Mapping the Domain of Medical Informatics

M. J. Schuemie[1]; J. L. Talmon[2]; P. W. Moorman[1]; J. A. Kors[1]
[1]Department of Medical Informatics, Erasmus MC, Rotterdam, The Netherlands
[2]School for Public Health and Primary Care, CAPRI, Maastricht University, Maastricht, The Netherlands

Keywords

Medical informatics, topic map, scientific journals, semantic similarity

Summary

Objectives: The domain of medical informatics (MI) is not well defined. It covers a wide range of research topics. Our objective is to characterize the field of MI by means of the scientific literature in this domain.

Methods: We used titles and abstracts from MEDLINE records of papers published between July 1993 and July 2008, and extracted uni-, bi- and trigrams as features. Starting with the ISI category of medical informatics, we applied a semi-automated procedure to identify the set of journals and proceedings pertaining to MI. A clustering algorithm was subsequently applied to the articles from this set of publications.

Results: MI literature can be divided into three subdomains: 1) the organization, application, and evaluation of health information systems, 2) medical knowledge representation, and 3) signal and data analysis. Over the last fifteen years, the field has remained relatively stable, although most journals have shifted their focus somewhat.

Conclusions: We identified the scientific literature pertaining to the field of MI, and the main areas of research. We were able to show trends in the field, and the positioning of different journals within this field.

Correspondence to:

M. J. Schuemie
Department of Medical Informatics
Erasmus University Medical Center
Rotterdam
Dr. Molewaterplein 50
3015GE Rotterdam
The Netherlands
E-mail: m.schuemie@erasmusmc.nl

Methods Inf Med 2009; 48: 76–83
doi: 10.3414/ME0576
received: May 23, 2008
accepted: September 24, 2008
prepublished:

Introduction

Medical informatics (MI) as a discipline has not been defined in a uniform, distinct way. The Handbook of Medical Informatics, for example, states: "*Medical informatics is located at the intersection of information technology and the different disciplines of medicine and health care*" [1]. Later on in the same reference work in MI is summarized as: "*In medical informatics we develop and assess methods and systems for the acquisition, processing, and interpretation of patient data with the help of knowledge that is obtained in scientific research.*" These definitions are not clear-cut and give only a rather broad description of what MI entails. For their strategic plan [2] the International Association of Medical Informatics (IMIA) describes the field of MI in a 'scientific map' consisting of 109 terms organized into six categories (appendix 2 in [3]), but an effort to map the literature to this map showed that many MI articles do not contain any of these terms [4]. Recently, the situation has become even more unclear. Since a couple of years, bioinformatics has become a major discipline. Groups that formerly were identified as medical informatics groups call themselves biomedical informatics groups now. Whether biomedical informatics is another discipline is not (yet) clear. Some advocate that the future of MI lies in bringing results from bioinformatics in the clinical domain. Others have argued that techniques developed in MI should be brought to bioinformatics as to better facilitate translational research. An overview of these arguments can be found in the first issue of *Methods of Information in Medicine* of 2002 [5].

Rather than enter in a discussion on what the definition of MI should be, it is our objective to define the field of MI by means of the scientific literature in this domain. As a basis for the selection of the MI literature, we use the ISI Web of Knowledge 2007 list of 20 journals under the subject heading 'Medical Informatics'. This list contains several journals that do not seem to address the domain of MI, whilst relevant journals and proceedings are not included in this list. We have therefore applied several techniques to both filter and enrich this list in a systematic way.

Based on the selected literature, we are able to answer several questions concerning the field of MI:

1. Which topics have received most attention in the last fifteen years?
2. Has the focus of the field remained the same during that period?
3. What are the upcoming trends that can be detected in the last three years?
4. How are the different journals in the field positioned relative to the topics in MI, and to one another?
5. Is this positioning stable over the last fifteen years?

In the past, others have sought to create a map of the field of MI, either by using co-citation analysis [6], intercitation analysis [7], or using a factor analysis of a unigram noun-phrase representation of documents from a manually selected set of journals [8]. Recently, the overlapping field of health information systems was analyzed using co-citation analysis [9]. Our methodology differs from the aforementioned studies in that we have attempted to make our literature selection process as objective as possible, have used a richer feature set, and applied automatic clustering to the literature.

Methods

Preprocessing

On the August 13, 2008, we retrieved all 6,287,660 records from Medline that

- contained an abstract;
- were published between July 1, 1993 and July 1, 2008;
- did not belong to one of these publication types: comment, editorial, news, historical article, congresses, biography, newspaper article, practice guideline, interview, bibliography, legal cases, lectures, consensus development conference, addresses, clinical conference, patient education handout, directory, technical report, festschrift, retraction of publication, retracted publication, duplicate publication, scientific integrity review, published erratum, periodical index, dictionary, legislation or government publication.

For each of these records, we retrieved the title, abstract, publication date, and journal in which the article was published. Because several journals have changed their name over time, we mapped their old name to the new name. The articles from the book series entitled 'Studies in Health Technology and Informatics' that could not be mapped to the proceedings of MIE conferences and Medinfo were removed from our set, because these covered not only MI topics, but also a wide diversity of other topics. The mapping of old to new journal names is available from the authors on request.

Feature Extraction

We concatenated the title and abstract of each Medline record, and from this text we extracted n-grams as features. N-grams are sequences of words that occur in the text. In this study, we used unigrams (n = 1, i.e. all single words), bigrams (n = 2), and trigrams (n = 3). We ignored all n-grams that crossed any form of punctuation or parentheses, or contained one or more words that belong to a set of stopwords. The set of stopwords consisted of all stopwords used by PubMed, all the numbers ranging from 0 to 99, and a set of upper-case words such as 'AIM', 'APPROACH', and 'CONCLUSIONS' that are used to denote the structure of an abstract. The complete set of stopwords is available from the authors on request. Lexical variations were removed by applying the LVG Normalizer [10] to all words that contained a majority of lower case characters. Abbreviations, which typically consist mainly of upper case characters, were not altered.

The process of feature extraction is illustrated in ▶Table 1, showing an example sentence and the extracted n-grams. The n-gram profile of a document is the set of n-grams occurring in that document.

Document Set Profiling

In order to characterize a set of documents, such as all articles belonging to a journal, we calculated a n-gram profile for the set. This profile consisted of a vector of weights corresponding to all n-grams occurring in the set. The weight of each n-gram was taken as the symmetric uncertainty coefficient [11]. The symmetric uncertainty coefficient (UC) is a normalized variation of the mutual information measure:

$$UC = \frac{2\,I(X;Y)}{H(X)+H(Y)}$$

where *X* and *Y* represent the two discrete random variables for which we want to calculate the UC. In this case, *X* is a binary variable representing the occurrence of an n-gram in a document, and *Y* is a binary variable representing the membership of a document to the set:

$$X = (Ngram,\ notNgram)$$
$$Y = (Setmember,\ notSetmember)$$

Table 1 Feature extraction from a Medline sentence. Stopwords are underlined in the example sentence. The remaining n-grams after lexical normalization of the words are listed.

SUMMARY: Electronic health records (EHRs) hold the potential to significantly improve the quality of care in long-term care (LTC) facilities.	
electronic	quality
health	care
electronic health	long
record	term
health record	long term
electronic health record	care
EHRs	term care
hold	long term care
potential	LTC
improve	facility

I(*X*; *Y*) is the mutual information between *X* and *Y*, defined as:

$$I(X;Y) = \sum_{y\in Y}\sum_{x\in X} p(x,y)\log\left(\frac{p(x,y)}{p(x)\,p(y)}\right)$$

where *p(Ngram)* is the fraction of documents containing the n-gram, and *p(notNgram)* is its complement. *p(Setmember)* is the fraction of documents belonging to the set, and *p(notSetmember)* is its complement. *p(x,y)* represents the fractions of documents belonging to the combinations of variables. *H(X)* and *H(Y)* are the entropies of variables *X* and *Y* respectively:

$$H(X) = -\sum_{x\in X} p(x)\log\left(p(x)\right)$$

$$H(Y) = -\sum_{y\in Y} p(y)\log\left(p(y)\right)$$

Profile Similarity

The similarity between two sets of documents, represented by n-gram profiles *a* and *b*, was calculated using the cosine:

$$\cos(a,b) = \frac{a\cdot b}{|a|\,|b|}$$

Clustering

In the past, many algorithms have been described for detecting clusters in sets of documents. However, most methods require the

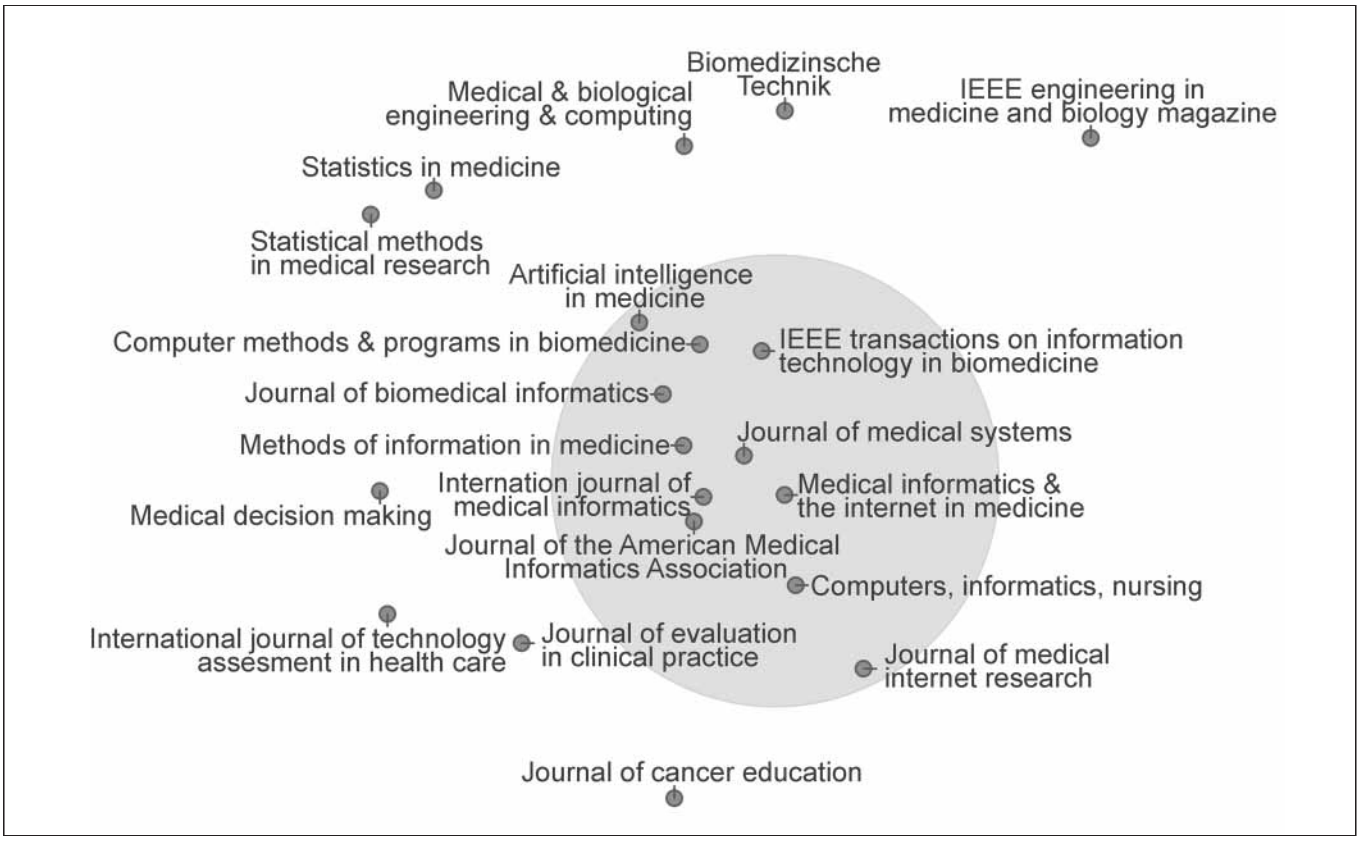

Fig. 1 All 20 journals belonging to the ISI category 'Medical Informatics'. The distance between journals approximates the dissimilarity between the n-gram profiles of the journals. The large circle indicates a coherent set of similar journals.

number of clusters to be defined beforehand. Other methods use information-based criteria such as Bayes Information Criteria or Minimum Description Length (e.g. [12]) to determine the optimum number of clusters. However, this presupposes that the information gained by dividing the corpus in clusters can be compared with the information needed to describe the clusters. This still requires some a priori estimations of parameters, which we would like to avoid since we do not know the nature of the underlying data that we are trying to cluster. We therefore used another approach: we optimize the average entropy reduction per cluster.

We define the entropy of a clustering as the sum of the entropy of all its clusters. The entropy of a single cluster is defined as the sum of the entropy of each feature (in our case n-grams) of the cluster. We determined the minimal possible entropy for a clustering with a given number of clusters using these steps:

1. Random initialization: documents are randomly assigning to a cluster.
2. Greedy hill-climber optimization: for all the documents we calculate the reduction in entropy resulting from moving a single document to another cluster. The move resulting in the most entropy reduction is effectuated and step 2 is repeated.
3. 3.Stop criterion: The optimization stops when the entropy cannot be reduced further.

We use the entropy of a single cluster containing all the documents as a baseline. We proceed to calculate the minimal entropy for an increasing number of clusters, and compare this to the baseline. By dividing the reduction in entropy by the number of clusters, we get the average reduction in entropy per cluster. The process is stopped when the average reduction stops increasing. The clustering with the highest average reduction is selected as the optimal clustering.

The combination of a random initialization and a hill-climbing algorithm could result in the system getting stuck in a local optimum. We therefore repeated all clusterings ten times, and selected the clustering with overall highest average reduction.

Visualization

Two-dimensional visualization of journals or clusters was performed using the spring-force layout algorithm of the Prefuse Toolkit for interactive information visualization [13]. The distance between two document sets with profiles *a* and *b* was defined as $-\log(\cos(a,b))$.

To provide a short description of a cluster, we selected the seven n-grams with the highest UC scores in the cluster profile. If an n-gram was a substring of another n-gram in the set of seven, we removed that n-gram and added the n-gram with the next highest UC score.

Similar Journals Search

In order to expand a 'seed' set of journals to include similar journals, we used the following method:

1. For each journal in Medline, we calculated the sum of the similarity scores between the profile of that journal and the journals in the seed set. We then ranked the journals based on these sums.
2. We selected all journals that had a score similar to or higher than the score of the lowest ranking seed journal.
3. We used this new set of journals as a seed for the next iteration, repeating steps 1 and 2.
4. The iterations were stopped when the ranking did not change.

Results

Journal Selection

We created a journal map of all the 20 journals belonging to the ISI category 'Medical Informatics' by creating the n-gram profiles for each journal, and visualizing these using the visualization algorithm, as shown in ▶Figure 1.

The circle indicates the journals that form a coherent set of similar journals. A hierarchical cluster analysis confirmed that this cluster is indeed distinct from the surrounding journals. The cluster was used as seed for the search of similar journals. After two iterations, the ranking stabilized to the list shown in ▶Table 2. We may conclude that these 16 journals and proceedings cover the domain of MI.

Table 2 shows several proceedings and two journals that have a higher score than the lowest ISI journal in our original seed set. Also, some of the journals of the ISI category show a very low similarity with the seed set when compared to other literature, as can be deduced from the ranks of these journals in this table.

Table 2 Journals most similar to the set of seed journals after two iterations. All journals belonging to the ISI category 'Medical Informatics' are shown (ISI), as well as the seed set (Seed).

Rank	Sum	Seed	ISI	Name
1	9.03	●		Medinfo
2	8.67	●	●	International journal of medical informatics
3	8.59	●		Proceedings of the AMIA Symposium
4	8.32	●		Proceedings of the MIE conferences
5	8.01	●	●	Methods of information in medicine
6	7.61	●	●	Journal of the American Medical Informatics Association
7	7.00	●	●	Medical informatics and the Internet in medicine
8	6.90	●	●	Journal of biomedical informatics
9	6.32	●	●	Journal of medical systems
10	6.15	●	●	Computer methods and programs in biomedicine
11	5.83	●	●	IEEE transactions on information technology in biomedicine
12	5.23	●	●	Computers, informatics, nursing
13	5.08	●	●	Artificial intelligence in medicine
14	4.98	●		BMC medical informatics and decision making
15	4.88	●		Computers in biology and medicine
16	4.08	●	●	Journal of medical Internet research
17	3.55			Proceedings of the IEEE Eng. in Med. and Bio. Society
18	3.38			Journal of healthcare information management
19	3.28			Informatics in primary care
20	3.26			Topics in health information management
21	3.10			BMC bioinformatics
22	2.84			Pacific Symposium on Biocomputing
23	2.63			Behavior research methods, instruments, & computers
24	2.59			Yearbook of medical informatics
25	2.52			Bioinformatics
…				
41	1.97		●	Medical & biological engineering & computing
50	1.72		●	Journal of evaluation in clinical practice
101	1.21		●	Medical decision making
107	1.19		●	Biomedizinische Technik
151	1.05		●	Journal of cancer education
162	1.01		●	Statistics in medicine
187	0.91		●	International journal of technology assessment in health care
275	0.72		●	Statistical methods in medical research
914	0.31		●	IEEE engineering in medicine and biology magazine

Topics in MI Literature

To analyze the topics dealt with in the domain of MI we analyzed the 14,885 articles belong-

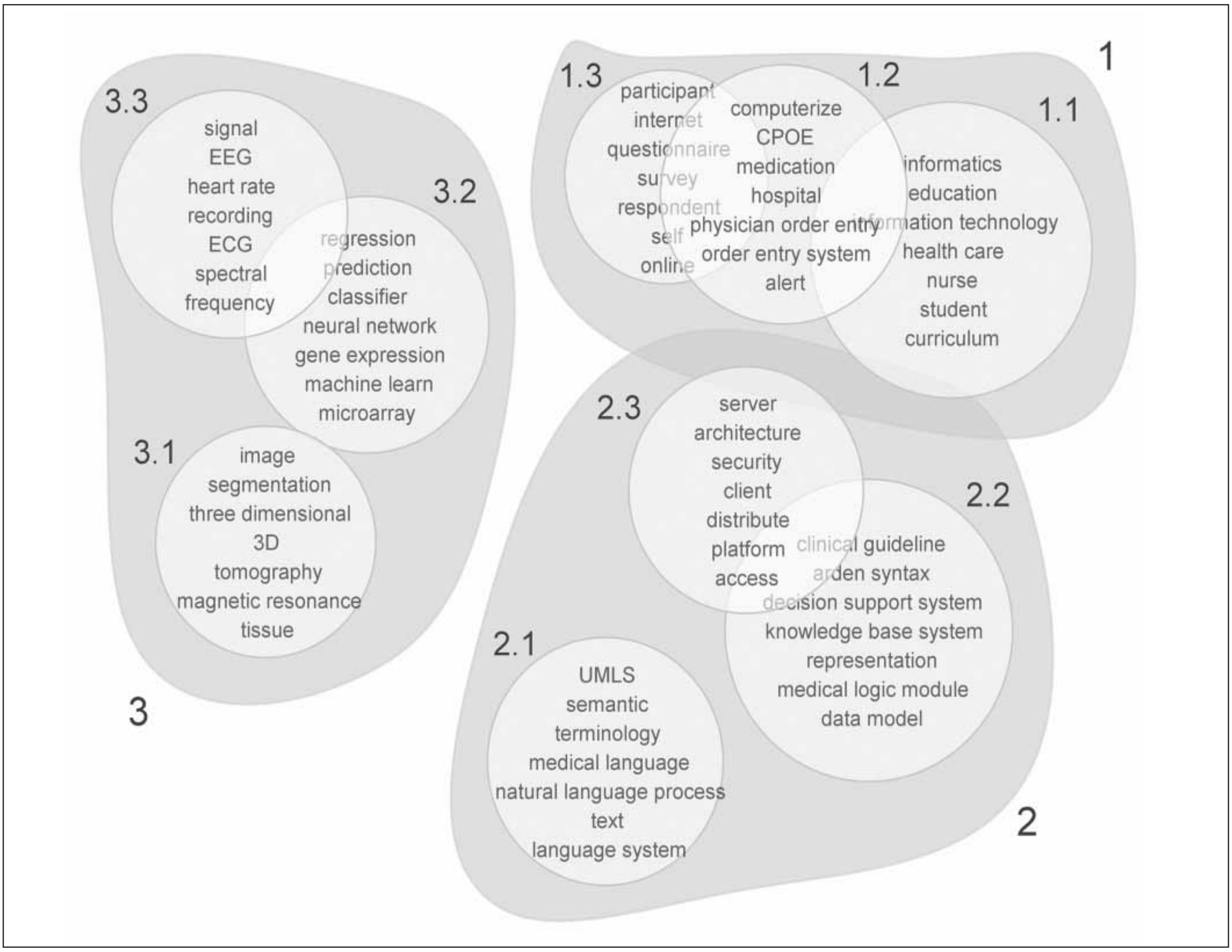

Fig. 2 Clustering and subclustering of 14,885 articles belonging to 16 MI journals. The size of the circles is proportional to the size of the subclusters. For each subcluster the seven n-grams with the highest uncertainty coefficient are shown.

ing to the 16 MI journals and proceedings to create the cluster map shown in ▶Figure 2.

To reduce visual clutter, the labels for the main categories are not shown in Figure 2, but are listed in ▶Table 3.

Cluster 1 appears to deal mainly with health information systems, their application, evaluation, and organization. An investigation of cluster 1.3 showed that this cluster contains many documents describing user evaluations of health information systems. Cluster 2 deals mainly with medical knowledge representation in the form of clinical guidelines, ontologies and databases. Also included is a subcluster dealing more specifically with the analysis of medical language. Cluster 3 deals with data analysis, with subclusters for classification techniques and statistical modeling, signal analysis, microarray analysis, and the field of image analysis.

Cluster 1	Cluster 2	Cluster 3
nurse	terminology	neural network
survey	semantic	algorithm
education	UMLS	signal
health care	concept	parameter
interview	natural language	method
questionnaire	architecture	estimation
IT	XML	linear

Table 3 Labels for the three main clusters

To study whether the field of MI has been stable over the last fifteen years, we divided the documents used for generating Figure 2 into five intervals of three years based on their publication date. For each interval, we determined the distribution of the documents over the nine MI subclusters, as shown in ▶Table 4.

Additionally, we investigated recent topics in MI by showing our original clustering, but only using the articles from the last three years to determine the cluster size, distances, and labels.

►Figure 3 shows several of the topics being discussed in the literature in the past three years. In cluster 1, we see Clinical Provider Order Entry (CPOE) systems remaining a major topic. User evaluations also appear to be a constant topic. In cluster 2 natural language processing remains a topic of re-

Table 4 Distribution of the articles from three-year time periods over the nine subclusters. Count indicates the total number of publications in a period, the other numbers indicate the percentage of these publications that are assigned to a particular cluster. Darker colors indicate higher percentages.

Period	Count	1.1	1.2	1.3	2.1	2.2	2.3	3.1	3.2	3.3
'93-'96	2453	14	9	4	8	23	10	8	10	13
'96-'99	2613	18	9	5	10	18	13	8	9	10
'99-'02	2680	14	10	7	10	18	11	9	11	10
'02-'05	3479	15	14	10	13	13	10	8	10	7
'05-'08	3660	13	13	11	10	8	8	11	14	11

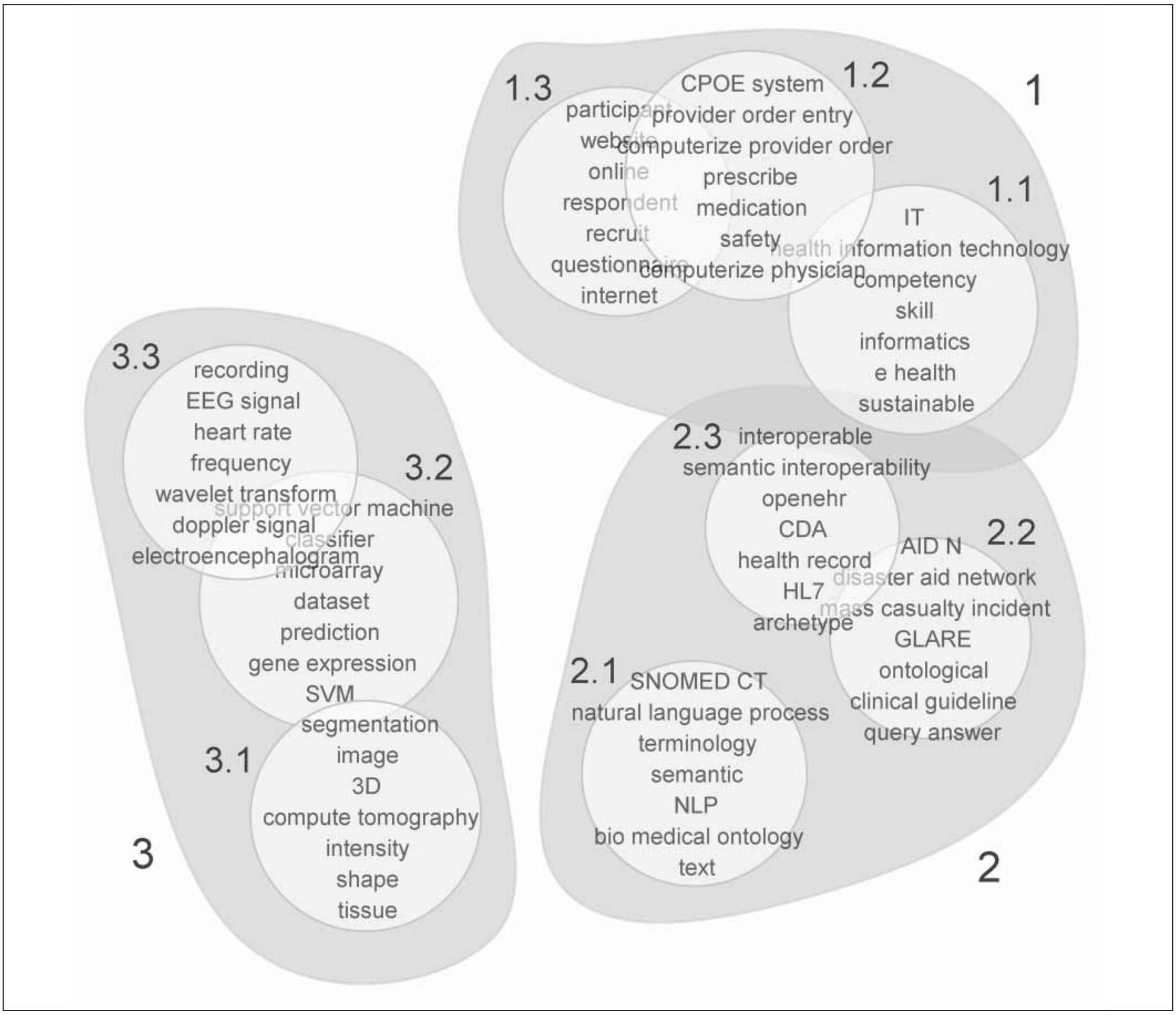

Fig. 3 Based on the clustering of Figure 2 the intercluster distances, their size and the seven n-grams with the highest uncertainty coefficient are shown for the 3660 articles that appeared in the last three years.

Table 5 Distribution of the articles from specific journals over the subclusters. Count indicates the total number of publications in a serial, the other numbers indicate the percentage of these articles that are assigned to a particular cluster. Darker colors indicate higher percentages.

Journal	count	1.1	1.2	1.3	2.1	2.2	2.3	3.1	3.2	3.3
Computers, informatics, nursing	444	62	16	12	2	5	1		1	1
Methods of information in med.	1091	15	7	5	8	13	12			
Med. informatics and the Internet in med.	365	11	7	16	7	15	20	9	8	6
Proceedings of the AMIA symposium	2971	13	19	5	21	24	9	2	7	2
Medinfo	1592	21	10	7	13	22	14	3	6	3
Proceedings of the MIE conferences	1068	18	8		11	25	17	4	7	5
International journal of med. informatics	1372	22	13	10	7	14	14	5	7	8
Journal of med. systems	690	22	20	9	1	9	7		8	18
JAMIA	847	19	28	14	17	8	10	1	4	
Artificial intelligence in med.	627	2	1	1	7	19	2	9	40	18
Journal of biomed. informatics	614	6	8	2	15	11	3	13	25	15
IEEE transactions on info. tech. in biomed.	550	5	3	3	3	8	16	35	13	13
Computers in biology and medicine	877	2	1		2	6	3	32	14	40
Computer methods and programs in biomed.	1317	3	3	2	1	10	10	27	19	25
Journal of med. Internet research	261	12		79	1	1	6			
BMC med. informatics and decision making	199	5	13	37	15	2	9	1	18	2

search, as well as the formalization of guidelines using the GuideLine Acquisition, Representation and Execution (GLARE) system. Furthermore, the development of standards for patient records, such as the openEHR initiative, and the clinical information exchange standard HL7 and its Clinical Document Architecture (CDA) are mentioned. Interestingly, the development of the advanced health and disaster aid network (AID N) is an upcoming topic. In cluster 3 we see mostly small changes in focus, such as the shift from neural networks to Support Vector Machines (SVM), and the introduction of wavelet theory.

Journal Positioning

▶Table 5 shows the distribution of the papers of the various journals over the subclusters. Most journals have defined their own focus within the field. This is exemplified by the differences in distribution of their papers over the various clusters. To gain insight into changes of focus of the journals, we have calculated the distribution of the papers in each journal for each of the three-year intervals, and subsequently calculated the correlation between their distributions in the different periods. These correlations are shown in ▶Table 6, and indicate to what extent journals have remained in the same clusters during the past fifteen years.

Most journals appear to have changed their focus over the last fifteen years, with the exception of *Computers, Informatics, Nursing*, the *Proceedings of the AMIA Symposium*, *Computers in Biology and Medicine*, and the *Journal of Medical Internet Research*.

Discussion

Our analysis of the ISI category 'Medical Informatics' shows that it contains a subset of similar journals that pertain to MI according to the broad definitions quoted in the introduction. There are also journals in this category that arguably do not belong to the field of MI, such as for instance the *IEEE Engineering in Medicine and Biology Magazine*, which has the lowest similarity to the MI subset, and deals primarily with topics other than information technology.

The ISI category 'Medical Informatics' does not contain typical bioinformatics journals, and our similarity analysis does not show any of the bioinformatics journals to be similar to the subset of MI-related journals. It therefore seems that MI and bioinformatics are distinct fields, although our cluster analysis does show one subcluster pertaining to the analysis of gene expression data (see ▶Figures 2 and 3, cluster 3.2). Journals that deal mainly with bioinformatics are relatively high on the ranked list of journals though (▶see Table 2); for example *BMC Bioinformatics* ranks 21st, and *Bioinformatics* ranks 25th. An explanation for these high ranks is that bioinformatics uses several computer science and mathematical methods and techniques that are also used in MI, especially in the subdomain represented by cluster 3 in our analysis: signal and data analysis. This explanation is in agreement with an earlier comparison of the two fields [8].

Our analysis shows MI to be divided into three subdomains: 1) the organization, application, and evaluation of health information systems, 2) medical knowledge representation, and 3) signal and data analysis. Over the last fifteen years, there has been little change in the focus of the field at the level of main clusters. At the level of subclusters, we see a

Table 6 Correlation between the cluster distributions of journals for consecutive time periods. The last column compares the last period to the first period. Darker colors indicate higher correlations.

Period 1:	93–96	96–99	99–02	02–05	05–08
Journal Period 2:	96–99	99–02	02–05	05–08	93–96
Computers, informatics, nursing	1.00	1.00	0.99	0.98	0.94
Methods of information in medicine	0.74	0.55	0.46	0.86	–0.23
Med. informatics and the Internet in med.	0.56	0.25	0.20	0.62	–0.33
Proceedings of the AMIA symposium	0.93	0.92	0.91	0.99	0.90
Medinfo	0.92	0.92	0.65	0.90	0.64
Proceedings of the MIE conferences		0.71	0.73	0.82	
International journal of medical informatics	0.13	0.87	0.88	0.79	–0.74
Journal of medical systems	0.78	0.62	0.90	0.85	0.47
JAMIA	0.76	0.93	0.82	0.98	0.62
Artificial intelligence in medicine	0.87	0.91	0.75	0.99	0.56
Journal of biomedical informatics	0.95	0.84	–0.01	0.67	0.02
IEEE transactions on info. tech. in biomed.		0.71	0.91	0.86	
Computers in biology and medicine	0.93	0.98	0.99	0.98	0.93
Computer methods and programs in biomed.	0.83	0.94	0.95	0.98	0.57
Journal of medical Internet research			0.95	1.00	
BMC med. informatics and decision making			0.80	0.85	

strong reduction in the number of publications in cluster 2.2, which is related to clinical guidelines and decision support systems. A possible explanation for this decline is that cluster 2.2 represents mainly the fundamental and theoretical components of this field, and that attention has shifted over time to practical applications that are published elsewhere. Indeed, the growing cluster 1.2 contains many articles describing the application of decision support systems in a practical setting, often in combination with CPOE.

An analysis of publications of the last three years shows several 'hot topics' that currently receive a great deal of attention. Most journals and conferences cover only parts of the entire field of MI. Most journals show slight shifts in focus over the past fifteen years, whilst some journals, most notably the *Journal of Biomedical Informatics*, show a larger change of course in the past.

We defined the field of MI through articles from a set of journals that were selected semi-automatically. We feel this selection process is less vulnerable to subjective bias than the direct selection of journals as done for instance by Bansard et al. [8]. It should be noted that any selection of literature based on journals may miss articles that could be considered to belong to the field of MI, but are published in non-MI journals. Although we do not know the extent of this problem, we assume that the remaining set is representational for the MI field.

The feature extraction method used in this study was completely automatic, in contrast to Bansard et al. [8] who used manual curation of their features, or Rebholz-Schuhman et al. [14] who used manually determined document frequency thresholds to identify relevant terms. Additionally, in contrast to the aforementioned studies, we used a combination of uni-, bi- and trigram, instead of depending either on unigrams or bigrams. The proliferation of all three classes of n-grams in our results indicate that these are all relevant and should not be omitted.

Conclusions

In this study we identified literature pertaining to MI, and the topics discussed therein, allowing for an empirical definition of the field. Given this definition, we were able to show trends in the field, and the positioning of different journals within this field.

Acknowledgments

This study was supported in part by the Biorange project sp 4.1.1. of the Netherlands Bioinformatics Centre.

References

1. van Bemmel JH, Musen MA (eds). Handbook of Medical Informatics. 2nd edition. Heidelberg: Springer Verlag; 2002.
2. Lorenzi NM. Towards IMIA 2015 – the IMIA Strategic Plan. Methods Inf Med 2007; 46 (Suppl 1): 1–5.
3. IMIA. http://www.imia.org/strategic/PDF/IMIA_Strategic_Plan_Final.pdf 2007.
4. Elkin P. Personal communication. 2008.
5. Musen MA, van Bemmel JH. Challenges for Medical Informatics as an Academic Discipline. Methods Inf Med 2002; 41 (1): 1–3.
6. Andrews JE. An author co-citation analysis of medical informatics. J Med Libr Assoc 2003 ; 91 (1): 47–56.
7. Morris TA, McCain KW. The structure of medical informatics journal literature. J Am Med Inform Assoc 1998; 5 (5): 448–466.
8. Bansard JY, Rebholz-Schuhmann D, Cameron G, Clark D, van Mulligen E, Beltrame E, et al. Medical informatics and bioinformatics: a bibliometric study. IEEE Trans Inf Technol Biomed 2007; 11 (3): 237–243.
9. Raghupathi W, Nerur S. Research Themes and Trends in Health Information Systems. Methods Inf Med 2008; 47 (5): 435–442.
10. McCray AT, Srinivasan S, Browne AC. Lexical methods for managing variation in biomedical terminologies. Proc Annu Symp Comput Appl Med Care 1994. pp 235–239.
11. Jelier R, Schuemie MJ, Roes PJ, van Mulligen EM, Kors JA. Literature-based concept profiles for gene annotation: The issue of weighting. Int J Med Inform 2007, Sep 7.
12. Ludl M, Widmer G. Towards a Simple Clustering Criterion Based on Minimum Length Encoding. Proceedings of the European Conference on Machine Learning; 2002; Helsinki, Finland: Springer; 2002. pp 258–269.
13. Heer J, Card SK, Landay JA. Prefuse: a toolkit for interactive information visualization. CHI (Computer Human Interaction) Conference Proceedings; 2005 April 2–7; Portland, OR; 2005.
14. Rebholz-Schuhman D, Cameron G, Clark D, van Mulligen E, Coatrieux JL, Del Hoyo Barbolla E, et al. SYMBIOmatics: synergies in Medical Informatics and Bioinformatics – exploring current scientific literature for emerging topics. BMC Bioinformatics. 2007; 8 (Suppl 1): S18.