INFORMATYKA EKONOMICZNA BUSINESS INFORMATICS 3(25) · 2012
ISSN 1507-3858

Marco A. Palomino, Tim Taylor, Richard Owen
University of Exeter, the United Kingdom
e-mail: m.palomino@exeter.ac.uk; timothy.taylor@pcmd.ac.uk; r.j.owen@exeter.ac.uk

Geoff McBridge
STFC Futures Program, the United Kingdom
e-mail: geoff.mcbride@stfc.ac.uk

WEB-BASED HORIZON SCANNING: RECENT DEVELOPMENTS WITH APPLICATION TO HEALTH TECHNOLOGY ASSESSMENT[1]

Abstract: Horizon scanning is an increasingly important part of management decision-making. It involves the systematic search for incipient trends, opportunities and constraints that might affect the probability of achieving management goals and objectives. It requires the continuous acquisition of up-to-date information to anticipate issues and thus inform decisions. As an example of the application of horizon scanning in the public sector, the UK National Horizon Scanning Center provides policy-makers with forewarning about emerging and new health technologies. In this paper, we describe the implementation of a horizon scanning system centered 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 implementation has been developed, and we discuss its potential in the context of a case study on proton beam therapy commissioned by the UK Department of Health.

Keywords: horizon scanning, Web data mining, strategic planning, information management, knowledge management.

1. Introduction

Horizon scanning has been defined as "the systematic search for incipient trends, opportunities, challenges and constraints – henceforth 'issues' – that might affect the probability of achieving societal goals and objectives. The objectives of horizon scanning are to anticipate issues, accumulate reliable data and knowledge about them, and thus inform policy making and implementation" [Sutherland et al. 2012].

[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 2012* (see [Palomino, Taylor, Vincenti, Owen 2012]).

Processes to identify incipient trends in a timely fashion have always been valued in the business environment. For instance, a Scan programme has been in place at SRI International, and subsequently SRI Consulting-Business Intelligence (SRIC-BI), since 1979 [SRI 2012]. Such a program evolved as a process that relied heavily on human expertise to recognize early signals of change, discontinuities, inflection points and disruptive forces in the business environment. Every month, for at least 25 years, SRIC-BI professionals assembled 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 [Patton 2005]. SRIC-BI changed its name to Strategic Business Insights (SBI) in November 2009. We cannot confirm whether the same Scan program used by SRIC-BI remains in operation at present, but SBI does employ a methodology – called "organizational radar" – to identify early signals of change [Strategic Business Insights 2012].

Up until the last decade, the scanning processes 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 [Patton 2005].

Currently, the contexts within which horizon scanning is carried out vary widely, but they normally involve 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) [CCI 2012], for example, 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 suitable choice to acquire knowledge about emerging issues in specific areas, such as those of interest to the CCI. Nevertheless, this can be time and resource consuming, particularly if it is done frequently and 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 [Stevens et al. 1999]. When used appropriately, the Web can corroborate information, increase the accuracy of forecasts and augment the amount of useful data available for review [Robert, Stevens, Gabbay 1999]. Consider Google Flu Trends [Google Flu Trends 2012]: by analyzing aggregated search data to estimate the occurrences of influenza – this means data based on search terms that Google [Google Inc. 2012] has identified as 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 [Ginsberg et al. 2009]. This information has been successfully used by the US Centers for Disease Control and Prevention [CDC 2012] to support decision-making and target resource deployment.

Even though Web-based horizon scanning has the potential to be a powerful tool, it should be observed that the vast amount of data available on the Web makes it difficult to identify what is indeed relevant and credible to inform decisions.

We describe here the development of an automated horizon scanning system that is based on the use of commercial, keyword-based, Web search engines and a visualization toolkit developed by Ontopia [Ontopia Forge 2012] 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. The addition of a visualization toolkit aids the recognition of patterns and enables the exploration of those patterns. A prototype system has been developed, and we will discuss the particulars of its implementation. As a proof-of-concept demonstration, we have included a case study to exemplify the use of our prototype system.

The remainder of this paper is organized as follows: Section 2 reviews the state-of-the-art in horizon scanning and discusses related work. Section 3 explains the concept of Web-based horizon scanning and the general methodology followed by this type of scan. Section 4 describes a case study undertaken to review the latest developments in proton beam therapy technology and details the implementation of our prototype system. Section 5 reports on the list of contexts where we have tested our prototype. Section 6 discusses future work and, finally, Section 7 states our conclusions.

2. Related work

In order to make the best possible decisions about the future, governments and businesses are increasingly turning to horizon scanning [Stonebridge 2008]. As stated above, horizon scanning searches for early signs of new trends, opportunities and risks that might become relevant. The results produced by horizon scanning are typically used as 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, horizon scanning has its roots in the establishment of the Office of Science and Technology's Foresight program in 1994 [Schultz 2006], 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* [Chancellor 1993]. In 1999, the Prime Minister's Performance and Innovation Unit – later on renamed as the Strategy Unit [Strategy Unit 2012] – produced a white paper on modernizing 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 [Schultz 2006].

In the UK, horizon scanning has also been used in procurement, where valuable information obtained opportunely 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 [Dstl 2012]. 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 program [Wilson, Holland-Smith 2008].

In the health sector, horizon scanning has helped to identify potentially significant technologies that might become available on the market in the next five years [Douw et al. 2003, 2004]. The UK National Horizon Scanning Center (NHSC) [NHSC 2012], 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 three years before their launch [Ward 2011]. Approximately 1,000 technologies are investigated by the NHSC every year [Fung, Simpson, Packer 2010].

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 program [Packer 2005]. The focused routine scanning is designed to identify significant and urgent advances, regardless of clinical speciality, 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 [Packer 2005].

In the case of environmental policy, horizon scanning has been employed to develop scenarios in a recent study on land use futures [Foresight 2010]. 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 [Dwyer 2011].

The UK Foresight Horizon Scanning Center (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 [BIS 2012]. Two pilot projects were launched by the HSC after its foundation: the Sigma Scan [Sigma Scan 2012] 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. The Sigma Scan has lately become a research portal and 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) [NHSU 2012] has been established as a part of the Australia and New Zealand Horizon Scanning Network (ANZHSN) [ANZHSN 2012]. The NHSU is a member of the International Information Network on New and Emerging Health Technologies, sometimes referred to as EuroScan [EuroScan 2012]. 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 anticipate their consequences [Packer, Simpson 2005]. EuroScan also supports the exchange of information, experience and research in the field of horizon scanning [Wild, Langer 2008].

With regard to other countries, the Conference Board of Canada [Stonebridge 2008] was recently engaged in exploring the use of horizon scanning as a tool by governments and organizations 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,[2] 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) program 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 allowing them to have more time for analyzing and making sense of the collected information.

In the United States, a large number of 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 [CSPO 2012].

In terms of new research and academic areas, 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 [ERFF 2012] and the horizon scanning activities of the

[2] Scenario planning – also called scenario thinking or scenario analysis – is a strategic planning method that some organizations use to make flexible, long-term plans. Scenario planning is, largely, an adaptation and generalization of classic methods used by military intelligence [JISC 2012].

Natural Environment Research Council (NERC) [NERC 2012]. On the European Commission side, efforts have included the European Foresight Platform as a network of networks on foresight activities in Europe [EFP 2012].

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 [Lloyd's… 2011] 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 [Schoemaker, van der Heijden 1992].

A number of companies offer foresight and futures planning as a service. For example, Hayes, Inc. [Hayes 2012] and ECRI [ECRI 2012], 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 [Recorded Future 2012], aim to track down what the world knows about the future and make it available for analysis. Recorded Future continuously collects and analyzes news content from more than 40,000 online sources, ranging from media and government websites to individual blogs and selected twitter streams [Truvé 2011]. 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 [Shaping Tomorrow 2012]. Shaping Tomorrow offers services including awareness of breaking information and full assistance for companies interested in foresight at a considerable level of detail [Shaping Tomorrow 2012]. The Future Analysts' Network [FAN Club 2012] – or FAN Club – a forum for those who have an interest in horizon scanning and futures analysis in the UK, is supported by Shaping Tomorrow, and runs a series of regular meetings to exchange ideas, innovative thinking and good practice.

3. 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, any review on emerging developments becomes obsolete 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 realize that other organizations referred to in Section 2 have worked towards the same goal before, achieving success at varying levels. However, until computers can read, process, integrate and analyze the full breadth of data available online, horizon scanning will remain a very human-intensive task [Patton 2005], 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, taken from Palomino et al.'s horizon scanning review [Palomino et al. 2012], shows a generalized approach to Web-based horizon scanning for strategic decision support. It emphasizes the iterative nature of horizon scanning, noting that the processes of retrieving or receiving documents, i.e. information retrieval, and extracting, categorizing, analyzing and archiving information are repeated continuously. The outputs are then communicated periodically in the form of newsletters or reports. 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 [Owen et al. 2009] and scenario development.

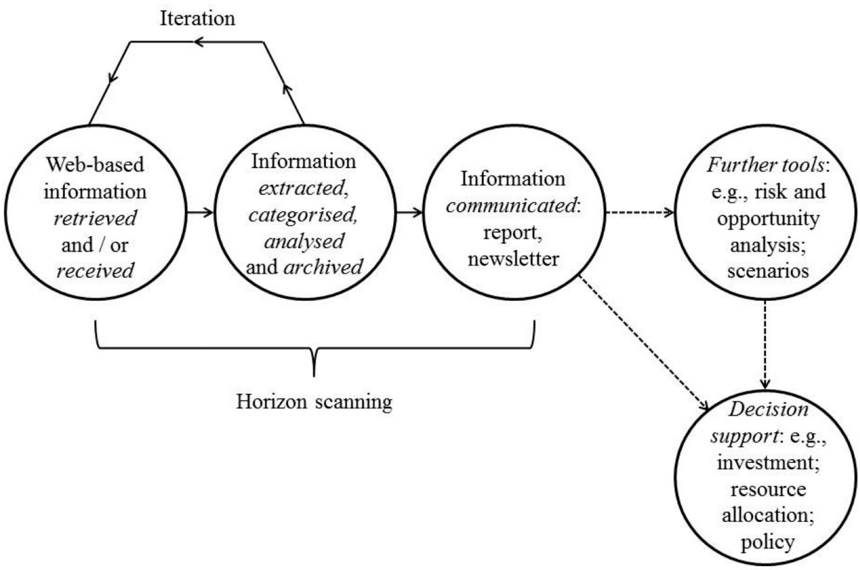

Figure 1. A generalized, Web-based horizon scanning approach for strategic decision support

Source: [Palomino et al. 2012].

In the following subsections, we detail each of the steps included in the particular methodology that we propose to implement a Web-based, horizon scanning system.

Our methodology builds upon some of the suggestions documented by Nie et al. [2009] in their practical approach to collecting information that is relevant to the issues of concern within an organization. Nevertheless, Nie et al.'s approach is limited to electronic journals, whereas ours is potentially open to all Web-based resources. We will exemplify each step of the implementation with references to a case study.

4. Proton beam therapy. A case study

Proton beam therapy (PBT) is an emerging form of radiation therapy that maximizes radiation doses to the target tumors, while preserving adjacent healthy tissue [Levin et al. 2005]. Due to the physics of proton particles, PBT can be directed at cancer cells more accurately than other forms of therapy, which translates into better tumor control, fewer radiation-induced complications and better outcomes overall. PBT was introduced in the United States on an experimental basis in the 1950s, but was not approved as a radiation treatment option by the US Food and Drug Administration (FDA) until 1988, and then only for localized tumors [US Food and Drug… 1988]. In 1990, Loma Linda University opened the first hospital-based proton beam clinic in the United States, followed by the Northeast Proton Therapy Center at Massachusetts General Hospital in 2001 [Jung 2010]. At present, there are several PBT facilities in operation in the world.

Because of the nature of the equipment involved, PBT facilities require sizeable spaces and the construction of suitable centers may be particularly expensive – potentially up to £150 million for a single center, as estimated by the University College London Hospitals – NHS Foundation Trust [University College…]. As part of the process of considering PBT for implementation in the NHS, the Science and Technology Facilities Council (STFC) [STFC 2012] was commissioned by the Department of Health [DH 2012] to undertake a review of proposed developments in PBT technology in June 2011. Specific requirements for this report were listed as follows:

- enumerate current and proposed developments in PBT technology in the market place;
- comment on the technical feasibility of new PBT developments;
- suggest a realistic timescale for new developments and, especially, for the implementation of the necessary equipment for service delivery in 2015;
- identify areas where PBT technology may offer cost savings from other existing systems.

Deliberately, the Department of Health did not engage directly with manufacturers of PBT technological solutions during the preparation of the review. However, the STFC was expected to undertake discussions with them. Indeed, if necessary, the STFC was allowed to enter in non-disclosure agreements to validate assumptions. The goal was to give the Department of Health an unbiased assessment of developments, availability and potential costs compared to existing alternatives. The manufacturers included for this review were IBA, Sumitimo, Varian, Hitachi, Protom, Still River, Tomotherapy, Procure and Siemens.

Considering that decision-making on the uptake and use of emerging health technology needs to be supported by opportune and high quality information [Douw, Vondeling 2006], the STFC made use of horizon scanning to produce the review. Previous attempts to use horizon scanning to identify new or emerging health technologies have been identified by Varela-Lema et al. [Varela-Lema et al. 2012]. Indeed, Varela-Lema et al. have developed a bibliographic search strategy to systematically identify new or emerging health technologies. Whereas their system is based on PubMed [PubMed 2012], the horizon scan that we undertook with the STFC intended to cover the whole Web. The particulars of our work are described below.

4.1. Key word selection

A suitable set of key words to search the Web for information related to PBT technology was selected. Key words are words and phrases that are consistently mentioned in relation to a topic and help to characterize it. We chose our key words with the support of software for the automated extraction of key words from relevant documents. The particular choice of software was Yahoo!'s Term Extraction Web Service [Yahoo! Term Extraction 2012], and the particular choice of documents is listed in Table 1. We refer to these documents as seed documents, because the key words extracted from them provided the starting point for the search process.

Seed documents must be reputable sources. Hence, as the first seed, we selected a piece on PBT, which is part of a textbook on Principles and Practice of Radiation Oncology [Perez, Brady 2008]. We also employed *Wikipedia*'s page on proton therapy [Proton therapy 2012], because it comprises a good summary of background information and relevant references. Lastly, we picked up together with the STFC analysts four webpages derived from the top 10 results retrieved by Google when searching for "proton beam therapy".

We submitted a copy of the text contained in each of the seed documents to Yahoo!'s Term Extraction Web Service to obtain an initial set of key words. We limited this set to the first 100 key words returned by Yahoo!'s Term Extraction Web Service. Then, these key words were presented to the STFC analysts, who selected the key words that they considered more useful.

Two sets of key words were identified, which we termed context-specific and generic key words. Context-specific key words are descriptors of institutions and fields involved in the research and practice of PBT. Context-specific key words may be, for example, anderson cancer center – which is the name of an institution dedicated to cancer patient care, research, education and prevention in the form of various treatments, including PBT – and radiation oncology – which is the medical speciality concerned with prescribing radiation with the intent to cure or for adjuvant therapy.

Generic key words describe areas or issues impacted by the context-specific key words. We refer to them as “generic” because they can be impacted by many other types of healthcare technologies and are not exclusively related to PBT. Examples of generic key words, in the context of the report commissioned by the Department of Health, may be cancer imaging and health benefit. Table 2 displays the finalized set of context-specific and generic key words for our work. Since the key words are later on used to search the Web for documents, we aimed to keep the sets of key words as small as possible; otherwise, the number of documents retrieved may become unmanageable.

To further improve the precision of our search for documents on the Web, we extended the sets of generic and context-specific key words by appending synonym rings. A synonym ring, or synset, is a group of words or phrases that are considered semantically equivalent for the purposes of information retrieval [Belew 2001]. For instance, the synonym ring for the key word cancer treatment, which we considered as a generic key word (see Table 2) contains the phrase oncology treatment. Searching specifically for cancer treatment would be unlikely to allow the retrieval of documents containing, exclusively, the phrase oncology treatment. Similarly, the acronym rsna was added to the synonym ring for the radiological society of north america – it would be unlikely to find documents containing the acronym rsna if we search specifically for documents containing the full name radiological society of north america, which is a context-specific key word in our list (see Table 2). Other words and phrases were added to the synonym rings of the various key words included in Table 1.

4.2. Document retrieval

We employ commercial search engines to search for information on the Web. One single search engine is not sufficient [Sterling 2007], but we combined the results of two different engines: Google and Yahoo! [Yahoo! 2012]

According to Alexa Traffic Rankings, our choice of engines contemplates two of the most used commercial choices in the world [Alexa 2012]. Besides, both Google and Yahoo! have developed freely-available application programming interfaces (APIs) for programmatic access to their indexes.

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 that API [Application programming… 2012]. APIs serve as an interface between software 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 [Google’s JSON/Atom… 2012]. Yahoo! provides Yahoo!’s Search BOSS API [Yahoo! Developer Network 2012] to grant access to Yahoo!’s investments in crawling, Web indexing, ranking and relevancy algorithms.

Table 1. Seed documents

Document title	Location
Principles and Practice of Radiation Oncology – Chapter 18 [Perez, Brady 2008]	Google Books
Proton therapy	http://en.wikipedia.org/wiki/Proton_therapy
The Promise of Proton-Beam Therapy	http://health.usnews.com/health-news/cancer/articles/2008/04/16/the-promise-of-proton-beam-therapy
Proton imaging provides more accuracy, less radiation to pediatric cancer patients therapy	http://cancer scienceblog.com/44850/proton-imaging-provides-more-accuracy-less-radiation-to-pediatric-cancer-patients/
OSU improving its radiation treatment for cancer	http://www.dispatch.com/content/stories/local/2011/04/25/osu-improving-its-radiation-treatment-for-cancer.html
Emory explores new proton therapy facility to offer advanced care to cancer patients	http://www.physorg.com/news/2011-03-emory-explores-proton-therapy-facility.html

Source: own elaboration.

Table 2. Context-specific and generic key words

Context-specific key words	Generic key words
anderson cancer center	cancer imaging
cone beam	cancer patient
proton beam therapy	cancer treatment
proton energy	Center
radiation dosage	Dependency
radiation oncology	health benefit
radiation oncology department	Research
radiation therapy	System
radiation therapy group	
radiological society of north america	

Source: own elaboration.

We programmatically released queries containing combinations of context-specific and generic key words via Google's JSON/Atom Custom Search API and Yahoo!'s Search BOSS API. Given the selection of key words displayed in Table 2, we were able to make 80 different queries in total, by combining all of the context-specific and generic key words. We released the queries once per day, between 15 June 2011 and 13 July 2011. Each query was released first via Google's JSON/Atom Custom Search API and then via Yahoo!'s Search BOSS API, and we kept a record of all the documents retrieved every day.

It is well known that the intersection of the results produced by different 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 [Bharat, Broder 1998].

One would expect that the percentage of URLs indexed by all the major search engines has increased 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 [Sterling 2007].

Since the intersection of the results produced by Google and Yahoo! is not particularly high, it is sensible to combine the results produced by Google's JSON/Atom Custom Search API and Yahoo!'s Search BOSS API, rather than adopting one and neglecting the other.

4.3. Document ranking

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.

Each daily programmatic release of queries containing combinations of context-specific and generic key words via Google's JSON/Atom Custom Search API between 15 June 2011 and 13 July 2011 resulted in 10,000 to 11,000 documents, and between 6,000 and 7,000 using Yahoo!'s Search BOSS API. To reduce the number of documents gathered daily to a manageable collection for review, we sorted them by means of a measure of importance based on two hypotheses:

– First, we hypothesized that the documents retrieved by both engines in their top 10 results are those of most importance, i.e. those of greatest relevance, because both engines have identified them as highly relevant results.
– Second, we hypothesized that the documents whose cumulative retrieval occurrences over the period in which we were releasing our queries daily was higher were the most relevant ones.

We employed a similar strategy to measure the importance of documents in a previous study where we tested our prototype within the context of the insurance industry. The goal of that study was to use our prototype for framing decision-making on novel risks – specifically risks associated with space weather and how these might affect the insurance market. The study was carried out in collaboration with Lloyd's of London – a global leader in the insurance sector. Our measure of importance appeared to significantly improve the number of highly relevant documents retrieved and presented to the risk analysts when benchmarked against their current practice [Palomino, Vincenti, Owen 2012]. Hence, we expected the same measure to yield a satisfying outcome in our search for information associated with PBT technology.

In order to prepare the review commissioned by the Department of Health, we selected those documents that Google and Yahoo! simultaneously retrieved at the top of their search results most frequently between 15 June 2011 and 13 July 2011.

4.4. Document clustering

The number of unique documents retrieved by both engines Google and Yahoo! between 15 June 2011 and 13 July 2011 was 1,277. However, due to time constraints, the analysts at STFC were unable to revise more than 100 documents derived from the Web-based horizon scan undertaken to prepare the review commissioned by the Department of Health. Hence, we limited the output of the Web-based horizon scan to 100 documents. As explained above, we chose those documents that were retrieved more frequently by both engines. Yet, in an attempt to better support the analysts, we clustered the documents automatically into categories. Our goal was to help the analysts to focus on documents that best characterize their interests, so that analysts could focus on particular clusters rather than revising all the documents.

4.5. Visualization

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 key words that we employed to discover them. We expect this graphical representation to provide a more direct and easier way to understand data and uncover hidden patterns behind data and relationships.

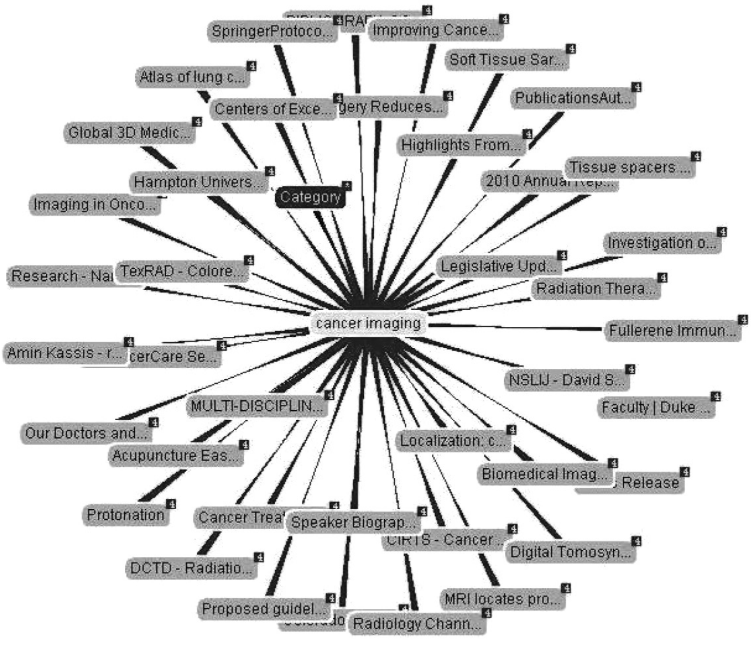

Figure 2. Cancer imaging topic map

Source: own elaboration.

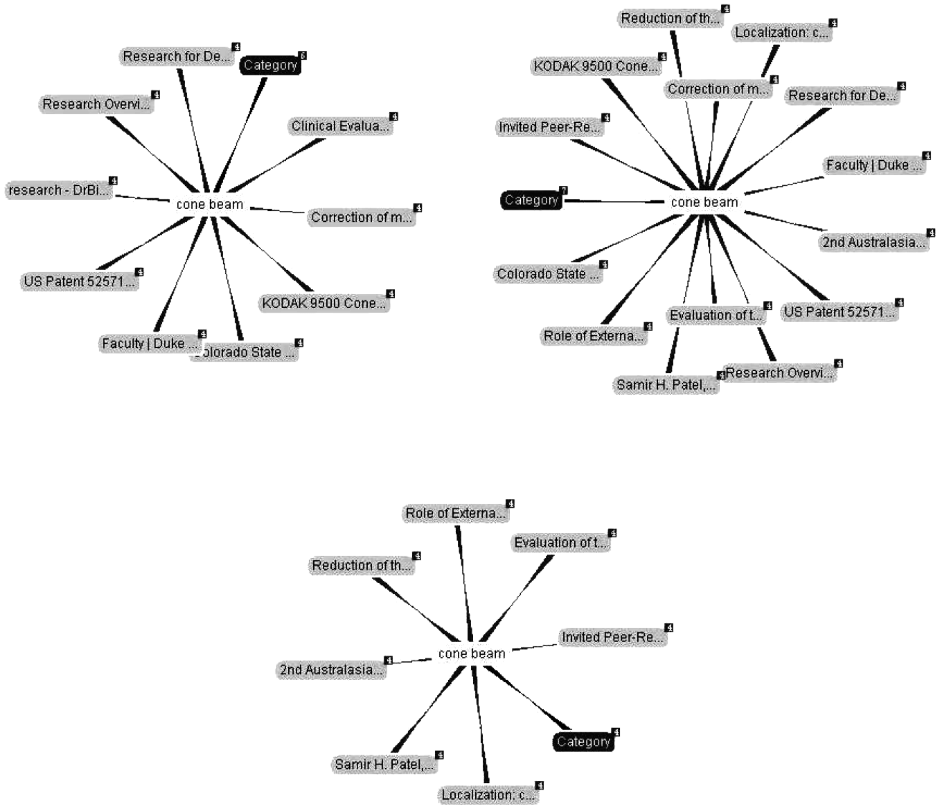

Figure 3. Difference map

Source: own elaboration.

Currently, the output of our horizon scanning prototype system is delivered in the form of topic maps [Standardization IOf 1999…], an ISO/IEC standard for describing knowledge structures and associating them with information resources [Maicher, Park 2005]. Topic maps are a kind of combination of subject indexing and semantic network knowledge representation [Frické 2012]. The particular piece of software that we employ to render topic maps is Ominigator [Ontopia… 2007], a freely-available topic map navigator powered by Ontopia [Ontopia Forge 2012].

Ominigator helps to visualize the evolution of 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 topic map rendered by Omnigator. Specifically, Figure 2 displays a set of documents associated with the key word cancer imaging, i.e. a set of documents retrieved by our prototype that contain the key word cancer imaging as part of their text. The documents are represented by text boxes containing their titles or part of their titles. By appropriately choosing the user-interface options provided by Omnigator, an analyst can render a picture containing the documents associated with more than one key word, uncovering relationships that may be hard to highlight when reviewing large lists of documents.

We also offer 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 the topic map corresponding to a previous moment in time.

For illustration purposes, Figure 3 shows an example of a difference map. Specifically, the difference map can be seen at the bottom of Figure 3, whereas today's map is displayed on the top right hand side of Figure 3, and the previous week's map appears on the top left hand side.

4.6. Outputs

The horizon scanning work that we undertook was used by the STFC as a knowledge baseline. Later on, the STFC formed a small group of subject matter experts from its internal staff and facility user community and generated a review together with the Department of Health, using the knowledge baseline and additional research – the additional research was derived from consultation with specific manufacturers of PBT technology who provided details in the form of whitepapers, brochures, quotes and other relevant material.

The final review highlighted the following issues:

– the current and proposed PBT developments in the market place;
– the technical feasibility of new developments;
– a realistic timescale for new developments and in particular the ability to implement the proposed PBT equipment for service delivery in 2015;
– areas where PBT technology may offer cost savings from existing systems.

The review enabled the PBT Board of the Department of Health to make an intelligence-based decision. In the near future, 1,500 cancer patients a year, approximately, will benefit from the PBT technology that will be available in London, at University College London Hospitals NHS Foundation Trust,and Manchester – at the Christie NHS Foundation Trust Hospital [Department of Health, *Centres*… 2012].

5. Other applications

To date we have used the horizon scanning prototype system that we have implemented to inform decision-making in distinct contexts beyond health technology assessment:

a) **Risk analysis within the insurance industry:** 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. Details of our work on risk analysis within the insurance industry can be found in [Palomino, Taylor, Owen 2012; Palomino, Vincenti, Owen 2012].

b) **Academia:** Our prototype can identify new research topics and help prioritize the research agenda. Currently, we are trialling this within the European Center for Environment and Human Health [ECEHH 2012], which is part of the University of Exeter Medical School.

6. Future work

6.1. Website authority

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 Web-based 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 the work described in Section 4, 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 to the STFC analysts. 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 based on the number of times that a document was retrieved by two search engines – Google and Yahoo! Even though the relevance model limited the information retrieved to hundreds rather than thousands 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, for example, assigns an authority score to a blog based on the number of other blogs linking to it in the last six months [Technorati 2012]. Similarly, BlogPulse ranks blogs based on the number of times they are cited by others over the last 30 days [BlogPulse 2012].

If we consider the Web as a network of directed edges indicating the links between websites and pages, we can apply other measures. For instance, the authority of a page can be characterized depending on the number and authority of other pages that link to it using the PageRank algorithm [Page et al. 1999], while the influence of a page can be captured by the degree to which the page contributes to the flow of information as determined by the flow-betweenness algorithm [Brandes, Fleischer 2005].

6.2. 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 synthesize information and pick up important, higher-level concepts and trends that summarize the discussions. This is precisely the ultimate objective of a computer-enhanced horizon scanning system: to automate the human-intensive process of detecting and summarizing patterns that are emerging on the Web. The detection of emerging trends requires some analysis of the way past trends have emerged. This is an area that needs further 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 modeling 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, and this is one of the strategies that we wish to test in the future.

7. Conclusion

This paper puts forward an approach centered 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, key word 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 horizon scanning approach that we propose to implement. This prototype has been tested in a risk analysis application for the insurance industry (see [Palomino, Taylor, Owen 2012; Palomino, Vincenti, Owen 2012]) and a health technology assessment study which we have described in detail in Section 4.

Currently, the prototype is being tested in an academic environment to discover new research topics and help prioritize the research agenda – both in terms of identifying potential "hot topics" for research and informing the writing of research proposals.

Acknowledgment

The European Center for Environment and Human Health, part of the University of Exeter Medical School, is supported by investment from the European Regional Development Fund and the European Social Fund Convergence Program for Cornwall and the Isles of Scilly.

References

Alexa, Alexa Internet, Inc., 2012, http://www.alexa.com/.

ANZHSN, Australia and New Zealand Horizon Scanning Network, 2012, http://www.horizonscanning.gov.au/.

Application programming interface, *Wikipedia*, 2012, http://en.wikipedia.org/wiki/Application_programming_interface.

Belew R.K., *Finding Out About: A Cognitive Perspective on Search Engine Technology and the WWW*, Cambridge University Press, Cambridge 2001.

Bharat K., Broder A., A technique for measuring the relative size and overlap of public web search engines, [in:] *Proceedings of the 7th International World Wide Web Conference*, Brisbane 1998, pp. 379–388.

BIS, Department for Business, Innovation and Skills, 2012, http://www.bis.gov.uk/foresight/our-work/horizon-scanning-centre.

BlogPulse, *Wikipedia*, 2012, http://en.wikipedia.org/wiki/BlogPulse.

Brandes U., Fleischer D., Centrality measures based on current flow, [in:] *Proceedings of the 22nd Annual Conference on Theoretical Aspects of Computer Science*, Heidelberg 2005, pp. 533–544.

CCI, Cambridge Conservation Initiative, 2012, http://www.conservation.cam.ac.uk/.

CDC, Centers for Disease Control and Prevention, 2012, http://www.cdc.gov/.

Chancellor of the Duchy of Lancaster, *Realising Our Potential: A Strategy for Science, Engineering and Technology*, London, UK, White Paper Cm 2250, 1993.

CSPO, Consortium for Science, Policy and Outcomes at Arizona State University 2012, http://www.cspo.org/.
DH, Department of Health, 2012, http://www.dh.gov.uk/.
Department of Health, *Centres Selected to Host Cutting-Edge Cancer Services*, April 2012, http://mediacentre.dh.gov.uk/2012/04/05/centres-selected-to-host-cutting-edge-cancer-services/.
Douw K., Vondeling H., Selection of new health technologies for assessment aimed at informing decision making: A survey among horizon scanning systems, *International Journal of Technology Assessment in Health Care* 2006, Vol. 22, No. 2, pp. 177–183.
Douw K., Vondeling H., Eskildsen D.,Simpson S., Use of the Internet in scanning the horizon for new and emerging health technologies: Asurvey of agencies involved in horizon scanning, *Journal of Medical Internet Research* 2003, Vol. 5, No. 1, p. 6
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, *International Journal of Technology Assessment in Health Care* 2004, Vol. 20, No. 3, pp. 342–350.
Dstl. Defence Science and Technology Laboratory, 2012, http://www.dstl.gov.uk/.
Dwyer J., UK land use futures: Policy influence and challenges for the coming decades, *Land Use Policy* 2011, Vol. 28, No. 4, pp. 674–683.
ECEHH, European Center for Environment and Human Health, 2012, http://www.ecehh.org/.
ECRI Institute, Health Technology Assessment, 2012, http://www.ecri.org.uk/techasmt.htm.
EFP, European Foresight Platform, 2012, http://www.foresight-platform.eu/.
ERFF, Environment Research Funders' Forum, 2012, http://www.erff.org.uk/.
EuroScan, International Information Network on New and Emerging Health Technologies, 2012, http://www.euroscan.org.uk/.
FAN Club, Future Analysts' Network,2012, http://www.bis.gov.uk/foresight/our-work/horizon-scanningcentre/fan-club
Foresight, *Land Use Futures: Making the Most of Land in the 21st Century*, Final Project Report 10/631, London 2010.
Frické M., *Logic and the Organization of Information*,Springer, London 2012.
Fung M., Simpson S., Packer C., Identification of innovation in public health, *Journal of Public Health* 2010, Vol. 33, No. 1, pp. 123–130.
Ginsberg J. et al., Detecting influenza epidemics using search engine query data, *Nature* 2009, Vol. 457, pp. 1012–1014.
Google Flu Trends, 2012, http://www.google.org/flutrends/.
Google's JSON/Atom Custom Search API, 2012, http://code.google.com/apis/customsearch/v1/overview.html.
Google Inc., Google, 2012, http://www.google.com/intl/en/about.html.
Hayes, Inc., Hayes: Transforming Healthcare with Evidence, 2012, http://www.hayesinc.com/hayes/.
JISC infoNet, *Scenario Planning*, London 2012, http://www.jiscinfonet.ac.uk/tools/scenario-planning.
Jung T., Acting Director, Division of Health Facility Planning, Memorandum: Proton Beam Therapy, March 2010.
Levin W.P., Kooy H., Loeffler J.S., DeLaney T.F., Proton beam therapy, *British Journal of Cancer* 2005, Vol. 93, No. 8, pp. 849–854.
Lloyd's Emerging Risks SIG, Emerging Risks Special Interests Group, 2011, http://www.lloyds.com/The-Market/Tools-and-Resources/Research/Exposure-Management/Emerging-risks/Emerging-risks-special-interests-group.
Maicher L., Park J., *Charting the Topic Maps Research and Applications Landscape: First International Workshop on Topic Map Research and Applications*, Springer, Leipzig 2005.
NERC, Natural Environment Research Council, 2012, http://www.nerc.ac.uk/.
NHSC, National Horizon Scanning Center, 2012, http://www.nhsc-healthhorizons.org.uk/.

NHSU, Adelaide Health Technology Assessment (AHTA), 2012, http://www.adelaide.edu.au/ahta/whatwedo/horscan/.

Nie K., Ma T., Nakamori Y., An approach to aid understanding emerging research fields – the case of knowledge management, *Systems Research and Behavioral Science* 2009, Vol. 26, No. 6, pp. 629–643.

Ontopia Forge, Ontopia, 2012, http://www.ontopia.net/.

Ontopia, Omnigator, 2007, http://www.ontopia.net/omnigator/models/index.jsp.

Owen R., Baxter D., Maynard T., Depledge M., Beyond regulation: Risk pricing and responsible innovation, *Environmental Science and Technology* 2009, Vol. 43, No. 18, pp. 6902–6906.

Packer C., The national horizon scanning centre (NHSC): Early warning for new and emerging health technologies in England, *Evidence-Based Healthcare & Public Health* 2005, Vol. 9, pp. 410–413.

Packer C., Simpson S., *European Information Network on New and Changing Health Technologies: EuroScan – Status Report*, Birmingham 2005.

Page L., Brin S., Motwani R., Winograd T., *The PageRank Citation Ranking: Bringing Order to the Web*, Technical Report: 422, Stanford 1999.

Palomino M.A. et al., Web-based horizon scanning: Concepts and practice, *Foresight* 2012, Vol. 14, No. 5, pp. 355–373.

Palomino M.A., Taylor T., Owen R., Towards the development of an automated, web-based, horizon scanning system, [in:] M. Ganzha, L. Maciaszek, M. Paprzycki (Eds.), *Proceedings of the Federated Conference on Computer Science and Information Systems FedCSIS 2012*, Polskie Towarzystwo Informatyczne, IEEE Computer Society Press, Warsaw, Los Alamitos, CA 2012, pp. 1009–1016.

Palomino M.A., Vincenti A., Owen R., Optimising web-based information retrieval methods for horizon scanning, *Foresight* 2012, Vol. 14, pp. 355–373.

Patton K.M., The role of scanning in open intelligence systems, *Technological Forecasting and Social Change* 2005, Vol. 72, No. 9, pp. 1082–1093.

Perez C.A., Brady L.W., Particle beam radiotherapy, [in:] *Principles and Practice of Radiation Oncology*, 5th ed., Lippincott Williams & Wilkins, Philadelphia 2008, pp. 407–422.

Proton therapy, *Wikipedia*, 2012, http://en.wikipedia.org/wiki/Proton_therapy.

PubMed, PubMed Central, 2012, http://www.ncbi.nlm.nih.gov/pubmed.

Recorded Future, Recorded Future: Unlock the Predictive Power of the Web, 2012, https://www.recordedfuture.com/.

Robert G., Stevens A.J., Gabbay J., "Early warning systems" for identifying new healthcare technologies, *Health Technology Assessment* 1999, Vol. 3, No. 13, pp. 1–108.

Schoemaker P.J.H., van der Heijden C.A.J.M., Integrating scenarios into strategic planning at royal Dutch/shell, *Strategy & Leadership* 1992, Vol. 20, No. 3, pp. 41–46.

Schultz W.L., The cultural contradictions of managing change: Using horizon scanning in an evidence-based policy context, *Foresight* 2006, Vol. 8, No. 4, pp. 3–12.

Shaping Tomorrow, Shaping Tomorrow Limited, 2012, http://www.shapingtomorrow.com/.

Sigma Scan, Foresight Horizon Scanning Center, 2012, http://www.sigmascan.org/.

SRI, SRI International, 2012, http://www.sri.com/.

Standardization IOf,ISO/IEC 13250: Topic Maps, December 1999, http://www1.y12.doe.gov/capabilities/sgml/sc34/document/0129.pdf.

Sterling G., Search Engine Land, 2007, http://searchengineland.com/dogpile-search-results-on-major-engines-diverging-11364.

Stevens A., Milne R., Lilford R., Gabbay J., Keeping pace with new technologies: Systems needed to identify and evaluate them, *British Medical Journal* 1999, Vol. 319, No. 7220, pp. 1291–1294.

STFC, Science and Technology Facilities Council, 2012, http://www.stfc.ac.uk/.

Stonebridge C., *Horizon Scanning: Gathering Research Evidence to Inform Decision Making*, Briefing, Ontario 2008.

Strategic Business Insights, 2012, http://www.strategicbusinessinsights.com/scan/.

Strategy Unit, Prime Minister's Strategy Unit, 2012, http://webarchive.nationalarchives.gov.uk/20061101004827/cabinetoffice.gov.uk/strateg/.

Sutherland W.J. et al., A horizon scan of global conservation issues for 2012, *Trends in Ecology & Evolution* 2012, Vol. 27, No. 1, pp. 12–18.

Technorati, Inc, 2012, http://technorati.com/.

Truvé S., *Big Data for the Future: Unlocking the Predictive Power of the Web*, Technical Report, Cambridge 2011.

US Food and Drug Administration, FDA Home, 22 February 1988.

University College London Hospitals NHS Foundation Trust (UCLH), Proton Beam Therapy [online],University College London Hospitals NHS Foundation Trust (UCLH).

Varela-Lema L., Punal-Riobóo J., Casal Acción B., Ruano-Ravina A., López García M., Making processes reliable: A validated pubmed search strategy for identifying new or emerging technologies, *International Journal of Technology Assessment in Health Care* 2012, FirstView, pp. 1–8.

Ward D., *An Early Awareness and Alert System for Health Technologies: Methods and Outputs of the NHSC*, July 2011.

Wild C., Langer T., Emerging health technologies: Informing and supporting health policy early, *Health Policy* 2008, Vol. 87, No. 2, pp. 160–171.

Wilson J.C., Holland-Smith D.J., *Dstl S&T Horizon Scanning*, Defence Science and Technology Laboratory (Dstl), Portsdown West, White Paper 2008.

Yahoo! Term Extraction, Yahoo! Developer Network, 2012, http://developer.yahoo.com/search/content/V1/termExtraction.html.

Yahoo! 2012, http://www.yahoo.com.

Yahoo! Developer Network, 2012, http://developer.yahoo.com/search/boss/.

MONITOROWANIE PRZESTRZENI WWW: OCENA NOWYCH KIERUNKÓW ROZWOJU TECHNOLOGII MEDYCZNYCH

Streszczenie: Monitorowanie przestrzeni WWW staje się znaczącą funkcją w procesie podejmowania decyzji. Polega ono na systematycznym poszukiwaniu powstających trendów, możliwości i ograniczeń, które mogą mieć wpływ na prawdopodobieństwo osiągnięcia celów zarządzania. Wymaga to ciągłego pozyskiwania aktualnych informacji do przewidywania zagadnień mających wpływ na decyzje. Przykładem instytucji monitorowania WWW w sektorze publicznym jest National Horizon Scanning Center w Wielkiej Brytanii. Zapewnia ono decydentom informację o powstających, a zarazem nowych, technologiach w medycynie. W niniejszym artykule opisano realizację systemu monitorowania przestrzeni WWW wykorzystującą słowa kluczowe (pojęcia) internetowych wyszukiwarek. Opisano możliwości systemu na przykładzie studium przypadku dotyczącego terapii promieniowania protonowego, zrealizowanego na zlecenie Brytyjskiego Ministerstwa Zdrowia.

Słowa kluczowe: monitorowanie przestrzeni WWW, eksploracja danych w sieci Web, planowanie strategiczne, zarządzanie informacją, zarządzanie wiedzą.