Beyond the Parameters:

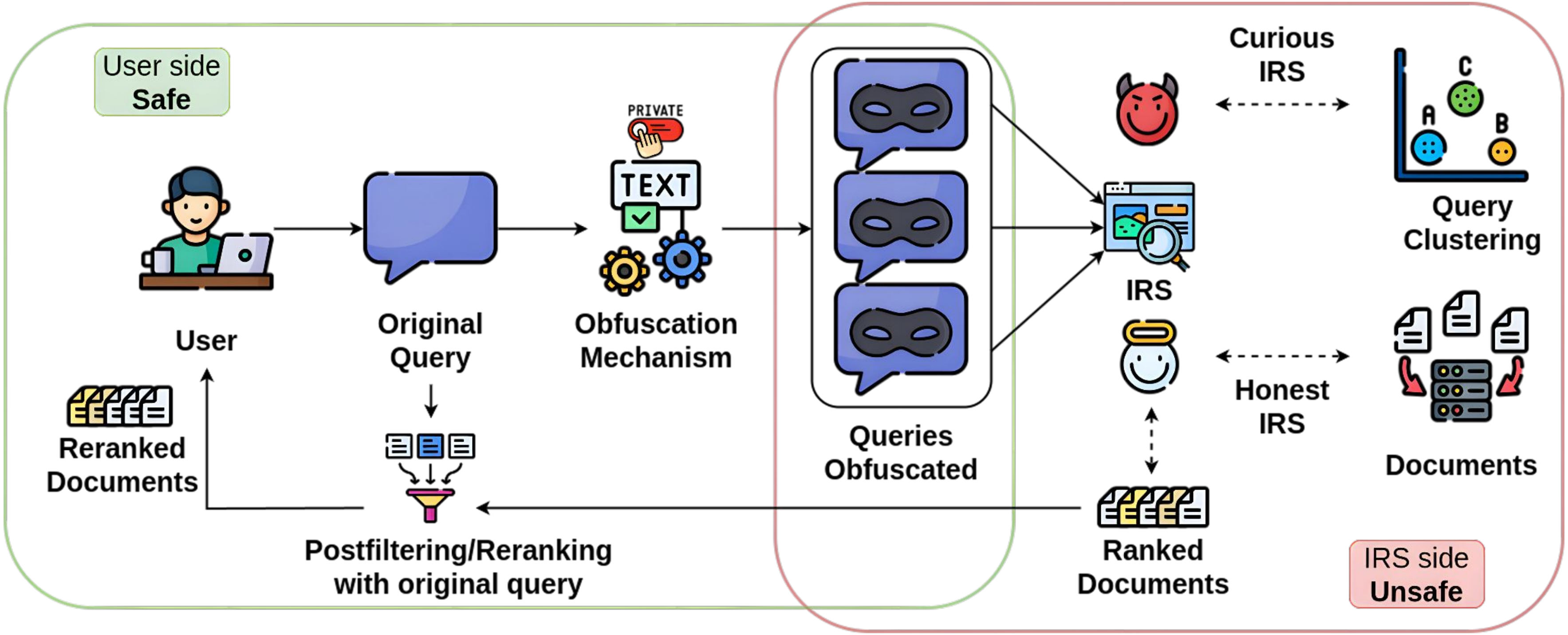
Measuring Actual Privacy in Obfuscated Texts

Francesco L. De Faveri
Guglielmo Faggioli
Nicola Ferro

IIR 2024
14th Italian Information Retrieval Workshop
September 5-6, 2024 - Udine, Friuli Venezia Giulia, Italy

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

IR & Privacy


User side
Safe
User
Original
Query
PRIVATE
TEXT
Obfuscation
Mechanism
Queries
Obfuscated
Curious
IRS
C
A
B
Query
Clustering
IRS
Honest
IRS
Documents
Reranked
Documents
Postfiltering/Reranking
with original query
Ranked
Documents
IRS side
Unsafe

Generating the obfuscated queries

Obfuscation Mechanism

ε - Differential Privacy Based

Heuristic Based

E.g. Using the CMP mechanism with ε = 12.5:
“do goldfish grow” –> “do xlvi grow”, “host frangieh expands”, “do goldfish grow”

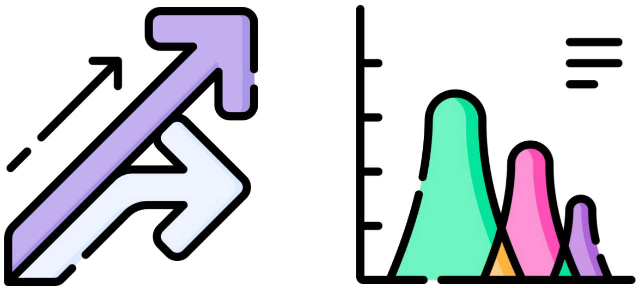
ε - Differential Privacy obfuscation

Considering any pair neighbouring datasets*, *D* and *D'*, a **privacy budget** $\varepsilon \in \mathbb{R}^+$, a mechanism is ε - Differentially Private if it holds:

$$\Pr\left[\mathcal{M}(D) \in S\right] \leq e^{\varepsilon} \cdot \Pr\left[\mathcal{M}(D') \in S\right] \quad \forall S \subset \mathrm{Im}(\mathcal{M})$$

Applied to **textual data**:

- *Embedding Perturbation*: CMP (2020), Mahalanobis (2020), Vickrey (2021)
- *Sampling*: CusText (2023), SanText (2021), TEM (2023)



*Datasets that differs for at most one record.

Heuristic based obfuscation

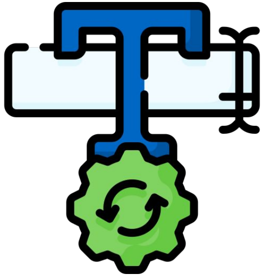
Arampatzis et al. (2013)

Obfuscate original query terms with **synonyms**, **hypernyms**, and **holonyms** from Wordnet.

E.g.:

"Cat" –> "**Feline**"

"Cancer" –> "**Disease**"

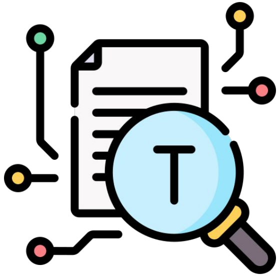


Fröbe et al. (2021)

Generates **keyword** queries using a small local corpus to *obfuscate the original information need of the user*.

E.g.:

"A user wants health advice while <u>hiding a potential disease</u>." submits to the IRS queries like "**lower heart rate**", "**forearm pain**", "**symptoms heart attack**".



What is the privacy?

Privacy

Formal Privacy

Privacy provided by the **definition** of the mechanism.
(e.g. the ε in Differential Privacy)

Actual Privacy

The amount of privacy that is provided by the mechanism **in reality**.

How to measure privacy?

Privacy

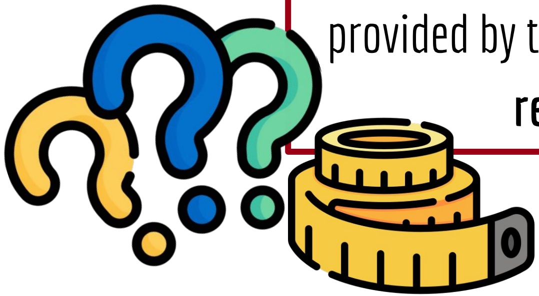
Formal Privacy

Privacy provided by the **definition** of the mechanism.

(e.g. the ε in Differential Privacy)

Actual Privacy

The amount of privacy that is provided by the mechanism **in reality**.



Strategies to measure actual Privacy

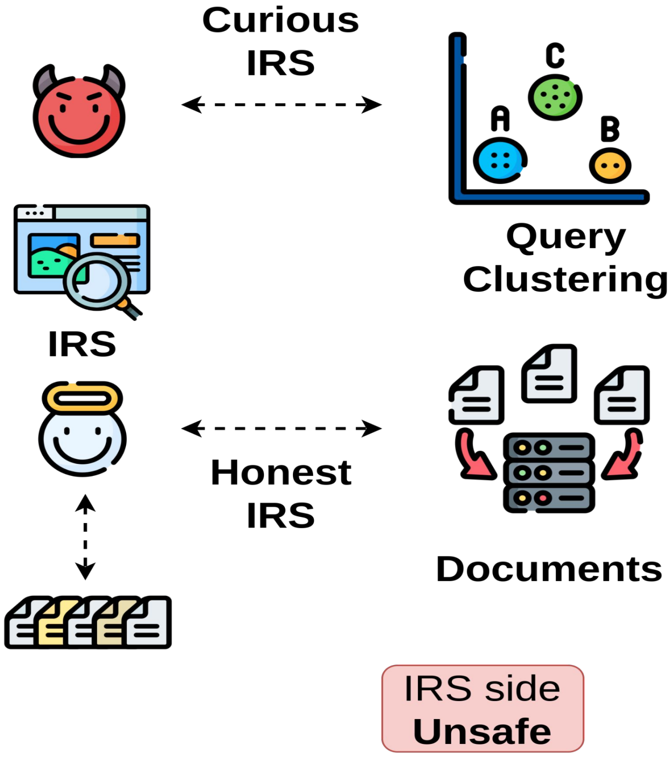
Technique	Measure	PROs	CONs
Translation Based	*BLEU*	- Effective short n-grams comparison - Computationally efficient	- No semantic similarity - Long distance dependences
	ROUGE	- n-gram comparison - Computationally efficient	- No semantic similarity - Long distance dependences
	METEOR	- Synonyms and stemming - Paraphrasing	- Heuristic-based - Computational expensive
Lexical Based	*Jaccard Index*	- Lexical similarity - Computationally efficient	- No semantic similarity - No synonyms or paraphrasing
	N_w and S_w	- Lexical similarity - Measure of mechanism failure	- No semantic similarity - Easy to deceive
Contextual Based	*BERTScore*	- Semantic similarity - Long distance dependences	- Pre-trained model dependent - Computational expensive
	Transformers Sentence Embeddings Similarity	- Semantic similarity - Long distance dependences	- Pre-trained model dependent - Computational expensive

Adversarial Risk of Breaking Privacy

How can we model the risk?

Risk = Probability of the attack x Vulnerability x Impact

Can the privacy metrics be **good indicators** for such risk?



Future directions

- Are the current measure enough to evaluate privacy?
- Do we need "human assessment" to evaluate privacy (e.g. Privacy relevance judgments) ?
- How can we understand if the measure is a good proxy for the probability of success for a class of attacks?

Beyond the Parameters:

Measuring Actual Privacy in Obfuscated Texts

Thanks for the attention!
Question Time

Francesco L. De Faveri [1]
Guglielmo Faggioli [1]
Nicola Ferro [1]

1: Department of Information Engineering, University of Padova, Padova, Italy

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Backup Slides

Beyond the Parameters:

Measuring Actual Privacy in
Obfuscated Texts

Francesco L. De Faveri [1]
Guglielmo Faggioli [1]
Nicola Ferro [1]

1: Department of Information Engineering, University of Padova, Padova, Italy

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Backup 1 - DIfferential Privacy Mechanisms

Strategy	Mechansim	Params	Description
Embedding	CMP	-	The noise is sampled from an n - dimensional Laplace distribution of scale $\frac{1}{\varepsilon}$.
	Mhl	λ	The noise is sampled from an n - dimensional Normal distribution defined by the λ regularized Mahalanobis norm of the term embedding.
	Vickrey	t, λ	The noise is sampled as defined by the parent mechanism (CMP or Mhl) and the obfuscation term is set based on a free parameter t.
Sampling	CusText	K	Sampling of a new term is bounded to K possible terms picked using the scores computed using the distances among word embeddings.
	SanText	-	Sampling of a new term is computed with a score based on the distances among embeddings, with terms closer to the obfuscation having a higher probability.
	TEM	β	Noise sampled from an n - dimensional Gumbel distribution is added to the scores, and the final obfuscation term is sampled accordingly to the maximum noisy score.