Query Obfuscation for Information Retrieval through Differential Privacy

Guglielmo Faggioli[1][0000−0002−5070−2049] and Nicola Ferro[1][0000−0001−9219−6239]

University of Padua, Padua, Italy

Abstract. Protecting the privacy of a user querying an Information Retrieval (IR) system is of utmost importance. The problem is exacerbated when the IR system is not cooperative in satisfying the user's privacy requirements. To address this, obfuscation techniques split the user's sensitive query into multiple non-sensitive ones that can be safely transmitted to the IR system. To generate such queries, current approaches rely on lexical databases, such as WordNet, or heuristics of word co-occurrences. At the same time, advances in Natural Language Processing (NLP) have shown the power of Differential Privacy (DP) in releasing privacy-preserving text for completely different purposes, such as spam detection and sentiment analysis. We investigate for the first time whether DP mechanisms, originally designed for specific NLP tasks, can effectively be used in IR to obfuscate queries. We also assess their performance compared to state-of-the-art techniques in IR. Our empirical evaluation shows that the Vickrey DP mechanism based on the Mahalanobis norm with privacy budget $\epsilon \in [10, 12.5]$ achieves state-of-the-art privacy protection and improved effectiveness. Furthermore, differently from previous approaches that are substantially on/off, by changing the privacy budget ϵ, DP allows users to adjust their desired level of privacy protection, offering a trade-off between effectiveness and privacy.

1 Introduction

Information Retrieval (IR) systems are a commodity used for many tasks, including searching for personal information, such as symptoms and diseases [7], political opinions, or egosurfing, i.e., searching the own name or social profile, among others. Such searches can be used to profile the user and can put at risk their privacy [6]. For example, an insurance company might try to access the user's queries to determine if they have any disease, or a malicious employee of a search engine might access the query log to blackmail them. To alleviate this, proxy obfuscation approaches hide the sensitive information need, by breaking it down into multiple non-sensitive queries that are less exposing and can be safely transmitted to the IR system. To this end, some approaches rely on replacing words with generalizations, i.e., hypernyms [2, 4]. Other strategies use a local corpus to determine which words, by co-occurring in the documents with those in the query, induce the same ranked list [3, 5, 19]. We investigate for the first time whether Differential Privacy (DP) mechanisms, originally designed for specific Natural Language Processing (NLP) tasks, can effectively be

used in IR to obfuscate queries. DP [15] is a state-of-the-art framework meant to release privately sensitive information. The general idea is to use a randomized mechanism that introduces noise into the computation. Thanks to this, the user can "plausibly deny" the output: it is impossible to prove that the output corresponds to the input of the user and is not due to the randomness of the mechanism. DP is particularly effective in the NLP domain. A line of research [18, 37, 45, 46] operationalizes DP to release text by obfuscating each word individually. Such mechanisms work as follows: i) each word in the text is mapped to a non-contextual embedding space; ii) the embeddings are perturbed with noise drawn from a specific distribution; iii) each word is replaced with the word closest to the noisy embedding. A major advantage of DP is that it allows setting the privacy budget based on the needs of the user. This is different from current obfuscation mechanisms in IR, which are either active or not and cannot be tuned based on the user's needs.

In this work, we focus on three of such mechanisms: the Calibrated Multivariate Perturbation (CMP) [18], the Mahalanobis [45] and the Vickrey [46]. These approaches were originally devised and tested for NLP tasks that include text classification and sentiment analysis. When it comes to the NLP scenario, the model can be trained directly on the obfuscated documents and this allows the model itself to learn how to account for the noise within the documents. However, this is not the case in IR: we assume the IR system to not preserve user privacy, and to possibly be malicious. In our use case, users are the ones concerned about their privacy. They don't want to reveal their real information needs and prefer to transmit obfuscated queries to the IR system while still retrieving relevant documents. Thus, in our use case, the IR system cannot be trained on obfuscated queries or documents, nor should be aware that an obfuscation mechanism has been used. Therefore, to operationalize our mechanism, we assume each user to locally obfuscate their query and transmit the obfuscated query, or possibly multiple queries, to the IR system instead of their real query.

Our goal is to determine if the DP mechanisms introduced above can successfully obfuscate users' information needs while still retrieving relevant documents. More in detail, our research questions are as follows:

RQ1 *Privacy Guarantees*: How much the original query leaks within queries obfuscated using DP obfuscation approaches, originally developed for NLP?

RQ2 *Relevant Documents Retrieved*: Is it feasible to exploit such DP mechanisms to retrieve relevant documents?

RQ3 *Comparing DP and non-DP Approaches*: What is the equivalent DP level for the scrambling approaches, which are the current state-of-the-art in IR?

The paper is organized as follows: Section 2 reports the main related works while Section 3 introduces the approaches to text obfuscation developed for IR and those developed for NLP. Section 4 delineates our experimental methodology and Section 5 details our findings. Finally, Section 6 draws our conclusion and outlines the future work.

2 Related Work

In this work, we assume the user to be interested in querying a search engine, without disclosing their real interest. Furthermore, the IR system is not cooperative and does not operate toward protecting the privacy of the user. Therefore, the query needs to be obfuscated on the user side and transmitted to the search engine so that the latter cannot understand the real user's interest. Three main approaches to obfuscate the query from the IR system exist: i) Approaches based on dummy queries; ii) Approaches based on unlinkability; iii) Approaches based on proxy queries. Approaches based on dummy queries [14, 16, 44, 47] transmit to the IR system, along the user's query, a set of unrelated queries, but syntactically similar to the user's query. Approaches based on unlinkability [8, 12, 13, 38] rely on cryptographic and Private Information Retrieval (PIR) primitives to allow a federation of users to exchange the query with each other so that each user submits the query of someone else and the system cannot profile the user. Finally, approaches based on proxy queries [2–5, 19] rely on breaking down the query into multiple non-sensitive queries, whose combined results might provide an answer to the user's query. For example, the query "throat cancer" could be transformed into "neck", and "tumour", reducing the information disclosed by each query. The disadvantage of dummy queries and unlinkability approaches is that the original query is sent to the IR system. Therefore, approaches based on dummy queries and unlinkability are vulnerable to machine learning attacks that aim to identify automatically generated queries: as shown by Khan et al. [24], Peddinti and Saxena [31, 32] this is relatively easy, with access to a small number of real user queries. Additionally approaches based on unlinkability move the problem from the IR system to a different user within the federation. This vulnerability does not occur with approaches based on proxy queries. The downside of approaches based on proxy queries is that there is a physiological decrease in effectiveness, which does not occur for dummy queries and unlinkability approaches. Nevertheless, we argue that a user wanting to achieve strong privacy guarantees should be able to do so, even though this might mean renouncing part of the effectiveness in favour of privacy. Therefore we focus on the obfuscation approaches that rely on proxy queries.

The usage of DP in IR involves some efforts to use it to release the local updates in the Federated Learning scenario [25, 41] and to release privacy-preserving query logs [20, 26, 36, 48]. We are not aware of any work employing it as the framework to obfuscate queries to be sent to the central server. Therefore, this can be considered the first work using DP for query obfuscation.

3 Approaches

We describe here two approaches designed for IR obfuscation query generation. We also introduce the DP framework. Finally, we present the DP mechanisms, designed for privately releasing text, considered in this work.

3.1 Native IR Approaches

We describe here the two major proxy query obfuscation approaches designed for IR tasks. Arampatzis et al. [2] propose a obfuscation method based on WordNet [29]. For each query term, Arampatzis et al. use WordNet to extract the set of synonyms, hypernyms, and holonyms. The approach considers sets of terms that are two steps away on the WordNet hierarchy. The candidate obfuscation queries are the cartesian product of the term sets. To filter out exposing queries, a similarity measure between each candidate query and the original query is considered. The similarity is the average wup similarity [43] between each term of the obfuscation query with all the terms of the original query. Arampatzis et al. empirically select obfuscated queries within the range (1., 0.7] of wup similarity.

Fröbe et al. [19] extended the earlier work of Arampatzis et al. [3, 5] and develop a statistical query obfuscation method. The approach consists of using a local corpus to select and filter candidate obfuscation queries. Using the user's query on the local corpus, Fröbe et al. consider all the possible combinations of n terms from a sliding window of terms within the top-k documents. To enhance privacy, all the candidates containing a query term or a synonym, hypernym, or hyponym, of it are dropped. To decide which candidate queries to submit, the top-k documents retrieved from the local corpus using the original query are considered pseudo-relevant. The candidate queries are ranked according to nDCG achieved in retrieving the pseudo-relevant documents from the local corpus.

3.2 Differential Privacy

Differential Privacy (DP) is considered a state-of-the-art approach for data release. Assume we have a private dataset D and we wish to compute and release some statistics $f(D)$, e.g., documents' scores in response to a query. A "randomized mechanism" $\mathcal{M}_f$ is a function that takes in input a dataset D and outputs the privacy-preserving result of $f(D)$, by introducing some noise. Two datasets D and D' are defined as "neighbouring" if differ by at most one record. A randomized mechanism $\mathcal{M}_f$ satisfies ϵ-DP iff, given a privacy budget $\epsilon \in \mathbb{R}^+$, for any pair of neighbouring datasets D and D':

$$\frac{Pr[\mathcal{M}_f(D) \in \mathcal{S}]}{Pr[\mathcal{M}_f(D') \in \mathcal{S}]} \leq \exp(\epsilon), \forall \mathcal{S} \subseteq \text{Image}(\mathcal{M}_f)$$

The smaller ϵ is, the larger the privacy guarantees, but also the larger the noise introduced in the data. For a DP mechanism, thanks to the introduced noise, the output of the mechanism on the two neighbouring datasets is likely similar.

Metric DP To achieve DP in a metric space, an obfuscation mechanism should have an equal probability of obfuscating any pair of points as the same point, irrespective of their distance. While this grants the highest level of privacy, it also requires high levels of noise, decreasing the utility of the data. In the case of metric spaces, it is often sufficient if the probability of obfuscating two points with the same one is proportional to the distance between the two points. Or,

alternatively, the proportion of sampling a certain noise is inversely proportional to the norm of the noise itself. To this end, a relaxation of DP, called Metric-DP, has been introduced. Metric-DP [1, 9, 27] is defined as follows: given a privacy budget ϵ and a distance measure $d : \mathbb{R}^p \times \mathbb{R}^p \rightarrow [0, \infty)$, a randomized mechanism $\mathcal{M} : \mathbb{R}^p \rightarrow \mathbb{R}^p$ defined over a geometric space is Metric-DP iff, for any three points in the space $w, w', \hat{w} \in \mathbb{R}^p$, the following holds:

$$\frac{Pr\{\mathcal{M}(w) = \hat{w}\}}{Pr\{\mathcal{M}(w') = \hat{w}\}} \leq \exp(\epsilon d\{w, w'\})$$

If the $d\{w, w'\}$ is small, w and w' are more likely to be obfuscated with the same point. Vice-versa, far apart points might be obfuscated with different points, without violating privacy constraints.

3.3 DP Mechanisms Designed for NLP Tasks

In this work, we evaluate if it is possible to adapt to the IR scenario three DP mechanisms that obfuscate text on a per-word basis, originally developed in the NLP context. More in detail, these approaches take as input a sequence of words. Each word is mapped into a non-contextual embedding, such as GloVe [33]. Then, the embedding is obfuscated by adding some appositely sampled noise to it. To ensure that Metric-DP is achieved, the noise vector z is expected to be sampled from a distribution f such that the probability of observing z is $f(z) \propto \exp(-\epsilon||z||)$, i.e., the probability of sampling a noise with norm $||z||$ is inversely proportional to $||z||$. Finally, the closest word to the noisy embedding is used to obfuscate the corresponding word in the original text. We propose to use these approaches in the IR scenario to perturb the queries instead of the documents, as done for NLP tasks. We chose to use approaches that work on a per-word basis, and not directly on the actual representation vector used by the IR system. Indeed, this would require the user to have access to the encoding procedure of the query (which can be computationally expensive). Moreover, knowing how such representation vectors are computed, would mean that the IR system cooperates to protect the user privacy, which is the opposite of our case. Methods based on encoding the words locally, i.e., Local DP, ensure that the system will not be aware of the privacy mechanism being in place, as it will receive just a query composed of terms, as usual. Finally, being transparent to the IR system, this approach is applicable to any IR system and portable, avoiding the need to develop ad-hoc solutions for a specific system.

Calibrated Multivariate Perturbation (CMP) Mechanism The obfuscation of a word according to the CMP mechanism, defined by Feyisetan et al. [18], is based on sampling a noise vector following an n-dimensional Laplace distribution. Such sampling works in two phases: i) an n-dimensional unitary vector $p \in \mathbb{R}^n$ is sampled uniformly. This vector represents the direction of the perturbation. ii) the radius of the perturbation $r \in \mathbb{R}^+$ is sampled from a Gamma distribution. To sample p, a vector $N \in \mathbb{R}^n$ is sampled from a multivariate normal distribution,

with location 0 and identity covariance matrix $\mathbf{I}_n$: $N \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$. Then $p = N/||N||_2$. The radius of the noise is sampled from a Gamma distribution with shape n and scale $\frac{1}{\epsilon}$ as $r \sim Gam(n, \frac{1}{\epsilon})$. It is possible to observe that, the larger the privacy requirement, i.e., the smaller the ϵ, the bigger the noise. The noise vector z is such that $z = p \cdot r$ and, as proven by Fernandes et al. [17], z corresponds to a random vector sampled from a multivariate Laplace distribution with scale $1/\epsilon$, thus $f(z) \propto \exp(-\epsilon||z||_2)$. To perturb a word w, the noise vector z, is added to the original word embedding $\phi(w) \in \mathbb{R}^n$, and the word closest to the noisy word embedding is used as obfuscation. Feyisetan et al. [18] demonstrate that for any word sequence $\mathcal{W}^l$ of length $l \geq 1$ and any $\epsilon > 0$, the mechanism CMP: $\mathcal{W}^l \rightarrow \mathcal{W}^l$ satisfies ϵd-privacy with respect to d, where d is the Euclidean distance. The CMP mechanism [18] to obfuscate a single word is presented in Algorithm 1.

Algorithm 1: The CMP mechanism [18]

Data: a string w, privacy parameter ϵ
$v = \phi(w)$;
sample z such that $f(z) \propto \exp(-\epsilon||z||_2)$ as described above;
$\hat{v} = v + z$;
$w = \text{argmin}\, ||\phi(u) - \hat{v}||$;
return w;

Feyisetan et al. [18] originally employed this mechanism to release documents for NLP use cases. In the IR scenario, we employ it to perturb the query by applying Algorithm 1 to each query term.

Mahalanobis Mechanism Xu et al. [45] noticed how the perturbation induced by CMP mechanism tends to be weak, especially for high ϵ. In particular, they consider this to be caused by how the direction of the noise is chosen. Their hypothesis is that sampling the direction of the perturbation on a circumference ($||p||_2 = 1$) increases the risk of sampling a point on an empty region. If this occurs, the embedding for the original word remains the closest to the noisy vector and the word is obfuscated with itself. Therefore, Xu et al. adapt the CMP mechanism to increase the likelihood that the sampled noise will be toward the direction where most of the embeddings are. Practically, this corresponds to transforming the direction of the noise from a circumference to an ellipsis whose orientation can be set to be towards the other embeddings. To do so, it is necessary to modify the sampling mechanism, so that, instead of sampling p such that $||p||_2 = 1$, p is sampled so that $||p||_M = 1$ where $||\cdot||_M$ is the Mahalanobis norm [28]. The Mahalanobis norm $||\cdot||_M$ is defined as follows: for a positive definite matrix Σ, the Mahalanobis norm of a vector $x \in \mathbb{R}^n$ is $||x||_M = \sqrt{x^T \Sigma^{-1} x}$. By properly setting the matrix Σ and normalizing x by its Mahalanobis norm, we can change the orientation and eccentricity of the ellipse around x. Xu et al. [45] define $\Sigma \in \mathbb{R}^{n \times n}$ as the covariance matrix of all the

word embeddings, divided by the mean of the variances of each embedding so that the trace of Σ is equal to n. Using Σ defined as above ensures that the noise is stretched toward the direction that corresponds to the largest variability in the embedding space: where it is more likely to find other embeddings. Notice that, the procedure provides a single matrix Σ for all the word embeddings, regardless of the words we are trying to obfuscate. To ensure that the noise z is sampled such that its probability distribution is $f(z) \propto \exp(-e||z||_M)$ a vector N is sampled from the multivariate normal distribution $N \sim \mathcal{N}(\mathbf{0}, I_n)$. Then, p is such that $p = \Sigma^{1/2} \cdot (N/||N||_2)$. The sampling of the norm of the noise r remains the same as for CMP. The Mahalanobis Mechanism, which we refer to as Mhl, obfuscates every single word as in Algorithm 1, with the only difference that, in step 2, z is sampled so that $f(z) \propto \exp(-e||z||_M)$. To obfuscate a query composed of multiple words, Mhl is applied independently to each word. Xu et al. [45] demonstrate that, for any $\epsilon > 0$ and for any sequence of words $\mathcal{W}^l$ of length l, Mhl satisfies ϵd-privacy with respect to the Mahalanobis distance.

Vickrey Mechanism The Mhl still tends to obfuscate a word with itself for large ϵ. To reduce the probability of masking a token with itself, Xu et al. [46] defines the Vickrey [39] DP mechanism (we refer to it as Vkr). The Vickrey mechanism draws upon the Vickrey auction, a type of auction in which the highest bidder wins but the price paid is the second-highest bid. Vkr is based on two steps: in the first step, a noisy vector is sampled using any of the mechanisms described above – Xu et al. [46] illustrate their approach using Mhl as the instantiating mechanism (we indicate it with Vkr_{Mhl}), but the results hold also for CMP (Vkr_{CMP}). In the second step, with probability Pr the word corresponding to the closest embedding to the noisy vector is used as the obfuscation word. Vice versa, with probability $1 - Pr$ the word corresponding to the second closest embedding is used as obfuscation. The probability Pr is defined as follows. We call $\phi(u_1)$ and $\phi(u_2)$ respectively the closest and second closest word embeddings to $\hat{v}$, the perturbed embedding of w, and t an additional free parameter. $Pr(t, \hat{v}) = \frac{(1-t)||\phi(u_2)-\hat{v}||_2}{t||\phi(u_1)-\hat{v}||_2+(1-t)||\phi(u_2)-\hat{v}||_2}$. If $t = 0$, then the Vkr mechanism falls back to the instantiating mechanism. Intermediate values of t allow selecting either the first or the second nearest neighbour depending on the size of t, but also on the distance of the second closest neighbour to the noisy vector. We set $t = 0.75$, being the most performing according to Xu et al. [46].

4 Evaluation Methodology

We would like to point out that, interpreting the results when it comes to privacy requires weighing the risk of information leakage and the effectiveness. Full privacy protection is achieved only by transmitting white noise to the IR system while to preserve entirely the effectiveness it is necessary to destroy the privacy. A privacy-preserving approach is as good as it is capable of reducing privacy leakage, while still obtaining satisfactory effectiveness, but also according to how easy it is for the user to tune such tradeoff, based on their needs.

Answering RQ1 To investigate how effective DP approaches are in protecting the user information need, we compute the average similarity of the obfuscation queries with respect to the original query. High similarity indicates that the approach does not protect the user's privacy, as the information need can be inferred also by looking at the obfuscation queries. We propose to adopt two similarity measures: the Jaccard similarity between the terms of the queries and a neural sentence similarity approach relying on MiniLM [42]. The former similarity allows us to verify to what extent the terms of the two queries overlap, while the latter allows us to verify that query terms have not simply been replaced with synonyms. This second strategy consists of encoding each obfuscation query and the original query using MiniLM and obtaining an embedding for each query. then, the average cosine similarity is computed between the embeddings for each obfuscation query and the one for the original query.

Answering RQ2 We investigate whether DP approaches produce obfuscation queries that retrieve relevant documents. Using each DP approach, we produce a set of obfuscation queries and use it to retrieve the documents, we then evaluate the number of relevant documents for the original user's query retrieved by at least one obfuscation query. If this holds, in a real-life scenario, the user interested in retrieving documents while protecting their privacy can issue the obfuscation queries in place of the real one, obtain the results and rerank or reindex them locally to improve the precision, as proposed by Arampatzis et al. [2].

Answering RQ3 We consider obfuscation approaches originally devised explicitly for the IR task and measure to what level of ϵ they can be considered equivalent. We take into consideration the seminal work by Arampatzis et al. [2], labelled AED, and the recent state-of-the-art solution by Fröbe et al. [19], labelled FSH. We will compare these approaches with the DP mechanisms, based on three axes: i) the *obfuscation measure*, which we define as 1 minus the sentence similarity computed using the MiniLM representations; ii) the pooled recall; iii) the nDCG@10 observed if we re-rank the documents pooled by the obfuscation queries. Upon receiving 100 documents for each obfuscation query, we rerank them using TAS-B and evaluate the quality of this ranked list. This goes beyond the current state-of-the-art [2, 19], which does not evaluate the final rank that the user observes. For each approach, these measures are reported on a radar plot where, as a rule of thumb, a larger area corresponds to more desirable results.

5 Experimental Results

5.1 Experimental setup

We considered two different collections TREC Robust '04 [40] and TREC Deep Learning (DL '19) [10]. The former relies on a corpus of documents: disks 4 and 5 of the TIPSTER collection, minus the congressional records. The latter is based on the MS MARCO [30] passages corpus. As word embeddings, we used GloVe [33] with 300 dimensions trained on the Common Crawl. We also

Table 1: Average Jaccad similarity and MiniLM-based sentence similarity between the original query and 20 obfuscation queries generated with different approaches.

	Robust '04									DL '19								
ϵ	1	5	10	12.5	15	17.5	20	50	No DP	1	5	10	12.5	15	17.5	20	50	No DP
Jaccard Similarity																		
CMP	0.000	0.006	0.225	0.512	0.772	0.915	0.965	0.988		0.000	0.002	0.109	0.299	0.537	0.731	0.855	0.976	
Mhl	0.000	0.005	0.101	0.259	0.470	0.679	0.841	0.988		0.000	0.004	0.051	0.145	0.291	0.475	0.648	0.975	
Vkr_{CMP}	0.000	0.005	0.096	0.159	0.188	0.196	0.195	0.239		0.000	0.002	0.054	0.099	0.139	0.171	0.167	0.200	
Vkr_{Mhl}	0.000	0.005	0.049	0.096	0.147	0.179	0.186	0.231		0.000	0.002	0.030	0.068	0.103	0.135	0.157	0.194	
AED									0.200									0.338
FSH									0.000									0.000
scrambling	MiniLM Sentence Similarity																	
CMP	0.074	0.100	0.396	0.672	0.871	0.961	0.987	0.996		0.024	0.032	0.214	0.458	0.681	0.824	0.903	0.952	
Mhl	0.077	0.095	0.244	0.427	0.627	0.794	0.907	0.996		0.020	0.034	0.119	0.241	0.427	0.610	0.750	0.951	
Vkr_{CMP}	0.077	0.100	0.278	0.412	0.511	0.578	0.622	0.760		0.028	0.032	0.137	0.211	0.308	0.372	0.413	0.565	
Vkr_{Mhl}	0.076	0.096	0.188	0.282	0.382	0.472	0.533	0.746		0.023	0.026	0.084	0.149	0.215	0.284	0.333	0.553	
AED									0.487									0.509
FSH									0.203									0.077

experiment with other sizes of vectors, obtaining substantially identical findings, not reported for space reasons. In terms of retrieval models, we consider two sparse bag-of-word models, BM25 [34] and Vector Space Model (TF-IDF) [35], and two dense bi-encoders, TAS-B [21] and Contriever [23]. The choice of using these retrieval models stems from the fact that BM25 is a widely adopted lexical method, mostly based on exact matching: by changing the terms in the query we might end up losing specific terms that allow us to retrieve relevant documents. Vice versa, both TAS-B and Contriever are dense IR models that project and compute the similarity of queries and documents in a dense space and thus do not rely explicitly on the exact matching of terms. Nevertheless, by obfuscating the query terms, we lose the semantics of the query, and this might impair the retrieval phase. By using these IR systems, we can observe how the obfuscation approaches interact with IR systems based on different rationales.

In AED, to avoid bias, i.e., too similar or different queries from the original one, we select for each query as obfuscation queries the 10 queries above median wup similarity [43] and the 10 queries below. For FSH we employ the sliding window candidate generator with a window size of 16 as it is the best performing as originally observed by Fröbe et al. [19]. We use the parametrization reported by Fröbe et al. [19], considering the first 10 documents retrieved by TF-IDF from the local corpus as target documents, obfuscation queries of at most three terms, and remove queries retrieving less than 100 documents. We use 50,000 documents randomly sampled from the MS MARCO collection and from the TIPSTER disks 4 and 5, as local corpora for the Robust '04 and DL '19 respectively. For each query, we generate 20 obfuscation queries. The code is publicly available at: https://github.com/guglielmof/24-ECIR-FF.

5.2 RQ1: Privacy Guarantees

Table 1 shows, as a proxy of the privacy achieved by the mechanisms, the similarity between the original query and the obfuscation queries generated to hide

it. As expected from a differential privacy mechanism, the higher the ϵ the higher the similarity between the queries – with $\epsilon = 50$ for both Robust '04 and DL '19, CMP and Mhl achieve a Jaccard similarity higher than 97.5%. This indicates that overall the generated queries are almost identical to the original ones and there is no substantial privacy protection. Depending on the collection, CMP and Mhl obtain Jaccard similarity which falls in the range 20%-30%, with ϵ in the range $[10, 12.5]$. This indicates that, on average, 1 in 3 to 5 words remain equal to the original query. Similar results can also be observed for AED. Interestingly, when it comes to Vkr-based mechanisms, they tend to be much safer, as they obtain, with $\epsilon = 50$ less than 0.24 of Jaccard similarity on the Robust '04 collection, and 0.20 on the DL '19, 40.8% less than AED. Notice that, according to Jaccard similarity, FSH achieves perfect privacy, i.e., zero similarity, thanks to the fact that the words of the query are removed from the vocabulary of words that can be used to generate obfuscation queries. However, the approach based on the Jaccard similarity fails in assessing privacy leakage when synonyms are used to obfuscate the words of the query. Therefore, we also measure the similarity between the obfuscation queries and original queries using a more semantic-oriented approach, as described in Section 4. All the approaches have a much higher semantic similarity than what was observed for the Jaccard: in most of the cases, words are replaced with synonyms or highly correlated words. As for the Jaccard similarity, FSH, which explicitly removes synonyms and hypernyms from the queries, is particularly safe and corresponds to a DP Vkr_{CMP} mechanism with $\epsilon \in [5, 10]$ or a Vkr_{Mhl} with $\epsilon \in [10, 12.5]$ for the Robust '04, and DP Vkr_{CMP} and a Vkr_{Mhl} mechanism with $\epsilon \in [5, 10]$ for the DL '19. As observed for the Jaccard similarity, the privacy achieved by AED can be achieved with ϵ in the range $[10; 12.5]$ by CMP and Mhl on both collections. ϵ values that grant a comparable level of privacy are much higher for Vkr-based mechanisms, especially Vkr_{Mhl}, on both collections. As a general observation, privacy is a trade-off between noise and performance. It is reasonable that privacy is almost negligible for high values of ϵ and maximized for low values of ϵ. Furthermore, it is also intuitive that there are ϵ levels for which the DP mechanisms perform better, at least in terms of privacy, than any other baseline. We argue that, besides granting lower similarity at specific ϵ levels, the major advantage of DP is that it allows to meet the privacy requirements of the user, who can specify the privacy they would like to obtain and adapt the obfuscation mechanism consequently.

5.3 RQ2: Relevant Documents Retrieved

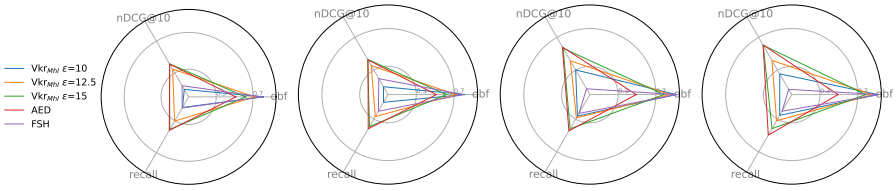
To assess the effectiveness of approaches based on DP, we measure the number of relevant documents retrieved, by pooling 100 documents from the 20 obfuscation queries representing the same information need. Not retrieving enough relevant documents would render the obfuscation approach unusable. As a reference point, we report the recall observed for the top 100 documents retrieved with the original query. Notice that, the number of used obfuscation queries is the same for DP-based approaches and for both AED and FSH. Using multiple obfuscation queries is generally a widely adopted procedure [19**?**]. Since the

Table 2: Mean recall achieved by pooling 100 documents retrieved for each obfuscation query.

		Robust '04									DL '19								
		ϵ									ϵ								
model	mechanism	1	5	10	12.5	15	17.5	20	50	No DP	1	5	10	12.5	15	17.5	20	50	No DP
BM25	CMP	0.020	0.146	0.483	**0.510**	0.489	0.442	0.421	0.407		0.011	0.135	0.384	**0.534**	0.517	0.514	0.480	0.444	
	Mhl	0.032	0.089	0.398	0.500	**0.512**	0.501	0.466	0.407		0.000	0.118	0.294	0.411	0.515	**0.529**	**0.529**	0.444	
	Vkr_{CMP}	0.041	0.152	0.407	0.506	0.548	0.554	**0.561**	0.518		0.000	0.073	0.282	0.363	0.498	**0.539**	0.498	0.533	
	Vkr_{Mhl}	0.021	0.133	0.304	0.409	0.493	0.544	**0.556**	0.530		0.016	0.039	0.263	0.286	0.419	0.479	**0.514**	0.506	
	AED									0.420									0.445
	FSH									0.140									0.231
	Original									0.410									0.454
TF-IDF	CMP	0.020	0.146	0.487	**0.512**	0.491	0.444	0.423	0.408		0.011	0.135	0.386	**0.535**	0.515	0.515	0.478	0.442	
	Mhl	0.032	0.089	0.398	0.504	**0.516**	0.504	0.468	0.408		0.000	0.118	0.295	0.412	0.515	**0.530**	0.527	0.442	
	Vkr_{CMP}	0.039	0.151	0.407	0.506	0.551	0.557	**0.563**	0.521		0.000	0.070	0.274	0.363	0.497	0.534	0.499	**0.533**	
	Vkr_{Mhl}	0.021	0.132	0.305	0.411	0.494	0.547	**0.559**	0.532		0.016	0.039	0.263	0.285	0.419	0.479	**0.512**	0.505	
	AED									0.420									0.443
	FSH									0.139									0.231
	Original									0.411									0.451
Contriever	CMP	0.034	0.106	0.469	**0.507**	0.481	0.433	0.406	0.392		0.000	0.077	0.446	0.615	**0.644**	0.628	0.576	0.512	
	Mhl	0.026	0.088	0.345	0.473	**0.503**	0.497	0.460	0.392		0.000	0.057	0.264	0.476	0.601	0.641	**0.650**	0.512	
	Vkr_{CMP}	0.038	0.125	0.397	0.486	0.510	0.518	**0.520**	0.475		0.000	0.010	0.280	0.430	0.537	0.598	0.579	**0.608**	
	Vkr_{Mhl}	0.024	0.094	0.269	0.392	0.462	0.504	**0.518**	0.481		0.000	0.027	0.254	0.321	0.418	0.522	0.580	**0.604**	
	AED									0.419									0.497
	FSH									0.204									0.204
	Original									0.392									0.528
TAS-B	CMP	0.027	0.080	0.434	**0.477**	0.444	0.398	0.369	0.356		0.000	0.063	0.392	0.615	**0.636**	0.622	0.575	0.498	
	Mhl	0.028	0.064	0.310	0.438	**0.464**	0.460	0.423	0.356		0.000	0.042	0.275	0.455	0.584	**0.645**	0.638	0.499	
	Vkr_{CMP}	0.025	0.086	0.355	0.448	0.483	0.490	**0.495**	0.446		0.000	0.025	0.245	0.398	0.534	0.585	0.579	**0.600**	
	Vkr_{Mhl}	0.023	0.073	0.237	0.355	0.423	0.476	**0.482**	0.452		0.000	0.019	0.206	0.267	0.375	0.503	0.553	**0.603**	
	AED									0.387									0.491
	FSH									0.161									0.238
	Original									0.358									0.518

queries are noisy, the adversarial is not able to recognize what was the topic of interest for the user. In turn, we expect each obfuscation query to return some relevant documents, most likely in low positions of the ranking, as it is not directly linkable to the original query. Once the results are available on a secure machine (e.g., the user's client) they can be reordered using the original query.

Following what was observed for the similarity, as observable in Table 2, the effectiveness varies widely over different mechanisms, with CMP and Mhl achieving higher recall for lower ϵ compared to Vkr-based mechanisms. Interestingly, the best-pooled recall is seldom achieved with $\epsilon = 50$ – exclusively using Vkr mechanisms and on the DL '19 collection. Moreover, for all the mechanisms, there is at least one value of ϵ for which the recall is higher than the one observed using the original queries. This is due to the fact that DP approaches with intermediate levels of ϵ automatically implement query rewriting. The usage of query variations has a strong impact on the performance of a system [11]: by automatically changing the words within a query with terms that are correlated but not identical, we can pool relevant documents that are lost when we use queries that are almost identical to the original ones, i.e., with $\epsilon = 50$ and CMP or Mhl mechanisms. By selecting either the closest or the second closest term, the Vkr-based mechanisms still apply implicit query rewriting also for higher levels of ϵ. Regardless of the collection considered, we notice that for ϵ ranging between 10 and 17.5, almost all mechanisms are able to overcome the IR state-of-the-art approaches in terms of recall. In particular, the DP approaches overcome FSH already with low ϵ, while $\epsilon \geq 15$ allows for overcoming AED as



(a) Robust '04, BM25 (b) Robust '04, Cnt. (c) DL '19, BM25 (d) DL '19, Cnt.

Fig. 1: Performance of different obfuscation mechanisms over three axes: pooled recall, nDCG@10 of the reranked documents, obfuscation (obf), measured as 1-similarity. Cnt. stands for "Contriever".

well. This suggests that using DP mechanisms that were originally thought to be used in NLP scenarios, and with additional training of the models, successfully allows us to obtain satisfactory performance in retrieval.

5.4 RQ3: Comparing DP and non-DP Approaches

As a final analysis, we compare the most promising DP approach, the Vickrey mechanism based on the Mahalanobis norm, with the current state-of-the-art approaches in IR. To avoid cluttering, we focus only on BM25 and Contriever as IR systems. Figure 1 reports the radar plots, showing the performance of different obfuscation approaches over the three axes mentioned above. We notice that the area corresponding to AED approach (in red) is encompassed within the area corresponding to Vkr$_{Mhl}$ with $\epsilon = 15$ (green). In fact, on the Robust '04 collection, AED achieves nDCG@10 of 0.410 and 0.424 for BM25 and Contriever respectively, recall of 0.420 and 0.419, and obfuscation of 0.513. Vice versa Vkr$_{Mhl}$ with $\epsilon = 15$ obtains nDCG@10 of 0.416 and 0.431, recall of 0.493 and 0.462, and obfuscation of 0.618. The exception is DL '19 with Contriever as the IR system, where AED has higher recall than Vkr$_{Mhl}$ (0.497 against 0.418). Nevertheless, this larger recall does not correspond to much larger nDCG@10, indicating that Vkr$_{Mhl}$ is preferable over AED, as it has comparable nDCG@10 (0.604 for Vkr$_{Mhl}$ against 0.607 for AED), with improved obfuscation (0.785 against 0.491). When it comes to FSH (purple), the behaviour depends on the collection. In the DL '19, using Vkr$_{Mhl}$ with $\epsilon = 10$ (blue) provides an edge over FSH: they have comparable obfuscation (0.916 the former, 0.923 the latter), but Vkr$_{Mhl}$ has much larger nDCG@10 (0.254 compared to 0.064). On the Robust '04 collection, to observe an improvement in terms of nDCG@10, it is necessary to use Vkr$_{Mhl}$ with $\epsilon = 12.5$ (nDCG@10 of 0.349 and 0.355 for BM25 and Contriever respectively) to overcome FSH in terms of nDCG@10 (0.140 and 0.194). Nevertheless, while Vkr$_{Mhl}$ with $\epsilon = 12.5$ exhibits nDCG@10 performance slightly lower than AED, it also has obfuscation (0.719) relatively close to FSH, which has obfuscation of 0.797, much closer than AED, with obfuscation 0.513. As a general guideline, our proposal is to use Vkr$_{Mhl}$ as the obfuscation mechanism, with ϵ chosen in the interval [10, 15], depending on the optimal trade-off between privacy and effectiveness, as chosen by the user.

An important aspect that should be discussed, is whether the user could be profiled or identified by looking at the documents retrieved. While the list of documents returned in response to an obfuscated query contains some relevant documents (see subsection 5.3), it is also true that, as the query is obfuscated, not all the documents retrieved will be strictly related to the topic of interest. Therefore, each obfuscated query will contain some "speck of gold", the relevant documents, which will be filtered on the user side, so that the adversarial cannot reconstruct the information need of the user by looking at the retrieved documents. Similarly, in more search engine-oriented scenarios, we could consider that the system might profile the user, based on which documents they click. We argue that, while present, this is an orthogonal problem to the task investigated here. In fact, there exist approaches, such as TrackMeNot [22], which simulate user clicks on random or non-relevant documents, to prevent the adversarial from using the clicks done by the user to profile them.

6 Conclusion and Future Work

In this work, we analyzed for the first time the performance of three DP mechanisms, originally designed for NLP, in the proxy query obfuscation IR task. These mechanisms are the Calibrated Multivariate Perturbation, the Mahalanobis, and the Vickrey mechanisms. We evaluated these mechanisms on the IR setting by considering three aspects: their obfuscation capabilities, their effectiveness in terms of recall, and their ability in allowing to retrieve highly relevant documents. To measure the obfuscation, we considered the dissimilarity between the original query and the obfuscation queries produced by different approaches. To measure their recall and effectiveness, we generated 20 obfuscation queries and used them to retrieve documents from Robust '04 and DL '19. Our findings highlight that the Vickrey mechanism with $\epsilon \in [10, 12.5]$ achieves higher privacy guarantees, with improved effectiveness, than current state-of-the-art approaches. Furthermore, lower or higher levels of ϵ allow for better satisfy the user, either in terms of privacy or accuracy, depending on their inclinations. As a future work, we plan to investigate how to perturb dense representations of the queries and combine them with generative language models to produce obfuscation queries with the same dense representation, but different terms.

Acknowledgments

This work has received support from CAMEO, PRIN 2022 n. 2022ZLL7MW.

Bibliography

[1] M. E. Andrés, N. Bordenabe, K. Chatzikokolakis, and C. Palamidessi. Geo-indistinguishability: differential privacy for location-based systems. In A. Sadeghi, V. D. Gligor, and M. Yung, editors, *2013 ACM SIGSAC Conference on Computer and Communications Security, CCS'13, Berlin, Germany, November 4-8, 2013*, pages 901–914. ACM, 2013. https://doi.org/10.1145/2508859.2516735.

[2] A. Arampatzis, P. S. Efraimidis, and G. Drosatos. Enhancing deniability against query-logs. In P. D. Clough, C. Foley, C. Gurrin, G. J. F. Jones, W. Kraaij, H. Lee, and V. Murdock, editors, *Advances in Information Retrieval - 33rd European Conference on IR Research, ECIR 2011, Dublin, Ireland, April 18-21, 2011. Proceedings*, volume 6611 of *Lecture Notes in Computer Science*, pages 117–128. Springer, 2011. https://doi.org/10.1007/978-3-642-20161-5_13. URL https://doi.org/10.1007/978-3-642-20161-5_13.

[3] A. Arampatzis, G. Drosatos, and P. Efraimidis. A versatile tool for privacy-enhanced web search. In P. Serdyukov, P. Braslavski, S. O. Kuznetsov, J. Kamps, S. M. Rüger, E. Agichtein, I. Segalovich, and E. Yilmaz, editors, *Advances in Information Retrieval - 35th European Conference on IR Research, ECIR 2013, Moscow, Russia, March 24-27, 2013. Proceedings*, volume 7814 of *Lecture Notes in Computer Science*, pages 368–379. Springer, 2013. https://doi.org/10.1007/978-3-642-36973-5_31. URL https://doi.org/10.1007/978-3-642-36973-5_31.

[4] A. Arampatzis, P. S. Efraimidis, and G. Drosatos. A query scrambler for search privacy on the internet. *Inf. Retr.*, 16(6): 657–679, 2013. https://doi.org/10.1007/s10791-012-9212-1. URL https://doi.org/10.1007/s10791-012-9212-1.

[5] A. Arampatzis, G. Drosatos, and P. S. Efraimidis. Versatile query scrambling for private web search. *Inf. Retr. J.*, 18(4): 331–358, 2015. https://doi.org/10.1007/s10791-015-9256-0. URL https://doi.org/10.1007/s10791-015-9256-0.

[6] M. Barbaro and T. Zeller. A Face Is Exposed For AoL Searcher No. 4417749. *New York Times*, 2006.

[7] S. Bavadekar, A. M. Dai, J. Davis, D. Desfontaines, I. Eckstein, K. Everett, A. Fabrikant, G. Flores, E. Gabrilovich, K. Gadepalli, S. Glass, R. Huang, C. Kamath, D. Kraft, A. Kumok, H. Marfatia, Y. Mayer, B. Miller, A. Pearce, I. M. Perera, V. Ramachandran, K. Raman, T. Roessler, I. Shafran, T. Shekel, C. Stanton, J. Stimes, M. Sun, G. Wellenius, and M. Zoghi. Google COVID-19 search trends symptoms dataset: Anonymization process description (version 1.0). *CoRR*, abs/2009.01265, 2020. URL https://arxiv.org/abs/2009.01265.

[8] J. Castellà-Roca, A. Viejo, and J. Herrera-Joancomartí. Preserving user's privacy in web search engines. *Comput. Commun.*, 32(13-14):

1541–1551, 2009. https://doi.org/10.1016/j.comcom.2009.05.009. URL https://doi.org/10.1016/j.comcom.2009.05.009.

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

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

[11] J. S. Culpepper, G. Faggioli, N. Ferro, and O. Kurland. Topic difficulty: Collection and query formulation effects. *ACM Trans. Inf. Syst.*, 40(1):19:1–19:36, 2022. https://doi.org/10.1145/3470563. URL https://doi.org/10.1145/3470563.

[12] J. Domingo-Ferrer and Ú. González-Nicolás. Rational behavior in peer-to-peer profile obfuscation for anonymous keyword search. *Inf. Sci.*, 185(1):191–204, 2012. https://doi.org/10.1016/j.ins.2011.09.010. URL https://doi.org/10.1016/j.ins.2011.09.010.

[13] J. Domingo-Ferrer, M. Bras-Amorós, Q. Wu, and J. A. Manjón. User-private information retrieval based on a peer-to-peer community. *Data Knowl. Eng.*, 68(11):1237–1252, 2009. https://doi.org/10.1016/j.datak.2009.06.004. URL https://doi.org/10.1016/j.datak.2009.06.004.

[14] J. Domingo-Ferrer, A. Solanas, and J. Castellà-Roca. h(k)-Private Information Retrieval from Privacy-Uncooperative Queryable Databases. *Online Inf. Rev.*, 33(4):720–744, 2009. https://doi.org/10.1108/14684520910985693. URL https://doi.org/10.1108/14684520910985693.

[15] C. Dwork and A. Roth. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9(3-4):211–407, 2014. https://doi.org/10.1561/0400000042.

[16] Y. Elovici, B. Shapira, and A. Maschiach. A New Privacy Model for Web Surfing. In A. Y. Halevy and A. Gal, editors, *Next Generation Information Technologies and Systems, 5th International Workshop, NGITS 2002, Caesarea, Israel, June 24-25, 2002, Proceedings*, volume 2382 of *Lecture Notes in Computer Science*, pages 45–57. Springer, 2002. https://doi.org/10.1007/3-540-45431-4_5. URL https://doi.org/10.1007/3-540-45431-4_5.

[17] N. Fernandes, M. Dras, and A. McIver. Generalised differential privacy for text document processing. In F. Nielson and D. Sands, editors, *Principles of Security and Trust - 8th International Conference, POST 2019, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2019, Prague, Czech Republic, April 6-11, 2019, Proceedings*, volume 11426 of *Lecture Notes in Computer Science*, pages 123–

148. Springer, 2019. https://doi.org/10.1007/978-3-030-17138-4_6. URL https://doi.org/10.1007/978-3-030-17138-4_6.

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

[19] M. Fröbe, E. O. Schmidt, and M. Hagen. Efficient query obfuscation with keyqueries. In J. He, R. Unland, E. S. Jr., X. Tao, H. Purohit, W. van den Heuvel, J. Yearwood, and J. Cao, editors, *WI-IAT '21: IEEE/WIC/ACM International Conference on Web Intelligence, Melbourne VIC Australia, December 14 - 17, 2021*, pages 154–161. ACM, 2021. https://doi.org/10.1145/3486622.3493950. URL https://doi.org/10.1145/3486622.3493950.

[20] M. Götz, A. Machanavajjhala, G. Wang, X. Xiao, and J. Gehrke. Publishing search logs - A comparative study of privacy guarantees. *IEEE Trans. Knowl. Data Eng.*, 24(3): 520–532, 2012. https://doi.org/10.1109/TKDE.2011.26. URL https://doi.org/10.1109/TKDE.2011.26.

[21] S. Hofstätter, S. Lin, J. Yang, J. Lin, and A. Hanbury. Efficiently teaching an effective dense retriever with balanced topic aware sampling. In F. Diaz, C. Shah, T. Suel, P. Castells, R. Jones, and T. Sakai, editors, *SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11-15, 2021*, pages 113–122. ACM, 2021. https://doi.org/10.1145/3404835.3462891. URL https://doi.org/10.1145/3404835.3462891.

[22] D. Howe and H. Nissenbaum. Trackmenot: Resisting surveillance in web search. Technical Report queries,, 2009.

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

[24] R. Khan, M. Ullah, A. Khan, M. I. Uddin, and M. Al-Yahya. NN-QuPiD Attack: Neural Network-Based Privacy Quantification Model for Private Information Retrieval Protocols. *Complex.*, 2021: 6651662:1–6651662:8, 2021. https://doi.org/10.1155/2021/6651662. URL https://doi.org/10.1155/2021/6651662.

[25] E. Kharitonov. Federated online learning to rank with evolution strategies. In J. S. Culpepper, A. Moffat, P. N. Bennett, and K. Lerman, editors, *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, WSDM 2019, Melbourne, VIC, Australia, February 11-15, 2019*, pages 249–257. ACM, 2019. https://doi.org/10.1145/3289600.3290968. URL https://doi.org/10.1145/3289600.3290968.

[26] A. Korolova, K. Kenthapadi, N. Mishra, and A. Ntoulas. Releasing search queries and clicks privately. In J. Quemada, G. León, Y. S. Maarek, and

W. Nejdl, editors, *Proceedings of the 18th International Conference on World Wide Web, WWW 2009, Madrid, Spain, April 20-24, 2009*, pages 171–180. ACM, 2009. https://doi.org/10.1145/1526709.1526733. URL https://doi.org/10.1145/1526709.1526733.

[27] P. Laud, A. Pankova, and M. Pettai. A framework of metrics for differential privacy from local sensitivity. *Proc. Priv. Enhancing Technol.*, 2020(2):175–208, 2020. https://doi.org/10.2478/popets-2020-0023.

[28] P. C. Mahalanobis. On the generalized distance in statistics. *Sankhyā: The Indian Journal of Statistics, Series A (2008-)*, 80:pp. S1–S7, 2018. ISSN 0976836X, 09768378. URL https://www.jstor.org/stable/48723335.

[29] G. A. Miller. Wordnet: A lexical database for english. *Commun. ACM*, 38(11):39–41, 1995. https://doi.org/10.1145/219717.219748. URL https://doi.org/10.1145/219717.219748.

[30] T. Nguyen, M. Rosenberg, X. Song, J. Gao, S. Tiwary, R. Majumder, and L. Deng. MS MARCO: A human generated machine reading comprehension dataset. In T. R. Besold, A. Bordes, A. S. d'Avila Garcez, and G. Wayne, editors, *Proceedings of the Workshop on Cognitive Computation: Integrating neural and symbolic approaches 2016 co-located with the 30th Annual Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain, December 9, 2016*, volume 1773 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2016. URL https://ceur-ws.org/Vol-1773/CoCoNIPS_2016_paper9.pdf.

[31] S. T. Peddinti and N. Saxena. On the Effectiveness of Anonymizing Networks for Web Search Privacy. In B. S. N. Cheung, L. C. K. Hui, R. S. Sandhu, and D. S. Wong, editors, *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security, ASIACCS 2011, Hong Kong, China, March 22-24, 2011*, pages 483–489. ACM, 2011. https://doi.org/10.1145/1966913.1966984. URL https://doi.org/10.1145/1966913.1966984.

[32] S. T. Peddinti and N. Saxena. Web search query privacy: Evaluating query obfuscation and anonymizing networks. *J. Comput. Secur.*, 22(1):155–199, 2014. https://doi.org/10.3233/JCS-130491. URL https://doi.org/10.3233/JCS-130491.

[33] J. Pennington, R. Socher, and C. D. Manning. Glove: Global vectors for word representation. In A. Moschitti, B. Pang, and W. Daelemans, editors, *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*, pages 1532–1543. ACL, 2014. https://doi.org/10.3115/v1/d14-1162. URL https://doi.org/10.3115/v1/d14-1162.

[34] S. E. Robertson, S. Walker, S. Jones, M. Hancock-Beaulieu, and M. Gatford. Okapi at TREC-3. In D. K. Harman, editor, *Proceedings of The Third Text REtrieval Conference, TREC 1994, Gaithersburg, Maryland, USA, November 2-4, 1994*, volume 500-225 of *NIST Special Publication*, pages 109–126. National Institute of Standards and Technology (NIST), 1994. URL http://trec.nist.gov/pubs/trec3/papers/city.ps.gz.

[35] G. Salton, A. Wong, and C. Yang. A vector space model for automatic indexing. *Commun. ACM*, 18(11):613–620, 1975. https://doi.org/10.1145/361219.361220. URL https://doi.org/10.1145/361219.361220.

[36] D. Sánchez, M. Batet, A. Viejo, M. Rodriguez-Garcia, and J. Castellà-Roca. A semantic-preserving differentially private method for releasing query logs. *Inf. Sci.*, 460-461:223–237, 2018. https://doi.org/10.1016/j.ins.2018.05.046. URL https://doi.org/10.1016/j.ins.2018.05.046.

[37] J. Tang, T. Zhu, P. Xiong, Y. Wang, and W. Ren. Privacy and utility trade-off for textual analysis via calibrated multivariate perturbations. In M. Kutylowski, J. Zhang, and C. Chen, editors, *Network and System Security - 14th International Conference, NSS 2020, Melbourne, VIC, Australia, November 25-27, 2020, Proceedings*, volume 12570 of *Lecture Notes in Computer Science*, pages 342–353. Springer, 2020. https://doi.org/10.1007/978-3-030-65745-1_20. URL https://doi.org/10.1007/978-3-030-65745-1_20.

[38] M. Ullah, M. A. Islam, R. Khan, M. Aleem, and M. A. Iqbal. Obsecure logging (oslo): A framework to protect and evaluate the web search privacy in health care domain. *J. Medical Imaging Health Informatics*, 9(6):1181–1190, 2019. https://doi.org/10.1166/jmihi.2019.2708. URL https://doi.org/10.1166/jmihi.2019.2708.

[39] W. Vickrey. Counterspeculation, auctions, and competitive sealed tenders, 1961.

[40] E. M. Voorhees. Overview of the TREC 2004 robust track. In E. M. Voorhees and L. P. Buckland, editors, *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.

[41] S. Wang, B. Liu, S. Zhuang, and G. Zuccon. Effective and privacy-preserving federated online learning to rank. In F. Hasibi, Y. Fang, and A. Aizawa, editors, *ICTIR '21: The 2021 ACM SIGIR International Conference on the Theory of Information Retrieval, Virtual Event, Canada, July 11, 2021*, pages 3–12. ACM, 2021. https://doi.org/10.1145/3471158.3472236. URL https://doi.org/10.1145/3471158.3472236.

[42] W. Wang, F. Wei, L. Dong, H. Bao, N. Yang, and M. Zhou. MiniLM: Deep Self-Attention Distillation for Task-Agnostic Compression of Pre-Trained Transformers. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020.

[43] Z. Wu and M. S. Palmer. Verb semantics and lexical selection. In J. Pustejovsky, editor, *32nd Annual Meeting of the Association for Computational Linguistics, 27-30 June 1994, New Mexico State University, Las Cruces, New Mexico, USA, Proceedings*, pages 133–138. Morgan Kaufmann Publishers / ACL, 1994. https://doi.org/10.3115/981732.981751. URL https://aclanthology.org/P94-1019/.

[44] Z. Wu, S. Shen, X. Lian, X. Su, and E. Chen. A Dummy-Based User Privacy Protection Approach for Text Information Retrieval. *Knowl. Based Syst.*, 195:105679, 2020. https://doi.org/10.1016/j.knosys.2020.105679. URL https://doi.org/10.1016/j.knosys.2020.105679.

[45] Z. Xu, A. Aggarwal, O. Feyisetan, and 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. https://doi.org/10.18653/v1/2020.privatenlp-1.2.

[46] Z. Xu, A. Aggarwal, O. Feyisetan, and N. Teissier. On a utilitarian approach to privacy preserving text generation. *CoRR*, abs/2104.11838, Apr. 2021. https://doi.org/10.48550/ARXIV.2104.11838.

[47] P. Yu, W. Ahmad, and H. Wang. Hide-n-Seek: An Intent-aware Privacy Protection Plugin for Personalized Web Search. In K. Collins-Thompson, Q. Mei, B. D. Davison, Y. Liu, and E. Yilmaz, editors, *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR 2018, Ann Arbor, MI, USA, July 08-12, 2018*, pages 1333–1336. ACM, 2018. https://doi.org/10.1145/3209978.3210180. URL https://doi.org/10.1145/3209978.3210180.

[48] S. Zhang, G. H. Yang, and L. Singh. Anonymizing query logs by differential privacy. In R. Perego, F. Sebastiani, J. A. Aslam, I. Ruthven, and J. Zobel, editors, *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, SIGIR 2016, Pisa, Italy, July 17-21, 2016*, pages 753–756. ACM, 2016. https://doi.org/10.1145/2911451.2914732. URL https://doi.org/10.1145/2911451.2914732.