City of Disguise: A Query Obfuscation Game on the ClueWeb

Maik Fröbe,[1] Nicola Lea Libera,[2] Matthias Hagen[1]

[1] Martin-Luther-Universität Halle-Wittenberg
[2] Bauhaus-Universität Weimar

Abstract We present City of Disguise, a retrieval game that tests how well searchers are able to reformulate some sensitive query in a 'Taboo'-style setup but still retrieve good results. Given one of 200 sensitive information needs and a relevant example document, the players use a special ClueWeb12 search interface that also hints at potentially useful search terms. For an obfuscated query, the system assigns points depending on the result quality and the formulated query. In a pilot study with 72 players, we observed that they find obfuscations to retrieve relevant documents but often only when they relied on the suggested terms.

Keywords: Query obfuscation · Private information retrieval · Gamification

1 Introduction

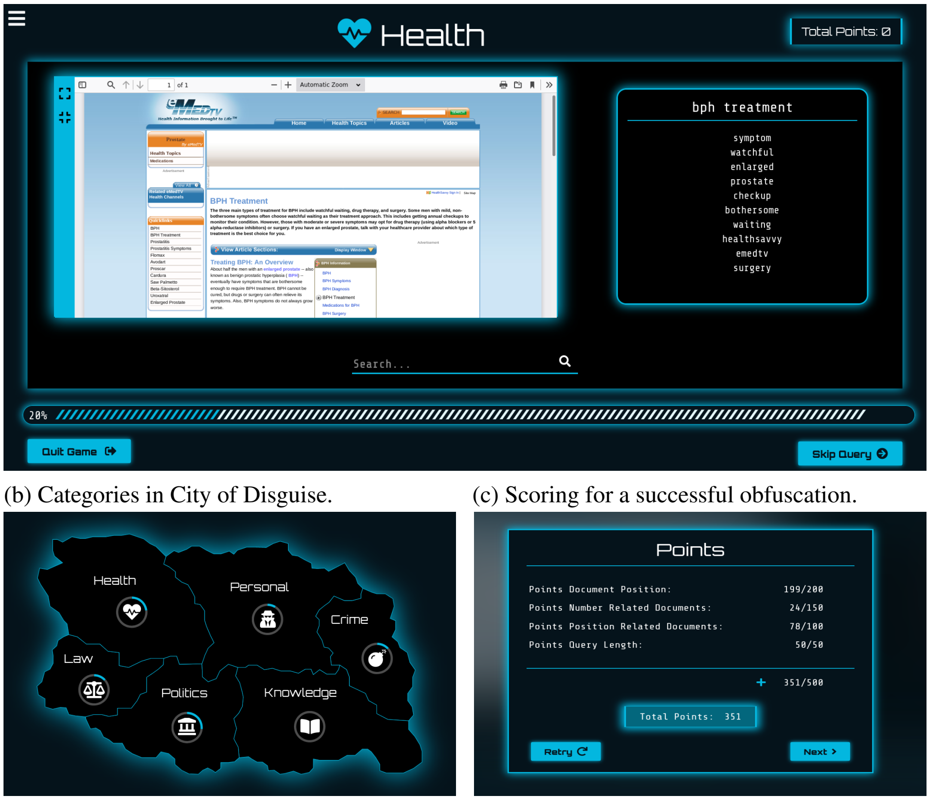
Retrieving relevant results without revealing private or confidential information is a current challenge in information retrieval [9]. Search engines can use innovative techniques to collect data while ensuring privacy [12, 14, 23]. Still, those privacy techniques are applied on the side of the search engines, requiring searchers to trust them. This trust might be unacceptable for searchers with a very sensitive information need, especially given the recent news that the police can access query logs.[3]

An option for someone not trusting a search engine but still wanting to retrieve results for some sensitive information need is to try to submit less sensitive but similar queries [3, 11]. In a way, this resembles the popular game Taboo where players try to explain a private word without using the word or some related ones. In this spirit, we present City of Disguise,[4] a game inspired by Taboo and PageHunt [15]. Players have to obfuscate a sensitive query and only submit "harmless" queries that still retrieve relevant results. The idea is that the sensitive query itself never appears in the search engine's log, which would happen for other privacy techniques like hiding the actual query in a stream of fake queries [1, 10, 17, 18, 19, 22]—an attacker would then still know that the sensitive query exists. In a gamification sense, the game's point system will particularly reward less sensitive alternative queries that still return results relevant to the original query.

[3] cnet.com/news/google-is-giving-data-to-police-based-on-search-keywords-court-docs-show/

[4] Demo: https://demo.webis.de/city-of-disguise
Screencast: https://demo.webis.de/city-of-disguise/screencast
Code and Data: https://github.com/webis-de/ecir22-query-obfuscation-game

(a) The search interface in City of Disguise for the sensitive query `bph treatment`.



(b) Categories in City of Disguise.

(c) Scoring for a successful obfuscation.

```
Points Document Position:                199/200
Points Number Related Documents:          24/150
Points Position Related Documents:        78/100
Points Query Length:                      50/50

                                   +     351/500

             Total Points:  351
```

Figure 1. The main elements in City of Disguise: (a) the search interface, (b) the city map where players select a category, and (c) the scoring scheme shown to players after query submission.

Playing City of Disguise is pretty simple. In the city map (cf. Figure 1 (b)), a player chooses a category (e.g., health-related topics) and the search interface (cf. Figure 1 (a)) is opened for a random "unsolved" sensitive information need from that category. In the search interface, the to-be-obfuscated sensitive query, a relevant target document for the underlying information need, and a list of suggested terms are shown (terms from the target document with highest TF · IDF scores). When a player submits a query, a score is derived (cf. Figure 1 (c)) with which they may compete in the public leaderboard or they may choose to try to further improve their score.

All interactions in the game are logged in an anonymized way. Besides showing scores in the leaderboards, the logging also enables analyses of human obfuscation strategies. Successful players' strategies can then be compared to automatic query obfuscation approaches [2, 3, 4, 5, 11] which even might inspire improvements.

2 Search System and Game Design

Our new query obfuscation game comes with 200 sensitive queries in six categories that we have manually selected from a pool of 700 candidate queries. Each sensitive query has one relevant document assigned that we show as a target document to players to simplify the obfuscation process. To test their obfuscations with fast response times, players submit their queries to a search engine with 0.6 million ClueWeb12 documents.

Document Collection and Rendering Target Documents. We use the ClueWeb as a resource because it is widely used in research in information retrieval [7, 8], and since we can reuse relevance judgments. We show a relevant document for each sensitive query to the players to simplify the query obfuscation process. Since the ClueWeb does not include all resources to nicely display a page, we replace links in the documents to stylesheets and images with links to the Wayback machine to the corresponding snapshots (if available). We render the final HTML into a PDF document (using the Wayback Machine) so that the target pages still render nicely inside our preview image and allow zooming, while the players can still copy terms from the relevant document.

Selection of Sensitive Queries. We select the 200 pairs of sensitive queries with a respective relevant document for the obfuscation game from 750 candidate queries that we extract from the 96 sensitive queries published by Arampatzis et al. [3], 65 sensitive TREC Web track queries on the ClueWebs, and 589 sensitive queries from the AOL query log [16]. Especially from the 96 sensitive queries by Arampatzis et al. [3], we remove those with pornographic or hateful intent because we show a document relevant to each query in our game. For each of the remaining sensitive queries, we retrieve the top-3 ClueWeb12 documents with ChatNoir [6] and render the documents (including CSS and image resources from the Wayback Machine). For each sensitive query, we review the top-3 documents and omit documents looking odd (due to missing CSS or images) or documents that are irrelevant to the sensitive query, retaining only the most relevant document for each query. We assign the remaining 204 valid query-document pairs into six categories, selecting 200 final queries that provide the best balance of all six categories (Figure 1 (b) gives an overview of the available categories).

Scoring Query Obfuscations. To motivate players to improve their obfuscated query multiple times, we show a score composed of four subscores to suggest potential ways for improvement. We calculate the score by submitting the obfuscated query (not allowing any queries that reuse terms from the sensitive query) against a test search engine indexing a 0.6 million document sample. The score combines the position of the relevant document, the query length, the recall, and the mean average precision for the 100 documents retrieved for the sensitive query. We show all four subscores to the players to indicate whether a query could be improved (cf. Figure 1 (c)).

To allow fast feedback cycles for players, we use a setup similar to Arampatzis et al. [3] and submit obfuscated queries against a search engine with a small sample of 0.6 million ClueWeb12 documents. To ensure that each of our 200 sensitive queries has enough relevant documents, we include the top-1000 ChatNoir [6] results for each sensitive query into the sample. We complement those 0.2 million documents by sampling 0.4 million documents from the ClueWeb12 with the sampling strategy of Arampatzis et al. [3]. We index this document sample with the BM25 implementation of Anserini [20] using the default settings (stemming with the Porter stemmer and removing stopwords using Lucene's default stopword list for English).

Table 1. Overview of the effectiveness of obfuscated queries in ChatNoir and the games' document sample ('Sample'). We report the MRR, the number of documents retrieved for the original query ('Ori.'), and the number of retrieved relevant documents ('Rel.'). We show results for automatically obfuscated queries and four different types of queries submitted by players.

					Our Sensitive Queries				**Sensitive Web Track Queries**			
		Obfuscated Queries			ChatNoir		Sample		ChatNoir		Sample	
		Count	Length	Time	MRR	Ori.	MRR	Ori.	MRR	Rel.	MRR	Rel.
Players	Only Suggestions	130 / 21	2.42	40.50 s	0.093	5.223	0.325	67.592	0.010	3.094	0.152	3.691
	Some Suggestions	556 / 125	4.53	42.45 s	0.046	4.667	0.258	85.829	0.013	3.632	0.038	3.568
	No Suggestions	576 / 157	2.88	44.39 s	0.029	2.935	0.082	38.932	0.015	1.783	0.024	3.316
	New Word	559 / 158	3.57	46.27 s	0.002	1.517	0.051	49.992	0.002	1.235	0.005	2.790
	Automatic	1025 / 327	2.91	—	0.088	9.229	0.420	84.264	0.014	2.872	0.042	3.743

3 Evaluation

We test our query obfuscation game and the ability of players to obfuscate sensitive information needs in a pilot study with 72 participants. We recruited players from two information retrieval courses and mailing lists at our universities. We logged 1,462 obfuscated queries, with an average of 43 seconds to formulate an obfuscated query.

Table 1 compares the effectiveness of the obfuscated queries that the players formulated in our pilot study to queries automatically obfuscated by formulating keyqueries for the target document [11] using the suggested terms as vocabulary. We report the MRR for finding the given relevant document, the number of documents retrieved from the top-100 ranking when submitting the sensitive query (Ori.), and the number of retrieved relevant documents (Rel.) for queries with relevance judgments from the ClueWeb tracks. We split the human obfuscations into four categories: (1) queries where all terms come from the game's suggestions (Only Suggestions), (2) queries with at least one term from the game's suggestions and one new term (Some Suggestions), (3) queries without suggestions (No Suggestions), and (4) queries with a term outside the shown relevant document (New Word). Overall, we find that the obfuscation effectiveness decreases the more the players deviate from the game's term suggestions. While players who use only suggestions slightly improve upon automatic obfuscation (MRR of 0.093 vs. 0.088), creative obfuscations that include new words are rather ineffective (MRR of 0.002). Those observations are also confirmed by the evaluations using the sensitive Web Track queries with real relevance judgments.

4 Conclusion and Future Work

Our new query obfuscation game tests a player's ability to hide sensitive information needs from a search engine while still retrieving relevant results. We plan to maintain the game as part of ChatNoir [6] and to add more topics and different search engines in the future (e.g., Transformer-based re-rankers [21] or dense retrieval models [13]).

From the game's logs, we want to learn how searchers manually obfuscate sensitive information needs. This knowledge could help to improve automatic query obfuscation approaches that one could apply when querying an untrusted search engine for a sensitive information need that should not appear "unencrypted" in the engine's log files.

Bibliography

[1] Ahmad, W.U., Rahman, M., Wang, H.: Topic model based privacy protection in personalized web search. In: Perego, R., Sebastiani, F., Aslam, J.A., Ruthven, I., Zobel, J. (eds.) Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, SIGIR 2016, Pisa, Italy, pp. 1025–1028, ACM (2016)

[2] Arampatzis, A., Drosatos, G., Efraimidis, P.S.: A versatile tool for privacy-enhanced web search. In: Serdyukov, P., Braslavski, P., Kuznetsov, S.O., Kamps, J., Rüger, S.M., Agichtein, E., Segalovich, I., Yilmaz, E. (eds.) Advances in Information Retrieval - 35th European Conference on IR Research, ECIR 2013, Moscow, Russia, Lecture Notes in Computer Science, vol. 7814, pp. 368–379, Springer (2013)

[3] Arampatzis, A., Drosatos, G., Efraimidis, P.S.: Versatile query scrambling for private web search. Inf. Retr. J. **18**(4), 331–358 (2015)

[4] Arampatzis, A., Efraimidis, P.S., Drosatos, G.: Enhancing deniability against query-logs. In: Advances in Information Retrieval - 33rd European Conference on IR Research, ECIR 2011, Dublin, Ireland, pp. 117–128 (2011)

[5] Arampatzis, A., Efraimidis, P.S., Drosatos, G.: A query scrambler for search privacy on the internet. Inf. Retr. **16**(6), 657–679 (2013)

[6] Bevendorff, J., Stein, B., Hagen, M., Potthast, M.: Elastic ChatNoir: Search engine for the ClueWeb and the Common Crawl. In: Azzopardi, L., Hanbury, A., Pasi, G., Piwowarski, B. (eds.) Advances in Information Retrieval. 40th European Conference on IR Research (ECIR 2018), Lecture Notes in Computer Science, Springer, Berlin Heidelberg New York (Mar 2018)

[7] Collins-Thompson, K., Bennett, P.N., Diaz, F., Clarke, C., Voorhees, E.M.: TREC 2013 Web track overview. In: Voorhees, E.M. (ed.) Proceedings of The Twenty-Second Text REtrieval Conference, TREC 2013, Gaithersburg, Maryland, USA, November 19-22, 2013, NIST Special Publication, vol. 500-302, National Institute of Standards and Technology (NIST) (2013)

[8] Collins-Thompson, K., Macdonald, C., Bennett, P.N., Diaz, F., Voorhees, E.M.: TREC 2014 Web track overview. In: Voorhees, E.M., Ellis, A. (eds.) Proceedings of The Twenty-Third Text REtrieval Conference, TREC 2014, Gaithersburg, Maryland, USA, November 19-21, 2014, NIST Special Publication, vol. 500-308, National Institute of Standards and Technology (NIST) (2014)

[9] Culpepper, J.S., Diaz, F., Smucker, M.D.: Research frontiers in information retrieval: Report from the third strategic workshop on information retrieval in Lorne (SWIRL 2018). SIGIR Forum **52**(1), 34–90 (2018)

[10] Domingo-Ferrer, J., Solanas, A., Castellà-Roca, J.: H(k)-private information retrieval from privacy-uncooperative queryable databases. Online Inf. Rev. **33**(4), 720–744 (2009)

[11] Fröbe, M., Schmidt, E.O., Hagen, M.: Efficient Query Obfuscation with Keyqueries. In: 20th International IEEE/WIC/ACM Conference on Web Intelligence (WI-IAT '21), ACM (Dec 2021), https://doi.org/10.1145/3486622.3493950, URL https://dl.acm.org/doi/10.1145/3486622.3493950

[12] Hong, Y., He, X., Vaidya, J., Adam, N.R., Atluri, V.: Effective anonymization of query logs. In: Cheung, D.W., Song, I., Chu, W.W., Hu, X., Lin, J.J. (eds.) Proceedings of the 18th ACM Conference on Information and Knowledge Management, CIKM 2009, Hong Kong, China, pp. 1465–1468, ACM (2009)

[13] Karpukhin, V., Oguz, B., Min, S., Lewis, P.S.H., Wu, L., Edunov, S., Chen, D., Yih, W.: Dense passage retrieval for open-domain question answering. In: Webber, B., Cohn, T., He, Y., Liu, Y. (eds.) Proceedings of the 2020 Conference on Empirical Methods in

Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020, pp. 6769–6781, Association for Computational Linguistics (2020)

[14] Kumar, R., Novak, J., Pang, B., Tomkins, A.: On anonymizing query logs via token-based hashing. In: Proceedings of the 16th International Conference on World Wide Web, WWW 2007, Banff, Alberta, Canada, pp. 629–638 (2007)

[15] Ma, H., Chandrasekar, R., Quirk, C., Gupta, A.: Page Hunt: Improving search engines using human computation games. In: Allan, J., Aslam, J.A., Sanderson, M., Zhai, C., Zobel, J. (eds.) Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2009, Boston, MA, USA, July 19-23, 2009, pp. 746–747, ACM (2009)

[16] Pass, G., Chowdhury, A., Torgeson, C.: A picture of search. In: Jia, X. (ed.) Proceedings of the 1st International Conference on Scalable Information Systems, Infoscale 2006, Hong Kong, May 30-June 1, 2006, ACM International Conference Proceeding Series, vol. 152, p. 1, ACM (2006)

[17] Peddinti, S.T., Saxena, N.: On the privacy of web search based query obfuscation: A case study of TrackMeNot. In: Atallah, M.J., Hopper, N.J. (eds.) Privacy Enhancing Technologies, 10th International Symposium, PETS 2010, Berlin, Germany, Lecture Notes in Computer Science, vol. 6205, pp. 19–37, Springer (2010)

[18] Peddinti, S.T., Saxena, N.: Web search query privacy: Evaluating query obfuscation and anonymizing networks. J. Comput. Secur. **22**(1), 155–199 (2014)

[19] Toubiana, V., Subramanian, L., Nissenbaum, H.: TrackMeNot: Enhancing the privacy of web search. CoRR **abs/1109.4677** (2011)

[20] Yang, P., Fang, H., Lin, J.: Anserini: Enabling the use of Lucene for information retrieval research. In: Kando, N., Sakai, T., Joho, H., Li, H., de Vries, A.P., White, R.W. (eds.) Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, Shinjuku, Tokyo, Japan, pp. 1253–1256, ACM (2017)

[21] Yates, A., Nogueira, R., Lin, J.: Pretrained transformers for text ranking: BERT and beyond. In: Diaz, F., Shah, C., Suel, T., Castells, P., Jones, R., Sakai, T. (eds.) SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11-15, 2021, pp. 2666–2668, ACM (2021)

[22] Yu, P., Ahmad, W.U., Wang, H.: Hide-n-Seek: An intent-aware privacy protection plugin for personalized web search. In: Collins-Thompson, K., Mei, Q., Davison, B.D., Liu, Y., Yilmaz, E. (eds.) The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR 2018, Ann Arbor, MI, USA, pp. 1333–1336, ACM (2018)

[23] Zhang, S., Yang, G.H., Singh, L.: Anonymizing query logs by differential privacy. In: Perego, R., Sebastiani, F., Aslam, J.A., Ruthven, I., Zobel, J. (eds.) Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, SIGIR 2016, Pisa, Italy, pp. 753–756, ACM (2016)