Preserving Privacy When Processing User Queries in IR.

Doctoral Consortium Paper

Francesco Luigi, De Faveri[1]

[1]Department of Information Engineering, University of Padova, Padova, Italy

Abstract

Data represents one of the most crucial assets of today's digital age. Privacy-preserving strategies play a crucial role in safeguarding the confidentiality of sensitive user data during the overall processing pipeline in Natural Language Processing (NLP) and Information Retrieval (IR) tasks. This paper presents an overview of obfuscation strategies and evaluation metrics employed to process users' textual information privately when interacting with IR systems, framing these solutions within the formal framework of ε-Differential Privacy (DP). The methodologies and findings presented in this paper describe the author's preliminary studies in his current PhD activity.

Keywords

Privacy Preserving Information Access, Differential Privacy, Information Retrieval, Information Security.

1. Introduction

Data has become one of the most valuable resources for researchers and industry in today's digital age. In such a scenario, an ever-growing amount of data for training, validation, and testing is needed to enhance the performance of NLP and IR systems. This includes highly sensitive and personal information, such as health records [1, 2], financial situations [3], and individual preferences [4], all of which are used to refine and enhance models' performance. For example, when a user interacts with an IR system, like a search engine, the information need is formulated into a natural language query. When the search engine processes such a query to retrieve relevant documents, confidential information, like the motivations of the search and personal identifiers, e.g., social security number and other personal attributes for *ego-surfing* [5], can be extracted and analysed [6, 7], thus presenting the user with the dilemma of exchanging personal information in order to retrieve relevant ones.

Recent works in NLP and IR [8, 9, 10] have shown the potentialities of applying the formal ε-Differential Privacy (DP) framework [11] to provide privacy guarantees to textual data employing obfuscation mechanisms. In this context, an obfuscation mechanism is an algorithm that, upon receiving a text as input, randomly produces another text composed of different words as output. In ε-DP, the number of changed words and the semantic relations with the original texts depend on the ε value, which sets the statistical noise used during the output computation. However, introducing ε-DP mechanisms to obfuscate the real meaning of texts poses some open research challenges. State-of-the-Art DP mechanisms do not guarantee that given a term x is changed in loose privacy regimes, i.e., high values of ε. In addition, standard evaluation measures pivot the analysis on varying the formal privacy budget ε, leading to extreme cases where a low ε, i.e., a strong privacy setting, may result in preserving the original text, thus giving a false perception of the privacy granted [12].

In this paper, submitted to the Doctoral Consortium track, the author reports the methodology proposed in previous works [13, 14, 15, 16] to address the above open challenges, providing robust privacy guarantees for texts proposing an obfuscation mechanism based on the ε-DP framework which ensures removing original words from the obfuscated output produced. Moreover, to address the problem of measuring actual privacy beyond the formal privacy budget ε, we report the privacy analysis in an adversarial scenario where the attacker exploits a public query log to infer the original query.

SEBD 2025: 33rd Symposium on Advanced Database Systems, June 16–19, 2025, Ischia, Campania, Italy

francescoluigi.defaveri@phd.unipd.it (F. L. De Faveri)
https://www.dei.unipd.it/~defaverifr/ (F. L. De Faveri)
0009-0005-8968-9485 (F. L. De Faveri)

The paper is structured as follows: Section 2 presents the related work on providing and measuring text privacy, also presenting the query obfuscation protocol in IR. Section 3 explains the methodology used to ensure privacy for textual queries and the method proposed to evaluate actual privacy. Finally, Section 4 outlines the findings of prior studies, and Section 5 concludes, highlighting open challenges.

2. Related Work and Background

Obfuscating Texts with ε-DP. ε-DP framework was introduced by Dwork et al. [11] to formalize the privacy guarantees when releasing data publicly. Given a privacy budget $\varepsilon \in \mathbb{R}^+$, and any pair of neighbouring datasets D, D', i.e., datasets that differ for only one entry, an obfuscation mechanism $\mathcal{M}$ is DP if it holds the inequality $\Pr[\mathcal{M}(D) \in S] \leq e^{\varepsilon} \cdot \Pr[\mathcal{M}(D') \in S] \ \forall S \subset \mathrm{Im}(\mathcal{M})$. DP introduces calibrated noise levels during output computation using the privacy budget ε, which controls the balance between data privacy and utility. The adoption of the DP framework for metric spaces, and therefore for NLP tasks, has been proposed in [17]. Metric-DP extends the traditional DP definition by ensuring that the probability of obfuscating two distinct points x, x' is proportional to the distance $d(x, x')$ between them. The DP framework has enabled the privacy research community to develop two main obfuscation strategies: either based on leveraging noisy embeddings or randomly sampling a new obfuscation term. The former approaches involve introducing statistical noise into text term embeddings based on the ε budget like in the Cumulative Multivariate Perturbation (CMP), Mahalanobis (Mhl), and their respective Vickrey-based variant mechanisms [18, 19, 20]. The latter employs random sampling to select a term as the obfuscated text, like in the Custumized Text (CusText), Sanitization Text (SanText), and Truncated Exponential (TEM) mechanisms [21, 22, 23]. For full details, we recommend the original papers.

Measuring Privacy. Wagner and Eckhoff [24] systematically classified over eighty privacy metrics, offering a comprehensive framework for assessing privacy across different domains, e.g., communication, databases, and social networks. The work proposes specific aspects of privacy that a metric aims to quantify, suggesting nine guiding questions for selecting the appropriate privacy measures. Specifically, the authors underlined the importance of considering the adversary's knowledge and capability when evaluating privacy. In addition, Sousa and Kern [25] described how different mechanisms developed for NLP tasks provide privacy for textual data with Habernal [26] stressing the importance of not relying strictly on formal analysis of DP in its application on NLP, encouraging research towards new privacy metrics. Traditional privacy measures focus on calculating the failure rates of obfuscation mechanisms [27] or assessing the similarities between original and obfuscated texts [28, 9]. Uncertainty measures such as N_x and S_x [18, 19] estimate the probability that a term x remains unchanged after obfuscation and the minimum cardinality of the set of words to which the mechanism maps x, respectively. The similarity between the input and output texts is commonly estimated using metrics like the Jaccard or cosine similarity between sentence embeddings computed by a Transformer.

The Query Obfuscation Protocol in IR. Figure 1 reports a high-level view of the query obfuscation protocol, considering two distinct sides: one for the user ("Safe Side") and one for the IR system ("Unsafe Side"). On the user side, the original query is formulated considering the User information need and privatized using an obfuscation mechanism, i.e., an algorithm that, given an original sensitive query, generates different non-sensitive obfuscated queries that (theoretically) prevent the unveiling of the original information need and still can retrieve relevant documents from the system for the user, without explicitly disclosing their information need. On the IR system side, documents are retrieved considering the queries received. If the obfuscation has been correctly performed, relevant documents to the user's original query are placed at a lower rank in the resultant document list (yellow documents in Figure 1), thus masking the actual intentions of the user. Once the list returns to the user, the latter can privately use its original query to re-rank the documents, placing the correct relevant ones first in the final list. The scenario studied works under the assumption of an IR system that does not collaborate to

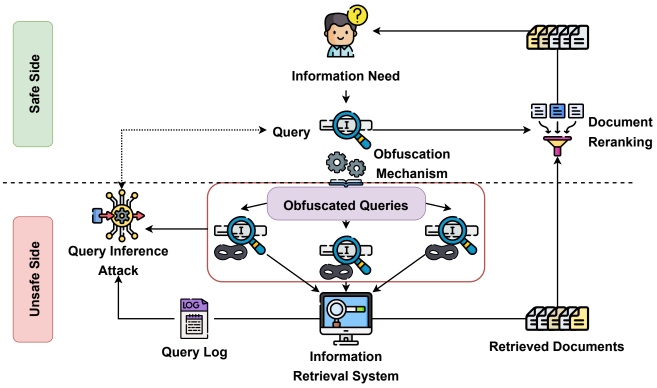

Figure 1: Query Obfuscation protocol overview. On the "Safe Side", the user information need is formulated into a natural language query privatized by a mechanism, producing different obfuscations. Such obfuscated queries are sent to the IR System, which retrieves documents relevant to the obfuscated variants received in the "Unsafe Side". Such documents are sent to the user as the response to the retrieval process. In an adversarial setting, the obfuscated queries are used to perform a Query Inference Attack, employing an available query log.

protect the privacy of the received user query. Therefore, the user is willing to renounce part of the effectiveness of the search to protect his privacy.

A final remark to consider is the use of cryptographic protocols, such as Private Information Retrieval (PIR) protocols, to ensure privacy when interacting with IR systems. This approach introduces open challenges and limitations considering the higher computational demand and time needed to retrieve relevant information from the systems [29, 30]. However, implementing PIR protocols can be seen as complementary to query obfuscation protocols. While query obfuscation focuses on concealing the user's true intent and altering the original query sent to the system, PIR protocols can interact with the system's index, ensuring that the documents are retrieved without revealing sensitive information.

3. Proposed Methodology

3.1. Obfuscating a Text: The Words Blending Boxes (WBB) ε-DP mechanism.

Current state-of-the-art obfuscation mechanisms either ensure the privacy of obfuscated queries by providing formal privacy guarantees under the DP framework or account for the presence of synonyms and holonyms. The WBB mechanism [13] addresses the limitations of these approaches by integrating both strategies. Specifically, the mechanism ensures that the top-k most semantically similar words–i.e., synonyms and holonyms closely positioned to the original term in the embedding space–are excluded from the obfuscation process. Instead, it selects the n words that are similar but do not belong to the top-k set as obfuscation candidates. The final obfuscation term is then sampled according to the DP exponential mechanism [31], which defines the selection probability based on the privacy parameter ε.

3.2. Evaluating Privacy: The Query Inference for Privacy and Utility (QuIPU) Score.

Traditional methods (see Section 2) often rely on theoretical privacy guarantees, such as those provided by the ε in DP, which may not accurately reflect the real-world privacy risks associated with obfuscated queries. The QuIPU score [15] addresses this gap by assessing the extent to which an obfuscated query hides the user's original intent from potential adversaries. Specifically, the score evaluates different obfuscation strategies by examining both the utility of the obfuscated query in performing the intended task and the risk of re-identification. The computation of risk probabilities in the QuIPU framework is grounded in assessing the effectiveness of adversarial strategies that attempt to reverse engineer the original user intent using a Transformer model to cluster obfuscated queries and an available query log. The probability of successfully reconstructing the original query is computed based on its rank among the most similar queries within the log, following the adversary's clustering of the obfuscated queries.

4. Preliminary Experimental Findings

The mechanisms based on ε-DP are tested on TREC collections using the Python package `ir_datasets`[1]. Specifically, we used the TREC Deep Learning‘19 [32] (DL‘19) and Deep Learning‘20 [33] (DL‘20) collections, thus considering 43 and 54 queries. In addition, to understand the impact on a different distribution of the queries, we also employed the obfuscations on the TREC Robust collection [34] (Robust ‘04), containing 250 queries. For each privacy setting of the mechanisms, i.e., $\varepsilon \in \{1, 5, 10, 15, 20, 25, 30, 50\}$, each query produces 20 different variant obfuscations, as done in [9]. To generate such obfuscations and measure the privacy guarantees provided, we employed the pyPANTERA framework [14], leaving as default vocabulary the words and embeddings 300-d from GloVe [35]. Moreover, to compute the QuIPU score, we analysed the scenarios described in [15] of three different attackers, i.e., *Lazy-Active-Motivated*, using as query log the AOL-dataset[2]. To avoid encumbering, we report the performance analysis only on the DL‘19, using as IR system the Contriever model [36] for the retrieval and reranking. We refer to the original papers [13, 14, 15] for the full results.

4.1. Performance Analysis

Evaluating obfuscation mechanisms, Table 1, across different privacy budgets ε reveals a clear trade-off between formal privacy and utility gained by the user during the retrieval pipeline, measured as Precision (P) and normalized Discounted Cumulative Gain (nDCG) at cut-off point 10. At low ε values, the obfuscation is performed in a strong privacy regime, reducing performance in both ranking metrics for all the mechanisms analysed, except for CusText and WBB. Among the tested mechanisms, embedding-based methods show significant improvements as ε increases, achieving stable performance at higher ε values. On the other hand, sampling-based mechanisms offer different behaviours, with TEM maintaining consistently high performance across all privacy budgets. Generally speaking, for the ε considered, the sampling mechanisms are not influenced by the formal parameter ε above 5.

Table 1
Performance analysis comparing different mechanisms at different levels of Privacy, guaranteed by the parameter ε, obtained by employing as retrieval and renaker model the Contriever systems [36] on the DL‘19 collection [32].

		P@10								nDCG@10							
Obfuscation	**Mechanism**	ε - Privacy Budget								ε - Privacy Budget							
		1	5	10	15	20	25	30	50	1	5	10	15	20	25	30	50
Embedding	CMP	0.002	0.047	0.686	0.709	0.714	0.712	0.712	0.712	0.005	0.039	0.585	0.595	0.601	0.599	0.599	0.599
	Mahalanobis	0.000	0.033	0.488	0.707	0.709	0.709	0.712	0.712	0.000	0.034	0.410	0.593	0.595	0.595	0.598	0.599
	VickreyCMP	0.000	0.056	0.595	0.707	0.716	0.716	0.719	0.709	0.000	0.047	0.493	0.594	0.603	0.604	0.604	0.595
	VickreyMhl	0.002	0.086	0.326	0.681	0.714	0.709	0.709	0.709	0.003	0.068	0.280	0.558	0.601	0.595	0.595	0.595
Sampling	CusText	0.707	0.709	0.709	0.709	0.709	0.709	0.709	0.709	0.592	0.595	0.595	0.595	0.595	0.595	0.595	0.595
	SanText	0.000	0.709	0.716	0.709	0.709	0.709	0.709	0.709	0.000	0.595	0.603	0.595	0.595	0.595	0.595	0.595
	TEM	0.005	0.772	0.772	0.772	0.772	0.772	0.772	0.772	0.008	0.674	0.674	0.674	0.674	0.674	0.674	0.674
	WBB	0.614	0.607	0.626	0.628	0.628	0.637	0.630	0.621	0.557	0.552	0.542	0.572	0.572	0.560	0.570	0.557

These experiments demonstrate that ranking performance deteriorates significantly under stringent privacy constraints (i.e., low ε). Moreover, utility improves as privacy constraints decrease (i.e., high ε), with most mechanisms achieving utility levels comparable to non-private settings. Furthermore, another insight is related to the obfuscation strategy: sampling-based mechanisms achieve higher performance at lower ε, while noisy embedding methods require higher ε values to reach saturation.

4.2. Privacy Analysis

Table 2 compares two different aspects of privacy. The average failure rate N_x of the mechanism $\mathcal{M}$ assesses the probability that a term x is mapped to itself over T obfuscations, with higher values indicating weaker privacy. Conversely, the QuIPU score measures how well the mechanism resists a query inference attack [15] from different attackers. The higher the score, the better the resistance.

[1] https://ir-datasets.com/

[2] https://ir-datasets.com/aol-ia.html

Table 2

Failure rate N_x of an obfuscation mechanism $\mathcal{M}$ when obfuscating $n = 400$ terms x for $T = 100$ times at different ε budgets. The WBB mechanism obtains a null probability of failure by design since the term x is *a priori* removed from the set of possible candidate obfuscations. The words x are randomly sampled from the GloVe 300-d Vocabulary. On the right part of the Table is reported the QuIPU Score [15] evaluates actual privacy considering the different obfuscation parametrization when subject to different attacker models, cf. Section 3.2.

		$N_x = \mathbb{P}[\mathcal{M}(x) = x]$								**QuIPU Score**								
		ε - Privacy Budget								*Lazy Attacker*			*Active Attacker*			*Motivated Attacker*		
Obfuscation	**Mechanism**	1	5	10	15	20	25	30	50	DL‘19	DL‘20	Robust‘04	DL‘19	DL‘20	Robust‘04	DL‘19	DL‘20	Robust‘04
Embeddings	CMP	<0.01	<0.01	0.15	0.65	0.94	0.99	1.00	1.00	0.299	0.372	0.175	0.257	0.323	0.001	0.283	0.353	0.170
	Mahalanobis	<0.01	<0.01	0.06	0.39	0.79	0.96	0.99	1.00	0.280	0.381	0.202	0.258	0.363	0.090	0.272	0.371	0.200
	VickreyCMP	<0.01	<0.01	0.07	0.21	0.26	0.27	0.28	0.31	0.341	0.430	0.194	0.310	0.411	0.103	0.334	0.424	0.193
	VickreyMhl	<0.01	<0.01	0.03	0.15	0.24	0.26	0.27	0.31	0.342	0.426	0.199	0.318	0.410	0.119	0.335	0.421	0.199
Sampling	CusText	0.15	0.56	0.94	0.99	1.00	1.00	1.00	1.00	0.041	0.109	0.034	-0.014	0.010	-0.084	0.020	0.074	0.028
	SanText	<0.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	-0.247	-0.222	0.046	-0.277	-0.252	-0.237	-0.255	-0.231	0.043
	TEM	<0.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	-0.264	-0.274	0.028	-0.264	-0.274	-0.329	-0.264	-0.274	0.027
	WBB	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.005	-0.002	0.001	-0.001	-0.022	-0.011	-0.006	-0.010	0.001

Noisy embedding-based methods such as CMP and Mhl show a gradual loss of privacy as ε increases, while VickreyCMP and VickreyMhl maintain lower N_x values, indicating more robust privacy guarantees. Sampling-based methods exhibit a different behaviour: CusText, SanText, and TEM rapidly lose their obfuscation capability, reaching $N_x = 1.00$ for relatively low ε. WBB, in contrast, preserves complete privacy with $N_x = 0$ across all budgets by design: the original word is always changed.

The QuIPU scores demonstrate the robustness of these mechanisms against different modelizations of the attackers [15]. Vickrey-based embedding methods offer better resistance among all the mechanisms studied, while the sampling-based methods, particularly SanText and TEM, do not perform well in all the adversarial settings. WBB provides a null QuIPU score, which means an equal performance-utility trade-off. Future research is needed to improve the robustness against the query inference attack.

In conclusion, WBB and Vickrey-based embeddings are more suitable for scenarios requiring stringent privacy guarantees. In contrast, CMP and Mhl obfuscations provide a more balanced trade-off between privacy and utility. Finally, sampling-based approaches indicate lower effectiveness in adversarial environments, considering their probability of failure and resilience against inference attacks.

5. Conclusion

The paper presented the privacy problems faced when providing privacy to textual data in the author's first studies during his initial works during the PhD studies. The paper shows the methodology adopted to evaluate the privacy provided to the texts analysed and the strategies adopted to assess the privacy guarantees obtained. Possible Open Research Discussions (RD1-3) that will be proposed during the Doctoral Consortium session are formulated as follows:

RD1. Current privacy-preserving obfuscation techniques often operate independently of the underlying IR models. How can obfuscation methods be optimized to leverage the characteristics of specific retrieval models while maintaining formal privacy guarantees?

RD2. The trade-off between privacy and utility in obfuscated queries remains a critical challenge for the WBB mechanism. Can we design adaptive obfuscation mechanisms to dynamically balance privacy and retrieval effectiveness based on user needs and system constraints?

RD3. The effectiveness of privacy-preserving obfuscation methods can vary depending on the structure and semantics of different query types. How can obfuscation techniques be adapted to different query characteristics while ensuring consistent privacy guarantees? Can we adapt the obfuscation to domain-specific sensitive context like health scenarios?

Declaration on Generative AI

During the preparation of this work, the author(s) used Grammarly for Readability and Spelling checks. After using this tool, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] A. L. Martínez, M. G. Pérez, A. Ruiz-Martínez, A comprehensive review of the state-of-the-art on security and privacy issues in healthcare, ACM Comput. Surv. 55 (2023) 249:1–249:38. URL: https://doi.org/10.1145/3571156. doi:10.1145/3571156.

[2] R. Sawhney, A. T. Neerkaje, I. Habernal, L. Flek, How much user context do we need? privacy by design in mental health NLP applications, in: Y. Lin, M. Cha, D. Quercia (Eds.), Proceedings of the Seventeenth International AAAI Conference on Web and Social Media, ICWSM 2023, Limassol, Cyprus, June 5-8, 2023, AAAI Press, 2023, pp. 766–776. URL: https://doi.org/10.1609/icwsm.v17i1.22186. doi:10.1609/ICWSM.V17I1.22186.

[3] O. Akanfe, R. Valecha, H. R. Rao, Design of an inclusive financial privacy index (INF-PIE): A financial privacy and digital financial inclusion perspective, ACM Trans. Manag. Inf. Syst. 12 (2021) 7:1–7:21. URL: https://doi.org/10.1145/3403949. doi:10.1145/3403949.

[4] J. Cohn, My tivo thinks i'm gay: Algorithmic culture and its discontents, Television & New Media 17 (2016) 675–690. URL: https://doi.org/10.1177/1527476416644978. doi:10.1177/1527476416644978. arXiv:https://doi.org/10.1177/1527476416644978.

[5] W. U. Ahmad, M. M. Rahman, H. Wang, Topic model based privacy protection in personalized web search, in: R. Perego, F. Sebastiani, J. A. Aslam, I. Ruthven, J. Zobel (Eds.), Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, SIGIR 2016, Pisa, Italy, July 17-21, 2016, ACM, 2016, pp. 1025–1028. URL: https://doi.org/10.1145/2911451.2914753. doi:10.1145/2911451.2914753.

[6] B. Poblete, M. Spiliopoulou, R. Baeza-Yates, Privacy-preserving query log mining for business confidentiality protection, ACM Trans. Web 4 (2010) 10:1–10:26. URL: https://doi.org/10.1145/1806916.1806919. doi:10.1145/1806916.1806919.

[7] A. Tigunova, Extracting personal information from conversations, in: A. E. F. Seghrouchni, G. Sukthankar, T. Liu, M. van Steen (Eds.), Companion of The 2020 Web Conference 2020, Taipei, Taiwan, April 20-24, 2020, ACM / IW3C2, 2020, pp. 284–288. URL: https://doi.org/10.1145/3366424.3382089. doi:10.1145/3366424.3382089.

[8] O. Klymenko, S. Meisenbacher, F. Matthes, Differential privacy in natural language processing: The story so far, in: O. Feyisetan, S. Ghanavati, P. Thaine, I. Habernal, F. Mireshghallah (Eds.), Proceedings of the Fourth Workshop on Privacy in Natural Language Processing, Association for Computational Linguistics, Seattle, United States, 2022, pp. 1–11. URL: https://aclanthology.org/2022.privatenlp-1.1/. doi:10.18653/v1/2022.privatenlp-1.1.

[9] G. Faggioli, N. Ferro, Query obfuscation for information retrieval through differential privacy, in: N. Goharian, N. Tonellotto, Y. He, A. Lipani, G. McDonald, C. Macdonald, I. Ounis (Eds.), Advances in Information Retrieval - 46th European Conference on Information Retrieval, ECIR 2024, Glasgow, UK, March 24-28, 2024, Proceedings, Part I, volume 14608 of *Lecture Notes in Computer Science*, Springer, 2024, pp. 278–294. URL: https://doi.org/10.1007/978-3-031-56027-9_17. doi:10.1007/978-3-031-56027-9_17.

[10] L. Hu, I. Habernal, L. Shen, D. Wang, Differentially private natural language models: Recent advances and future directions, in: Y. Graham, M. Purver (Eds.), Findings of the Association for Computational Linguistics: EACL 2024, St. Julian's, Malta, March 17-22, 2024, Association for Computational Linguistics, 2024, pp. 478–499. URL: https://aclanthology.org/2024.findings-eacl.33.

[11] C. Dwork, F. McSherry, K. Nissim, A. Smith, Calibrating noise to sensitivity in private data

analysis, in: S. Halevi, T. Rabin (Eds.), Theory of Cryptography, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006, pp. 265–284.

[12] A. Blanco-Justicia, D. Sánchez, J. Domingo-Ferrer, K. Muralidhar, A critical review on the use (and misuse) of differential privacy in machine learning, ACM Comput. Surv. 55 (2023) 160:1–160:16. URL: https://doi.org/10.1145/3547139. doi:10.1145/3547139.

[13] F. L. De Faveri, G. Faggioli, N. Ferro, Words blending boxes. obfuscating queries in information retrieval using differential privacy, CoRR abs/2405.09306 (2024). URL: https://doi.org/10.48550/arXiv.2405.09306. doi:10.48550/ARXIV.2405.09306. arXiv:2405.09306.

[14] F. L. De Faveri, G. Faggioli, N. Ferro, pyPANTERA: A python package for natural language obfuscation enforcing privacy & anonymization, in: E. Serra, F. Spezzano (Eds.), Proceedings of the 33rd ACM International Conference on Information and Knowledge Management, CIKM 2024, Boise, ID, USA, October 21-25, 2024, ACM, 2024, pp. 5348–5353. URL: https://doi.org/10.1145/3627673.3679173. doi:10.1145/3627673.3679173.

[15] F. L. De Faveri, G. Faggioli, N. Ferro, Measuring actual privacy of obfuscated queries in information retrieval, in: Proceedings of the 47th European Conference on Information Retrieval, Lucca, Italy, 2025. URL: https://www.dei.unipd.it/~defaverifr/papers/25_ECIR_DFF_MeasuringActualPrivacyQuIPU_CameraReady.pdf.

[16] F. L. De Faveri, G. Faggioli, N. Ferro, A comparative study of large language models and traditional privacy measures to evaluate query obfuscation approaches, in: Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '25, Association for Computing Machinery, New York, NY, USA, 2025, p. 2711–2716. URL: https://doi.org/10.1145/3726302.3730158. doi:10.1145/3726302.3730158.

[17] K. Chatzikokolakis, M. E. Andrés, N. E. Bordenabe, C. Palamidessi, Broadening the scope of differential privacy using metrics, in: E. D. Cristofaro, M. K. Wright (Eds.), Privacy Enhancing Technologies - 13th International Symposium, PETS 2013, Bloomington, IN, USA, July 10-12, 2013. Proceedings, volume 7981 of *Lecture Notes in Computer Science*, Springer, 2013, pp. 82–102. URL: https://doi.org/10.1007/978-3-642-39077-7_5. doi:10.1007/978-3-642-39077-7_5.

[18] O. Feyisetan, B. Balle, T. Drake, T. Diethe, Privacy- and utility-preserving textual analysis via calibrated multivariate perturbations, in: J. Caverlee, X. B. Hu, M. Lalmas, W. Wang (Eds.), Proceedings of the 13th International Conference on Web Search and Data Mining, ACM, 2020, pp. 178–186. doi:10.1145/3336191.3371856.

[19] Z. Xu, A. Aggarwal, O. Feyisetan, N. Teissier, A differentially private text perturbation method using regularized mahalanobis metric, in: Proceedings of the Second Workshop on Privacy in NLP, Association for Computational Linguistics, 2020. doi:10.18653/v1/2020.privatenlp-1.2.

[20] Z. Xu, A. Aggarwal, O. Feyisetan, N. Teissier, On a utilitarian approach to privacy preserving text generation, CoRR abs/2104.11838 (2021). doi:10.48550/ARXIV.2104.11838. arXiv:2104.11838.

[21] S. Chen, F. Mo, Y. Wang, C. Chen, J.-Y. Nie, C. Wang, J. Cui, A customized text sanitization mechanism with differential privacy, in: A. Rogers, J. Boyd-Graber, N. Okazaki (Eds.), Findings of the Association for Computational Linguistics: ACL 2023, Association for Computational Linguistics, Toronto, Canada, 2023, pp. 5747–5758. URL: https://aclanthology.org/2023.findings-acl.355. doi:10.18653/v1/2023.findings-acl.355.

[22] X. Yue, M. Du, T. Wang, Y. Li, H. Sun, S. S. M. Chow, Differential privacy for text analytics via natural text sanitization, in: C. Zong, F. Xia, W. Li, R. Navigli (Eds.), Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, Association for Computational Linguistics, Online, 2021, pp. 3853–3866. URL: https://aclanthology.org/2021.findings-acl.337. doi:10.18653/v1/2021.findings-acl.337.

[23] R. S. Carvalho, T. Vasiloudis, O. Feyisetan, K. Wang, TEM: high utility metric differential privacy on text, in: S. Shekhar, Z. Zhou, Y. Chiang, G. Stiglic (Eds.), Proceedings of the 2023 SIAM International Conference on Data Mining, SDM 2023, Minneapolis-St. Paul Twin Cities, MN, USA, April 27-29, 2023, SIAM, 2023, pp. 883–890. URL: https://doi.org/10.1137/1.9781611977653.ch99. doi:10.1137/1.9781611977653.CH99.

[24] I. Wagner, D. Eckhoff, Technical privacy metrics: A systematic survey, ACM Comput. Surv. 51 (2018) 57:1–57:38. URL: https://doi.org/10.1145/3168389. doi:10.1145/3168389.

[25] S. Sousa, R. Kern, How to keep text private? A systematic review of deep learning methods for privacy-preserving natural language processing, Artif. Intell. Rev. 56 (2023) 1427–1492. URL: https://doi.org/10.1007/s10462-022-10204-6. doi:10.1007/S10462-022-10204-6.

[26] I. Habernal, When differential privacy meets NLP: the devil is in the detail, in: M. Moens, X. Huang, L. Specia, S. W. Yih (Eds.), Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021, Association for Computational Linguistics, 2021, pp. 1522–1528. URL: https://doi.org/10.18653/v1/2021.emnlp-main.114. doi:10.18653/V1/2021.EMNLP-MAIN.114.

[27] S. Clauß, S. Schiffner, Structuring anonymity metrics, in: A. Juels, M. Winslett, A. Goto (Eds.), Proceedings of the 2006 Workshop on Digital Identity Management, Alexandria, VA, USA, November 3, 2006, ACM, 2006, pp. 55–62. URL: https://doi.org/10.1145/1179529.1179539. doi:10.1145/1179529.1179539.

[28] S. J. Meisenbacher, N. Nandakumar, A. Klymenko, F. Matthes, A comparative analysis of word-level metric differential privacy: Benchmarking the privacy-utility trade-off, in: N. Calzolari, M. Kan, V. Hoste, A. Lenci, S. Sakti, N. Xue (Eds.), Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation, LREC/COLING 2024, 20-25 May, 2024, Torino, Italy, ELRA and ICCL, 2024, pp. 174–185. URL: https://aclanthology.org/2024.lrec-main.16.

[29] H. Seo, H. Lee, W. Choi, Fundamental limits of private information retrieval with unknown cache prefetching, IEEE Transactions on Communications 69 (2021) 8132–8144. doi:10.1109/TCOMM.2021.3117936.

[30] G. Persiano, K. Yeo, Limits of preprocessing for single-server PIR, in: J. S. Naor, N. Buchbinder (Eds.), Proceedings of the 2022 ACM-SIAM Symposium on Discrete Algorithms, SODA 2022, Virtual Conference / Alexandria, VA, USA, January 9 - 12, 2022, SIAM, 2022, pp. 2522–2548. URL: https://doi.org/10.1137/1.9781611977073.99. doi:10.1137/1.9781611977073.99.

[31] F. McSherry, K. Talwar, Mechanism design via differential privacy, in: 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2007), October 20-23, 2007, Providence, RI, USA, Proceedings, IEEE Computer Society, 2007, pp. 94–103. URL: https://doi.org/10.1109/FOCS.2007.41. doi:10.1109/FOCS.2007.41.

[32] N. Craswell, B. Mitra, E. Yilmaz, D. Campos, E. M. Voorhees, Overview of the TREC 2019 deep learning track, CoRR abs/2003.07820 (2020). URL: https://arxiv.org/abs/2003.07820. arXiv:2003.07820.

[33] N. Craswell, B. Mitra, E. Yilmaz, D. Campos, Overview of the TREC 2020 deep learning track, in: E. M. Voorhees, A. Ellis (Eds.), Proceedings of the Twenty-Ninth Text REtrieval Conference, TREC 2020, Virtual Event [Gaithersburg, Maryland, USA], November 16-20, 2020, volume 1266 of *NIST Special Publication*, National Institute of Standards and Technology (NIST), 2020. URL: https://trec.nist.gov/pubs/trec29/papers/OVERVIEW.DL.pdf.

[34] E. M. Voorhees, Overview of the TREC 2004 robust track, in: E. M. Voorhees, L. P. Buckland (Eds.), Proceedings of the Thirteenth Text REtrieval Conference, TREC 2004, Gaithersburg, Maryland, USA, November 16-19, 2004, volume 500-261 of *NIST Special Publication*, National Institute of Standards and Technology (NIST), 2004. URL: http://trec.nist.gov/pubs/trec13/papers/ROBUST.OVERVIEW.pdf.

[35] J. Pennington, R. Socher, C. D. Manning, Glove: Global Vectors for Word Representation, in: A. Moschitti, B. Pang, W. Daelemans (Eds.), Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL, ACL, 2014, pp. 1532–1543. URL: https://doi.org/10.3115/v1/d14-1162. doi:10.3115/v1/d14-1162.

[36] G. Izacard, M. Caron, L. Hosseini, S. Riedel, P. Bojanowski, A. Joulin, E. Grave, Unsupervised dense information retrieval with contrastive learning, Trans. Mach. Learn. Res. 2022 (2022). URL: https://openreview.net/forum?id=jKN1pXi7b0.