Research Subjects and Research Trends in Medical Informatics

Kemal Hakan Gülkesen1  Reinhold Haux1

1Peter L. Reichertz Institute for Medical Informatics, TU Braunschweig and Hannover Medical School, Braunschweig, Germany

Address for correspondence Kemal Hakan Gülkesen, MD, PhD, Peter L. Reichertz Institute for Medical Informatics, TU Braunschweig and Hannover Medical School, Mühlenpfordtstraße 23, 38106 Braunschweig, Germany (e-mail: hgulkesen@gmail.com).

Abstract

Objectives To identify major research subjects and trends in medical informatics research based on the current set of core medical informatics journals.

Methods Analyzing journals in the Web of Science (WoS) medical informatics category together with related categories from the years 2013 to 2017 by using a smart local moving algorithm as a clustering method for identifying the core set of journals. Text mining analysis with binary counting of abstracts from these journals published in the years 2006 to 2017 for identifying major research subjects. Building clusters based on these terms for the complete time period as well as for the periods 2006–2008, 2009–2011, 2012–2014, and 2015–2017 for identifying trends.

Results The identified cluster includes 17 core medical informatics journals. By text mining of these journals, 224,992 different terms in 14,414 articles were identified covering 550 specific key terms. Based on these key terms five clusters were identified: “Biomedical Data Analysis,” “Clinical Informatics,” “EHR and Knowledge Representation,” “Mobile Health,” and “Organizational Aspects of Health Information Systems.” No shifts in the clusters were observed between the first two 3-year periods. In the third period, some terms like “mobile phone,” “mobile apps,” and “message” appear. Also, in the third period, a “Clinical Informatics” cluster appears and persists in the fourth period. In the fourth period, a rearrangement of clusters was observed.

Conclusions Beside classical subjects of medical informatics on organizing, representing, and analyzing data, we observed new developments in the context of mobile health and clinical informatics. These subjects tended to grow over the past years, and we can expect this trend to continue.

Keywords
► medical informatics
► biomedical and health informatics
► research

Background

Medical informatics (MI), or more generally Biomedical and Health Informatics, has been most variously and often inconsistently defined.1 According to one definition, it is “concerned with the optimal use of information, often aided by the use of technology, to improve individual health, health care, public health, and biomedical research.”2 According to another it is “a discipline, concerned with the systematic organization, representation, and analysis of data, information, and knowledge in biomedicine and health care.”3 Recommendations for MI education, which were revised by the International Medical Informatics Association (IMIA), can also be a clue for defining the field of MI.4

On the other hand, MI is frequently referred to by other names, with different yet closely related meanings. “Biomedical and Health Informatics,” “Biomedical Informatics,” Health-care Informatics,” and “Clinical Informatics” are some of them.5 As the name MI is more frequently used in journal classifications (such as Institute for Scientific Information [ISI] and Science-Metrix), terminologies (PubMed), or in the names of non-governmental organizations (such as IMIA and American
Medical Informatics Association (AMIA), in the present text, we preferred to use MI as a term.

A systematic approach to help define MI could improve our understanding of its research contents by analyzing its patterns of communication through publications produced by the MI community. This would also help in designing and reshaping MI education. And it will additionally help to support management decisions and to design future research agendas.6

Several studies have been published examining the MI literature. To select MI articles or journals, the authors of these studies mainly used four approaches.

- Using the Medical Subject Headings (MeSH) indexing to define MI articles.7–11
- Composing a core MI journal set by expert opinions.12
- Clustering journals by co-citation data to determine a core MI journal set.13
- Text mining of abstracts and clustering journals by the help of the obtained terms to determine a core MI journal set.14

We also wanted to examine research subjects in MI. Our approach is to use footprints of scientific knowledge: references. Direct citation (intercitation),15 bibliographic coupling,16 and co-citation17,18 are three main approaches for clustering similar articles or journals. When an article cites another article, the relationship is denoted as direct citation. When two different articles cite an article, the relationship is called bibliographic coupling. When articles from two different journals are included in the same reference list, it is called co-citation (Fig. 1).13,19 Direct citation was reported to be more successful than the other two methods for clustering similar articles in the analysis of historical data.20 However, bibliographic coupling is possibly better in relatively short-term data analysis.15,20

A scientific community communicates its research in its scientific journals, and the contents of these journals reflect the main areas of interest of the community. The content of the MI literature has been examined in several earlier papers, with authors taking two main approaches to classify or cluster the content of MI articles or journals.

- Documentation of MeSH terms used for indexing the articles.9,11,12,21
- Extraction of terms by text mining of the abstracts.10,14

Another approach involves examining an author’s choice of keywords.22 However, both MeSH terms and author keywords are affected by human factors involving subjective bias. In the present study, we preferred text mining as a somewhat more objective method for examining and finding possible groupings for the MI literature.

Some of the previous studies divided the literature into time periods to make comparisons between them, and to identify trends.9,11,12,14 We also examine publications across different time periods to examine and compare their contents.

**Objectives**

Two questions motivated us to conduct this study:

Q1: What are the major research subjects in MI?

Q2: Do these subjects change over time? If they change, what do these changes look like?

Before being able to provide answers to these questions, another question arose. Assuming that MI research is often communicated through core MI journals:

Q0: What might be the current core MI journals?

The third and fourth sections concentrate on Q0, while our main questions, Q1 and Q2, will be examined in the fifth and sixth sections.

**Study Design, Methods, and Tools for Identifying Core Medical Informatics Journals (Q0)**

For the years 2013 to 2017, we considered all journals listed in the Web of Science (WoS) under the categories “Medical Informatics,” “Biochemical Research Methods,” “Biotechnology and Applied Microbiology,” “Mathematical and Computational Biology,” “Statistics and Probability,” “Computer Science: Information Systems,” “Health Care Sciences and Services,” “Engineering: Biomedical,” “Computer Science: Interdisciplinary Applications,” “Computer Science: Theory and Methods,” “Computer Science: Artificial Intelligence,” and “Public, Environmental and Occupational Health.” We included all papers of all journals that published 40 or more articles during this time period. The data were downloaded between June 4 and June 18, 2018. The tool used for clustering was VOSviewer (version 1.6.8).23 The reference data were extracted automatically by VOSviewer. As clustering technique the smart local moving algorithm, introduced by Waltman and Van Eck,24 was used with bibliographic coupling analysis and fractional counting.25 VOSviewer also enabled visualizing the obtained results. For this visualization, association strength was used as the normalization method, because it was recommended for bibliometric studies.26 Because too large clusters were obtained by using the default resolution value (1.00), it was increased to 4.50. We tested

---

**Fig. 1** Visual representation of citation relations. A, B, and C represent different journals and arrows represent citations. Green ovals show how the documents are clustered by each approach.
the most satisfactory clustering by 0.5 increments in resolution
value. At level 4.0 there were 39 clusters and most of the
MI journals, such as JAMIA, were in the largest cluster,
containing 89 journals, mostly health management journals.
At level 5.0 there were 54 clusters. The MI cluster was
composed of 16 journals, with all journals being the same
as level 4.5, except Artificial Intelligence in Medicine, which
was in another cluster as a single journal. We considered 4.5
as the most satisfactory resolution value. Attraction and
repulsion values (these values do not affect results but
aesthetic appearance of figures) were taken as 2 and 0,
respectively. The authors selected and named related clus-
ters independently and then came together to reach a con-
sensus decision on the final naming of these clusters. The
first author (K.H.G.) received his PhD 9 years ago and has
14 years of MI research and teaching experience in MI
departments of universities. The second author (R.H.)
received his PhD 35 years ago. He has been working as a
university professor in MI departments of universities for
31 years, and wrote several MI textbooks. Both of the authors
are or have been in editorial boards of various journals. Most
of these journals are in the field of MI.

This research does not involve human subjects, human
material, or human data.

Results for Identifying Core Medical Informatics Journals (Q0)

Downloaded data amounted to a total of 427,012 articles,
published in 867 journals, of which 807 contained 40 or
more articles.

Forty-seven clusters were obtained (see Online Supple-
mentary Material 1), and one of them included the core MI
journals (► Fig. 2). This MI cluster includes 15 journals out of
the 25 journals in the MI category in WoS, plus two addi-
tional telemedicine journals (► Table 1).

As related clusters, we defined those clusters being close to
the identified MI cluster, the clusters containing journals of
the WoS MI category, and the cluster containing bioinfor-
matics journals. Six related clusters have been found and
defined: “Bioinformatics,” “Biomedical Engineering: Imaging
and Information Technology,” “Biomedical Engineering: Bio-
mechanics and Medical Technology,” “Health Management,”
“Information Management,” and “Statistics” (► Fig. 2). For
these six journal clusters, the authors independently gave
the same names to “Bioinformatics,” “Statistics,” “Health
Management,” and “Information Management” clusters
(66.7% initial agreement rate) and two clusters of biomedical
engineering were named by a consensus decision.

Study Design, Methods, and Tools for Identifying Subjects and Trends

For the 17 core MI journals identified we downloaded the
abstracts of all these journals from articles published from
the years 2006 to 2017 in the WoS—on July 5, 2018. We
searched only “article” type documents in “Science Citation
Index Expanded.” We performed text mining analysis with
binary counting to obtain terms. Terms were one to four
word expressions. Then we were clustering these terms by
using the smart local moving algorithm with the co-word
method with binary counting.27

As the tool for text mining and co-word clustering we used
VOSviewer (version 1.6.8),18 which was also used for visual-
izing the results. Weights of links, weight of total link
strength, and weights of occurrences for each term are given
in online supplementary materials. For explanation of
meanings of these weights, please refer VOSviewer Manual.28

The text mining module of VOSviewer is based on the
Apache OpenNLP toolkit.29 Its text mining functionality is
described in the work of Van Eck and Waltman.30 Text mining
functionality of VOSviewer does not need preprocessing. It
automatically imports abstracts and processes data in five
steps: (1) removal of copyright statements; (2) sentence
detection; (3) part-of-speech tagging (using this algorithm,
each word is assigned a part of speech, such as verb, noun,
adjunctive, preposition, and so on); (4) noun phrase identi-
fication (it defines a noun phrase as a sequence of one or more
consecutive words within a sentence such that the last word
in the sequence is a noun and each of the other words is
either a noun or an adjective); and (5) noun phrase unifica-
tion (unification of noun phrases is accomplished by remov-
ing most nonalphanumeric characters, by removing accents
from characters, by converting upper case characters to
lower case, and by converting plural noun phrases to sin-

ular). For visualization of our results, association strength was
used as a normalization method. The resolution parameter
value was chosen as 1.00, and the attraction and repulsion
parameter values were taken as 2 and 1, respectively.

We selected frequently used terms—which were
found in at least 50 articles (and also used 25 as noted below).
The authors gave names to the clusters independently and
then came together to reach a consensus on the final naming
of these clusters.

We divided the articles into four groups, each covering a
2015–2017. We analyzed each group in the same way as
mentioned above. We selected those terms which are used
in at least 25 articles. Resolution values for clustering were
set as 1.00, 1.10, 1.20, and 1.20 for the periods, respectively.
Result of the default resolution value 1.0 was satisfactory
for us in the first period, but we needed to increase the
resolution values in the following periods to obtain similar
clusters. This may be due to increasing number of articles
with time.

This research does not involve human subjects, human
material, or human data.

Results for Identifying Subjects and Trends
On Major Research Subjects in medical informatics
(Q,)

By text mining of abstracts of articles from the 17 core MI
journals, we detected 224,992 different terms in 14,414
articles. We scanned all terms, which were used in more
than 50 articles and found the terms with the same or similar
meaning. We converted the terms with the same or similar meaning to the most frequent one by the “replace by” function of VOSviewer to combine them as a single term (Online Supplementary Material 2). We obtained 1,334 such combined terms at the end of this process. A relevance score was calculated for each term, and 800 terms—that is, approximately 60% of the most relevant terms—were selected for manual processing. There we eliminated non-specific terms (Online Supplementary Material 3). In the end, 550 distinct terms were obtained. Cluster analysis of these terms revealed five different clusters (Online Supplementary Material 4). We named these five clusters as (1) “Mobile Health,” (2) “Organizational Aspects of Health Information Systems,” (3) “Biomedical Data Analysis” (4) “EHR and
Knowledge Representation,” and (5) “Clinical Informatics” (Fig. 3). The authors gave the same names to all clusters except “EHR and Knowledge Representation” (80.0% initial agreement rate).

The most frequent 24 terms in each cluster are presented in Table 2.

How Subjects can Change over Time (Q2)

The 2006 to 2008 Period
There were 402 terms in 2,001 articles and 241 of them were in the most relevant 60%. After elimination of nonspecific terms, 142 of them were selected for cluster analysis. At the end of our analysis, we obtained four groups (Online Supplementary Material 5; Fig. 4).

Pattern of Clusters for the 2009 to 2011 Period
In 2,765 articles in this period, there were 553 terms and 332 of them were in the most relevant 60%. After nonspecific terms were eliminated, the remaining 205 terms were used for cluster analysis (Online Supplementary Material 6; Fig. 5).

Pattern of Clusters for the 2012 to 2014 Period
There were 4,378 articles and 916 terms. A total of 550 were in the most relevant 60%. After nonspecific terms were eliminated, the remaining 368 terms were used for cluster analysis. (Online Supplementary Material 7; Fig. 6).

Pattern of Clusters for the 2015 to 2017 Period
For 5,270 articles in this period, there were 1,100 terms and 660 were in the most relevant 60%. After nonspecific terms were eliminated, the remaining 449 terms were used for cluster analysis (Online Supplementary Material 8; Fig. 7).

Comparison of the Clustering Groups over the Different Periods
The first two periods seem to be similar to each other in both numbers and content of the clusters. In the third period,
some terms like "mobile phone," "mobile apps," and "message" appear in the cluster, which we named telehealth in the previous periods. We named this cluster as mobile health in periods three and four. In the third period, a relatively small clinical informatics cluster appears and persists in the fourth period. In the fourth period, we observe a rearrangement of clusters.

Discussion

According to our cluster analysis for journals, the core MI journals we identified overlap only partially with the WoS MI category. Ten journals in the WoS MI category were identified as belonging to other clusters, and two journals in the "Healthcare Sciences and Services" category were identified as belonging to the MI cluster. Another classification of scientific journals was made by Science-Metrix. Its categories were modeled on those of existing journal classifications, and their groupings of journals acted as "seeds" or attractors for journals in the new classification. Individual journals were assigned to single, mutually exclusive categories via a hybrid approach combining algorithmic methods (using citation data and author addresses) and expert judgment. MI category contains 30 journals. Three journals in our MI cluster were published after this classification, so they are not present in the list. Other three well-known MI journals in our MI cluster (CIN—Computers Informatics Nursing, Health Information Management Journal, and Informatics for Health and Social Care) are also not present in the list. The list contains eight medical education journals, a few MI journals which are not included by WOS, and a few journals which are in different clusters in our clustering results. Our results only partially overlap with Science-Metrix classification.

Our cluster analysis revealed 47 clusters of journals. Considering the six related clusters, according to our analysis, the "Bioinformatics" and "Statistics" clusters were close to each other, whereas, to our surprise, they did not have a close relationship with the MI cluster. However, "Health management," "Information Management," and "Biomedical Engineering: Imaging and Information Technology" clusters are three close neighbors of the MI cluster. Although they are under the "Biomedical Informatics" umbrella, the MI and Bioinformatics scientific communities have divergent features related to their scientific conferences and journals. Deeper insight into this situation as well as some suggestions to increase communication between these scientific fields have been discussed previously. On the other hand, with...
<table>
<thead>
<tr>
<th>Cluster 1 (168 terms)</th>
<th>Cluster 2 (126 terms)</th>
<th>Cluster 3 (112 terms)</th>
<th>Cluster 4 (107 terms)</th>
<th>Cluster 5 (39 terms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobile health</td>
<td>Organizational aspects of health information systems</td>
<td>Biomedical data analysis</td>
<td>EHR and knowledge representation</td>
<td>Clinical informatics</td>
</tr>
<tr>
<td>Term</td>
<td>n</td>
<td>%</td>
<td>Term</td>
<td>n</td>
</tr>
<tr>
<td>Web</td>
<td>1,288</td>
<td>8.9</td>
<td>Provider</td>
<td>914</td>
</tr>
<tr>
<td>Internet</td>
<td>1,062</td>
<td>7.4</td>
<td>Nurse</td>
<td>910</td>
</tr>
<tr>
<td>Education</td>
<td>907</td>
<td>6.3</td>
<td>Barrier</td>
<td>790</td>
</tr>
<tr>
<td>Behavior</td>
<td>868</td>
<td>6.0</td>
<td>Organization</td>
<td>778</td>
</tr>
<tr>
<td>Clinics</td>
<td>766</td>
<td>5.3</td>
<td>Adoption</td>
<td>761</td>
</tr>
<tr>
<td>Visit</td>
<td>727</td>
<td>5.0</td>
<td>Perception</td>
<td>678</td>
</tr>
<tr>
<td>Web site</td>
<td>661</td>
<td>4.6</td>
<td>Concern</td>
<td>668</td>
</tr>
<tr>
<td>Feedback</td>
<td>641</td>
<td>4.4</td>
<td>Telehealth</td>
<td>668</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>620</td>
<td>4.3</td>
<td>Staff</td>
<td>567</td>
</tr>
<tr>
<td>Mobile phone</td>
<td>584</td>
<td>4.1</td>
<td>Information system</td>
<td>481</td>
</tr>
<tr>
<td>Home</td>
<td>568</td>
<td>3.9</td>
<td>Consultation</td>
<td>478</td>
</tr>
<tr>
<td>Health information</td>
<td>556</td>
<td>3.9</td>
<td>Workflow</td>
<td>446</td>
</tr>
<tr>
<td>Mobile apps</td>
<td>531</td>
<td>3.7</td>
<td>Acceptance</td>
<td>432</td>
</tr>
<tr>
<td>Student</td>
<td>530</td>
<td>3.7</td>
<td>Privacy</td>
<td>415</td>
</tr>
<tr>
<td>Diabetes</td>
<td>521</td>
<td>3.6</td>
<td>Policy</td>
<td>410</td>
</tr>
<tr>
<td>Symptom</td>
<td>514</td>
<td>3.6</td>
<td>Case study</td>
<td>381</td>
</tr>
<tr>
<td>Message</td>
<td>490</td>
<td>3.4</td>
<td>Health care provider</td>
<td>381</td>
</tr>
<tr>
<td>Attitude</td>
<td>475</td>
<td>3.3</td>
<td>Security</td>
<td>377</td>
</tr>
<tr>
<td>Skill</td>
<td>462</td>
<td>3.2</td>
<td>Infrastructure</td>
<td>345</td>
</tr>
<tr>
<td>Children</td>
<td>437</td>
<td>3.0</td>
<td>Health record</td>
<td>322</td>
</tr>
<tr>
<td>Face</td>
<td>417</td>
<td>2.9</td>
<td>Collaboration</td>
<td>321</td>
</tr>
<tr>
<td>Engagement</td>
<td>414</td>
<td>2.9</td>
<td>Focus group</td>
<td>318</td>
</tr>
<tr>
<td>Preference</td>
<td>411</td>
<td>2.9</td>
<td>Exchange</td>
<td>299</td>
</tr>
<tr>
<td>Efficacy</td>
<td>399</td>
<td>2.8</td>
<td>Consumer</td>
<td>275</td>
</tr>
</tbody>
</table>

Abbreviations: ADE, adverse drug events; CDSS, clinical decision support systems; COPD, chronic obstructive lung disease; CPOE, computerized physician order entry; EHR, electronic health records; ICU, intensive care unit; SVM, support vector machines.
Fig. 4 Cluster map of the terms obtained by text mining from articles from core MI journals published in 2006–2008. Group topics involve: red: TH, Telehealth, green: OA, Organizational Aspects of Health Information Systems, yellow: KR, Knowledge Representation, and blue: BMDA, Biomedical Data Analysis.

Fig. 5 Cluster map of the terms obtained by text mining from articles of core MI journals published in 2009–2011. Topics are colored as: red: TH, Telehealth, green: OA, Organizational Aspects of Health Information Systems, yellow: KR, Knowledge Representation, and blue: BMDA, Biomedical Data Analysis.
Fig. 6 Cluster map of the terms obtained by text mining from articles of core MI journals published in 2012–2014. Topics: red: MH, Mobile Health, green: OA, Organizational Aspects of Health Information Systems, yellow: KR, Knowledge Representation, blue: BMDA, Biomedical Data Analysis, and purple: CI, Clinical Informatics.

increasing efforts to integrate molecular data with those from electronic health records, we can expect a closer relation between bioinformatics and MI to develop in the future.

Although there is only some overlap between the MI cluster as identified in this analysis with the WoS MI category, in our opinion our clustering analysis results are more reasonable in some respects. For example, journals like “Statistical Methods in Medical Research” and “Statistics in Medicine” assigned to the WoS MI category are clearly not MI journals, and somehow conflate the well-known fact that statistical analyses are frequently reported in informatics papers with the notion that journals focused on statistical methods in medicine are likely to be informatics-related, which is clearly not the case. On the other hand, a journal can be assigned to more than one category in the WoS categorization system. Because our clustering method assigns each journal to only one cluster, some journals—such as “Computer Methods and Programs in Biomedicine”—with relatively low MI content may be assigned to other clusters and not to the MI cluster.

We also tried to compare our journal clustering results with previous studies, although comparisons of this type are problematic, among other reasons, because of the different time periods covered and ever-shifting professional practices, which often make the content of the journals change over time.14

- Using MeSH: Some of studies using MeSH have not mentioned journal names and other such studies have produced veryquestionably relevant results. For example, according to one study, the journals considered to be the most prominent ones publishing MI articles are “Proceedings of IEEE Engineering in Medicine and Biology Society Conference,” “IEEE Transactions on Image Processing,” and “Medical Physics.”9 This clear overspecific focus on journals that emphasize engineering or computational methodologies for analysis and design, and not the informatics methods used in most studies (which may, however, implicitly rely on engineering and computational implementations), may arise from using MeSH inappropriately for defining MI articles, and related to problems in the MeSH indexing structure and implementations. For example, in one article, 63% of the articles, indexed by the telemedicine term, were found to have not been indexed by MI or bioinformatics terms.35 According to another study, the sensitivity of MeSH-term-based search is 60% and one-third of the obtained articles were found to be irrelevant to the intended subject.36 On the other hand, searching by MeSH terms has the capability of detecting some important papers in other journals because of core and scatter phenomenon. Core and scatter is the distinctive pattern of concentration and dispersion that appears in collections of papers when relative frequencies of entities are counted. In the context of mapping specialties, core and scatter has a significant effect on gathering a collection of papers to cover the specialty. On the one hand, it is usually easy to find a group of highly relevant papers that cover the core of the specialty, but on the other, it becomes increasingly laborious to gather all papers with some significant relevance, and impossible to gather all papers that are marginally relevant to the specialty.15 According to a MeSH-term-based research, 30 journals represented the first third of the total published articles in the MI field.11

- Composing a core MI journal set by expert opinion: In another study published in 2017, the authors defined which journals “belonged” to the MI category according to expert opinions.12 They made a list of 36 MI journals. This list includes all of our core MI journals except for two very new Journal of Medical Internet Research (JMIR) journals. According to their classification, “Computer Methods and Programs in Biomedicine” and “IEEE Journal of Biomedical and Health Informatics” are also in the group of MI journals. The remaining journals were MI journals or proceedings which are not covered by ISI, or health information management journals which are not
covered by ISI or in the health management cluster in our classification.

- **Clustering journals by co-citation data to determine a core MI journal set**: The study, which was based on the co-citation method, is rather old and therefore hardly comparable to our study (1993–1995 vs. 2013–2017).  

- **Text mining of abstracts and clustering journals by the help of the obtained terms to determine a core MI journal set**: In a study published in 2009, the results were similar except they included “Computer Methods and Programs in Biomedicine,” and the IEEE Journal of Biomedical and Health Informatics (under its previous name “IEEE Transactions on Information Technology in Biomedicine”) in the set of core MI journals, and they did not include two telemedicine journals. Naturally, new journals such as “Applied Clinical Informatics” were not included in this study. The difference may be due to a difference in the way in which research methods for clustering were applied—such as clustering based on terms versus clustering based on citations or by different time periods of the studies (1993–2008 vs. 2013–2017).

It seems that clustering by text mining gives the closest results to our method. The main difference of the two methods is that text mining presents a classification based on the use of words, terms, and concepts, whereas bibliographic coupling presents a classification based on the flow of scientific information, knowledge, and ideas. In other words, the first method answers the question of “how can one classify journals according to the use of terminology,” whereas the latter method answers the question of “how can one classify journals according to similarity of information or knowledge, which they present.” The latter, of course, depends crucially on the definition of “similarity” and how it is computed in relation to the individual and groups of items being classified—journals in this case. So, the preferred method may change according to the point of view, choice of methods, and even the techniques of implementation chosen by a researcher.

We preferred using text mining instead of using a controlled vocabulary or author keywords for analyzing MI subjects. The advantage of text mining is its objectivity (absence of intervention by an author or an indexer) and capability of detecting new terms. On the other hand, it has a disadvantage of resulting in a disorganized bunch of terms. In a controlled vocabulary, such as MeSH, synonyms are collected under the same term and the terms are organized ontologically. Results of text mining require some (more or less) manual work to collect synonyms under the same umbrella and interpretation of the results is more difficult.

As a result of the terms obtained by text mining, the five clusters of terms in the examined period 2006 to 2017 can be described as follows.

- **Mobile Health**: Typical terms in this cluster are “web,” “education,” “mobile phone,” “home,” “mobile apps,” “diabetes,” and “message.” This cluster is the result of the effects of introducing new mobile technologies in health care applications. It seems that there is substantial research on mobile apps, homecare, online education, and diabetes. We were observing this trend, but it is still surprising to see them as a separate big cluster. This cluster was called “Telehealth” in the first two periods (2006–2011). By the appearance of the “mobile health” term, the “telehealth” term migrated to the “Clinical Informatics” cluster in the third period (2012–2014), and finally to the “Organizational Aspects” cluster in the fourth period (2015–2017). Probably this change corresponds to the increasing integration of telehealth practices in routinely used information systems.

- **Organizational Aspects of Health Information Systems**: we observed terms such as “provider,” “nurse,” “barrier,” “organization,” “adoption,” “perception,” “concern,” “telehealth,” and “privacy” in this cluster. This cluster represents an important aspect of MI. It reflects the studies on the relation of information systems with organizations and people. This cluster is the only one which ceased to enlarge in the last period (2015–2017). This may be a result of the widespread use of health information systems in health care institutions and the increasing acceptance of them, which means that early introduction problems are no longer central, while often-heard complaints about the inadequacies of health care systems and their detriments to clinical practice and workflows are not included under organizational aspects, possibly due to the socio-economic complexities involved and the sensitivities of industry and governments to such complaints.

- **EHR and Knowledge Representation**: Terms such as “EHR,” “term,” “concept,” “methodology,” “structure,” “identification,” and “integration” were most prominent in this cluster. This cluster also reflects an important field of MI. The “EHR” term was in the “Organizational Aspects” cluster at the beginning, but it migrated to this cluster in the last period (2015–2017). It may show that the acceptance phase of EHR as a concept is coming to an end, and researchers are concentrating on the technical aspects of EHR.

- **Biomedical Data Analysis**: Typical terms in this cluster are “algorithm,” “diagnosis,” “dataset,” “classification,” “image,” “detection,” “classifier,” “prediction,” and “machine.” It seems that this cluster contains mostly terms related to theoretical as well as a few practical aspects of decision support and data analysis systems for biomedical research, including machine learning and imaging informatics.

- **Clinical Informatics**: We observe “drug,” “CDSS,” “alert,” “emergency department,” “patient safety,” “admission,” “heart failure,” and “CPOE” as typical terms in this cluster. This is a new and relatively small cluster, and reflects the ultimate aim of MI, i.e., to support better health care services, though one might expect some of the organizational issues that are arising in acceptance of such systems to migrate to this cluster in the future.

Studying MI terms in the literature has been performed in several studies. However, we found only two studies, which also dealt explicitly with the clustering of these terms. In the
first study, MeSH terms for articles published in 20 MI journals in the period of 1995 to 1999 were clustered. The authors found eight clusters, namely “Imaging Techniques,” “Diagnostic Imaging,” “Science and the Art of Medicine,” “Statistical Analysis,” “Biological Communications,” “Cognitive and Physiological Communication Concepts,” “Immunology,” and “Molecular Genetics.” Results of this article are not comparable to those from our study because of differences in research methods (MeSH terms vs. text mining) and the many technological and practice changes between the periods studied (1995–1999 vs. 2006–2017).16

In another study, abstracts of 16 MI journals published in the period of 1993 to 2008 were text mined and the obtained terms were clustered. The authors obtained three main clusters. They did not name them but described: “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 in 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.”14 These clusters are similar to our clusters “Organizational Aspects of Health Information Systems,” “EHR and Knowledge Representation,” and “Biomedical Data Analysis,” respectively. We found two additional clusters in our analysis—namely “Mobile Health” and “Clinical Informatics.” When we consider that these two clusters were not present in articles from our first two periods, we can conclude that these subjects represent rising subfield trends which are likely to continue.

Our study has several limitations, mostly due to the complex nature of our research subject.

One limitation is that we have limited ourselves by only considering sources, which are indexed by WoS. Therefore, our clustering approach did not include proceedings of important MI meetings such as MEDINFO, MIE, and the AMIA Annual Symposium. There are also a few MI journals which are not covered by ISI, and we could not include them, because our clustering is based on reference data in the ISI database. In addition, we are aware of the fact that MI articles are also published in a wide range of journals, often in related but frequently only loosely related disciplines in their relationship to MI. However, since we are convinced that most of high-quality MI research is communicated through core MI journals, we nevertheless wanted to focus on these journals.

The clustering method itself imposes several limitations besides the foundational one of choosing a similarity measure for the clustering. The size of the clusters can be chosen to be either smaller or larger, and to include fewer or more journals. This depends on just how “loosely and generally” one wishes to define such a heterogeneous and complex field of study and application as MI. There has been long-standing discussion in the discipline initiated by van Bemmelen17 and others38–40 on the very definition of MI as art versus science and implicitly the problems of clinical practice versus bio-

Conflict of Interest
Both of the authors are or have been in editorial boards of various journals. Most of these journals are in the field of MI.

Acknowledgment
The authors would like to thank Casimir Kulikowski for his support during the initial stage of manuscript preparation. Through his edits, he not only turned the text into a much better readable one, but also helped to further reflect on methodological approach and on its limitations.

References
Klavans R, Boyack KW. Which type of citation analysis generates the most accurate taxonomy of science? J Assoc Inf Sci Technol 2010;60(08):1536–59.


Klavans R, Boyack KW. Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge? J Assoc Inf Sci Technol 2017;68(04):984–998.


