Using Recall and Elimination Terms in Separate Runs for High Volume Document Sorting

Harvey Hyman, Terry Sincich, Rick Will, and Warren Fridy III

***Abstract*—This paper reports on a study undertaken to explore the problem of volume in searching large scale digital collections. An experiment is conducted using elimination terms as a method to reduce the number of non-relevant documents in the information retrieval (IR) result. The goal is to provide insight into how elimination terms can be used as a sorting method to reduce volume. The results of the experiment demonstrate that modifying the search structure with an elimination term component can significantly reduce the number of non-relevant documents in the retrieval set to address the problem of high volume in electronic document sorting and searching tasks.**

***Index Terms*—Information retrieval, knowledge discovery, search methods, document sorting.**

I. Introduction

Volume is one of the "three Vs" associated with large scale digital information (volume, variety and veracity), [1]. In this paper we focus on the volume problem by asking the question: How can we apply sorting methods to reduce the volume of electronic documents for sorting and searching tasks?

A current trending problem is that, given the large volumes of information contained in electronic stores, search tools need to support the retrieval of relevant documents from large collections without producing too many non-relevant documents [2]-[4].

A relevant document is defined as a document that meets the user's needs (see VanRijsbergen, 1979). A non-relevant document is defined as a document that does not meet the user's needs. To address the problem of high volume in digital collections, an effective method for sorting documents is helpful to return relevant documents, and not return non-relevant documents, in the IR result [5].

An additional concern is about retrieving documents that are unauthorized, such as privileged or private documents (treated as non-relevant for sorting purposes). Also, the high cost associated with manual review of documents, both paper and electronic, increases the priority of using an automated method for sorting relevant and non-relevant documents, and reducing the number of documents in the retrieval set that must ultimately be settled by human review–the most expensive part of the process.

Precision, which we define in this paper as the *percentage of relevant documents returned in the search result*, is the measure used in this reported experiment to determine efficiency in a search result. Our goal here is that, if precision can be improved, then a smaller collection, with a greater percentage of relevant documents, can be produced by the automated system for human review.

The research reported in this paper is part of an ongoing exploration into the application of IR processes in the legal domain. In this case it turns out that precision (as a measure of IR efficiency) is a particularly acute goal in the domain of litigation and Legal-IR, where a party may often seek documents that are not relevant but may lead to relevance. This is an interesting task, given that the initial goal is not relevance, but learning the context of what might be relevant. In some cases, litigators are accused of "fishing" for information [3]. Fishing is an interesting metaphor for the volume problem, because if one pictures casting a large net to catch relevant fish, this can lead to the over inclusion in the IR result (too many non-relevant documents in the retrieval set).

II. Techniques Considered

During the initial evaluation of possible algorithms to apply to this task, we considered several techniques that have been proposed in the literature.

One such effective technique in the literature that can be adapted to this problem is called *stop words*. Stop-words traditionally are *non*-informative words such as "the, a, is" usually ignored by an IR algorithm [6]. Informative stop words could be used as filters. This study evaluates the use of filters to represent user selected elimination terms to remove a document from consideration.

We decided to conduct an experiment to determine if, in fact, user selected elimination criteria, such as the use of terms specifically intended to remove a document from consideration by the classifier, can improve precision simply by preventing a non-relevant document from being considered by the system. For example, the user may know of a certain type of header or footer, the number of words in a document, or specific terms (elimination terms) that will eliminate a document (containing positive search terms) from consideration otherwise retrieved by a query using Recall terms alone.

An example of a technique used to produce broader recall (inclusion of documents) but can lead to the problem of producing too many non-relevant documents (over-inclusion) is *stemming*. Stemming is the reduction of different forms of the same word down to its stem or root

Manuscript received October 16, 2015; revised February 4, 2016.

Harvey Hyman is with New College of Florida, 5800 Bay Shore Road, Sarasota, Florida 34243, USA (e-mail: hhyman@NCF.edu).

Terry Sincich and Rick Will are with University of South Florida, 4202 E. Fowler Avenue, Tampa, Florida 33620, USA (e-mail: tsincich@USF.edu, rwill@usf.edu).

Warren Fridy III is with H2 & WF3 Research, LLC 701 S. Howard Avenue, Tampa, Florida 33606, USA (e-mail: warren@h2wf3.com).

doi: 10.18178/ijmlc.2016.6.2.585

[6]. For instance, stemming "fall, falling, and falls" to find documents associated with a person falling (slip and fall cases). The limitation with stemming is that it can lead to non-relevant documents being included due to the polysemy problem (a single word having multiple meanings); in a "slip and fall" case the user does not want documents about an autumn day in the month of September.

Therefore, in order to be effective, recall terms chosen for document inclusion must translate some characteristic that distinguishes the relevant documents from the rest of the collection (the non-relevant documents). This concept has been identified as "term discrimination" [7]. The associated hypothesis is that term discrimination can be achieved by separating the *internal* context of the relevant document from the *external* content of the corpus (Precision) – we believe this can be achieved by separating recall terms for inclusion from elimination terms for exclusion, and classifying the documents into corresponding separate bins, thereby eliminating non-relevant documents from the retrieval set.

III. Background Literature on Filtering Approaches

There has also been a significant amount of research done in the areas of using both filtering and weighting methods to reduce the number of non-relevant items in a retrieval set. Early work in this area was done by Robertson and Jones in their 1976 paper which explored the use of "statistical techniques for exploiting relevance information to weight search terms" [8]. This work was still relevant 25 years later and extended in their 2000 paper reporting on a series of experiments using a probabilistic model [9].

During the intervening years, starting in 1989, Wong and Yao applied a probability distribution to an information retrieval model [10]; in 1995, Amanda Spink applied term relevance feedback [11], and in 1996, Losee applied "syntactical rules and tags" [12].

Since 2000, there have also been several experiments reporting the use of incorporated search behavior for relevance feedback [13], and exploring term dependencies [14] and Bayesian networks [15].

We shift our focus to the experiments reported in this paper. We extend the concept of relevance feedback techniques and term weighting methods as mentioned above, and also incorporate a behavioral approach, by offering the user the opportunity to divide their search model into inclusionary terms and exclusionary terms. The underlying thinking in our approach is that the user is in the best position to make their assessment on what is the most relevant and what is the least relevant, prior to employing an automated learning algorithm.

IV. Research Methods

The method used in the study reported herein is a controlled experiment. The data collected from the participants in the study are divided into *Recall* terms and *Elimination* terms.

Recall terms are words or phrases that are inclusionary. Elimination terms are words or phrases that are exclusionary. Our hypothesis is that by offering the user the opportunity to split their search structure into two methods of sorting, the automated search mechanism can be more discrete in its execution. This makes intuitive sense when one considers that quite often individuals engaged in an IR task may be unclear about what they specifically are seeking, but may in fact be very clear on what they wish to avoid, and therefore search by *elimination*, sometimes referred to as "culling." Once again, not a method to achieve perfection in search, but instead a method to address the question of how to reduce the volume of the search space and increase the percentage of relevant documents in the retrieval set.

The data collected in this experiment is analyzed using a paired differences test, also called a random block design (RBD). A user interface prototype has been specifically designed to support this data collection effort. A mockup of the user interface screen appears in the appendix.

The IR task in this case requests the participants to provide inclusionary (recall) and exclusionary (elimination) search terms with the goal of sorting relevant documents to respond to an IR request. The automated system prompts the participant to provide *Recall terms* and *Elimination terms* using the interface screen displayed in the Appendix.

The experiment is designed to evaluate how *elimination terms* as a separate module of an algorithm can impact performance in terms of *precision* as measured by the difference in non-relevant documents produced between participant samples. By having the participants provide both *Recall terms* and *Elimination terms*, the study avoids the possibility that some other input produced the reduction in non-relevant documents. The experiment also has each participant engage in 10 sessions to increase the likelihood that their selections are that of critical analysis and not random chance guessing. The prototype user interface built for the study is housed on a small private server and accessed by participants using a URL link from their self-provided laptop computers.

V. Task/Treatment

The **task** is an information retrieval task designed to approximate an IR task from the legal domain. The task description is reproduced in the Appendix. The **treatment** in this experiment is the use of elimination terms. The **dependent** variable in this case is *Precision* as measured by the percentage of relevant and non-relevant documents in the retrieval set. There are no covariates tracked in this experiment. The **effect** evaluated in this study is the use of *elimination terms* and the difference in total non-relevant documents retrieved in the task. The variables declared in this experiment are listed below in Table I.

TABLE I: List of Variables and Descriptions

Variable	Description
Recall Term (IV)	User chosen term for document inclusion.
Elimination Term (IV)	User chosen term for document exclusion.
Precision (DV)	Percentage of relevant documents in the retrieval set measured by the difference in number of non-relevant documents between samples.

VI. Dataset

The data set used is the EDRM version 2 of the Enron collection. The full corpus of this version contains approximately 650,000 to 680,000 email objects depending on the counting of attachments. This data set has been previously validated in the literature at several conferences on text retrieval [8]. In this experiment, a random sample of 10,000 email files from the larger corpus was used.

VII. Process

The description of sorting process and system architecture are depicted in Fig. 1. A file watcher is used to begin the procedure calls and process the user selected terms for the IR execution. The IR result set is saved to separate system bin files for export, usually in an Excel spreadsheet, but sometimes raw output is reviewed during the tuning process.

The system is designed to run the IR selections based on the user's chosen **recall terms** (saved in the R1 bin) and **elimination terms** (saved in the R2 bin). A depiction of the system model is displayed below in Fig. 1.

The retrieval task described in the Appendix is presented to the participants via their laptop access to the server application using a URL link.

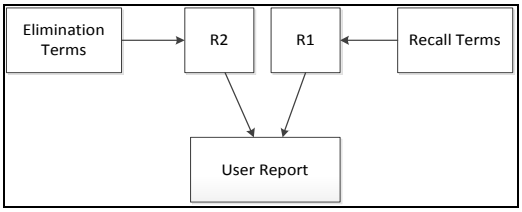


Fig. 1. System process design model.

VIII. Pilot Studies

There have been two pilots performed in this experiment to refine the system and the data collection instrument in preparation for the full study. The first pilot study consisted of 5 paralegals selected by convenience from the legal community. These 5 participants acted more like a focus group than a controlled population. Much of the interaction with the focus group was during open interviews about how they perform their IR tasks and how an automated process might assist them in reducing the volume of documents to review.

Feedback from the pilot study helped us address shortcomings in the initial design as well as some underlying assumptions we had about the potential users. For example, our initial design measured across four methods; it proved to be difficult to measure and hard to isolate effects. Also, in the pilot study, participants were told that there were up to 100 relevant documents in the collection. This caused confusion among the participants. Additionally, on a separate evaluation, our panel of experts felt that this may introduce a bias into the study if not changed.

Based on the feedback and trial runs, the experiment was redesigned to measure differences between the applications of search terms alone (Recall terms) versus the application of search terms combined with Elimination terms. After feedback was received from both the pilot participants and our panel of experts, we redesigned the experiment to the current format reported herein. We use a random block design (RBD) for data analysis. In this case the blocks are the participants. Each block contains two observations. The observations themselves are actually averages of 10 sessions conducted by each participating user. We are seeking to measure the differences between the two observations: paired differences.

A second pilot was conducted with the current design using the artifact with RBD for data analysis; it consisted of 10 paralegals who have worked as document reviewers. The purpose here was to confirm that the experiment design was in fact valid and capable of producing adequate results for evaluation. The participants from the second pilot were individuals who volunteered their time to assist with the project. The IR results produced from the second pilot are reported in Table II.

IX. Results of Pilot

Most interesting in the reported results of the second pilot is that a significant effect was detected with as few as 10 participants.

Precision significantly improved with the addition of using the elimination component over using recall search terms alone. The total number of Non-relevant documents was reduced on average from 40.4 to 30. Average reduction in non-relevant documents was 10.8, with the greatest reduction being 18 documents and the least reduction being 4 documents. The standard deviation between the samples of non-relevant documents was 4.4 for the use of *Recall terms* alone, and 4.3 for the addition of *Elimination terms*. Using this information, we obtain the following 95% confidence interval for the true mean reduction in non-relevant documents: 10.8 +/- 1.56 = (9.24, 12.36).

TABLE II: Reduction in Non-Relevant Documents

Participant	Non-Relevant Documents Using Recall Terms Alone	Non-Relevant Documents Using Elimination Terms	Reduction in Non-Relevant Documents
1	36	27	9
2	35	30	5
3	42	34	8
4	44	30	14
5	45	32	13
6	**42**	**38**	**4**
7	**45**	**27**	**18**
8	44	31	13
9	37	29	12
10	34	22	12
Average	40.4	30	**10.8**
STDev	4.4	4.3	4.3
CI			$10.8 +/- 2.26*\sqrt{4.3/9} = \underline{\mathbf{1.56}}$

X. Full Study Methods and Analysis

The method of analysis in the full study is also a random block design (RBD). Like before, we are measuring paired

differences; there are two rounds of observations for each participant. The study utilizes 30 participants. Each round consisted of 10 sessions for a total of 300 samples per round. The 10 sessions were averaged for each round. Each round was analyzed for differences in results produced between the observed rounds. The user selections have been captured using the same server based application as described previously.

Round One consisted of using recall terms alone to approximate inclusion of documents. Round Two consisted of recall terms combined with elimination terms to approximate the combined effect of inclusion conditions and exclusion conditions.

The difference lies in the implementation of the Recall terms alone versus the Recall terms along with Elimination terms. The 10 sessions for each round were averaged for each participant. The *blocks* in the random block design, are the participants.

SAS 9.2 was used to perform the statistical analysis for the random block design/paired difference test. The test is represented by the following model:

$$E(y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_9 X_9 + \beta_{10} X_{10} + \beta_{11} X_{11} + \beta_{12} X_{12} + \beta_{13} X_{13} + \beta_{14} X_{14} + \beta_{15} X_{15} + \beta_{16} X_{16} + \beta_{17} X_{17} + \beta_{18} X_{18} + \beta_{19} X_{19} + \beta_{20} X_{20} + \beta_{21} X_{21} + \beta_{22} X_{22} + \beta_{23} X_{23} + \beta_{24} X_{24} + \beta_{25} X_{25} + \beta_{26} X_{26} + \beta_{27} X_{27} + \beta_{28} X_{28} + \beta_{29} X_{29} + \beta_{30} X_{30}$$

where:

X_1 = A dummy variable representing (0) for Recall and (1) for Elimination,

X_2 through X_{30} = Dummy variables representing the averaged sessions for the participants are (1) for each level of participant and (0) for not.

The null and alternative hypotheses are as follows:

H_0: $B_1 = 0$, meaning there is no difference between the mean number of non-relevant documents using recall terms and the mean number of documents using elimination terms (e.g. $\mu_{recall} = \mu_{elimination}$).

H_a: $B_1 \neq 0$ meaning there is a significant difference between the mean number of non-relevant documents using *Recall* terms and the mean number of documents using *Recall* along with *Elimination* terms (e.g. $\mu_{recall} \neq \mu_{elimination}$).

XI. Results of Full Study

The findings produced by the experiment indicate that the use of elimination terms produces a statistically significant reduction in non-relevant documents in the retrieval set, resulting in an improvement in precision. This means that, on average, the search result contained more relevant documents and less non-relevant documents.

The effect detected was significant at alpha .01, with a 95% confidence interval for the true mean reduction in non-relevant documents: 10.37+/- .949 = (9.42, 11.32, meaning the average reduction in documents we expect to see would be a high of 11.32 and a low of 9.42.

The SAS 9.2 printout reporting the results of the RBD/Paired Difference analysis has been reproduced in Table III and the reduction in non-relevant documents have been reproduced in Table IV.

TABLE III: Paired Difference/RBD

The GLM Procedure
Class Level Information

Number of Observations Read 60
Number of Observations Used 60

The GLM Procedure
Dependent Variable: NON-RELEVANT DOCS

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	30	245916.3667	8197.2122	628.08	<.0001
Error	29	378.4833	13.0511		
Corrected Total	59	246294.8500			

R-Square	Coeff Var	Root MSE	NON_RELE Mean
0.998463	1.917027	3.612637	188.4500

Source	DF	Type I SS	Mean Square	F Value	Pr > F
PARTIC	29	244304.3500	8424.2879	645.48	<.0001
SAMPL	1	1612.0167	1612.0167	123.52	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
PARTIC	29	244304.3500	8424.2879	645.48	<.0001
SAMPL	1	1612.0167	1612.0167	123.52	<.0001

TABLE IV: Reduction in Non-Relevant Documents

Participant	Non-Relevant Docs Recall **Obs 1**	Non-Relevant Docs Elimination **Obs 2**	Difference in Docs
1	80	76	4
2	66	61	5
3	61	56	5
4	70	65	5
5	116	107	9
6	114	106	8
7	131	121	10
8	140	133	7
9	168	160	8
10	257	245	12
11	245	233	12
12	235	224	11
13	234	223	11
14	266	254	12
15	236	225	11
16	252	240	12
17	226	215	11
18	220	210	10
19	239	231	8
20	223	216	7
21	195	189	6
22	189	183	6
23	169	163	6
24	238	230	8
25	205	198	7
26	216	199	17
27	288	264	24
28	273	251	22
29	247	227	20
30	210	193	17

Average	**10.37**
Standard Deviation	**5.11**
Confidence Interval	10.37+/- 2.26* √(5.11/ 29)= **.949**

XII. Discussion

There were similarities and overlap in the recall term selections across the participants. This turns out to be consistent with other findings that have been reported in this domain [16]. Not surprisingly, some search terms produced significantly better recall results than others. For instance, participants submitting the recall term 'EOL' produced a retrieval result that was significantly higher in recall (relevant documents) than the other participants in their group who did not submit that term. This is consistent with other research in the field supporting the notion that retrieval is highly sensitive to choice of search terms and can often be effected by the context of the subject matter being searched [3], [4].

The significant results produced for reduction in non-relevant documents due to the use of elimination terms are encouraging given that the time and cost associated with manual, human review can be reduced if there are significantly fewer non-relevant documents in the retrieval result [4]. In other words, a smaller total set to review, and fewer false positives to have to wade through.

One explanation for why precision was significantly improved could be due to the fact that *Elimination terms* are applied as exclusionary without regard to weights or threshold. Perhaps, the use of *Elimination terms* allowed for non-relevant documents to be "eliminated" from the retrieval that otherwise would have been included by using inclusionary (Recall) search terms alone. However, there were instances when relevant documents were eliminated from the retrieval resulting in a reduction of relevant documents in the result. This is a possible drawback for IR tasks where recovery of as many relevant documents as possible is the supreme goal, even if it means wading through a greater number of non-relevant documents. This will need to be addressed in future studies to determine if the loss in relevant documents can be controlled to avoid this negative consequence.

XIII. Limitations

Like any study, there are limitations in the study reported here. First, our participant populations for both the pilot and the full study were small and we acknowledged that we only had 30 users in the full study. To compensate for that, we had each participant conduct 10 sessions for each round of selections in order to produce more usable observations. We suggest that this may have produced an unintended benefit, insofar as the participants had the opportunity for multiple iterations that were averaged into each round, rather than having a larger group of participants but fewer chances to attempt the task. In our next study we will attempt to accomplish both.

Ultimately in this case, the sample size of 30 participants with 10 sessions for each round, was large enough to produce a detectable significant result. The study as designed, establishes that there is an effect here, worthy of future exploration.

A more important limitation with this experiment might be that the results are ad hoc – meaning that the results produced may be peculiar to this data set or the design of the task itself. Given that we used a single task from the TREC Conference competition, the confidence in being able to generalize the results to other IR domains is somewhat reduced. We plan to address this limitation in our next series of experiments, and in a future study using a different data set, with a different IR task to improve the universality for explaining the phenomenon claimed to be observed in this experiment.

XIV. Future Work

The initial results from the pilots and the full study are encouraging for the use of elimination terms as a combined sorting component for large scale IR. Clearly there is an effect, but how generalizable it is will remain to be seen.

Our next steps will be to conduct cross-validation studies with other populations and different domains, starting with repeating this experiment on a dissimilar data set, using alterative search tasks, to see if the results produced here are repeatable across circumstances, populations, and environments.

XV. Contributions

The series of experiments reported in this study demonstrate that *non-relevant* documents can be reduced by the use of an *elimination term* function, thereby improving precision in the retrieval set. This effect can be used to gain leverage in IR tasks oriented more toward precision (fewer documents and reduced volume) rather than recall (maximizing relevant documents in spite of greater size return sets).

If non-relevant documents can be reduced (increase precision) without a significant loss of relevant documents (recall), use of a separately applied elimination component can successfully reduce the time and cost associated with manual, human review. Further study of the relationship of *elimination terms* with recall and precision is certainly warranted.

A particularly significant discovery reported here is that recall of relevant documents did not vary widely by participant. This is consistent with prior findings, that *relevance* is very *context* and *content* dependent. The new insight here is the finding that *precision* can be manipulated through the use of an algorithm modified to accept elimination terms as a separate module in the IR algorithm.

XVI. Conclusion

The goal of the study reported here was to explore an experimental method to address the volume problem in large scale digital collections, associated with manual,

human review. The results produced during the series of experiments as reported, demonstrate that it is feasible to address large volume by using a sorting mechanism based on a combination of recall terms and elimination terms.

APPENDIX

Information Retrieval task

Task adapted from TREC 2011 Legal Track Topic 401

The purpose of this task is to retrieve documents that match the below request for production. The company in this case is Enron. The company is a now defunct energy trading company that was the subject of a large body of litigation both civil and criminal.

The following is the request for production:

You are requested to produce all documents or communications that describe, discuss, refer to, report on, or relate to the design, development, operation, or marketing of enrononline, or any other online service offered, provided, or used by the Company (or any of its subsidiaries, predecessors, or successors-in-interest), for the purchase, sale, trading, or exchange of financial or other instruments or products, including but not limited to, derivative instruments, commodities, futures, and swaps.

Additional Guidance for Relevance:

The above request broadly seeks documents concerning Enron online, the Company's general purpose trading system, or any other online financial or commodities services offered, provided, or used by the Company and its agents.

In this case attorney-client communication or otherwise privileged information is not an issue.

This request is seeking information specifically about an online system for trading financial instruments. A document is not relevant if it refers to the purchase, sale, trading, or exchange of a financial instrument or product, but does not involve the use of an online system.

A document is relevant if it describes, discusses, refers to, reports on, or relates to: the design, development, operation, or marketing of "enrononline," or any other online services offered, provided or used. This includes, how the system was set up, how the system worked on a day-to-day basis, how the Company developed or modified the system, how the Company marketed or advertised the system, and the actual use of the system by the Company, its subsidiaries, predecessors, or successors in interest.

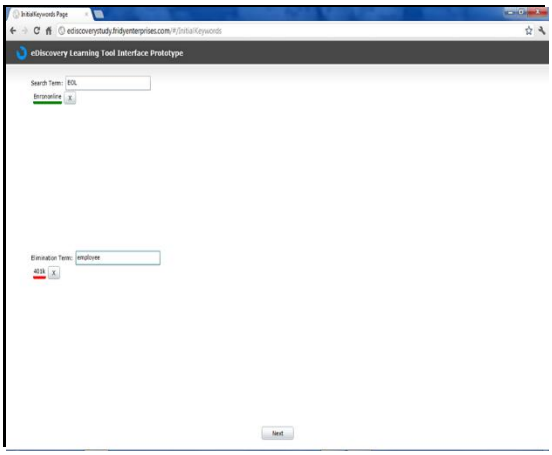
A relevant document can be for the purchase, sale, trading, or exchange of: financial instruments, financial products, including, derivative instruments, commodities, futures, or swaps. These instruments and products are distinguished from other goods and services by the fact that their value depends on future events and their purchase incurs financial risk.

A document is relevant even if it makes only implicit reference to these parameters. No particular transaction (i.e., purchase or sale) need be cited specifically. If the document generally references such activities, transactions, or a system whose function is to execute such transactions, and it otherwise meets the criteria, it is relevant.

Examples of responsive documents include: Correspondence, Policy statements, Press releases, Contact lists, or Enronline guest access emails.

Additional Guidance for Non-Relevance

Examples of non-relevant documents include: Purchase, sale, trading or exchange of products or services other than financial instruments or products, or any documents referring to employee stock options or stock purchase plans offered as incentives or compensation, or the exercise thereof. Also documents relating to structured finance deals or swaps that are specified explicitly by written contracts, even if the contracts themselves are electronic or electronically signed are not relevant. Also documents related to the use of online systems by Enron employees for their personal use are outside this request and are not relevant.



User interface screen.

REFERENCES

[1] J. Ward and A. Barker, *Undefined By Data: A Survey of Big Data Definitions*, 2013.

[2] M. Oussalaleh, S. Khan, and S. Nefti, "Personalized information retrieval system in the framework of fuzzy logic," *Expert Systems with Applications*, vol. 35, p. 423, 2008.

[3] D. W. Oard, J. R. Baron, B. Hedin, D. D. Lewis, and S. Tomlinson, "Evaluation of information retrieval for e-discovery," *Artificial Intelligence and Law*, vol. 18, no. 347, 2010.

[4] M. R. Grossman and G. V. Cormack, "Technology-assisted review in e-discovery can be more effective and more efficient than exhaustive manual review," *Richmond Journal of Law and Technology*, vol. 27, iss. 3, 2011.

[5] H. S. Hyman, T. Sincich, R. Will, M. Agrawal, B. Padmanabhan, and W. Fridy III, "A process model for information retrieval context learning and knowledge discovery," *Artificial Intelligence and Law*, vol. 23, no. 2, 2015.

[6] A. Singhal, *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, 2001.

[7] K. Spark-Jones, "A statistical interpretation of term specificity and its application in retrieval," *Journal of Documentation*, vol. 28, no. 1, 1972.

[8] S. E. Robertson and K. S. Jones, "Relevance weighting of search terms," *J. Am. Soc. Inf. Sci*., vol. 27, pp. 129–146, 1976.

[9] K. Sparck-Jones and S. E. Robertson, "A probabilistic model of information retrieval: development and comparative experiments," *Information Processing & Management*, vol. 36, no. 6, p. 779, 2000.

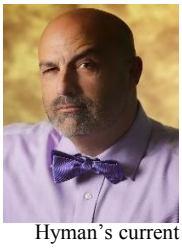
[10] S. K. M. Wong and Y. Y. Yao, "A probability distribution model for information retrieval," *Information Processing & Management*, vol. 25, no. 1, p. 39, 1989.

[11] A. Spink, "Term relevance feedback and mediated database searching: Implications for information retrieval practice and systems design," *Information Processing & Management*, vol. 31, no. 2, p. 161, 1995.

[12] R. M. Losee, "Learning syntactic rules and tags with genetic algorithms for information retrieval and filtering: An empirical basis for grammatical rules," *Information Processing & Management*, vol. 32, no. 2, p. 185, 1996.

[13] I. Ruthven, M. Lalmas and K. van Rijsbergen, "Incorporating user search behavior into relevance feedback," *Journal of the American Society for Information Science and Technology*, vol. 54, no. 6, 2003.

[14] B. Cho, C. Lee, and G. Lee, "Exploring term dependences in probabilistic information retrieval model," *Information Processing & Management*, vol. 39, no. 4, p. 505, 2003.
[15] S. Acid, L. De Campos, J. Fernández-Luna, and J. Huete, "An information retrieval model based on simple Bayesian networks," *International Journal of Intelligent Systems*, vol. 18, no. 2, 2003.
[16] E. Vorhees and L. Buckland, *TREC Proceedings (2009, 2010) NIST Special Publication*.



Harvey Hyman holds a PhD in information systems and decision sciences from University of South Florida, an MBA from Charleston Southern University, and a law degree (JD) from University of Miami, Florida.

He holds or has held teaching appointments at Georgia Southern University, Florida Polytechnic University, University of South Florida, New College of Florida, and Saint Leo University.

Hyman's current appointment is with the Library of Congress, National Library Service where he is leading the effort to migrated their content delivery system to a cloud-based service architecture.

Terry L. Sincich is an associate professor in the Information Systems Decision Sciences Department. He teaches introductory statistics at the undergraduate level and advanced statistics courses at the doctoral level. He has won numerous teaching awards, including the Kahn Teaching Award and Outstanding Teacher Award, recognizing excellence as an educator.

Dr. Sincich's areas of research interest include applied statistical analysis and statistical modeling. He has co-authored statistics textbooks, book chapters, and research papers in top-rated journals in a variety of disciplines, including the Journal of the American Statistical Association, Academy of Management Journal, Auditing: A Journal of Theory & Practice, Research on Accounting Ethics, Demography, and International Journal of Forecasting. He has presented his work numerous times at academic conferences.

Dr. Sincich received a BS degree in mathematics and MS and PhD in statistics from the University of Florida, where he also taught for several years before joining USF in 1988. He is a member of the American Statistical Association and the Decision Sciences Institute.

Rick Will is an associate professor in the Information Systems and Decision Sciences department at the University of South Florida. He holds a doctorate (PhD) in management information systems from The University of Houston, College of Business, where he also received his M.B.A. His B.S. was received from the College of Technology, also at the University of Houston.

He has published in journals such as Communications of the ACM, International Journal of Human Machine Studies, Computers in Human Behavior, Interfaces, Expert Systems with Applications, Journal of Computer Information Systems, and the Journal of Information Technology Education: Discussion Cases. He is teaching and researching in the areas of IT Project Management, Systems Analysis and Design and Technology Innovations in Education.

Rick is currently working with the Jimmy Carter National Historic Site – Education Program with the design and implementation of multimedia applications.

Warren Fridy III is a co-founder, chief technology officer, and director of Product Design and Development at RetrivikaTM a cloud based eDiscovery software service innovator.

His love of computers and technology began at a very early age. When he was just a junior in high school, he opened his first computer consulting company. That passion for technology has continued through his bachelor and master degrees in computer science and into his professional career. His knowledge and experience reaches a wide variety of fields including insurance, financial, and education.