ISSN 1507-3858

Marco A. Palomino, Tim Taylor
University of Exeter Medical School, United Kingdom
e-mail: {m.palomino; timothy.j.taylor}@exeter.ac.uk

Geoff McBride
Science and Technology Facilities Council, United Kingdom
e-mail: geoff.mcbride@stfc.ac.uk

Hugh Mortimer
RAL Space, United Kingdom
e-mail: hugh.mortimer@stfc.ac.uk

Richard Owen, Michael Depledge
University of Exeter Business School, United Kingdom
e-mail: {r.j.owen; m.depledge}@exeter.ac.uk

THE IMPACT OF RELEVANCE FEEDBACK ON WEB-BASED INFORMATION RETRIEVAL FOR HORIZON SCANNING APPLICATIONS[1]

Abstract: Horizon scanning is being increasingly regarded as an instrument to support strategic decision making. It requires the systematic examination of data to identify potential threats and opportunities to improve resilience and decrease risk exposure. Horizon scanning may benefit from various retrieval techniques to augment the acquisition of data, 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 the results. As a proof-of-concept demonstration, we have carried out a horizon scanning exercise which showed that our utilisation of relevance feedback for horizon scanning applications was able to maintain the retrieval of relevant documents constant over the entire length of the experiment, without any reduction. This represents an improvement over previous studies where relevance feedback was not considered.

Keywords: horizon scanning, web mining, strategic planning, search engines.

1. Introduction

The use of the World Wide Web for futures research has been gaining increasing attention [Chan, Franklin 2011; Gheorghiu et al. 2009; Linstone, Turoff 2011].

[1] Selected parts of this article were published under non-exclusive copyright in *Proceedings of the Federated Conference on Computer Science and Information Systems FedCSIS 2013* – see [Palomino et al. 2013].

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 [Sutherland, Woodroof 2009]. 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" [Chairman of the Joint Intelligence Committee 2012]. Explicit objectives of horizon scanning are to anticipate issues, accumulate reliable data and knowledge about those issues and thus inform policy making and implementation [Sutherland et al. 2013].

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 he/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. [Palomino et al. 2012], 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* [Salton, Buckley 1990], a technique utilised by some information retrieval systems.

Relevance feedback is an automatic process designed to produce improved query formulations following an initial retrieval operation. It was introduced in the field of information retrieval more than 40 years ago [Chang, Cirillo, Razon 1971; Rocchio 1971], but, to the best of our knowledge, it has never been tested in the context of horizon scanning. The rationale behind relevance feedback consists of choosing important terms, or expressions, attached to certain previously retrieved documents that have been identified as relevant by the users, and of enhancing the importance of these terms in a new query formulation. Analogously, terms included in previously retrieved non-relevant documents could be deemphasised in future query formulations. The effect of such an alteration process is to point the query in the direction of relevant items and away from non-relevant ones, with the expectation of retrieving more wanted and fewer non-wanted items in later searches.

The aim of this paper is to assess the use of relevance feedback as part of a horizon scanning system. The remainder of this paper is organised as follows: Section 2 reviews related work on relevance feedback and briefly outlines previous research on Web-based horizon scanning. Section 3 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 4 discusses two horizon scanning exercises: first, an exercise that was conducted for a European Framework Programme 7 (FP7) project in association with *RAL Space* (RAL Space, 2013) – a world-class space research centre – to review current and future technologies for detecting and monitoring diseases in vegetation; and, second, an exercise carried out in collaboration with Lloyd's of London, a global leader in the insurance market, to use horizon scanning to frame decision making on novel risks. We used the first exercise to assess our implementation of relevance feedback. Section 5 reports on the evaluation of the results of RAL Space's exercise and Section 6 states our conclusions.

2. Related work

Relevance feedback has been extensively studied since its development in the mid-1960s [Harper 1980; Robertson, Jones 1976; Rocchio 1971; Salton, Buckley 1990]. 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 [Robertson, Jones 1976; Rocchio 1971; Salton, Buckley 1990], where the key is to optimally balance the original query and the feedback information [Lv, Zhai 2009] – 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) [Text REtrieval Conference 2013]. 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 [Carlsson, Jorgensen 1998; Douw, Vondeling, Oortwijn 2006; O'Malley, Jordan 2009; Palomino et al. 2012; Sutherland et al. 2012; Sutherland et al. 2011; Sutherland et al. 2013; Sutherland et al. 2010; Sutherland, Woodroof 2009], as the interest in the subject increases, but only a few academic papers describe the methodology to carry out an automated scan [Palomino et al. 2012; Palomino, Taylor, Owen 2012; Palomino, Vincenti, Owen 2013], and the combined use of horizon scanning and relevance feedback has not been documented yet.

Shaping Tomorrow [Shaping Tomorrow 2013] and *Recorded Future* [Recorded Future 2013] 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. 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 [Jackson 2013].

Recorded Future is a "big data" start-up company that uses online data to make predictions about events, people, and entities. Primarily, it serves government intelligence agencies, but it has some private sector clients too [Recorded Future 2013]. Essentially, Recorded Future is established on the premise that all the information available on the Web is useful to support forecasting methods. 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 [Truvé 2011]. As opposed to Recorded Future, we are not interested in predicting future events, 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 [Douw et al. 2003; Douw et al. 2004; Palomino et al. 2012]. However, due to the large amount of information published online, it is difficult to recognise valuable data [Wild, Langer 2008]. In an attempt to establish how exactly the Web should be used in health technology assessments, Douw et al. [Douw et al. 2003] 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.

The basic steps of horizon scanning have been summarised by Choo [Choo 2001] – see Table 1. With Choo's steps in mind, we have developed our own method to scan the horizon, which is largely based on information collected from the Web, and comprises several interlinked components formerly described by Palomino et al. [Palomino et al. 2012]: emerging information, which is relevant to an organisation, is retrieved – manually or otherwise – and/or received – e.g., via selected RSS feeds – from a variety of Web-based sources – such as, news websites and publishers of online scientific and peer-reviewed literature. Key parts of the retrieved information may be extracted and later on categorised in some way – in its simplest form classified within a specific topic area. Afterwards, the information is often archived in a database. Periodically, outputs are presented to either decision makers or more generally through one or more communication mechanisms – typically a report or newsletter.

Table 1. The basic steps of horizon scanning

Managing Information as the Core of the Scanning Function: Six Interrelated Processes	
1.	Identifying information needs. Key groups of users are identified and situations in which they will use the scanning information are thoroughly understood.
2.	Acquiring information. Includes a widely distributed organisation activity where virtually everyone participates. A single locus usually coordinates the information collection network.
3.	Organising and storing information. Usually occurs through a central repository or database where information is structured to facilitate searching, retrieving and browsing.
4.	Developing information products/services. Products should be "relevant" – on topic – but also "right" in focus, orientation, and format, and could include face-to-face briefings, workshops, written reports, and special exhibits.
5.	Using information in three linked areas. Sense-making (what do the external signals mean?); knowledge creating (what knowledge do we need and how can we develop it?) and decision making (what course of action is best for the organization?). Information from scanning is a vital resource in all three arenas.

Source: Horizon Scanning: Gathering Research Evidence to Inform Decision Making [Stonebridge 2008], adapted from Choo [Choo 2001].

It should be noted that scanning sources, either available online or not, are abundant and keep growing incessantly, but the scan budget, the allocated time-frame to generate results and the number of dedicated staff, limit, inevitably, the extent of any horizon scanning exercise. Dedicated scanners continuously search, sift through, and scrutinise information based on pre-established criteria, and then prioritise this information according to its potential impact on the organisation [Albright 2004]. Performance among organisations, from corporations to hospitals, has improved with horizon scanning, particularly when the scanning is carried out continuously [Choo 2001].

3. 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.

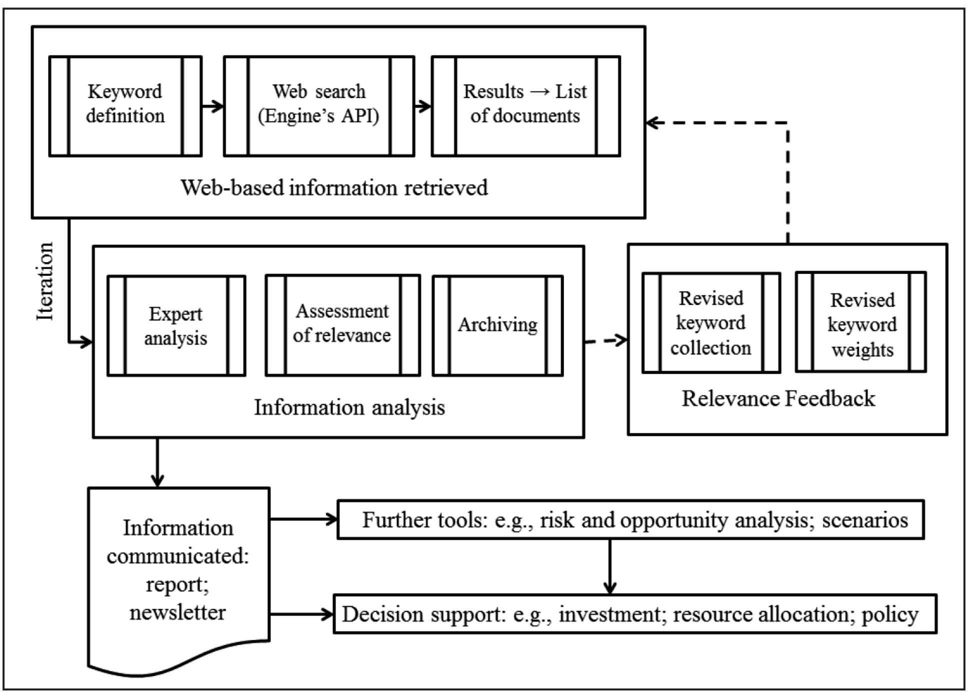

Figure 1. A generalised approach to Web-based horizon scanning for decision support using relevance feedback – based on Palomino et al. [Palomino et al. 2012].

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 [Text REtrieval Conference 2013] and scenario development.

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 the 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 requiring any former analysis of the availability of data on the Web.

(iv) It features a controlled query alteration process designed to emphasise some keywords and de-emphasise 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

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

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

$$Q' = (w'_1, w'_2, \ldots, w'_n),$$

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 a deterioration in retrieval performance [Salton, Buckley 1987]. 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, for the purpose of the scanning exercises described in Section 4, we used *Yahoo!'s Content Analysis Web Service* [Yahoo! Developer Network 2013].

The mechanisms for making relevance assessments are important. Although relevance feedback can improve a search, it is not always the case that the users are willing to make relevance assessments. Partly this may be due to a lack of awareness, on the part of the user, as to how relevance feedback works. But it may also be the case that complex mechanisms to provide feedback discourage the users to make relevance assessments: asking users to spend time marking documents that are not relevant to their search may be difficult to achieve in a practical setting [Ruthven, Lalmas 2003]. In actual conditions, even the first retrieval of documents as a result of a search is expected to provide more relevant information than non-relevant, which can then be refined to improve the retrieval.

3.1. Introducing relevance feedback into horizon scanning

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`, `first time`, and others which have been suggested by the *UK Defence Science and Technology Laboratory* (Dstl) as descriptors of emerging issues [Wilson, Holland-Smith 2008] – see Table 2. 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.

Table 2. Dstl descriptors of emerging issues

breakthrough
closer to reality
first time
groundbreaking
new development
new threat
novel
paves the way
previously impossible
previously unknown
revolutionary
unprecedented
world's first

Source: own elaboration.

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 [Yahoo! Developer Network 2013].

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.

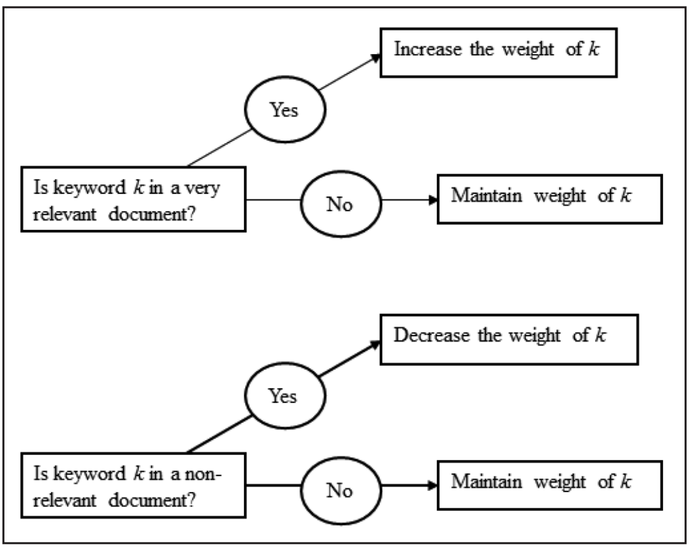

Figure 2. Keyword weight adjustment

Source: own elaboration.

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* [Google 2012]. We chose Google's Custom Search API, because Google is the most popular search engine [Purcell, Brenner, Rainie 2012]; 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.

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. Despite the presence of all these types of documents and formats, Google's method of ranking on the basis of the *PageRank* citation algorithm [Page et al. 1999] 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.

3.2. 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, titles are more descriptive of the contents of a document than the rest of the text [Page et al. 1999].

(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. 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 [Page et al. 1999].

(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 3 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 3. 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

Source: own elaboration.

In order to illustrate the relevance feedback process in detail, we will use an example. The example derives from a horizon scanning exercise executed with the collaboration of RAL Space in October 2012, which is explained in the following section. Then we will analyse the results of this exercise in view of previous work where relevance feedback was not employed.

4. Horizon scanning exercises

4.1. RAL space scanning exercise

In October 2012, RAL Space, based at the *Rutherford Appleton Laboratory* (RAL), undertook a review of current and proposed developments in technologies for the European Framework 7 project *Q-Detect: Developing Quarantine Pest Detection Methods for use by National Plant Protection Organizations (NPPO) and Inspection Services* (Q-DETECT, 2013). RAL Space, part of the *Science and Technology Facilities Council*, works alongside the *UK Space Agency* (UKSA) to perform research and technology development, in the areas of space science [RAL Space 2013].

The motivation behind the review for the Q-Detect project was to assess the potential for technology to be used to detect and monitor non-indigenous invasive plant pests and pathogens, which is a threat to the future of global food security: non-indigenous invasive plant pests and pathogens are the largest threat to the future of global food security [Chakraborty et al. 1998; Gregory et al. 2009; Wittwer 1995] with world-wide economic losses estimated to be in excess of €32.8 billion. In the USA alone, the direct costs from production losses in agriculture and forestry, and indirect costs from management and control, amount to €36.3 billion annually [Oerke 2006; Pimentel, Zuniga, Morrison 2005].

New tools to enable the early detection of infected material and the ability to monitor disease are necessary to minimise the impact of these pathogens and pests. One emerging area, that could provide inspection services with a powerful method of detecting disease is the use of remote sensing techniques. It has been shown that through the use of remote sensing from small unmanned aircraft and Earth observation satellites, it is possible to assess and monitor the health of vegetation on different spatial scales over long periods of time with little human interaction [Hatfield, Pinter 1993; Pinter et al. 2003]. 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 remote detection of quarantined disease. 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 4 – were mostly academic papers chosen by RAL Space. The text of all the abstracts of the academic papers in Table 4 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`.

Table 4. 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" 2001, 88.
Cloutis E.A., *Agricultural crop monitoring using airborne multi-spectral imagery and C-band synthetic aperture radar*, "International Journal of Remote Sensing" 1999, vol. 20, issue 4.
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" 2006, 32.
Lelong C.D., 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" 2008, 8.
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 damag*e, "BioScience" 1986, 36.
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.

Source: own elaboration.

(iii) *Descriptors of emerging issues*, which are keywords defined by Dstl to capture "fresh" information on relevant subjects. Table 2 provides the entire list of descriptors that we employed in this horizon scanning exercise.

Table 5 shows the precise set of keywords that we used 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 5. Initial sets of keywords (RAL Space)

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`

Source: own elaboration.

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 – 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. [Palomino et al. 2013] 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 of 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.

4.2. Lloyd's scanning exercise

In collaboration with *Lloyd's of London*, one of the global leaders in the insurance market, we carried out a study to use horizon scanning to frame decision making on novel risks – specifically risks associated with *space weather* and how these might affect terrestrial and near-Earth insurable assets. Space weather refers to the changing environmental conditions in near-Earth space – examples of space weather events are, for instance, solar flares and coronal mass ejections. It has been demonstrated that space weather has an impact on insurable near-Earth and terrestrial assets – like satellites, power distribution lines, aviation and telecommunications [Hapgood, Thomson 2010]. As part of the study, we benchmarked our prototype against current practice within *Lloyd's Emerging Risks Group* [Lloyd's of London 2013]. The results highlighted not only the potential of Web-based horizon scanning, but also the challenges of undertaking this effectively [Palomino et al. 2013].

We used the report by Hapgood and Thomson [Hapgood, Thomson 2010] as the seed document – the aim of this report was to increase awareness of space weather as a global risk, and explore the threats posed by space weather events to different businesses, drawing attention to ways in which organisations can manage this growing risk. To obtain a list of keywords suitable for this exercise from the seed document, we also employed Yahoo!'s Content Analysis Web Service. Such a list was subsequently presented to Lloyd's Emerging Risks Group, which then selected those that they considered more useful to characterise their interests in space weather from an insurance and risk analysis perspective. Table 6 displays the finalised set of keywords, separated into *context-specific*, which are descriptors of space weather events that may pose risks to the insurance industry – for instance, `Carrington Event` and `solar flare` – and *generic*, which describe impacts of relevance to the insurance industry that the events described by the context-specific keywords may have.

Table 6. Set of keywords (Lloyd's Emerging Risks Group)

Generic keywords	Context-specific keywords
business interruption	proton flares
insurance	space weather
navigation	solar activity
satellite navigation systems	solar wind
air traffic control	solar flares
gps signals	solar maximum
power distribution network	coronal mass ejection
radio transmissions	van allen radiation
power transformers	solar storms
power grids	electromagnetic storms
satellites	galactic cosmic rays
radiation	solar radio bursts
pipelines	
aviation	
oil and mineral industries	
finance	
financial impact	
electricity grids	

Source: own elaboration.

Given that the total number of generic keywords was 18, and the total number of context-specific keywords was 12, the total number of queries automatically released during the exercise was 216 – i.e. each query combined one generic keyword and one context-specific keyword. This still allowed up to 13,824 documents to be retrieved each time we released the queries. In fact, the results showed that between 9,000 and 10,000 unique documents, approximately, were retrieved per week from a daily programmatic search using the keywords in Table 6. Since this would be unmanageable for the Emerging Risks Group to review in the time allocated to this activity, we agreed to restrict the output threshold of the system to 100 documents per iteration for presentation to the Emerging Risks Group.

To reduce the list of documents to a manageable size for the evaluation of the Emerging Risks Group, we sorted and filtered the documents by means of a *measure of importance*. Our hypothesis, which we subsequently tested, was that the documents of most importance – i.e. those of greatest relevance – were the ones that consistently appear at the top of Google's search results. We thus presented a ranked list of documents once a week to Lloyd's analysts, with the ranking being based on the number of times that the document was retrieved by Google over the course of that week – i.e. *cumulative retrieval occurrences* from seven daily programmatic releases of queries. The same approach was used in the RAL Space study, except that the length of the experiment was shorter and thus the number of programmatic releases of queries per iteration was smaller.

Table 7 summarises the two horizon scanning exercises described above in terms of the subject of the scans, their context, the time when they took place and the method that we employed to rank and select the results.

Table 7. Horizon scanning exercises

	Lloyd's Study	RAL Space Study
Subject	Space weather	Remote monitoring of plant diseases
Context	Risk analysis in the insurance industry	Collection of reliable data and knowledge
Time	14 September – 12 October 2010	12-19 October 2012
Method	Cumulative retrieval occurrences from daily programmatic releases of queries	Cumulative retrieval occurrences from programmatic releases of queries aided by relevance feedback

Source: own elaboration.

5. Results

As part of the Lloyd's exercise, we identified several documents that Lloyd's Emerging Risks Group considered very relevant to assess insurance exposure, yet, the number of very relevant documents retrieved per iteration decreased as the experiment progressed, while the number of non-relevant documents increased [Palomino et al. 2013].

Table 8 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 this study and the relevance of the documents was evaluated according to the criteria developed by Lloyd's analysts – see Palomino et al. [Palomino et al. 2013] for full details.

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 RAL Space's 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 remained constant over the length of the experiment.

Table 8. Lloyd's evaluation results

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

Source: own elaboration.

The scanning exercise undertaken with RAL Space comprised three iterations between 12 and 19 October 2012. Table 9 shows the exact number of very relevant, relevant and non-relevant documents retrieved per iteration. Note 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 8.

Table 9. RAL Space evaluation results

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

Source: own elaboration.

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 10.

Due to the involvement of RAL Space in the Q-Detect project, academic papers were considered of particular importance for the review. Table 10 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 [Manning, Raghavan, Schütze 2008]. Precision is defined as the

Table 10. 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

Source: own elaboration.

Table 11. Precision measured per iteration

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

Source: own elaboration.

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 11 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 decreased over the experiment – see Table 9.

5.1. Discussion

A possible explanation as to why the number of very relevant documents was retrieved per iteration in our Lloyd's study, may be the timescale of the evolution of space weather documents on the Web. A period of four weeks may be insufficient to capture a significant number of additional newly published documents on space weather after the first programmatic query release. This would suggest that those very relevant documents retrieved in the first week of the experiment were likely to be the most relevant for the entire experimental period of one month, and might therefore imply that the automated scanning need only be undertaken at these relatively longer intervals – at least for this topic and at this time. Supporting this, Figure 3 displays the number of unique new documents on space weather discovered day-by-day during the experimental period. While the first day of the experiment produced 7,802 unique new documents – any unique document found on the first day was of course considered new – the second day of the experiment produced only 392 unique new documents. From that point on, the number of unique new documents discovered on any given day was below 300 for most of the experiment, as indicated by the horizontal trend-line placed on the graph in Figure 3 exactly at 300 new documents.

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.

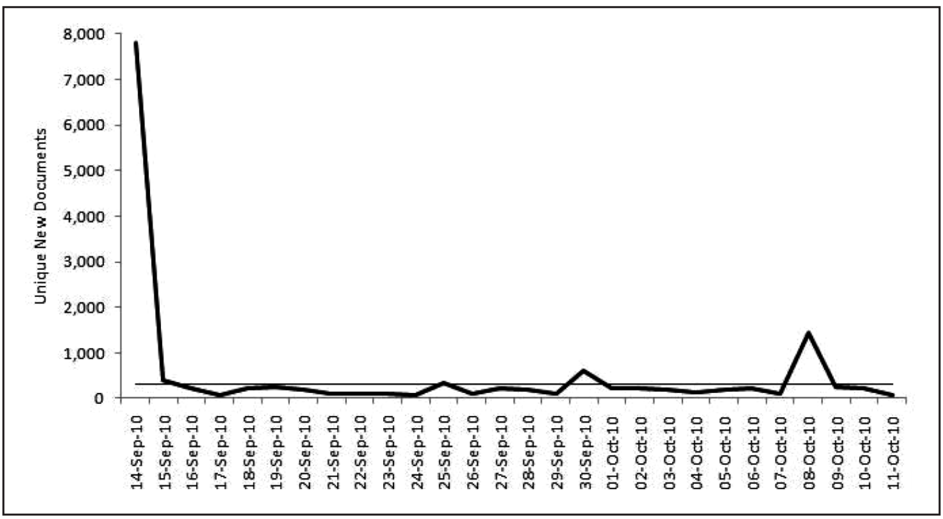

Figure 3. Unique new documents discovered per day over the experiment

Source: own elaboration.

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* [Manning et al. 2008] as well to evaluate the performance of the prototype in both horizon scanning exercises. 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.

6. 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 a 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 5.

As an opportunity for future work, we are considering mining social networks – particularly *Twitter* [Twitter 2013] – 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* [Wikipedia 2013] and *Palantir Technologies* [Palantir Technologies 2013], whose foundations rely on the work by Bollen et al. [Bollen, Mao, Zeng 2011], 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 such streams for horizon scanning purposes.

Web-based horizon scanning offers significant possibilities for the identification of risks and opportunities. Relevance feedback improves the efficiency of the process – and as such will save time for horizon scanners in the longer run if it is appropriately applied.

References

Albright K.S., *Environmental Scanning: Radar for Success*, "Information Management Journal", 2004, 38(3), pp. 38-45.

Bollen J., Mao H., Zeng X., *Twitter mood predicts the stock market*, "Journal of Computational Science" 2011, 2(1), 1-8, http://dx.doi.org/10.1016/j.jocs.2010.12.007.

Carlsson P., Jorgensen T., *Scanning the Horizon for Emerging Health Technologies: Conclusions from a European Workshop*, "International Journal of Technology Assessment in Health Care" 1998, 14(04), pp. 695-704, doi:10.1017/S0266462300012010.

Chairman of the Joint Intelligence Committee, *Review of Cross-Government Horizon Scanning*, London, Cabinet Office, 2012.

Chakraborty S., Murray G.M., Magarey P.A., Yonow T., O'Brien R.G., Croft B.J., Emmett, R.W., *Potential impact of climate change on plant diseases of economic significance to Australia*, "Australasian Plant Pathology"1998, 27(1), pp. 15-35, 10.1071/ap98001.

Chan S.W.K., Franklin J., *A text-based decision support system for financial sequence prediction*, "Decision Support Systems" 2011, 52(1), pp. 189-198, http://dx.doi.org/10.1016/j.dss.2011.07.003.

Chang Y.K., Cirillo C., Razon J., *Evaluation of feedback retrieval using modified freezing, residual collection, and test and control groups*, *processing* [in:] *The SMART retrieval system – experiments in automatic document,* ed. G. Salton, Prentice-Hall, Inc., 1971, pp. 355-370.

Choo C.W., *Environmental scanning as information seeking and organizational learning*, "Information Research" 2001, 7(1).

Douw K., Vondeling H., Eskildsen D., Simpson S., *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), http://dx.doi.org/10.2196/jmir.5.1.e6.

Douw K., Vondeling H., Oortwijn W., *Priority setting for horizon scanning of new health technologies in Denmark: Views of health care stakeholders and health economists*, "Health Policy" 2006, 76(3), pp. 334-345, http://dx.doi.org/10.1016/j.healthpol.2005.06.016.

Douw K., Vondeling H., Sørensen J., Jørgensen T., Sigmund H., *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), pp. 342-350.

Gheorghiu R., Curaj A., Paunica M., Holeab C., *Web 2.0 and the emergence of future oriented communities*, "Economic Computation & Economic Cybernetics Studies & Research" 2009, 43(2), p. 1.

Google, *Custom Search*, https://developers.google.com/custom-search/, 2012.

Gregory P.J., Johnson S.N., Newton A.C., Ingram J.S.I., *Integrating pests and pathogens into the climate change/food security debate*, "Journal of Experimental Botany" 2009, 60(10), pp. 2827-2838, 10.1093/jxb/erp080.

Hapgood M., Thomson A., *Space weather: Its impact on Earth and implications for business: Lloyd's 360 Risk Insight*, 2010.

Harper D.J., *Relevance Feedback in Document Retrieval Systems: An Evaluation of Probabilistic Strategies*: University of Cambridge, 1980.

Hatfield P.L., Pinter Jr, P.J., *Remote sensing for crop protection*, "Crop Protection" 1993, 12(6), pp. 403-413, http://dx.doi.org/10.1016/0261-2194(93)90001-Y.

Jackson M., Personal communication, 2013.

Linstone H.A., Turoff M., *Delphi: A brief look backward and forward*, "Technological Forecasting and Social Change" 2011, 78(9), pp. 1712-1719, http://dx.doi.org/10.1016/j.techfore.2010.09.011.

Lloyd's of London, *Emerging Risks Special Interests Group*, http://www.lloyds.com/the-market/tools-and-resources/research/exposure-management/emerging-risks/emerging-risks-special-interests-group, 2013.

Lv Y., Zhai C., *Adaptive relevance feedback in information retrieval,* Paper presented at the CIKM '09: Proceedings of the 18th ACM conference on Information and knowledge management, Hong Kong, China 2009.

Manning C.D., Raghavan P., Schütze H., *Introduction to Information Retrieval*, Cambridge University Press, 2008.

O'Malley S.P., Jordan E., *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), pp. 374--382, 10.1017/S0266462309990031.
Oerke E.C., *Crop losses to pests*, 10.1017/S0021859605005708, “The Journal of Agricultural Science” 2006, 144(01), pp. 31-43.
Page L., Brin S., Motwani R., Winograd T., *The PageRank Citation Ranking: Bringing Order to the Web*, Stanford InfoLab, 1999.
Palantir Technologies, *Palantir*, http://www.palantir.com/, 2013.
Palomino M.A., Bardsley S., Bown K., Lurio J.D., Ellwood P., Holland-Smith, D., Owen R., *Web-based horizon scanning: Concepts and practice. [General review]*, “Foresight” 2012, 14(5), pp. 355-373.
Palomino M.A., McBride G., Taylor T., Owen R., *Web-based horizon scanning: Recent developments with application to health technology assessment*, “Business Informatics” 2012, 3(25), pp. 139-159.
Palomino M.A., Taylor T., McBride G., Mortimer H., Owen R., Depledge M., *Optimising Web-Based Information Retrieval Methods for Horizon Scanning Using Relevance Feedback*, 2013 Federated Conference on Computer Science and Information Systems (FedCSIS), in press, 2013.
Palomino M.A., Taylor T., Owen R., *Towards the development of an automated, Web-based, horizon scanning system*, Paper presented at the 2012 Federated Conference on Computer Science and Information Systems (FedCSIS), 9-12 Sept. 2012.
Palomino M.A., Vincenti A., Owen R., *Optimising Web-based information retrieval methods for horizon scanning,* “Foresight” 2013, 15(3), 159-176, http://dx.doi.org/10.1108/fs-10-2011-0045.
Pimentel D., Zuniga R., Morrison D., *Update on the environmental and economic costs associated with alien-invasive species in the United States*, “Ecological Economics” 2005, 52(3), pp. 273-288, doi: http://dx.doi.org/10.1016/j.ecolecon.2004.10.002.
Pinter Jr P. J., Hatfield J.L., Schepers J.S., Barnes E.M., Moran M.S., Daughtry C.S.T., Upchurch D.R., *Remote sensing for crop management*, “Photogrammetric Engineering and Remote Sensing” 2003, 69(6), pp. 647-664.
Purcell K., Brenner J., Rainie L., Search Engine Use 2012, The Pew Research Center’s Internet & American Life Project, Washington 2012.
Q-DETECT, *Developing Tools for On-Site Phytosanitary Inspection*, http://www.qdetect.org/0_home/index.php, 2013.
RAL Space, http://www.stfc.ac.uk/ralspace/default.aspx, 2013.
Recorded Future, *Recorded Future: Creating an Insightful World*, https://www.recordedfuture.com/, 2013.
Robertson S.E., Jones K.S., *Relevance weighting of search terms*, “Journal of the American Society for Information Science” 1976, 27(3), pp. 129-146, 10.1002/asi.4630270302.
Rocchio J., *Relevance Feedback in Information Retrieval,* “The SMART Retrieval System” 1971, pp. 313-323.
Ruthven I., Lalmas M., *A survey on the use of relevance feedback for information access systems*, “Knowledge Engineering Review” 2003, 18(2), pp. 95-145.
Salton G., Buckley C., *Term Weighting Approaches in Automatic Text Retrieval*, Cornell University 1987.
Salton G., Buckley C., *Improving retrieval performance by relevance feedback*, “Journal of the American Society for Information Science” 1990, 41(4), pp. 288-297.
Shaping Tomorrow, Shaping Tomorrow website, from http://www.shapingtomorrow.com/, 2013.
Stonebridge C., *Horizon Scanning: Gathering Research Evidence to Inform Decision Making*, The Conference Board of Canada, Ontario 2008.
Sutherland W.J., Aveling R., Bennun L., Chapman E., Clout M., Côté I. M., Watkinson A.R., *A horizon scan of global conservation issues for 2012*, “Trends in Ecology & Evolution” 2012, 27(1), pp. 12-18, http://dx.doi.org/10.1016/j.tree.2011.10.011.

Sutherland W.J., Bardsley S., Bennun L., Clout M., Côte I.M., Depledge M.H., Watkinson A.R., *Horizon scan of global conservation issues for 2011*, "Trends in Ecology & Evolution" 2011, 26(1), pp. 10-16.
Sutherland W.J., Bardsley S., Clout M., Depledge M.H., Dicks L.V., Fellman L., Watkinson A.R., *A horizon scan of global conservation issues for 2013*, "Trends in Ecology & Evolution" 2013, 28(1), pp. 16-22.
Sutherland W.J., Clout M., Côté I.M., Daszak P., Depledge M.H., Fellman L., Watkinson A.R., *A horizon scan of global conservation issues for 2010*, "Trends in Ecology & Evolution" 2010, 25(1), pp. 1-7.
Sutherland W.J., Woodroof H.J., *The need for environmental horizon scanning*, "Trends in Ecology & Evolution" 2009, 24(10), pp. 523-527, http://dx.doi.org/10.1016/j.tree.2009.04.008.
Text REtrieval Conference (TREC), *TREC Tracks*, http://trec.nist.gov/, 2013.
Truvé S., *Big Data for the Future – Unlocking the Predictive Power of the Web*, Recorded Future, Cambridge 2011.
Twitter, *Twitter: The fastest, simplest way to stay close to everything you care about*, https://twitter.com/about, 2013.
Wikipedia, *Derwent capital markets*, http://en.wikipedia.org/wiki/Derwent_Capital_Markets, 2013.
Wild C., Langer T., *Emerging health technologies: Informing and supporting health policy early*, "Health Policy" 2008, 87(2), pp. 160-171, http://dx.doi.org/10.1016/j.healthpol.2008.01.002.
Wilson J.C., Holland-Smith D.J., *White Paper: Dstl S&T horizon scanning: Defence Science and Technology Laboratory (Dstl)*, 2008.
Wittwer S.H., *Food, Climate, and Carbon Dioxide: The Global Environment and World Food Production*, Lewis Publishers, 1995.
Yahoo! Developer Network, *Yahoo! Content Analysis Web Service*, http://developer.yahoo.com/search/content/V2/contentAnalysis.html, 2013.

WPŁYW TRAFNOŚCI SPRZĘŻENIA ZWROTNEGO WYSZUKIWANIA INFORMACJI W SIECI WEB NA APLIKACJE MONITOROWANIA PRZESTRZENI

Streszczenie: Monitorowanie przestrzeni coraz częściej jest traktowane jako instrument wspierania strategicznego podejmowania decyzji. To wymaga systematycznej analizy danych w celu identyfikacji potencjalnych zagrożeń i szans w celu poprawy elastyczności i zmniejszenia ryzyka. Monitorowanie przestrzeni może korzystać z różnych technik wyszukiwawczych, aby poszerzyć możliwość pozyskiwania danych, choć wiąże się to z przeszukiwaniem nowych oraz pojawiających się wcześniej nie znanych zagadnień. Aby zoptymalizować takie poszukiwania, proponujemy użycie odpowiedniego sprzężenia zwrotnego, co wiąże się z interakcją w procesie wyszukiwania informacji w taki sposób, aby poprawić wyniki. Udowadniając koncepcję, zaprezentowano przeprowadzone monitorowania przestrzeni, które wykazały, że nasze wykorzystanie trafności sprzężenia zwrotnego dla aplikacji monitorowania przestrzeni pozwala utrzymać pobieranie odpowiednich dokumentów na całej długości eksperymentu na stałym poziomie. Oznacza to poprawę względem poprzednich badań, w których trafność sprzężenia zwrotnego nie była rozpatrywana.

Słowa kluczowe: monitorowanie przestrzeni WWW, eksploracja w sieci Web, planowanie strategiczne, wyszukiwarki.