

Evolution of Health Web certification, through the HONcode experience

Vincent Baujard ^a, Celia Boyer ^a, and Antoine Geissbühler ^{a,b}

^a*Health on the Net Foundation, Geneva, Switzerland*

^b*Medical Informatics Service, University Hospital of Geneva, Switzerland*

Abstract. Today, the Web is a media with increasing pervasiveness around the world. Its use is constantly growing and the medical field is no exception. With this large amount of information, the problem is no longer about finding information but assessing the credibility of the publishers as well as the relevance and accuracy of the documents retrieved from the web. This problem is particularly relevant in the medical area which has a direct impact on the well-being of citizens and in the Web 2.0 context where information publishing is easier than ever. To address the quality of the medical Internet, the HONcode certification proposed by the Health On the Net Foundation (HON) is certainly the most successful initiative. The aims of this paper are to present certification activity through the HONcode experience and to show that certification is more complex than a simple code of conduct. Therefore, we first present the HONcode, its application and its current evolutions. Following that, we give some quantitative results and describe how the final user can access the certified information.

Keywords. Information quality, Certification, Natural Language Processing, Internet.

Introduction

In recent years the ease of publishing on the Internet has been further increased with the advent of the Web 2.0 phenomenon. Thus, despite the wealth of content available, the question is not just about finding information but also whether the information provided is credible. The problem is particularly acute in the medical information domain (explanation of the disease, recommended treatment, type of medicine ...) which has a direct impact on the health of public [1-3]. In response to the lack of transparency of the health information, many theoretical and practical initiatives have marked the short history of the Web. The history of the quality of information on the Web is intimately linked to developments in information retrieval. The most significant trends that have been applied to the Web on the quality of information (medical or not) are: the selection of webpages (e.g. Yahoo), self-regulation (e.g. Discern[4]), the popularity of webpages (e.g. Page Rank[5-7]), the certification of websites (e.g. URAC[8], HONcode[9]), education of the user (e.g. OMNI[10]) and the collaboration of users (e.g. Google Co-op[11]). Simple at first glance, the HONcode certification is based on four basic and indispensable elements. Indeed, experience has shown that 1/ **ethical charter** was only one of the visible elements of certification despite being essential. Thus 2/ **procedures** according to the HONcode assessments are vital because they reflect the practical and equitable implementations of the HONcode in the daily work of HONcode reviewers. Following this, 3/ **task certification** is

emerging as the key asset of certification, in fact other health codes have not been sustained because few people had volunteered to implement them. And finally, 4/ the means used to provide access to this information on the Internet.

1. Method and development of the HONcode

1.1. Certification

It is possible to cite the most prominent initiatives with their most salient features (see also [12]): Initiated in 1995, the implementation of the HONcode (third party certification, 8 principles audited annually) began in 1996, Discern (self-evaluation) in 1998, WebMedica in 1998 (certification only for Spanish), MedCERTAIN (detailed meta-data system) and Hi-Ethics (third party certification) in 2000, eHealth Code of Ethics (self-evaluation) in 2001, URAC in 2001 (very detailed but expensive) European Guidelines in 2002 (Eq. HONcode principles of the HON which participated in the development) and AFGIS in 2003 (dedicated to German sites). While some initiatives have disappeared or others do not have many candidates, the HONcode has been translated into 34 languages, had over 6500 sites certified by the end of 2008 in 118 countries and had been selected in 2007 by France to be the official certification body of French health websites.

1.2. Certification Process

HONcode certification [13] is a voluntary act on the part of the site applicant; the first step is submitting the application form on the HON website. A pre-assessment is proposed to the webmaster in order to identify the missing principles. Once the certification request is submitted, HON experts evaluate the websites that have applied. Each ethical principle which is not being complied by and should be added to the content of the webpages is indicated. Once the changes have been made, a seal of certification and unique identification is issued. All HONcode sites are certified for a period of 1 year and are reviewed annually in a systematic way. If a website no longer respects the HONcode principles, the webmaster receives a warning and if required changes are not made, the site may then lose its certification. As you can see, the certification process is an interactive process that provides a constructive contact between auditors (HON) and the webmaster. Indeed, the aim is not only to separate the wheat from the chaff but also help to bring up sites to a certain level of quality and transparency. In keeping with this aim, some additions have been made in the regulation implementing the HONcode principles to address the peculiarities of Web 2.0 (detailed Web 2.0 guidelines see Table 2. and [13]). The collaborative platform in addition to the current guidelines should respect as well the ones added specific to the Web 2.0.

Table 2. Web 2.0 specificity of HONcode

Principles	Regulations for implementations of principles
1. Authoritative	It must be clearly stated if the platform is moderated or not(...). It must be stated if platform users are warned or notified before being banned.
2. Complementarity	The complementarity statement should be added to the platform main page.

3.Privacy	It should be clearly mentioned that any messages could be used in other messages. It must be clearly stated if the platform user has the possibility to modify or erase his posts.
4.Attribution	A statement asking platform users to give references (links for ex.) to the health/medical information they provide, when is not from personal experience , must be provided. All posts must be automatically dated.
5.Justifiability	It must be clearly stated that platform users must post information which are true and correct to their knowledge.
6.Transparency	Same
7.Financial disclosure	It must be clearly stated if the moderators are volunteers or not.
8.Advertising policy	An advertising and editorial policy about platform user's posts must be provided. It must be stated if advertisement (links, banners, content...) is permitted or not on the platform.

1.3. Automatic detection of principles

In view of the dynamics of the Web, the certification is in continuous expansion. Recently, initiatives based on algorithms of criteria recognition, based on rules (originally proposed by Price [14]) or by automatic learning were presented to give an indication of quality and ethics to the pages of Web health. While the model of supervised learning Aphinyanaphongs [15] is based on static examples of good and bad pages and therefore dependent on fields, the HON approach is more generic since it is based on the model of the HONcode [16,17]. This last approach offer good results with 78% of accuracy over all principles, and its integration in HON daily activity is in progress.

2. Results of the certification and access to the final user

Currently the database contains more than 7,200 HONcode certified websites across 118 countries representing more than 1 million pages indexed in Google. 52% of the certified sites are in English and about 20% in French, followed by sites in Spanish, Italian, Australian and German (table 1 to see evolution). For each evaluated site, the following information are collected: 1/ the HONcode principles respected, 2/ text extracted corresponding to the 8 principles, 3/ URLs of the principles on the site, 4/ MeSH terms keyword[18] selected from the site and 5 / more general site label.

Table 1. Number of HONcode submissions between 2005 - 2009

Year	French Subscription	Subscription for all countries
2005	57	895
2006	42	898
2007	217	1142
2008	442	1409
2009	488	1604

In early 1996, a simple seal was introduced, allowing users to identify a certified site from a non-certified. However, the HONcode seal quickly became an additional safeguard for the Internet by requiring the sites to link the HONcode seal to the unique

HONcode certificate on the HON site. The idea is to limit misuse of the HONcode seal, as the final verification is done on the HON site. The new basic principle is that custody by HON ultimately enables control of the display of the HONcode seal (the single image generated for a given site and is also hosted at HON, figure 1).



Figure 1: Dynamic HONcode logo following the current status of the HONcode certification process

Google is the search engine most used by the Internet; it could become the perfect tool for the promotion and awareness of the quality of medical information on the Internet. In 2006 Google launched its collaborative platform Google Co-op. In our case, the procedure is to propose to Google.com annotations of webpages compared with predetermined labels (in agreement with Google) such as "treatment", "type of audience" etc. Thus the latest information provided by Google through HON consisted of 83,243 annotations for a total of 49,305 URLs (or patterns) (see Figure 2 point 4 green). HON Toolbar [19] is the most integrated way to access HONcode certified sites (Figure 2 point 1, 2, and 3). It is composed of 3 features (shown in red Figure 2) that are 1/ **identification of the HONcode membership** in real time while browsing the Web. 2/ The search tool, **HONcodeHunt**, exclusively dedicated to certified HONcode sites is accessible from the search bar of the browser. 3/ The **emphasis of certified sites in popular search tools** such as Google, Yahoo, MedlinePlus [20] and Wikipedia.

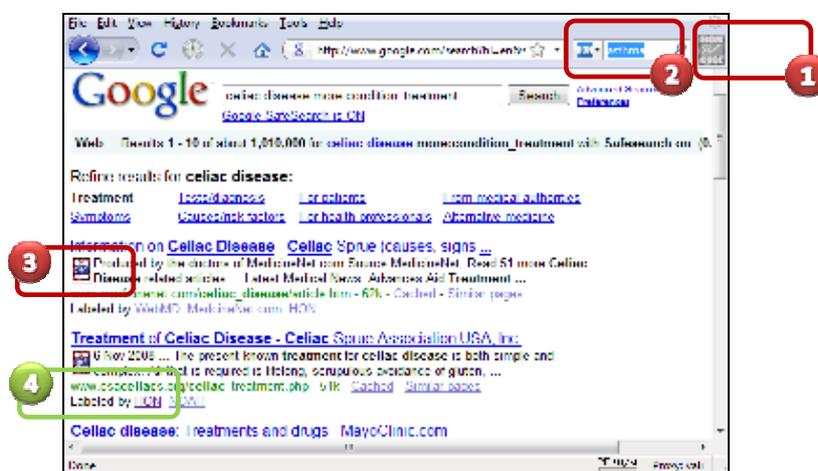


Figure 2: HON Toolbar within Google

3. Conclusion and perspectives

The aim of this paper was to show the many facets of the HONcode through its history, its evolution, implementation and use. During the past 14 years, HON has sought to promote the quality of medical information on the Web on a global scale. To

meet the quantitative requirements of the Web, human expertise is assisted by many automated systems for a systematic, reliable and faster evaluation of websites. It is very important to expand distribution channels to reach as many potential users. Thus the realization of collaborations, to share our information, our philosophy and our vision, with major players such as the National Library of Medicine (USA) or Google is essential. The approach led by the HON is comprehensive and covers more than 35 languages around the world. However, HON must respond to local needs, the variety of languages, cultural differences and different regulations. The creation of local branches in different parts of the world, such as those initiated in Africa, Italy or Spain, should enable us to think locally and act globally to improve the quality of medical information on the Internet. France is the pioneer in quality eHealth by legislating on the issue of quality of health sites. A similar approach in other European countries would be welcomed to continue to promote the quality of medical information on the Internet for the benefit of Internet users.

Célia Boyer
Fondation HON
81 Boulevard de la Cluse
1206 Genève
SUISSE
Celia.Boyer@healthonnet.org

References

- [1] S. Fox, Online Health Search 2006, Online Health Search 2006. Washington, DC: Pew Internet & American Life Project, October 29, 2006.
- [2] P. Wilson, (2002), "How to find the good and avoid the bad or ugly: a short guide to tools for rating quality of health information on the Internet", British Medical Journal, Vol. 324 No. 7337, pp. 598-602.
- [3] S. Fox, (2007) E-patients with a Disability or Chronic Disease. Pew Internet & American Life Project, 2007.
- [4] D. Charnock, The DISCERN Handbook. 1998. Radcliffe Medical Press.
- [5] B. Amento, L. Terveen, W. Hill, in: Proceedings of the Twenty-Third Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Does "Authority" Mean Quality? Predicting Expert Quality Ratings of Web Documents, 2000.
- [6] H. Borges, M. Cervi, P.T. Alvarez de Arcaya, G. Guardado, R. Rabaza, J. Sosa, Rate of compliance with the HON code of conduct versus number of inbound links as quality markers of pediatric web sites, in: Proceedings of the Sixth World Congress on the Internet in Medicine, Udine, Italy, 29 November—2 December 2001.
- [7] L. Page, S. Brin, R. Motwani and T. Winograd, The pagerank citation ranking: Bringing order to the web. Technical report, Stanford Digital Library Technologies Project, 1998.
- [8] URAC: <http://www.urac.org/MMandQualityChasm.asp>, Nov 2008.
- [9] C. Boyer, V. Baujard and J.R. Scherrer, HONcode: a standard to improve the quality of medical/health information on the internet and HON's 5th survey on the use of internet for medical and health purposes. In 6th Internet World Congress for Biomedical Sciences (INABIS 2000), 1999.
- [10] OMNI: omni.ac.uk, Dec 2008.
- [11] Google Co-op: <http://www.google.com/coop>, Dec 2008.
- [12] A. Risk, J. Dzenowagis, Review of Internet information quality initiatives. Journal of Medical Internet Research. 2001; 3(4):e28.
- [13] HONcode Guidelines: <http://www.hon.ch/HONcode/Guidelines/guidelines.html>, Nov 2008.
- [14] S.L. Price, W.R. Herish. Filtering web pages for quality indicators: an empirical approach to finding high quality consumer health information on the world wide web. Proc AMIA Symp 1999:911-5.
- [15] Y. Aphinyanaphongs, C. Aliferis, Text categorization models for identifying unproven cancer treatments on the web. Stud Health Technol Inform, 2007.

- [16] A. Gaudinat, N. Grabar, C. Boyer, Machine learning approach for automatic quality criteria detection of health web pages. *Medinfo*. 2007;129:705-9.
- [17] A. Gaudinat, N. Grabar, C. Boyer, Automatic retrieval of web pages with standards of ethics and trustworthiness within a medical portal: What a name page tells us - 11th Conference on Artificial Intelligence in Medicine (AIME 07) - 07-11 July 2007 Amsterdam, The Netherlands.
- [18] National Library of Medicine, Bethesda, Maryland. Medical Subject Headings, 2001. <http://www.nlm.nih.gov/mesh/meshhome.html>.
- [19] HON Toolbar: <http://www.hon.ch/HONcode/Plugin/Plugins.html>
- [20] N. Miller, E.M. Lacroix and J.E. Backus, MedlinePlus: building and maintaining the National Library of Medicine.