Query Obfuscation by Semantic Decomposition

Danushka Bollegala Tomoya Machide Ken-ichi Kawarabayashi

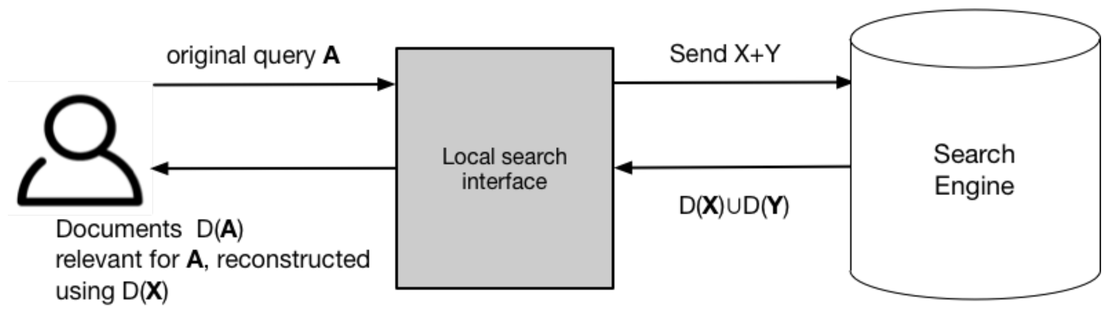
UNIVERSITY OF LIVERPOOL

NII 大学共同利用機関法人 情報・システム研究機構 国立情報学研究所 National Institute of Informatics

Motivation

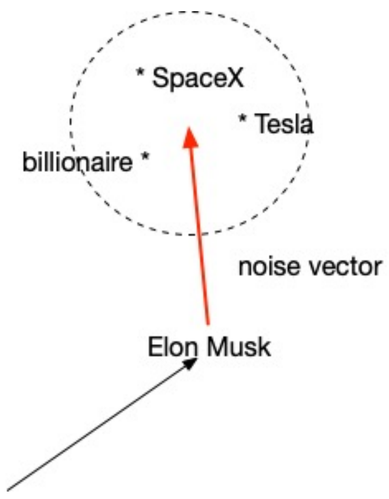
- How can we protect the privacy of Web search engine users?
- The Trade-off
  - We would like to retrieve relevant search results for our (secret) information needs [relevancy]
  - However, we do not want the search engine companies to "know" about our (secret) information needs [privacy]
- We propose a method that
  - hides the information intent of a user from the search engine by obfuscating the search queries
  - At the same time returns relevant results

Overview of the Proposal



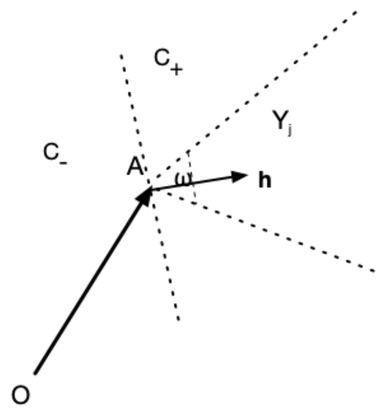
Word Embeddings to the Rescue!

- We use pretrained static word embeddings to find noisy-related terms (via cosine similarity in the random noise-added embedding space) to *decompose* a user query *A*.
- Ideally, the related terms semantically decompose the the user query *A* such that by aggregating the search results for each related term we can reconstruct the search results for *A*.



Finding distractor terms

- Sending only the noisy-related terms alone to the search engine is still risky because the search engine can apply some denoising method and still predict the original user query *A*
- Therefore, we also find a set of unrelated distractor terms for each user query *A* by random sampling from the vocabulary.



Reconstructing Search Results

- For a given user query *A*, we find related terms, X_1,…, X_n and distractor terms, Y_1,…, Y_m using the above-described methods and retrieve search results for all those terms. [Idea: the search engine will find it difficult to guess A because of the distractor terms.]
- We will discard search results for the distractor term, and reconstruct the search results for *A* as the union of the search results retrieved for the related terms

$$\mathcal{D}'(A) = \bigcup_{i=1}^{n} \mathcal{D}(X_i).$$

Obfuscity vs. Reconstructability

- We define obfuscity as the (dis)similarity between the original query and the set of queries sent to the search engine (limited to related terms)

$$\alpha = 1 - \frac{1}{|\mathcal{Q}(A)|} \sum_{q \in \mathcal{Q}(A)} \text{sim}(v(A), v(q))$$

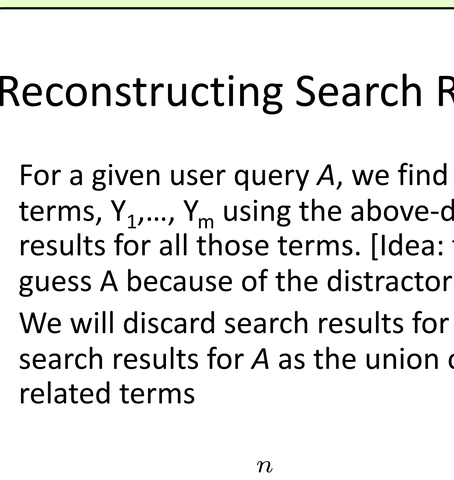
- We define reconstructability as the overlap between the set of documents obtained via the reconstruction process and the set of documents we would have obtained if we had sent the original query to the search engine

$$\rho = \frac{|\mathcal{D}(A) \cap \mathcal{D}'(A)|}{|\mathcal{D}(A)|}$$

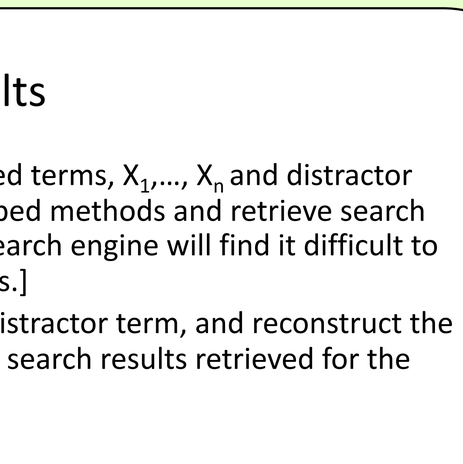
- We prove the following trade-off between these measures

$$\log \rho = \frac{cl}{2d} \left(c + 2(1-\alpha) \left|\left|v(A)\right|\right|_2\right) - \log Z$$

Experimental Results



Relationship between obfuscity and reconstructability. No noise added and no distractor terms used.



Hit rate for the k-mean clustering attacks for increasing number of clusters (k) with 20 distractor terms.

Qualitative Examples

Query	Hitler
noise	high-level
related terms	nazi, führer, gun, wehrmacht, guns, nra, pistol, bullets
obfuscity	0.867
reconstructability	0.831
Clustering Attack	Revealed Query
k=1	motagomery
k=2	albany, george
k=3	smith, albany
k=4	smith, fresno
k=5	rifle, albany

Query	mass murder
noise	high-level
related terms	terrorism, killed, wrath, full-grown
obfuscity	0.789
reconstructability	0.747
Clustering Attack	Revealed Query
k=1	richmond
k=2	fremont, death
k=4	pasadena, words
k=4	pasadena, words
k=5	pasadena, anderson

- A human evaluation shows that even human annotators find it difficult to predict the original user queries using the terms found by the proposed method.

-The proposed method can be extended other types of data such as document or images, using embeddings to implement anonymized multimodal retrieval methods.

Contact
danushka@liverpool.ac.uk
@Bollegala
https://danushka.net