Improving Classification Based Off-topic Search Detection via Category Relationships

Alana Platt
Computer Science Department
Illinois Institute of Technology
Chicago, Illinois, U.S.A.
platt@ir.iit.edu

Saket S. R. Mengle
Computer Science Department
Illinois Institute of Technology
Chicago, Illinois, U.S.A.
saket@ir.iit.edu

Nazli Goharian
Computer Science Department
Illinois Institute of Technology
Chicago, Illinois, U.S.A.
nazli@ir.iit.edu

ABSTRACT

The illegitimate access of documents by insiders (also known as *off-topic* search) is an increasingly prevalent and largely ignored problem. We propose an approach that uses text classification for off-topic search detection. Our empirical results indicate that off-topic search detection effectiveness improves by considering only a subset of documents that are retrieved for a given user query. Furthermore, we also show that the effectiveness of off-topic search detection improves by using the ontological information of document categories. Our empirical results demonstrate that utilizing sibling relationship information and relationships derived from misclassification information statistically significantly improves the results over the baseline in most cases.

Categories and Subject Descriptors

K.6.5 [**Management of Computing and Information Systems**]: Security and Protection - Unauthorized Access

General Terms

Algorithms, Experimentation, Security

Keywords

Insider misuse, user profiles, need-to-know, ontology

1. INTRODUCTION

In this work, we explore the use of text classification to detect deviations in users' search patterns. Oftentimes this indicates malicious information access. This problem is known as *off-topic search* or *insider misuse*. Although employees of an organization might have the authorization to access certain documents, it does not give them the privilege access all documents. A recent survey [12] indicates that *insider misuse* is a more prevalent form of computer crime than computer viruses. Although organizations wish to protect sensitive records from nefarious insiders, restricting access rights to a document may not always be desirable. (For example, users may have cause to occasionally look at documents in an area that is not related to their direct areas of interest.) Content-based off-topic search detection systems are used to gather evidence that fraudulent information access has occurred. Furthermore, the record of off-topic searches can be used to make informed decisions about the subsequent user's queries, such as in [11].

We explore the use of text classification to detect off-topic search and favorably demonstrate the detection effectiveness on various types of user profiles. Our empirical results indicate that using information about relationships among categories, and using only a subset of the retrieved documents significantly improves the effectiveness of our system in most cases.

2. PRIOR WORK

The objective of off-topic search detection is to determine if the user's search is within the scope of his or her predefined area of interest. One common strategy for off-topic search detection is a two phase process: building a user profile that models the user's legitimate behavior, and then comparing the current behavior to the profile [1, 5, 17]. After a threshold of allowable off-topic searches is exceeded (as defined by the organization), the system generates a warning. We, too, use a two phase strategy in our work.

In the first phase, a profile is built based on the legitimate scope of work in the organization for each user. User profiles are of interest to many areas of research, including personalization [15, 16, 18], peer-to-peer information retrieval [9], malicious data access detection [4], and off-topic search detection [2]. The user profile may be represented in various ways, including (but not limited to) bag of words derived from persumably valid user queries [5, 6], bag of words that map to the legitimate user behavior or organizational tasks [7, 11], or topics (categories) mapping to the legitimate user interests [13, 14]. Similarly, in this work we assume that the user profiles consist of one or more topics based on user's interest.

In the second phase, a query issued by a user is marked as on-topic or off-topic based on the user's profile. The user issued queries (as in [5, 6]) or documents retrieved by the query (as in [7, 11] and our current approach) are compared to the user's profile. The similarity between the query and the user's profile or the retrieved documents and the user's profile is measured. As the result a possible off-topic search may be detected.

Some state-of-the-art approaches for detecting off-topic search that use this two-phase approach are as follows. [3] and [8] use clustering to group users' web search results to form user profiles with which they perform anomaly detection. Retrieved documents are clustered in [7] for detecting off-topic search and it is further refined in [11], where segments of queries are grouped

SAC'09, March 8-12, 2009, Honolulu, Hawaii, U.S.A.

into windows, and their similarities to each other are computed. If there are sufficient similarities among the queries, then the system may opt to modify its off-topic evaluation of the query.

Among other efforts in off-topic search detection is an ontology-based approach [1]. The access to a document is considered illegitimate if a user's profile does not have a semantic association with the documents retrieved by the search. Authors in [6] compare the user profile terms with the query terms and relevance feedback terms generated by that query. In [5], using classification, the authors refine the work presented in [6] to demonstrate that a subset of features (terms) can be used and still achieve equivalent effectiveness. Unlike in [5, 6] that use query terms (and relevance feedback terms) to determine if off-topic search has occurred, our effort classifies retrieved documents and compares those classifications to the user's profile.

In [13], the authors propose a multi-agent system for implementing "need-to-know" access policies. The users issue requests to an authorization system for accesing confidential documents. The authorization system then compares the user's request to the profile of the user, which is represented by one category. Before retrieving a document, the system classifies the document. If the predicted category matches the user's profile, the request for the document is approved. The authors further refine their approach in [14]. When training the text classifier, more cost is associated with wrongly allowing access to a "confidential" document than denying access to an innoculous document. Unlike [13], we aim to identify off-topic queries rather than off-topic documents. We assign categories to the retrieved documents and based on voting identify the majority category. This selected category is then assigned to the query. The closest comparison between [13] and our approach would be achieved by implementing [13] in accordance with a search engine. Thus, when the user's query is issued to the search engine, each of the top retrieved documents is treated as if the user attempted to access it. Specifically, each of the top retrieved documents is classified and compared to the user's profile. To determine if off-topic search has occurred, the most frequently predicted class label is identified, and compared to the profile. We adopt this method as our baseline approach.

In [17], the authors propose a fusion-based (hybrid) approach to detect off-topic search. By fusing role based monitoring methods, social network analysis, and semantic content analysis, an approach for detecting inappropriate information exchange is developed. In a continuation of the effort described in [17], a natural language processing approach involving entity tagging for detecting off-topic search is presented in [19], and the authors favorably compare their approach against a described "bag of words" solution.

3. METHODOLOGY

Our system evaluates off-topic search based upon the similarity of the documents retrieved by a query to the user's profile. However, one cannot assume that a user's query that is genuinely on-topic only retrieves documents that are on-topic. It was demonstrated that using a subset of retrieved documents through clustering query results improves off-topic search detection [7]. In this work, we are interested in using supervised learning and group sets of documents based on classifier predictions of their categories. We call these groups of documents "segments". After distributing the retrieved documents into segments, a subset of segments are selected to be used in off-topic search evaluation, while the rest are discarded.

Additionally, situations may arise where the retrieved documents belong to a category, which is outside of the user's profile, but are tangentially related to a legitimate interest of that user. To address this issue, we check if the predicted category is *related* to any categories that are in the user's profile. The existence of such relationships is considered in the evaluation of a query as an *on-topic* search. We provide a detailed explanation of the detection process as follows.

Step 1: Document Retrieval

The top 100 relevant documents to each user's queries are retrieved. We used the Google API to issue queries on the web and download the top 100 documents.

Step 2: Text Classification

Each retrieved document is labeled with a category as determined by the text classifier. Our collection contains single labeled documents.

Step 3: Document Selection

A subset of retrieved documents that are deemed to be the most representative of the user's intent is selected. We vary the number of retrieved documents that we keep in this process as: *d* = 25, 50, 75 or 100. (All top 100 retrieved documents are used only in our baseline.) The documents are selected by one of the three following methods:

Top Retrieved Documents

In the *top retrieved documents* selection algorithm, we select documents based on the search engine's rankings. We assume that the higher ranked documents are more closely related to the query. Thus we vary the number of selected documents from the top d =25, 50, and 75 retrieved documents.

Largest Segments

Even though a document is highly ranked in the list of retrieved results, it may be a false positive. Thus, for the *largest segments* document selection algorithm, we employ segmentation. The description of the technique follows:

When classifying the retrieved documents, one category is assigned to each document. However, we also have information on how likely it is that each document belongs to the other categories. After assigning each document to the most probable category, we record the next two most likely categories as candidates. We then segment all documents based on the top three predicted categories. (We empirically determined three categories as the optimal number of categories based on our datasets.) We believe that when there are many retrieved documents that have the same top three categories, it is likely that this combination of categories are closely related to the user's intent. We rank the segments based on the number of retrieved documents that belong to the segment. Then, we add the highest ranked segments to our subset of documents to be used in evaluation, and keep a count of the number of selected documents. We select the segments until there are at least *d* documents in the subset.

Least Ambiguous Segments

In *least ambiguous segments*, rather than using the size of the segment as our selection criterion, we consider the largest average number of *keywords*. We employ the same document segmentation method as in the *largest segments* approach. While performing text classification on the documents, we keep track of the number of *keywords* [10] in each document. *Keywords* are terms that are found to consistently occur in documents of a category, and are not frequently found in documents of other categories.

We compute the average number of *keywords* for each segment. We recognize that some documents with many *keywords* may have *keywords* that belong also to the other categories. However, our documents have one primary category. This indicates that the majority of the keywords most likely are from that main category. Following this intuition, one would expect that segments with a higher average number of *keywords* are more likely to contain documents that have been categorized correctly. Thus, we rank the segments from the highest average number of *keywords* to the lowest average number of *keywords*. Again, we select documents in descending order of average *keywords*, and keep track of the documents that are selected. We add documents to our subset of selected documents until reaching a threshold.

Step 4: Off-Topic Search Evaluation

In this step, we identify the most commonly predicted category among our selected documents. This category is assigned to the query, and compared to the user's profile (section 4.3). If there is a match between the legitimate interests of the user (represented by the profile) and the category assigned to the majority of the selected documents, we deem the query to be on-topic for the user. Otherwise, we consider the query *off-topic*. At this point, we either designate the query as off-topic for the user, or opt for *warning level reevaluation,* as step 5.

Step 5: Ontological Evaluation

In our fifth step, we use the ontology information to reevaluate the assigned warning levels. In an off-topic search detection process, it is crucial to lower the false positive rate. Off-topic queries are relatively rare compared to the number of legitimate queries, and thus a high false alarm rate is troublesome for both the user and any organization. We apply information from the category ontology to decrease the incidence of false positives. Due to the relationships among categories, it is shown that the related categories are frequently misclassified as each other. We leverage the category relationships to decrease the instances of false positives, as explained below:

Sibling Information

In the taxonomy of document categories, the categories that share a parent are considered to be related to one another. We call these categories *siblings*. Since *siblings* are a part of a common, over-arching category, it is likely that they share many common terms.

The category with the most number of the documents is identified along with its *siblings*. We then compare all of the *siblings* to the user's profile. If any of the *sibling* categories are on-topic for the user, we label the query as on-topic. If none of the *sibling* categories are on-topic for a user, we continue to assume the search was off-topic.

Misclassification Information

In *sibling information*, we utilize information from a defined ontology to decrease false alarms. However, some datasets do not have a defined ontology. Hence, we explore an automatic method to discover relationships among categories [10]. The *misclassification information* generated during the process of text classification is used to predict relationships among categories. We are interested in leveraging these category relationships to lower the incidence of false alarms. Thus, if a related category exists in the user's profile we change the prediction to on-topic.

Once we complete the five-step process, we compare our system's off-topic search prediction to the ground-truth values. For each of our datasets, three computer science students have assessed the relevancy of each query to each category in the dataset. A query is considered *on-topic* if the human assessors evaluated it as relevant to any of the categories in the user's profile. Otherwise, the query is evaluated as *off-topic*. These ground-truth relevancy judgments are compared to our system's judgments, as explained in detail in section 5.

4. EXPERIMENTAL SETUP

4.1 Datasets

We use 20 Newsgroup and a subset of the Open Directory Project datasets for training the text classifier. The 20 News Groups (20NG) dataset[1] is already divided into a priori known categories. It consists of 20,000 documents categorized into 20 different categories. Each category has 1000 documents assigned to it. ODP46 is a subset of the Open Directory Project[2] tree. This dataset contains 46 categories. We select 500 documents per category in ODP46 dataset. We do not list the categories in ODP46 subset to maintain brevity.

4.2 Queries

For each dataset, we construct two sets of 50 queries: one set of long queries (6-10 terms), and one set of short queries (1-5 terms). Each query corresponds to at least one category in the ontology. The relevance of a query to each of the categories is determined by human evaluators. These queries are then issued to the Google search engine, and the top 100 retrieved results are downloaded.

The documents from 20NG and ODP46 are each used to train the text classifier. The testing set consists of the retrieved documents from our two sets of 50 queries.

4.3 Profiles

Since an organization has authorized users with a wide range of tasks, it is necessary to simulate various possible profiles. Thus, for a dataset with *N* categories, we built profiles that range from 1 to *N-1* categories on-topic.

For each of the two datasets described in section 4.1, we constructed a set of profiles. Within a profile, each category is enumerated and assigned positive or negative relevancy for that user. For each of our two datasets, we created sets of profiles that contain *N*-1 types of profiles, where *N* is the number of

[1] 20 News Groups dataset. (http://people.csail.mit.edu/ jrennie/20Newsgroups.)

[2] Open Directory Project (http://dmoz.org).

categories in a given dataset. Each type of profile corresponds to a different number of categories that the user is legitimately allowed to view – from 1 category to *N-1* categories. Profiles that consist of very few on-topic categories model situations where the legitimate interest of the user is narrowly defined. However, it may be desirable to allow users less restrictive user profiles (for example, individuals whose legitimate interests are inherently broad, or a degree of curiosity into other categories is allowable). These less restrictive user profiles are modeled by profiles with more on-topic categories. Thus, we create *10*(N-1)* profiles per dataset (10 profiles of each type).

5. EVALUATION

Five levels of off-topic search are considered, which are determined by obtaining the user's search deviation from a valid profile. The levels are "off-topic" (L5), "probably off-topic" (L4), "undetermined" (L3), "probably on-topic" (L2) and "on-topic" (L1). The ranking levels generated by the detection system, i.e., predicted level (columns in table 1), are compared against the actual level (rows in table 1), as determined by human evaluators.

In contingency matrix (table 1), true positive (*TP)* is the number of off-topic queries that the system correctly identified as off-topic. False negative (*FN)* is the number of off-topic search queries that were not identified as such. False positive (*FP)* is the number of queries the system wrongly identified as off-topic. True negative (*TN*) is the number of queries that were on-topic and the system correctly identified as such.

Table 1. Contingency matrix – predicted level vs. actual level

Stringent		Predicted				
		L5	L4	L3	L2	L1
Actual	L5	TP	TP	FN	FN	FN
	L4	TP	TP	FN	FN	FN
	L3	FP	FP	TN	TN	TN
	L2	FP	FP	TN	TN	TN
	L1	FP	FP	TN	TN	TN

We evaluate the effectiveness of our detection system using the standard metrics of recall, precision, and F1-measure. Recall measures the rate of detection, while precision is defined as the ratio of the cases detected correctly as off-topic to the total of the true and false detections. Similar to [11], we use true negatives to calculate precision and recall, as the objective of our method is to decrease false positives and increase true negatives. The F1-measure is a harmonic mean of precision and recall. False alarm rate is defined as the ratio of false positive to the total number of predictions. The formulae for calculating precision, recall, F1-measure, and false alarm rate are as follows:

$$Recall=\frac{TP+TN}{TP+FN+TN}$$

$$Precision=\frac{TP+TN}{TP+FP+TN}$$

$$F1-Measure=\frac{2*Recall*Precision}{Recall+Precision}$$

$$False\,Alarm\,Rate=\frac{FP}{TP+FP+TN}$$

6. RESULTS

In this section, we present the effects of various methods for improving the effectiveness of the off-topic search detection task. We compare our methods with the scenario where no document selection algorithms or ontology information is used.

6.1 Effects of Number of On-Topic Categories in a Profile

Figures 1-4 present the F1-measure for profiles of various sizes. The y-axis is F1 measure and the x-axis corresponds to different ontological evaluation approaches and different types of profiles. F1 values are also given below the figures. We plot a polynomial trend line to observe the effects of different methods for different profile sizes. The trend indicates that the F1-measure is higher when the number of categories is either small or large. When a profile consists of almost all the categories as on-topic, the likelihood of false negative is low as the majority of the predicted categories are in that profile. Conversely, when a profile consists of very few categories as on-topic, very few false positives are generated. When the number of on-topic categories and off-topic categories with respect to a given profile is equivalent, there is a higher likelihood of generating both false positives and false negatives. Hence, the F1-measure in such cases is lower (an average of 20.83%) than profiles with only one category on topic or one category off-topic for the user.

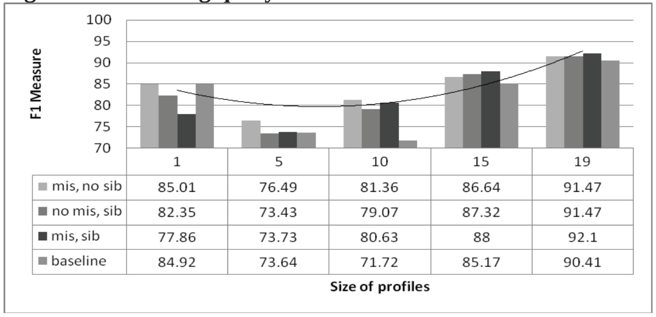
6.2 Document Selection Algorithms

For 20NG dataset, the *top retrieved documents* selection algorithm outperforms other document selection algorithms when less than half of the categories deemed as on-topic for the user. Since these documents are most relevant to the query, the *top retrieved documents* selection algorithm performs well. The only exception to this trend that should be noted is for cases where profiles have only one category that is on-topic. However, for profiles that have at least half the categories as on-topic for the user, segmentation techniques – particularly *least ambiguous segments* tend to improve performance. The *least ambiguous segments* technique favors segments that have documents with many keywords. Documents with many keywords have a high probability of being classified correctly. These document selection techniques improved F1-measure by an average of 5.59%. For ODP46 dataset, the *top retrieved documents* selection algorithm consistently outperforms the other methods. The *largest segments* selection algorithm fails to significantly improve the F1-measure for all types of user profiles for both the 20NG and ODP46 datasets.

6.3 Effects of Sibling Information

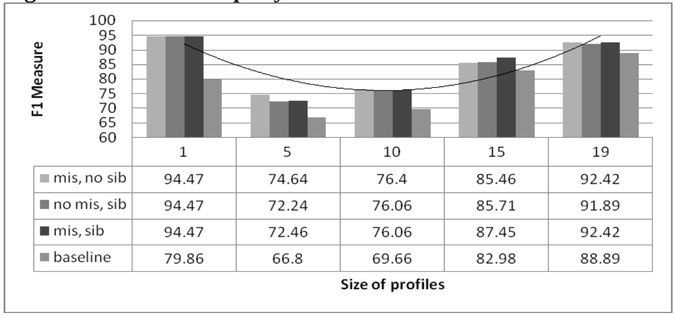
Our results indicate that using *sibling* categories statistically significantly improves the F1-measure by up to 14.61% on the 20NG dataset and by 10.33% on the ODP46 dataset for the short queries. However, our results indicate that using *sibling* categories does not improve F1-measure for the long queries. Long queries are more detailed and likely to retrieve very precise documents. Our results indicate that baseline for the long queries perform statistically significantly better (an average of 5.7% higher F1-

Figure 1. 20NG long query results



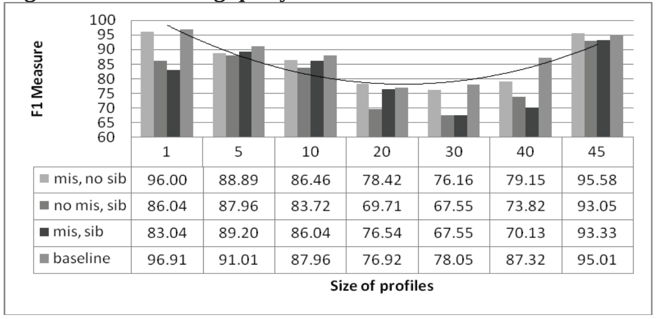
	1	5	10	15	19
mis, no sib	85.01	76.49	81.36	86.64	91.47
no mis, sib	82.35	73.43	79.07	87.32	91.47
mis, sib	77.86	73.73	80.63	88	92.1
baseline	84.92	73.64	71.72	85.17	90.41

Figure 2. 20NG short query results



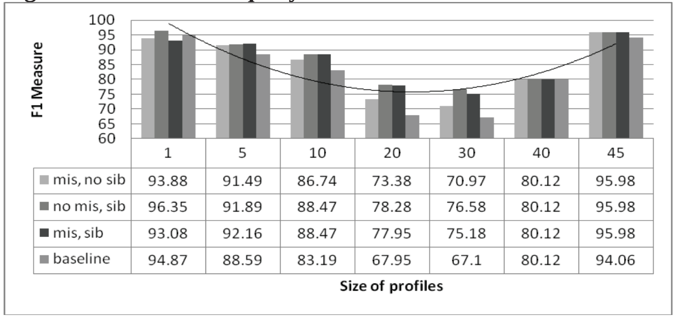
	1	5	10	15	19
mis, no sib	94.47	74.64	76.4	85.46	92.42
no mis, sib	94.47	72.24	76.06	85.71	91.89
mis, sib	94.47	72.46	76.06	87.45	92.42
baseline	79.86	66.8	69.66	82.98	88.89

Figure 3. ODP46 long query results



	1	5	10	20	30	40	45
mis, no sib	96.00	88.89	86.46	78.42	76.16	79.15	95.58
no mis, sib	86.04	87.96	83.72	69.71	67.55	73.82	93.05
mis, sib	83.04	89.20	86.04	76.54	67.55	70.13	93.33
baseline	96.91	91.01	87.96	76.92	78.05	87.32	95.01

Figure 4. ODP46 short query results



	1	5	10	20	30	40	45
mis, no sib	93.88	91.49	86.74	73.38	70.97	80.12	95.98
no mis, sib	96.35	91.89	88.47	78.28	76.58	80.12	95.98
mis, sib	93.08	92.16	88.47	77.95	75.18	80.12	95.98
baseline	94.87	88.59	83.19	67.95	67.1	80.12	94.06

measure) than the short queries as the retrieved documents are more specific to a query. Hence, if warning levels are lowered,

more false negatives are generated for long queries and consequently F1-measure decreases.

6.4 Effects of Misclassification Information

As shown in figures 1 and 2, for both short and long queries with the 20NG dataset, for the profiles that have at least five on-topic categories, *misclassification information* statistically significantly (90% confidence) improves F1-measure (an average of 6.34% increase). This trend is consistent regardless of using *sibling information*.

For profiles with less than five categories on-topic, *misclassification information* does not improve the detection. Since there are few categories that are on-topic for such profiles, instances of true negatives are rare. As such, when a query's warning level is lowered, it is much more likely to have an *undetected misuse* rather than correcting a *false alarm*. As shown in figures 3 and 4, *misclassification information* does not improve results in the ODP46 dataset. One should note that much more relationships were identified between categories in the ODP46 dataset than in the 20NG dataset. This finding was reported in [10] and was also validated by human evaluators. As the precision for predicting the relationships in the 20NG dataset (59.0%) is higher than in the ODP46 dataset (31.8%), the probability of wrongly lowering a *true positive* is higher in the OPD46 dataset. Thus, *misclassification information* is only useful in an ontology where the number of relationships between categories is relatively low.

We also evaluated combining both *sibling* relationship information and relationships derived from *misclassification information*. Our results indicate that though this hybrid method improves the F1-measure in some cases, the improvements are not statistically significant.

Table 2. Average decrease in false alarm rate

Query type	Warning level reevaluation	20NG	ODP46
		Avg. FA Rate	Avg. FA Rate
Short	Baseline	32.26	35.21
	Misclass	29.43	34.14
	Sibling	31.31	33.03
	Misclass + Sibling	**28.31**	**33.00**
Long	Baseline	16.79	21.87
	Misclass	12.07	20.96
	Sibling	13.81	21.05
	Misclass + sibling	**9.96**	**20.39**

6.5 Effect of Using Ontological Information on False Alarm Rate

As shown in Table 2, using information from the category ontology lowers false alarm rates for all configurations. Using *misclassification information*, *sibling information,* or a combination of the two (labeled "Misclass + Sibling" in Table 2) each statistically significantly reduces the false alarm rate (~ 6% net gain) with a confidence level of 95%. Our baseline is the average false alarm rate while using neither *misclassification information* nor *sibling information.* Our improvements are statistically significant (95% confidence level) for both the long and the short queries, and across all configurations and profiles.

Using both *misclassification information* and *sibling information* ("Misclass + Sibling") improves the false alarm rate by 3.95% net gain for short queries and 6.83% net gain for long queries on 20NG dataset. Similarly, using both *misclassification information* and *sibling information* ("Misclass + Sibling") improves the false alarm rate by 2.21% net gain for short queries and 1.48% net gain for long queries on ODP46 dataset. The relationships found between the categories in OPD dataset are less strong than between categories in 20NG dataset, as explained in section 6.4, thus, leading to a higher improvement in the false alarm rate in 20NG than in ODP.

7. CONCLUSION

We described and evaluated a classification-based approach for off-topic search detection. We demonstrated the applicability of this approach for various types of users with the profiles ranging from one search subject to many subjects.

We demonstrated that by performing text classification on only a subset of the retrieved results, we are able to achieve higher F1-measure and lower false alarm rates. We demonstrated that using *sibling* categories statistically significantly improves the baseline for short queries with a confidence of 95%, up to 14.61%. The category relationships derived from *misclassification information* statistically significantly (90% confidence) improves the F1-measure over the baseline for both long and short queries on 20NG dataset (an average of 6.34%). Furthermore, our analysis showed that using specific profiles (fewer categories) or very broad profiles (almost all categories) achieves the highest effectiveness in terms of F1-measure. The effectiveness decreases in situations where the number of on-topic and off-topic categories in a profile is equal. Using ontological information showed a reduction in the false alarm rate on 20NG dataset (by up to 6.83% net gain) and on ODP46 dataset (by up to 2.21% net gain).

In our future work, we plan to use the sequence of user queries, as in [11], to evaluate the effectiveness of our ontology-based detection system.

8. REFERENCES

[1] B. Aleman-Meza, P. Burns, M. Eavenson, D. Palaniswami, A. Sheth. An ontological approach to the document access problem of insider threat. *IEEE Intl. Conf. on Intelligence and Security Info. (ISI)*, May 2005.

[2] R. Cathey, L. Ma, N. Goharian, D. Grossman. Misuse Detection for Information Retrieval Systems. *ACM Conference on Info. And Knowledge Management (CIKM),* Nov. 2003.

[3] Y. Elovici, et al. Content-based detection of terrorists browsing the web using an advanced terror detection system (ATDS). *IEEE Intl. Conf. on Intelligence and Security Info. (ISI)*, May 2005.

[4] J. Fonseca, M. Vieira, H. Maderia. Online Detection of Malicious Data Access Using DBMS Auditing. *ACM Symp. on Applied Computing (SAC)*, March 2008.

[5] N. Goharian, L. Ma, Off-Topic Access Detection In Information Systems, *ACM Conf. on Info. and Knowledge Management (CIKM)*, Nov. 2005.

[6] N. Goharian and L. Ma. Query length impact on misuse detection in information retrieval systems. A*CM Symp. on Applied Computing (SAC)*, March 2005.

[7] N. Goharian and A. Platt. DOTS: Detection of Off-Topic Search Via Result Clustering. *IEEE Intl. Conf. on Intelligence and Security Info. (ISI),* May 2007.

[8] M. Last, et al. Content-based methodology for anomaly detection on the Web. *Lecture Notes in Computer Science, Intl. Atlantic Web Intelligence Conf.*, May 2003.

[9] J. Lu, J. Callan. User modeling for full-text federated search in Peer-to-Peer networks. *ACM Conf. on Research and Development in Info. Retrieval (SIGIR),* Aug. 2006.

[10] Mengle, N. Goharian, A. Platt. Discovering Relationships among Categories using Misclassification Information. A*CM Symp. on Applied Computing (SAC)*, March 2008.

[11] A. Platt, N. Goharian, S. Mengle. Using User Query Sequence to Detect Off-Topic Search. *ACM 22nd Symp. on Applied Computing (SAC),* March 2007.

[12] R. Richardson. 2007 CSI Computer Crime and Security Survey. (http://i.cmpnet.com/v2.gocsi.com/pdf/CSISurvey 2007.pdf). 2007

[13] Y. Seo, J. Giampapa, and K. Sycara. A multi-agent system for enforcing Need-To-Know security policies. *Intl. Conf. on Auto. Agents and Multi Agent Systems Workshop on Agent Oriented Info. Systems (AOIS-04)*, July 2004.

[14] Y. Seo and K. Sycara. Cost-Sensitive Access Control for Illegitimate Confidential Access by Insiders. *IEEE Intl. Conf. on Intelligence and Security Info. (ISI)*, May 2006.

[15] X. Shen, B. Tan, C. Zhai. Context sensitive information retrieval using implicit feedback. *ACM Conf. on Research and Development in Info. Retrieval (SIGIR)*, Aug. 2005.

[16] K. Sugiyama, K. Hatano, M. Yoshikawa. Adaptive web search based on user profile constructed without any effort without users. *Intl. World Wide Web Conf. (WWW)*, May 2004.

[17] S. Symonenko, L. Liddy, O. Yilmazel, R. Del Zoppo, E. Brown, M. Downey. Semantic analysis for monitoring insider threats. *IEEE Intl. Conf. on Intelligence and Security Info. (ISI)*, May 2004.

[18] J. Teevan, S. Dumais, E. Horvitz. Personalizing search via automated analysis of interests and activities. *ACM Conf. on Research and Development in Info. Retrieval (SIGIR)*, Aug. 2005.

[19] O. Yilmazel, S. Symonenko, N. Balasubramanian, E. Liddy. Leveraging One-Class SVM and Semantic Analysis to Detect Anomalous Content. *IEEE Intl. Conf. on Intelligence and Security Info. (ISI)*, May 2005.