Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 1127–1134

Optimising Web-Based Information Retrieval Methods for Horizon Scanning Using Relevance Feedback

Marco A. Palomino and Tim Taylor
University of Exeter
Medical School
Truro, UK
[m.palomino, tjt205]@exeter.ac.uk

Geoff McBride and Hugh Mortimer
STFC Futures Programme
Rutherford Appleton Laboratory
Didcot, UK
[geoff.mcbride, hugh.mortimer]@stfc.ac.uk

Richard Owen
University of Exeter
Business School
Exeter, UK
r.j.owen@exeter.ac.uk

Michael Depledge
University of Exeter
Medical School
Truro, UK
m.depledge@exeter.ac.uk

Abstract—**Horizon scanning is being increasingly regarded as an instrument to support strategic decision making. It requires the systematic examination of information to identify potential threats, emerging issues and opportunities to improve resilience and decrease risk exposure. Horizon scanning can use the Web to augment the acquisition of information, though this involves a search for novel and emerging issues without knowing them beforehand. To optimise such a search, we propose the use of relevance feedback, which involves human interaction in the retrieval process so as to improve results. As a proof-of-concept demonstration, we have carried out a horizon scanning exercise which showed that our implementation of relevance feedback was able to maintain the retrieval of relevant documents constant over the length of the experiment, without any reduction. This represents an improvement over previous studies where relevance feedback was not considered.**

I Introduction

The use of the World Wide Web for futures research has been gaining increasing attention [1-3]. Largely, the aim of futures research is to anticipate and prepare for new and changing risks, and to consider the implications that emerging issues will have on the distribution of resources and existing priorities. Given the current environment of change and uncertainty, both public and private sectors have identified the need to strengthen futures research and integrate it into strategic thinking and planning.

In the UK, the importance of futures research has been highlighted by a series of perceived failures in science and policy, such as the failure to recognise the concerns of the public about genetically modified crops until they emerged in the media, and the inadequate reaction to the outbreak of the foot/hoof and mouth disease in 2001 [4]. As a consequence of these setbacks, the UK Government has emphasised its use of horizon scanning, "the systematic examination of information to identify potential threats, risks, emerging issues and opportunities, beyond the Parliamentary term, allowing for better preparedness and the incorporation of mitigation and exploitation into the policy making process" [5]. Explicit objectives of horizon scanning are to anticipate issues, accumulate reliable data and knowledge about those issues and thus inform policy making and implementation [6].

Data collection associated with horizon scanning has blossomed with the availability of electronic databases and Web search engines. Regrettably, the process of searching for potential threats and emerging issues is not transparent. While searching is a retrieval process where the searcher knows in advance what she is looking for, horizon scanning is a process where we are trying to discover what is novel and surfacing without knowing it ahead of time. As explained by Palomino et al [7], we have access to ''search engines'' on the Web, but not to ''scanning engines''.

The impossibility of establishing precisely what is being sought before beginning the search makes it difficult to formulate information queries that are well designed for horizon scanning purposes. This suggests that the first retrieval operation involved in the process of scanning the horizon should be conducted with a tentative, initial query, and should be treated as a trial only, designed to locate a few useful items, which could then be examined for relevance so that later on new and improved query formulations can be constructed with the expectation of retrieving additional useful items in subsequent search operations. This is the reason why we have decided to explore the use of a controlled, automatic process for query reformulation, namely, relevance feedback, a technique utilised by some information retrieval systems [8].

The aim of this paper is to assess the use of relevance feedback as part of a horizon scanning system. To this extent, the remainder of this paper is organised as follows: Section II reviews related work on relevance feedback and briefly outlines previous research on Web-based horizon scanning. Section III details our implementation of relevance feedback in the context of a horizon scanning prototype which we are employing as a proof-of-concept demonstration. Section IV discusses a horizon scanning exercise that was conducted for a European Union Framework 7 project in association with RAL Space [9]—a world-class space research centre—to review current and future technologies for detecting and monitoring diseases in vegetation. We used this exercise as a case study to test our implementation of relevance feedback. Section V reports on the evaluation of the results of RAL Space's exercise, and, finally, Section VI states our conclusions.

978-1-4673-4471-5/$25.00

II RELATED WORK

Relevance feedback has been extensively studied since its development in the mid-1960s [8, 10-12]. It refers to an interactive process that helps to improve retrieval performance: when a user submits a query, an information retrieval system would first return an original set of documents that satisfy the query and then ask the user to judge whether these documents are relevant or not; after that, the system would reformulate the query based on the user's judgments, and return a new set of documents. To some extent, relevance feedback is an alternative to save users from articulating queries in a trial-and-error manner.

Most of the research on relevance feedback undertaken thus far has approached its implementation as a supervised learning problem [8, 10, 11], where the key is to optimally balance the original query and the feedback information [13]—a special track to look into the effects of different factors on the success of relevance feedback has been organised by the Text Retrieval Conference (TREC) [14]. However, the use of relevance feedback in the context of horizon scanning has not been investigated yet. References to the applications of horizon scanning and the results of specific scans keep growing [4, 6, 15-21], as the interest in the subject increases, but only a few academic papers describe the methodology to carry out an automated scan [7, 22, 23], and the combined use of horizon scanning and relevance feedback has not been documented until now.

Shaping Tomorrow [24] and Recorded Future [25] are two private firms using Web-based scanning tools. Shaping Tomorrow helps organisations make better decisions through anticipating and preparing for the future. It uses a variety of manual, semi-manual and automated scanning processes to track and share information from around the world. It is first supported by a virtual network of volunteer and client researchers who "scan the scanners"—experts in the field—for material. Shaping Tomorrow also employs its own purpose-built Web-robot to scrape high value future websites and its service has accumulated 100,000 scan hits on emerging change, gathered over ten years from 5,000 plus sources, and 3,600 issues—trends, uncertainties and surprises—evidenced and linked to the scan hits. Shaping Tomorrow will soon release software to read the scan hit and do almost all of the researchers work automatically [26].

Recorded Future is established on the premise that all the information available on the Web is useful to support forecasting methods, financial or otherwise. Recorded Future continuously harvests news from more than 40,000 online sources, ranging from media and government websites to individual blogs and selected twitter streams [25]. Recorded Future aims to create and maintain a database of facts—pairs of timed entities and event instances—to track trends and historical developments and predict future events. As opposed to Recorded Future, we are not interested in predicting the future, but rather in improving resilience and the capability to react to new risks and opportunities.

In the public sector, horizon scanning has proved useful to identify new and emerging health technologies [20, 27, 28]. However, due to the large amount of information published online, it is difficult to recognise valuable data [29]. In an attempt to establish how exactly the Web should be used in health technology assessments, Douw et al [27] circulated a questionnaire among organisations known to use the Web for horizon scanning purposes. The questionnaire focussed on the type of websites scanned, the frequency of the scanning, and the importance of the Web for the identification of new health technologies. Responses to the questionnaire indicated that the organisations surveyed found new information through word of mouth, and links found on websites that they monitor continuously. Even though this highlights the importance of personal networking in horizon scanning, and the expertise of the scanners to choose the best links to follow, our work is directed towards the automation of the human-intensive practice of detecting and summarising emerging information. Hence, rather than surveying organisations, we have concentrated on the methodology to carry out a Web-based scan of the horizon.

Our methodology for Web-based horizon scanning comprises several interlinked components, as described by Palomino et al [7]: emerging information is retrieved—manually or otherwise—and / or received—e.g., via selected RSS feeds—from a variety of Web-based sources—such as, online scientific, peer-reviewed literature and news websites, which were sources of high importance for the work with RAL Space that we will describe in Section IV. Key parts of the retrieved information may be extracted and later on categorised. Afterwards, the information is often archived in a database. Periodically, outputs are presented to decision makers or used to write up reports or newsletters.

III RELEVANCE FEEDBACK

The main idea behind our implementation of relevance feedback consists of choosing important keywords attached to certain previously retrieved documents that have been characterised as relevant by the users, and of enhancing the importance of those keywords in future queries. Correspondingly, keywords included in previously retrieved non-relevant documents could be deemphasised in any future query formulation. Ideally, the effect of this query alteration process is to "steer" the query in the direction of the relevant documents and away from the non-relevant ones, with the expectation of retrieving more useful and fewer non-useful documents in later steps of the search.

Figure 1 shows a general Web-based horizon scanning approach for strategic decision support that uses relevance feedback. It accentuates the importance of the continuous scanning, noting that the processes of retrieving documents, and analysing, categorising and archiving information are iterated as part of a continuous process—static or sporadic scans become outdated quickly. The outputs of horizon scanning can be interfaced with further tools for opportunity and risk analysis [14] and scenario development.

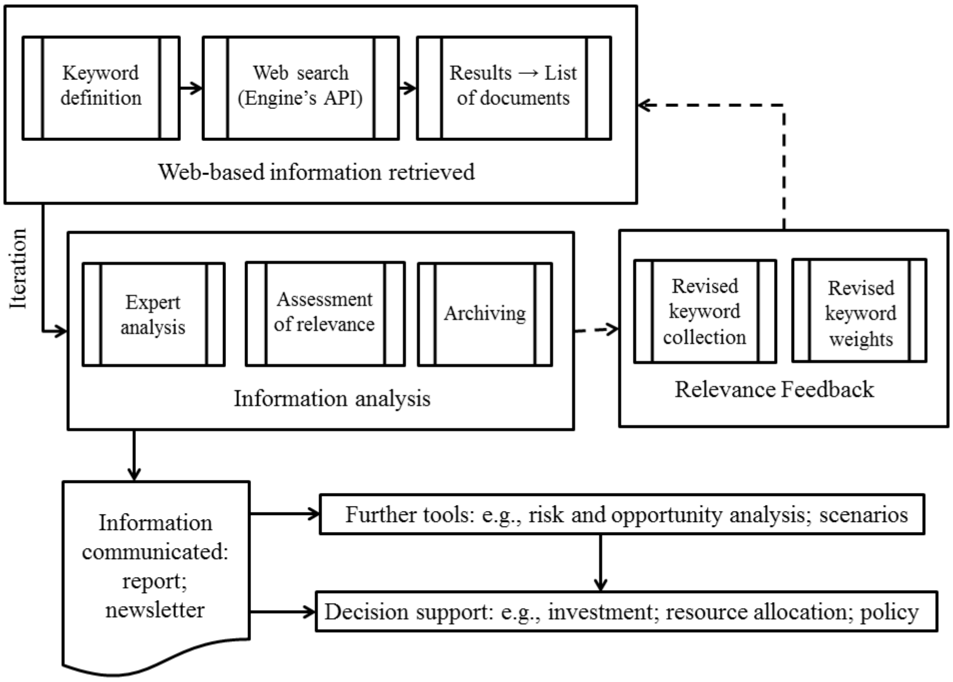

Figure 1. A generalised approach to Web-based horizon scanning for decision support using relevance feedback—based on Palomino et al [7]

Relevance feedback offers the following advantages to the analysts in charge of scanning the horizon:

(i) It frees the analysts from the details of the query formulation process—especially in late stages of the search.
(ii) It splits the search into an organised sequence of steps to reach the desired information gradually.
(iii) It devises useful queries without former analysis of the availability of data on the Web.
(iv) It features a controlled query alteration process designed to emphasise some keywords and deemphasise others, as required to accomplish a particular search.

Relevance feedback was originally developed as a technique to be used in conjunction with vector queries—i.e., queries represented by vectors with as many entries as keywords comprised in the query. Each entry refers to a "weight" symbolising the importance of the corresponding keyword within the query. For example, a particular query Q composed of n keywords may be written as

$$Q = (w_1, w_2, \dots, w_n),$$

where w_i is the weight of the i-th keyword. Keyword weights are restricted to the range 0 to 1, where 0 means the corresponding keyword is absent from the query and 1 means it is so critical to the query that it has a full weight.

Given a vector such as Q, the relevance feedback process starts by generating a new vector

$$Q' = (w_1', w_2', \dots, w_n'),$$

where w_i' represents a modified weight for the i-th keyword in the query—new keywords can be introduced to the query, and old keywords can be removed by reducing to 0 its weight. The process continues by creating yet another vector Q'' by modifying the weights of Q' according to new feedback, and so on and so forth until the required documents are found or the process reaches a pre-established number of iterations. Graphically, the relevance feedback process can be depicted as a relocation of the query vector from one place to another in the n-dimensional space defined by the n keywords under consideration.

A poorly conceived query reformulation can result in deterioration in retrieval performance [30]. Hence, a suitable set of keywords to search for information should be selected at each step in the process. We always choose our keywords with the support of software for the automatic extraction of keywords. Specifically, we use Yahoo!'s Content Analysis Web Service [31].

Normally, a scan of the horizon begins by defining the goals of the scan with a few sentences. We then submit those sentences to Yahoo!'s Content Analysis Web Service to automatically extract keywords—when available, entire documents relevant to the scan, called seed documents, are submitted to extract keywords.

These keywords are used to create the initial queries to search the Web for information. Normally, these keywords are combined with terms and phrases such as `new development`, `revolutionary`, `first time`, and others which have been suggested by the UK Defence Science and Technology Laboratory (Dstl) as descriptors of emerging issues [32]. These combinations of automatically extracted keywords and descriptors of emerging issues constitute the queries employed to bootstrap the relevance feedback process—i.e., these are the queries whose formulation we will attempt to refine along the process.

Once we have retrieved a first list of documents as a result of releasing our queries, we proceed to collect feedback. Usually, an expert, or a group of experts, in the field of the scan, or the same people who developed the requirements for the scan, are asked to indicate, for each document in our results, whether it is relevant, very relevant or non-relevant. The documents that are marked as very relevant are submitted to Yahoo!'s Content Analysis Web Service to extract new keywords. Keywords that were not considered in the initial queries, but are at the top of the new list of keywords yielded by Yahoo!'s Content Analysis Web Service are added to the original keywords and used to formulate new queries—keywords at the top of the list are expected to be more characteristic of the documents submitted than those near the bottom [31].

For each document that we retrieve, we keep a record of the keywords that were included in the queries used to retrieve it—note that a particular document can be retrieved as a result of more than one query and therefore be associated with several keywords. The weights of keywords used in queries that retrieved documents that were marked as very relevant are increased by a factor proportional to the number of very relevant documents associated with them. Likewise, the weights of keywords associated with documents marked as non-relevant is decreased by a factor proportional to the number of non-relevant documents associated with them—see Figure 2. The weights of keywords associated with documents marked as relevant—but not very relevant—is not modified and remains the same for the following iteration.

Once the set of keywords has been amended to integrate the initial feedback received, and the weight of each keyword has been adjusted to reflect the number of relevant, very-relevant and non-relevant documents retrieved with them, we proceed to release new queries, whose formulation can be thought of as a refinement of the initial ones, and the entire process can be repeated again until we complete a pre-established number of iterations. To automate our search for documents on the Web, we programmatically released our queries via Google's Custom Search API [33]. We chose Google's Custom Search API, because Google is the most popular search engine [34]; yet, other engines with an API interface could be used too—in other words, we will focus on Google for testing purposes, but the approach described here is not restricted to a specific search engine.

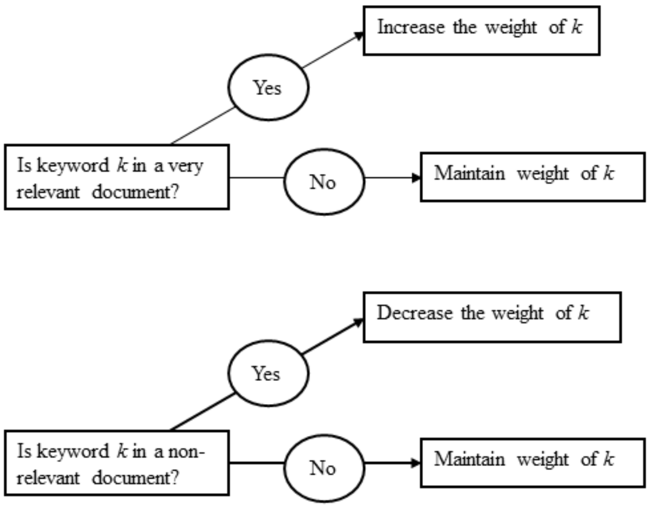

Figure 2. Keyword weight adjustment

Google has one of the largest databases of Web pages, including many types of documents—blog posts, wiki pages, group discussion threads—and document formats—PDF, Microsoft Word or PowerPoint documents, among many others. Despite the presence of all these types of documents and formats, Google's method of ranking on the basis of the PageRank citation algorithm [35] often places relevant documents near the top of the search results, and Google's Custom Search API allows us to query Google's repository directly and frequently in an automated way. Indeed, the frequency with which we query Google's repository can be adapted to the particular needs of the scan.

A. Queries with weighted keywords

A critical aspect of our relevance feedback implementation is the use of weights to express the importance keywords. Appropriately using those weights is what guarantees that our process reaches the desired information gradually; otherwise, the continuous extraction of keywords from newly retrieved documents would simply increase the number of keywords and queries, which would in turn increase the number of collected documents, without guaranteeing that we are actually gathering more useful information. Devising a way to adequately use the weights so that subsequent queries assign higher importance to keywords with greater weights is one of the most challenging features to accomplish.

Our implementation is based on using the weights of the keywords to decide how we should employ those keywords to look for documents:

(i) Keywords with low weights are used to search for documents that include the keywords anywhere in the text—not necessarily in prominent places.

(ii) Keywords with high weights are used to search for documents that include the keywords in their titles—according to Page et al [35], titles are more descriptive of the contents of a document than the rest of the text.

(iii) Keywords with very high weights are used to search for documents which are referenced to by hyperlinks whose text includes the keywords—Page et al [35] have stated that the text contained in the hyperlinks that point to a document, also known as the anchor text, link text, or link title, is greatly descriptive of the contents of the document referred to.
(iv) Keywords whose weights have been reduced to 0, which means that they have no relevance at all to the search, are preceded by the "minus" operator in our queries to explicitly indicate that they must not appear in the retrieved documents.
(v) All keywords have the same weight at the start, when the first search takes place and no feedback has been gathered yet. For the first iteration, all keywords are used to search for documents that contain them anywhere in the text.
(vi) Keywords that are meant to be descriptors of emerging issues—for instance, `ground breaking` and `closer to reality`—have constant weights that are never modified through the entire process. We always search for documents that contain these keywords anywhere in the text.

Our implementation of relevance feedback ensures that keywords with higher weights are looked for in places which are expected to have higher importance and therefore be more descriptive of the documents that contain them. Table I displays the association between weight ranges for keywords and the locations—hyperlinks, titles, or general text—where we search for those keywords to retrieve new documents that contain them.

TABLE I. KEYWORD RANGES AND KEYWORD LOCATIONS

Weight range	Keyword location
0	Nowhere in the document
(0,0.33]	Anywhere in the text
(0.33,0.66]	In the title
(0.66,1]	In the anchor text

In order to illustrate the relevance feedback process in detail, we will use an example. The example derives from a horizon scanning exercise proposed by RAL Space in October 2012 and it is explained in the following section.

IV RAL SPACE SCANNING EXERCISE

In October 2012, RAL Space, based at the Rutherford Appleton Laboratory (RAL), undertook a review for the European Union Framework 7 project Q Detect: Developing Quarantine Pest Detection Methods for use by National Plant Protection Organizations (NPPO) and Inspection Services [36]. In this review, RAL Space looked into current and future aerial platform technologies and instrumentation options for detecting and monitoring diseases in vegetation, and the mapping of pests through the use of aerial platforms.

The review aimed to assess the efficacy of remote sensing techniques—such as direct imaging and spectrally resolving reflected light—from different aerial platforms—ranging from small unmanned aircraft to low altitude satellites—to evaluate and monitor the health of plant life over long periods of time with little human inspection. The report was not meant to target specific plant diseases, but to provide an overview of various, if not all, potential diseases, whilst providing a thorough examination of the state-of-the-art in remote sensing instrumentation and platform technology.

As part of the review, RAL Space assessed how low, medium and high-altitude platforms integrated with high spectral and spatial resolution instrumentation could be used to come up with different performance metrics within a specific user requirement framework, which included cost, endurance, spatial resolution and frequency of measurement. RAL Space's review contributed to compare the economic benefit and practical realisation of present and forthcoming technology to assist in the detection of quarantined disease remotely. Since decision making on the uptake and use of emerging technology for disease monitoring has to be supported by timely and high quality information, RAL Space made use of horizon scanning to produce the review.

The horizon scanning exercise began by establishing the seed documents. These documents—listed in Table II—were mostly academic papers chosen by RAL Space.

TABLE II. SEED DOCUMENTS

Carter, G.A. & Knapp, A.K. 'Leaf optical properties in higher plants: linking spectral characteristics to stress and chlorophyll concentration.' American Journal of Botany, 88: (2001)
Cloutis, E.A. 'Agricultural crop monitoring using airborne multi-spectral imagery and C-band synthetic aperture radar'. International Journal of Remote Sensing. Volume 20, Issue 4. (1999)
Coops, N.C., Goodwin, C., Stone, C. Sims, N. 'Assessment of forest plantation canopy condition from high spatial resolution digital imagery'. Canadian Journal of Remote Sensing. 32: (2006)
Lelong, CD, Burger, C., Jubelin, G. Roux, Labbé, S. & Baret, F. 'Assessment of Unmanned Aerial Vehicles Imagery for Quantitative Monitoring of Wheat Crop in Small Plots'. Sensors 8. (2008)
Moran, S.M. 'Thermal Infrared Measurement as an Indicator of Plant Ecosystem Health.' Journal remote sensing. (2003)
Rock, B., Vogelmann, J., Williams, D., Vogelmann, A., & Hoshizaki, T. 'Remote Detection of Forest Damage.' BioScience, 36. (1986)
Sharples, J.A. 'The Corn Blight Watch Experiment: Economic implications for use of remote sensing for collecting data on major crops'. LARS information note 110173.

The text of all the abstracts of the academic papers in Table II was submitted to Yahoo!'s Content Analysis Web Service, and a large list of keywords was produced in return. Together with an analyst from RAL Space, we chose the keywords that we considered most useful and grouped them into three different categories:

(i) Subject keywords, which refer to the main subject of RAL Space's review—for example, `crop monitoring` and `plant health`.
(ii) Technology keywords, which refer to different technological alternatives for detecting and monitoring diseases in vegetation—for example, `satellite` and `remote sensing`.
(iii) Descriptors of emerging issues, which are keywords defined by Dstl to capture "fresh" information on relevant subjects.

Table III shows the precise set of keywords that we use to start the process. Combinations of these keywords produced a total of 140 queries: each query included one, and only one, keyword from each category. Those 140 queries were used to start the search.

TABLE III. INITIAL SETS OF KEYWORDS

Subject	Technology	Emerging issues
crop disease	aerial platforms	breakthrough
crop monitoring	remote sensing	closer to reality
environmental monitor	satellite	first time
forest monitoring	unmanned aerial vehicle	ground breaking
plant health		new development
		novel
		revolutionary

Although we set up our prototype to limit to 64 the number of results per query, this still allowed up to 8,960 documents to be retrieved for each automatic release of the 140 queries employed in the initial search—indeed, nearly 4,000 unique documents, approximately, were retrieved per iteration. It would be unmanageable for a RAL Space analyst to review all those documents, given the short time allocated to this activity. Hence, we committed to deliver 50 documents, exclusively, per iteration to RAL Space, because this was the number of estimated documents that could be reviewed by a RAL Space analyst per iteration.

We assumed that the documents of most importance—i.e., those of greatest relevance—would be the ones that consistently appear at the top of the search results. We thus presented a ranked list of documents to RAL Space, with the ranking being based on the number of times that the document was retrieved by Google's Custom Search API over the course of each iteration—i.e., cumulative retrieval occurrences from programmatic releases of queries—see Palomino et al [22] for more details regarding the use of Google's Custom Search API.

Once the top-ranked 50 documents per iteration were chosen, we divided them into three different categories: academic papers, news articles and standard documents. The academic papers comprised, mostly, peer-reviewed papers relevant to the scan. The news articles were, mostly, press releases and news articles available on the Web; and the list of standard documents consisted of documents retrieved as a result of our queries that were not published by news websites or online academic journals. All the documents that we delivered, regardless of the category, were published between 2010 and 2012, exclusively.

V RESULTS

We previously conducted a benchmarking study between September and October 2010 in collaboration with Lloyd's of London [37], one of the global leaders in the insurance market. The goal of that study was to use our prototype for framing decision making on novel risks—specifically risks associated with space weather and how these might affect terrestrial and near-Earth insurable assets [22]. As part of the study, we were able to identify several documents that Lloyd's Emerging Risks Group analysts considered very relevant to assess insurance exposure; yet, the number of very relevant documents retrieved per week decreased as the experiment progressed, while the number of non-relevant documents retrieved increased over the same period [22].

Table IV displays the precise numbers of very relevant, relevant and non-relevant documents retrieved weekly in our study with Lloyd's of London—relevance feedback was not employed in that study and the relevance of the documents was evaluated according to the criteria developed by Lloyd's analysts —see Palomino et al [22] for full details.

TABLE IV. LLOYD'S EVALUATION RESULTS

	Week 1	Week 2	Week 3	Week 4
Very relevant	29	19	11	5
Relevant	66	64	74	74
Non-relevant	5	17	15	21

Although there were reasons to justify why most of the very relevant documents retrieved in our Lloyd's study were discovered in the first week, one of the major goals of the current study, and a motivation for our interest in relevance feedback, was to improve the performance of our prototype to make sure that the retrieval of relevant documents remains constant over the length of the experiment.

The scanning exercise undertaken with RAL Space comprised three iterations between 12 and 19 October 2012. Table V shows the exact number of very relevant, relevant and non-relevant documents retrieved per iteration. Table V shows that the number of very relevant documents decreased by one in the second iteration but then remained constant, which is an improvement over the results of the Lloyd's experiment, where the number of very relevant documents decreased by 10 after the first set of results and kept decreasing afterwards—see the first row in Table IV.

TABLE V. RAL SPACE EVALUATION RESULTS

	Iteration 1	Iteration 2	Iteration 3
Very relevant	16	15	15
Relevant	15	23	24
Non-relevant	19	12	11

As explained above, the 50 documents that we delivered per iteration to RAL Space were divided into academic papers, news articles and standard documents—all of them published between 2010 and 2012, exclusively. The specific breakdown per category and iteration is shown in Table VI.

TABLE VI. RAL SPACE EVALUATION RESULTS PER CATEGORY

First iteration

	Very relevant	Relevant	Non-Relevant
Academic	8	8	6
Standard	6	5	4
News	2	2	9

Second iteration

	Very relevant	Relevant	Non-Relevant
Academic	7	15	5
Standard	7	6	5
News	1	2	2

Third iteration

	Very relevant	Relevant	Non-Relevant
Academic	8	14	4
Standard	5	7	3
News	2	3	4

Due to the involvement of RAL Space in the Q-Detect project, academic papers were considered of particular importance for RAL Space's review. Table VI shows that the number of very relevant academic papers discovered by our prototype decreased only in the second week—decreased by one—but remained almost constant for the entire length of the experiment, which shows the potential of relevance feedback for searches within online journals.

To further evaluate the performance of our prototype, we used precision, one of the most common measures for evaluating the performance of information retrieval systems [38]. Precision is defined as the fraction of retrieved documents that are relevant to the search. For this experiment, we computed precision by considering all the documents evaluated by the analyst as being relevant or very relevant to be at least relevant, and compared these to the total number of documents presented to RAL Space each week—i.e., 50. Table VII displays the precision of our prototype per iteration. The final column shows the overall precision value for the entire experiment—namely, 72%. Note that the precision of the prototype actually increased on a weekly basis. Also note that the number of non-relevant documents—as indicated in Table V—decreased over the experiment, though not by much.

TABLE VII. PRECISION MEASURED PER ITERATION

	Iteration 1	Iteration 2	Iteration 3	Overall
Precision	62%	76%	78%	72%

A possible explanation as to why the number of relevant documents decreased as the Lloyd's experiment progressed is related to the timescale of the evolution of space weather documents on the Web. A period of four weeks might be insufficient to capture a significant number of additional newly published documents on space weather after our first search—i.e., after the first release of queries has been made. Consequently, the very relevant documents retrieved in the first week of the experiment were likely to be the most relevant ones for the entire experimental period of one month. To support this, we were able to verify that most of the documents marked as very relevant by Lloyd's Emerging Risks Group analysts were discovered in the first week of the experiment, but we could not include them in the results for the first week because we were restricted to a maximum of 100 documents per week.

As opposed to the case of the Lloyd's experiment, in the horizon scanning exercise undertaken with RAL Space, where we experimented with the use of relevance feedback, we can confirm that none of the documents delivered to RAL Space in the final iteration was discovered previously, and only two of the relevant documents delivered in the second iteration were discovered in the first week. The reason why we were able to find new documents and maintain the number of very relevant documents per week was that our relevance feedback implementation allowed us to modify the queries to reach different areas of the Web that we would not have been able to approach by releasing the same queries for all the iterations of the experiment.

Ideally, we would have liked to use recall as well to evaluate the performance of the prototype [38]. However, it is infeasible to measure recall for a Web-based system, since it is very difficult to determine all the existing documents on a given topic that are available online at a particular time. In addition, it should be noted that the horizon scanning prototype proposed here is not designed to return all relevant documents, but instead 50 documents per iteration.

VI CONCLUSIONS

Relevance feedback provides a method for reformulating queries based on previously retrieved relevant and non-relevant documents. A simple vector modification process that adds new keywords to queries and scales up or down the importance of existing keywords seems very useful. In view of its simplicity, we recommend that this process should be incorporated into operational text retrieval for horizon scanning systems and applications. Poorly processed feedback may lead to deterioration in retrieval effectiveness, which is a major limitation for relevance feedback implementations, but, when properly employed, the overall precision is improved, as shown in Section V.

As an opportunity for future work, we are considering mining social networks—particularly Twitter [39]—as a potential source of data for horizon scanning work. We are aware of the use of Twitter in financial applications, such as those employed by Derwent Capital Markets [40] and Palantir Technologies [41], whose foundations rely on the work by Bollen et al [42], and we realise that relevant information for horizon scanning that has been published originally by science and technology websites has appeared in Twitter streams. Thus, it is worth contemplating the monitoring of such streams for horizon scanning purposes.

ACKNOWLEDGMENTS

We are very grateful to Michael Jackson—founder member, ambassador and consultant of Shaping Tomorrow—for reading our manuscript and clarifying text on Shaping Tomorrow.

Palomino, Taylor and Depledge are staff at the European Centre for Environment and Human Health—part of the University of Exeter Medical School. This is part financed by the European Regional Development Fund Programme 2007 to 2013 and European Social Fund Convergence Programme for Cornwall and the Isles of Scilly.

REFERENCES

[1] Chan, S.W.K. and J. Franklin, A text-based decision support system for financial sequence prediction. Decision Support Systems, 2011. **52**(1): p. 189-198.
[2] Gheorghiu, R., et al., Web 2.0 and the emergence of future oriented communities. Economic Computation & Economic Cybernetics Studies & Research, 2009. **43**(2): p. p1.
[3] Linstone, H.A. and M. Turoff, Delphi: A brief look backward and forward. Technological Forecasting and Social Change, 2011. **78**(9): p. 1712-1719.
[4] Sutherland, W.J. and H.J. Woodroof, The need for environmental horizon scanning. Trends in Ecology & Evolution, 2009. **24**(10): p. 523-527.
[5] Chairman of the Joint Intelligence Committee, Review of cross-government horizon scanning, 2012, Cabinet Office: London, UK.
[6] Sutherland, W.J., et al., A horizon scan of global conservation issues for 2013. Trends in Ecology & Evolution, 2013. **28**(1): p. 16-22.
[7] Palomino, M.A., et al., Web-based horizon scanning: Concepts and practice. Foresight, 2012. **14**(5): p. 355-373.
[8] Salton, G. and C. Buckley, Improving retrieval performance by relevance feedback. Journal of the American Society for Information Science, 1990. **41**(4): p. 288-297.
[9] RAL Space. RAL Space website. 2013; Available from: http://www.stfc.ac.uk/ralspace/default.aspx.
[10] Robertson, S.E. and K.S. Jones, Relevance weighting of search terms. Journal of the American Society for Information Science, 1976. **27**(3): p. 129-146.
[11] Rocchio, J., Relevance Feedback in Information Retrieval, in The SMART Retrieval System. 1971. p. 313-323.
[12] Harper, D.J., Relevance Feedback in Document Retrieval Systems: An Evaluation of Probabilistic Strategies. 1980: University of Cambridge.
[13] Lv, Y. and C. Zhai. Adaptive relevance feedback in information retrieval. in CIKM '09: Proceedings of the 18th ACM conference on Information and knowledge management. 2009. Hong Kong, China: ACM.
[14] Text REtrieval Conference (TREC). TREC Tracks. 2013; Available from: http://trec.nist.gov/.
[15] Sutherland, W.J., et al., A horizon scan of global conservation issues for 2010. Trends in Ecology & Evolution, 2010. **25**(1): p. 1-7.
[16] Sutherland, W.J., et al., Horizon scan of global conservation issues for 2011. Trends in Ecology & Evolution, 2011. **26**(1): p. 10-16.
[17] Sutherland, W.J., et al., A horizon scan of global conservation issues for 2012. Trends in Ecology & Evolution, 2012. **27**(1): p. 12-18.
[18] Carlsson, P. and T. Jorgensen, Scanning the Horizon for Emerging Health Technologies: Conclusions from a European Workshop. International Journal of Technology Assessment in Health Care, 1998. **14**(04): p. 695-704.
[19] Douw, K., H. Vondeling, and W. Oortwijn, Priority setting for horizon scanning of new health technologies in Denmark: Views of health care stakeholders and health economists. Health Policy, 2006. **76**(3): p. 334-345.
[20] Palomino, M.A., et al., Web-based Horizon Scanning: Recent Developments with Application to Health Technology Assessment. Business Informatics, 2012. **3**(25): p. 139-159.
[21] O'Malley, S.P. and E. Jordan, Horizon scanning of new and emerging medical technology in Australia: Its relevance to Medical Services Advisory Committee health technology assessments and public funding. International Journal of Technology Assessment in Health Care, 2009. **25**(03): p. 374-382.
[22] Palomino, M.A., A. Vincenti, and R. Owen, Optimising Web-based information retrieval methods for horizon scanning. foresight, 2013. **15**(3): p. 159-176.
[23] Palomino, M.A., T. Taylor, and R. Owen. Towards the development of an automated, Web-based, horizon scanning system. in Computer Science and Information Systems (FedCSIS), 2012 Federated Conference on. 2012.
[24] Shaping Tomorrow. Shaping Tomorrow website. 2013; Available from: http://www.shapingtomorrow.com/.
[25] Truvé, S., Big Data for the Future - Unlocking the Predictive Power of the Web, 2011, Recorded Future: Cambridge, MA.
[26] Jackson, M., Personal communication, 2013.
[27] Douw, K., et al., Use of the Internet in scanning the horizon for new and emerging health technologies: A survey of agencies involved in horizon scanning. Journal of Medical Internet Research, 2003. **5**(1): p. e6.
[28] Douw, K., et al., "The future should not take us by surprise": preparation of an early warning system in Denmark. Int J Technol Assess Health Care, 2004. **20**(3): p. 342-350.
[29] Wild, C. and T. Langer, Emerging health technologies: Informing and supporting health policy early. Health Policy, 2008. **87**(2): p. 160-171.
[30] Salton, G. and C. Buckley, Term Weighting Approaches in Automatic Text Retrieval, 1987, Cornell University.
[31] Yahoo! Developer Network. Yahoo! Content Analysis Web Service. 2013; Available from: http://developer.yahoo.com/search/content/V2/contentAnalysis.html.
[32] Wilson, J.C. and D.J. Holland-Smith, White Paper: Dstl S&T horizon scanning, 2008, Defence Science and Technology Laboratory (Dstl).
[33] Google. Custom Search. 2012; Available from: https://developers.google.com/custom-search/.
[34] Purcell, K., J. Brenner, and L. Rainie, Search engine use 2012, 2012, The Pew Research Center's Internet & American Life Project: Washington, DC.
[35] Page, L., et al., The PageRank Citation Ranking: Bringing Order to the Web, in Stanford InfoLab1999.
[36] Q-DETECT. Q-DETECT: Developing tools for on-site phytosanitary inspection. 2013; Available from: http://www.qdetect.org/0_home/index.php.
[37] Lloyd's of London. Emerging Risks Special Interests Group. 2013; Available from: http://www.lloyds.com/the-market/tools-and-resources/research/exposure-management/emerging-risks/emerging-risks-special-interests-group.
[38] Manning, C.D., et al., Introduction to Information Retrieval. 2008: Cambridge University Press. 496.
[39] Twitter. Twitter: The fastest, simplest way to stay close to everything you care about. 2013; Available from: https://twitter.com/about.
[40] Wikipedia. Derwent Capital Markets. 2013; Available from: http://en.wikipedia.org/wiki/Derwent_Capital_Markets.
[41] Palantir Technologies. Palantir. 2013; Available from: http://www.palantir.com/.
[42] Bollen, J., H. Mao, and X. Zeng, Twitter mood predicts the stock market. Journal of Computational Science, 2011. **2**(1): p. 1-8.