic characteristics of its user base, and modes of email access.

The Radicati Group, Inc. (2019) report[1] states that the total number of emails sent and received per day will have exceeded 300 billion in 2020, and that email will be used by 4 billion people, over half of the world's population. Despite email being a mature technology, the report projects steady year-over-year growth of roughly 4% for the next several years. The Radicati Group, Inc. (2015) report also breaks down these statistics by business and consumer users, finding that the number of business emails exceeds the number of consumer emails sent and received, with both numbers projected to grow. The growth in the consumer email traffic is cited to be mainly due to machine-generated email, not interpersonal communication, which is consistent with other reports (Maarek, 2017).

These statistics demonstrate the importance of email in the business setting, and allow to draw a clear distinction between the personal email use case, and the enterprise use case (Narang *et al.*, 2017). This puts business email search and discovery in a clear connection to the existing work on enterprise search (Kruschwitz and Hull, 2017), with the added constraint that the corpora (user mailboxes) are private, rather than shared across the organization.

Narang *et al.* (2017) investigate email usage in a large organization (*Microsoft*) and report on the activities performed by a large sample of close to 300,000 US employees. In particular, they note that as mailbox size increases, people are much more likely to spend time on its organization by deleting, moving or marking email. Search activity also has strong positive correlation with the mailbox size. Activity analysis shows that $20\% - 35\%$ of all email activity involves organization, and $10\% - 20\%$ involves search, with the variation mainly attributed to the mailbox size and email deletion rate.

[1] The Radicati Group is an analyst firm specialized in tracking emerging communication and collaboration technologies, providing quantitative and qualitative market research. In this survey, we are quoting statistics provided in their 2015, 2018 and 2019 executive summaries, which are available online at www.radicati.com.

Full text available at: http://dx.doi.org/10.1561/1500000069

For the personal email use-case, Carmel *et al.* (2017b) provide a fascinating peek into the demographics of the *Yahoo! Mail* US user base. Overall, they find that email users are older and more affluent than both the average web searchers, as well as the overall US population. They are also more likely to be female – 58.4% of all email searches come from women, as opposed to 49.7% of web searches.

While in the early days of email desktop clients using POP or IMAP were more prevalent, today many users use webmail or mobile clients to access their email. Both webmail and mobile email clients are usually controlled by a large email provider that also controls a centralized secure storage for all user mailboxes. A recent Litmus Email Analytics (2019) report indicates that only 18% of the email opens today can be attributed to desktop clients. The same report lists *Gmail*, *Yahoo! Mail* and *Outlook.com* as the most used webmail clients. Examples of international webmail providers also include, among others, *QQ Mail* by Tencent, *163/126 Mail* by NetEase, *Mail.Ru*, *Yandex*, *ProtonMail* and *GMX Mail*.

As most people access their email today through one of these large-scale centralized email providers, in the remainder of this survey we shall assume that the mailboxes are centrally and securely stored and managed. This setting provides the opportunity to develop new search and discovery capabilities using a large-scale dataset containing millions of user mailboxes. It also carries the challenge of developing these capabilities while maintaining user trust through audited access, data anonymization, and data erasure compliance.

Indeed, breaking user trust has been shown to have major implications for email providers. This is evidenced by negative public reaction to services like *Google Buzz*, which "automatically searched the user's most emailed contacts and added them as followers, thereby inadvertently exposing potentially sensitive communications" (Nowak, 2010), or Oath (Yahoo mail owner) purportedly selling consumer preferences gleaned from promotional emails to advertisers (Liao, 2010).

Therefore, the tension between the opportunities for novel user experiences and the challenge to preserve user trust is a major recurrent theme that runs throughout the email search and discovery research, and is discussed extensively in this survey. In particular, we dedicate

Chapter 7 to the challenges of privacy-preserving management of user data.

1.2 Email Management and Finding Strategies

In the previous section, we established the scale of email usage, and the importance of mailboxes as personal and organizational information repositories. In this section, we focus on the way that users keep track of and find information in these repositories in real-world settings.

The goal of the majority of email searches is re-finding information in previously seen emails, which relates it to the *known-item search* problem (Craswell *et al.*, 2005), where only one particular, known in advance item can fully satisfy the user information need. It is not surprising, therefore, that emails are frequently revisited, and most of the revisits are information seeking (Alrashed *et al.*, 2018).

Some information types that users seek during email revisits are listed in Table 1.1. Interestingly, finding task-related instructions is the most common reason for email revisit, which is in line with the prevalence of email usage for task management that is noted by other researchers as well (Whittaker, 2005; Lampert *et al.*, 2010).

Table 1.1: Distribution of information types users are looking for in email revisits, as reported in a survey of 395 corporate email users (Alrashed *et al.*, 2018).

Type of Information	Percent
Instructions to perform a certain task	24.1%
A document (e.g., attachment, link)	22.0%
An answer to a question that was previously asked	16.3%
status update	10.2%
A solution to a problem	9.0%
A task request to you	4.9%
A person/customer (e.g., contact information)	2.0%
An appointment/event	2.0%
Machine generated message (e.g., reservation)	0.8%
Other	8.6%

Ai *et al.* (2017) conduct a survey of 324 users to examine what message attributes facilitate searcher recall. They find that, unlike in web search, email searchers tend to remember more details about the

Full text available at: http://dx.doi.org/10.1561/1500000069

provenance of the messages they are interested in (e.g., sender and sent date – see Figure 1.1). This reflects greater familiarity with email than web pages, and re-affirms the known-item approach to email search. Ai *et al.* (2017) also find that this good attribute recall is not always reflected in the search query length and structure. Based on a sample of 2 million queries from Outlook email search logs, they report that advanced syntax is used in only 18% of the queries, and most of these advanced queries contain either `from:` or `to:` filters.

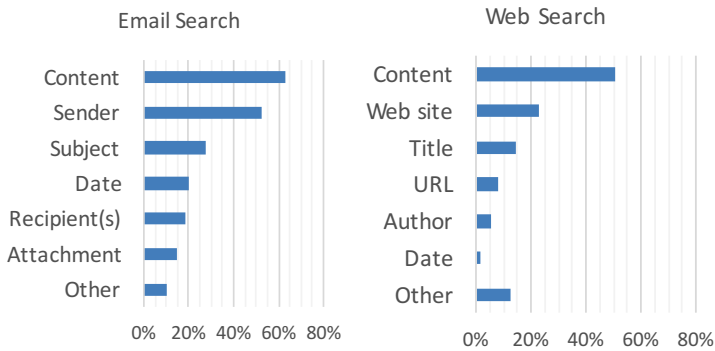

Figure 1.1: Percentage of searchers who remembered certain attributes, compared between email and web searches, based on a survey of 324 regular email users, conducted by Ai *et al.* (2017).

Users may also use other *email discovery* mechanisms beyond search to find the relevant information in their mailboxes. Examples of email discovery mechanisms include content recommendation, classification and information extraction. For instance, some email services can automatically tag emails with labels such as "Travel" or "Finance" (Grbovic *et al.*, 2014) and extract useful information like bill due dates or hotel check-in times (Sheng *et al.*, 2018). This can help with relevant information discovery without the need for conducting an explicit search.

Broadly speaking, most email search and discovery mechanisms discussed in this survey are in the realm of *personal information management* (PIM). PIM studies the organization and maintenance of information items stored for the purpose of completing personal or work-related tasks and activities. In fact, Whittaker *et al.* (2006) argue that email plays a critical role in three key PIM areas, including task management, personal archiving, and contact management.

Full text available at: http://dx.doi.org/10.1561/1500000069

One notable exception to viewing email search and discovery as a sub-field of PIM, is access to mailboxes by third-parties who are not the persons to whom the email was addressed. Such access is conducted in cases such as legal e-discovery for litigation or government investigations (Oard and Webber, 2013), historical research (Task Force on Technical Approaches for Email Archives, 2018), or logging by organizational mail auditing tools (Microsoft 365, 2020). As this survey takes a user-centric approach to email search and discovery, most of these cases are outside of our scope. However, some of the described techniques are likely to be helpful in finding relevant information by third-parties as well.

1.3 Survey Scope and Organization

The majority of the research on email search and discovery that this survey draws upon has appeared over the past decade in a broad spectrum of information retrieval and data mining conferences including (but not limited to)

- ACM SIGIR Conference on Research and Development in Information Retrieval – https://dl.acm.org/conference/sigir
- ACM International Conference on Web Search and Data Mining – https://dl.acm.org/conference/wsdm
- ACM SIGKDD Conference on Knowledge Discovery and Data Mining – https://dl.acm.org/conference/kdd
- The Web Conference (formerly known as International World Wide Web Conference, or WWW) – https://dl.acm.org/conference/www
- The Conference on Information and Knowledge Management – https://dl.acm.org/conference/cikm
- Text REtrieval Conerence – https://trec.nist.gov/

We made our best attempt to provide a comprehensive survey of this large body of research, providing some historical perspective,

Full text available at: http://dx.doi.org/10.1561/1500000069

organizing it into broad themes, and finally suggesting some directions for future research. We also attempted to provide a perspective – based on the existing research, as well as our own experience – on the unique challenges facing the researchers in this field, contrasting it to the more commonly known web search setting.

Prerequisites This survey assumes minimal prior knowledge and should be relatively self-contained. We keep most of the discussions at a high level of abstraction, and refer the readers to the original research papers for technical details. However, some grasp of standard notation, concepts and techniques in information retrieval and machine learning can be beneficial for getting the most out of this survey. We suggest the following introductory and freely available books as useful accompanying references:

- Introduction to Information Retrieval, by Schütze *et al.* (2008)
- The Elements of Statistical Learning, by Hastie *et al.* (2009)

Target Audience We hope that the following audiences will find this survey useful:

- Search practitioners and engineers who want to be exposed to the scientific fundamentals of email search (or other personal search scenarios)
- Industry and academic researchers and graduate students in the fields of information retrieval, machine learning or natural language processing who are interested in better understanding the state-of-the-art and the emerging trends in email search and discovery.

Outline The remainder of this survey is organized as follows. In Chapter 2 we provide a high-level overview of the architecture of a standard email search engine. As there is no previously published work that summarizes such architecture, we do our best to synthesize multiple

Full text available at: http://dx.doi.org/10.1561/1500000069

disparate research avenues, and compare the different design and architecture choices to web search engines, which are likely to be more familiar to our readers.

Chapter 3 is dedicated to the evolution of interfaces for email search and discovery, from manually defined folders and exact search to relevance-based ranking and knowledge panels. In Chapters 4 and 5, we discuss the various aspects of email and query understanding, respectively. In these chapters, due to the heterogeneous topics discussed, we often go beyond the realm of email search and delve into other aspects of email management and discovery, including spam detection, labeling and templatization.

In Chapter 6, we once again broaden our scope beyond search and discuss various assistive applications that allow users to effectively find, manage and create email content. In this chapter, we also often go beyond the boundaries of the mailbox, and discuss how assistance can work across multiple personal content types (e.g., email, calendar entries or personal files).

Chapter 7 is dedicated to management of user data in email search and discovery. We discuss the best practices of privacy-preserving treatment of user data, as well as learning from sparse and biased click data in email search. We speculate on possible future research directions in personal search and discovery in Chapter 8, and conclude the survey in Chapter 9.

Special considerations There are three important considerations that we would like our readers to keep in mind as they make their way through this survey.

First and foremost, data and user privacy is an important *leitmotif* in email search and discovery. Our goal is to elucidate the importance of these topics, and the degree to which they affect how the research in the field is conducted. Therefore, we include a chapter dedicated to privacy-preserving user data management, and return to this topic throughout the survey.

Second, when possible, we try to draw parallels between email and web search. The latter may be a more familiar territory to many of our readers, as it has been one of the focal points of information retrieval

Full text available at: http://dx.doi.org/10.1561/1500000069

research for the past two decades. Contrasting email and web search also aids in highlighting the unique aspects of email search and discovery algorithms.

Finally, the readers are likely to notice that while the title of the survey focuses on email search, some chapters broaden their scope beyond email to other types of personal content, and modes of content management and discovery that go beyond search. This is by design, rather than mere lack of focus. We strongly believe that the future of personal content search and discovery lies in integrative approaches that seamlessly combine personal information across various content silos to best assist the users in completing their personal or work tasks.

Full text available at: http://dx.doi.org/10.1561/1500000069

References

Abadi, M., A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang. (2016). "Deep Learning with Differential Privacy". In: *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. ACM. 308–318.

AbdelRahman, S., B. Hassan, and R. Bahgat. (2010). "A New Email Retrieval Ranking Approach". *International Journal of Computer Science and Information Technology*. 2(5): 44–63. DOI: 10.5121/ijcsit.2010.2504.

Aberdeen, D., O. Pacovsky, and A. Slater. (2010). "The Learning Behind Gmail Priority Inbox". In: *LCCC: NIPS 2010 Workshop on Learning on Cores, Clusters and Clouds*.

Abu-Nimeh, S., D. Nappa, X. Wang, and S. Nair. (2007). "A Comparison of Machine Learning Techniques for Phishing Detection". In: *Proceedings of the Anti-Phishing Working Groups 2nd Annual eCrime Researchers Summit*. ACM. Pittsburgh, PA, USA. 60–69. DOI: 10.1145/1299015.1299021.

Agarwal, A., X. Wang, C. Li, M. Bendersky, and M. Najork. (2019a). "Addressing Trust Bias for Unbiased Learning-to-Rank". In: *Proceedings of the World Wide Web Conference. WWW '19*. San Francisco, CA, USA: Association for Computing Machinery. 4–14. DOI: 10.1145/3308558.3313697.

Full text available at: http://dx.doi.org/10.1561/1500000069

Agarwal, A., I. Zaitsev, X. Wang, C. Li, M. Najork, and T. Joachims. (2019b). "Estimating Position Bias without Intrusive Interventions". In: *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining. WSDM '19*. 474–482.

Aggarwal, C. C. (2005). "On k-anonymity and the Curse of Dimensionality". In: *Proceedings of the 31st International Conference on Very Large Data Bases (VLDB 2005)*. VLDB Endowment. 901–909.

Agichtein, E., E. Brill, and S. Dumais. (2006). "Improving Web Search Ranking by Incorporating User Behavior Information". In: *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '06*. Seattle, Washington, USA: ACM. 19–26. DOI: 10.1145/1148170.1148177.

Ai, Q., K. Bi, C. Luo, J. Guo, and W. B. Croft. (2018). "Unbiased Learning to Rank with Unbiased Propensity Estimation". In: *The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '18*. Ann Arbor, MI, USA: Association for Computing Machinery. 385–394. DOI: 10.1145/3209978.3209986.

Ai, Q., S. T. Dumais, N. Craswell, and D. J. Liebling. (2017). "Characterizing Email Search using Large-scale Behavioral Logs and Surveys". In: *Proceedings of the 26th International Conference on World Wide Web, WWW 2017, Perth, Australia, April 3-7, 2017*. 1511–1520. DOI: 10.1145/3038912.3052615.

Ai, Q., T. Yang, H. Wang, and J. Mao. (2021). "Unbiased Learning to Rank: Online or Offline?" *ACM Transactions on Information Systems (TOIS)*. 39(2): 1–29.

Ailon, N., Z. S. Karnin, E. Liberty, and Y. Maarek. (2013). "Threading Machine Generated Email". In: *Sixth ACM International Conference on Web Search and Data Mining, WSDM 2013, Rome, Italy, February 4-8, 2013*. 405–414. DOI: 10.1145/2433396.2433447.

Alrashed, T., A. H. Awadallah, and S. Dumais. (2018). "The Lifetime of Email Messages: A Large-Scale Analysis of Email Revisitation". In: *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval. CHIIR '18*. New Brunswick, NJ, USA: ACM. 120–129. DOI: 10.1145/3176349.3176398.

Full text available at: http://dx.doi.org/10.1561/1500000069

Anguera, X., J. Xu, and N. Oliver. (2008). "Multimodal Photo Annotation and Retrieval on a Mobile Phone". In: *Proceedings of the 1st ACM International Conference on Multimedia Information Retrieval*. ACM. 188–194.

Apple. (2017). "Learning with Privacy at Scale". URL: https://machinelearning.apple.com/docs/learning-with-privacy-at-scale/appledifferentialprivacysystem.pdf.

Arguello, J., F. Diaz, and J. Callan. (2011). "Learning to Aggregate Vertical Results into Web Search Results". In: *Proceedings of the 20th ACM International Conference on Information and Knowledge Management. CIKM '11*. Glasgow, Scotland, UK: ACM. 201–210. DOI: 10.1145/2063576.2063611.

Arya, D., V. Ha-Thuc, and S. Sinha. (2015). "Personalized Federated Search at LinkedIn". In: *Proceedings of the 24th ACM International Conference on Information and Knowledge Management*. ACM. 1699–1702.

Asadi, N., D. Metzler, T. Elsayed, and J. Lin. (2011). "Pseudo Test Collections for Learning Web Search Ranking Functions". In: *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '11*. Beijing, China: ACM. 1073–1082. DOI: 10.1145/2009916.2010058.

Ashkan, A. and D. Metzler. (2019). "Revisiting Online Personal Search Metrics with the User in Mind". In: *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2019)*. Paris, France: Association for Computing Machinery. 625–634. DOI: 10.1145/3331184.3331266.

Aumüller, M., E. Bernhardsson, and A. Faithfull. (2017). "ANN-benchmarks: A benchmarking tool for approximate nearest neighbor algorithms". In: *International Conference on Similarity Search and Applications*. Springer. 34–49.

Avigdor-Elgrabli, N., M. Cwalinski, D. Di Castro, I. Gamzu, I. Grabovitch-Zuyev, L. Lewin-Eytan, and Y. Maarek. (2016). "Structural Clustering of Machine-Generated Mail". In: *Proceedings of the 25th ACM International Conference on Information and Knowledge Management. CIKM '16*. Indianapolis, Indiana, USA: ACM. 217–226. DOI: 10.1145/2983323.2983350.

Full text available at: http://dx.doi.org/10.1561/1500000069

Azarbonyad, H., R. Sim, and R. W. White. (2019). "Domain Adaptation for Commitment Detection in Email". In: *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*. ACM. 672–680.

Balog, K. and T. Kenter. (2019). "Personal Knowledge Graphs: A Research Agenda". In: *Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval*. ACM. 217–220.

Bar-Yossef, Z., I. Guy, R. Lempel, Y. S. Maarek, and V. Soroka. (2006). "Cluster Ranking with an Application to Mining Mailbox Networks". In: *Sixth International Conference on Data Mining (ICDM'06)*. 63–74. DOI: 10.1109/ICDM.2006.35.

Bawa, M., R. J. Bayardo Jr, R. Agrawal, and J. Vaidya. (2009). "Privacy-preserving Indexing of Documents on the Network". *The VLDB Journal—The International Journal on Very Large Data Bases*. 18(4): 837–856.

Bekkerman, R. (2004). "Automatic Categorization of Email into Folders: Benchmark Experiments on Enron and SRI Corpora". *Tech. rep.* No. 218. Computer Science Department Faculty Publication Series, University of Massachusetts Amherst.

Bell, R. M. and Y. Koren. (2007). "Lessons from the Netflix Prize Challenge." *SiGKDD Explorations*. 9(2): 75–79.

Bendersky, M., X. Wang, D. Metzler, and M. Najork. (2017). "Learning from User Interactions in Personal Search via Attribute Parameterization". In: *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. ACM. 791–799.

Bendersky, M., X. Wang, M. Najork, and D. Metzler. (2018). "Learning with Sparse and Biased Feedback for Personal Search". In: *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI 2018)*. Stockholm, Sweden. 5219–5223. DOI: 10.24963/ijcai.2018/725.

Bengio, Y., R. Ducharme, P. Vincent, and C. Jauvin. (2003). "A Neural Probabilistic Language Model". *Journal of Machine Learning Research*. 3(Feb): 1137–1155.

Full text available at: http://dx.doi.org/10.1561/1500000069

Bennett, P. N. and J. G. Carbonell. (2007). "Combining Probability-Based Rankers for Action-Item Detection". In: *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference*. Rochester, New York: Association for Computational Linguistics. 324–331. URL: https://www.aclweb.org/anthology/N07-1041.

Berry, M. W., M. Browne, and B. Signer. (2001). "Topic annotated Enron email data set". *Philadelphia: Linguistic Data Consortium*.

Bhole, A. and R. Udupa. (2015). "On Correcting Misspelled Queries in Email Search." In: *Proceedings of the 29th AAAI Conference on Artificial Intelligence*.

Bhushan, A., K. Pogran, R. Tomlinson, and J. White. (1973). "Standardizing Network Mail Headers". URL: https://tools.ietf.org/html/rfc561.

Birrell, A. D., E. P. Wobber, and M. D. Schroeder. (1997). *Pachyderm*. URL: http://birrell.org/andrew/pachywww/.

Brin, S. and L. Page. (1998). "The Anatomy of a Large-scale Hypertextual Web Search Engine". *Computer Networks and ISDN Systems*. 30(1-7): 107–117.

Broder, A. Z., N. Eiron, M. Fontoura, M. Herscovici, R. Lempel, J. McPherson, R. Qi, and E. Shekita. (2006). "Indexing Shared Content in Information Retrieval Systems". In: *Advances in Database Technology - EDBT 2006*. Ed. by Y. Ioannidis, M. H. Scholl, J. W. Schmidt, F. Matthes, M. Hatzopoulos, K. Boehm, A. Kemper, T. Grust, and C. Boehm. Berlin, Heidelberg: Springer Berlin Heidelberg. 313–330.

Burges, C. J. (2010). "From RankNet to LambdaRank to LambdaMART: An overview". *Learning*. 11(23–581): 81.

Carenini, G., R. T. Ng, and X. Zhou. (2007). "Summarizing Email Conversations with Clue Words". In: *Proceedings of the 16th International Conference on World Wide Web (WWW 2007)*. Banff, Alberta, Canada: Association for Computing Machinery. 91–100. DOI: 10.1145/1242572.1242586.

Full text available at: http://dx.doi.org/10.1561/1500000069

Carlini, N., C. Liu, Ú. Erlingsson, J. Kos, and D. Song. (2019). "The Secret Sharer: Evaluating and Testing Unintended Memorization in Neural Networks". In: *Proceedings of the 28th USENIX Conference on Security Symposium (SEC 2019)*. 267–284.

Carmel, D., G. Halawi, L. Lewin-Eytan, Y. Maarek, and A. Raviv. (2015). "Rank by Time or by Relevance?: Revisiting Email Search". In: *Proceedings of the 24th ACM International Conference on Information and Knowledge Management*. ACM. 283–292.

Carmel, D., L. Lewin-Eytan, A. Libov, Y. Maarek, and A. Raviv. (2017a). "Promoting Relevant Results in Time-Ranked Mail Search". In: *Proceedings of the 26th International Conference on World Wide Web. WWW '17*. Perth, Australia: International World Wide Web Conferences Steering Committee. 1551–1559. DOI: 10.1145/3038912.3052659.

Carmel, D., L. Lewin-Eytan, A. Libov, Y. Maarek, and A. Raviv. (2017b). "The Demographics of Mail Search and Their Application to Query Suggestion". In: *Proceedings of the 26th International Conference on World Wide Web (WWW 2017)*. International World Wide Web Conferences Steering Committee. Perth, Australia: Association for Computing Machinery. 1541–1549. DOI: 10.1145/3038912.3052658.

Caruana, R. (1997). "Multitask Learning". *Machine Learning*. 28(1): 41–75.

Carvalho, V. R. and W. W. Cohen. (2004). "Learning to Extract Signature and Reply Lines from Email". In: *Proceedings of the Conference on Email and Anti-Spam*. Vol. 2004.

Carvalho, V. R. and W. W. Cohen. (2007). "Preventing Information Leaks in Email". In: *Proceedings of the 2007 SIAM International Conference on Data Mining*. SIAM. 68–77.

Carvalho, V. R. and W. W. Cohen. (2008). "Ranking Users for Intelligent Message Addressing". In: *European Conference on Information Retrieval*. Springer. 321–333.

Full text available at: http://dx.doi.org/10.1561/1500000069

Chen, F., K. Xia, K. Dhabalia, and J. I. Hong. (2019a). "MessageOnTap: A Suggestive Interface to Facilitate Messaging-related Tasks". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI 2019). CHI '19*. Glasgow, Scotland UK: Association for Computing Machinery. 575:1–575:14. DOI: 10.1145/3290605.3300805.

Chen, M. X., B. N. Lee, G. Bansal, Y. Cao, S. Zhang, J. Lu, J. Tsay, Y. Wang, A. M. Dai, Z. Chen, T. Sohn, and Y. Wu. (2019b). "Gmail Smart Compose: Real-Time Assisted Writing". In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. KDD '19*. Anchorage, AK, USA: Association for Computing Machinery. 2287–2295. DOI: 10.1145/3292500.3330723.

Chui, M., J. Manyika, J. Bughin, R. Dobbs, C. Roxburgh, H. Sarrazin, G. Sands, and M. Westergren. (2012). "The Social Economy: Unlocking Value and Productivity Through Social Technologies". *Tech. rep.* McKinsey Global Institute. URL: https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/the-social-economy.

Chuklin, A., I. Markov, and M. de Rijke. (2015). *Click Models for Web Search*. Morgan & Claypool. DOI: 10.2200/S00654ED1V01Y201507ICR043.

Cincotta, A. V. (1983). "Navy Electronic Mail Service User's Guide." *Tech. rep.* No. CMLD-83-22. Bethesa, MD 20084: David W. Taylor Naval Ship Research and Development Center.

Claude, F., A. Fariña, M. A. Martínez-Prieto, and G. Navarro. (2011). "Indexes for Highly Repetitive Document Collections". In: *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*. ACM. 463–468.

Cohen, W. W. *et al.* (1996). "Learning Rules That Classify E-mail". In: *AAAI Spring Symposium on Machine Learning in Information Access*. Vol. 18. Stanford, CA. 25.

Cohen, W. W., V. R. Carvalho, and T. M. Mitchell. (2004). "Learning to Classify Email into "Speech Acts"". In: *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*. 309–316.

Full text available at: http://dx.doi.org/10.1561/1500000069

Cohen, W. W. (2015). "Enron Email Dataset". URL: http://www.cs.cmu.edu/~enron/.

Cormack, G. V. (2007). "TREC 2007 Spam Track Overview". In: *Proceedings of the Sixteenth Text REtrieval Conference (TREC 2007)*.

Crammer, K., A. Kulesza, and M. Dredze. (2009). "Adaptive Regularization of Weight Vectors". In: *Advances in neural information processing systems*. 414–422.

Craswell, N., A. P. de Vries, and I. Soboroff. (2005). "Overview of the TREC 2005 Enterprise Track". In: *Proceedings of the Fourteenth Text REtrieval Conference (TREC 2005)*. Vol. 5. 199–205.

Cui, Q., G.-V. Jourdan, G. V. Bochmann, R. Couturier, and I.-V. Onut. (2017). "Tracking Phishing Attacks Over Time". In: *Proceedings of the 26th International Conference on World Wide Web. WWW '17*. Perth, Australia: International World Wide Web Conferences Steering Committee. 667–676. DOI: 10.1145/3038912.3052654.

Culpepper, J. S., F. Diaz, and M. D. Smucker. (2018). "Report from the Third Strategic Workshop on Information Retrieval (SWIRL)".

Cutrell, E., S. T. Dumais, and J. Teevan. (2006). "Searching to Eliminate Personal Information Management". *Commun. ACM*. 49(1): 58–64.

Dabbish, L. A., R. E. Kraut, S. Fussell, and S. Kiesler. (2005). "Understanding Email Use: Predicting Action on a Message". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. 691–700.

Dai, N., M. Shokouhi, and B. D. Davison. (2011). "Learning to Rank for Freshness and Relevance". In: *Proceedings of the 34th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. URL: https://www.microsoft.com/en-us/research/publication/learning-to-rank-for-freshness-and-relevance/.

Di Castro, D., I. Gamzu, I. Grabovitch-Zuyev, L. Lewin-Eytan, A. Pundir, N. R. Sahoo, and M. Viderman. (2018). "Automated Extractions for Machine Generated Mail". In: *Companion Proceedings of the The Web Conference 2018. WWW '18*. Lyon, France: International World Wide Web Conferences Steering Committee. 655–662. DOI: 10.1145/3184558.3186582.

Full text available at: http://dx.doi.org/10.1561/1500000069

Di Castro, D., Z. Karnin, L. Lewin-Eytan, and Y. Maarek. (2016a). "You've got Mail, and Here is What you Could do With It!: Analyzing and Predicting Actions on Email Messages". In: *Proceedings of the 9th ACM International Conference on Web Search and Data Mining (WSDM 2016)*. San Francisco, CA, USA: ACM. 307–316. DOI: 10.1145/2835776.2835811.

Di Castro, D., L. Lewin-Eytan, Y. Maarek, R. Wolff, and E. Zohar. (2016b). "Enforcing k-anonymity in Web Mail Auditing". In: *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*. ACM. 327–336.

Dumais, S., E. Cutrell, J. Cadiz, G. Jancke, R. Sarin, and D. C. Robbins. (2003). "Stuff I've Seen: A System for Personal Information Retrieval and Re-use". In: *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM. 72–79.

Dwork, C. (2008). "Differential Privacy: A Survey of Results". In: *Proceedings of the 5th International Conference on Theory and Applications of Models of Computation (TAMC 2008)*. Vol. 4978. *Lecture Notes in Computer Science*. Xi'an, China: Springer Verlag. 1–19. URL: https://www.microsoft.com/en-us/research/publication/differential-privacy-a-survey-of-results/.

Elsayed, T. and D. W. Oard. (2006). "Modeling Identity in Archival Collections of Email: A Preliminary Study." In: *Proceedings of the 3rd Conference on Email and Anti-Spam (CEAS 2006)*. 95–103.

European Parliament and Council of the European Union. (2016). "General Data Protection Regulation". URL: http://data.europa.eu/eli/reg/2016/679/oj.

Fang, Z., A. Agarwal, and T. Joachims. (2019). "Intervention Harvesting for Context-Dependent Examination-Bias Estimation". In: *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2019)*. Paris, France: Association for Computing Machinery. 825–834. DOI: 10.1145/3331184.3331238.

Full text available at: http://dx.doi.org/10.1561/1500000069

Feild, H. A., J. Allan, and J. Glatt. (2011). "CrowdLogging: distributed, private, and anonymous search logging". In: *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. 375–384.

Fette, I., N. Sadeh, and A. Tomasic. (2007). "Learning to Detect Phishing Emails". In: *Proceedings of the 16th International Conference on World Wide Web (WWW 2007)*. Banff, Alberta, Canada: ACM. 649–656. DOI: 10.1145/1242572.1242660.

Foley, J., M. Zhang, M. Bendersky, and M. Najork. (2018). "Semantic Location in Email Query Suggestion". In: *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR 2018, Ann Arbor, MI, USA, July 08-12, 2018*. 977–980. DOI: 10.1145/3209978.3210116.

Gamzu, I., Z. S. Karnin, Y. Maarek, and D. Wajc. (2015). "You Will Get Mail! Predicting the Arrival of Future Email". In: *Proceedings of the 24th International Conference on World Wide Web Companion, WWW 2015, Florence, Italy, May 18-22, 2015 - Companion Volume*. 1327–1332. DOI: 10.1145/2740908.2741694.

Gamzu, I., L. Lewin-Eytan, and N. Silberstein. (2018). "Unsubscription: A Simple Way to Ease Overload in Email". In: *Proceedings of the 11th ACM International Conference on Web Search and Data Mining (WSDM 2018)*. Marina Del Rey, CA, USA: ACM. 189–197. DOI: 10.1145/3159652.3159698.

Ganin, Y., E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky. (2016). "Domain-Adversarial Training of Neural Networks". *The Journal of Machine Learning Research*. 17(1): 2096–2030.

Gao, N., M. Dredze, and D. W. Oard. (2016). "Knowledge base population for organization mentions in email". In: *Proceedings of the 5th Workshop on Automated Knowledge Base Construction*. 24–28.

Gibney, E. (2017). "Ethics of Internet Research Trigger Scrutiny". *Nature News*. 550(7674): 16.

Gmail Help. (2018). "Search operators you can use with Gmail". URL: https://support.google.com/mail/answer/7190?hl=en.

Google Search. (2018). "How Search Organizes Information". URL: https://www.google.com/search/howsearchworks/crawling-indexing/.

Full text available at: http://dx.doi.org/10.1561/1500000069

Graham, P. (2003). "Better Bayesian Filtering". URL: http://www.paulgraham.com/better.html.

Grammarly. (2018). "5 Ways Grammarly Helps You Learn While You Write". URL: https://www.grammarly.com/blog/grammarly-helps-you-learn/.

Graus, D., P. N. Bennett, R. W. White, and E. Horvitz. (2016). "Analyzing and Predicting Task Reminders". In: *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization. UMAP '16*. Halifax, Nova Scotia, Canada: ACM. 7–15. DOI: 10.1145/2930238.2930239.

Grbovic, M., G. Halawi, Z. S. Karnin, and Y. Maarek. (2014). "How Many Folders Do You Really Need?: Classifying Email into a Handful of Categories". In: *Proceedings of the 23rd ACM International Conference on Information and Knowledge Management, CIKM 2014, Shanghai, China, November 3-7, 2014*. 869–878. DOI: 10.1145/2661829.2662018.

Gupta, J., Z. Qin, M. Bendersky, and D. Metzler. (2019a). "Personalized Online Spell Correction for Personal Search". In: *The World Wide Web Conference. WWW '19*. San Francisco, CA, USA: ACM. 2785–2791. DOI: 10.1145/3308558.3313706.

Gupta, R., R. Kondapally, and S. Guha. (2019b). "Large-Scale Information Extraction from Emails with Data Constraints". In: *International Conference on Big Data Analytics*. Springer. 124–139.

Hangal, S., M. S. Lam, and J. Heer. (2011). "MUSE: Reviving Memories Using Email Archives". In: *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology. UIST '11*. Santa Barbara, California, USA: ACM. 75–84. DOI: 10.1145/2047196.2047206.

Hastie, T., R. Tibshirani, and J. Friedman. (2009). *The Elements of Statistical Learning*. Springer. DOI: 10.1007/978-0-387-84858-7.

Hawking, D. (2010). "Enterprise Search". In: *Modern Information Retrieval, 2nd Edition*. Ed. by R. Baeza-Yates and B. Ribeiro-Neto. Addison-Wesley. 645–686. URL: http://david-hawking.net/pubs/ModernIR2_Hawking_chapter.pdf.

Full text available at: http://dx.doi.org/10.1561/1500000069

He, J., J. Zeng, and T. Suel. (2010). "Improved Index Compression Techniques for Versioned Document Collections". In: *Proceedings of the 19th ACM International Conference on Information and Knowledge Management. CIKM '10*. Toronto, ON, Canada: ACM. 1239–1248. DOI: 10.1145/1871437.1871594.

Heer, J. (2005). "Exploring Enron: A Sketch of Visual Data Mining of Email". In: *Email Archive Visualization Workshop, University of Maryland*.

Horovitz, M., L. Lewin-Eytan, A. Libov, Y. Maarek, and A. Raviv. (2017). "Mailbox-Based vs. Log-Based Query Completion for Mail Search". In: *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM. 937–940.

Hu, J., R. J. Passonneau, and O. Rambow. (2009). "Contrasting the interaction structure of an email and a telephone corpus: A machine learning approach to annotation of dialogue function units". In: *Proceedings of the SIGDIAL 2009 Conference*. 357–366.

Jiang, L., Y. Kalantidis, L. Cao, S. Farfade, J. Tang, and A. G. Hauptmann. (2017). "Delving Deep into Personal Photo and Video Search". In: *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. ACM. 801–810.

Joachims, T. (2002). "Optimizing Search Engines Using Clickthrough Data". In: *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM. 133–142.

Joachims, T., L. Granka, B. Pan, H. Hembrooke, F. Radlinski, and G. Gay. (2007). "Evaluating the Accuracy of Implicit Feedback from Clicks and Query Reformulations in Web Search". *ACM Trans. Inf. Syst.* 25(2). DOI: 10.1145/1229179.1229181.

Joachims, T., A. Swaminathan, and T. Schnabel. (2017). "Unbiased Learning-to-rank With Biased Feedback". In: *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. ACM. 781–789.

Full text available at: http://dx.doi.org/10.1561/1500000069

Kamvar, M., M. Kellar, R. Patel, and Y. Xu. (2009). "Computers and iPhones and Mobile Phones, Oh My!: A Logs-based Comparison of Search Users on Different Devices". In: *Proceedings of the 18th international conference on World wide web*. 801–810.

Kannan, A., K. Kurach, S. Ravi, T. Kaufmann, A. Tomkins, B. Miklos, G. Corrado, L. Lukács, M. Ganea, P. Young, and V. Ramavajjala. (2016). "Smart Reply: Automated Response Suggestion for Email". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*. 955–964. DOI: 10.1145/2939672.2939801.

Kaushal, G. (2016). "Inbox by Gmail: Find Answers Even Faster". URL: https://www.blog.google/products/gmail/inbox-by-gmail-find-answers-even-faster/.

Kiritchenko, S. and S. Matwin. (2001). "Email Classification with Co-training". In: *Proceedings of the 2001 Conference of the Center for Advanced Studies on Collaborative Research (CASCON 2001)*. Toronto, Ontario, Canada: IBM Press. 301–312. URL: http://dl.acm.org/citation.cfm?id=782096.782104.

Kleinberg, J. M. (1999). "Authoritative Sources in a Hyperlinked Environment". *J. ACM*. 46(5): 604–632. DOI: 10.1145/324133.324140.

Kokkalis, N., T. Köhn, C. Pfeiffer, D. Chornyi, M. S. Bernstein, and S. R. Klemmer. (2013). "EmailValet: Managing Email Overload Through Private, Accountable Crowdsourcing". In: *Proceedings of the 2013 conference on Computer supported cooperative work*. ACM. 1291–1300.

Konečný, J., H. B. McMahan, D. Ramage, and P. Richtarik. (2016). "Federated Optimization: Distributed Machine Learning for On-Device Intelligence". arXiv: 1610.02527 [cs.LG].

Koren, Y., E. Liberty, Y. Maarek, and R. Sandler. (2011). "Automatically Tagging Email by Leveraging Other Users' Folders". In: *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '11*. San Diego, California, USA: ACM. 913–921. DOI: 10.1145/2020408.2020560.

Kruschwitz, U. and C. Hull. (2017). "Searching the Enterprise". *Foundations and Trends® in Machine Learning*. 11(1): 1–142. DOI: 10.1561/1500000053.

Full text available at: http://dx.doi.org/10.1561/1500000069

Kulkarni, A., J. Teevan, K. M. Svore, and S. T. Dumais. (2011). "Understanding Temporal Query Dynamics". In: *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining. WSDM '11.* Hong Kong, China: ACM. 167–176. DOI: 10.1145/1935826.1935862.

Kuzi, S., D. Carmel, A. Libov, and A. Raviv. (2017). "Query Expansion for Email Search". In: *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2017).* ACM. Shinjuku, Tokyo, Japan. 849–852. DOI: 10.1145/3077136.3080660.

Kveton, B., C. Szepesvari, Z. Wen, and A. Ashkan. (2015). "Cascading Bandits: Learning to Rank in the Cascade Model". In: *International Conference on Machine Learning.* 767–776.

Lam, D. S. (2002). "Exploiting E-mail Structure to Improve Summarization". *PhD thesis.* Massachusetts Institute of Technology.

Lampert, A., R. Dale, and C. Paris. (2009). "Segmenting Email Message Text into Zones". In: *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing.* Singapore: Association for Computational Linguistics. 919–928. URL: https://www.aclweb.org/anthology/D09-1096.

Lampert, A., R. Dale, and C. Paris. (2010). "Detecting Emails Containing Requests for Action". In: *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics.* Los Angeles, California: Association for Computational Linguistics. 984–992. URL: https://www.aclweb.org/anthology/N10-1142.

Lempel, R. and S. Moran. (2001). "SALSA: The Stochastic Approach for Link-Structure Analysis". *ACM Trans. Inf. Syst.* 19(2): 131–160. DOI: 10.1145/382979.383041.

Li, C., M. Zhang, M. Bendersky, H. Deng, D. Metzler, and M. Najork. (2019a). "Multi-view Embedding-based Synonyms for Personal Search". In: *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2019).* Paris, France: ACM. 575–584. DOI: 10.1145/3331184.3331250.

Full text available at: http://dx.doi.org/10.1561/1500000069

Li, P., Z. Qin, X. Wang, and D. Metzler. (2019b). "Combining Decision Trees and Neural Networks for Learning-to-Rank in Personal Search". In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '19*. Anchorage, AK, USA: Association for Computing Machinery. 2032–2040. DOI: 10.1145/3292500.3330676.

Liang, J., L. Jiang, L. Cao, Y. Kalantidis, L. Li, and A. G. Hauptmann. (2019). "Focal Visual-Text Attention for Memex Question Answering". *IEEE Transactions on Pattern Analysis and Machine Intelligence*: 1–1. DOI: 10.1109/TPAMI.2018.2890628.

Liao, S. (2010). "Yahoo Mail is still scanning your emails for data to sell to advertisers". URL: https://www.theverge.com/2018/8/28/17792522/yahoo-mail-email-scan-data-advertisers-opt-out.

Lin, C.-C., D. Kang, M. Gamon, and P. Pantel. (2018). "Actionable Email Intent Modeling with Reparametrized RNNs". In: *Thirty-Second AAAI Conference on Artificial Intelligence*.

Linden, G., B. Smith, and J. York. (2003). "Amazon.com Recommendations: Item-to-item Collaborative Filtering". *IEEE Internet Computing*. 7(1): 76–80.

Litmus Email Analytics. (2019). "The 2019 Email Client Market Share". URL: https://litmus.com/blog/infographic-the-2019-email-client-market-share.

Littlestone, N. (1988). "Learning Quickly when Irrelevant Attributes Abound: A New Linear-threshold Algorithm". *Machine Learning*. 2(4): 285–318.

Maarek, Y. (2017). "Web Mail is not Dead!: It's Just Not Human Anymore". In: *Proceedings of the 26th International Conference on World Wide Web (WWW 2017)*. International World Wide Web Conferences Steering Committee. Perth, Australia: Association for Computing Machinery. 5. DOI: 10.1145/3038912.3050916.

McMahan, B. and D. Ramage. (2017). "Federated Learning: Collaborative Machine Learning without Centralized Training Data". URL: https://ai.googleblog.com/2017/04/federated-learning-collaborative.html.

Full text available at: http://dx.doi.org/10.1561/1500000069

Mei, Q. and K. Church. (2008). "Entropy of Search Logs: How Hard is Search? With Personalization? With Backoff?" In: *Proceedings of the 2008 International Conference on Web Search and Data Mining (WSDM 2008)*. Palo Alto, California, USA: Association for Computing Machinery. 45–54. DOI: 10.1145/1341531.1341540.

Microsoft 365. (2020). *Manage mailbox auditing*. URL: https://docs.microsoft.com/en-us/microsoft-365/compliance/enable-mailbox-auditing.

Mikolov, T., M. Karafiát, L. Burget, J. Černocký, and S. Khudanpur. (2010). "Recurrent Neural Network Based Language Model". In: *11th Annual Conference of the International Speech Communication Association (INTERSPEECH 2010)*. Makuhari, Chiba, Japan.

Minkov, E., R. C. Wang, and W. W. Cohen. (2005). "Extracting Personal Names from Email: Applying Named Entity Recognition to Informal Text". In: *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*. Association for Computational Linguistics. 443–450.

Mitra, B. and N. Craswell. (2018). "An Introduction to Neural Information Retrieval". *Foundations and Trends® in Information Retrieval*. 13(1): 1–126.

Mitra, B., E. Nalisnick, N. Craswell, and R. Caruana. (2016). "A Dual Embedding Space Model for Document Ranking". arXiv: 1602.01137 [cs.IR].

Mohammad, S. M. and T. Yang. (2011). "Tracking Sentiment in Mail: How Genders Differ on Emotional Axes". In: *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis. WASSA '11*. Portland, Oregon, USA: Association for Computational Linguistics. 70–79. URL: http://dl.acm.org/citation.cfm?id=2107653.2107662.

Mukherjee, S., S. (Mukherjee, M. Hasegawa, A. Hassan Awadallah, and R. W. White. (2020). "Smart To-Do : Automatic Generation of To-Do List from Emails". In: *Annual Conference of the Association for Computational Linguistics (ACL 2020)*.

Najork, M. (2009). "Web Spam Detection." *Encyclopedia of Database Systems*. 1: 3520–3523.

Full text available at: http://dx.doi.org/10.1561/1500000069

Najork, M. (2018). "Training On-Device Ranking Models from Cross-User Interactions in a Privacy-Preserving Fashion". In: *Proceedings of the First Biennial Conference on Design of Experimental Search & Information Retrieval Systems, Bertinoro, Italy, August 28-31, 2018*. 108. URL: http://ceur-ws.org/Vol-2167/short11.pdf.

Najork, M. A. (2007). "Comparing the Effectiveness of Hits and Salsa". In: *Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management. CIKM '07*. Lisbon, Portugal: ACM. 157–164. DOI: 10.1145/1321440.1321465.

Naragon, K. (2018). "We Still Love Email, But We're Spreading the Love with Other Channels". URL: https://theblog.adobe.com/love-email-but-spreading-the-love-other-channels/.

Narang, K., S. T. Dumais, N. Craswell, D. Liebling, and Q. Ai. (2017). "Large-Scale Analysis of Email Search and Organizational Strategies". In: *Proceedings of the 2017 Conference on Computer Human Information Interaction and Retrieval. CHIIR '17*. Oslo, Norway: ACM. 215–223. DOI: 10.1145/3020165.3020175.

Naranyanan, A. and V. Shmatikov. (2008). "Robust De-anonymization of Large Datasets". In: *Proceedings of the 2008 IEEE Symposium on Security and Privacy, May 2008*.

Nowak, P. (2010). "Privacy commissioner reviewing Google Buzz". URL: https://www.cbc.ca/news/technology/privacy-commissioner-reviewing-google-buzz-1.899267.

Oard, D., W. Webber, D. Kirsch, and S. Golitsynskiy. (2015). "Avocado Research Email Collection". *Philadelphia: Linguistic Data Consortium*.

Oard, D. W. and W. Webber. (2013). "Information retrieval for e-discovery". *Foundations and Trends in Information Retrieval*. 7(2-3): 99–237.

Ogilvie, P. and J. Callan. (2005). "Experiments with Language Models for Known-Item Finding of E-mail Messages". In: *Proceedings of the Fourteenth Text REtrieval Conference (TREC 2005)*. National Institute of Standards and Technology. Gaithersburg, Maryland, USA.

Full text available at: http://dx.doi.org/10.1561/1500000069

Olston, C. and M. Najork. (2010). "Web Crawling". *Foundations and Trends® in Machine Learning*. 4(3): 175–246. DOI: 10.1561/1500000017.

Oosterhuis, H. and M. de Rijke. (2018). "Ranking for Relevance and Display Preferences in Complex Presentation Layouts". In: *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. ACM. 845–854.

Oosterhuis, H. and M. de Rijke. (2021). "Unifying Online and Counterfactual Learning to Rank: A Novel Counterfactual Estimator that Effectively Utilizes Online Interventions". In: *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*. 463–471.

Pantel, P. and D. Lin. (1998). "Spamcop: A Spam Classification & Organization Program". In: *Proceedings of AAAI-98 Workshop on Learning for Text Categorization*. 95–98.

Perronnin, F. C. (2009). "Statistical Language-model Based System for Detection of Missing Attachments".

Radlinski, F., M. Kurup, and T. Joachims. (2008). "How Does Clickthrough Data Reflect Retrieval Quality?" In: *Proceedings of the 17th ACM Conference on Information and Knowledge Management (CIKM 2008)*. Napa Valley, California, USA: ACM. 43–52. DOI: 10.1145/1458082.1458092.

Ramarao, P., S. Iyengar, P. Chitnis, R. Udupa, and B. Ashok. (2016). "InLook: Revisiting Email Search Experience". In: *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '16*. Pisa, Italy: ACM. 1117–1120. DOI: 10.1145/2911451.2911458.

Rambow, O., L. Shrestha, J. Chen, and C. Lauridsen. (2004). "Summarizing Email Threads". In: *Proceedings of HLT-NAACL 2004*. USA. 105–108.

Ramzan, Z. (2010). "Phishing Attacks and Countermeasures". In: *Handbook of Information and Communication Security*. Springer. 433–448.

Ravi, S. and Q. Diao. (2016). "Large Scale Distributed Semi-supervised Learning Using Streaming Approximation". In: *Artificial Intelligence and Statistics*. 519–528.

Full text available at: http://dx.doi.org/10.1561/1500000069

Rogers, S. K. (2016). "Item-to-item Recommendations at Pinterest". In: *Proceedings of the 10th ACM Conference on Recommender Systems*. ACM. 393–393.

Sahami, M., S. Dumais, D. Heckerman, and E. Horvitz. (1998). "A Bayesian Approach to Filtering Junk E-mail". In: *Proceedings of AAAI-98 Workshop on Learning for Text Categorization*. Vol. 62. Madison, Wisconsin. 98–105.

Samarati, P. and L. Sweeney. (1998). "Protecting Privacy When Disclosing Information: k-anonymity and its Enforcement Through Generalization and Suppression". *Tech. rep.* technical report, SRI International.

Santos, R. L. T., C. Macdonald, and I. Ounis. (2015). "Search Result Diversification". *Foundations and Trends® in Information Retrieval*. 9(1): 1–90. DOI: 10.1561/1500000040.

Sayed, M. F., W. Cox, J. L. Rivera, C. Christian-Lamb, M. Iqbal, D. W. Oard, and K. Shilton. (2020). "A Test Collection for Relevance and Sensitivity". In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1605–1608.

Schütze, H., C. D. Manning, and P. Raghavan. (2008). *Introduction to Information Retrieval*. Vol. 39. Cambridge University Press.

Sculley, D. and G. Wachman. (2007). "Relaxed Online SVMs in the TREC Spam Filtering Track." In: *Proceedings of the Sixteenth Text REtrieval Conference (TREC 2007)*.

Segal, R. B. and J. O. Kephart. (2000). "Swiftfile: An Intelligent Assistant for Organizing E-mail". In: *AAAI 2000 Spring Symposium on Adaptive User Interfaces*. Palo Alto, CA, USA: AAAI. 107–112.

Shao, J., S. Ji, and T. Yang. (2019). "Privacy-aware Document Ranking with Neural Signals". In: *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2019)*. Paris, France: Association for Computing Machinery. 305–314. DOI: 10.1145/3331184.3331189.

Full text available at: http://dx.doi.org/10.1561/1500000069

Shen, J., M. Karimzadehgan, M. Bendersky, Z. Qin, and D. Metzler. (2018). "Multi-task Learning for Email Search Ranking with Auxiliary Query Clustering". In: *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. ACM. 2127–2135.

Shen, J., O. Brdiczka, and J. Liu. (2013). "Understanding Email Writers: Personality Prediction from Email Messages". In: *Proceedings of the 21st International Conference on User Modeling, Adaptation, and Personalization (UMAP 2013)*. Ed. by S. Carberry, S. Weibelzahl, A. Micarelli, and G. Semeraro. Vol. 7899. *Lecture Notes in Computer Science*. Rome, Italy: Springer. 318–330. DOI: https://doi.org/10.1007/978-3-642-38844-6_29.

Sheng, Y., S. Tata, J. B. Wendt, J. Xie, Q. Zhao, and M. Najork. (2018). "Anatomy of a Privacy-Safe Large-Scale Information Extraction System Over Email". In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD 2018)*. London, United Kingdom: ACM. 734–743. DOI: 10.1145/3219819.3219901.

Silvestri, F. (2010). "Mining Query Logs: Turning Search Usage Data into Knowledge". *Foundations and Trends® in Machine Learning*. 4(1 & 2): 1–174. DOI: 10.1561/1500000013.

Singh, A. and T. Joachims. (2018). "Fairness of Exposure in Rankings". In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM. 2219–2228.

Sordoni, A., M. Galley, M. Auli, C. Brockett, Y. Ji, M. Mitchell, J.-Y. Nie, J. Gao, and B. Dolan. (2015). "A Neural Network Approach to Context-Sensitive Generation of Conversational Responses". In: *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Denver, Colorado: Association for Computational Linguistics. 196–205. DOI: 10.3115/v1/N15-1020.

Spirin, N. and J. Han. (2012). "Survey on Web Spam Detection: Principles and Algorithms". *ACM SIGKDD Explorations Newsletter*. 13(2): 50–64.

Full text available at: http://dx.doi.org/10.1561/1500000069

Sun, Y., L. Garcia-Pueyo, J. B. Wendt, M. Najork, and A. Broder. (2018). "Learning Effective Embeddings for Machine Generated Emails with Applications to Email Category Prediction". In: *2018 IEEE International Conference on Big Data (Big Data)*. IEEE. 1846–1855.

Sutskever, I., O. Vinyals, and Q. V. Le. (2014). "Sequence to sequence learning with neural networks". *Advances in neural information processing systems*. 27: 3104–3112.

Sutton, C. and A. McCallum. (2012). "An Introduction to Conditional Random Fields". *Foundations and Trends® in Machine Learning*. 4(4): 267–373.

Swaminathan, S., R. Fok, F. Chen, T.-H. (Huang, I. Lin, R. Jadvani, W. S. Lasecki, and J. P. Bigham. (2017). "WearMail: On-the-Go Access to Information in Your Email with a Privacy-Preserving Human Computation Workflow". In: *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology. UIST '17*. Québec City, QC, Canada: ACM. 807–815. DOI: 10.1145/3126594.3126603.

Tang, J. C., E. Wilcox, J. A. Cerruti, H. Badenes, S. Nusser, and J. Schoudt. (2008). "Tag-It, Snag-It, or Bag-It: Combining Tags, Threads, and Folders in e-Mail". In: *CHI '08 Extended Abstracts on Human Factors in Computing Systems*. 2179–2194.

Task Force on Technical Approaches for Email Archives. (2018). *The Future of Email Archives*. URL: https://www.clir.org/pubs/reports/pub175/.

Tata, S., A. Popescul, M. Najork, M. Colagrosso, J. Gibbons, A. Green, A. Mah, M. Smith, D. Garg, C. Meyer, and R. Kan. (2017). "Quick Access: Building a Smart Experience for Google Drive". In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2017)*. Halifax, NS, Canada. 1643–1651. DOI: 10.1145/3097983.3098048.

Taylor, B., D. Fingal, and D. Aberdeen. (2007). "The War Against Spam: A Report From the Front Line". In: *NIPS 2007 Workshop on Machine Learning in Adversarial Environments for Computer Security*.

Full text available at: http://dx.doi.org/10.1561/1500000069

The Radicati Group, Inc. (2015). "Email Statistics Report 2015-2019". URL: https://www.radicati.com/wp/wp-content/uploads/2015/02/Email-Statistics-Report-2015-2019-Executive-Summary.pdf.

The Radicati Group, Inc. (2018). "Email Statistics Report 2018-2022". URL: https://www.radicati.com/wp/wp-content/uploads/2018/01/Email_Statistics_Report,_2018-2022_Executive_Summary.pdf.

The Radicati Group, Inc. (2019). "Email Statistics Report 2019-2023". URL: https://www.radicati.com/wp/wp-content/uploads/2018/12/Email-Statistics-Report-2019-2023-Executive-Summary.pdf.

Tran, B., M. Karimzadehgan, R. K. Pasumarthi, M. Bendersky, and D. Metzler. (2019). "Domain Adaptation for Enterprise Email Search". In: *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR'19*. Paris, France: ACM. 25–34. DOI: 10.1145/3331184.3331204.

Tsotsis, A. (2011). "ComScore Says You Don't Got Mail: Web Email Usage Declines, 59% Among Teens!" URL: https://techcrunch.com/2011/02/07/comscore-says-you-dont-got-mail-web-email-usage-declines-59-among-teens/.

Turtle, H. and J. Flood. (1995). "Query evaluation: strategies and optimizations". *Information Processing & Management*. 31(6): 831–850.

US Department of Health and Human Services. (2012). "Guidance regarding methods for de-identification of protected health information in accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule". URL: https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.%20html.

Uzuner, Ö., Y. Luo, and P. Szolovits. (2007). "Evaluating the state-of-the-art in automatic de-identification". *Journal of the American Medical Informatics Association*. 14(5): 550–563.

Full text available at: http://dx.doi.org/10.1561/1500000069

Van Gysel, C., B. Mitra, M. Venanzi, R. Rosemarin, G. Kukla, P. Grudzien, and N. Cancedda. (2017). "Reply With: Proactive Recommendation of Email Attachments". In: *Proceedings of the 2017 ACM Conference on Information and Knowledge Management (CIKM 2017)*. Singapore, Singapore: Association for Computing Machinery. 327–336. DOI: 10.1145/3132847.3132979.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. (2017). "Attention is all you need". In: *Advances in Neural Information Processing Systems*. 5998–6008.

Wan, S. and K. McKeown. (2004). "Generating Overview Summaries of Ongoing Email Thread Discussions". In: *Proceedings of the 20th International Conference on Computational Linguistics*. Association for Computational Linguistics. 549–555.

Wang, X., M. Bendersky, D. Metzler, and M. Najork. (2016a). "Learning to Rank with Selection Bias in Personal Search". In: *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2016)*. Pisa, Italy. 115–124. DOI: 10.1145/2911451.2911537.

Wang, X., N. Golbandi, M. Bendersky, D. Metzler, and M. Najork. (2018). "Position Bias Estimation for Unbiased Learning to Rank in Personal Search". In: *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. 610–618.

Wang, Y., D. Yin, L. Jie, P. Wang, M. Yamada, Y. Chang, and Q. Mei. (2016b). "Beyond Ranking: Optimizing Whole-Page Presentation". In: *Proceedings of the 9th ACM International Conference on Web Search and Data Mining (WSDM 2016)*. San Francisco, California, USA: Association for Computing Machinery. 103–112. DOI: 10.1145/2835776.2835824.

Watson Marketing. (2018). "2018 Marketing Benchmark Report: Email and Mobile Metrics for Smarter Marketing". URL: https://www.ibm.com/downloads/cas/L2VNQYQ0.

Full text available at: http://dx.doi.org/10.1561/1500000069

Wendt, J. B., M. Bendersky, L. G. Pueyo, V. Josifovski, B. Miklos, I. Krka, A. Saikia, J. Yang, M. Cartright, and S. Ravi. (2016). "Hierarchical Label Propagation and Discovery for Machine Generated Email". In: *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining, San Francisco, CA, USA, February 22-25, 2016*. 317–326. DOI: 10.1145/2835776.2835780.

Wenyin, L., G. Huang, L. Xiaoyue, Z. Min, and X. Deng. (2005). "Detection of Phishing Webpages Based on Visual Similarity". In: *Special Interest Tracks and Posters of the 14th International Conference on World Wide Web. WWW '05*. Chiba, Japan: ACM. 1060–1061. DOI: 10.1145/1062745.1062868.

Whittaker, S. (2005). "Supporting Collaborative Task Management in E-mail". *Human–Computer Interaction*. 20(1-2): 49–88. DOI: 10.1080/07370024.2005.9667361. eprint: https://www.tandfonline.com/doi/pdf/10.1080/07370024.2005.9667361.

Whittaker, S., V. Bellotti, and J. Gwizdka. (2006). "Email in personal information management". *Communications of the ACM*. 49(1): 68–73.

Whittaker, S., T. Matthews, J. Cerruti, H. Badenes, and J. Tang. (2011). "Am I Wasting My Time Organizing Email?: A Study of Email Refinding". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2011)*. ACM. Vancouver, BC, Canada. 3449–3458. DOI: 10.1145/1978942.1979457.

Whittaker, S. and C. Sidner. (1997). "Email Overload: Exploring Personal Information Management of Email". *Culture of the Internet*: 277–295.

Wu, Y. (2018). "Smart Compose: Using Neural Networks to Help Write Emails". URL: https://ai.googleblog.com/2018/05/smart-compose-using-neural-networks-to.html.

Xu, X., A. H. Awadallah, S. Dumais, F. Omar, B. Popp, R. Rounthwaite, and F. Jahanbakhsh. (2020). "Understanding User Behavior For Document Recommendation". In: *Proceedings of the Web Conference 2020 (WWW 2020)*. Taipei, Taiwan: Association for Computing Machinery. DOI: 10.1145/3366423.3380071.

Full text available at: http://dx.doi.org/10.1561/1500000069

Yue, Y., R. Patel, and H. Roehrig. (2010). "Beyond Position Bias: Examining Result Attractiveness as a Source of Presentation Bias in Clickthrough Data". In: *Proceedings of the 19th International Conference on World Wide Web (WWW 2010)*. Raleigh, North Carolina, USA: Association for Computing Machinery. 1011–1018. DOI: 10.1145/1772690.1772793.

Zamani, H., M. Bendersky, X. Wang, and M. Zhang. (2017). "Situational Context for Ranking in Personal Search". In: *Proceedings of the 26th International Conference on World Wide Web (WWW 2017)*. International World Wide Web Conferences Steering Committee. Perth, Australia: Association for Computing Machinery. 1531–1540. DOI: 10.1145/3038912.3052648.

Zhang, A., L. Garcia Pueyo, J. B. Wendt, M. Najork, and A. Broder. (2017). "Email Category Prediction". In: *Companion Proc. of the 26th International World Wide Web Conference*. 495–503.

Zhang, S., H. Yang, and L. Singh. (2016). "Anonymizing Query Logs by Differential Privacy". In: *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2016)*. Pisa, Italy: Association for Computing Machinery. 753–756. DOI: 10.1145/2911451.2914732.

Zhang, Y., W. Chen, D. Wang, and Q. Yang. (2011). "User-Click Modeling for Understanding and Predicting Search-Behavior". In: *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2011)*. San Diego, California, USA: Association for Computing Machinery. 1388–1396. DOI: 10.1145/2020408.2020613.

Zhao, Q., P. N. Bennett, A. Fourney, A. L. Thompson, S. Williams, A. D. Troy, and S. T. Dumais. (2018). "Calendar-Aware Proactive Email Recommendation". In: *The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2018)*. Ann Arbor, MI, USA: ACM. 655–664. DOI: 10.1145/3209978.3210001.

Zhuang, J. and Y. Liu. (2019). "PinText: A Multitask Text Embedding System in Pinterest". In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM. 2653–2661.