Designing a multilingual, quality-aware search engine for medical practitioners and patients

Matthias Samwald
Information Retrieval Facility & Medical University of Vienna
Vienna, Austria
m.samwald@ir-facility.org

Jimison Iavindrasana
Health-on-the-Net Foundation
Geneva, Switzerland
nlpresearcher@healthonnet.org

Veronika Stefanov
Information Retrieval Facility
Vienna, Austria
v.stefanov@ir-facility.org

Celia Boyer
Health-on-the-Net Foundation
Geneva, Switzerland
celia.boyer@healthonnet.org

Allan Hanbury
Information Retrieval Facility
Vienna, Austria
a.hanbury@ir-facility.org

ABSTRACT
The majority of patients in the industrial world uses the web for health advice, and a growing number of medical practitioners rely on web-based information as well. However, there are concerns that health-related information on the web could be of low quality and could harm patients. We analyze the current concerns about the quality of medical information on the web and introduce the European Khresmoi project, which aims to provide a high-quality multilingual search engine for medical practitioners and patients.

Categories and Subject Descriptors
H.1.2 **[MODELS AND PRINCIPLES]**: User/Machine Systems - *Human factors*

General Terms
Design, Reliability, Human Factors.

Keywords
Information retrieval, Medicine, World Wide Web, Multilinguality, Trustability, Categorization.

1. BACKGROUND
Sixty-one percent of American adults seek out health advice online. Around a third of those surveyed admitted that they changed their thinking about how they should treat a condition based on what they found on the web [11]. At present, most of these patients choose to find their medical information by using one of the major search engine providers, leading to information of varying quality. Medical professionals and other commentators are voicing concerns that questionable health-related information on the web could harm patients.

Furthermore, medical professionals themselves are increasingly open to using web-based systems to inform medical decision making [10,13]. However, medical practitioners usually do not have time to judge the validity of claims, so they require reliable tools to indicate the quality of medical information they are viewing. Basing medical decisions on faulty or outdated information could have grave consequences.

Various studies on the quality of online information have come to different conclusions, depending on the region of medicine under consideration. About 70% of the top websites with information on oral cancers gathered by Google and Yahoo searches were found to have serious deficiencies [8]. Among other things, the websites failed to attribute authorship, cite sources and report conflicts of interest. A study of 'googled' advice on how to deal with heartburn examined 16 websites and found that "the evidence for most of the recommendations is weak to nonexistent" [9]. Often the source of the information and the website sponsoring are problematic. A study on internet searches for obstetrically related terms [5] found that on the first page of results, “lawyers were the most common sponsors of websites retrieved by the terms cerebral palsy (52%), birth trauma (48%), and shoulder dystocia (43%). Only 3.6% of websites on the first page of results were created or sponsored by obstetrician-gynecologists.” A study examining 48 UK websites on attention deficit hyperactivity disorder [1] came to the conclusion that “Information on UK websites about drug treatment for ADHD is basic and incomplete.” In a study on internet use by cancer survivors [12], 60% of a sample of 261 cancer survivors used the internet mainly to obtain information about cancer and its treatment. The patients preferred to get their information from reliable sites, such as Web sites of hospitals and oncologists. Nevertheless, asking the patients about which websites they visited revealed that websites that are completely financed and created by pharmaceutical industries were mentioned most often.

Better experiences were reported in a study on patient information about general anaesthesia on the world wide web [14]. Using Google, the authors searched for information on general anaesthesia, using four synonyms of the term in four languages and analysing the top 20 results, yielding a total of 80 results examined for each language. For the individual language search terms, the following percentage of search results contained relevant information: 45% with English (UK); 16% with Swedish; and 20% with Finnish. 'Good' websites were found with the search terms in all languages: 12 with English (UK); two with Swedish; and one with Finnish. However, few good websites showed a reading grade level that is recommended for consumer health information. For opthalmology, the transparency of 20 websites aimed at the general public was evaluated [6]. The websites performed differently on different transparency criteria: All websites maintained separation between advertising and editorial content. Between 80 and 90% of the websites studied contained complete information about the provider, aims and target audience, and feedback from users. The criteria for privacy were met by 70% of websites, 40% met those for currency of content and data, 35% those for methods of quality assurance, 15% those for financing and sponsoring and 0% met the requirements for authors and sources of information.

Web 2.0 sources are becoming more widely used by the general public as sources of medical information. Wikipedia articles appear in the top 10 results for more than 70% of medical queries in four different search engines tested in [7]. It also gets more hits than corresponding pages on the US National Library of Medicine's MedlinePlus service. Wikipedia ranked highest for rare diseases, although its incidence in several categories had decreased. Page views increased parallel to the occurrence of 20 seasonal disorders and news of three emerging health concerns. Whereas Wikipedia medical articles have been found to be accurate, they are also incomplete. For example, a study on drug information comparing Wikipedia to the Medscape Drug Reference [2] found that "no factual errors were found in Wikipedia, whereas 4 answers in Medscape conflicted with the answer key." On the other hand, "Wikipedia was able to answer significantly fewer drug information questions (40.0%) compared with MDR (82.5%)." An advantage of Wikipedia was that "there was a marked improvement in Wikipedia over time, as current entries were superior to those 90 days prior."

2. PRELIMINARY RESULTS

We can distinguish between two fundamental dimensions of trust for medical information: the trustworthiness of the publishing entity (*quality of the medium*), and the trustworthiness, methodological soundness of the medical data presented by this publishing entity (*quality of the evidence*).

The quality of the medium can be assessed according to the HONcode ethical codes used by the Health On the Net Foundation (HON) for classifying medical websites. These consist of eight criteria:

- Authoritativeness (indicating the qualifications of the authors)
- Complementarity (patient relationship information should support, not replace, the doctor)
- Privacy (data submitted to the site by the visitor, respect the privacy and confidentiality of personal information)
- Attribution (citing the sources of published information, references related to medical content mentioned and date)
- Justifiability (backing up claims)
- Transparency (accessible presentation, correct contact information)
- Sponsorship (identifying funding sources)
- Advertising policy (clearly distinguishing advertising from editorial content)

To obtain the HONcode certification, webmasters have to submit a formal application via the HON website. The demand is evaluated and then the HONcode reviewer committee assesses the site according to the criteria mentioned above. The adherence to each criterion is verified and amendments are requested when necessary. This is an iterative task as none of the websites asking for HONcode certification has fulfilled all eight criteria. HON has certified more than 7300 web sites in 32 languages since 1996.

However, the ever-growing quantity of online health information makes the manual review process time consuming. HON investigated the use of an automatic system to help certification experts with their tasks, as HON guarantees the accuracy of the conformity with the eight criteria by the manual evaluation. In order to identify the adherence to the eight criteria, the HONcode Review Committee has to navigate the site in order to find and identify the pages mentioning the text related to a given criterium. An automated system will reduce this time consuming work and retrieve the pages related to a given criterium.

The automatic detection of the eight criteria is based on machine learning techniques [3,4]. The content of webpages related to each criterium and their URL addresses were used as learning material. Classical algorithms such as Naïve Bayes, SVM, k-NN and decision trees were analyzed for multi-classes supervised learning. These studies covered four languages: English, French, Spanish and Italian. The studies were limited to these languages because the creation of features is based on natural language processing techniques such as stemming or the use of stop words. Figure 1 shows the results obtained in the classification of 375 English web pages from 35 web sites [4]. This contingency table between automatically **assigned** ("*Ass*") and **correct** ("*Cor*") classes shows how the system recognizes the principles in terms of precision and recall. The diagonal line in blue shows the good categorization results. The two squares in pink identify insufficient categorization results, and squares in orange identify inaccurately assigned classes.

Additional studies will be conducted in order to automatically identify if certain criteria are fulfilled or not. Increasing the coverage of the languages is also an important task to pursue. These efforts will be conducted within the European FP7 project Khresmoi. Other websites available in other European languages will also be included in the learning material. The final aim is the integration of the automatic principle detection into a health search engine for practical evaluation.

Ass \ Cor	Authority	Complem.	Privacy	Reference	Justif.	Trans.	Sponsor.	Advert.	Date
Authority	**64 / 72**	05 / 05	01 / 01	19 / 34	01 / 09	04 / 13	04 / 09	01 / 01	00 / 01
Complem.	05 / 05	**80 / 82**	05 / 03	01 / 02	06 / 44	00 / 00	03 / 05	00 / 01	00 / 00
Privacy	02 / 03	02 / 04	**92 / 90**	00 / 01	00 / 03	01 / 02	01 / 02	02 / 06	00 / 00
Reference	24 / 13	03 / 02	03 / 01	**64 / 57**	02 / 08	01 / 01	02 / 02	00 / 00	01 / 02
Justif.	06 / 01	32 / 03	06 / 00	06 / 01	**45 / 33**	02 / 01	00 / 00	00 / 00	00 / 00
Trans.	06 / 02	02 / 01	08 / 02	02 / 01	00 / 00	**81 / 81**	00 / 00	01 / 00	00 / 00
Sponsor.	05 / 03	04 / 02	02 / 01	01 / 01	00 / 02	02 / 02	**69 / 69**	16 / 17	00 / 00
Advert.	01 / 01	02 / 01	05 / 01	00 / 00	00 / 02	00 / 00	13 / 12	**77 / 73**	00 / 00
Date	00 / 00	01 / 00	01 / 00	06 / 03	00 / 00	00 / 00	01 / 01	01 / 01	**90 / 98**

Figure 1: Results of the automatic classification of English web pages

Even among media that meet all of the criteria above (such as refereed literature) the quality of reported findings can be vastly different. The methodologies and evidence used to yield medical knowledge vary greatly. While some of them (such as systematic reviews) represent very strong evidence, others (such as simple case reports or animal studies) represent only very weak evidence. Ongoing research in the Khresmoi project is focused on aligning measures of the quality of the medium with measures of the quality of the evidence, and presenting such combined quality measures through a unified web search interface.

Another aspect to be considered is the sustainability of quality-control mechanisms. The size as well as the number of relevant data sources is growing, and existing sources are updated and changed. It becomes increasingly important to find ways to minimize the amount of human intervention that is necessary for effective quality judgments. Even seemingly automated approaches based on machine learning require manual curation of the training data, and an increased need for re-training can be expected as the characteristics of the data change over time. This implies a need for monitoring the data sources at the origin, in order to recognize when the content has changed enough to require re-inspection. Automatic triggering of evaluation workflows will be necessary. Also, borders between information sources become blurred with the proliferation of Web 2.0 features. A website that is perceived as one entity can contain evidence of very different quality, with comments, reviews and external content becoming an integral part of the whole, making it harder to differentiate between static and dynamic content. On the one hand this presents a challenge to users and quality-control, but on the other hand the enriched information, sharing and commenting provides yet another level of data that can be mined for quality indicators.

3. CONCLUSION

The web has become an important resource for health-related decisions, making it likely that the quality of information available on the web can have a noticeable effect on public health. How can the quality-improving measures discussed in the paper have an impact in the real world? Most users – especially patients – will keep on using the search engines that they already use, such as Google. One solution could be that quality-aware resources such as Khresmoi act not only as search engines, but also as providers of aggregated, curated information. This would make it possible to position the information in Khresmoi as highly-ranked search results for various medical terms. This form of 'benevolent' search engine optimization needs to be investigated further

4. ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 257528 (KHRESMOI).

5. REFERENCES

[1] Akram, G. u. a. 2008. Characterisation and evaluation of UK websites on attention deficit hyperactivity disorder. *Archives of Disease in Childhood*. 93, 8 (2008), 695-700.

[2] Clauson, K.A. u. a. 2008. Scope, Completeness, and Accuracy of Drug Information in Wikipedia. *Annals of Pharmacotherapy*. 42, 12 (2008), 1814-1821.

[3] Gaudinat, A. u. a. 2007. Combination of heterogeneous criteria for the automatic detection of ethical principles on health web sites. *AMIA ... Annual Symposium Proceedings / AMIA Symposium. AMIA Symposium*. (2007), 264-268.

[4] Gaudinat, A. u. a. 2007. Machine learning approach for automatic quality criteria detection of health web pages. *Studies in Health Technology and Informatics*. 129, Pt 1 (2007), 705-709.

[5] Kaimal, A.J. u. a. 2008. Google obstetrics: who is educating our patients? *American Journal of Obstetrics and Gynecology*. 198, 6 (Juni. 2008), 682.e1-5.

[6] Lüchtenberg, M. u. a. 2009. Transparency of Information on Eye Diseases on the Internet. *Ophthalmologica*. 223, 3 (2009), 145-154.

[7] Laurent, M.R. und Vickers, T.J. 2009. Seeking Health Information Online: Does Wikipedia Matter? *Journal of the American Medical Informatics Association*. 16, 4 (2009), 471-479.

[8] López-Jornet, P. und Camacho-Alonso, F. 2009. The quality of internet sites providing information relating to oral cancer. *Oral Oncology*. 45, 9 (2009), e95-e98.

[9] Nocon, M. und Müller-Riemenschneider, F. 2009. Internet-based lifestyle advice for heartburn patients. *European Journal of Integrative Medicine*. 1, 2 (2009), 99-102.

[10] Parekh, N. u. a. 2009. Connecting with Physicians Online - Searching for Answers; Hall & Partners and Google Custom HCP Study.

[11] Pew Internet & American Life Project. *http://www.pewinternet.org/Reports/2009/8-The-Social-Life-of-Health-Information/01-Summary-of-Findings.aspx*. Accessed: 02-07-2011.

[12] van de Poll-Franse, L.V. und van Eenbergen, M.C.H.J. 2008. Internet use by cancer survivors: current use and future wishes. *Supportive Care in Cancer: Official Journal of the Multinational Association of Supportive Care in Cancer*. 16, 10 (Okt. 2008), 1189-1195.

[13] Sim, M. u. a. 2006. Does general practice Google? *Australian Familiy Physician*. 37, 6 (Juni. 2006), 471-474.

[14] Tallgren, M. und Bäcklund, M. 2009. Patient information about general anaesthesia on the internet. *Anaesthesia*. 64, 4 (Apr. 2009), 408-415.