Search Among Sensitive Content

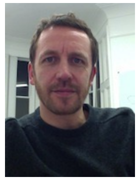
ECIR 2021 Tutorial

Graham McDonald, University of Glasgow, UK
Graham.mcdonald@glasgow.ac.uk

Douglas W. Oard, University of Maryland, USA
oard@umd.edu

Search-Among-Sensitive-Content.GitHub.io

About The Presenters

Graham McDonald 

- Sensitivity classification
  - Active-learning strategies
- Technology-Assisted Sensitivity Review
  - Decision Support
  - Reviewing time predictions
  - Resource Allocation
- Fair IR

Doug Oard 

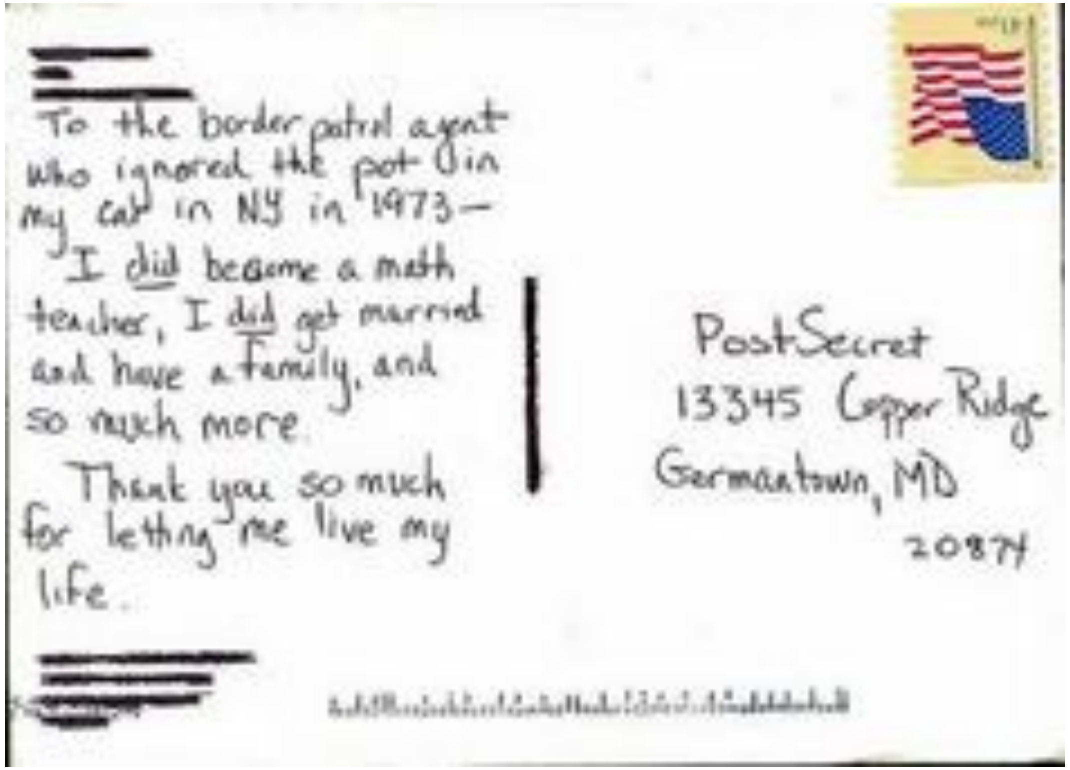
- Searching human language
  - Cross-Language IR
  - Speech Retrieval
  - Document Image Retrieval
  - Email Search (E-discovery)
- Evaluation design
- Privacy-protecting ranked retrieval

Tutorial Outline

CET

- 14:15 Background
- 14:45 Evaluation
- 15:20 Detecting sensitive content
- 16:00 Protecting Sensitive Content
- **16:15 Break**
- 16:45 Protecting Sensitive Content
- 17:00 Other Issues
- 17:20 Two Design Sprints (“choose your ending”)
- 17:55 Wrap up
- **18:15 End!**

We all have secrets …



Context Collapse: Everything's all mixed up

Search Enron | night stand

8429 results found

Relevance | Date ascending | Date descending

Subject:
Date: 2001-11-26T19:23:24
From: Dawn Carter <dcarter@allmort.com>
To: bill.williams@enron.com

So . . . you were looking for a one night stand afterall . . . ??

DC

Subject: RE: Moving
Date: 2002-02-20T09:42:35
From: Germany, Chris <chris.germany@enron.com>
To: germanj@basf-corp.com

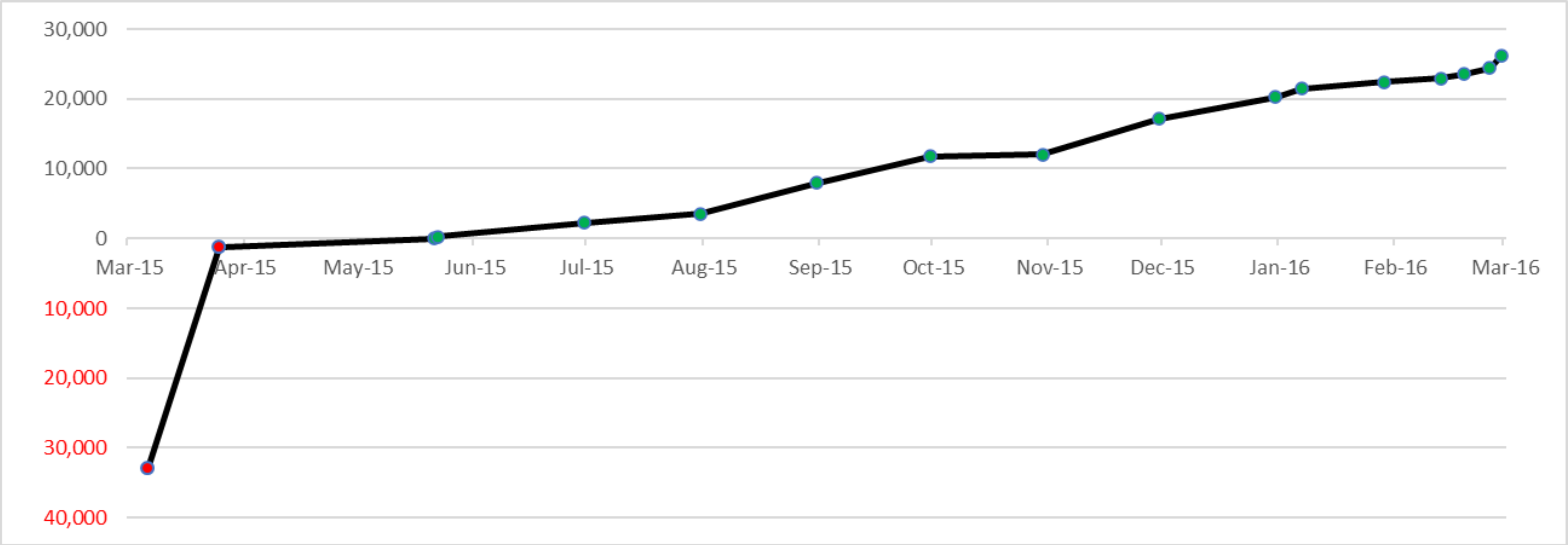
POINTY HEAD!!!!!!   I KNEW IT! Poor little fella.

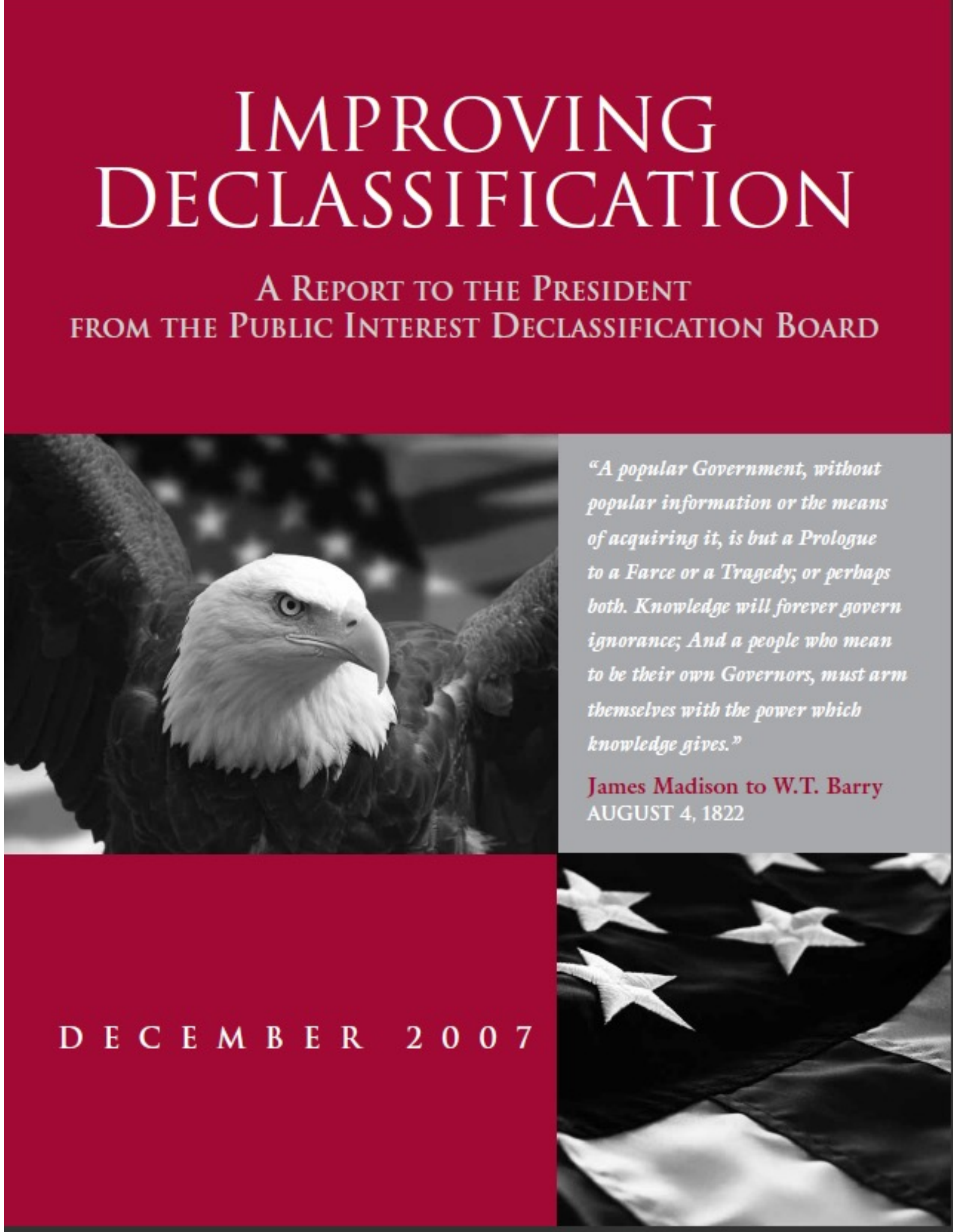
I like the apartment but the walk to the parking garage is terrible. I don't think Immer is going to like it very well either when she's lugging the baby around. Talk about space, I GOT IT! For now anyway. Yeah, I've emptied my car of certain items because I didn't want to carry them around. Hey we need to discuss how to divy those up. I've always wanted them even though I don't use them. Dad gave me the single shot 12, but there is still the little one, the double (the prize possesion) and the new one that you wanted to use.

The apartment is still pretty empty. I didn't want to empty my boxes until we got Immer's stuff in there. I don't think she has that much either, living room furniture, bed, 2 bedroom night stands, armwour, kitchen talble, 1 stupid cat.....

The Scope of the Problem: Clinton Email

- 59,171 emails generated over 4 years, stored on a personal server
  - 31,830 deleted as personal and not turned over
  - 1,200 entirely withheld by the State Department as personal
  - 23 entirely withheld for containing national security or law enforcement content
  - 26,118 released over ~10 months after review (>2,000 with redactions)

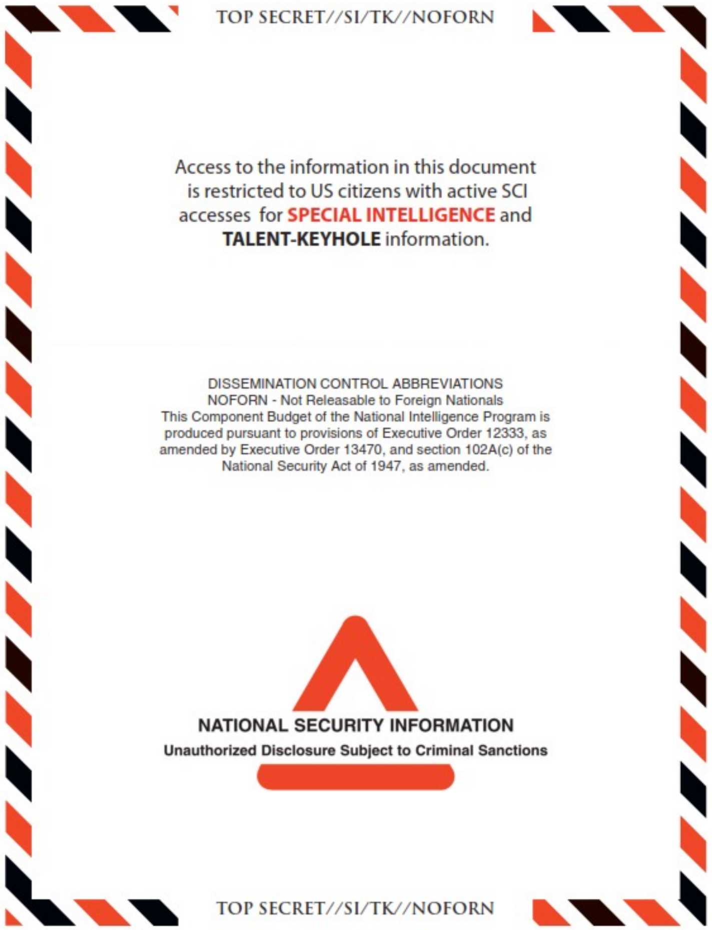




ISSUE NO. 2: Prioritizing the Declassification Review of Historically Significant Information. There is no satisfactory means at present of identifying historically significant information within the vast body of information that is being reviewed and declassified. Accordingly, no priority is given to the declassification and release to the public of such information.

Legal Regimes

- Formal specifications
  - National security ("classified" information)
  - Health records (e.g., HIPAA)
  - Personally Identifiable Information (e.g., GDPR)
- Categorical descriptions
  - Freedom of information Act exemptions
  - Attorney-client privilege
  - Right to be forgotten
- Personal privacy



Some Sensitivity Categories

- Personally Identifiable Information (PII)
- Student
- Health
- Employment
- Legal
- Crime
- Drug use
- Personal

Some Concerns of Donors to Email Archives

- Memberships and beliefs

  “his grandfather or great-grandfather…was a member of the Klan and he was scandalized about that”
- Evidence of stigmatized activity (e.g. drug use).
- Indiscretion
  - Gossiping, making “unfiltered” or “very very frank” remarks, using foul language
- Expressing emotional content in professional situations
- Battles

  “Usually, [sensitivity] is almost entirely going to be something that happened in their career that was contentious. Some controversy that they were part of, some event where they were at loggerheads with another person … and they would prefer not to have that made public.”
- Reputations of others

  “So for example, we have the papers of a very prominent religious speaker and she gets a lot of letters from people about spiritual crises they’re going through. And in some cases that involve… heavy things like abortions, she has asked that the identifying information, the name of the person who sent her that letter be anonymized.”

K. Shilton, et al., Protecting Sensitive Content in Email: Archival Views on Challenges and Opportunities, Workshop on Privacy-Sensitive Collections for Digital Scholarship, 2017

Stakeholders

- The searcher
  - Who wants to (at least) find relevant content
- The current owner of the content
  - Who wants their content used **<u>and</u>** their sensitivities protected
- The original creators of the content
  - Who want **<u>their</u>** sensitivities protected
- People or organizations described by the content
  - Who want **<u>their</u>** sensitivities protected

NEWS

CIA Realizes It's Been Using Black Highlighters All These Years

11/30/05 12:55PM • SEE MORE: POLITICS

LANGLEY, VA—A report released Tuesday by the CIA's Office of the Inspector General revealed that the CIA has mistakenly obscured hundreds of thousands of pages of critical intelligence information with black highlighters.

CIA Director Porter Goss.

According to the report, sections of the documents— "almost invariably the most crucial passages"—are marred by an indelible black ink that renders the lines impossible to read, due to a top-secret highlighting policy that began at the agency's inception in 1947.

CIA Director Porter Goss has ordered further internal investigation.

"Why did it go on for this long, and

Three Core Tasks

- Detect documents that contain sensitive content

- Detect sensitive content in a document

- Find relevant documents without exposing sensitive content

Tutorial Outline

CET

- 14:15 Background
- 14:45 Evaluation
- 15:20 Detecting sensitive content
- 16:00 Protecting Sensitive Content
- **16:15 Break**
- 16:45 Protecting Sensitive Content
- 17:00 Other Issues
- 17:20 Two Design Sprints ("choose your ending")
- 17:55 Wrap up
- **18:15 End!**

Section Outline

- Three Public Test Collections

- Protecting Private Test Collections
    - K-Anonymity
    - Differential privacy
    - Algorithm deposit

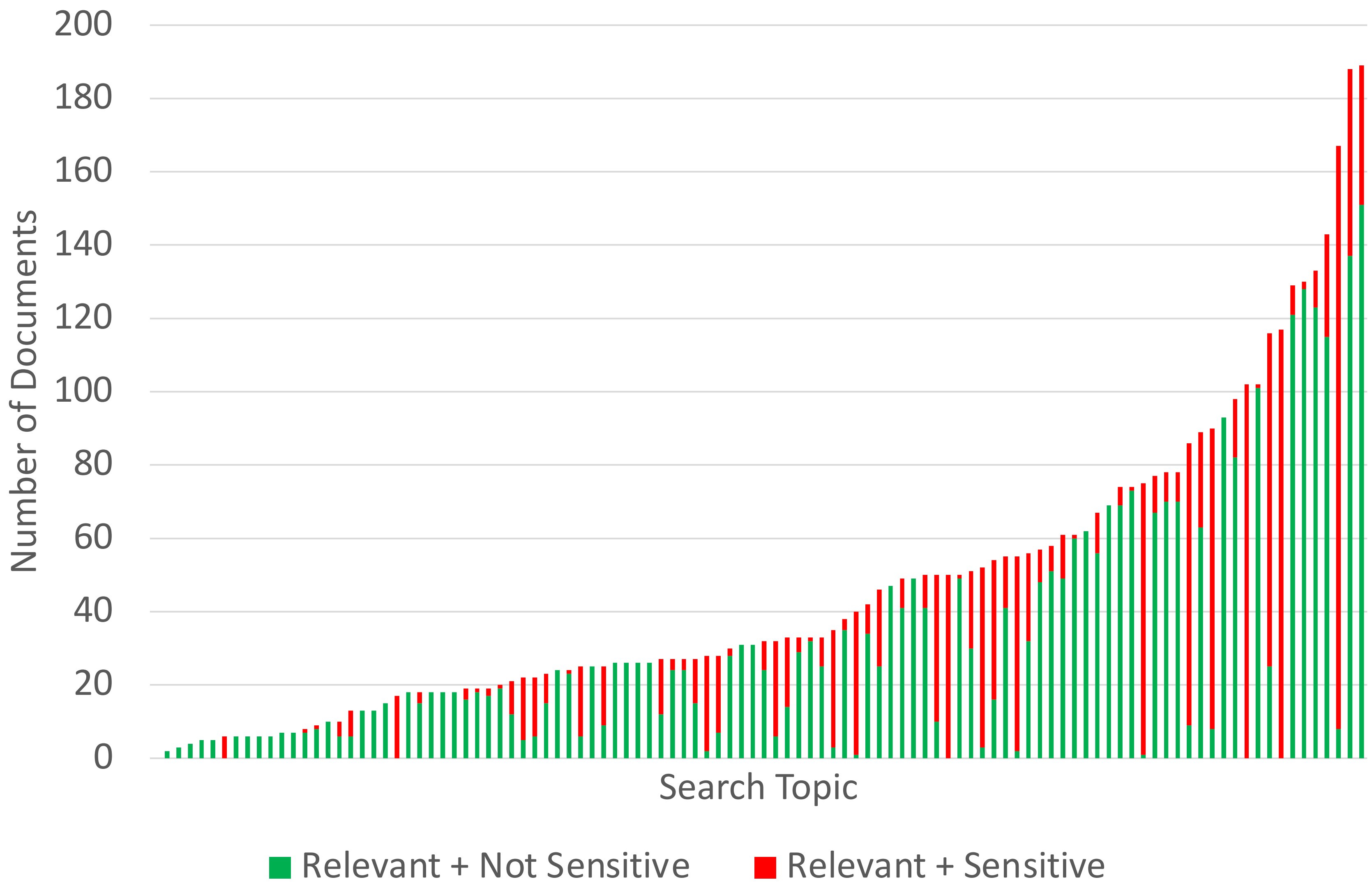
- Evaluation Measures
    - Classification
    - Redaction
    - Ranked retrieval

LETOR OHSUMED Sensitivity Test Collection

- 348,566 MEDLINE titles and abstracts
  - 334,136 training documents for sensitivity classifier training
  - 14,430 test documents w/relevance judgments for evaluation
- 106 Topics
  - Example: sigmoidoscopy in preventive care; whether the recommended frequency of sigmoidoscopy is effective and sensitive in detecting cancer
- Relevance Judgments
  - Complete in the test set
  - On average, 0.3% of documents are relevant
- Simulating sensitivity
  - Union of 2 Medical Subject Headings (MeSH):
    - Male genital diseases; Female genital diseases
  - 12.2% of judged documents are sensitive

M. Sayed, D. Oard, Jointly Modeling Relevance and Sensitivity for Search Among Sensitive Content. SIGIR, 2019

LETOR OHSUMED Sensitivity Test Collection



M. Sayed, D. Oard, Jointly Modeling Relevance and Sensitivity for Search Among Sensitive Content. SIGIR, 2019

Avocado Email Sensitivity Test Collection

- Avocado Email Research Collection (Licensed from LDC)
  - ~500K (deduped) messages, with attachments
- Two sensitivity personas
  - One with many sensitive documents, one with fewer
- 65 topics
  - 35 per sensitivity category (5 in common)
- Relevance Judgments
  - Pooled highly ranked documents from several systems
  - Additional documents from interactive searching
  - Annotate each for relevance and sensitivity

M. Sayed, et al., A Test Collection for Relevance and Sensitivity, SIGIR, 2020

“

User Quote

Motivations for email donation

Holly is reluctant to donate her emails to an archive because she worries that, her emails will be taken out of context when accessed in archives. She might want to pass her emails on to her family. However, she does not feel her emails hold enough importance that it’s worth the time and effort to curate them.

Why are their emails useful?

Holly has led several important research advances in the field of economics. Historians and economists would likely be interested in emails documenting her collaborations, research ideas, and research process.

Holly Palmer (Reluctant Professor)

Professor of Economics
Johns Hopkins University

Background

Palmer is a distinguished professor at Johns Hopkins University. She has won prestigious awards for her work in economics. Her papers and books are easily accessible on the internet but her communications and collaborations over the years are largely documented only in her email.

Perceptions & Use of Emails

She has used her work email to communicate with both personal and professional connections. She feels that her communication with them has been about work, home, logistics, gossips, and trade secrets (since she has consulted with Data and Tech companies for policy decisions). She feels that she has to pass on this collection to her family if no one else.

Pain points

• She is worried about the time and effort it will take to filter and delete conversations about her family and personal life.

• She is also mentioned travel and other receipts, and professional reviews of colleagues as information that she would not want shared.

• She is worried about potential harm to others (family and colleagues) if they discover unflattering things she has written about them.

• She is worried that her emails will be taken out of context.

• She does not understand why her emails would be useful to historians and scholars.

Goals

• Palmer is a busy professor, who does not have time or motivation to solve this problem - She wants a quick solution for her worst problem: unchecked social conversations.

• She wants to use the platform once a year for maintenance.

• She wants to understand why donation and archiving are important.

• She prefers to stay safe rather than save emails.

“

User Quote

Motivations for email donation

John is aware of the importance of his innovations, and motivated to donate his emails to an archive.

Why are their emails useful?

John is a respected senior engineer with a long and important career. He invented several products important to the history of computing.

John Snibert (Expert Engineer)

Retired Senior Computer Engineer
AVOCADO, Inc.

Background

John Snibert recently retired as a top engineer at AVOCADO, Inc. He was the inventor behind some of AVOCADO's most important products. He now gives numerous talks across the US and the world.

Pain points

- John is aware that there are sensitive emails in his collection that include his conversations with his family and romantic partners, peer reviews and collaboration on projects that contain proprietary information and trade secrets.
- He is not able to reliably find and delete emails when required.
- He worries about the intentions of the people who might access his emails, like journalists looking for a story.

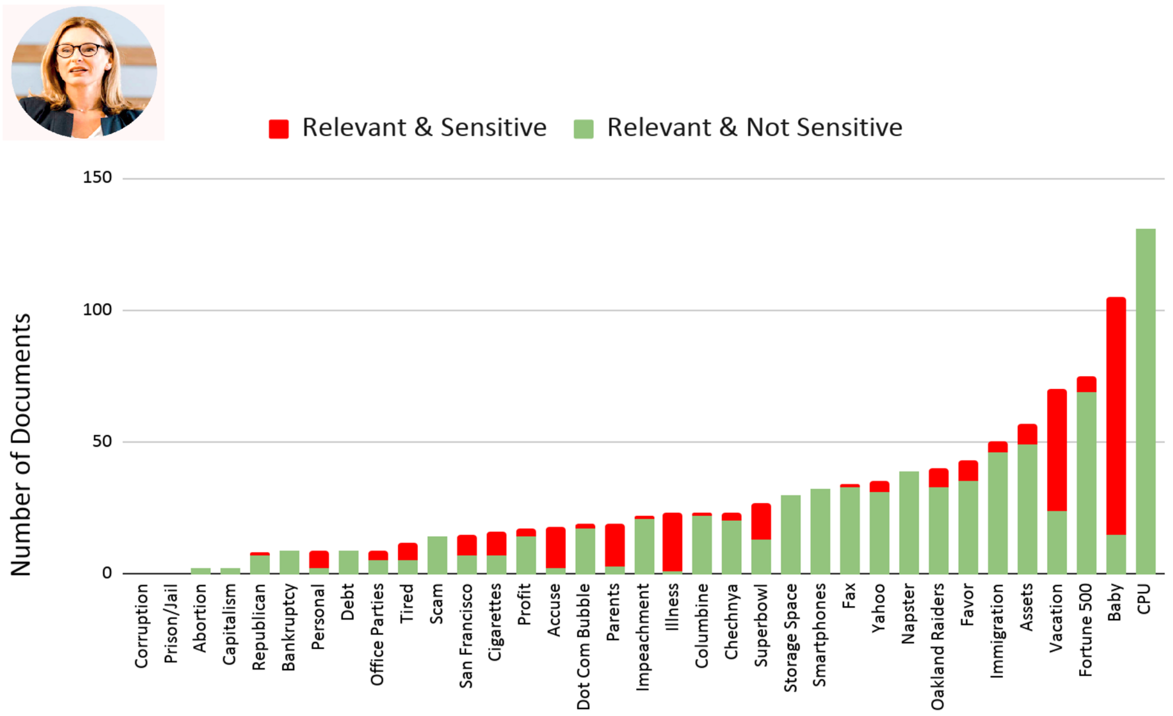
Perceptions & Use of Emails

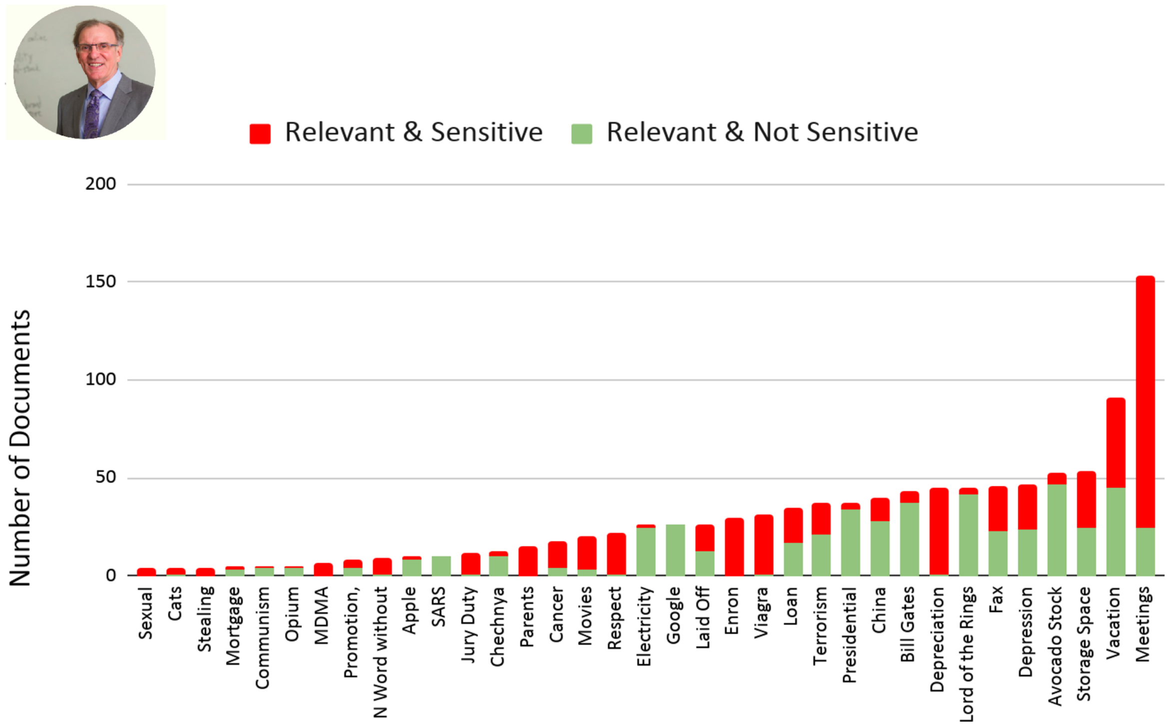
John has used email extensively for his work, including coordinating projects, planning presentations, and having conversations with other people important to the history of computer technology. He has also used email for sensitive matters like conversing with family and romantic partners. He believes he has been careful about what he puts in his email, and he has already done some curating and deleting of sensitive information. However, he finds it very difficult to find old emails and is worried that something he has missed might come to light.

Goals

- Snibert spends a lot of time organizing his emails and plans to donate his emails for both the common good of people who need this collection and to preserve his legacy - he wants to retain his reputation as an influential researcher.
- He wants to be able to search his older emails quickly.
- He wants to easily filter any deleterious emails he might have in collections.
- He wants to use emails as a memory aid for many other things (by checking his visits, calendar invites etc.)
- He wants to save the right emails and ratain his reputation.

Avocado Email Sensitivity Test Collection





M. Sayed, et al., A Test Collection for Relevance and Sensitivity, SIGIR, 2020

Deliberative Process Privilege Test Collection

- Documents
  - 509 OCR’d documents from 2 lawyers advising President Clinton
  - All exempted from public release for 12 years because they contained advice
- Annotations
  - 2 expert FOIA lawyers annotated for Deliberative Process Privilege exemption
  - All documents were marked at **document** level
  - Possibly exempt documents were also marked at the **paragraph** level

J. Baron, et al., Providing More Efficient Access to Government Records: A Use Case Involving Application of Machine Learning to Improve FOIA Review for the Deliberative Process Privilege, CoRR abs/2011.07203, 2020

Deliberative Process Privilege Test Collection

0609661

THE WHITE HOUSE

WASHINGTON

November 1, 1993

MEMORANDUM FOR THE PRESIDENT

FROM: DAVID WATKINS D. Watkins

SUBJECT: Presidential Yacht Sequoia

We now have the opportunity to cooperate with a nonprofit organization and thereby assist them in returning the Presidential Yacht Sequoia to the people of the United States and their Presidents. This is an "of the moment opportunity," though the effort is not new.

The Presidential Yacht Trust has been working since 1981 to preserve the Sequoia and return it to service. Currently, however, the Trust is in debt from renovation expenses and the Sequoia will be sold unless action is taken immediately. The Trust has recently made arrangements with leading citizens of Kuwait to make donations to the Trust sufficient to cover the debts of the Trust and to create a fund to maintain the Sequoia without taxpayer expense.

Before the donors will proceed, they have sought reassurances from the Yacht Trust that the White House is favorably disposed to this effort. To assist the Trust in meeting this reservation, I propose the attached letter for your signature. I have discussed this with other senior White House officials, including Bruce Lindsey. We are in concurrence that this opportunity should not be missed and that this letter is appropriate.

Name checks by the NSC are pending on the individuals known to be involved; nothing will go forward until these results come back clear.

I recommend the accompanying letter for your signature and I am available to discuss this matter with you further.

Batch	Custodian	Files	Paragraphs	File Names	Reviewer(s)
K1	Elena Kagan	9	523	Superfund, Welfare Budget, Welfare-Blair Visit, Service Summit Policy, Service General, Veterans Affairs/Filipinos, Drugs Coerced Abstinence, Drugs Heroin Chic	A
K2	Elena Kagan	10	447	Education/ TIMSS Meeting, Education/Troops to Teachers, Education/Vouchers, Environment/Climate Change, Kids Executive Order, Family Child Care Policy, Social Security/Nazis, Social Security/Prisoners, Drugs/Drug Testing	A & B
K3	Elena Kagan	10	670	Emails Received, Health/Radiation Experiments, Health/ Organ Transplants, Health/ Nursing Homes, Health/Medicaid Cap, Health/Immunization, Health/Genetic Screening, Drugs/Southwest Border, Environment/Port Dredging	A
R4	Cynthia Rice	5	466	Child Support/Gambling, Child Support/License, Fathers/Bayh Bill, Budget 2001 FY New Ideas, Disability-Kennedy-Jeffords 1999	A
K5	Elena Kagan	3	631	Tax Proposals; Drugs/Media Campaign, Drugs/ Meth Report	A
E5	Elena Kagan	3	286	Tax Proposals; Drugs/Media Campaign, Drugs/ Meth Report	A

J. Baron, et al., Providing More Efficient Access to Government Records: A Use Case Involving Application of Machine Learning to Improve FOIA Review for the Deliberative Process Privilege, CoRR abs/2011.07203, 2020

Section Outline

- Three Public Test Collections

➢ Protecting Private Test Collections
  - K-Anonymity
  - Differential privacy
  - Algorithm deposit

- Evaluation Measures
  - Classification
  - Redaction
  - Ranked retrieval

A Reidentification Attack

AOL User 4417749


The New York Times

LIEBERMAN LOSES SENATE PRIMARY BUT STAYS IN RACE

August 9, 2006

No. 4417749 conducted hundreds of searches over a three-month period on topics ranging from "**numb fingers**" to "**60 single men**" to "**dog that urinates on everything**." And search by search, click by click, the identity of AOL user No. 4417749 became easier to discern. There are queries for "**landscapers in Lilburn, Ga**," **several people with the last name Arnold** and "**homes sold in shadow lake subdivision gwinnett county georgia**." It did not take much investigating to follow that data trail to Thelma Arnold, a 62-year-old widow who lives in Lilburn, Ga., frequently researches her friends' medical ailments and loves her three dogs. "Those are my searches," she said, after a reporter read part of the list to her. AOL removed the search data from its site over the weekend and apologized for its release, saying it was an unauthorized move by a team that had hoped it would benefit academic researchers. But the detailed records of searches conducted by Ms. Arnold and 657,000 other Americans, copies of which continue to circulate online, underscore how much people unintentionally reveal about themselves when they use search engines — and how risky it can be for companies like AOL, Google and Yahoo to compile such data.

K-Anonymity

- Provides a quantifiable level of anonymity for entities.
- Hide sensitive information among *k* similar copies of data
  - Any individual can not be distinguished from at least *k-1* other individuals whose information is also released

- Query Logs: To satisfy k-anonymity, only release the query click data for records appearing at least k times in the original query log.
  - Difficult to retain the utility of logs, due to data sparseness

- Conceptual limitation:
  - Assumes knowledge of all of the available data
  - Current and future data

Differential Privacy

- Hides information of terms by adding noise to the sample statistics in a dataset.
  - Provide a statistical proof of privacy guarantee.

- Goal: No more harm can come to a person than if they did not appear in the data set
  - Data *seems* to no longer exist. It should be impossible to identify an individual.

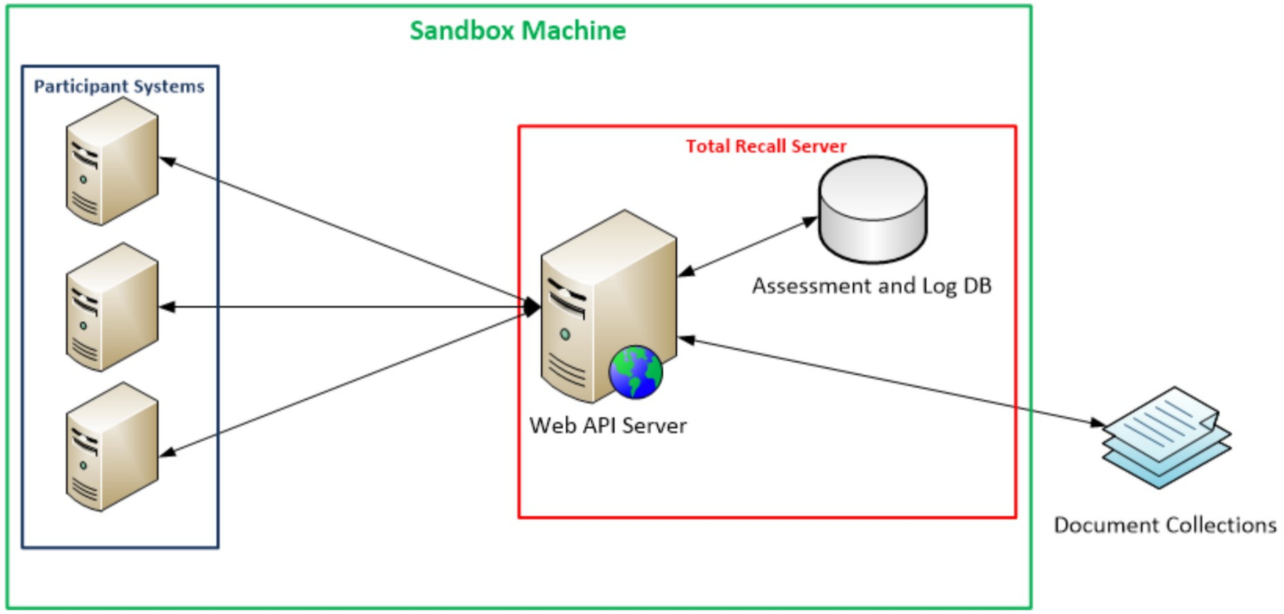
- Does not make assumptions on what knowledge an adversary has.

	Dataset 1	Dataset 2
Raw Data	Alice has 5 apples Bob has 4 apples Carol has 2 apples	Alice has 5 apples Bob has 4 apples
Sum of apples	5+4+2=11	5+4=9
Anonymized Sum	11+Noise=10	9+Noise =10

S. Zhang, G.H. Yang, Deriving differentially private session logs for query suggestion, ICTIR, 2017

TREC-2015 Total Recall Sandbox Task

- On-site access to former Governor Tim Kaine's email collection at the Library of Virginia.
- Sandbox used to conduct and evaluate experiments.
- Topics correspond to archival category labels
  - Not a Public Record
  - Open Public Record
  - Restricted Public Record
  - Virginia Tech Shooting Record



A. Roegiest, G. Cormack, C. Clarke, M.Grossman, TREC 2015 Total Recall Track Overview, TREC, 2015

Section Outline

- Three Public Test Collections
- Protecting Private Test Collections
  - K-Anonymity
  - Differential privacy
  - Algorithm Deposit
- ➢Evaluation Measures
  - Classification
  - Redaction
  - Ranked retrieval

Sensitivity Classification Metrics

Predicted As ➡	Sensitive	Not Sensitive
Sensitive	*TP*	*FN*
Not Sensitive	*FP*	*TN*

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}$$

$$BalancedAccuracy = \frac{\frac{TP}{TP+FN} + \frac{TN}{TN+FP}}{2}$$

K. Brodersen, C. Ong, K. Stephan, J. Buhmann, The balanced accuracy and its posterior distribution, ICPR, 2010

Sensitivity Classification Metrics

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

Sensitivity Classification Metrics

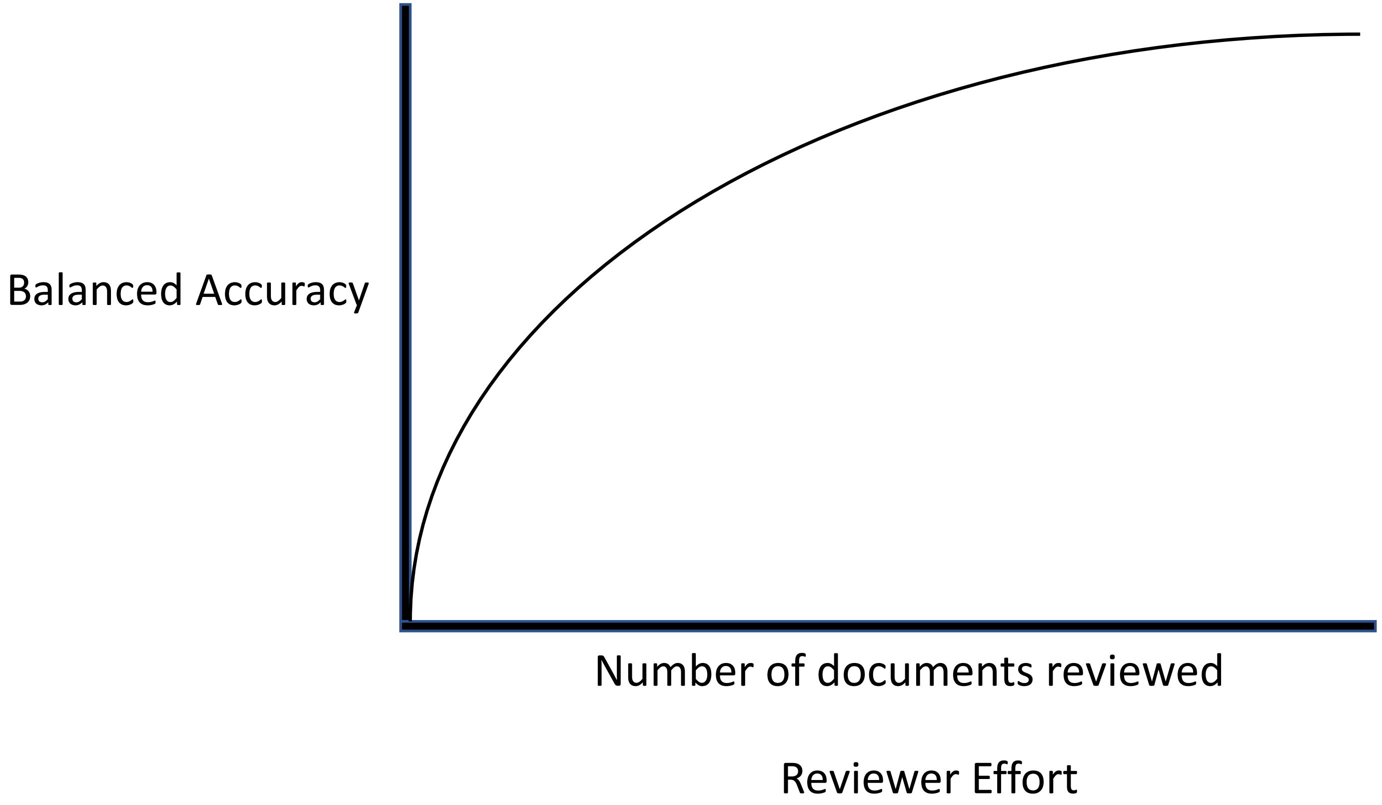
Parameterised harmonic mean of precision and recall

$$F_\beta = (1 + \beta^2) \cdot \frac{\mathrm{precision} \cdot \mathrm{recall}}{\beta^2 \cdot \mathrm{precision} + \mathrm{recall}}$$

$$F_\beta = \frac{(1 + \beta^2) \cdot \mathrm{true\ positive}}{(1 + \beta^2) \cdot \mathrm{true\ positive} + \beta^2 \cdot \mathrm{false\ negative} + \mathrm{false\ positive}}$$

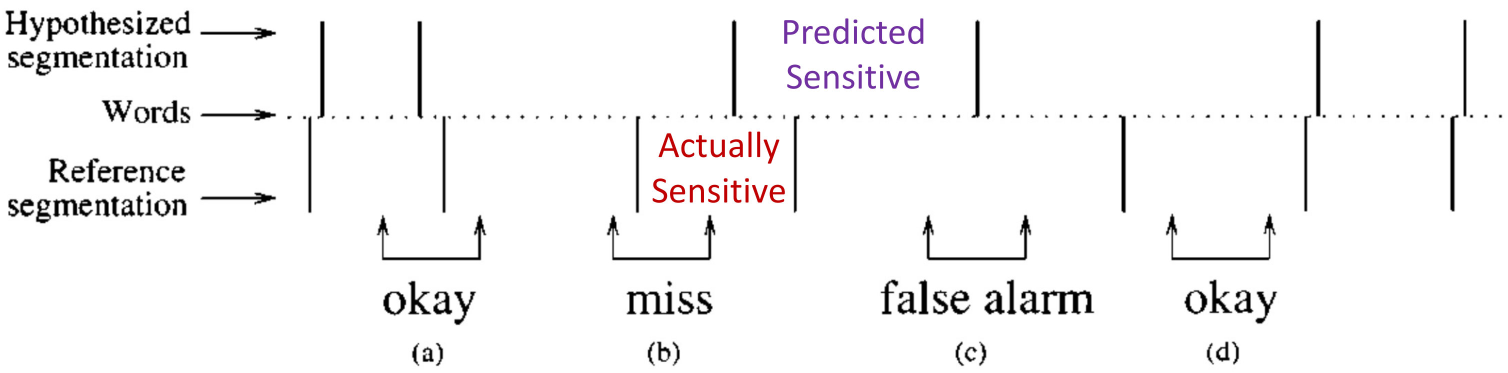
C.J. van Rijsbergen, Information Retrieval, Butterworth, 1979

Active-Learning Sensitivity Metrics



Span Detection Measures

- For two segmentations, reference (ref) and hypothesis (hyp), in a corpus of *n* sentences:



$$P_D(\mathtt{ref}, \mathtt{hyp}) = \sum_{1 \le i \le j \le n} D(i,j)\ (\delta_{\mathtt{ref}}(i,j)\ \overline{\oplus}\ \delta_{\mathtt{hyp}}(i,j))$$

D is a *distance probability distribution*

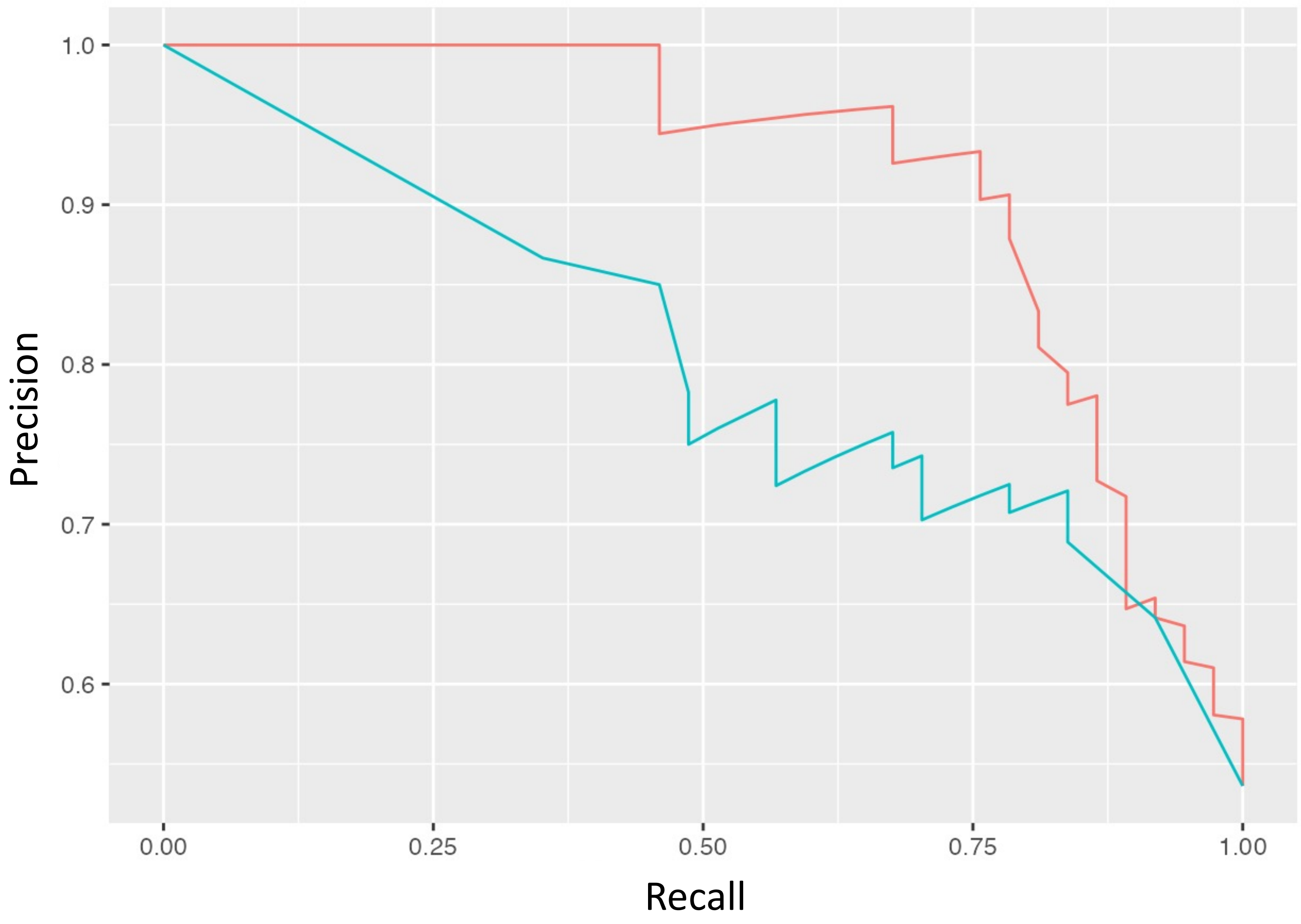
$\delta_{\mathtt{ref}}$ = 1 iff in reference span else 0

$\delta_{\mathtt{hyp}}$ = 1 iff in hypothesis span else 0

$\overline{\oplus}$ is XNOR (Both or neither)

D. Beeferman, A. Berger, J. Lafferty, Statistical models for text segmentation, Machine Learning, 34(1-3), 177–210, 1999

(Mean) Average Precision



Discounted Cumulative Gain

	Highly Relevant	Moderately Relevant	Not Relevant
RETRIEVED	+3	+1	0
NOT RETRIEVED	0	0	0

$$DCG_k = \sum_{i=1}^{k} \frac{g_i}{d_i}$$

Cost-Sensitive Discounted Cumulative Gain

	Highly Relevant	Moderately Relevant	Not Relevant
RETRIEVED	+3	+1	0
NOT RETRIEVED	0	0	0

$$DCG_k = \sum_{i=1}^{k} \frac{g_i}{d_i}$$

RETRIEVED	Highly Relevant	Moderately Relevant	Not Relevant
Not Sensitive	+3	+1	0
Sensitive	-5	-5	-5

NOT RETRIEVED	Highly Relevant	Moderately Relevant	Not Relevant
Not Sensitive	0	0	0
Sensitive (s)	0	0	0

$$CS - DCG_k = \sum_{i=1}^{k} \left(\frac{g_i}{d_i} + c_i\right)$$

M. Sayed, D. Oard, Jointly Modeling Relevance and Sensitivity for Search Among Sensitive Content. SIGIR, 2019

Tutorial Outline

CET

- 14:15 Background
- 14:45 Evaluation
- 15:20 Detecting sensitive content
- 16:00 Protecting Sensitive Content
- **16:15 Break**
- 16:45 Protecting Sensitive Content
- 17:00 Other Issues
- 17:20 Two Design Sprints ("choose your ending")
- 17:55 Wrap up
- **18:15 End!**

Section Outline

- Features: More than just words
- Sensitivity Classification
  - Context-dependent sensitivities
  - Active learning
- Decision-Support: Assisting Human Sensitivity Reviewers

From: H <hrod17@clintonemail.com>
Sent: Saturday, January 29, 2011 9:41 PM
To: 'verveerms@state.gov'
Subject: Re: Your husband

Thx for the firsthand report--I had heard he was unusually good in Davos. Must be the mountain air! I need some for sure.

----- Original Message Ambassador
From: Verveer, Melanne S <VerveerMS@state.gov>
To: H
Sent: Thu Jan 27 19:06:15 2011
Subject: Fw: Your husband

50 hours

Resending

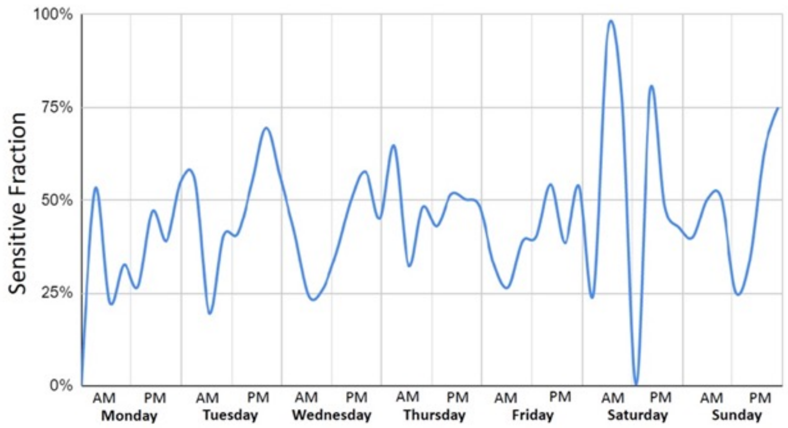
----- Original Message -----
From: Verveer, Melanne S
To: 'hdr22@clintonemail' <hdr22@clintonemail>
Sent: Thu Jan 27 19:03:46 2011
Subject: Your husband

Bill Clinton Klaus

Your husband's remarks in answer to questions from Schwab here in Davos today were exceptional in every way. He was reflective, expansive, knowledgeable, funny --and he looked terrific. He just gets better and better. Everyone seemed to be talking about him at once tonight.
At the end of a discourse that covered Israel and Palestine, the rioting in arab states, the off-year election, optimism about america, the deficit and the economy, health care, trade, etc., he was asked what his hopes were for the next 10

Features

- Words
- Time
  - Time of day, Day of week , Holidays, …
- Identity
  - Sender, recipients, mentions, relationships, organizational roles, …
- Interaction
  - Reply, forward, burstiness, …
- Specialized detectors
  - Spam, mailing list, confirmation, …



Case Study: Securing FOIA Sensitivities

the national archives


Freedom of Information Act 2000

2000 CHAPTER 36

(Office of Public Sector Information, 2000)

Exemptions

Section 21: Information Accessible by Other Means	Section 34: Parliamentary Privilege
Section 22: Information Intended for Future Publication	Section 37: Certain Aspects Relating to the Royal Family and Honours
Section 23: Bodies Dealing with Security Matters	Section 38: Health and Safety
Section 24: National Security	Section 39: Environmental Information
Section 26: Defence	Section 40: Personal Information
Section 27: International Relations	Section 41: Information Provided in Confidence
Section 29: The Economy	Section 44: Prohibitions on Disclosure
Section 31: Law Enforcement	

FOIA Sensitivity Test Collection

- 3800 Government documents
- Sensitivity reviewed by expert sensitivity reviewers from UK Government departments
- Text span-level ground truth annotations
  - Section 27 International Relations
  - Section 40 Personal Information

Context-Dependent Sensitive Information

Often, before reviewing a collection of documents, we do not know which text is likely to be sensitive.

Sensitivity can often arise as a result of the context in which the information is produced.

Context-Dependent Sensitive Information

Examples of FOIA context-dependent sensitivities include:

- Information that has been supplied with a reasonable expectation of confidentiality.
- Disparaging or inappropriate remarks about an *important* person.
- *Allegations* of inappropriate behaviour by a state or individual.
- Negative remarks from one country about the capabilities of another country or important person.
- Mentions of personal information, such as employment history, criminal activity, personal finances, ill health etc.

Context-Dependent Sensitive Information

Often dependent on a combination of multiple contextual factors, e.g.,

who ***said/did*** ***what*** ***when*** ***about/for*** ***whom***

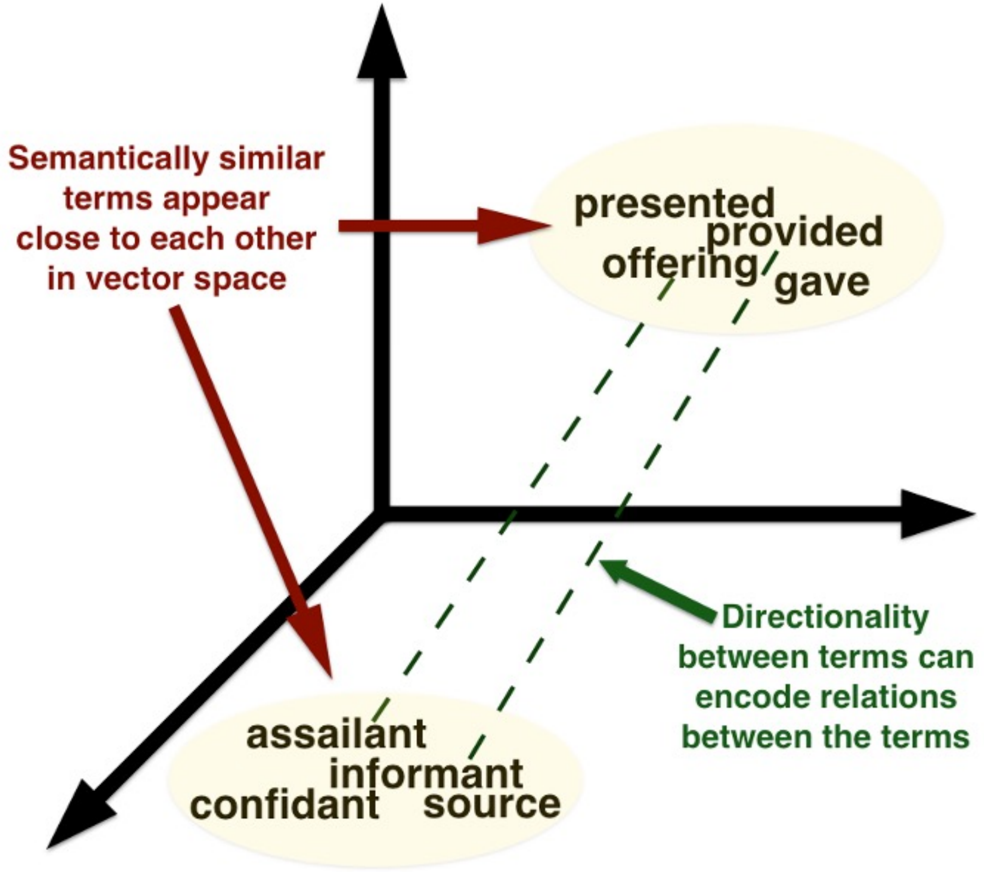
Swiss federal department of foreign affairs provided embassy on April 25 with an English translation by US interests station of Swiss embassy in Tehran of Rafsanjani's April 21 Friday message which described an alleged US espionage ring in Iran.

G. McDonald, A Framework for Technology-Assisted Sensitivity Review, PhD thesis, University of Glasgow, 2019

Sensitivity Classification

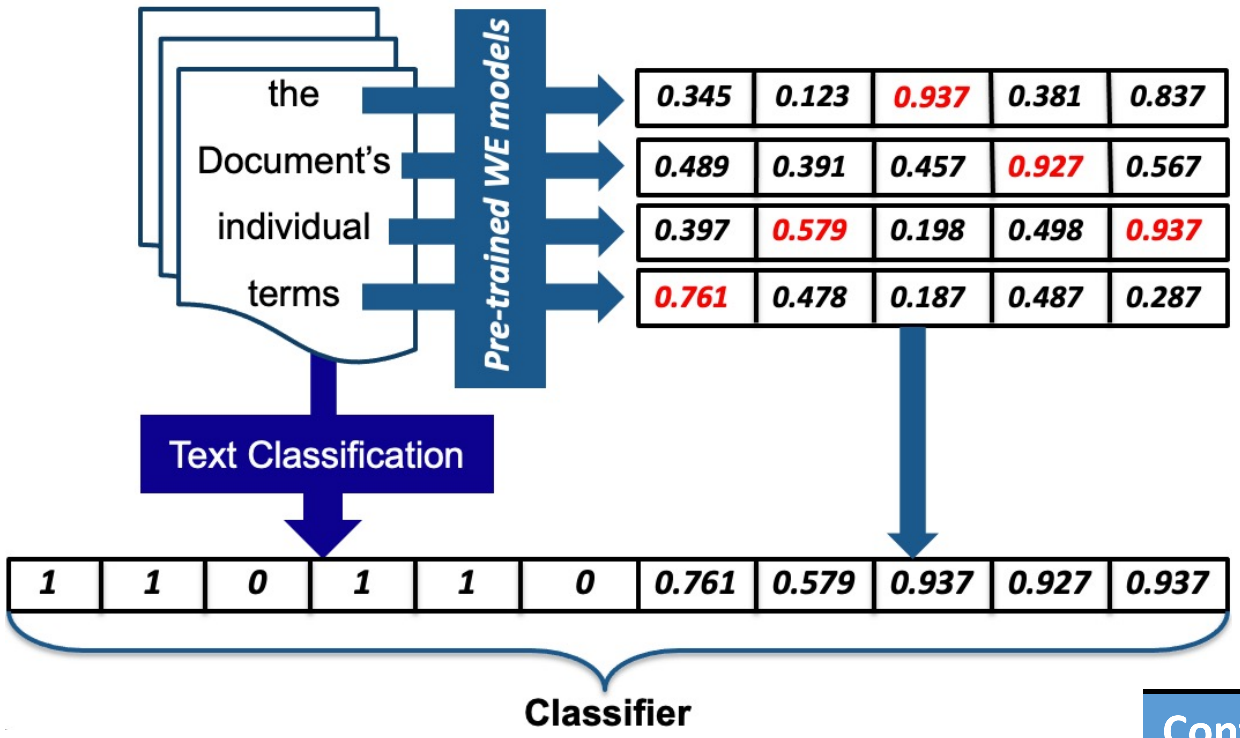
Need to be able to learn to automatically features that are indicative of contexts that are likely to be sensitive.

Embeddings can be effective at capturing such contexts.



G. McDonald, C. Macdonald, I. Ounis, Enhancing sensitivity classification with semantic features using word embeddings, ECIR, 2017

Sensitivity Classification



Combining semantic representations of documents with text classification improves sensitivity classification.

Configuration	Precision	recall	F_2	BAC
Text Classification (TC)	0.2410	0.6573	0.4874	0.6707
TC + Semantic Features	**0.2730**	**0.7229**	**0.5425**	**0.7149**

G. McDonald, C. Macdonald, I. Ounis, Enhancing sensitivity classification with semantic features using word embeddings, ECIR, 2017

Active-Learning for Sensitivity Classification

Each collection of documents that is reviewed will have different sensitivities that a classifier needs to learn to recognise.

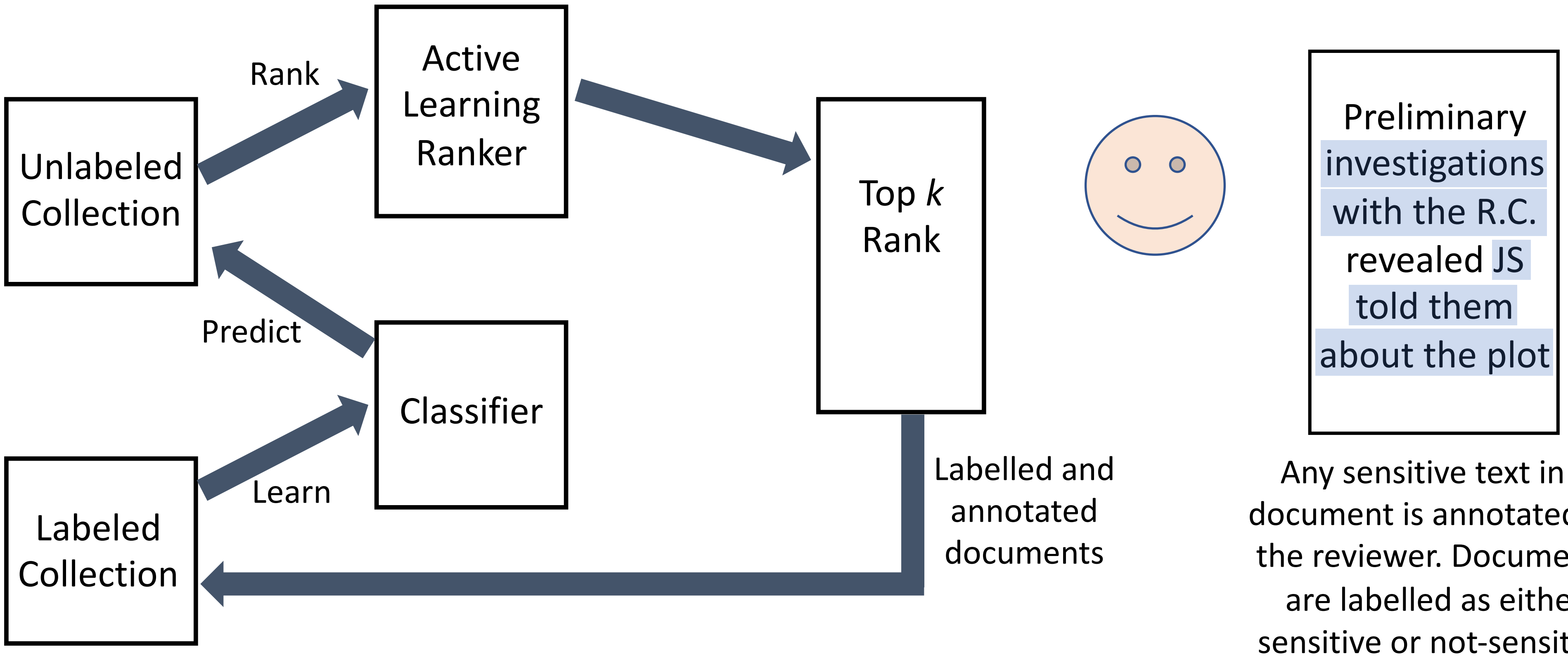
Human reviewers must review a set of documents and annotate the sensitivities to train the classifier. The aim is to reduce the number of documents that need to be manually reviewed, i.e., the reviewing effort, to train a sensitivity classifier.

Reviewing Effort = Number of documents that have to be reviewed to be able to learn a classifier that has an acceptable level of effectiveness (e.g., BAC, F_2).

G. McDonald, C. Macdonald, I. Ounis, Active learning strategies for technology assisted sensitivity review, ECIR, 2018

Active-Learning for Sensitivity Classification

Quickly learning to classify different types of context-dependent sensitivities:



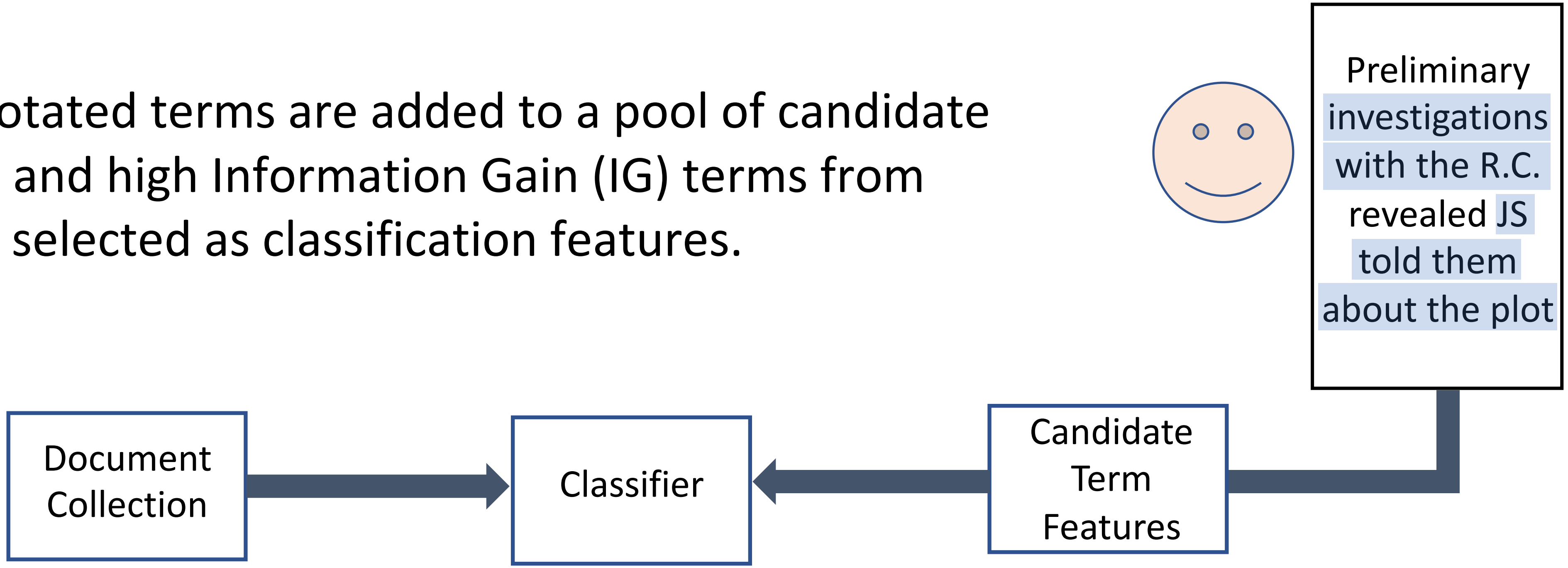
Any sensitive text in a document is annotated by the reviewer. Documents are labelled as either sensitive or not-sensitive

G. McDonald, C. Macdonald, I. Ounis, Active learning strategies for technology assisted sensitivity review, ECIR, 2018

Active-Learning for Sensitivity Classification

In each active learning iteration, the reviewer annotates any sensitive text within the documents being reviewed.

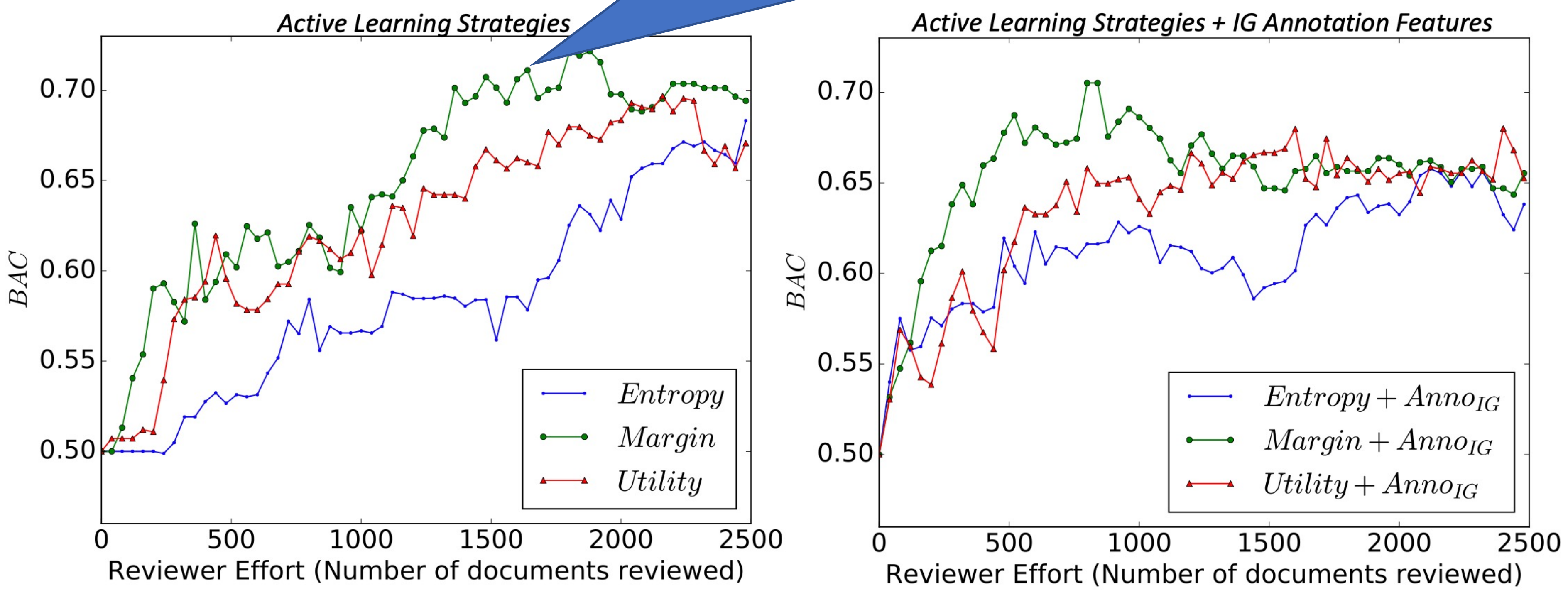
The annotated terms are added to a pool of candidate features and high Information Gain (IG) terms from pool are selected as classification features.



G. McDonald, C. Macdonald, I. Ounis, Active learning strategies for technology assisted sensitivity review, ECIR, 2018

Active-Learning for Sensitivity Classification

0.7 Balanced Accuracy (BAC) after ~1600 documents were reviewed.

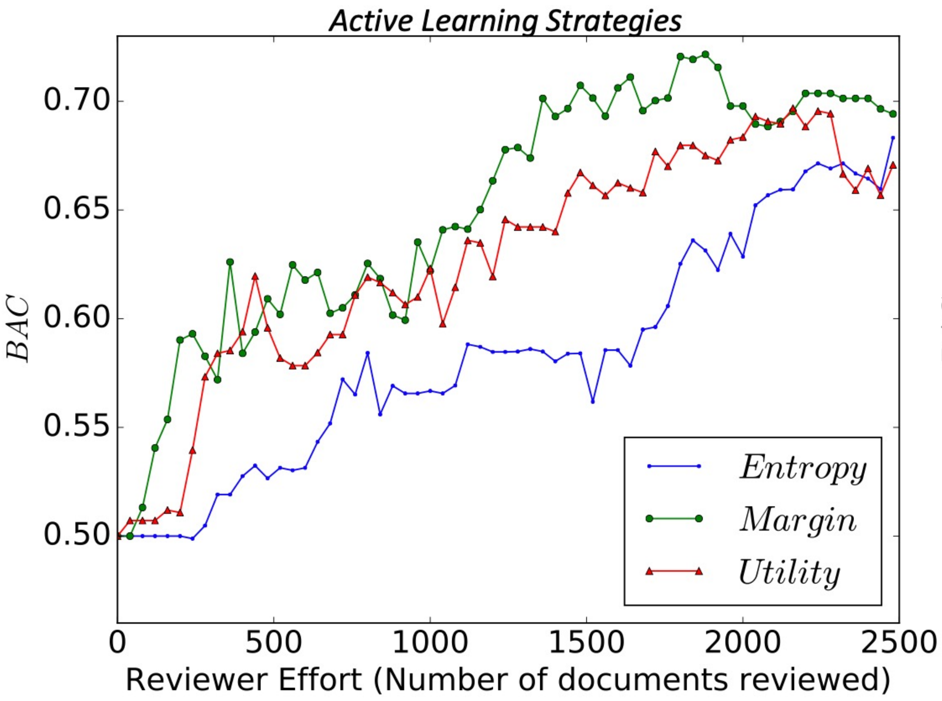


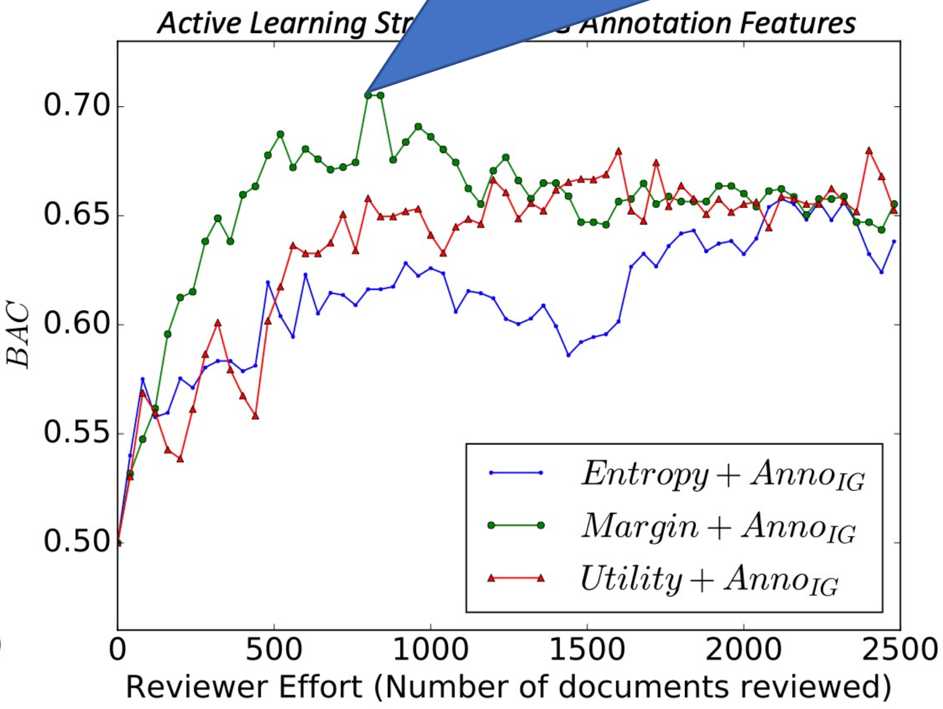
Integrating the high Information Gain (IG) term features (*Anno*$_{IG}$) results in reaching peak classification effectiveness using 51% less reviewing effort.

G. McDonald, C. Macdonald, I. Ounis, Active learning strategies for technology assisted sensitivity review, ECIR, 2018

Active-Learning for Sensitivity Classification

0.7 Balanced Accuracy (BAC) after ~800 documents were reviewed.



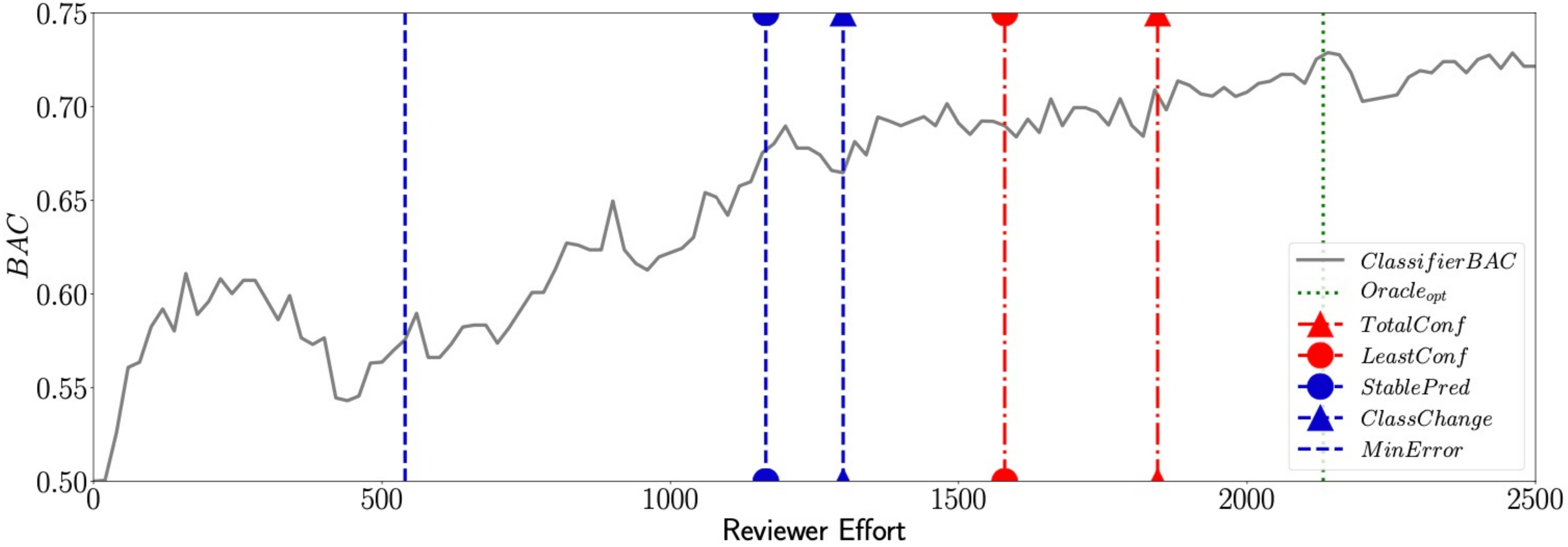


Integrating the high Information Gain (IG) term features (*$Anno_{IG}$*) results in reaching peak classification effectiveness using 51% less reviewing effort.

G. McDonald, C. Macdonald, I. Ounis, Active learning strategies for technology assisted sensitivity review, ECIR, 2018

Active-Learning for Sensitivity Classification

When to stop active-learning and switch to another presentation strategy?



Stop active-learning when:

Oracle$_{opt}$: the optimal classifier has been learned.

TotalConf*: the classifier's confidence stops increasing.*

LeastConf*: the probability of sensitive stops increasing*

StablePred*: the classifier's predictions stabilize.*

ClassChange*: the classifier's predictions stop changing.*

MinError*: the classifier correctly classifies the top k documents.*

G. McDonald, C. Macdonald, I. Ounis, Active learning stopping strategies for technology-assisted sensitivity review, SIGIR, 2020

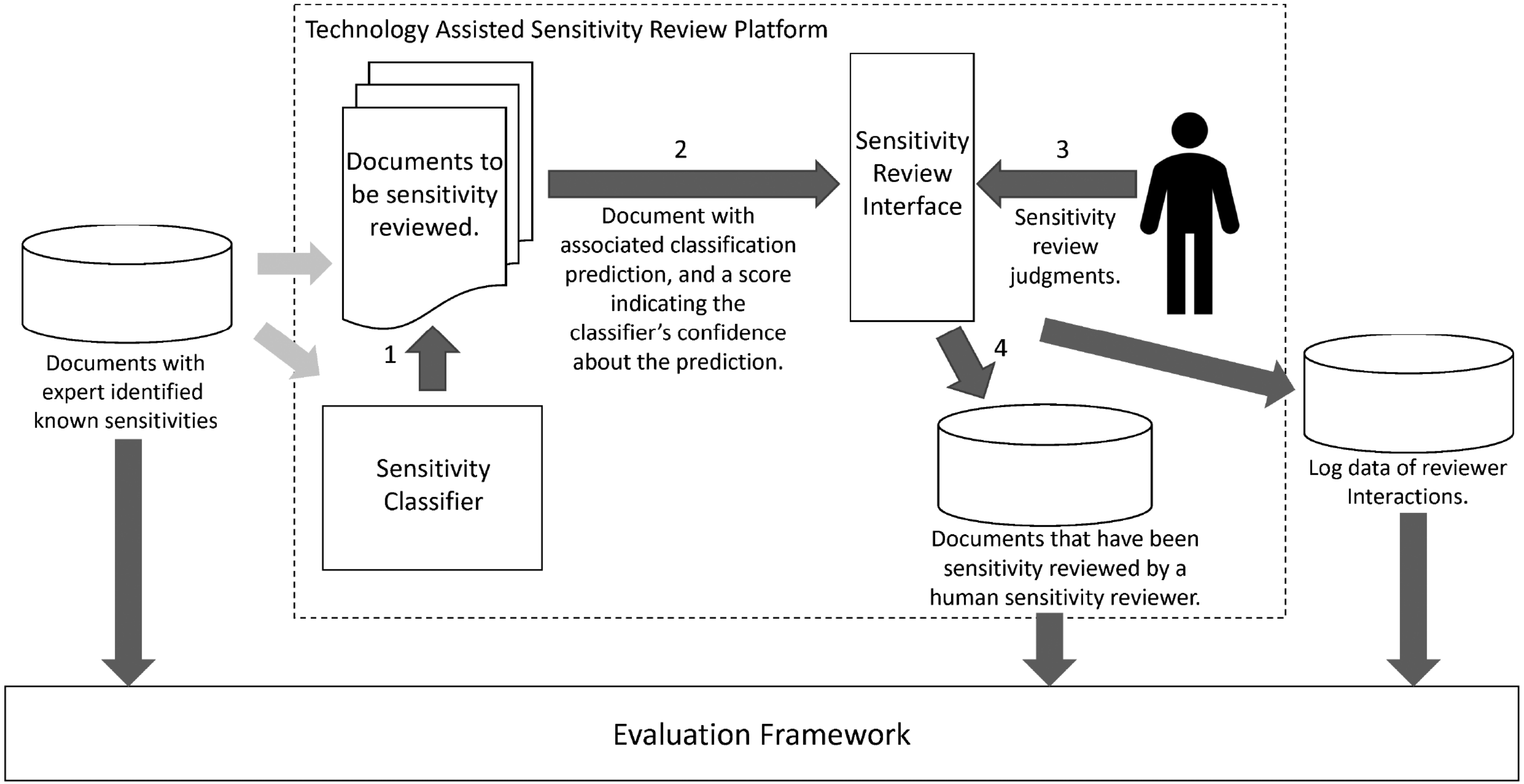
Decision-Support: Assisting Human Sensitivity Reviewers

Manually reviewing documents to identify context-dependent sensitivities is

- Labour intensive
- Time consuming
- Expensive

Professional sensitivity reviewers from five intelligence agencies were assigned to review Hillary Clinton's emails.

Decision-Support: Assisting Human Sensitivity Reviewers



G. Mcdonald, C. Macdonald, I. Ounis, How the accuracy and confidence of sensitivity classification affects digital sensitivity review, TOIS, 39(1), 1–34, 2020

Decision-Support: Assisting Human Sensitivity Reviewers

Aim: reduce the amount of time that it takes a reviewer to sensitivity review a document while maintaining (or increasing) the reviewer's accuracy.

Normalised Processing Speed (NPS):

- Measures the amount of time that reviewers require to review a document in *words per minute.*
- Accounts for: differences in reading speeds across reviewers using geometric averaging, and variations in document lengths.

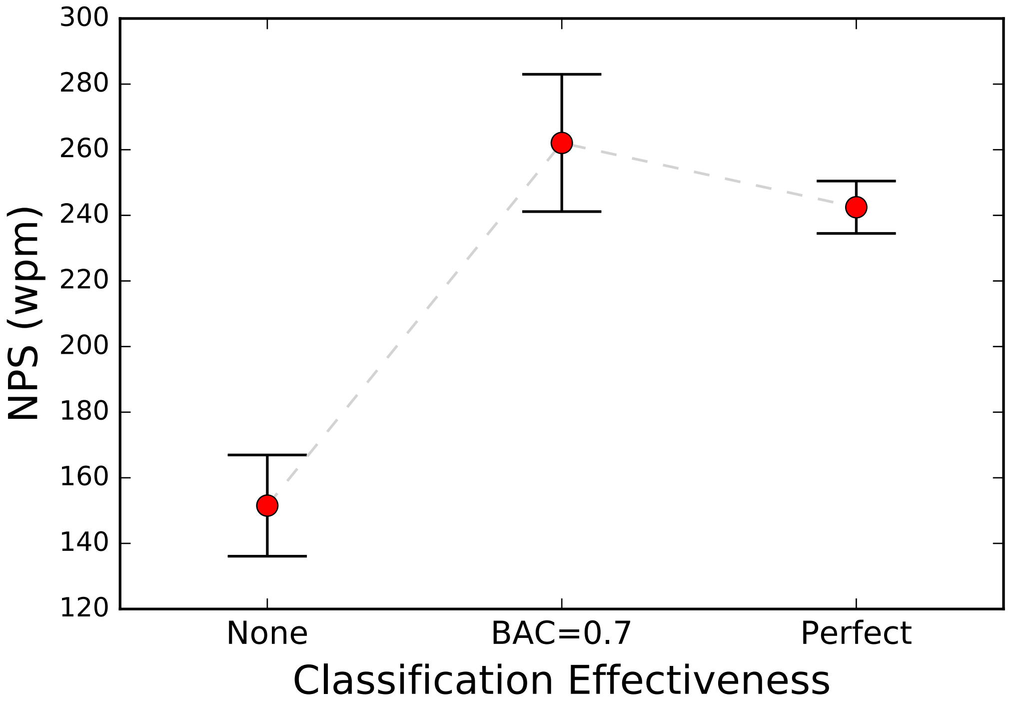
$$NPS = \frac{DocLength}{exp^{(log(time)+\mu-\mu_a)}}$$

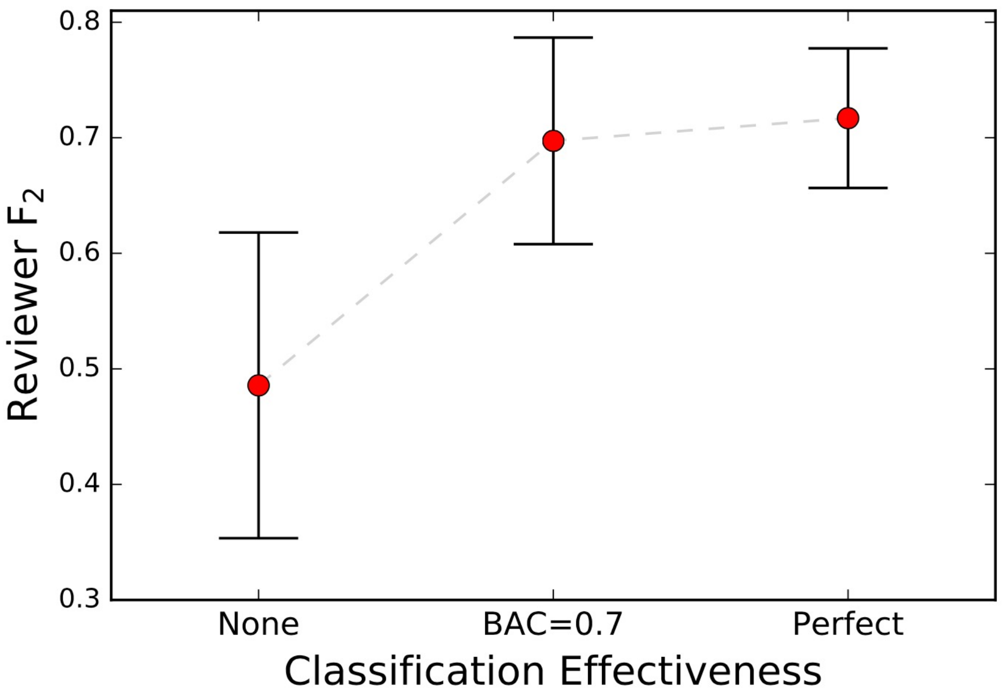
d_1	d_2	d_3	
5	7	6	μ_a = 18/3 = 6
2	3	2	μ_a = 7/3 = 2.33
			μ = 8.33/2 = 4.17

T. Damessie, F. Scholer, J.S. Culpepper, The influence of topic difficulty, relevance level, and document ordering on relevance judging, ADCS, 2016

Decision-Support: Assisting Human Sensitivity Reviewers

Aim: Increase the speed at which human reviewers can accurately sensitivity review a collection of documents.

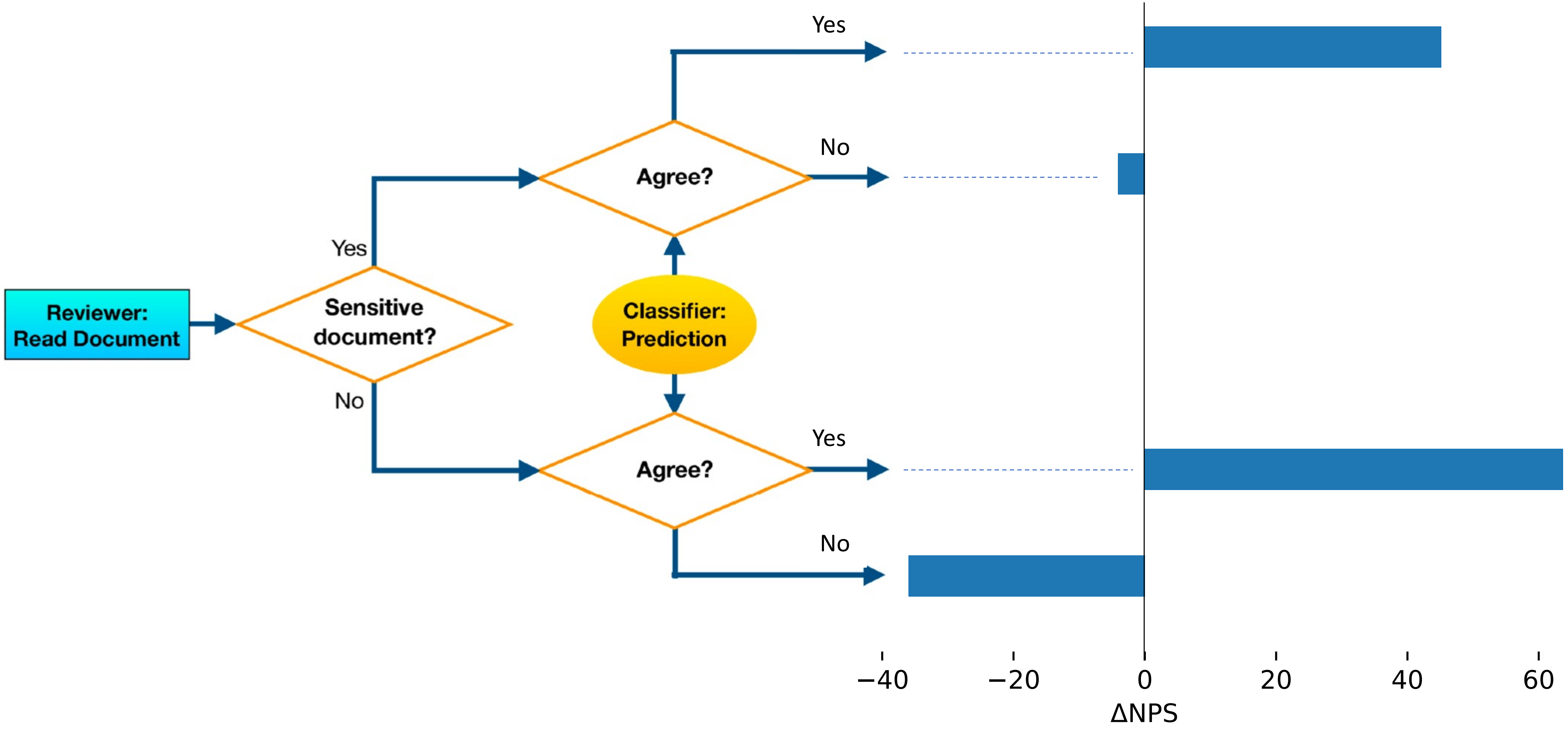




G. Mcdonald, C. Macdonald, I. Ounis, How the accuracy and confidence of sensitivity classification affects digital sensitivity review, TOIS, 39(1), 1–34, 2020

Decision-Support: Assisting Human Sensitivity Reviewers

Predicted As ➡	Sensitive	Not Sensitive
Sensitive	NPS_{SA}	NPS_{ND}
Not Sensitive	NPS_{ND}	NPS_{NA}



G. Mcdonald, C. Macdonald, I. Ounis, How the accuracy and confidence of sensitivity classification affects digital sensitivity review, TOIS, 39(1), 1–34, 2020

Tutorial Outline

CET

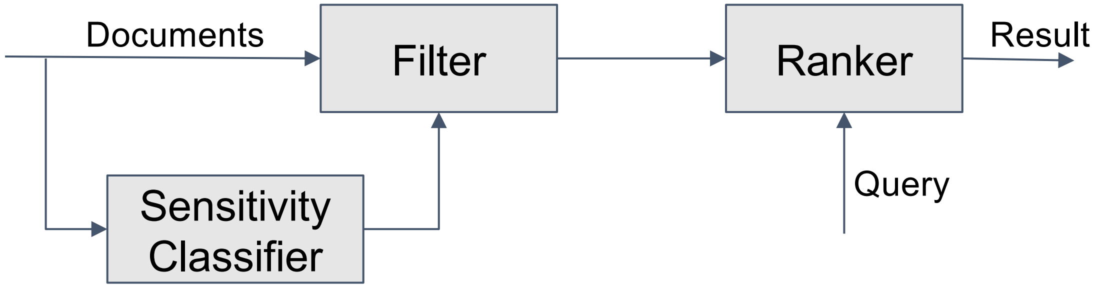
- 14:15 Background
- 14:45 Evaluation
- 15:20 Detecting sensitive content
- 16:00 Protecting Sensitive Content
- **16:15 Break**
- 16:45 Protecting Sensitive Content
- 17:00 Other Issues
- 17:20 Two Design Sprints ("choose your ending")
- 17:55 Wrap up
- **18:15 End!**

Section Outline

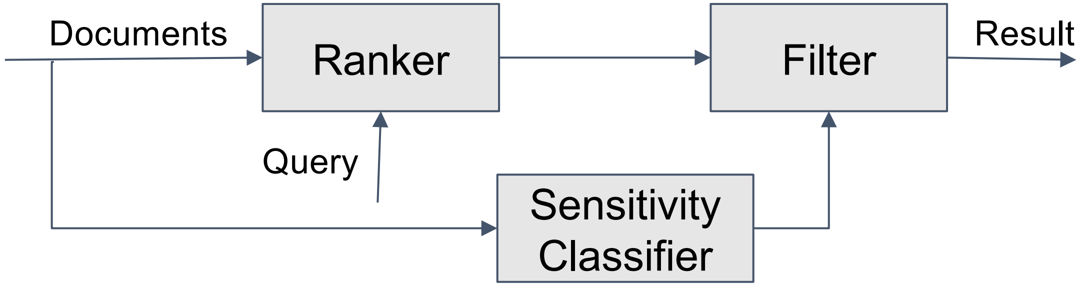
- End-User search
  - Sensitivity-aware ranked retrieval
- Intermediated search
  - Cost-sensitive prioritization
- Content protection
  - Redaction, sanitization

Sensitivity-Aware Ranked Retrieval

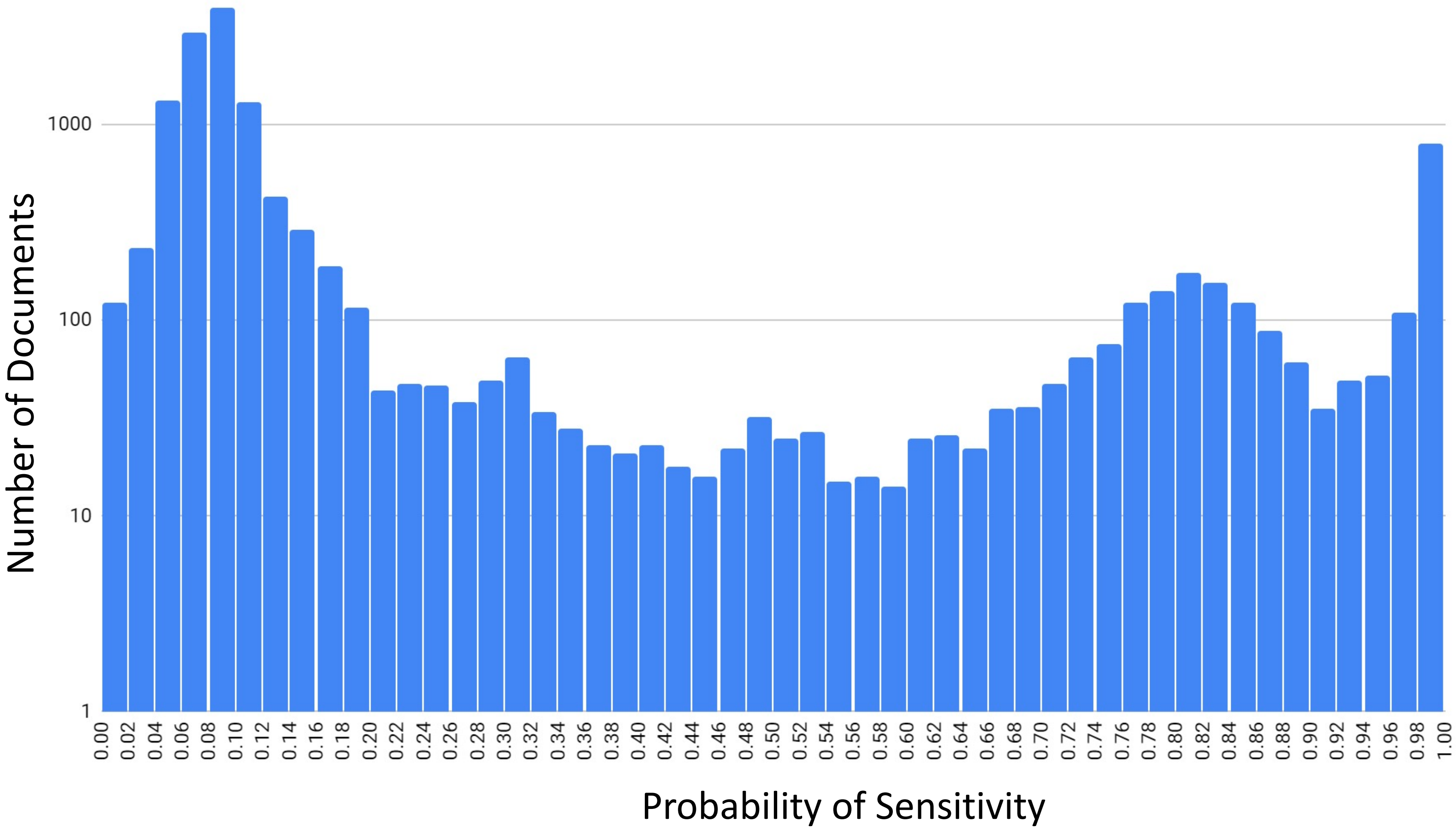
Prefilter (FOIA)



Postfilter (E-Discovery)

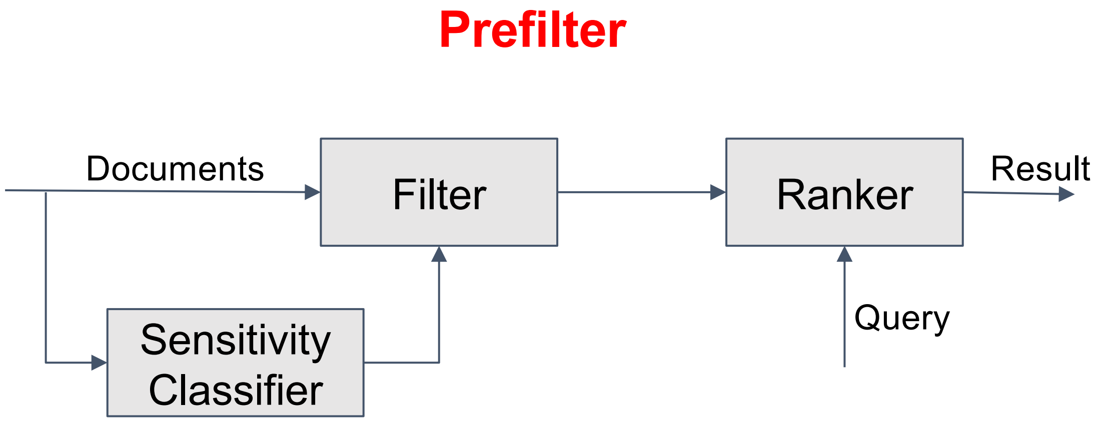


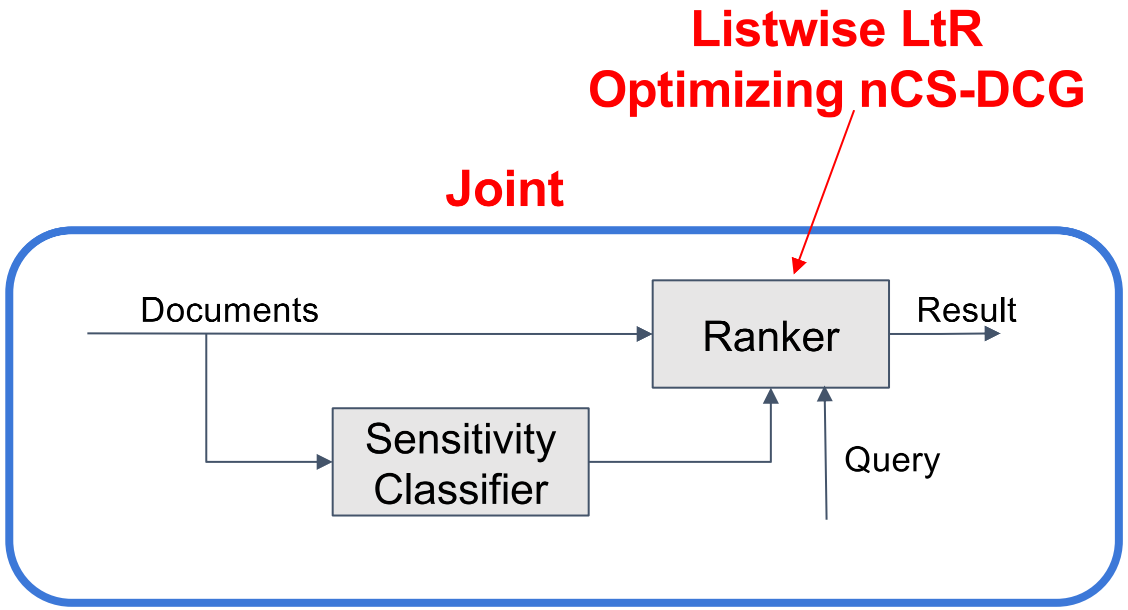
Sensitivity Probability Distribution

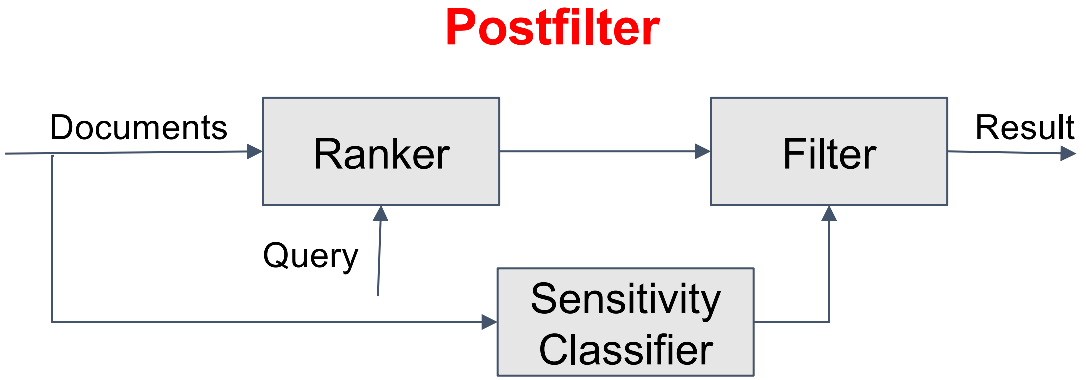


Linear kernel SVM, recall=0.61, precision=0.87

Sensitivity-Aware Ranked Retrieval

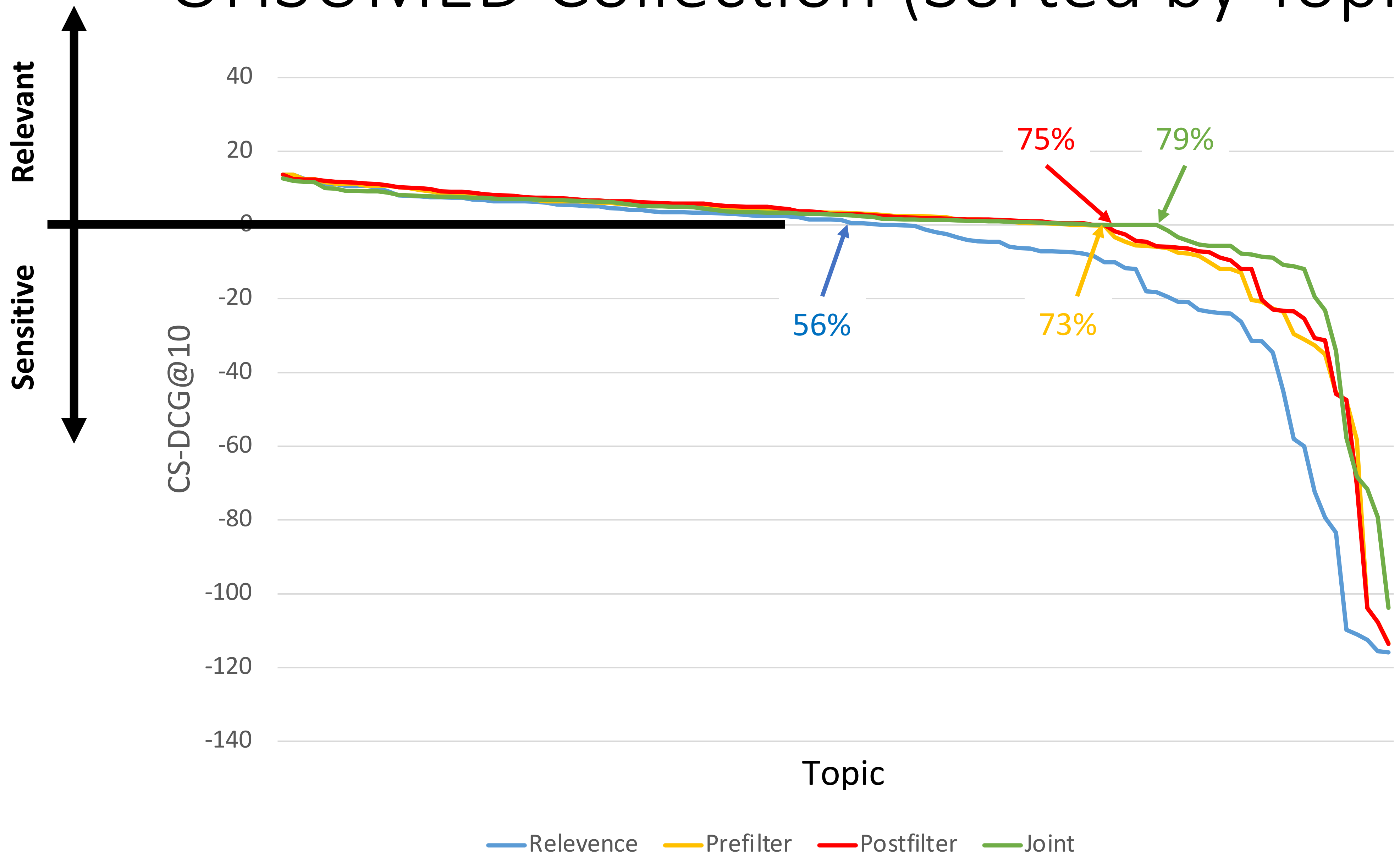






M. Sayed, D. Oard, Jointly Modeling Relevance and Sensitivity for Search Among Sensitive Content. SIGIR, 2019

OHSUMED Collection (Sorted by Topic)

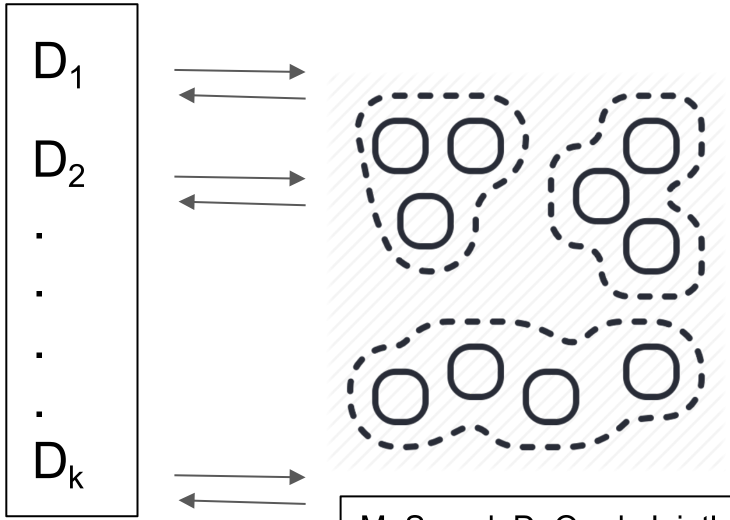


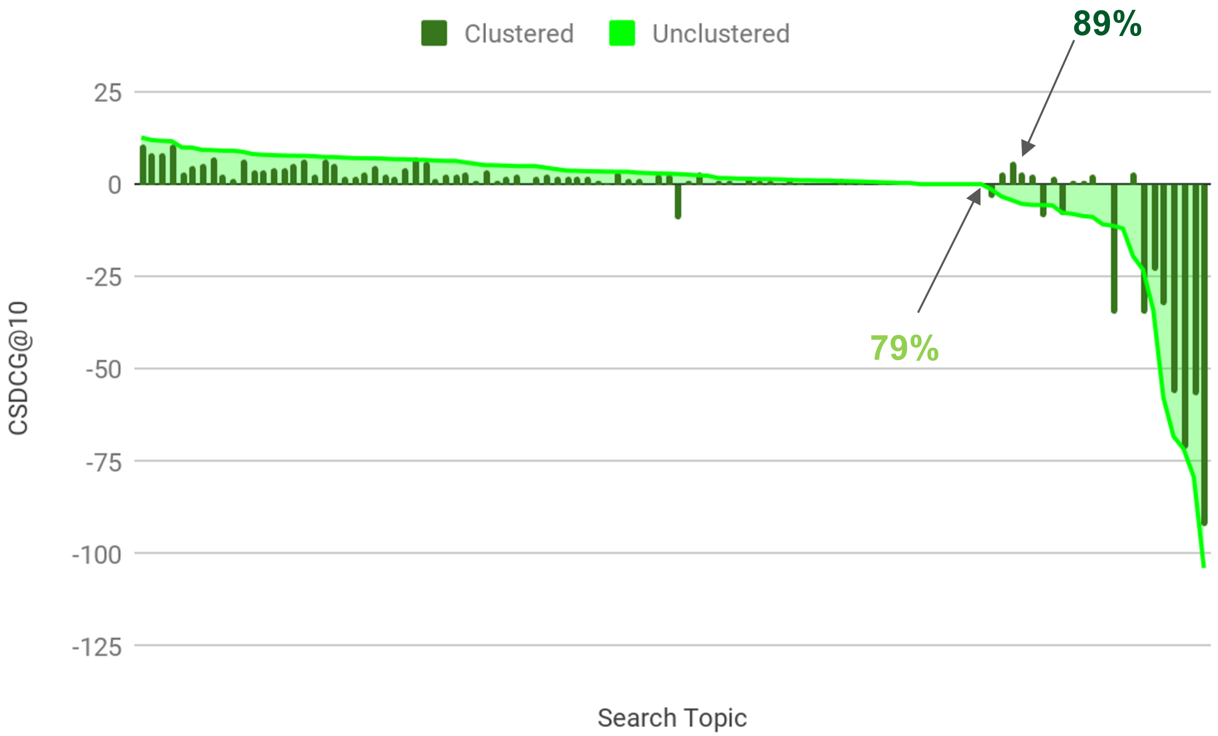
M. Sayed, D. Oard, Jointly Modeling Relevance and Sensitivity for Search Among Sensitive Content. SIGIR, 2019

Sensitivity Classifier F_1=0.75

Cluster-Based Replacement

- Similar to diversity ranking
  - Retrieved documents are clustered
  - For any potentially sensitive document in the result list is replaced with a document in the same cluster but less sensitive





20 clusters using repeated bisection

M. Sayed, D. Oard, Jointly Modeling Relevance and Sensitivity for Search Among Sensitive Content. SIGIR, 2019

Section Outline

- End-User search
  - Sensitivity-aware ranked retrieval

➢ Intermediated search
  - Cost-sensitive prioritization

- Content protection
  - Redaction, sanitization

Decision-Support: Assisting Human Sensitivity Reviewers

Rank automatically classified documents so as to optimize the cost-effectiveness of human reviewers post-checking.

Semi-Automated Text Classification for Sensitivity Identification

Giacomo Berardi[◇], Andrea Esuli[◇], Craig Macdonald[♣],
Iadh Ounis[♣], Fabrizio Sebastiani[♡*]
[◇]Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, Pisa, Italy
[♣]School of Computing Science, University of Glasgow, Glasgow, UK
[♡]Qatar Computing Research Institute, Hamad bin Khalifa University, Doha, Qatar

ABSTRACT

Sensitive documents are those that cannot be made public, e.g., for personal or organizational privacy reasons. For instance, documents requested through Freedom of Information mechanisms must be manually reviewed for the presence of sensitive information before their actual release. Hence, tools that can assist human reviewers in spotting sensitive information are of great value to government organizations subject to Freedom of Information laws. We look at sensitivity identification in terms of semi-automated text classification (SATC), the task of ranking automatically classified documents so as to optimize the cost-effectiveness of human post-checking work. We use a recently proposed utility-theoretic approach to SATC that explicitly optimizes the chosen effectiveness function when ranking the documents by sensitivity; this is especially useful in our case, since sensitivity identification is a recall-oriented task, thus requiring the use of a recall-oriented evaluation measure such as F_2. We show the validity of this approach by running experiments on a multi-label multi-class dataset of government documents manually annotated according to different types of sensitivity.

sensitive documents is attractive, since it can increase the efficiency of human reviewers. The possibility of treating sensitivity review as an automated text classification task has recently been shown in [7], where text classifiers were used in order to automatically detect sensitive documents, and where "sensitive" can have different interpretations (e.g., defence-related issues, or issues related to law enforcement).

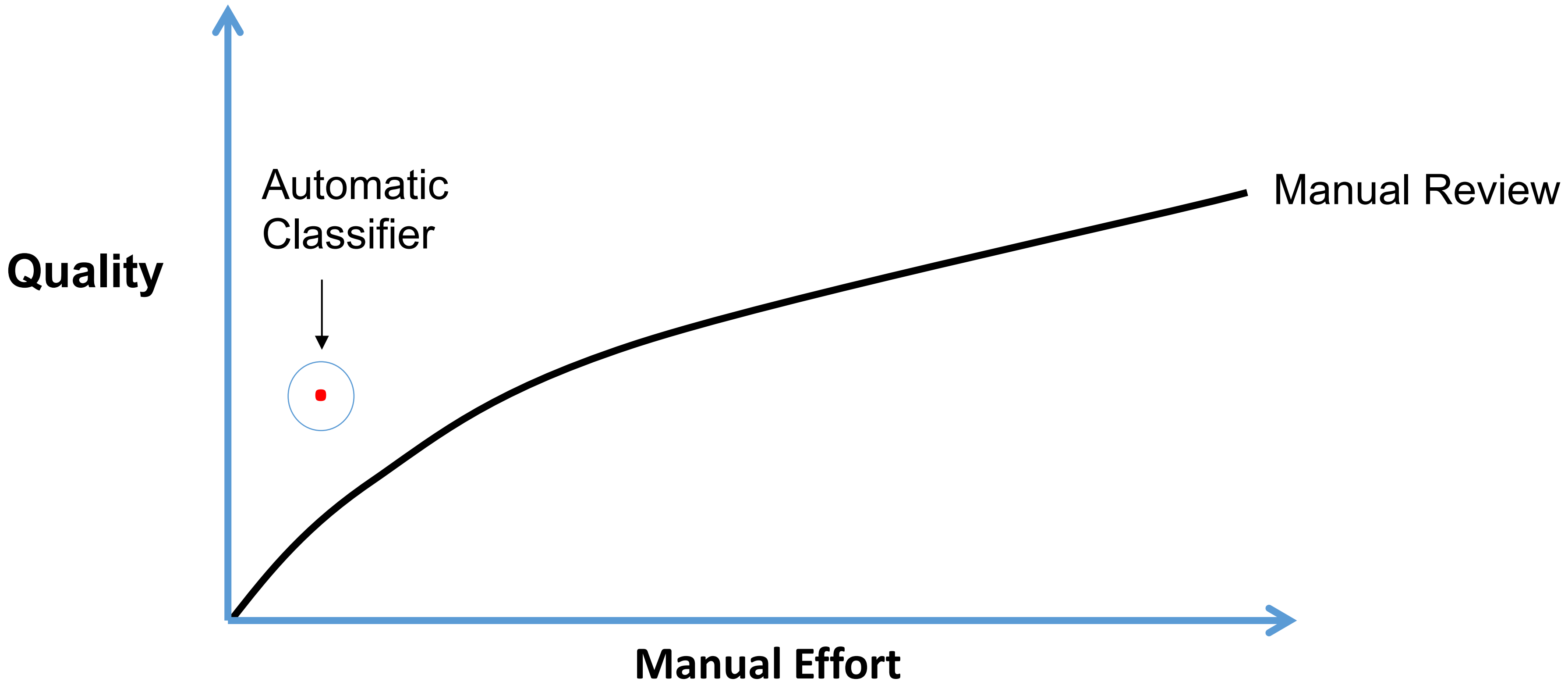
The task of sensitivity identification bears strong resemblances with "review for privilege" in e-discovery [8], where expert attorneys must check that "privileged" (i.e., sensitive) information is not accidentally disclosed to a requesting party in the context of a civil litigation process [3, 10]. Another task that bears resemblances with sensitivity identification is record anonymisation, as when e.g., medical records have to be anonymised before they are released for epidemiological studies; in this case, sensitive information such as patients' names and medical doctors' names have to be spotted in order to be redacted [9]. Sensitivity identification and privilege identification are text classification tasks, while record anonymisation is an information extraction task. Notwithstanding the differences, all these cases are characterized by the fact that the costs of accidental disclosure of sensitive information are high.

Two-Stage Intermediated Search

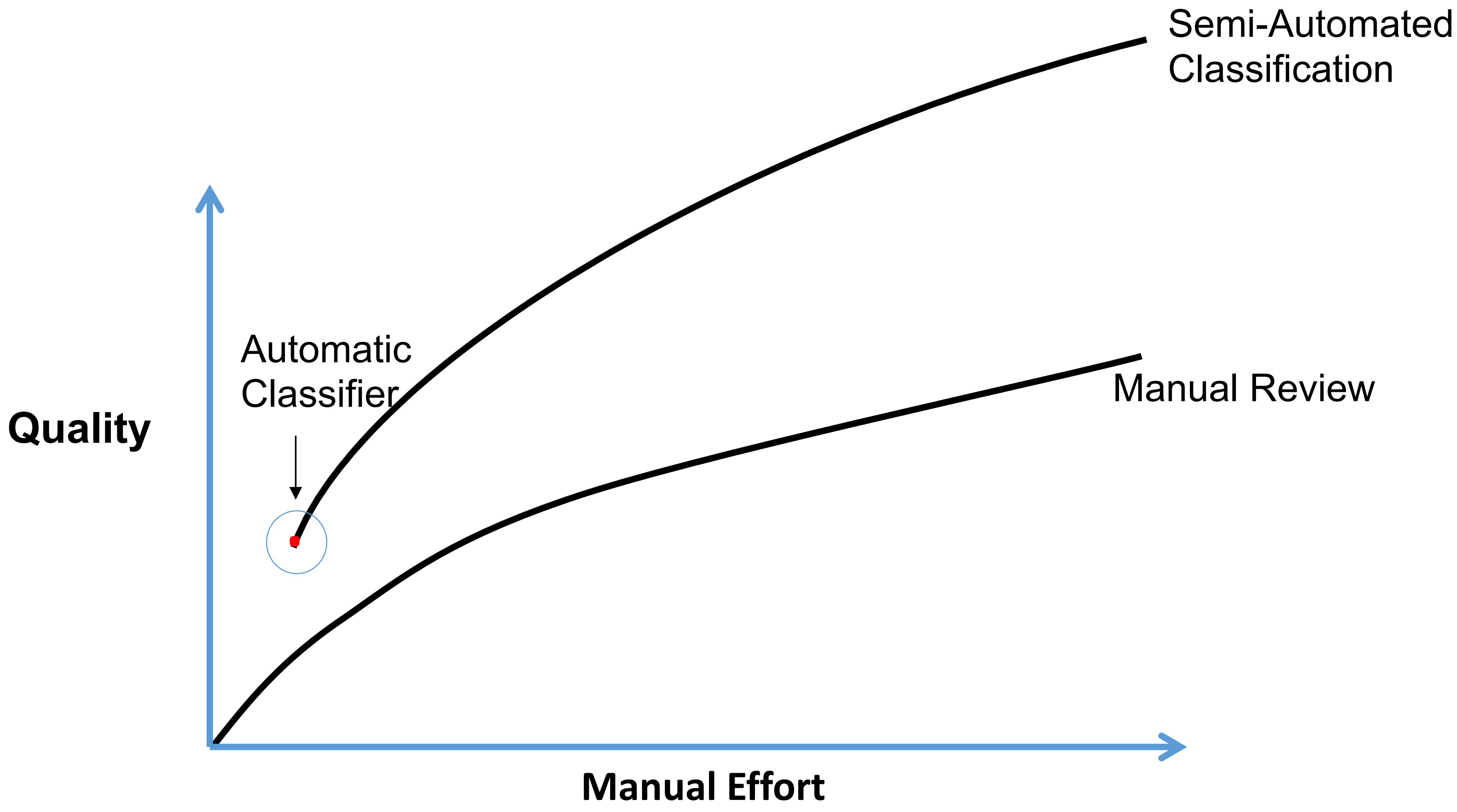
- E-Discovery requires that we employ a reasonable process to:
  - Identify documents that are relevant (i.e., “responsive”) to a request
  - Among the relevant documents, identify those that are privileged

- 3 possible actions:
  - **Produce** (i.e., disclose) documents that are **relevant and not privileged**
  - Enter on a Privilege **Log** documents that are **relevant and privileged**
  - **Withhold** documents that are **not relevant**

Idea #1: Finite Population Annotation

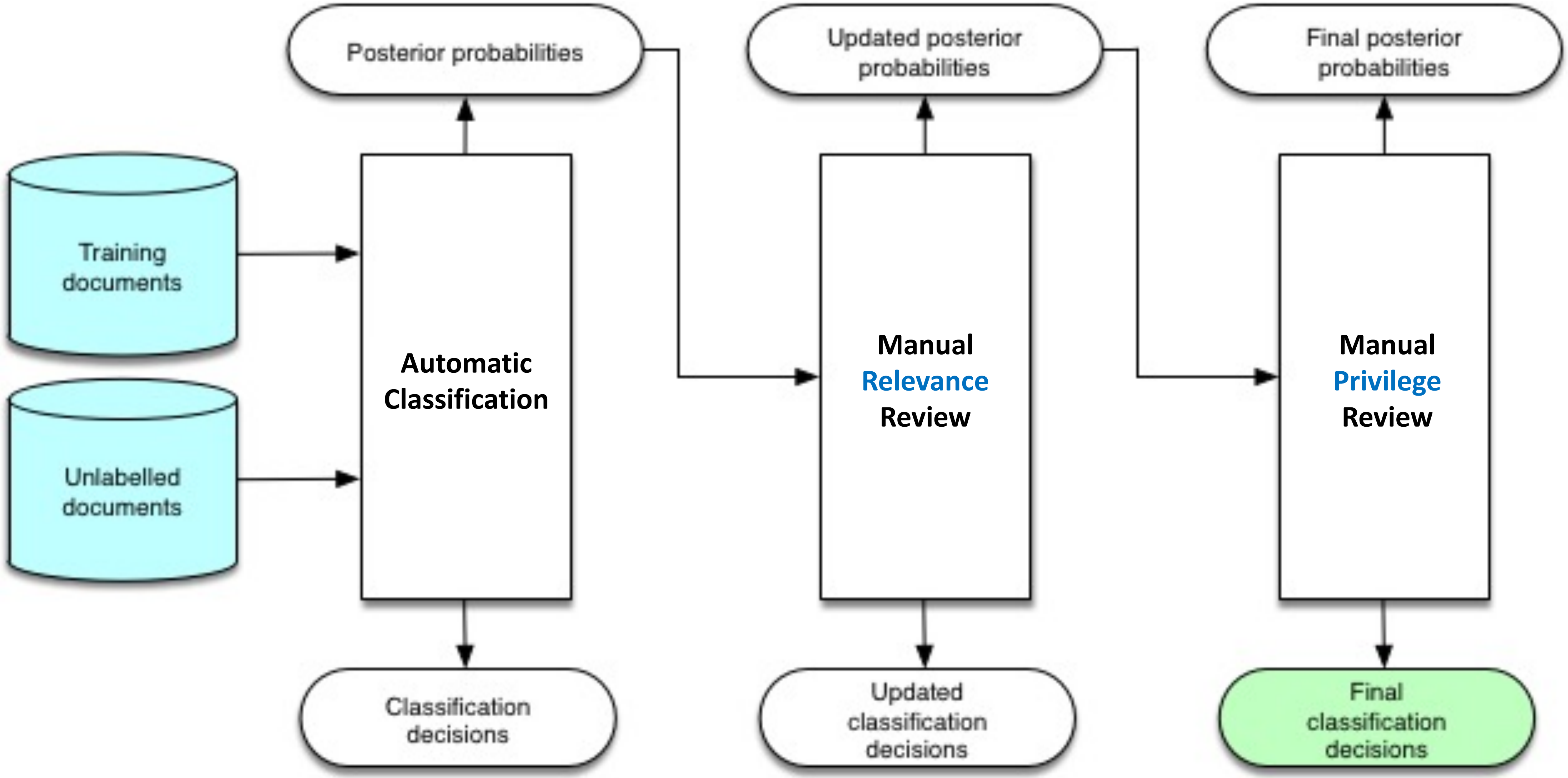


Idea #1: Finite Population Annotation



Berardi, Esuli, Sebastiani: Utility-Theoretic Ranking for Semiautomated Text Classification. ACM TKDD, 2015

Idea #2: Two Manual Review Stages



Idea #3: Task-Based Misclassification Cost

Correct Decision

Prediction	Produce	Log	Withhold
Produce		**$600**	$5
Log	$150		$3
Withhold	$15	$15	

Question # 2 *Consider two types of mistakes:*

LP *Situation:* *Document is responsive and nonprivileged (it should thus be produced)*
Mistake: *Document is erroneously reported on the privilege log and not produced*

PL *Situation:* *Document is responsive and privileged (it should thus be reported on the privilege log and not produced)*
Mistake: *Document is erroneously produced*

Is mistake LP more serious than mistake PL?

☐ *Yes, mistake LP is* ☐ *times more serious than mistake PL.*

☑ *No, mistake PL is* 4 *times more serious than mistake LP.*

☐ *They are equally serious.*

Question # 3 *Consider two types of mistakes:*

LW *Situation:* *Document is nonresponsive (it should thus be withheld)*
Mistake: *Document is erroneously reported on the privilege log (and not produced)*

WL *Situation:* *Document is responsive and privileged (it should thus be reported on the privilege log and not produced)*
Mistake: *Document is erroneously deemed nonresponsive (and thus withheld)*

Is mistake LW more serious than mistake WL?

☐ *Yes, mistake LW is* ☐ *times more serious than mistake WL.*

☐ *No, mistake WL is* ☐ *times more serious than mistake LW.*

☐ *They are equally serious.*

D. Oard, et al., Jointly Minimizing the Expected Costs of Review for Responsiveness and Privilege in E-Discovery, ACM Transactions on Information Systems, 37(1)11:1-11:35, 2018

Expected Misclassification Cost

Cost Per Mistake

Correct Decision

Prediction	Produce	Log	Withhold
Produce		**$600**	$5
Log	$150		$3
Withhold	$15	$15	

Expected Number of Mistakes

Correct Decision

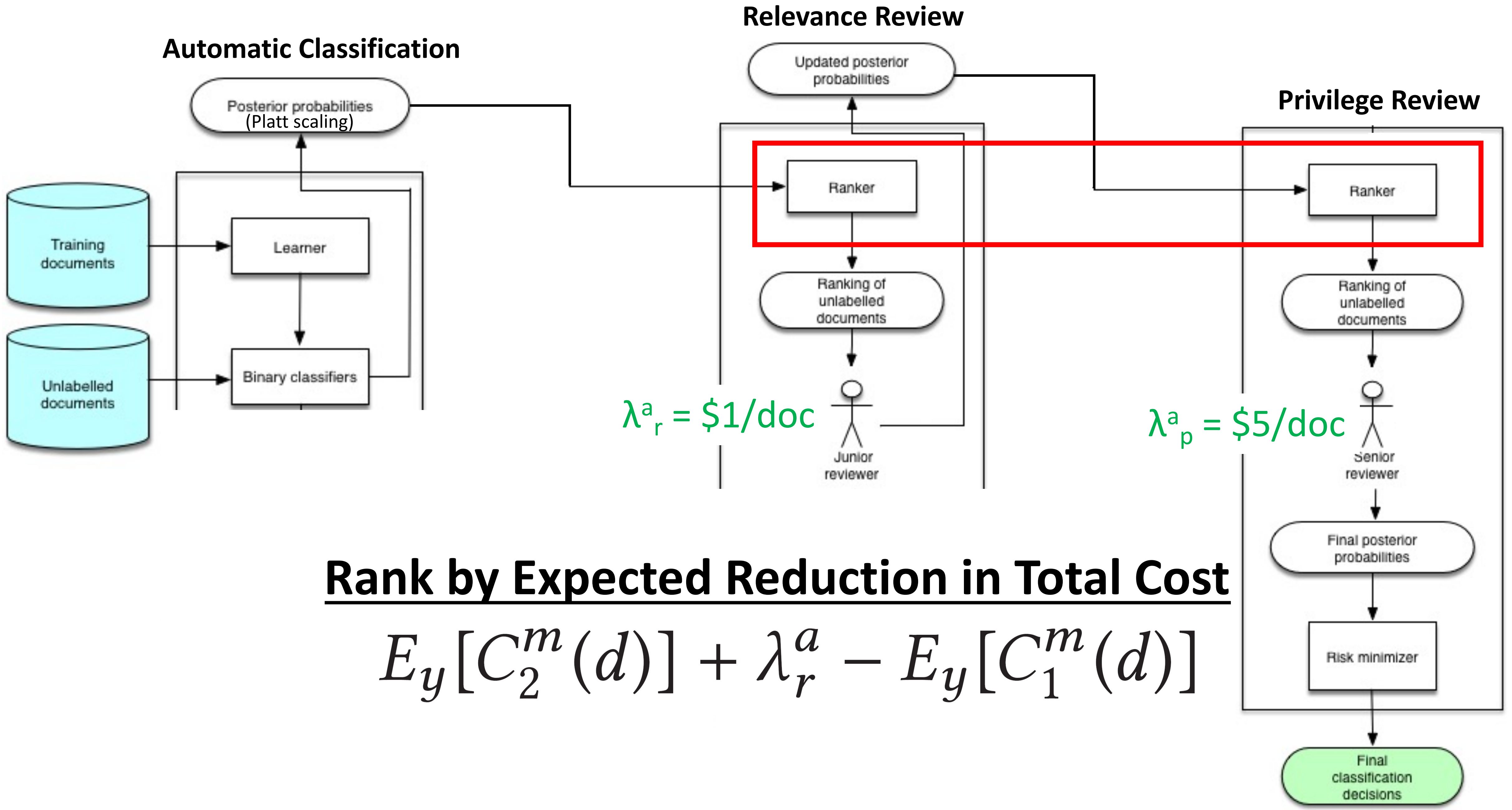
Prediction	Produce	Log	Withhold
Produce		**100**	5
Log	10		1
Withhold	5	1	

Expected Misclassification Cost

Correct Decision

Prediction	Produce	Log	Withhold
Produce		**$60,000**	$25
Log	$1,500		$3
Withhold	$75	$15	

$61,618



Rank by Expected Reduction in Total Cost

$$E_y[C_2^m(d)] + \lambda_r^a - E_y[C_1^m(d)]$$

Evaluation

- **Test Collection**
  - Reuters RCV1-v2 (news stories)
    - Mod-Apte training-test partition (23K train, 200K test)
  - 120 category pairs
    - 24 categories each represent **relevance** (3% to 7%) [e.g., M12: Bond Markets]
    - For each, 5 other categories represent **privilege** (1% to 20%) [e.g., E21: Government Finance]
- **Automatic Classifiers**
  - Linear-kernel SVMs for relevance and privilege
  - Standard term weights for this collection (tfidf:ltc, stemmed, stopped)
- **Manual review**
  - Simulated as perfect judgments (using ground truth)
- **Evaluation measure**
  - Expected Total Cost: manual annotation cost + misclassification cost

D. Oard, et al., Jointly Minimizing the Expected Costs of Review for Responsiveness and Privilege in E-Discovery, ACM Transactions on Information Systems, 37(1)11:1-11:35, 2018

Increase in cost over "Risk Minimization" Cascade

Risk Minimization	Fully Automatic	Active Learning Uncertainty Sampling	Active Learning Relevance Sampling	Fully Manual
0%	+29%	+47%	+52%	+235%

Reviewing as many documents as Risk Minimization

D. Oard, et al., Jointly Minimizing the Expected Costs of Review for Responsiveness and Privilege in E-Discovery, ACM Transactions on Information Systems, 37(1)11:1-11:35, 2018

Section Outline

- End-User search
  - Sensitivity-aware ranked retrieval
- Intermediated search
  - Cost-sensitive prioritization

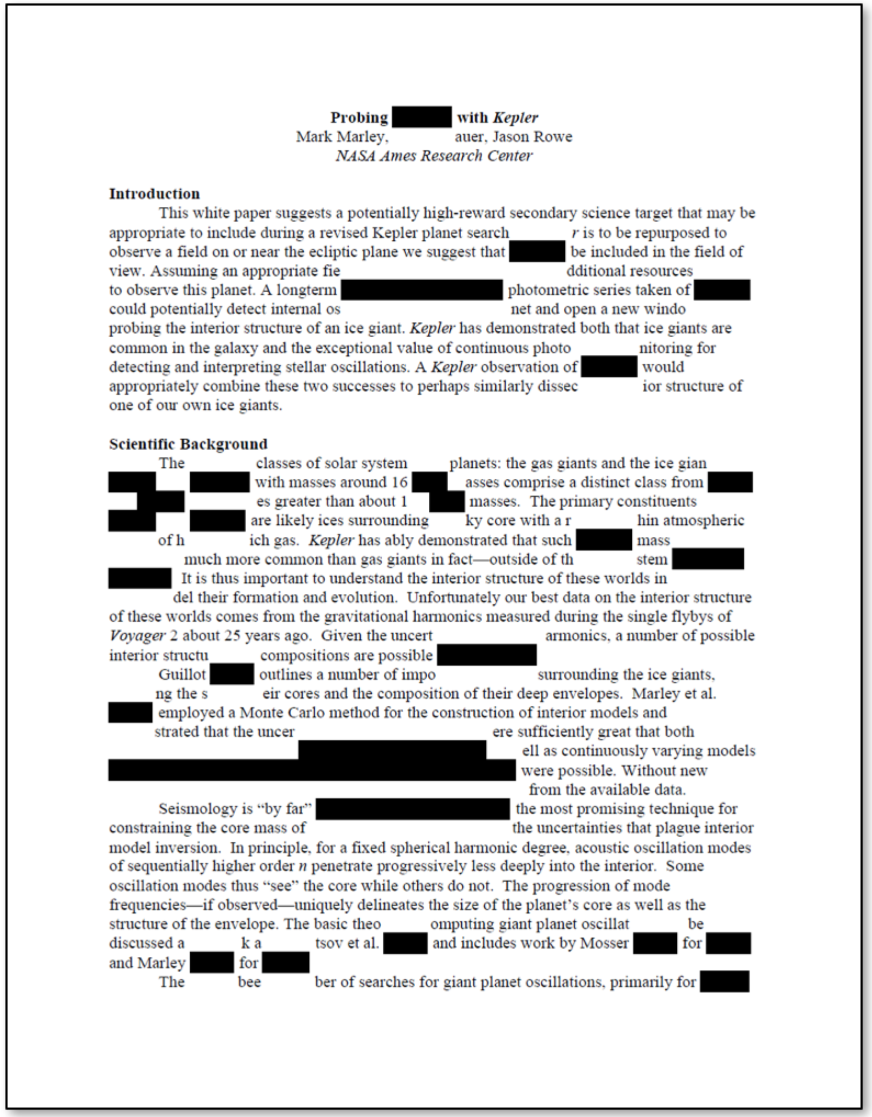
➢ Content protection
  - Redaction, sanitization

Redaction

How can we create automatic redactions that can comply with different redaction policies?

Different policies need to be applied for different types of sensitivity, e.g.,

- FOIA Personal information: redact only the terms which include personal information.
- FOIA International Relations: redact sensitive information and any context that alludes to the sensitivity.



Probing [redacted] with *Kepler*
Mark Marley, auer, Jason Rowe
NASA Ames Research Center

Introduction

This white paper suggests a potentially high-reward secondary science target that may be appropriate to include during a revised Kepler planet search r is to be repurposed to observe a field on or near the ecliptic plane we suggest that [redacted] be included in the field of view. Assuming an appropriate fie dditional resources to observe this planet. A longterm [redacted] photometric series taken of [redacted] could potentially detect internal os net and open a new windo probing the interior structure of an ice giant. *Kepler* has demonstrated both that ice giants are common in the galaxy and the exceptional value of continuous photo nitoring for detecting and interpreting stellar oscillations. A *Kepler* observation of [redacted] would appropriately combine these two successes to perhaps similarly dissec ior structure of one of our own ice giants.

Scientific Background

The classes of solar system planets: the gas giants and the ice gian [redacted] with masses around 16 [redacted] asses comprise a distinct class from [redacted] es greater than about 1 [redacted] masses. The primary constituents [redacted] are likely ices surrounding ky core with a r hin atmospheric of h ich gas. *Kepler* has ably demonstrated that such [redacted] mass much more common than gas giants in fact—outside of th stem [redacted] It is thus important to understand the interior structure of these worlds in del their formation and evolution. Unfortunately our best data on the interior structure of these worlds comes from the gravitational harmonics measured during the single flybys of *Voyager* 2 about 25 years ago. Given the uncert armonics, a number of possible interior structu compositions are possible [redacted] Guillot [redacted] outlines a number of impo surrounding the ice giants, ng the s eir cores and the composition of their deep envelopes. Marley et al. [redacted] employed a Monte Carlo method for the construction of interior models and strated that the uncer ere sufficiently great that both [redacted] ell as continuously varying models were possible. Without new from the available data.

Seismology is "by far" [redacted] the most promising technique for constraining the core mass of the uncertainties that plague interior model inversion. In principle, for a fixed spherical harmonic degree, acoustic oscillation modes of sequentially higher order *n* penetrate progressively less deeply into the interior. Some oscillation modes thus "see" the core while others do not. The progression of mode frequencies—if observed—uniquely delineates the size of the planet's core as well as the structure of the envelope. The basic theo omputing giant planet oscillat be discussed a k a tsov et al. [redacted] and includes work by Mosser [redacted] for [redacted] and Marley [redacted] for [redacted]

The bee ber of searches for giant planet oscillations, primarily for [redacted]

Redacting Personally Identifiable Information

Financial
BANK_ACCOUNT_NUMBER
BANK_ROUTING
CREDIT_DEBIT_NUMBER
CREDIT_DEBIT_CVV
CREDIT_DEBIT_EXPIRY
PIN

Personal
NAME
ADDRESS
PHONE
EMAIL
AGE

Information Systems
USERNAME
PASSWORD
URL
AWS_ACCESS_KEY
AWS_SECRET_KEY
IP_ADDRESS
MAC_ADDRESS

National
SSN
PASSPORT_NUMBER
DRIVER_ID

Other
DATE_TIME

https://aws.amazon.com/blogs/machine-learning/detecting-and-redacting-pii-using-amazon-comprehend/

(α,C)-Sanitization

The patient suffers from **acquired immunodeficiency syndrome** because of a **blood transfusion**. He was diagnosed when his **immune system** responded poorly to **influenza**.

The patient suffers from [redacted] because of a [redacted]. He was diagnosed when his [redacted] responded poorly to [redacted].

The patient suffers from **a long-term condition** because of a **medical procedure**. He was diagnosed when his **body** responded poorly to **an acute illness**.

Given:

- an input document D
- a set of sensitive entities C
- A protection degree $\alpha \geq 1$

We say that D' is a C-sanitized version of D if:

- D' does not contain any group of terms T that in aggregate have
- Pointwise Mutual Information (PMI) with any term $c \in C$
- greater than $-\log p(c)/\alpha$

David Sánchez and Montserrat Batet, C-sanitized: A privacy model for document redaction and sanitization, *JASIST*, 67(1), 2016

(α,C)-Redaction vs. (α,C)-Sanitization

Entity/Wikipedia article	Model instantiation	Redaction	Sanitization
HIV	(1.0, HIV)-sanitized	96.2%	97.2%
	(1.5, HIV)-sanitized	32.9%	66.6%
	(2.0, HIV)-sanitized	17.6%	61.2%
STD	(1.0, STD)-sanitized	95.3%	97.5%
	(1.5, STD)-sanitized	70.0%	85.8%
	(2.0, STD)-sanitized	65.5%	84.5%
Los Angeles	(1.0, Los Angeles)-sanitized	88.3%	94.6%
	(1.5, Los Angeles)-sanitized	54.4%	80.1%
	(2.0, Los Angeles)-sanitized	48.1%	77.5%
New York	(1.0, New York)-sanitized	97.2%	99.2%
	(1.5, New York)-sanitized	39.3%	64.2%
	(2.0, New York)-sanitized	20.1%	58.3%
Homosexuality	(1.0, Homosexuality)-sanitized	92.9%	97.5%
	(1.5, Homosexuality)-sanitized	55.3%	81.3%
	(2.0, Homosexuality)-sanitized	50.4%	77.2%
Catholicism	(1.0, Catholicism)-sanitized	96.3%	98.1%
	(1.5, Catholicism)-sanitized	41.3%	73.4%
	(2.0, Catholicism)-sanitized	30.9%	65.8%

David Sánchez and Montserrat Batet, C-sanitized: A privacy model for document redaction and sanitization, *JASIST*, 67(1), 2016

Correspondents

Gwen Adams (3)

Jill Carter K... (2)

Sender

Owner (2)

Accessions

ARCH 2019-070... (3)

ARCH 2019-070... (1)

SORT BY

Date: Sep 22, 2013 6:11pm

From: Maggie McLoughlin <mam@...>

To: gadams3702@...>

Bcc: <mam@...>

Subject: ....

ID 4 0

.... .,

(.) "..... ..." ....--., .. .:.. .. .:... [......... ..
Bogart/ Bacall "... ....." .. .:...]

(.) Jill ....
.... Kunsthistorisches Museum. ... ...
.. . .?

...., .

Tutorial Outline

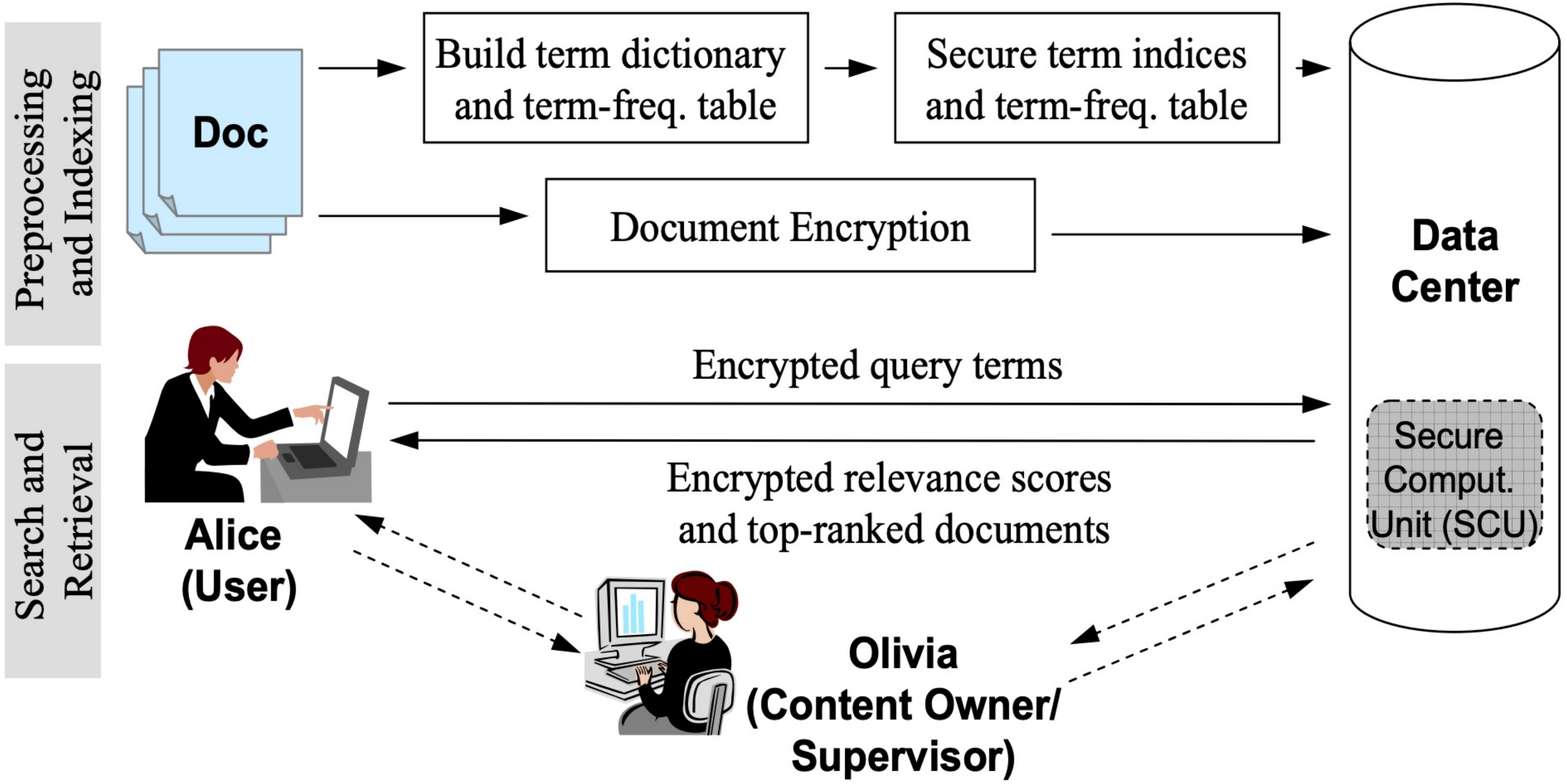
CET

- 14:15 Background
- 14:45 Evaluation
- 15:20 Detecting sensitive content
- 16:00 Protecting Sensitive Content
- **16:15 Break**
- 16:45 Protecting Sensitive Content
- 17:00 Other Issues
- 17:20 Two Design Sprints ("choose your ending")
- 17:55 Wrap up
- **18:15 End!**

Section Outline

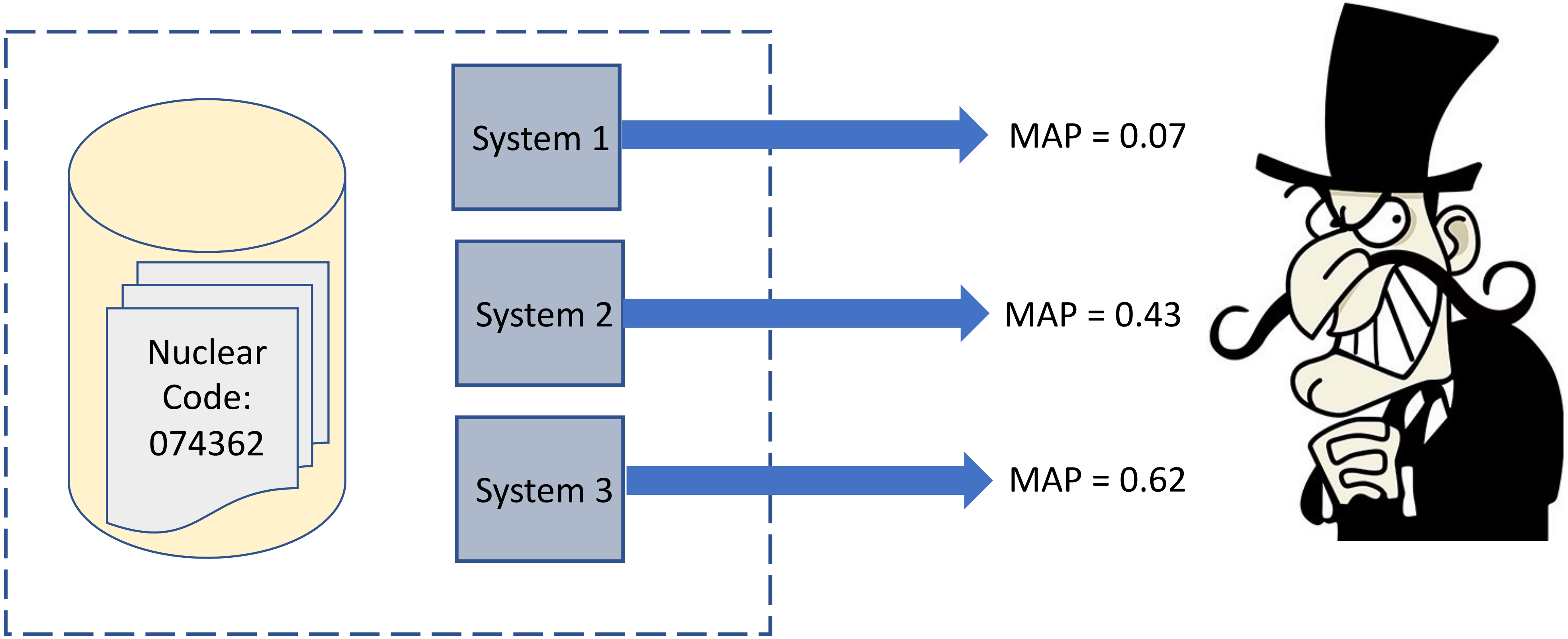
- Encrypted search
- Mosaicing
- Algorithm deposit leakage

Private Information Retrieval: Encrypted Search

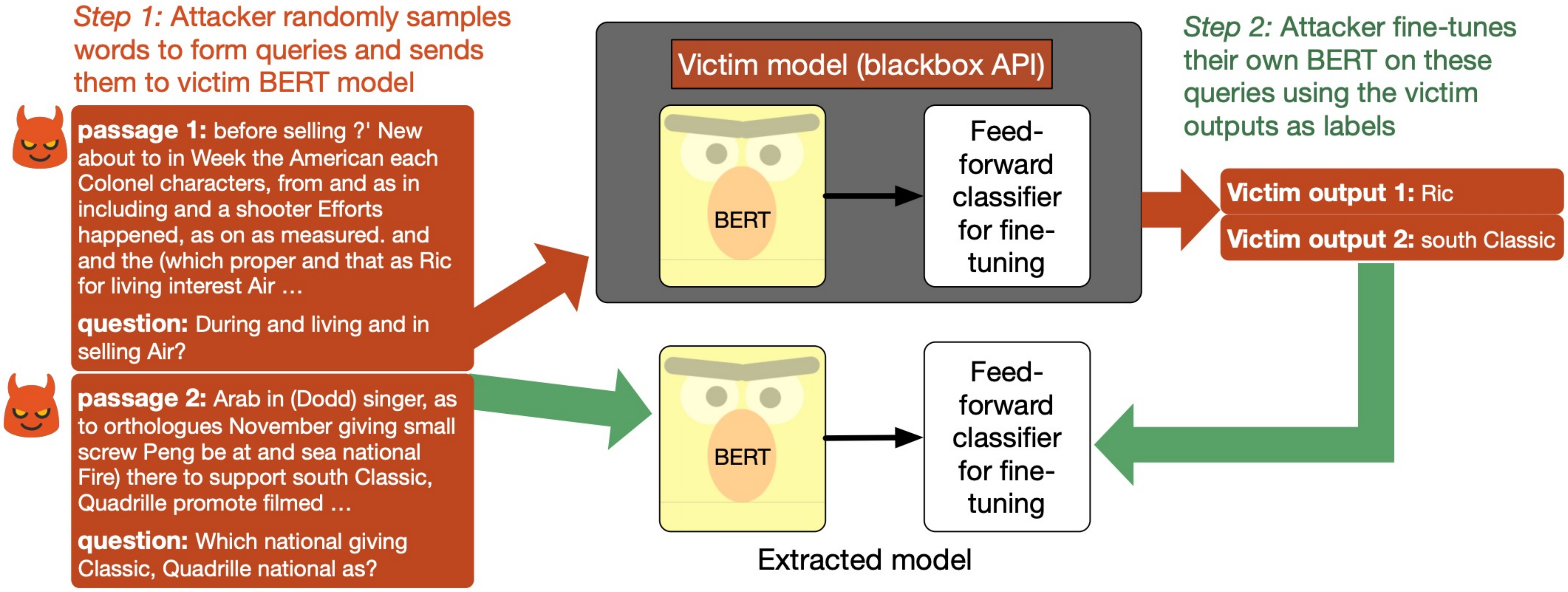


A. Swaminathan, et al., Confidentiality-Preserving Rank-Ordered Search, ACM Workshop on Storage, Security and Survivability, 2007.

Algorithm Deposit Leakage



Sesame Street-Based Retrieval



Tomar, et al., Thieves on sesame street! model extraction of BERT-based APIs, ICLR, 2020

Mosaicing

~~TOP SECRET~~

644..A

HISTORY OF THE CUSTODY AND DEPLOYMENT OF NUCLEAR WEAPONS (U)

JULY 1945 THROUGH SEPTEMBER 1977

Controlled Document
Certificate of Destruction Required

PREPARED BY
OFFICE OF THE ASSISTANT TO THE SECRETARY OF DEFENSE (ATOMIC ENERGY)
FEBRUARY 1978

"EXCISED UNDER ... OF (THE FREEDOM OF INFORMATION ACT) 5USC552 (b) (1) and (b)(3)"

TS-A94-0031
CLASSIFIED BY: ATSD(AE) AND CG-W-4

Copy 11

~~TOP SECRET~~

RESTRICTED DATA
This document contains Restricted Data as defined in the Atomic Energy Act of 1954, its dissemination or disclosure to any unauthorized person is prohibited.

#306

R-1 94-F 0572(?)
~~49-F-1525~~
~~94-TS-D0~~

~~TOP SECRET~~

COUNTRY	WEAPON	INITIAL ENTRY	WITHDRAWN
Guam (cont.)	Talos	Jul 65	Jun 69
	Astor	Nov 65	Mar 74
	ASROC	Jan 66	
	Terrier	Mar 66	Jan 67
	155mm Howitzer	May 66	
	Polaris	Jul 66	Aug 66
	Nike Hercules	Jun 68	Jun 69
Hawaii	Bomb	Jul 54	Jun 69
	Depth Bomb	Dec 55-Feb 56	
	Regulus	Mar-May 56	Jan-Mar 65
	Boar	Sep-Nov 56	Apr-Jun 63
	Honest John	Jun-Aug 57	Jun 75
	8-inch Howitzer	Oct-Dec 58	Jun 72
	ADM	Jan-Mar 59	Jun 75
	Hotpoint	Jan-Mar 60	Oct-Dec 64
	Nike Hercules	Jul-Sep 60	Jun 73
	Little John	Apr-Jun 62	Oct 68
	Talos	Oct-Dec 63	Aug 68
	ASROC	Oct-Dec 63	
	Astor	Apr-Jun 64	
	Davy Crockett	Apr-Jun 64	Jun 69
	155mm Howitzer	Oct-Dec 64	Jun 75
	Terrier	Mar 65	Sep 66
	Subroc	Aug 65	
	Falcon	May 66	Jun 67
[redacted]	Nonnuclear Bomb	Feb 56	Jun 66
	Bomb	Sep 56	Sep-Dec 59
[redacted]	Nonnuclear Bomb	Dec 54-Feb 55	Jun 65
Johnston Is.	Thor	Jul-Sep 64	Jun 71
[redacted]	Nike Zeus	Jul-Dec 63	Jul 66
Midway	Depth Bomb	Jul 61	Jun 65
[redacted]	Nonnuclear Bomb	Jul-Sep 53	Jun 65
	Bomb	May 54	Sep 63
	Depth Bomb	Sep-Nov 57	Mar 61

H

- Haiti
- Holy See
- Honduras
- Hong Kong
- Hungary

Top of Page

I

- Iceland
- India
- Indonesia
- Iran
- Iraq
- Ireland
- Israel
- Italy

Top of Page

J

- Jamaica
- Japan
- Jordan

The "Mosaic Theory"

Iceland. Iceland is another "non-nuclear" country whose nuclear history remains incomplete. In Appendix B, **<u>Iceland is clearly the first blacked out country listed after Hawaii and before Johnston Island.</u>** Non-nuclear components were stored at the American base at Keflavik for a decade, from February 1956 to June 1966, and complete nuclear bombs were deployed there from September 1956 to September-December 1959.

Norris et al. (1999), Where They Were, *Bulletin of Atomic Scientists*, 55(6), 25-35

Tutorial Outline

CET

- 14:15 Background
- 14:45 Evaluation
- 15:20 Detecting sensitive content
- 16:00 Protecting Sensitive Content
- **16:15 Break**
- 16:45 Protecting Sensitive Content
- 17:00 Other Issues
- 17:20 Two Design Sprints ("choose your ending")
- 17:55 Wrap up
- **18:15 End!**

Design the CLEF-2022 “Shhh Task”

Designing the Shhh Task

- Task(s)
  - Sensitive content detection? Sensitivity-aware ranking? Set retrieval?
- Evaluation Framework
  - Algorithm deposit? Distributable test collection?
- Test Collection
  - Government records? Business email? Conversational speech?
  - Queries
  - Sensitivities
  - Relevance judgments
- Training Data

Mosaicing Research Framework

Mosaicing Research Framework

- Design an experimentation system/platform/framework for developing and evaluating approaches for protecting against mosaicing attacks.
  - What are the motivating research questions?
  - System Architecture Diagram
  - Evaluation metrics?
  - Baselines approaches?
  - What test collections can be used?
    - How to collect annotations?

Tutorial Outline

CET

- 14:15 Background
- 14:45 Evaluation
- 15:20 Detecting sensitive content
- 16:00 Protecting Sensitive Content
- **16:15 Break**
- 16:45 Protecting Sensitive Content
- 17:00 Other Issues
- 17:20 Two Design Sprints (“choose your ending”)
- 17:55 Wrap up
- **18:15 End!**

The Technology – Policy Design Space

- Without adequate technology, some practices are impractical
- Without adequate policy, some technologies are insufficient

Closing Thoughts

- We’re still in the early days
- Existing work has been non-neural
- Its not just digital text; speech is the killer app
- It need not be perfect to be useful
  - But it does need to be pretty darn good

Closing Thoughts

- Who else should we be talking with?
    - What channels of communication need to be opened?
- Who do we need to work with?
    - Made progress working with Government, Layers
    - Who else? Information scientists, cryptography, politicians, social scientists, …
- What problems are of most interest to the IR community?
- What are the most important / timely problems to address
    - What's the next *low hanging fruit*?

Search Among Sensitive Content

ECIR 2021 Tutorial

Graham McDonald, University of Glasgow, UK
Graham.mcdonald@glasgow.ac.uk

Douglas W. Oard, University of Maryland, USA
oard@umd.edu

Reading list available from

Search-Among-Sensitive-Content.GitHub.io

Search Among Sensitive Content

ECIR 2021 Tutorial

Graham McDonald and Douglas W. Oard

University of Glasgow (UK), University of Maryland (USA)
graham.mcdonald@glasgow.ac.uk, oard@umd.edu

March 28, 2021

Bibliography

Anandan, B., Clifton, C., Jiang, W., Murugesan, M., Pastrana-Camacho, P. & Si, L. (2012), 't-Plausibility: Generalizing words to desensitize text', *Transactions on Data Privacy* **5**(3), 505–534.
URL: *https://www.cs.purdue.edu/homes/lsi/TDP_2012.pdf*

Bagga, A. & Baldwin, B. (1998), Entity-based cross-document coreferencing using the vector space model, *in* C. Boitet & P. Whitelock, eds, '36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, COLING-ACL '98, August 10-14, 1998, Université de Montréal, Montréal, Quebec, Canada. Proceedings of the Conference', Morgan Kaufmann Publishers / ACL, pp. 79–85.
URL: *https://www.aclweb.org/anthology/P98-1012/*

Supported in part by US NSF grant 1618695