Proceedings of the Federated Conference on
Computer Science and Information Systems pp. 1009–1016

ISBN 978-83-60810-51-4

Towards the Development of an Automated, Web-based, Horizon Scanning System

Marco A. Palomino
ECEHH
University of Exeter
Truro, Cornwall, TR1 3HD
Email: m.palomino@exeter.ac.uk

Tim Taylor
ECEHH
University of Exeter
Truro, Cornwall, TR1 3HD
Email: timothy.taylor@pcmd.ac.uk

Richard Owen
Business School
University of Exeter
Exeter, EX4 4PU
Email: r.j.owen@exeter.ac.uk

***Abstract*—Horizon scanning is an increasingly important part of management decision making in all sectors. It involves the systematic search for incipient trends, opportunities and constraints that might affect the probability of achieving management goals and objectives. This requires the continuous acquisition of up-to-date information to anticipate issues, collect data about them and thus inform critical decisions. Although horizon scanning has its roots in the pre-electronic information era, it has blossomed with the availability of electronic databases and Web-based information. In this paper, we propose the implementation of a horizon scanning system centred on the use of keyword-based, Web search engines. By leveraging the existing infrastructure of proven search engines, our system aims to automate the human-intensive process of seeking information and emerging trends. A prototype application that integrates the software that we plan to use has been developed to accompany this paper, and we discuss the potential for its application.**

I. Introduction

HORIZON *scanning* has been defined as *"the systematic search for incipient trends, opportunities and constraints that might affect the probability of achieving management goals and objectives. Explicit objectives of horizon scanning are to anticipate issues, accumulate data and knowledge about them, and thus inform crucial decisions"* [1].

Processes to identify emerging trends in a timely fashion have always been valued in the business environment to acquire foresight. For instance, a *Scan programme* has been in place at *SRI International*, and subsequently *SRI Consulting - Business Intelligence* (SRIC-BI) [2], since 1979. Such a programme evolved as a group process that relied heavily on human expertise to recognise early signals of change, discontinuities, inflection points, and disruptive forces in the business environment. Every month, for at least 25 years, SRIC-BI professionals used to assemble more than 100 abstracts of developments that they perceived to have an impact on the constant confluence of culture, commerce and technology that defines the business environment [3].

Up until the past decade, the scanning processes employed by most organisations have been largely arbitrary, depending on what concerned individuals and leaders are reading, thinking about, or sharing informally with each other. No foresight function, however, can operate with confidence without a disciplined process for spotting patterns of change and highlighting emerging issues for consideration and action [3].

Currently, the contexts within which horizon scanning is carried out vary widely, but it normally involves the use of one of two approaches: *network-based* and *Web-based* horizon scanning. In the case of the first approach, emerging trends, opportunities and constraints are identified via formal meetings, such as conferences and workshops, and informal networking, supplemented by material obtained from the literature and media. The *Cambridge Conservation Initiative* (CCI) [4], for instance, which was founded in 2007, brings together a global panel of academics and specialists annually to identify novel issues in biodiversity conservation. Meeting leading researchers and practitioners is a good way to acquire knowledge about emerging issues in specific areas, such as those related to the CCI. Nevertheless, this can be time and resource consuming, particularly if it is done frequently and it is extended to a number of areas of interest that may require a wider expertise base. In this context, the need for computer-enhanced decision making becomes increasingly apparent. Given the proliferation of automatic tools for processing information, horizon scanning should find a way to reduce costs and improve outcomes.

The *World Wide Web* has been recommended as a source of information for horizon scanning [5]. When used appropriately, the Web can corroborate information, increase the accuracy of forecasts, and augment the amount of useful data [6]. To give an example of the usefulness of Web-based horizon scanning, we may refer to *Google Flu Trends* [7]: by analysing aggregated search data to estimate the occurrences of influenza—this means data based on search terms that *Google* [8] has identified as good indicators of influenza activity—, Google Flu Trends has been able to predict by up to several weeks where influenza outbreaks are most likely to occur on a geographical basis [9]. This information has been successfully used by the *U.S. Centres for Disease Control and Prevention* [10] to support timely decision making and to target resource deployment.

Although Web-based horizon scanning has the potential to be a powerful tool, especially when coupled to risk assessment and management [11], it should be observed that the vast amount of data available on the Web makes it difficult to identify which pieces of information are indeed relevant and credible to inform decisions.

In this paper, we describe the development of an automated horizon scanning system that is based on the use of commercial, keyword-based, Web search engines, and a visualisation toolkit to help with the discovery of emerging trends. By leveraging the existing infrastructure of proven search engines, our system aims to automate the human-intensive process of searching for information and identifying emerging trends. To present results, we have enhanced our system with a visualisation toolkit to aid with the recognition of patterns and enable the exploration of those patterns. A prototype system has been developed to accompany this paper, and we discuss the particulars of its implementation.

The remainder of this paper is organised as follows: Section II reviews the state-of-the-art in horizon scanning and discusses related work. Section III details the implementation of our Web-based horizon scanning prototype system. Section IV comments on the contexts within which we have tested our prototype, and points out other areas of potential application. Finally, Section V states our conclusions.

II. Related Work

In order to make the best possible decisions about their future, governments and businesses are increasingly turning to horizon scanning [12]. As stated above, horizon scanning searches out for early signs of new trends, opportunities, and risks that might become relevant. The results produced by horizon scanning are typically used as an input to policy development, strategic planning and decision-support.

For the public sector, horizon scanning is useful to inform policy makers about risks and opportunities that may arise in the future and require forward planning. Within the UK Government, horizon scanning has its roots in the establishment of the *Office of Science and Technology's Foresight* programme in 1994 [13], following the publication of a white paper by the Chancellor of the Duchy of Lancaster, *"Realising our potential: A strategy for science, engineering and technology"* [14]. In 1999, the *Prime Minister's Performance and Innovation Unit*—later on renamed as the *Strategy Unit* [15]—produced a white paper on modernising government that prompted the creation of the *Strategic Challenges Team*. Their mandate was to identify key challenges facing the UK Government over the next 10 to 20 years, which eventually expanded to include coordinating and benchmarking foresight activities [13].

In the UK, horizon scanning has also been used in procurement, where timely information about future technologies allows for better purchasing decisions—particularly in technology dependent activities such as defence and health. Specifically in terms of defence, the UK *Ministry of Defence* (MOD) has been supporting systematic science and technology scanning using tools and techniques developed by the *Defence Science and Technology Laboratory* (Dstl) since 2006 [16]. Dstl scans a broad collection of technical literature, initially with no bias for particular fields. The scan highlights and disseminates developments that appear to be important for defence and may not otherwise be tracked through the MOD's existing research programme [17].

In the health sector, horizon scanning has helped to identify potentially significant technologies that might become available on the market in the next 5 years [18], [19]. The UK *National Horizon Scanning Centre* (NHSC) [20], for instance, provides advance notice to the UK *Department of Health*, the *National Institute for Health and Clinical Excellence* (NICE) and other policy makers of significant new and emerging health technologies up to 3 years before their launch [21]. Approximately, 1,000 technologies are investigated by the NHSC every year [22].

As described by Packer, the NHSC's identification process of new technologies that have the potential to make an impact on healthcare includes a focused routine scanning and a speciality-based work programme [23]. The focused routine scanning is designed to identify significant and urgent advances, regardless of clinical specialty, and includes the regular review of medical and pharmaceutical research literature, news and financial reports, commercial databases of products in development, liaison with other early warning and horizon scanning systems, commercial companies, and the consultation of manufacturers, clinicians and expert specialist groups [23].

In the case of environmental policy, horizon scanning has been employed to develop scenarios in a recent study on land use futures [24]. To prepare this study, evidence from trends and interviews with experts were used to develop narrative scenarios going ahead to 2060 to capture potential changes and challenges for future land use policy [25].

The UK *Foresight Horizon Scanning Centre* (HSC) was set up in 2005 to explore future issues and trends over the next 50 years that may have an impact on public policy [26]. Two pilot projects were launched by the HSC after its foundation: the *Sigma Scan* [27] and the *Delta (S&T) Scan*. The Sigma Scan—which covered the full public policy spectrum—was produced by scanning the world's leading sources of foresight information, including think-tanks, academic publications, mainstream media, corporate foresight, strategic thinkers, governments, alternative journals, charities, NGOs, blogs, minority communities and futurists. Lately, the Sigma Scan has become a research portal as well as a source of evidence and analysis in its own right. The Delta (S&T) Scan—incorporated into the Sigma Scan since November 2008—was developed as a collaborative editorial project looking at potential developments in science and technology over the next 50 years.

The developments highlighted above in the UK have been mirrored worldwide to varying degrees. In Australia, the *National Horizon Scanning Unit* (NHSU) [28] has been established as part of the *Australia and New Zealand Horizon Scanning Network* (ANZHSN) [29]. The NHSU is a member of the *International Information Network on New and Emerging Health Technologies*, sometimes referred to as *EuroScan* [30]. The primary aim of EuroScan is to share information on selected emerging health technologies, or new applications of existing ones to address their effects and anticipated consequences [31]. EuroScan also supports the exchange of information, experience and research in the field of horizon scanning [32].

With regard to other countries, the *Conference Board of Canada* [12] was recently engaged in exploring the use of horizon scanning as a tool by governments and organisations that have expertise with scanning. The Conference Board of Canada published a report in 2008, which was meant to assess the establishment of a horizon scanning team for the *Health System Planning and Research Branch of the Ontario Ministry of Health and Long-Term Care*.

In Asia, the government of Singapore has a long history of using *scenario planning*[1], but in the wake of several crises, such as SARS, a new methodology had to be taken into consideration. The goal of Singapore's *Risk Assessment and Horizon Scanning* (RAHS) programme is to uncover early-warning indicators of major events to keep policy-makers alerted and prepared for potential strategic-level surprises. RAHS's scanning process is automated by means of a *trawler* that searches the Web for information. The trawler is a piece of software that enriches an analyst's efforts by automating some of the time-intensive, scanning work and allowing them to have more time for analysing, interpreting, and making sense of the collected information. RAHS produces a variety of communications—for instance, *Vanguard*, a monthly bulletin with a detailed look at emerging trends.

In America, a large amount of academic publications on the subject of horizon scanning has been produced. The *Consortium for Science, Policy & Outcomes at Arizona State University* provides online access to plenty of material on horizon scanning research currently being undertaken [34].

In terms of informing new research areas for academia and policy, a range of network-based horizon scanning activities are conducted by research councils and the *European Commission*. Examples include the horizon scanning group of the UK *Environment Research Funders Forum* [35] and the horizon scanning activities of the *Natural Environment Research Council* (NERC) [36]. On the European Commission side, efforts have included the *European Foresight Platform* as a network of networks on foresight activities in Europe [37].

Horizon scanning has also sparked the interest of the private sector, especially to identify emerging changes in the social, natural and technological environment that occur at the margins of, or beyond, a company's current knowledge and planning. *Lloyd's of London*, for example, one of the global leaders in the insurance market, has established an *Emerging Risks Team* [38] to highlight any areas of emerging risks that might affect the insurance industry on a range of different timescales. Likewise, the *Royal Dutch/Shell Group* uses historical and recent information to develop global scenarios that cast light on the areas where Shell operates, identify emerging challenges, and foster adaptability to change. *Shell Global Scenarios to 2025* were built to address an ample variety of strategic and planning needs across the whole spectrum of relevant time horizons and contexts [39].

[1]*Scenario planning*—also called *scenario thinking* or *scenario analysis*—is a strategic planning method that some organisations use to make flexible, long-term plans. Scenario planning is, largely, an adaptation and generalisation of classic methods used by military intelligence [33].

A number of companies offer foresight and futures planning as a service. For example, *Hayes, Inc.* [40] and *ECRI* [41], both located in the United States, have developed health technology assessment and consulting to assist with better healthcare decisions. Other companies, such as *Recorded Future* [42], aim to record what the world knows about the future and make it available for analysis. Recorded Future continuously collects and analyses news content from more than 40,000 Web sources, ranging from media and government websites to individual blogs and selected twitter streams [43]. Recorded Future uses this information to spot historical developments and formulate hypotheses about likely future events. Certainly, our work differs from Recorded Future, because we are not interested in predicting future events, but rather in improving resilience and the capability to react to new risks and opportunities.

Another private company operating a Web-based horizon scanning approach is *Shaping Tomorrow* [44]. Shaping Tomorrow offers services including awareness of breaking information as well as full assistance for companies interested in foresight at a considerable level of detail [44]. The *Future Analysts' Network* [45]—or *FAN Club*—, a forum for those who have an interest in horizon scanning and futures analysis in the UK, is organised by Shaping Tomorrow, and runs a series of regular meetings to exchange ideas, innovative thinking and good practice.

III. System Implementation

Horizon scanning comprises a variety of bespoke approaches to discover emerging issues, potential threats, opportunities and likely future developments. Due to the rapid pace at which change is occurring in recent years, the results of any review on emerging developments becomes obsolete rather quickly. Hence, horizon scanning requires a continuous and systematic approach to ensure the relevance and timeliness of its output.

We aim to improve the way to perform horizon scanning by means of using the Web and automating different steps in the process to identify emerging trends. We realise that other organisations referred to in Section II have worked towards the same goal before, achieving success at varying levels. However, until computers can read, process, integrate, and analyse the full breadth of data available online, Web-based horizon scanning will remain a very human-intensive task [3], especially when it comes to the discovery of new trends. The approach that we are proposing to implement does not intend to eliminate the personal interaction from horizon scanning, but rather to augment such an interaction by taking advantage of electronic sources of information—in particular, the Web.

Figure 1 shows a generalised approach to Web-based horizon scanning for strategic decision support. It emphasises the iterative nature of horizon scanning, noting that the processes of retrieving or receiving documents—i.e., *information retrieval*—, and extracting, categorising, analysing and archiving information are repeated continuously.

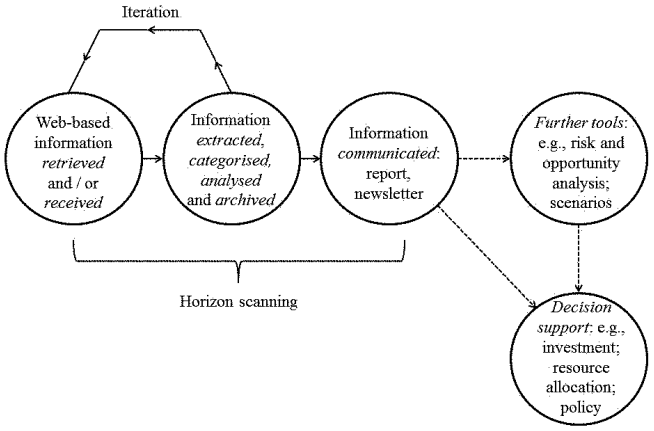

Fig. 1. A generalised, Web-based, horizon scanning approach for decision support—picture obtained from Palomino, *et al.*'s horizon scanning review [46], currently in press.

The outputs of horizon scanning are communicated periodically in the form of newsletters or reports, as indicated in Figure 1. Outputs can be provided directly to decision makers, or interfaced with further tools before this is done. These tools may include more detailed forms of opportunity and risk analysis [11] and scenario development.

In the following subsections, we detail each of the steps included in the particular methodology that we propose to implement an automated, Web-based, horizon scanning system. The system that we propose builds upon some of the suggestions documented by Nie, *et al.* in their practical approach to collecting information that is relevant to the issues of concern within an organisation [47]. Nevertheless, Nie, *et al.*'s approach is limited to electronic journals, whereas ours is potentially open to all Web-based resources.

A. Keyword Selection

To bootstrap the search for emerging information on the Web, a suitable set of *keywords* has to be selected. Keywords are descriptive words or phrases that are mentioned consistently in relation to the issues of concern which will lead the search for emerging information and may help to better characterise them. The selection of keywords may be supported by the utilisation of keyword extraction software, but it requires the additional choice of a reputable source, which may be a keynote, a journal paper or a website from which the keywords will be extracted. We refer to this source as the *seed*, since it provides the starting point for the whole scanning process.

Even though some recently devised keyword-extraction algorithms have proven to be very efficient, the selection of keywords does demand a certain amount of human intervention. An expert in the field has to review the list of keywords extracted automatically, refine the list and ascertain how many keywords should be employed to search for information.

In a previous study carried out in collaboration with the Emerging Risks Team of Lloyd's of London, which looked into the impacts of *space weather* on terrestrial and near-Earth insurable assets[2], we chose two different sets of keywords to search for information [49]. The first set of keywords included *events* associated with space weather—for example, `solar activity`, `solar flares` and `coronal mass ejections`—and the second one comprised the *areas impacted* by space weather that are of interest to the insurance industry—for example, `satellite navigation systems`, `air traffic control` and `electricity grids`. The two sets of keywords were initially determined using a seed document, which in this case was the report produced by the *Space Environment Group* at *RAL Space* [48]. The aim of the seed document was to increase awareness of space weather as a global risk, and explore the threats posed by space weather events to different businesses, highlighting ways in which organisations can manage this growing risk.

To extract keywords from the seed document, we employed *Yahoo!'s Term Extraction Web Service* [50] to generate a preliminary list of keywords. We presented these keywords to members of the Emerging Risks Team. Subsequently, they selected those that they considered useful to characterise their interests in space weather from an insurance and risk analysis perspective. Some new keywords were added too, which were not picked up by the keyword-extraction software, or were not mentioned in the seed document.

Once keywords have been selected, the next step in our search for emerging information consists of using keyword combinations to produce queries that are automatically submitted to commercial, Web search engines. In the case of the study described above, we gathered documents of relevance to the Emerging Risks Team by automatically submitting queries to a search engine with a combination of context-specific and generic keywords. As a result, we obtained documents of relevance to the insurance industry that relate to the risks of space weather—e.g., documents about the impacts and risks of solar flares on power grids.

B. Information Retrieval

We propose the use of commercial, Web search engines to seek information. Considering that one single search engine is not sufficient to search for everything on the Web [51], we combined the results of different engines. For the purpose of implementing a prototype that we could use as a proof-of-concept, we have chosen the following engines: Google, *Bing* [52], and a news service provider called *NewsCred* [53].

According to *Alexa Traffic Rankings*, our choice of search engines contemplates two of the most used commercial search engines in the world: Google and Bing [54]. Both Google and Bing have developed *application programming interfaces* (APIs) for programmatic access to their indexes.

[2]*Space weather* refers to the conditions on the sun and in the solar wind, magnetosphere, ionosphere and thermosphere that can influence the performance and reliability of space-borne and ground-based technological systems and endanger human life or health [48].

An API is a precise set of rules and specifications that a software program should follow to access and make use of the services and resources provided by another software program that implements it [55]. APIs serve as an interface between programs and facilitate their interaction, similar to the way in which the user interface facilitates the interaction between humans and computers. Google provides its *JSON/Atom Custom Search API* as an interface to facilitate the interaction between software programs and Google's repository [56]. Bing provides *Bing Search API* to grant access to Bing's investments in crawling, Web indexing, ranking and relevancy algorithms [57].

To automate our search for relevant documents on the Web, we programmatically released queries containing combinations of keywords via Google's JSON/Atom Custom Search API and Bing's Search API. In the study carried out in collaboration with the Emerging Risks Team of Lloyd's of London described above, we released the queries once per day [49]. However, the frequency of the release can be adapted to the particular needs of different users and cases.

It is well known that the intersection of the results produced by separate search engines is very small. Indeed, Bharat and Broder showed that only 1.4% of the URLs indexed by the four major search engines in 1997—namely, *AltaVista*, *Excite*, *HotBot* and *Infoseek*—were common to all of them [58]. One would expect that the percentage of URLs indexed by the major search engines has improved in recent times, but the truth is that it remains low. A more recent study showed that over 80% of the pages in a major search engine's database exist only in that database [51].

Considering that the intersection of the results produced by Google and Bing is not very high, it is sensible to combine the results produced by Google's JSON/Atom Custom Search API and Bing's Search API together, rather than adopting one and neglecting the other. The use of other services, such as NewsCred, adds further value to the search for information.

NewsCred licenses and curates full text articles, photos and video from over 5,000 global news sources like *The Economist*, *The Guardian*, *Reuters* and *Forbes* [53]. We have used NewsCred to retrieve "fresh" data. Although such data is not likely to be supported by a great deal of scientific evidence, we can guarantee that it is new and emerging.

As in the case of Google and Bing, NewsCred allows programmatic access to news articles, images and videos via an API. The *NewsCred Platform API* [59] is wrapped for an array of technologies—including Java, which is the main programming language for our implementation—and responses are formatted in XML, JSON and RSS.

C. Information Analysis

The objective of the next component of our horizon scanning system is to filter down the references returned by the search engines to a manageable collection of documents that are highly relevant to the identification of emerging trends and their supporting information.

It is useful to think of a Web page in a two-dimensional space defined by two metrics: *relevance* and *authority*. Both dimensions or metrics are important and they need to be considered in a unified view in order to provide adequate insight to the users of a horizon scanning system. While the first dimension, relevance, is specific to a page or even a small section of text contained in it, the notion of authority is most naturally assigned at the website level.

In our previous work undertaken in collaboration with the Emerging Risks Team of Lloyd's of London, we had to sort out and filter the documents that we retrieved from the Web by relevance, in order to present a manageable number of documents for evaluation purposes. Document relevance was determined 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 the search engines' results. We thus presented a ranked list of documents, with the ranking being based on the number of times that a document was retrieved by a particular search engine—namely, Google—over the course of a week. In other words, we delivered to the Emerging Risks Team those documents that Google retrieved at the top of its search results most frequently over each week during the length of the study.

Even though the relevance model helped us to limit the universe of information retrieved to thousands rather than millions of documents, the volume of information collected was still beyond a manageable limit. Therefore, in future work, we plan to use the level of authority of particular websites to filter down the number of documents. Influential websites may or may not be factual experts but nevertheless influence the opinions of others via discussions on particular topics. From a horizon scanning perspective, it is important to identify "expert" websites, since any sentiment that they express can spread far and wide.

In addition to authorities, websites that are very well connected are responsible for the spread of information in the Web. When presented with a large number of posts relevant to a topic, ordering them by the website's influence assists in information analysis, given that it is not feasible to read all the posts that are available.

Since reliable readership information is difficult to obtain, the links between websites are commonly used to determine a site's authority. *Technorati* assigns an authority score to a blog based on the number of other blogs linking to it in the last six months [60]. Similarly, *Blogpulse* ranks blogs based on the number of times they are cited by others over the last 30 days [61]. If we consider the Web as a network of directed edges indicating the links between websites and pages, we can apply more complex measures of prestige from social network analysis. For instance, the authority of a page can be characterised based on the number and authority of other pages that link to it using the well-known *PageRank* algorithm [62], while the influence of a page can be captured by the degree to which the page contributes to the flow of information determined by the *flow-betweenness* algorithm [63].

When working with the Emerging Risks Team of Lloyd's of London, we set the output threshold of the system to 100 documents per week for presentation to the Emerging Risks Team, since this was the number of estimated documents that could be reviewed by an analyst per week, given the time currently allocated by Lloyd's of London to work on horizon scanning. We thus presented a list of 100 documents once per week sorted by the number of times that those documents were retrieved by Google over the course of that week—i.e., cumulative retrieval occurrences from seven daily programmatic releases of queries.

The review of the documents that constituted the output of the system was performed manually by an analyst of the Emerging Risks Team. Comments made by the analyst during the review in relation to the institutions whose websites hosted the documents that we delivered will help us in the future to devise a *measure of authority*—certain documents were considered to be very relevant largely because they were hosted by a source known to be reputable by the analyst, such as the *National Aeronautics and Space Administration* (NASA) or a well-regarded newspaper—e.g., *The Telegraph*.

D. Emerging Trends

Relevance provides a useful way to focus our attention on the documents that we should read. As we read those documents, we naturally synthesise information and identify important, higher-level concepts and trends that summarise the discussions. This is precisely the ultimate objective of a computer-enhanced horizon scanning system: to automate the human-intensive process of detecting and summarising patterns that are emerging in the Web, and it is one of the issues that we aim to pursue in our future work.

Natural language processing approaches can identify collocations of consecutive words like `global` and `warming`. Nevertheless, such occurrences may not be particularly interesting if they are mentioned too frequently—this would mean that they are not "emerging" topics any more. Of greater interest are phrases that occur much more frequently in the past. Such an approach is more likely to capture phrases like `healthcare reform` when they are emerging from the background discussion.

At an even higher level of analysis, document clustering and topic modelling techniques can be used to identify collections of documents expressing cohesive patterns of discussion. Such models can be extended with notions of temporal continuity to provide a view of how dominant themes evolved over time. This can be further improved by incorporating feedback from an expert on which themes to track or discard.

E. Visualisation

We have enriched our prototype with the use of a graphical tool to visually display the relations among the documents that we retrieved and the keywords that we employ to discover them. We expect this graphical representation to provide a more direct and easier way to understand data and uncover hidden patterns behind the data and relationships.

Currently, the output of our horizon scanning prototype system is delivered in the form of *topic maps* [64], a new ISO/IEC standard for describing knowledge structures and associating them with information resources [65]. Topic maps are a kind of combination of subject indexing and semantic network knowledge representation [66]. However, their main purpose is to allow people to navigate from one node to another in a network to broaden or narrow the knowledge of a specific topic.

Topic maps help to visualise the evolution of emerging trends over time, which should support the work of the analysts, by reducing the amount of data that they read and classify, making the process more time effective. For illustration purposes, Figure 2 shows a snapshot of our visualisation toolkit. In Figure 2, we display a set of documents associated with the keyword `pollution`—i.e., a set of documents that contain the keyword `pollution` as part of its text and were retrieved by our horizon scanning system. The documents are represented by text boxes containing their titles, or part of their titles. By appropriately choosing the user-interface options provided by our visualisation toolkit, an analyst can render a picture containing the documents associated with more than one keyword, uncovering relationships that may be hard to highlight when reviewing large lists of documents.

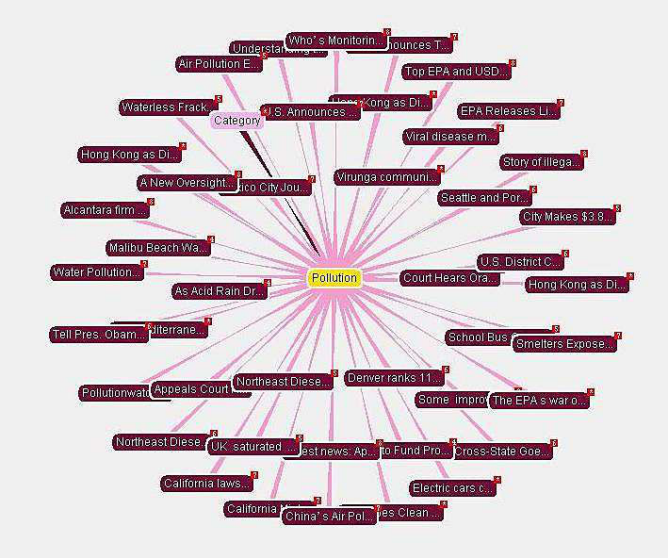

Fig. 2. Visualisation toolkit

Our visualisation toolkit also offers the capability of creating *difference maps*. A difference map is a topic map that shows documents retrieved from the Web at the present date that were not available a few days, weeks or months ago—i.e., a difference map is a representation of the difference between today's topic map and previous topic maps.

IV. Applications

To date we have used the horizon scanning prototype system that we have implemented to inform decision making in three distinct contexts:

- *Risk analysis within the insurance industry*: As stated earlier, our prototype has been used for framing decision making on novel risks in the insurance industry. Future risk analysis applications may include the use of Web-based methods for business intelligence—keeping abreast of the actions of competitors or risks in the supply chain.
- *eHealth*: Our prototype is currently being tested to continuously scan the Web for information on eHealth[3]. We are interested in identifying new developments which are likely to impact the provision of health services, and to draw attention to opportunities for future research.
- *Academia*: Our prototype offers potential for determining research topics and prioritise the research agenda—both in terms of identifying potential "hot topics" for research and informing the writing of research proposals. We are trialling this method within the *European Centre for Environment and Human Health* [70].

V. Conclusion

This paper puts forward an approach centred on commercial, keyword-based, Web search engines for horizon scanning in business, the public sector and academia. Web-based methods have a number of advantages over traditional ones, including significant cost savings and the potential for more regular information feeds than may be possible under other alternatives. With the proliferation of information available on the Web, keyword search using a range of search engines, and suitable methods to determine document relevance, is crucial to ensure that decision makers can be better informed, while not suffering from information overload. Additionally, Web-based methods represent a good opportunity for decision makers in terms of procurement, given the potential reductions in costs that it ensures, and the greater gains from having access to wider opinion than may be possible in the more traditional approaches. Web-based horizon scanning can also be used to complement existing futures strategies and lead to better informed decisions.

We have presented a prototype system that illustrates the Web-based approach that we propose to implement. Formerly, this prototype has been tested in a risk analysis application for the insurance industry. Currently, the prototype is being tested to search for innovative eHealth developments.

Acknowledgment

The European Centre for the Environment and Human Health—part of the Peninsula College of Medicine and Dentistry which is a joint entity of the University of Exeter, the University of Plymouth and the NHS in the South West—is supported by investment from the European Regional Development Fund (ERDF) and European Social Fund (ESF) Convergence Programme for Cornwall and the Isles of Scilly.

[3] *eHealth*—also written *e-health*—is a relatively recent term for healthcare practice supported by electronic processes and communication [67]. Broad definitions of the term cover all electronic processes in health [68], while narrower characterisations use it in the sense of healthcare practice employing the Internet [69].

References

[1] W. J. Sutherland, S. Bardsley, L. Bennun, M. Clout, I. M. Côté, M. H. Depledge, L. V. Dicks, A. P. Dobson, L. Fellman, E. Fleishman, D. W. Gibbons, A. J. Impey, J. H. Lawton, F. Lickorish, D. B. Lindenmayer, T. E. Lovejoy, R. M. Nally, J. Madgwick, L. S. Peck, J. Pretty, S. V. Prior, K. H. Redford, J. P. Scharlemann, M. Spalding, and A. R. Watkinson, "Horizon scan of global conservation issues for 2011," *Trends in Ecology & Evolution*, vol. 26, no. 1, pp. 10 – 16, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S016953471000265X

[2] SRI. (2012, May) SRI International. [Online]. Available: http://www.sri.com/

[3] K. M. Patton, "The role of scanning in open intelligence systems," *Technological Forecasting and Social Change*, vol. 72, no. 9, pp. 1082–1093, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/B6V71-4DXC85J-1/2/4556b430905d1aa4f7c9cb481579daa4

[4] CCI. (2012, May) Cambridge Conservation Initiative. [Online]. Available: http://www.conservation.cam.ac.uk/

[5] A. Stevens, R. Milne, R. Lilford, and J. Gabbay, "Keeping pace with new technologies: Systems needed to identify and evaluate them," *British Medical Journal*, vol. 319, no. 7220, pp. 1291–1294, Nov. 1999.

[6] G. Robert, A. Stevens, and J. Gabbay, "Early warning systems for identifying new healthcare technologies," *Health Technology Assessment*, vol. 3, no. 13, pp. 1–108, 1999. [Online]. Available: http://www.biomedsearch.com/nih/Early-warning-systems-identifying-new/10632623.html

[7] Google Flu Trends. (2012, May) Google Flu Trends. [Online]. Available: http://www.google.org/flutrends/

[8] Google. (2012, May) About Google. [Online]. Available: http://www.google.com/intl/en/about.html

[9] J. Ginsberg, M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, and L. Brilliant, "Detecting influenza epidemics using search engine query data," *Nature*, vol. 457, no. 7232, pp. 1012–4, 2009. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/19020500

[10] CDC. (2012, May) Centers for Disease Control and Prevention. [Online]. Available: http://www.cdc.gov/

[11] R. Owen, D. Baxter, T. Maynard, and M. Depledge, "Beyond regulation: Risk pricing and responsible innovation," *Environmental Science & Technology*, vol. 43, no. 18, pp. 6902–6906, 2009. [Online]. Available: http://pubs.acs.org/doi/abs/10.1021/es803332u

[12] C. Stonebridge, "Horizon scanning: Gathering research evidence to inform decision making," The Conference Board of Canada, Canada, Tech. Rep., March 2008.

[13] W. Schultz, "The cultural contradictions of managing change: Using horizon scanning in an evidence-based policy context," *Foresight*, vol. 8, no. 4, pp. 3–12, 2006.

[14] Chancellor of the Duchy of Lancaster, "Realising our potential: A strategy for science, engineering and technology," UK Government, London, UK, White Paper Cm 2250, May 1993.

[15] Strategy Unit. (2012, June) Prime Minister's Strategy Unit. [Online]. Available: http://webarchive.nationalarchives.gov.uk/20061101004827/cabinetoffice.gov.uk/strategy/

[16] DSTL. (2012, May) Defence Science & Technology Laboratory. [Online]. Available: https://www.dstl.gov.uk/

[17] J. C. Wilson and D. J. Holland-Smith, "Dstl S&T horizon scanning," Defence Science and Technology Laboratory (Dstl), Portsdown West, UK, Tech. Rep., March 2008.

[18] K. Douw, H. Vondeling, D. Eskildsen, and S. Simpson, "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*, vol. 5, no. 1, p. e6, March 2003. [Online]. Available: http://www.jmir.org/2003/1/e6/

[19] K. Douw, H. Vondeling, J. Sorensen, T. Jorgensen, and H. Sigmund, "The future should not take us by surprise: Preparation of an early warning system in Denmark," *International Journal of Technology Assessment in Health Care*, vol. 20, no. 3, pp. 342–50, 2004. [Online]. Available: http://www.biomedsearch.com/nih/future-should-not-take-us/15446764.html

[20] NHSC. (2012, May) National Horizon Scanning Centre. [Online]. Available: http://www.nhsc-healthhorizons.org.uk/

[21] D. Ward, "An early awareness and alert system for health technologies: Methods and outputs of the NHSC," Presentation delivered at the University of Exeter's Horizon Scanning Workshop, July 2011.

[22] M. Fung, S. Simpson, and C. Packer, "Identification of innovation in public health," *Journal of Public Health*, vol. 33, no. 1, pp. 123–130, June 2011.

[23] C. Packer, "The National Horizon Scanning Centre (NHSC): Early warning for new and emerging health technologies in England." *Evidence-Based Healthcare and Public Health*, vol. 9, no. 6, pp. 410–413, 2005. [Online]. Available: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed7&AN=2005539793

[24] Foresight, "Land use futures: Making the most of land in the 21st century," The Government Office for Science, London, UK, Final Project Report 10/631, February 2010. [Online]. Available: http://bis.ecgroup.net/Publications/Foresight/LandUseFutures/10631.aspx

[25] J. Dwyer, "UK Land use futures: Policy influence and challenges for the coming decades," *Land Use Policy*, vol. 28, no. 4, pp. 674–683, Oct. 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0264837710001225

[26] Department for Business, Innovation and Skills (BIS). (2012, May) Foresight Horizon Scanning Centre. [Online]. Available: http://www.bis.gov.uk/foresight/our-work/horizon-scanning-centre

[27] Foresight. (2012, May) The Sigma Scan 2.0. [Online]. Available: http://www.sigmascan.org/Live/

[28] NHSU. (2012, May) National Horizon Scanning Unit. [Online]. Available: http://www.horizonscanning.gov.au/internet/horizon/publishing.nsf/Content/horizon-scanning-reports-2

[29] ANZHSN. (2012, May) Australia and New Zealand Horizon Scanning Network. [Online]. Available: http://www.horizonscanning.gov.au/

[30] EuroScan. (2012, May) EuroScan International Network. [Online]. Available: http://euroscan.org.uk/

[31] C. Packer and S. Simpson, "EuroScan: Status report," European Information Network on New and Changing Health Technologies, Birmingham, UK, Tech. Rep., Jan. 2005. [Online]. Available: http://euroscan.org.uk/mmlib/includes/sendfile.php?id=44

[32] C. Wild and T. Langer, "Emerging health technologies: Informing and supporting health policy early," *Health Policy*, vol. 87, no. 2, pp. 160–171, Aug. 2008. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/18295925

[33] JISC infoNet. (2012, May) Scenario Planning. [Online]. Available: http://www.jiscinfonet.ac.uk/tools/scenario-planning

[34] CSPO. (2012, May) Consortium for Science, Policy & Outcomes at Arizona State University. [Online]. Available: http://www.cspo.org/

[35] ERFF. (2012, May) Environment Research Funders' Forum (ERFF). [Online]. Available: http://www.erff.org.uk/

[36] NERC. (2012, May) Natural Environment Research Council (NERC). [Online]. Available: http://www.nerc.ac.uk/

[37] EFP. (2012, May) European Foresight Platform (EFP). [Online]. Available: http://www.foresight-platform.eu/

[38] Lloyds. (2012, May) Lloyd's Emerging Risks Team. [Online]. Available: http://www.lloyds.com/The-Market/Tools-and-Resources/Research/Exposure-Management/Emerging-risks/Emerging-risks-special-interests-group

[39] Shell, "The Shell global scenarios to 2025 - The future business environment: Trends, trade-offs and choices," Royal Dutch Shell plc., The Hague, The Netherlands, Tech. Rep., June 2005.

[40] Hayes, Inc. (2012, June) Hayes: Transforming healthcare with evidence. [Online]. Available: http://www.hayesinc.com/hayes/

[41] ECRI Institute. (2012, June) Health technology assessment. [Online]. Available: http://www.ecri.org.uk/techasmt.htm

[42] Recorded Future. (2012, May) Recorded Future: Unlock the predictive power of the Web. [Online]. Available: https://www.recordedfuture.com/

[43] S. Truvé, "Big Data for the future: Unlocking the predictive power of the Web," Recorded Future, Cambridge, MA, Tech. Rep., 2011.

[44] ShapingTomorrow. (2012, May) Shaping Tomorrow Limited. [Online]. Available: http://www.shapingtomorrow.com/

[45] Foresight. (2012, June) Future Analysts' Network. [Online]. Available: http://www.bis.gov.uk/foresight/our-work/horizon-scanning-centre/fan-club

[46] M. Palomino, S. Bardsley, K. Bown, J. D. Lurio, P. Ellwood, D. Holland-Smith, B. Huggins, A. Vincenti, H. Woodroof, and R. Owen, "Web-based horizon scanning: Concepts and practice," Foresight Journal–Special Issue on Horizon Scanning Approaches and Methods (in press).

[47] K. Nie, T. Ma, and Y. Nakamori, "An approach to aid understanding emerging research fields—The case of knowledge management," *Systems Research and Behavioral Science*, vol. 26, no. 6, pp. 629–643, 2009. [Online]. Available: http://doi.wiley.com/10.1002/sres.926

[48] M. Hapgood. (2012) Space weather: Its impact on earth and implications for business. [Online]. Available: http://www.lloyds.com/News-and-Insight/Risk-Insight

[49] M. Palomino, A. Vincenti, and R. Owen, "Optimising Web-based information retrieval methods for horizon scanning," foresight Journal (in press).

[50] Yahoo! (2012, May) Yahoo!'s Term Extraction Web Service. [Online]. Available: http://developer.yahoo.com/search/content/V1/termExtraction.html

[51] G. Sterling. (2007, June) Dogpile: Search results on major engines diverging. [Online]. Available: http://searchengineland.com/dogpile-search-results-on-major-engines-diverging-11364

[52] Bing. (2012, May) Bing - Microsoft. [Online]. Available: http://www.bing.com/

[53] NewsCred. (2012, May) Newscred. [Online]. Available: http://www.newscred.com/

[54] I. Alexa Internet. (2012, May) Alexa Traffic Rankings. [Online]. Available: http://www.alexa.com/

[55] Wikipedia. (2012, May) Application Programming Interface (API) - Wikipedia Entry. [Online]. Available: http://en.wikipedia.org/wiki/Application_programming_interface

[56] Google. (2012, May) Google's JSON/Atom custom search API. [Online]. Available: http://code.google.com/apis/customsearch/v1/overview.html

[57] Bing. (2012, May) Bing API, Version 2. [Online]. Available: http://msdn.microsoft.com/en-us/library/dd251056.aspx

[58] K. Bharat and A. Broder, "A technique for measuring the relative size and overlap of public Web search engines," in *Proceedings of the seventh international conference on World Wide Web 7*. Amsterdam, The Netherlands: Elsevier Science Publishers B. V., 1998, pp. 379–388. [Online]. Available: http://dl.acm.org/citation.cfm?id=297805.297863

[59] NewsCred. (2012, May) NewsCred Developer Section. [Online]. Available: http://www.newscred.com/developer

[60] Technorati. (2012, May) Technorati, Inc. [Online]. Available: http://technorati.com/

[61] Wikipedia. (2012, May) BlogPulse - Wikipedia Entry. [Online]. Available: http://en.wikipedia.org/wiki/BlogPulse

[62] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the Web," Stanford InfoLab, Technical Report 1999-66, November 1999. [Online]. Available: http://ilpubs.stanford.edu:8090/422/

[63] U. Brandes and D. Fleischer, "Centrality measures based on current flow," in *Proceedings of the 22nd Annual Conference on Theoretical Aspects of Computer Science*. Berlin, Heidelberg: Springer-Verlag, 2005, pp. 533–544. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-31856-9_44

[64] Standardization IOf. (1999, December) ISO/IEC 13250: Topic maps. [Online]. Available: http://www1.y12.doe.gov/capabilities/sgml/sc34/document/0129.pdf

[65] L. Maicher and J. Park, "Charting the topic maps research and applications landscape," in *First International Workshop on Topic Map Research and Applications, TMRA*. Springer, October 2005. [Online]. Available: http://www.springer.com/computer/ai/book/978-3-540-32527-7

[66] M. Frické, *Logic and the Organization of Information*. New York: Springer, 2012. [Online]. Available: http://books.google.co.uk/books?id=DNfSwaCwyvMC

[67] V. Della Mea, "What is e-health (2): The death of telemedicine?" *Journal of Medical Internet Research*, vol. 3, no. 2, p. e22, June 2001. [Online]. Available: http://www.jmir.org/2001/2/e22/

[68] J.-C. Healy, "Implementing e-health in developing countries: Guidance and principles," ICT Applications and Cybersecurity Division CYB, White Paper, September 2008. [Online]. Available: http://www.itu.int/ITU-D/cyb/app/docs/e-Health_prefinal_15092008.PDF

[69] M. J. Ball and J. Lillis, "E-health: Transforming the physician/patient relationship," *International Journal of Medical Informatics*, vol. 61, no. 1, pp. 1–10, 2001. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1386505600001301

[70] Marco Palomino. (2012, May) Web-based horizon scanning for emergent themes in environment and human health. [Online]. Available: http://www.ecehh.org/publication/web-based-horizon-scanning-emergent-themes-environment-and-human-health