Permutation of Web Search Query Types for User Intent Privacy

Kato Mivule
Department of Computer Science
Norfolk State University
Norfolk, Virginia, USA

***Abstract*—Privacy remains a major concern when using search engines to find for information on the web due to the fact that search engines own massive resources in preserving search logs of each user and organizations. However, many of the present query search privacy practices require the very same search engine and third party to collaborate, making privacy even more difficult. Therefore, as a contribution, we present a heuristic, permutation of web search query types, a non-cryptographic heuristic that works by formation of obfuscated search queries via permutation of query keyword categories. Preliminary results from this study show that web search query and specific user intent privacy might be achievable from the user side without involvement of the search engine or other third parties by the permutation of web search query types.**

***Keywords*—Web search query privacy; user intent privacy; search engines; Information Retrieval**

I. Introduction

Privacy remains a major concern when using search engines to search for information on the Internet due to the fact that search engines own considerable resources in keeping user and organization search logs. However, many of the present query search privacy practices necessitate the very same search engine and third party applications to collaborate, making web search privacy a challenge.

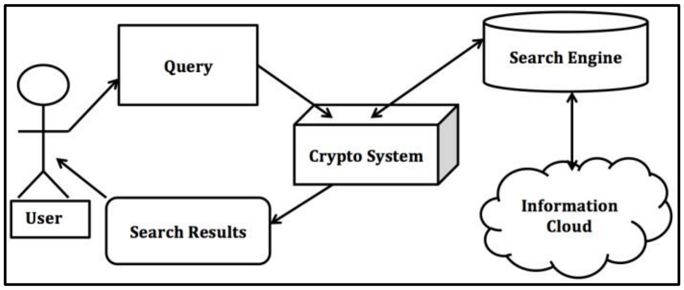


Fig. 1. Third party web search privacy system

As illustrated in Figure 1, many web search query privacy techniques require collaboration with a search engine or third party cryptographic system. In our proposed heuristic, as shown in Figure 2, the user has full control in obfuscating their search intent without the need of a third party. In this paper, we present a heuristic, permutation of web search query types, a non-cryptographic heuristic that works by formation of obfuscated search queries via permutation of query keyword categories. We also make a distinction between concept user intent and specific user intent, with the goal of giving users privacy controls over their specific user search intent without involvement of third parties. Concept user intent is concerned with general aspects that users search for, while specific user intent is concerned with the specific item the user intends to search for. Preliminary results from this study show that web search query and specific user intent privacy might be achievable from the user side without involvement of the search engine collaboration or other third parties by the permutation of web search query types. In this proposed heuristic, users initially generate obfuscated queries based on permutations of different query keyword types. The generated permutated queries are then combined with the original search terms to search for information in the search engine at the same time.

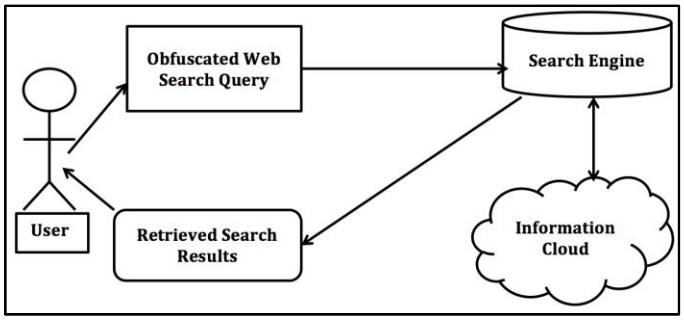


Fig. 2. User-side web search privacy system

The rest of the paper is organized as follows. In Section 2, we discuss the background and related work. In Section 3, we outline the suggested heuristic. In Section 4, we discuss preliminary results. In Section 5, a conclusion and future works is given.

II. Background and Related Work

A web search engine: This is a software application normally hosted by search engine organizations and used mainly to search for information and index documents on the web[1][2][3]. *Web search queries:* These are words and phrases input by a user into a search engine to retrieve relevant indexed documents stored by the search engine [4]. There are three main web search queries categories [5] [6] [7]: *Informational queries* – web search queries that deal with general subjects, and retrieve large related result numbers, e.g. “Trains” and “Tourism”. *Navigational queries* – these are queries that deal with searching for a specific website or webpage e.g. “Twitter” and “Yahoo Movies”. *Transactional* *queries* – these are queries in which the user seeks to make an online action like buy an item, stream music, or watch a movie; e.g., "Buy airline tickets". Web search queries in most cases are always concise, imprecise, contain subtopics, and can be viewed as in two major categories – faceted and ambiguous queries [6]: *Faceted queries:* Faceted queries can be comprised of subtopics, but are non-ambiguous, precise, and return particular and relevant results. *Ambiguous queries:* these types of queries typically have more than one denotation, and so the search engine returns results that might not be pertinent to the user.

Data privacy: This is the process in which individual or entity information is protected against unauthorized disclosure. In this case, user intent, which is the real purpose for an individual or entity issuing a query on a search engine, could be considered private information [8].

Single-party privacy search: Single-party privacy search is a privacy technique that works by permitting users to generate their own public profiles with need to make changes on the server side, by using cataloged topics of interests, generating false queries that are amalgamated with real queries, and implementing all produced queries simultaneously in the web search engine [9]. The phony queries are produced using a knowledge-base, such as, the open directory project, to interpret the singular intent based on the sematic distance between the false and real intent in the query outcomes [9]. We chart a parallel methodology to the single party privacy search, nevertheless, in our heuristic, focus is placed on permutation of topical query keywords in the production of disguised queries and the pseudo-user profile in this case could be created via deflecting URL clicks produced from retrieved search results.

Web search query disambiguation: This is the process in which query search terms undergo reformation and refinement to eliminate any ambiguity so as to better predict user intent in order to and retrieve highly relevant search results for o the user [10].

Web search query reformation: Similar to query disambiguation, reformation is the process in which search engines use query enhancement methods to modify and accurately capture user intent. The reformatted query is then presented as an alternative to the user in replacement for what the search engine perceived as ambiguous. It is important for privacy practitioners to note that search engines store both the original query issued by the user and the reformed query selected by the user, to correct future errors, typos, and for adjustments of the query for personalized results. Although search engines could yield improved and more unambiguous search results for the web user, search query reformation can be viewed as vulnerability against web search query obfuscation [11][12][13][14].

Precision and recall: these are the two main measures employed by search engines to calculate the efficiency of web search queries in regards to retrieved relevant documents. The value for these measures is between 0 and 1, with 1 being the best value for both precision and recall [10]. The formal expression for precision and recall are:

$$Precision\ (P) = \frac{Total\ of\ retrieved\ relevant\ articles}{Total\ of\ retrieved\ articles} \quad (1)$$

$$Recall\ (R) = \frac{Total\ of\ retrieved\ relevant\ articles}{Total\ of\ relevant\ articles} \quad (2)$$

The average precision metric: can be employed to measure how efficient the obfuscated web search queries are by measuring the precision and relevance of documents returned to a user. Suggested by Turpin and Scholer (2006), the average precision (AP) quantifies the efficiency of search queries in retrieving relevant documents and is formally expressed as follows [15]:

$$AP = \frac{1}{\sum_{i=1}^{k} r_i} \sum_{i=1}^{k} r_i \left(\frac{\sum_{j=1}^{i} r_j}{i} \right) \quad (3)$$

The symbol r_i returns a value of 1 if the retrieved document is relevant otherwise a 0 is returned. *k* symbolizes the amount of retrieved items, that is, the *top-k* retrieved items.

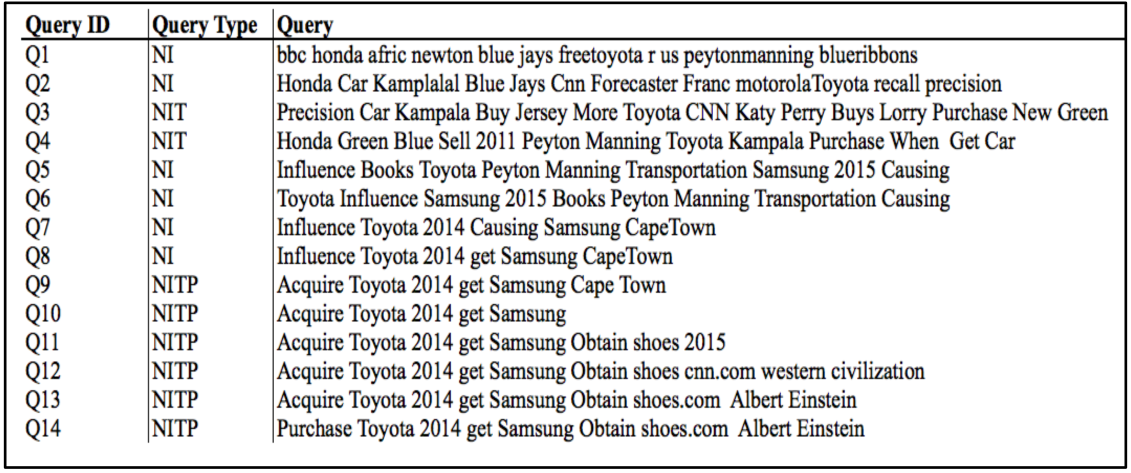
Plausible deniability: Plausible deniability search is web search query privacy process in which a set of *k-1* dummy queries with attributes analogous to the original but on dissimilar subjects, are produced and used to obscure original queries [16]. Plausible deniability search demands that every unprecedented query be replaced with a consistent but analogous dummy query with the intention to retrieve outcomes very comparable to those projected from the original query. Any subsection of *k* dummy queries will generate statistically indistinguishable outcomes to a corresponding unprecedented set of *k* queries [16]. The produced dummy queries are implemented at the identical time to cover the intent of the user, creating a difficulty in detecting which precise query was deliberated on by the user; leaving the burden of proof to the search engine to ascertain which query fits a particular user [16]. Formally plausible deniability search is expressed as follows [16]:

"The conditions for plausible deniability privacy (PD-Privacy) $\boldsymbol{Q_i}$ *are realized if: (i) the user can show that any query* $\boldsymbol{Q_j} \in \boldsymbol{S}$ *would have produced the set* $\boldsymbol{S}$ *with the same likelihood as* $\boldsymbol{Q_i}$*; (ii) all* $\boldsymbol{Q_j} \in \boldsymbol{S}$ *are on distinctive subjects; (iii) all* $\boldsymbol{Q_j} \in \boldsymbol{S}$ *are likely similar to the authentic query. Where* $\boldsymbol{S} = \{\boldsymbol{Q_1}, \dots, \boldsymbol{Q_k}\}$*, the set of queries at time* $\boldsymbol{t}$*, kept in a log on a server;* $\boldsymbol{Q_i}$*, is the authentic query by the user;* $\boldsymbol{Q_j}$ *is the set of dummy queries similar to the authentic query".*

Furthermore, plausible deniability necessitates that at the query implementation phase, all queries must be correlated to the leading theme but with low Euclidean distance in the semantic space [16]. Although our proposed heuristic meets some benchmarks of plausible deniability search, our methodology deviates from the standard plausible deniability search approach by producing dummy queries based on the query type instead of the query topic. We focus on creating dummy query keywords that comprise a permutation of navigational, informational, transactional, temporal, and natural language processing query keywords. The query keywords could be correlated or dissimilar to the unprecedented query but are joined together with the unprecedented query keywords during query implementation phase. Each permutation is expected to produce fluctuating outcomes.

III. Methodology

Permutation of Web Search Query Types Heuristic: In this section, we discuss how permutation is used in our suggested heuristic after identifying the major web search query categories. The typical search query groups as observed in literature is used in this proposed heuristic in the following way [5] [6] [7]: *Navigational* queries symbolized as *N*. *Informational* queries symbolized as *I*. *Transactional* queries symbolized as *T*. *Natural language processing* queries symbolized as *L*. *Temporal* queries symbolized as *P*. For effectual privacy, the user should create a set of dummy queries that comprise, navigational, informational, transactional, natural language processing, and temporal search query types.



Query ID	Query Type	Query
Q1	NI	bbc honda afric newton blue jays freetoyota r us peytonmanning blueribbons
Q2	NI	Honda Car Kamplalal Blue Jays Cnn Forecaster Franc motorolaToyota recall precision
Q3	NIT	Precision Car Kampala Buy Jersey More Toyota CNN Katy Perry Buys Lorry Purchase New Green
Q4	NIT	Honda Green Blue Sell 2011 Peyton Manning Toyota Kampala Purchase When Get Car
Q5	NI	Influence Books Toyota Peyton Manning Transportation Samsung 2015 Causing
Q6	NI	Toyota Influence Samsung 2015 Books Peyton Manning Transportation Causing
Q7	NI	Influence Toyota 2014 Causing Samsung CapeTown
Q8	NI	Influence Toyota 2014 get Samsung CapeTown
Q9	NITP	Acquire Toyota 2014 get Samsung Cape Town
Q10	NITP	Acquire Toyota 2014 get Samsung
Q11	NITP	Acquire Toyota 2014 get Samsung Obtain shoes 2015
Q12	NITP	Acquire Toyota 2014 get Samsung Obtain shoes cnn.com western civilization
Q13	NITP	Acquire Toyota 2014 get Samsung Obtain shoes.com Albert Einstein
Q14	NITP	Purchase Toyota 2014 get Samsung Obtain shoes.com Albert Einstein

Fig. 3. A Sample of Obfuscated Queries using the Permutation Heuristic

The suggested set of search query formation will be a permutation of items in set *Q = {NITLP}*. This entails that any obfuscated query formation will comprise a permutation of *N*, *I*, *T*, *L*, *P* keywords. Formally the quantity of permutations of any *k* items can be calculated as follows [17]:

$$P(n,k) = \frac{n!}{(n-k)!} \quad for\ 0 \leq k \leq n \qquad (4)$$

Where n is the number in set Q, and k is the quantity of permutations for any k items. Assuming the set *Q = {NITLP}*, the quantity of permutations of the five objects in the set Q at an occurrence is:

$$P(n,k) = \frac{n!}{(n-n)!} = \frac{n!}{0!} = n! \qquad (5)$$

$P(5,5) = 5 * 4 * 3 * 2 * 1 = 120.$ Consequently the quantity of permutations from the five objects in set Q produces 120 arrangements of conceivable obfuscated query formations. For the preliminary experiment done suggested heuristic, not every permutation of each k or all the k=n (where n=5) items was used in the query implementation phase. Nonetheless, an assortment was done from the *k=1, k=2, k=3, k=4* and *k=n* permutation categories. The subsequent permutations were chosen for the preliminary experiment to examine the hypothesis: *Q* $_p$ *={I, IT, IP, TP, IL, NI, NIT, NIP, IPL, ITP, NITP, ITPL, NIPL, NITL, NITPL}*. Because order is essential to the query formation, permutation is employed as core for producing several query sets. Although the permutations are determinate, the incentive is that for each search term that necessities obfuscation, the user has the prospect to produce permutations (*n!*) of obfuscated queries that are problematic for the search engine to disambiguate immediately. This is with the assumption that the search engine does not know apriori what that particular user intent is and how that same user intends to construct their search queries. A sample of the obfuscated queries using the suggested heuristic is shown in Figure 3.

IV. Experiment and Preliminary Results

The experiment: The aim of the experiment was to disguise queries using the suggested search heuristic of permutation of web search query types, namely, transaction, navigational, temporal, and informational queries. In the experiment, obfuscation of web search queries is done for a hypothetical entity that wants to unambiguously *"buy a Toyota"* automobile. In the experiment, real search queries are embedded in a set of distracting keywords but keeping with the permutations as suggested in the heuristic. The goal is that retrieved outcomes should allow the hypothetical user to read or surf Toyota related web snippets but without letting the search engine know the true intent of the search. The intention of the experiment was to apply the suggested heuristic to obfuscate the search query, "*Buy Toyota*", and explicit intent that the user wanted to "*buy a Toyota*". In the experiment, 121 search queries were produced after the permutation process. Each created query permutation was then implemented simultaneously with the original search query in the search engine. In the experiment, the user is logged into their browser/search engine account. The aim is that the browser/search engine logs the user query activity and generates a browsing and search history that further helps serve as a decoy and make it difficult for the search engine to decipher user intent. In this experiment, the Google search engine and Chrome browser were employed. A sum of 12,177 Google articles (snippets) was retrieved and analyzed using text mining tools for relevant and non-relevant documents.

Documents that were viewed as relevant contained phrases that were related to "*Buy Toyota*". Non-relevant articles were those that did not contain any phrase related to "Buy Toyota" and were viewed as diversionary.

Preliminary results: The examination of results from the experiment was to show the number of retrieved documents that were considered relevant. In Figure 4, retrieval search outcomes from the obfuscated queries are shown with overall 121 queries formulated and implemented. The amount of retrieved and relevant documents (snippets) is presented on the y-axis with each query (*Q1* to *Q121*) corresponding on the y-axis. Additionally Figure 4 shows a summary of how the obfuscated web search queries performed as compared to the number of retrieved relevant results. The *x*-axis shows the obfuscated queries *Q1* to *Q121*, whereas the *y*-axis depicts the quantity of retrieved relevant documents. The two series illustrated in the graph show the documents retrieved, with the top series showing overall retrieved documents whereas the lower series depicts the relevant documents. In this experiment, the key phrase "*Toyota*" was employed in selecting all documents that were viewed as relevant. For example, for the query *Q1*, 126 overall documents were retrieved but only ten documents were relevant, that is, only documents that comprised the key search phrase, "*Toyota*". The other residual non-relevant documents are associated to the dummy queries and deflecting key phrases employed to obfuscate the main intent and key search phrase, "*Toyota*". The specific intent of the user was to intentionally buy a Toyota. Yet deprived of any encryption methods, this experiment pursued if it were achievable to obfuscate particular user intent and yet retrieve relevant documents to the user. As shown in Figure 4, the amount of retrieved documents was cut off at an average of 100 documents per query. In this experiment, the *top-k* articles were chosen, with *k* being an average of 100 articles. The rational was grounded on research that the average user is inclined to glance the first page of the search engine results and browse the first 20 top articles (snippets) [18][19][20][21]. In certain occasions, particular search engine outcomes were less that 100. In such situations, owing to the type of the query formation employing a set of key words to obscure the intended key search term, retrieved results were abridged with demands to reformulate the query, a technique that Google and other search engines do in order to have the user to enter a more "correct" search term. However, query reformation was evaded in this experiment; instead focus was placed on observing the effect of the obfuscation in terms of the retrieved relevant documents. The other observation from Figure 4 is that some queries retrieved larger amounts of relevant documents in comparison with other queries. For example, queries *Q1* and *Q4* retrieved 126 and 106 documents correspondingly. However, only 10 and 17 were considered relevant for each of those queries.

Meanwhile, for queries *Q22* and *Q50*, a total of 105 and 107 documents were retrieved correspondingly. Yet only 74 and 82 were considered relevant documents for each respective query. One reason for such an effect in the various numbers of retrieved documents is that the obfuscated search query formation, in this case, could be viewed as an adjustable parameter. This is to say that the amount of obfuscation during query formation affects the number of relevant articles retrieved in context of the user's intended original query. Therefore, the reverse is also true; in that less obfuscated the query is the more retrieved relevant results one generates. Additionally, the low correlation value shown in Table 1, among the retrieved and relevant results, is at *0.21* signifying that there exists little relation. Yet this correlation value could be viewed, as a signal that the obfuscation methods used in the making of the web search query could be effectual. Although the low correlation value could be a good gauge for a better privacy, the difficulty of usability still remains. The relevancy of retrieved documents to the user in regards to the obfuscated queries remains essential to the user.

Moreover, as shown in Table 1, the null hypothesis that the entire retrieved articles are relevant is rejected given that the Chi-square shows the $p\text{-}value < 0.05$. However, caution should be taken when reading these results. Although the p-value is less that 0.05 and could signify that the permutation of the web search query type heuristic is effective in granting obfuscation and confidentiality for specific user intent, other factors such as user clicks on URLs and advertisement links could still reveal the specific user intent to a search engine. Furthermore, we are only presenting our suggested heuristic and preliminary results. More extended experiments need to be done that will take the user clicks on URLs and advertisement links into consideration and will be done as part of our future works.

Privacy verses usability: Furthermore, the type of query permutation determines the retrieved relevant results. This highlights the challenge of finding equilibrium between privacy and usability while showing the need to consider necessary trade-offs. To highlight this challenge, in Figures 6, 7, 8, and 9, four queries, *Q12*, *Q27*, *Q22*, and *Q102* are chosen – two with the least average precision values two with the highest average precision values as shown in Figure 5. The precision and recall values are depicted in each graph; the *x*-axis shows the amount of retrieved articles, while the *y*-axis shows the precision and recall values. The recall values continuously move near the value 1, since precision is computed basing on the *top-k* retrieved articles. Hence the average precision is that value at which the recall is 1. Thus for queries *Q12*, *Q27*, *Q22*, and *Q102*, the average precision values are 0.376, 0.342, 0.735, and 0.916 correspondingly as illustrated in Figure 5.

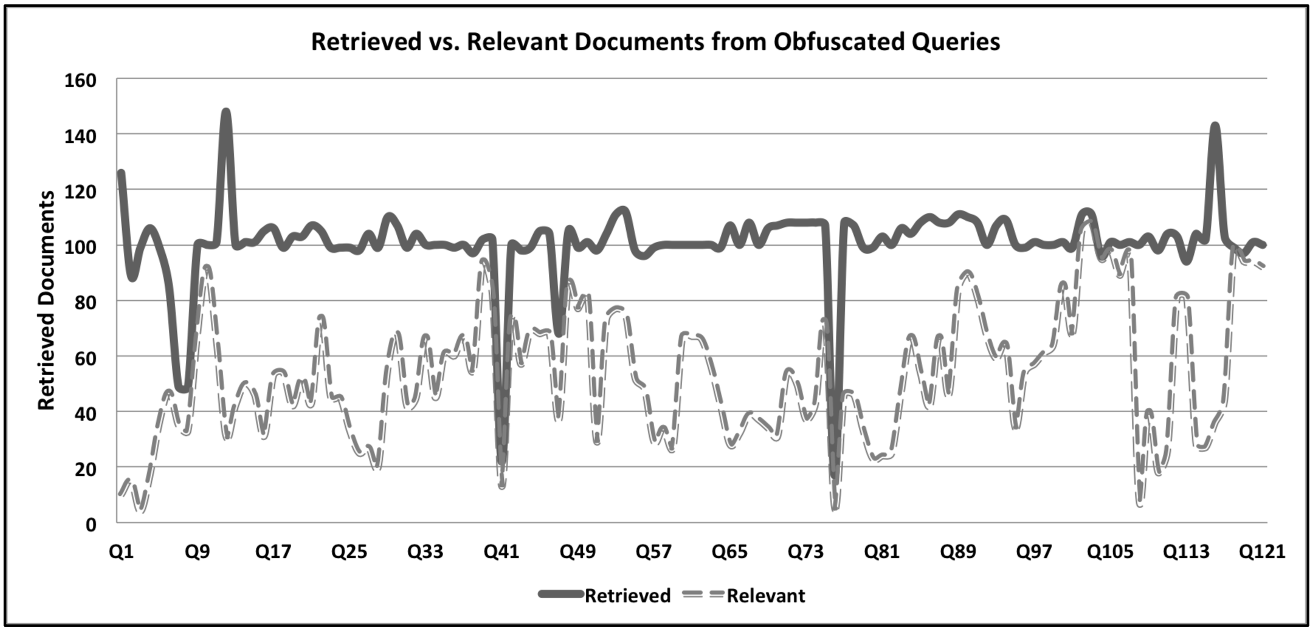


Fig. 4. Retrieved versus relevant documents as per search query Q1 to Q121

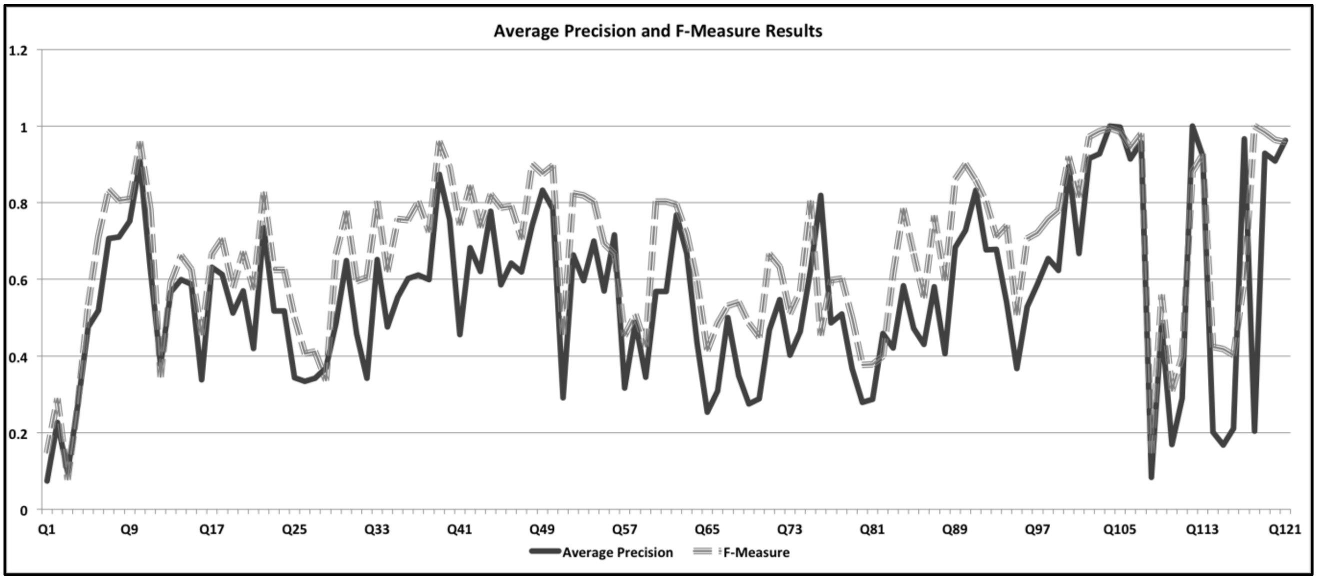


Fig. 5. Graphical presentation of Average Precision and F-Measure Results

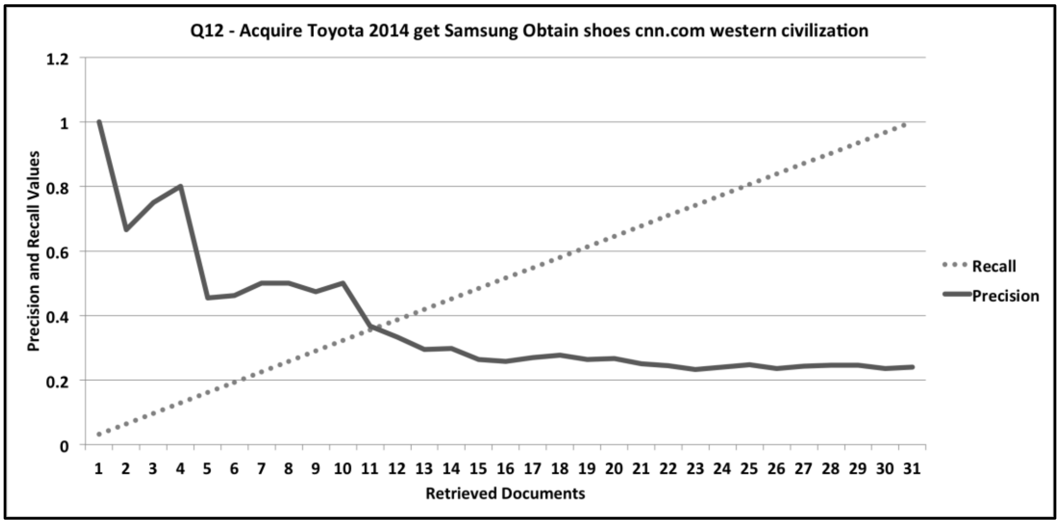


Fig. 6. Recall and Precision @ *k* for query Q12

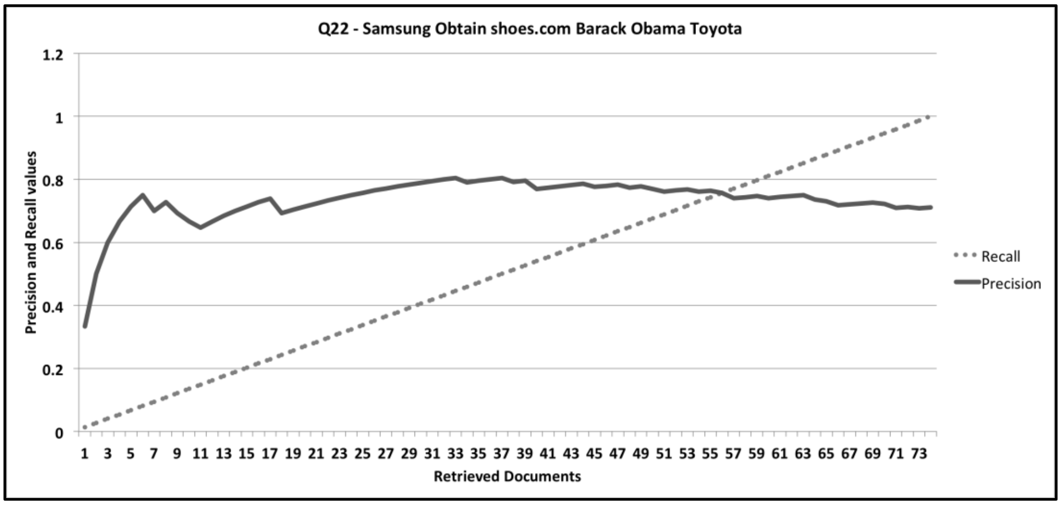


Fig. 7. Recall and Precision @ *k* for query Q27

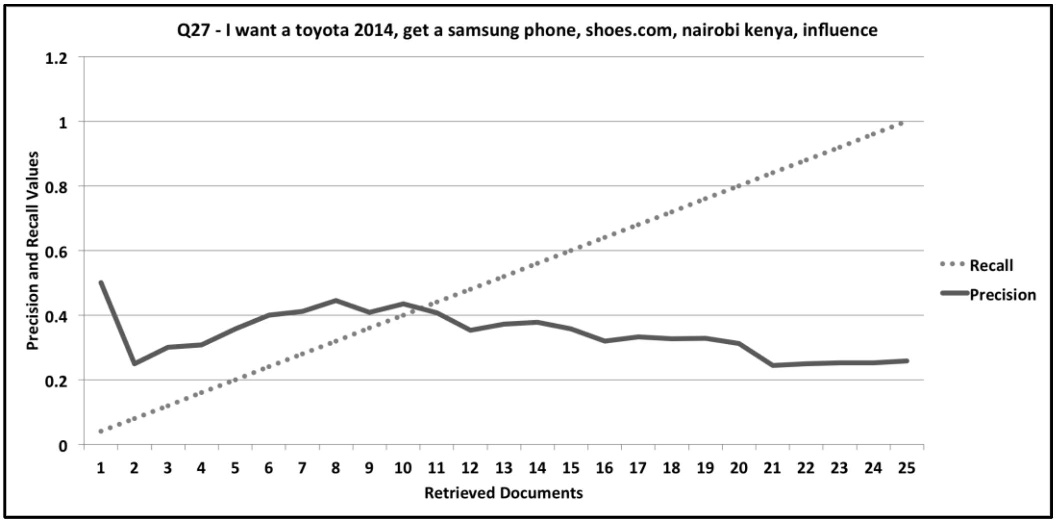


Fig. 8. Recall and Precision @ *k* for query Q22

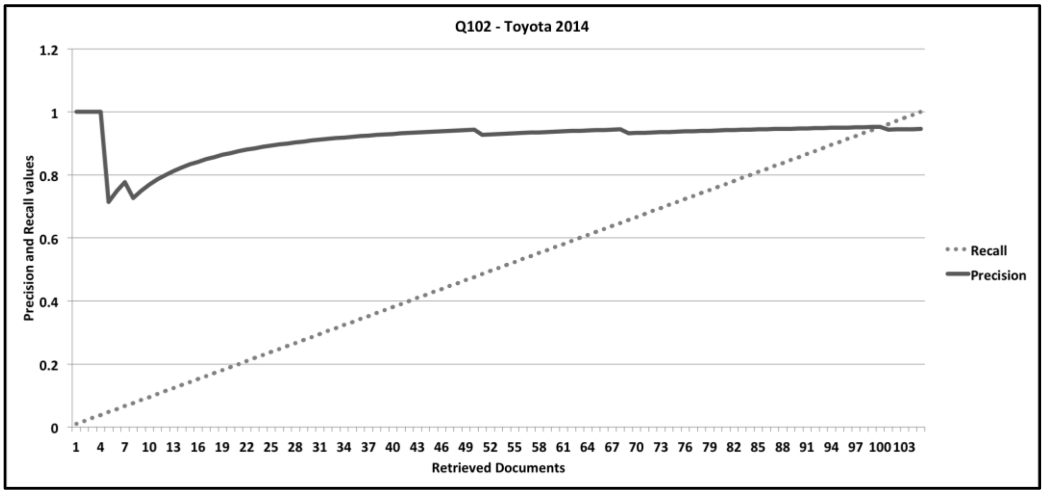
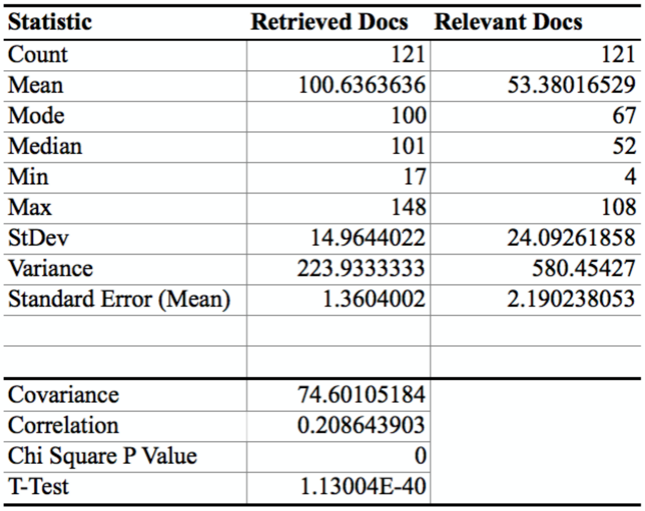


Fig. 9. Recall and Precision @ *k* for query Q102

Considering the average precision metric, the values vary between 0 and 1, where 1 as a signal for better performance. The queries *Q12* and *Q27* outcomes show low average precision values but effective obfuscation, which can be attributed to the level of permutation. This is additionally seen in Figure 4, where the retrieved articles for *Q12* and *Q27* are 148 and 104 correspondingly, yet the relevant articles – articles that comprised the keyword "*Toyota*", for *Q12* and *Q27* recorded values 31 and 27 in that order. It could be reasoned that this is a signal of good obfuscation due to a robust permutation set being chosen. However, this only brings the question of usability to the forefront. Retrieving articles that are of no meaning in regards to what the user is querying, does grant some level of user intent privacy yet the usefulness of such outcomes remains challenging.

TABLE I. CORRELATION – RETRIEVED VS. RELEVANT



Statistic	Retrieved Docs	Relevant Docs
Count	121	121
Mean	100.6363636	53.38016529
Mode	100	67
Median	101	52
Min	17	4
Max	148	108
StDev	14.9644022	24.09261858
Variance	223.9333333	580.45427
Standard Error (Mean)	1.3604002	2.190238053
Covariance	74.60105184	
Correlation	0.208643903	
Chi Square P Value	0	
T-Test	1.13004E-40	

The permutation of web search query types: In the case of the permutation of web search query types heuristic, the aim is to take advantage of the distortion caused by the different permutation of the query keywords to conceal user intent but at the expense of disregarding useful retrieved links associated with the original "*Toyota*" keyword. Yet still, one significant aspect that our preliminary results reveal is that various permutations of the web search query type have an effect on retrieved relevant results. For example the query *Q12* is of the formation, "*Acquire Toyota 2014 get Samsung Obtain shoes.com western civilization*"; *Q27* is of the formation, "*I want a Toyota, get Samsung phone, shoes.com, Nairobi Kenya, influence*". *Q12* and *Q27* do yield the least values in regards to relevant documents retrieved and the average precision. Though the main intended key search term by the user, "*Toyota*", is in the first segment of the query text, the search engine appear incapable of deciphering the obfuscation and retrieve the most relevant documents associated to "*Toyota*". Yet this result continues beyond queries *Q12* and *Q27*. Meanwhile queries *Q22* and *Q102* returned results of relevant retrieved documents at 74 out of 105 and 105 out of 111 correspondingly. The average precision results for queries *Q22* and *Q102* respectively returned values 0.735 and 0.916. The permutation of the web search query type NIT and IP were used for queries *Q22* and *Q120*, while NITP and NITPL were used for *Q12* and *Q27* in that order. The query *Q22* has the formation "*Samsung Obtain shoes.com Barack Obama Toyota*", the keyword "*Toyota*" is positioned at the end of the query text. Nevertheless, for query *Q102*, the formation is instinctive, "*Toyota 2014*". Queries *Q22* and *Q102* yield the highest retrieved relevant documents and uppermost average precision values, with query *Q102* returning a significant average precision value of 0.916. Nonetheless, while queries *Q22* and *Q102* yield some of the best results in terms of retrieved relevant documents, such yields come at the expense of privacy – signifying low user intent privacy. Finally, the process of the query formation indubitably plays an essential part in the retrieval of relevant documents, making it a challenge as to which query formation or permutation of the query type yields the most optimal results in context of privacy and usability.

V. CONCLUSION AND FUTURE WORK

Preliminary experimental outcomes from this study indicate that the permutation of web search query types heuristic might be an effective means of providing obfuscation and privacy for user intent. Preliminary outcomes indicate that web search query and intent privacy might be achievable from the user side using non-cryptographic means such as the suggested permutation of web search query types heuristic, without the requirement of third parties. However, the privacy versus usability challenge remains suggesting that more study is necessary on the formation of obfuscated web search queries in the context of both optimal privacy and usability. The practice of retrieving relevant documents yet preserving satisfactory levels of privacy also remain a challenge and necessitates a more comprehensive study that takes into consideration all aspects of security. Another issue is that trade-offs would be obligatory between user privacy and usability needs; yet finding optimal trade-off areas is another challenge. For instance, web search queries with lower levels of obfuscation might yield higher relevant retrieved documents but at the expense of revealing user intent and undermined web search privacy. Finally, it is essential that web search users and privacy custodians give considerable attention to the resource capabilities of search engines such as, computational, storage, query disambiguation, and semantic processing when employing any query obfuscation techniques. Search engines continuously analyze web search queries with the objective of decoding user intent by separating dummy from real queries. Consequently, a constant revision of search query obfuscation and permutation of query search types techniques would have to be done to prevent classification attacks, a topic to be investigated as part of our future. We intend to consider application of the suggested heuristic on big data and enterprise systems with multiple logged in users, simulating a real world scenario.

References

[1] M. Gordon and P. Pathak, “Finding information on the World Wide Web : the retrieval effectiveness of search engines,” vol. 35, 1999.

[2] S. Brin and L. Page, “The anatomy of a large-scale hypertextual Web search engine,” Comput. Networks ISDN Syst., vol. 30, no. 1–7, pp. 107–117, Apr. 1998.

[3] R. Ozcan, I. S. Altingovde, B. B. Cambazoglu, F. P. Junqueira, and Ö. Ulusoy, “A five-level static cache architecture for web search engines,” Inf. Process. Manag., 2011.

[4] R. Baeza-Yates and B. Ribeiro-Neto, Modern Information Retrieval. New York: ACM Press, 1999.

[5] A. Broder, “A taxonomy of web search,” ACM SIGIR Forum, vol. 36, no. 2, p. 3, Sep. 2002.

[6] M. Z. Ullah and M. Aono., “Query subtopic mining for search result diversification,” in IEEE International Conference of Advanced Informatics: Concept, Theory and Application (ICAICTA), 2014, no. 1, pp. 309–314.

[7] J. Zamora, M. Mendoza, and H. Allende., “Query intent detection based on query log mining,” J. web Eng., vol. 13, no. 1–2, pp. 24–52, 2014.

[8] K. Mivule, “Utilizing Noise Addition for Data Privacy , an Overview,” in Proceedings of the International Conference on Information and Knowledge Engineering (IKE 2012), 2012, pp. 65–71.

[9] A. Viejo, J. Castell, and O. Bernad, “Single-Party Private Web Search,” pp. 1–8, 2012.

[10] C. Mangold, “A survey and classification of semantic search approaches,” Int. J. Metadata, Semant. Ontol., vol. 2, no. 1, p. 23, 2007.

[11] V. Dang, W. B. Croft, and B. Croft, “Query reformulation using anchor text,” in Proceedings of the third ACM international conference on Web search and data mining., 2010, pp. 41–50.

[12] Y. Song, D. Zhou, and L. He, “Query suggestion by constructing term-transition graphs,” in Proceedings of the fifth ACM international conference on Web search and data mining - WSDM ’12, 2012, pp. 353–362.

[13] M. Gupta and M. Bendersky, “Information Retrieval with Verbose Queries,” in Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2015, pp. 1121–1124.

[14] L. Bing, W. Lam, T.-L. Wong, and S. Jameel, “Web query reformulation via joint modeling of latent topic dependency and term context.,” ACM Trans. Inf. Syst., vol. 33, no. 2, p. 6., 2015.

[15] A. Turpin and F. Scholer, “User Performance versus Precision Measures for Simple Search Tasks,” in Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2006, pp. 11--18.

[16] M. Murugesan, “Providing Privacy through Plausibly Deniable Search ∗,” pp. 768–779.

[17] E. W. and J. Brawner, Discrete Mathematics for Teachers. IAP, 2010.

[18] N. Matsuda and H. Takeuchi, “Do Heavy and Light Users Differ in the Web-Page Viewing Patterns ? Analysis of Their Eye-Tracking Records by Heat Maps and Networks of Transitions,” Int. J. Comput. Inf. Syst. Ind. Manag. Appl., vol. 4, pp. 109–120, 2012.

[19] T. Joachims, L. Granka, B. Pan, H. Hembrooke, and G. Gay, “Accurately interpreting clickthrough data as implicit feedback,” in Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 154–161.

[20] Y. Matsuda, H. Uwano, M. Ohira, and K. Matsumoto, “An Analysis of Eye Movements during Browsing,” Human-Computer Interact. New Trends, pp. 121–130, 2009.

[21] B. Pan, H. a Hembrooke, G. K. Gay, L. a Granka, M. K. Feusner, and J. K. Newman, “The determinants of web page viewing behavior: an eye-tracking study,” in Proceedings of the ETRA ’04 Symposium on Eye Tracking Research and Applications, 2004, pp. 147–154.