Just Google it! Exploring New Web-based Tools for Identifying Public Interest in Science and Pseudoscience

Ayelet Baram-Tsabari
Department of Education in
Technology and Science,
Technion, Israel
ayelet@technion.ac.il

Elad Segev
Research Institute for Law,
Politics and Justice,
Keele University, Keele, UK
e.segev@keele.ac.uk

Abstract

More knowledge of the scientific interests and needs of lay people ("citizen science") is needed for tailoring relevant and engaging formal and free-choice science environments for different publics. However, research-based evidence regarding what the public is interested to know about science is quite rare and superficial. This study aims to utilize two existing web-based tools - *Google Trends* and *Google Zeitgesit* - to identify interests in science and pseudoscience, and to describe their potential and limitations for educational data mining. On a global scale it was found that the share of S&T related searched decrease, both for general and specific search terms, with Global warming being an exception. Some science and pseudoscience terms share a similar level of interest on the global level, but their popularity differs greatly between countries. At the national level it was found that some non-English speaking countries display relatively higher percentage of science related searches. Our analysis reveals that a combination of *Google Trends* and *Google Zeitgesit* is instrumental and useful in examining scientific interests and needs worldwide.

Keywords: public interests of science, data mining, popular searches, Google Trends, Google Zeitgeist.

Introduction

Although science and technology (S&T) are involved in ever increasing parts of our lives, the degree of public ignorance in science revealed by surveys is very distressing, and many do not reliably distinguish between scientific and nonscientific knowledge claims (National Science Board, 2008). The need to create scientifically literate citizenry is a widely accepted educational goal (American Association for the Advancement of Science, 1990; Laugksch, 2000; National Science Board, 2008; Superior Committee on Science Mathematics and Technology Education in Israel, 1992). Yet, after over a decade of efforts to improve scientific literacy among the general population, mainly through the formal school system, surveys suggest that little has been achieved (Miller, 2004).

Interest is a powerful motivator (Deci, 1992), which plays a role in learning through its contribution to connection with the content, allowing to maintain that connection for long enough to be able to learn (Ainley, Hidi, & Berndorff, 2002). Much of what is known about the general public interest in science originates from surveys summarized by the Eurobarometer reports (Eurobarometer, 2005) and the bi-annual American *Science & Engineering Indicators* reports (National Science Board, 2008). When EU citizens are asked what news related issues they are interested in, scientific research is mentioned by 31% (for comparison - Sport heads the

overall ranking of interests at 40%), with medicine and the environment regarded as the most interesting topics in the field of scientific research (Eurobarometer, 2005). Space exploration has consistently ranked low both in the USA and around the world (National Science Board, 2008). Information regarding publics' knowledge, attitudes and concerns towards S&T in general, and specifically biotechnology, nanotechnology and the environment is more prevalent (e.g. Eurobarometer, 2005, 2008; Lujan & Todt, 2000; Macoubrie, 2006; National Science Board, 2008; Scheufele & Lewenstein, 2005; Schutz & Wiedemann, 2008), but research-based evidence regarding what the public is interested to know about science is quite rare and superficial.

When wishing to learn about specific scientific issues, over half of Americans choose the Internet as their main information source (National Science Board, 2008). Therefore, keywords given to search engines can be regarded as resources for detecting the information needs of publics (Murata, 2006; Segev & Ahituv, Forthcoming). This approach was used to study trends and interests worldwide (e.g. Jansen & Spink, 2003; Jansen, Spink, & Saracevic, 2000; Silverstein, Henzinger, Marais, & Moricz, 1999). This study aims to utilize existing web-based tools in order to identify interests in science and pseudoscience as reflected by search queries, and to describe the potential and limitations of these tools for educational data mining. Another contribution of this study is in providing a cross-national comparison of popular science- and pseudoscience-related searches, a comparison that has not yet been conducted in previous analyses of search queries.

Research Approach and Research Tools

Public interest in science and technology was traditionally measured using massively-distributed top-down surveys (e.g. Eurobarometer, 2005; National Science Board, 2008). Other, more learner/audience-centered data-collection methods, such as purchase patterns of books (Schummer, 2005) and self-generated questions sent to a popular science TV-show (Baram-Tsabari & Yarden, 2007, 2008) were also sporadically used.

The widely increasing use of the Internet for researching S&T related interests by the lay public calls for a closer look at existing web-based tools. *Google* is the most widely used search engine today with more than 500 million different users every month (comScore, 2008; Kopytoff, 2007; Sullivan, 2006). Being the most widely used search engine on the web, we chose *Google* and its' advanced features as our data source, using analysis of millions of search queries daily to learn about the Zeitgeist, "the spirit of the times", with regard to public interest in science. In particular, a combination of two data sources was used:

I. *Google Trends*. *Google Trends* (www.google.com/trends) analyzes and displays the share of searches done for the terms entered out of the total number of searches done on *Google* over a period of time (between 2004 and 2008). *Google Trends* shows also how frequently topics have appeared in *Google News* stories, and in which countries the share of searches is higher.
II. *Google Zeitgeist*. *Google Zeitgeist* (www.google.com/press/intl-zeitgeist.html) displays a monthly report on the most popular search queries in different countries. Zeitgeist by Country displays either the top gaining queries in a domain (the searches that have gained the most popularity with respect to the previous month) or the top search queries in a domain.

Choosing specific search queries. In order to use *Google Trends* one has to identify specific search queries to be compared. A list of potential science-related search queries was constructed

based on terms appearing in the Eurobarometer and Science and Engineering Indicators reports (Eurobarometer, 2005; National Science Board, 2008). Pseudoscience-related terms were extracted from relevant Internet sites. In addition, the extensive database of popular searches in *Google Zeitgeist*, providing a list of more than 8000 popular search queries from different countries since 2004, was examined, and all scientific and pseudoscientific search terms were identified. For a list of search queries used see Table 1.

Table 1. Search queries used in the study

| Topic	Search queries
General science words	Medicine, Physics, Biology, Astronomy, Chemistry, Biotechnology
Environment	Global warming, Climate change, Pollution
Biotechnology	Cloning, Stem cells, DNA
Health and Medicine	Antibiotics, Vaccine, Bacteria, Gene therapy
Alternative medicine	Acupuncture
Space exploration	NASA, Astronomy, Solar system, Big bang
Nanotechnology	Nanotechnology
Evolution	Evolution, Darwin, Intelligent design, Creationism
Modern physics	Quantum mechanics, Quantum computing
Miscellaneous	Radioactive, Dinosaur, Einstein, Newton, Britney Spears
Pseudoscience	Homeopathy, Astrology, Reiki, Horoscope

Results and Discussion

Three general science-related search queries: "physics", "biology" and "chemistry", were compared. These search queries were most popular during the Noble-prize winners announcements in October of each year, and least popular during the winter and summer vacations (see Figure 1). However, countries differed widely in the percentage of searches for these general science related terms out of the total searches. These science-related searches were more prevalent in the Philippines and India than in other countries. Although English is not the first language in either of these countries, the share of these general science related terms was higher than in English-speaking countries. One possible reason is the high percentage of students and educated users among *Google* users in these countries. The relatively low percentage of online users in the Philippines (15.4%), and India (5.3%) (Internet World Stats, 2008) supports this assumption (Segev & Ahituv, Forthcoming). This trend is further supported by findings from *Google Zeitgeist*, in which India had the highest percentage of education related searches among popular search queries in the world during 2004 and 2005, with more than 9 percent of the searches being education-related. This is compared to less than 1 percent in most other countries. In Israel the terms "physics", "biology" and "chemistry" (in Hebrew: "פיזיקה", "ביולוגיה", "כימיה") were most popular in the weeks preceding the matriculation exams.

The search terms "physics", "biology" and "chemistry" demonstrated a decrease in search volume during the years 2004-2008. This is probably due to the general increase in access and use of the Internet, which allows wider, less-educated population to search the web via *Google* for increasingly various purposes. Therefore, the search volume for specific S&T-related terms decreases relatively to all the searches conducted using *Google*. This trend is also apparent when searching for emerging technologies such as "quantum computing" (118 times less popular search term than "physics") or a specific professional search term such as "quantum mechanics" (33 times less popular search term than "physics") – both demonstrates a decrease in search volume relatively to the whole search volume using *Google*.
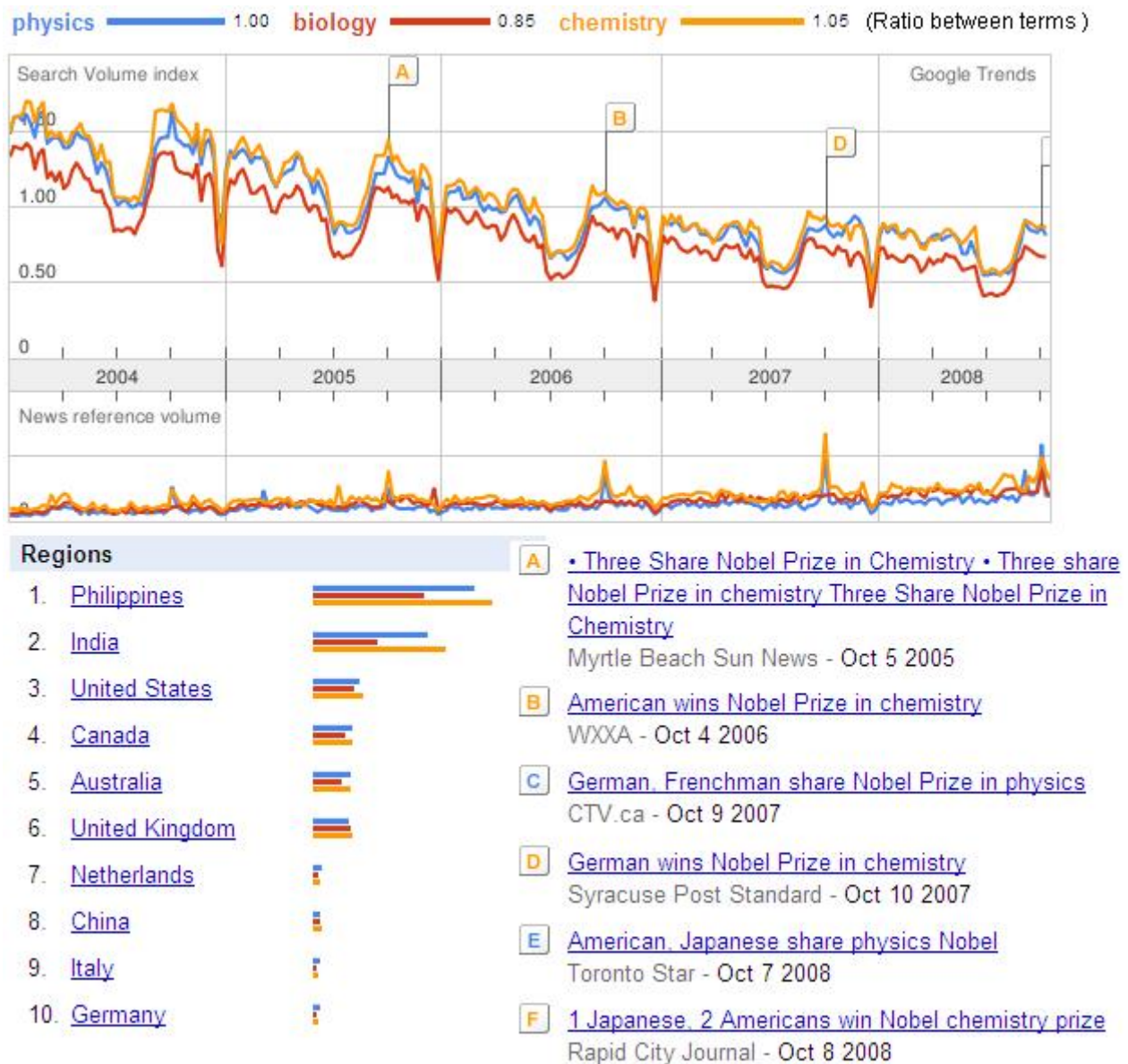
Figure 1. Trends in general science searches

Google Trends may be used to validate and support trends described by the literature in certain science and pseudoscience topics. A tempting comparison of the search queries "astrology" and "astronomy", both topics which were found to be of high interest to the lay public (National Science Board, 2008), showed dominance of searches for "astrology" over "astronomy" (see Figure 2). "astrophysics" is a far less popular search term than "astronomy", less popular even than the pseudoscience search query "homeopathy". The number of news items cataloged by *Google News*, however, is higher for "astronomy" (see Figure 2, middle part). India displays the highest share of searches for "astrology" (see Figure 2, bottom part). The USA was ranked in the sixth place with regard to searches for "astrology".
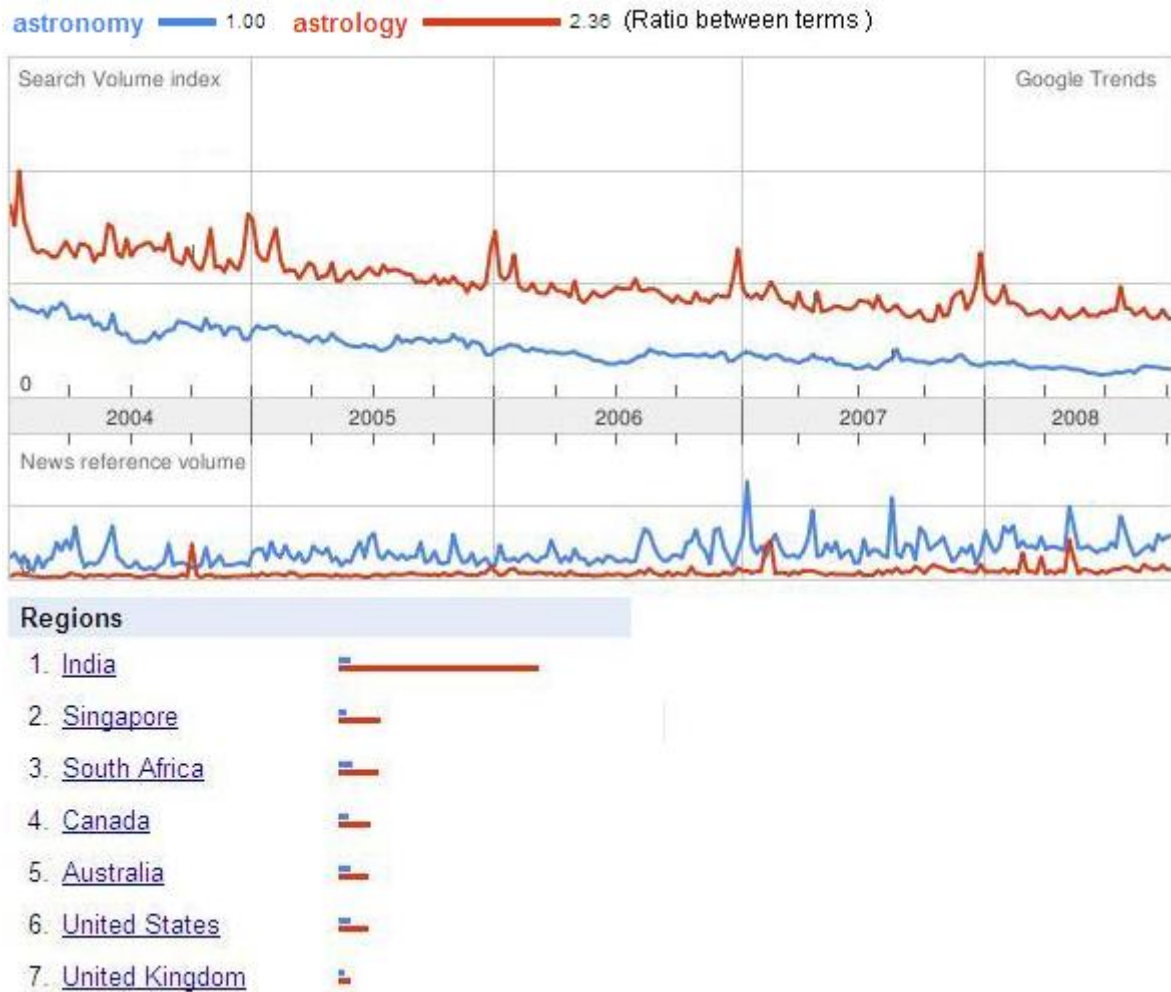
Figure 2. Trends in science- and pseudoscience-related searches

However, more specific terms for Astronomy- and Astrology-related information are used in different languages, and even English-speaking users may use different and often more specific search terms. *Google Zeitgeist* provides a possible methodological-remedy by pointing out the appropriate search terms used in different countries to search for these topics. Table 2 summarizes the most popular searches related to Astrology as appeared in *Google Zeitgeist* during 2004 and 2005.

Table 2. Astrology related popular searches according to *Google Zeitgeist*

| Country	Search Query	Share of searches (%)
Germany	Horoskop	3.4
Japan	占い (fortune telling)	2.7
Spain	Horoscopo	1.8
Brazil	Horoscopo	1.1
Italy	Oroscopo	0.9
Denmark	Horoskop	0.6
UK	horoscope	0.6

According to *Google Zeitgeist* many people use the term "horoscope" in different variations while referring to Astrology. In this case, *Google Zeitgeist* is a more useful tool for conducting a cross-national comparison, suggesting that Germany and Japan display relatively more popular searches related to Astrology than other countries. Similarly, the term "astronomy" is very general. People around the world use more specific terms such as "mars" and "NASA" to find Astronomy-related information (see Table 3).

Table 3. Astronomy related popular searches according to *Google Zeitgeist*

Country	Search Query
USA	Mars, nasa, spirit rover
UK	Mars
Germany	Mars, nasa, Saturn
Spain	Nasa
France	Mars
Italy	Marte
Norway	Mars
Russia	Первый человек на луне (first person on the moon)
India	Nasa

A comparison of the terms "horoscope" and "mars"/"NASA" on *Google Trends* reveals that all three terms display very similar portion of the search queries submitted on *Google*. However, searches for "NASA" and "mars" are influenced by events, such as the launch of Discovery, while searches for "horoscope" are more stable over time.

As mentioned earlier, all science-related search terms showed a decline in popularity since 2004. An exception of this rule-of-thumb was the search term "global warming", which increased in volume, and even became more popular than "astrology" and "reiki" in 2007 (see Figure 3). This finding is in agreement with real-world surveys which found that Americans have recently become somewhat more concerned about environmental quality (National Science Board, 2008), and that among Europeans, climate change is ranked as a top environmental concern (Eurobarometer, 2008). The growth in searches for "global warming" in 2007 is accompanied by an increase in media attention (see Figure 3, bottom part). This increase in 2007 also appears in specific countries, such as the Philippines, South Africa and India, while in many Western countries the increase in search volume starts approximately six months earlier.
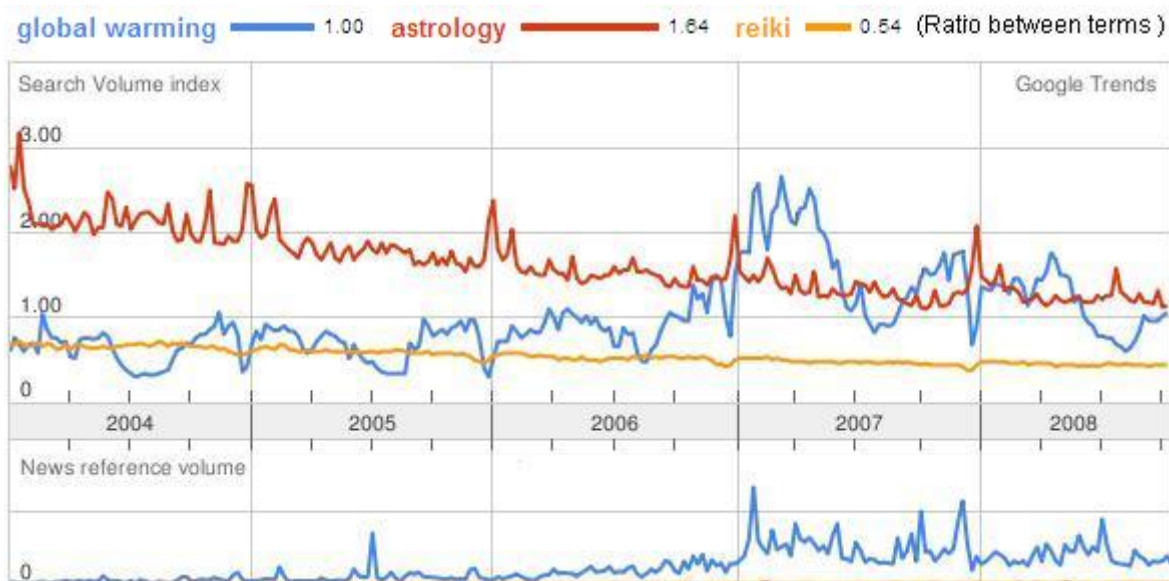
Figure 3. Trends in environment and pseudoscience searches.

More findings regarding interest in medicine, evolution, Biotechnology and Nanotechnology will be presented at the talk.

Conclusion

The expanding use of the Internet as an information source makes it increasingly possible to make use of web-based tools in order to learn more about users' needs and interests. Using publicly available web-based research tools, we were able to shed some light on specific public interests of science issues.

The potential of these new research tools for educational research is great. It can be summarized into three main contributions:

- Providing a bottom-up, up-to-date, authentic and learner-centered monitoring of public's interest in science and pseudoscience as reflected by search queries
- Allowing a cross-national comparison of popular science and pseudoscience related- search.
- Enabling triangulation of results found using surveys.

  All these will allow tailoring evidence-based formal and informal science learning environments to build on these genuine interests.

  However, the use of these tools for an academic research bears some limitations. *Google Trends* provides an arbitrary scale system, and thus making the comparison qualitative rather than quantitative. It returns results for terms that receive a significant amount of search traffic. Therefore, specific science-related terms in Hebrew, for example, could not be studied using this tool. *Google Trends* is a *Google Labs* product, meaning that it is still in its early stages of development. It may contain inaccuracies as a result of data-sampling considerations and certain approximations. Finally, our study indicates that *Google Trends* is a limited tool for cross-national comparison, since users from different countries use different languages to search for information. To that end, *Google Zietgeist* provides a complementary data on users' searches worldwide, and therefore can be useful to further examine, validate and support the result in *Google Trends*.

Notwithstanding these limitations, we argue that the combination of *Google Trends* and *Google Zeitgeist* was found to be instrumental for studying and comparing search trends worldwide: While *Google Trends* outlines the popularity of any specific search query in each country, *Google Zeitgeist* displays the most popular searches in different countries. Hence, the combination of these tools can help to identify and compare worldwide trends of interest in science and pseudoscience, encourage further investigations on specific issues and provide support to surveys and studies in the field.

References

Ainley, M., Hidi, S., & Berndorff, D. (2002). Interest, learning, and the psychological processes that mediate their relationship. *Journal of educational Psychology, 94*(3), 545-561.

American Association for the Advancement of Science. (1990). *Science for all Americans*. New York: Oxford University Press.

Baram-Tsabari, A., & Yarden, A. (2007). Interest in biology: A developmental shift characterized using self-generated questions. *The American Biology Teacher, 69*(9), 546-554.

Baram-Tsabari, A., & Yarden, A. (2008). Girls' biology, boys' physics: evidence from free-choice science learning settings. *Research in Science Technological Education, 26*(1), 75-92.

comScore. (2008). February 2008 U.S. Search Engine Rankings. Retrieved Apr 6, 2008, from http://www.comscore.com/press/release.asp?press=2119

Deci, E. L. (1992). The relation of interest to the motivation of behavior: a self-determination theory perspective. In K. A. Renninger, S. Hidi & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 43-70). Hillsdale, NJ: Lawrence Erlbaum Associates.

Eurobarometer. (2005). *Europeans, Science and Technology* (No. Special Eurobarometer 224): European Commission, Public Opinion Analysis sector.

Eurobarometer. (2008). *Attitudes of European citizens towards the environment* (No. Special EUROBAROMETER 295). Brussels: European Commission, Public Opinion Analysis sector.

Internet World Stats. (2008). World Internet Users and Population Stats. Retrieved September, 2008, from http://www.internetworldstats.com/stats.htm

Jansen, B. J., & Spink, A. (2003, 23-26 June). *An analysis of web information seeking and use: Documents retrieved versus documents viewed.* Paper presented at the 4th International Conference on Internet Computing, Las Vegas, Nevada.

Jansen, B. J., Spink, A., & Saracevic, T. (2000). Real life, real users and real needs: A study and analysis of users' queries on the Web. *Information Processing and Management, 36*(2), 207-227.

Kopytoff, V. (2007). Top secret, in plain view Google Earth may blur the image, but others don't. *Journal, 18 May 2007*. Retrieved from http://www.sfgate.com/cgi-bin/article.cgi?f=/c/a/2007/05/18/GOOGLE.TMP&feed=rss.news

Laugksch, R. C. (2000). Scientific literacy: A conceptual overview. *Science Education, 84*(1), 71-94.

Lujan, J. L., & Todt, O. (2000). Perceptions, attitudes and ethical valuations: The ambivalence of the public image of biotechnology in Spain *Public Understanding of Science, 9*(4), 383-392

Macoubrie, J. (2006). Nanotechnology: public concerns, reasoning and trust in government. *Public Understanding of Science 15*(2), 221-241.

Miller, J. D. (2004). Public Understanding of, and Attitudes toward, Scientific Research: What We Know and What We Need to Know. *Public Understanding of Science, 13*(3), 273-294.

Murata, T. (2006, December 18-22). *Towards the Detection of Breaking News from Online Web Search Keywords.* Paper presented at the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, Hong Kong.

National Science Board. (2008). Science and Technology: Public Attitudes and Understanding. In *Science and Engineering Indicators 2008*. Washington, D.C.: U.S. Government Printing Office.

Scheufele, D. A., & Lewenstein, B. V. (2005). The Public and Nanotechnology: How Citizens Make Sense of Emerging Technologies. *Journal of Nanoparticle Research, 7*(6), 659-667.

Schummer, J. (2005). Reading nano: the public interest in nanotechnology as reflected in purchase patterns of books. *Public Understanding of Science, 14*(2), 163-183.

Schutz, H., & Wiedemann, P. M. (2008). Framing effects on risk perception of nanotechnology. *Public Understanding of Science, 17*(3), 369-379.

Segev, E., & Ahituv, N. (Forthcoming). Popular Searches in Google and Yahoo: A Digital Divide of Information Uses. *The Information Society*.

Silverstein, C., Henzinger, M., Marais, H., & Moricz, M. (1999). Analysis of a very large Web search engine query log. *ACM SIGIR Forum, 33*(3).

Sullivan, D. (2006, August 21). comScore Media Metrix Search Engine Ratings. Retrieved June, 2008, from http://searchenginewatch.com/showPage.html?page=2156431

Superior Committee on Science Mathematics and Technology Education in Israel. (1992). *Tomorrow 98*. Jerusalem: Israeli Ministry of Education.