Medical Infobahn for Europe
A. Hasman et al. (Eds.)
IOS Press, 2000

HONselect: A multilingual and intelligent search tool integrating heterogeneous Web resources

C. Boyer[a], V. Baujard[a], V. Griesser[b], J.R. Scherrer[a]

[a] *Health On the Net Foundation (HON), Geneva, Switzerland*
[b] *Medical Informatics Division, University Hospital of Geneva, Geneva, Switzerland*

Background

The Internet carries a huge and growing amount of medical and health information. Users wishing to find specific information on a medical condition have two options. Either they consult topic lists that are often manually built or they use general or specialised full-text search engines. In both cases, they have to sort through and analyse the results they retrieve and only a few of these results correspond to the information sought. Further, the results in most cases are presented merely as a list of Web sites with no annotation or pre-information such as a basic description of the disease and its hierarchical structure, an indication of the medical context in which it belongs, a global overview of the term or links to additional sources of knowledge. To pull together disparate types of information like scientific articles, bibliographic references, news, medical illustrations and multimedia, the user until recently has had to undertake repetitive searches. This is no longer necessary. **Health On the Net Foundation** has developed **HONselect**, a new search integrator based on the hierarchical structure of medical terms of the MeSH® (Medical Subject Headings). It not only provides a solid foundation in the form of the MeSH® thesaurus [1, 2, 3], it also precludes time-wasting repetition by combining five information types – MeSH® classification, authoritative scientific articles, health and medical news, Web sites and multimedia – into one tool to focus and accelerate a user's search. The current version of HONselect is bilingual, in English and French, but HON's intention is to develop a multilingual tool.

1. Objective

Users looking for medical and health information tend to investigate and collect all types of sources in their effort to study medical explanations and clinical possibilities for a given disease or topic. They will usually seek all-round information on a specific topic, that is, not only a description of a disease or a condition but also scientific articles related to it, practice guidelines, continuing medical education, multimedia, news from the corresponding domain, Web sites, support groups, patient advocacy, conferences and more. HONselect is designed to meet this need and can be used by all – patients and their family members as well as researchers and medical and healthcare professionals [1,2]. This "bundling" of Internet resources affords users more time to explore the different databases and stay more concentrated in their reading. In the case of non-medical users, they gain an opportunity to learn about medical terminology and conceptual hierarchy as well as up-to-date details of a given disease, its context and related diseases and terms. It should be noted that HON is a not-for-profit organisation and the use of HONselect is entirely free of charge.

2. Method

HONselect first presents its users with the official MeSH context of any medical term being sought, its MeSH description and other "entry terms". Then, according to the MeSH term chosen, HONselect offers users direct access via hyperlinks, from the same on-screen page, to scientific bibliographic references, news, Web resources, and medical illustrations (multimedia). It is believed to have been the first system to integrate, in a single electronic interface, a selection of heterogeneous medical- and health-resource databases on the Internet [2]. The HONselect network architecture is totally adaptable, scalable and dynamic as shown in Figure 1. Other databases can thus be easily added to this network and the preformatted searches can be adjusted according to users' needs. The databases integrated by HONselect tool are independent of commercial influence, of high quality and fully-complementary: the Medical Subject Headings (MeSH) terminology thesaurus (a set of 33,000 controlled terms from the National Library of Medicine (NLM) [9], MEDLINE for scientific publication abstract (often with a link to the full paper), HON's MedHunt full-text search engine for Web resources, the Newspage service for daily news and HON's Media Gallery for medical images. It is an integrated system of Internet information types and databases offering an "intelligent" selection of resources available in each database [3, 4, 5, 6].

The National Library of Medicine (NLM) in the U.S.A. was the first to pioneer search integration with MEDLINEPlus [7, 8]. This consists of a complementary resources selection of Web sites plus bibliographical references for pre-defined diseases and conditions. It is in English only, and is dedicated mainly to patients.

Like MEDLINEPlus, the HONselect tool is carefully and regularly updated, but it also goes an important step further. It was built as a multilingual tool from the outset. It is dedicated to all Internet users – patients and care providers – with sustained medical curiosity and basic Web competence. It enables parallel searching of not two but five data types. It employs the MeSH hierarchy for standardised user orientation, offering the choice of its general "Browse" function [9] or its more specific "Search" function (see below).

HONselect is thus an entirely new concept that gives users the possibility of learning more about a given disease and its place in the accepted medical hierarchy with the added bonus of direct access to all information types on the subject that are available on the Internet. HONselect also permits the searching of medical terms in natural language, for example "back ache". It helps users find the corresponding scientific term, its standard definition, its position in the hierarchical structure of medical headings with broader and narrower concepts and terms, related terms, and, finally, other resources such as Web sites or scientific articles on the Internet.

The HONselect tool's specificity can be illustrated as follows:

- the Medical Subject Headings (MeSH) terminology in French and English;
- selections of bibliographical references in MEDLINE;
- Web resources with HON's MedHunt;
- multimedia with the HON Media Gallery;
- news with NewsPage;
- a conversion of natural-language search terms;
- a medical/health spell-checker.

Medical Subject Headings (MeSH) terminology in English and French:
The Medical Subject Headings consist of a classified and controlled medical vocabulary developed at the National Library of Medicine [7]. MeSH organises its 33,000 concepts into a hierarchical structure including anatomy, organisms, diseases, chemicals and drugs, psychiatry and psychology, etc. and provides the description of these terms [9]. The French organisation accredited to translate the MeSH terminology is the *Equipe "Centre Medlars" du réseau DIC-DOC de l'INSERM* [10]. Both MeSH versions available on-line have been entirely downloaded and are the principal elements of HONselect, thus enabling medical/health queries to be performed using MeSH terms [9]. Requests (common or scientific terms) result in the display of the relevant subset of the MeSH hierarchical structure. The user can then browse this structure in order to find appropriate concepts and then exploit the on-page offerings of HON's MedHunt, MEDLINE, NewsPage or HON's Media Gallery to click for access to selected Web sites, article references and abstracts, related daily news, or images and animation[3, 4, 5, 6]. Searches for a broad concept within the MeSH description utility can therefore include Web sites, articles, or information indexed to narrower concepts. See the example on "osteoporosis" in [11].

Bibliographical references in MEDLINE:
HONselect proposes bibliographical references when browsing or searching medical and health terms. Search pre-formatting is proposed via PubMed [12], the free official service for querying MEDLINE furnished by the United States government's National Library of Medicine (NLM). MEDLINE is the NLM's premier bibliographic database covering the fields of medicine, nursing, dentistry, veterinary medicine, the health care system and the pre-clinical sciences.

MEDLINE contains bibliographic citations (e.g., authors, title, and journal reference) and author abstracts from over 3,900 world-wide biomedical journals.

For each MeSH term, the selection is divided into two sections: one for individuals, the other for medical professionals or researchers. Each section offers four main categories based on the work of Haynes R.B. et al [13]: diagnosis, etiology, prognosis, and therapy. The "Individuals" section is dedicated to the general public and offers built-in queries with no restriction. The "Medical Professionals" section is devoted to clinicians with built-in clinical queries based upon the PubMed model. Each of the four study categories is proposed with the following options (an example for the term "osteoporosis" [14]):

- Sensitivity or recall, which retrieves most relevant articles but probably also some less relevant ones (broader focus);
- Specificity or precision, which retrieves most relevant articles but probably omits a few (narrower focus).

News with NewsPage:
HONselect offers the user a selection of free full-text news stories from the independent provider NewsPage, updated daily, in chronological order [4].

Web resources with MedHunt and multimedia resource with the HON's media gallery:
HONselect proposes a collection of Web resources retrieved automatically with MedHunt, HON's healthcare search engine. This indexes medical and health Web resources using the automatic Web spider MARVIN (Multi-Agent Retrieval Vagabond on Information Networks) [15], and features for its top search results a time-saving site description of a few lines, an

indication of whether the site is a HONcode subscriber [16] and the date of HON's last visit to it. A list of Web sites included in the MedHunt databases is automatically retrieved according to a given MeSH term [5]. The automatic selection is based on the relevance of the document for a given term. This method usually retrieves the best Web resources, though sometimes non-relevant documents are included. A similar technique is used to retrieve multimedia resources for a given MeSH term [6]. The HON's media gallery counts to date 1'300 images indexed according the MeSH classification.

The medical/health spell-checker:
Spell-checking is performed on the basis of the standard medical and health terms included in the MeSH glossary. If a query returns no results, then HONselect proposes glossary terms that have a similar spelling (at most four spelling differences). For example, when dealing with a misspelled entry such as "osteoroporosis", the spell checker verifies if there is a result with one spelling difference. If yes, it will display "osteoporosis". If not, the spell-checker will try with two, three and four spelling differences. The user can then redo the search by simply clicking on the terms proposed. If the spell-check propositions are not satisfactory (as would be the case for, say, the term "pain management"), HONselect proposes to search the term as a non-MeSH term with the HON's full-text medical and health search engine, MedHunt.

Search, Browse & Favourites:
HONselect proposes three ways to access the above resources: *Search*, *Browse* and *Favourites*. The *Search* section enables searching for the correct information with a definition and resources related to a given medical concept. The user may search any medical terms of the MeSH thesaurus, either only in the MeSH headings, or in the MeSH headings and related descriptions or in the French translation of the MeSH headings. The results propose the ten first terms retrieved from the MeSH concepts, the entry terms and related descriptions.

The *Search* facility retrieves the terms users are looking for and organises the results with the corresponding hierarchical context. In particular, the user may move up or down this hierarchy to either broaden or narrow his/her areas of interest. This can be done by clicking on the icon representing a concept (main heading , subheading , or terminal term). HONselect will show its hierarchical context, its definition, the narrow terms (subheadings) and eventual synonyms or entry terms.

The *Browse* section enables users first to browse a selection of terms (e.g. Disease, Anatomy, Chemicals and Drugs) to pinpoint the information sought with a definition and resources related to a given medical concept in the context of a hierarchy. From that point onwards, this section provides the same facilities as the *Search* section.

The *Favourites* section offers a regularly updated list of those medical terms which HON's own Web site traffic logs show to be those most frequently searched by HON users [14]. The list of these terms starts with AIDS, allergies, Alzheimer disease, arthritis, asthma, and so forth.

3. Results

HONselect in its English and French versions has been in use on the HON Web site since June, 1999. It was registering an average of 3,000 accesses per day at the time of writing (February 15, 2000). The English version is accessed about three times as frequently as the French version. The majority of its users appears to be health professionals, although

anecdotal evidence suggests it is also accessed by growing numbers of experienced lay users. User feedback indicates general satisfaction with the tool, in particular with the pre-formatted search facility. Other user comment highlights the fact that it permits access to requested information through simple clicking on a variety of direct links, avoids the need to access several different services to retrieve the same information and saves tiresome retyping of search requests.

4. Conclusion

As far as the authors know, HONSelect is the first system to integrate the MeSH terminology thesaurus [9], MEDLINE for scientific publication abstracts, a powerful search engine for Web resources, a recognised daily news service for current medical events and a variety of still, video and audio multimedia resources. It is an intelligent, user-targeted, integrated and easy-to-use index of Internet information types and databases. By encouraging lay users to familiarise themselves with standard medical terminology, it could promote health education and help facilitate patient-physician communication. When HONselect's multilingual functionality is enhanced with a German version, the authors hope it will set a new service standard for European users of the medical Intenet.

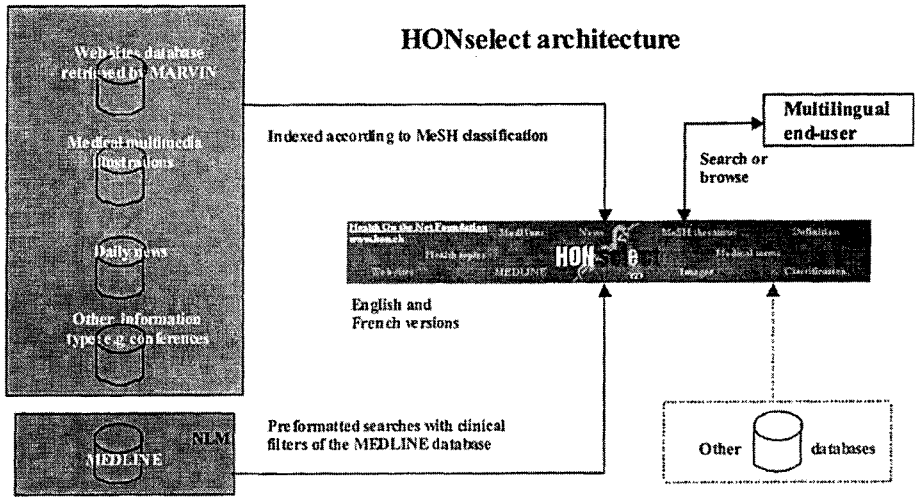

Figure 1 HONselect network architecture is adaptable and dynamic: other databases can be added to this architecture and the preformatted searches can be adjusted according to the users' needs.

References

[1] http://www.hon.ch/
[2] http://www.hon.ch/HONselect/
[3] http://www.nlm.nih.gov/mesh/
[3] http://www.nlm.nih.gov/databases/medline.html
[4] http://www.newspage.com/
[5] http://www.hon.ch/MedHunt/

[6] http://www.hon.ch/gallery/
[7] http://www.nlm.nih.gov/
[8] http://www.nlm.nih.gov/medlineplus/
[9] http://www.hon.ch/HONselect/Browse.html
[10] http://dicdoc.kb.inserm.fr:2010/basismesh99/mesh.html
[11] http://www.hon.ch/cgi-bin/HONselect?browse+C05.116.198.579
[12] http://www.nlm.nih.gov/pubmed/
[13] http://www.ncbi.nlm.nih.gov:80/entrez/query/static/clinical.html
[14] http://www.hon.ch/HONselect/Favourites.html
[15] O. Baujard et al, Trends in medical information retrieval on internet; Computers in Biology and Medicine 28 (1998), pages 589 - 601
[16] C. Boyer et al, The Health On the Net Code of Conduct for medical and health Web sites; Computers in Biology and Medicine 29 (1998), pages 603 - 610