Connecting Medical Informatics and Bio-Informatics
R. Engelbrecht et al. (Eds.)
ENMI, 2005

Representation of Medical Informatics in the Wikipedia and its Perspectives

Udo Altmann

Institute of Medical Informatics
University of Gießen, Heinrich-Buff-Ring 44, D-35392 Gießen

Abstract

A wiki is a technique for collaborative development of documents on the web. The Wikipedia is a comprehensive free online encyclopaedia based on this technique which has gained increasing popularity and quality. This paper's work explored the representation of Medical Informatics in the Wikipedia by a search of specific and less specific terms used in Medical Informatics and shows the potential uses of wikis and the Wikipedia for the specialty. Test entries into the Wikipedia showed that the practical use of the so-called WikiMedia software is convenient. Yet Medical Informatics is not represented sufficiently since a number of important topics is missing. The Medical Informatics communities should consider a more systematic use of these techniques for disseminating knowledge about the specialty for the public as well as for internal and educational purposes.

Keywords:
Wiki,;Wikipedia; Encyclopaedia; Medical Informatics

1. Introduction

"A wiki is a website (or other hypertext document collection) that allows a user to add content, like on an Internet forum, but also allows that content to be edited by anybody " (Definition for "wiki" in the Wikipedia [1]). The name "wiki" is based on an Hawaiian word for "quick". Web pages can be put into an edit mode which means that an HTML form (text area) is displayed where text is entered in a specific syntax. The syntax can easily be learned and is much more compact than HTML syntax (although some wiki engines also allow input of HTML). Thereby collaborative documents can be edited quickly with one author complementing and improving the other author's work.

Based on these ideas wikis are used for projects where documents have to be edited collaboratively, preferably for presentation on the Web or for the compilation of help pages. The first wiki initiated by Ward Cunningham was about sharing and developing ideas about programming in 1995. Meanwhile, a large variety of wikis exist, e.g., for different aspects of lifestyle, scientific issues, program languages, user documentation for programs, and, for encyclopaedic projects. The largest of the encyclopaedic projects is the "Wikipedia" that exists in many languages. Among these the English language project is the largest (437086 entries, start in January 2001, status on January 1st, 2005) and the German language project is the second largest one (183012 entries, start in May 2001). The project's mission is to

provide free encyclopaedic information with a license for using the text that has similarities to free software license models.

Since anybody can contribute to articles there is much concern about quality issues (including so-called "vandalism" of contents). A test carried out by a German computer magazine in October 2004 showed, that overall quality is good comparing the contents of the German Wikipedia to commercial products [2].

Other authors promote or use wikis for knowledge management in work groups [3] or as part of an interactive, web-based programming environment that enables biologists to analyze biological systems by combining knowledge and data through direct end-user programming [4].

This contribution shall explore the potential of Wiki techniques with respect to public and specialist knowledge about Medical Informatics.

2. Methods

To investigate the coverage of Medical Informatics topics by the English language Wikipedia, a number of terms (57) were selected mainly from the table of contents of the "Handbook of Medical Informatics" [5]. They comprised of terms closely related to Medical Informatics as well as of general informatics topics. For each search term, a full text search was performed to get a most specific (with respect to Medical Informatics) result. The full list of search terms can be seen in the results table.

The Wikipedia website (http://www.wikipedia.org) was visited systematically on two days (October 17th, 2004 and January 2nd, 2005). On these days the terms were searched and their existence or non-existence was recorded.

In case of existence the date of entry, the date of last change and the number of changes were recorded on occasion of the January visit according to the Wikipedia's built-in "history" function. Since changes of contents are often followed by a number of small immediate corrections, one author was only counted once per day. This count shall give a rough measurement of activity on this term.

In few cases the author added new entries on October 17th (Medical Subject Headings, Hospital Information System, Nursing Informatics, Medical Classification, Cancer Registry). The intention was not only to give new input to the encyclopaedia but to observe what happens to new articles.

Articles entered or modified by the author were put on a so-called "watchlist" that allows a quick overview of articles that were recently modified. Thus authors can quickly react if something undesired (for example vandalism) happens to these articles.

The selection of terms claims not be representative for Medical Informatics. Therefore results of the investigation are interpreted textually and the few counts and calculations have only limited value.

3. Results

Table 1 shows the result of the search. The column "Search Term" contains the initial term that was searched. The next column shows one ore more terms that were found during the search. For these terms, the date when they were entered and when they were changed for the last time is listed as well as the number of changes during this time. The last column contains the intensity of changes as calculation of: number of changes / (last change – date entered + 1).

Table 1 – Search Terms and Results

Search term	Found term	Entered	Last change	Days	Len-gth	Chan-ges	Inten-sity
Medical Informatics	Medical Informatics	28.10.2003	28.12.2004	428	4763	34	0,08
Nursing Informatics	Nursing Informatics	17.10.2004	04.11.2004	19	3508	8	0,42
Hospital Information System	VA Kernel (but not VistA)	22.02.2003	22.02.2003	1	828	0	0,00
	Hospital Information System	17.10.2004	17.10.2004	1	622	1	1,00
	Laboratory information system	16.12.2004	01.01.2005	17	5288	10	0,59
Clinical Information System	Not found						
Patient Record	Medical record	27.10.2004	31.12.2004	66	8094	11	0,17
	Medical privacy	28.11.2003	28.11.2003	1	1588	2	2,00
	HIPAA	01.12.2003	17.11.2004	353	2057	16	0,05
Electronic medical record	Electronic medical record	25.08.2004	25.12.2004	123	2566	11	0,09
Electronic Patient Record	see Electronic medical record						
Computer-based Medical Record	Not found						
Electronic Health Record	Not found						
Health Level Seven	HL7	01.12.2003	23.11.2004	359	2524	7	0,02
Primary Care	Primary Care	17.03.2004	01.08.2004	138	1493	2	0,01
Shared Care	Not found						
Information	Information	25.02.2002	15.04.2004	781	9434	92	0,12
Communication	Communication	07.11.2001	01.01.2005	1152	5311	120	0,10
Completeness	Completeness	10.06.2002	16.12.2004	921	4561	23	0,02
Accuracy	Accuracy and precision	26.02.2002	28.12.2004	1037	3395	11	0,01
Precision	Precision	25.02.2002	31.08.2004	919	1430	13	0,01
Coding	Coding	25.02.2002	17.10.2004	966	1839	11	0,01
Free Text	Not found						
Data Entry	see Data processing						
User Interface	User Interface	21.03.2002	08.12.2004	994	5134	16	0,02
Information Processing	Information Processing	10.09.2003	08.12.2004	456	2159	14	0,03
Data Processing	Data Processing	26.02.2002	02.12.2004	1011	1827	14	0,01
Data Presentation	Presentation (disambiguation)	25.04.2004	17.10.2004	176	1076	7	0,04
	Small multiple	16.03.2004	16.03.2004	1	594	0	0,00
	Markup (computing)	20.10.2003	21.11.2004	399	1231	11	0,03
Data Model	Data Model	09.09.2002	12.12.2004	826	2643	12	0,01
Data Structure	Data Structure	27.10.2001	24.11.2004	1125	3714	80	0,07
Classification	Classification	23.05.2003	27.12.2004	585	1651	23	0,04
Nomenclature	Nomenclature	07.12.2003	29.12.2004	389	1139	14	0,04
Thesaurus	Thesaurus	06.12.2001	31.12.2004	1122	2521	24	0,02
Taxonomy	Taxonomy	10.08.2001	18.12.2004	1227	3299	43	0,04
Nosology	Nosology	11.08.2002	01.11.2004	814	745	11	0,01
ICD	International Statistical Classification of Diseases and Related Health Problems	30.12.2001	06.11.2004	1043	3603	21	0,02
SNOMED	Medical Informatics						
ICD-O	Not found						
ICPM	Not found						
MeSH	Medical Subject Headings	17.10.2004	17.10.2004	1	892	2	2,00
DRG	Diagnosis-related group	13.03.2003	03.10.2004	571	3196	15	0,03
UMLS	SPECIALIST lexicon	15.05.2004	09.11.2004	179	5389	4	0,02
	Medical Classification	17.10.2004	06.11.2004	21	1794	3	0,14
Biosignal	Signal Processing	25.02.2002	25.12.2004	1035	922	14	0,01
Medical Imaging	Medical Imaging	27.05.2003	31.12.2004	585	4819	30	0,05
Ultrasound	Medical ultrasonography	02.11.2002	01.12.2004	761	6132	24	0,03
Computer Tomography	Computed axial tomography	01.05.2002	25.11.2004	940	4872	44	0,05
MRI	Magnetic resonance imaging	03.08.2001	20.12.2004	1236	10629	79	0,06
Medical Image Processing	Medical Imaging						
Decision Support	Decision support system	14.02.2004	19.12.2004	310	11223	18	0,06
Bayes' Rule	Bayes' theorem	18.04.2002	10.11.2004	938	13708	79	0,08
Arden Syntax	Not found						
Neural Network	Neural Network	02.10.2001	31.12.2004	1187	19649	128	0,11
Guideline	Guideline (medical)	18.12.2004	18.12.2004	1	2998	0	0,00
Data Protection	Not found						
Data Security	Privacy	02.01.2002	28.12.2004	1092	2919	31	0,03
	Data Security	11.11.2004	15.12.2004	35	688	2	0,06
	Medical privacy	28.11.2003	28.11.2003	1	1588	2	2,00
Confidentiality	Confidentiality	31.10.2003	28.12.2004	425	1982	15	0,04
Health Telematics	Telematics	29.07.2004	19.12.2004	144	1474	8	0,06
Registry	Registry	30.09.2003	31.12.2004	459	2376	24	0,05
Cancer Registry	Cancer Registry	17.10.2004	19.11.2004	34	1663	1	0,03
Hospital Cancer Registry	Not found						
Population-based Cancer Registry	Not found						

Coverage / Missing Terms

The more general "informatics" topics were covered well, but important rather specific "Medical Informatics" terms could not be found on the first visit:

- Nursing Informatics
- Hospital Information System, Clinical Information System
- ICD-O, ICPM
- Biosignal
- Arden Syntax, Guideline
- Health Telematics
- Cancer Registry

Some of the terms that were missing on the first visit were added by the author. "Medical guideline" was entered between the first and second visit by other authors as well as the terms "Laboratory information system" and "Data Security".

Variation of terminology

Some topics could not be found by the original term but by variations. The largest variation exists for "Electronic Medical Record".

Intensity of work on terms

Usually activity on terms is high during the first time after the term was entered. After excluding terms with a short period (20 days) between entry and last change, the arithmetic mean value for intensity is 0.047 (with a range from 0.01 – 0.17) which means that the terms of this set were changed approximately once per twenty days on average. Terms with high activity were

- Medical record, Medical Classification ("medical terms")
- Information, Neural Network, Communication ("informatics" terms)

From the terms that were entered by the author only "Nursing Informatics" was changed frequently.

Quality of entries

This paper has no focus on the quality of specific articles. Where possible improvements were obvious and the author felt competent they were introduced usually without negative response by other contributing authors. In some cases (Data model, Classification) the intention was to introduce a more common language use into rather expert/technical definitions. The Wikipedia software offers a possibility to discuss articles with other authors (Talk pages) that helped to explain this intention to the previous authors and to develop formulations that could be accepted by all active authors.

During the period of observance no vandalism could be observed to the articles on the author's watchlist.

4. Discussion

In October 2005, the German computer magazine c't published a comparison between two commercial encyclopaedias (Microsoft Encarta Pro and Brockhaus) and the Wikipedia (all

German versions). Terms with different degrees of difficulty from different areas of science, society and culture were searched and compared by length and quality. The overall quality of the Wikipedia was better than the one of the commercial products although there were also articles of poor quality and there exist shortcomings with respect to specific multimedia functions (e.g. interactive maps). One of the most evident disadvantages that also could frequently be observed by the author was a slow response time of the servers due to capacity overload. According to information from the project, this is mainly caused by limited hardware resources. To sum up the magazine's evaluation shows that the ideas "wiki" and "Wikipedia" work well, and that it is worthwhile thinking about their significance for Medical Informatics.

Representation of Medical Informatics in the Wikipedia

The results of this paper's work show that Medical Informatics topics are not very well represented in the Wikipedia currently. This can mean that

- there is low awareness of / few knowledge about the Wikipedia project in the Medical Informatics "community"
- there exist no resources to contribute to the project
- the project is not estimated as a possibility to transfer knowledge about Medical Informatics to the public

One can indeed question whether and to what extent Medical Informatics should be represented in such a project. What are the options to obtain knowledge about Medical Information on the internet? Web portals of organizations like the International Medical Informatics Association (IMIA, http://www.imia.org) and European Federation for Medical Informatics (EFMI, http://www.efmi.org) are often designed to support organisational issues and don't explain what this specialty contributes to society and science. Using search engines internet users can obtain extensive information but with the disadvantage that search engines cannot measure quality and the user himself has to decide what information is best for him within an often large search result. Although it doesn't guarantee quality per se an encyclopaedia offers a more systematic access to knowledge. Yet as the mentioned experiences with "Data model" and "Classification" show, there is a risk of using a technical language that cannot be understood by the readers. Actually we have to deal with at least two major groups of readers.

The first group are people with no or few knowledge about Medical Informatics who might come across terms e.g. in newspapers or consider studying the profession. Articles for this public should be easily understandable. Since they won't provide very detailed information the frequency of changes will be low.

The second group are people with at least basic knowledge about Medical Informatics who want to learn more about another area. They will expect more detailed and up to date information than the first group. The advantage of wikis is that everybody can contribute to them and that information can easily be actualised in contrast to books.

How To Build Up A Wiki For Medical Informatics

The own experiences affirm that the Wikipedia's project software (MediaWiki, open source) works well and allows convenient editing of articles. The culture of collaboration leads to usually continually improving articles. Since obviously the area of Medical Informatics has not been worked on systematically in the Wikipedia, it is more or less left to chance whether an article is further developed or not (e.g. "Nursing Informatics" versus "Cancer Registry").

Writing articles for the Wikipedia can be very labour-intensive. A starting point to a more systematic approach to put Medical Informatics contents into a wiki like the Wikipedia could be its use in education and training. Students often have to do homeworks or presentations on specific topics. The same techniques they learn by these works – learning through active acquiring and presentation of knowledge, use of neutral point of view language, correct composition of papers, developing exact and understandable formulations – can also be trained by building up articles (at least for parts of their work, e.g. introductions and glossaries). The review that is necessary to improve quality of articles can be done by the teachers who would have to revise and assess the work anyway or by a group of students collaboratively. Using the built-in revision history, the contributions of the author or the state of a document can easily be tracked.

Although such wikis could be hosted on local servers, the best integration would be implementing them as part of the Wikipedia project. In this case, providing some funding should be considered to support the expansion of the currently limited resources.

As soon as authors from different institutions contribute to one article, one can anticipate that conflicts will arise from different opinions regarding a topic, e.g. the correct definition of a term. The WikiMedia software offers a medium to discuss these opinions through the already mentioned talk pages. A discussion on a topic remains linked to it and helps following readers or authors of the topic to interpret the contents and to understand the crux of the matter.

5. Conclusion

Wiki techniques offer new perspectives for collaborative work including the dissemination of knowledge. The Wikipedia project has gained an impressive volume with an overall good quality. Yet this potential is currently not used sufficiently by Medical Informatics for its representation neither to the general public nor to specialists. A means to improve this situation efficiently could be to use wikis or the Wikipedia for education and get articles as by-product of student works.

6. References

[1] http://en.wikipedia.org/wiki/Wiki on January 2nd, 2005

[2] Kurzidim M. Wissenswettstreit – Die kostenlose Wikipedia tritt gegen Marktführer Encarta und Brockhaus an. c't Magazin für Computertechnik, Heise-Verlag, Hannover, Heft 21, pp 132-139

[3] Kosuch HT; Porth AJ. Wissensmanagement in der Medizin mit "Wikis". Kooperative Versorgung - Vernetzte Forschung - Ubiquitäre Information. 49. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (gmds), 19. Jahrestagung der Schweizerischen Gesellschaft für Medizinische Informatik (SGMI) und Jahrestagung 2004 des Arbeitskreises Medizinische Informatik (ÖAKMI) der Österreichischen Computer Gesellschaft (OCG) und der Österreichischen Gesellschaft für Biomedizinische Technik (ÖGBMT). Innsbruck, 26.-30.09.2004. Düsseldorf, Köln: German Medical Science; 2004. Doc 04gmds371 (http://www.egms.de:80/en/meetings/gmds2004/04gmds371.shtml)

[4] Massar JP, Travers M, Elhai J, Shrager J. BioLingua: a programmable knowledge environment for biologists. Bioinformatics. 2004 Aug 12; [Epub ahead of print]

[5] van Bemmel JH, Musen MA (eds). Handbook of Medical Informatics. Springer, Houten/Diegem 1997

Address for correspondence

Dr. med. Udo Altmann, Institute of Medical Informatics, University of Gießen, Heinrich-Buff-Ring 44
D-35392 Gießen, Tel. ++49 641 99-41380, Fax. ++49 641 99-41359, e-Mail:
Udo.Altmann@informatik.med.uni-giessen.de