Structural Relationships Within Medical Informatics

Theodore A. Morris, MLS, University of Cincinnati Medical Center, Cincinnati, OH USA

Abstract

This study seeks to increase our understanding of the structure of Medical Informatics. In particular, it focuses on the relationships between information science and information technology on the one hand, and biomedical research, clinical practice, and medical education on the other, that have defined "medical informatics." Using indexing terms and MeSH tree structures assigned to medical informatics literature covered by MEDLINE, co-occurrence analysis provides a "map" of the field. Major research and application focuses arrayed within the map elucidate a finer structure than reported previously. Dimensions "Techniques vs. Systems" and "Signs & Symptoms vs. Processes" form the two axes of the map and relate to the relationships underlying the indexing assignments given to the literature studied. Related studies underway using the INSPEC database will provide a complementary perspective on the structure of medical informatics as a field.

Introduction

Medical Informatics has been defined by many authors, but the consistent thread running through all definitions is the confluence of information science and information technology (IS/IT) with the research problems and applications within biomedical research, clinical practice, and teaching. Medical informatics, more broadly interdisciplinary than other fields, provides more difficulty for performing structural studies. As part of a larger work [1], this study considers "What constitutes medical informatics?" and "What is the relationship between biomedicine and information science/information technology, that has defined medical informatics?"

Background

The use of information technology is pervasive throughout medical research, practice, and education. To better support the information seeking and information management needs of knowledge workers in the field, we need to understand better what are the major research areas within the field, from what fields and disciplines it borrows, and to what fields and disciplines it contributes.

Previous studies have shown that co-occurrence analysis of scholarly and professional literature in a field can effectively yield insights into the internal structure of the field. Such studies may be performed using citations, indexing terms, or the uncontrolled text of authors' works. Analysis may be performed on groups of important documents, to discover underlying intellectual relationships among ideas expressed in those documents. Authors names may also serve as the subject of such analysis, where individual authors' collective outputs (oeuvres) "define" the underlying ideas among which relationships are sought. At an even higher level of aggregation, journals provide in their backruns a collective expression of intellectual scope that can be related to that of other journals.

Morris & McCain [2] "mapped" major research areas within medical informatics in a recent journal cocitation study. However, they noted that authors whose writings are published in medical informatics journals do not appear to cite heavily into the information science and information technology journal literature. The study at hand uses "co-word" analysis of indexing terms assigned to the medical informatics literature to provide a complementary picture of the field to that obtained by Morris & McCain.

Tijssen [3] suggests that such co-word analyses can yield structural details, especially in large multidisciplinary fields, and in those fields where citations "do not sufficiently cover research activities." [4] McCain [5] found complementary *retrieval results* using term co-occurrence and citation co-occurrence search strategies, while Braam *et al.* found the differing results from cocitation data and word profiles of documents studied to be useful in interpreting the changes over time the structural changes a field undergoes. [6] Other reports suggesting core literatures in medical informatics and related fields have not sought to provide a graphical visualization of the field [7-15].

Methodology

The core medical informatics journal set identified by Morris & McCain served as a starting point. Searches for indexing records in MEDLINE (DIALOG File 4), corresponding to publications in those journals during 1995-1999, yielded a retrieval set from which the MeSH terms (/DE) and tree structure numbers (DC=) assigned were ranked. Subject heading term retrieval counts were combined ignoring subheadings. Tree structure numbers were truncated after the fifth numeric segment as needed (e.g., E1.370.350.600.300+ *Image Enhancement*) and subsumed tree numbers found in the retrieval set were combined with their higher-level parent tree numbers in subsequent searches. Tree numbers were not exploded; only those lower-level tree numbers retrieved in the searches were combined into subsequent searches. The resulting set of ranked terms and codes were then used to search the entire MEDLINE database for publication years 1995-1999. This yielded a set of records for indexed publications throughout the database that shared important classification and indexing assignments with the core journals' output for the same period.

Tree numbers and tree number "groups" were identified subjectively as being "biomedical," "IS/IT" related, or other (demographic and geographic terms, check tags, etc.) In preparation for co-occurrence analysis, the highly-posted (frequently occurring) IS/IT terms were combined in a single search statement, to serve as a "filter" for IS/IT context. [16] Additional highly posted IS/IT terms were added to this set to expand its context. (The limiting factor in adding terms to the filter is the amount of processing overhead one is willing to impose on the search process.) The filter was applied to every pairwise comparison search performed for the final data set. This insured that the co-occurrence of two highly posted biomedical terms from earlier steps was only considered because they also shared an IS/IT perspective. Using the top 36 terms identified as biomedical and the top 36 IS/IT terms (72 terms total), final data results included 2556 search statements ($n(n-1)/2 = 72(71)/2$). The use of 72 terms overall sought to provide enough detail for meaningful interpretation while conserving the practical limits of the search system. The retrieved co-occurrence data were entered into Excel as a 72x72 square matrix, using mean co-occurrence values for the diagonal cells. The raw co-occurrence data were converted to proximity data (Pearson's r) using SPSS®, indicating the similarities among the co-occurrence patterns in the data set. The proximity data were analyzed with cluster analysis (complete linkage method) and multidimensional scaling (ALSCAL).

Results

The map in Figure 1 shows the results of the multidimensional scaling analysis. This two-dimensional model explains 94% of the variance in the data, with low stress (Kruskal's Stress Factor 1) of only .13. Eight clusters were identified, and cluster "boundaries" have been overlaid on the map for clarity. Figures 2-6 show more detail of individual clusters.

Diagnostic Imaging (S03) and Image Enhancement (S08) [17] are representative of their respective clusters on the left side of the map, anchoring a dimension with one pole representing Signs & Symptoms. The clusters containing Biomedical Communications terms and Immunology terms on the right side of the display are examples of a focus on the patient at an organ or system level, anchoring the other end of this dimension with the sense of "Underlying Processes."

"Techniques" vs. "Systems:" focus appears to distinguish terms along the vertical dimension. For example, statistics and outcomes terms included in the "Statistical Analysis" cluster focus on activities, while terms in the "Cognitive and Physiological Communications Concepts" and "Molecular Genetics" clusters treat more on information systems inherent to study of their respective topics. See Figure 7.

Discussion

Visualization techniques for interpreting large data sets are increasingly important. They give those "interested in the literature additional aids to judgment and decision." [18] The representation of the structure of medical informatics shown in Figure 1 shows a perspective on the field as derived from MEDLINE indexing. Not surprisingly, the use of computers in medicine is both central and pervasive to this perspective. Certain "core" activities dependent on medical informatics, such as cardiovascular disease, cancer, etc., appear in the center of the map. Arrayed about the center are specific interest areas such as psychology (cognition), communication, statistics, bio-molecular phenomena, and imaging.

The larger study of which this work is a part includes a similar treatment of indexing records from the INSPEC database (produced by the Institute of Electrical Engineers) will provide an IS/IT perspective on the structure of medical informatics, complementary to the biomedical perspective described here. Together, they are intended to increase our appreciation for the structure, scope, and breadth of medical informatics.

References

[1] Morris TA. Structural relationships within medical informatics: A classification/indexing

co-occurrence analysis (working title). PhD dissertation, Drexel University (in progress).

[2] Morris TA, McCain KW. The structure of medical informatics journal literature. J Am Med Inform Soc. 1998;5:448-466.

[3] Tijssen RJW. A scientometric cognitive study of neural network research: Expert mental maps versus bibliometric maps. Scientometrics. 1993;28:111-136.

[4] Tijssen RJW, van Raan AFJ. Mapping changes in science and technology: bibliometric co-occurrence analysis of the R&D literature. Eval Rev. 1994;18:98-115.

[5] McCain KW. Descriptor and citation retrieval in the medical behavioral sciences literature; Retrieval overlaps and novelty distribution. J Am Soc Inform Sci. 1989;40:110-114.

[6] Braam RR, Moed HF, van Raan AFJ. Mapping of science by combined co-citation and word analysis. I. structural aspects. J Am Soc Inform Sci. 1991;42:233-251.

[7] Miller MD. Citation patterns of articles published in the Journal of Medical Education. J Med Ed. 1982;5:797-799

[8] Beck JR, Pyle KI, Lusted LB. A citation analysis of the field of medical decision making, 1959-1982: Computer-aided diagnosis and clinical decision analysis. Med Decis Making. 1984;4:449-468.

[9] Pyle KI, Lobel RW, Beck JR. Citation analysis of the field of medical decision making: Update, 1959-1985. Med Decis Making, 1988;8:155-164.

[10] Greenes RA, Siegel ER. Characterization of an emerging field: Approaches to defining the literature and disciplinary boundaries of medical informatics. In: Proceedings, Symposium on Computer Applications in Medical Care, Eleventh, Washington, DC. Los Angeles: IEEE Computer Society Press. 1987:411-415.

[11] Sittig DF, Kaalaas-Sittig J. A quantitative ranking of the biomedical informatics serials, Meth Inform Med. 1995;34:397-410.

[12] Sittig DF, Kaalaas-Sittig J. A citation analysis of medical informatics journals, Medinfo '95: Proceedings, World Congress on Medical Informatics, 8th, Vancouver, B.C., Canada, July 23-27, 1995. Edmonton, Albert: International Medical Informatics Association; Healthcare Computing & Communications Canada. 1995:1452-1456.

[13] Sittig DF. Identifying a core set of medical informatics serials: An analysis using the MEDLINE database. Bull Med Libr Assoc. 1996;84:200-204.

[14] Sittig DF. Use of fuzzy set theory to extend Dhawan's journal selection model: Ranking the biomedical informatics serials. Bull Med Libr Assoc. 1999;87:43-9.

[15] Vishwanatham R. Citation analysis in journal rankings: Medical informatics in the library and information science literature. Bull Med Libr Assoc. 1998;86:518-522.

[16]Rogers LA, Anderson J. A new approach to defining a multidisciplinary field of science: The case of cardiovascular biology. Scientometrics. 1993;28:61-77.

[17] S03, S08, etc. are variable labels assigned that were subject to processing limitations. Variable labels and their corresponding concepts appear in Table 1.

[18] White HD, McCain KW. Visualization of literatures. In: Williams ME, ed. Annual Review of Information Science and Technology (ARIST). 1997;32:99-168.

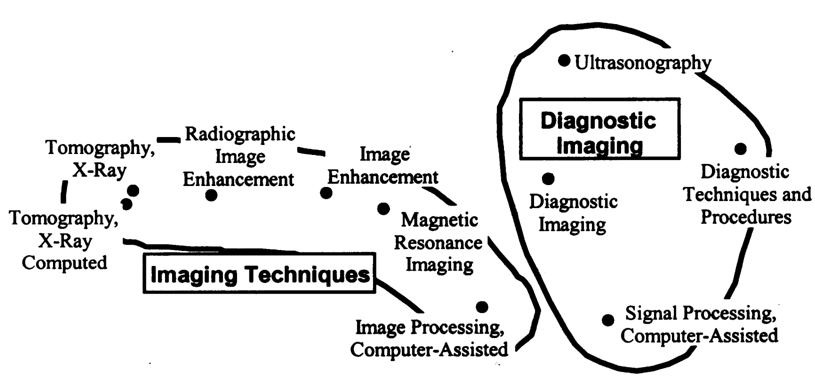

Figure 1.
Detail of Clusters
"Imaging Techniques"
And "Diagnostic Imaging"

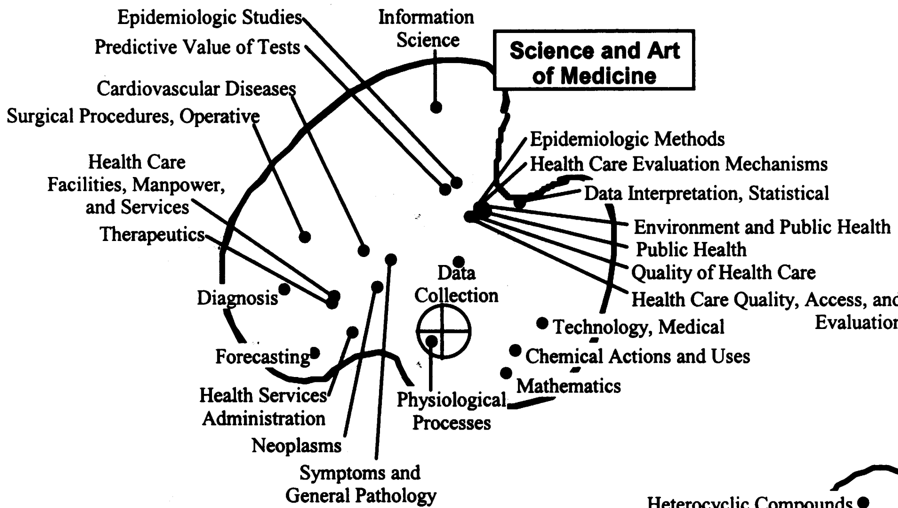

Figure 2. Detail of Cluster "Science and Art of Medicine"

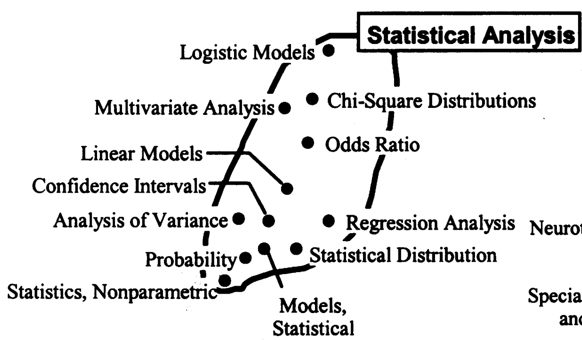

Figure 3. Detail of Cluster "Statistical Analysis"

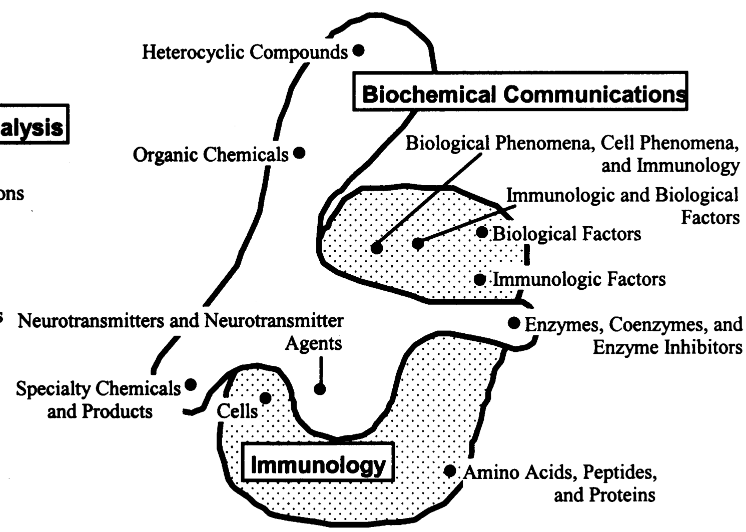

Figure 4. Detail of Clusters "Biochemical Communications" And "Immunology"

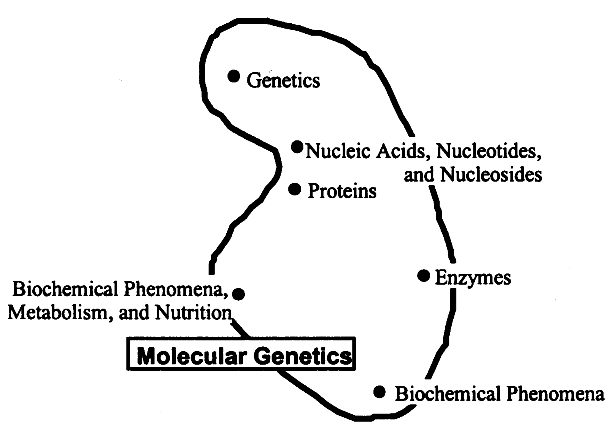

Figure 5. Detail of Cluster "Molecular Genetics"

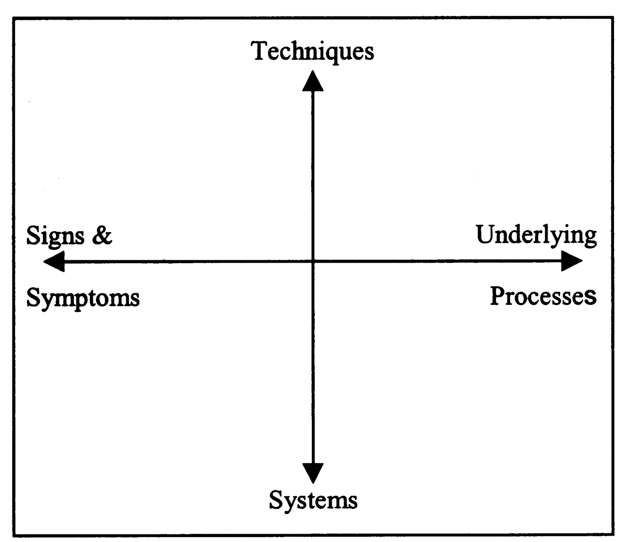

Figure 6. Underlying Dimensions

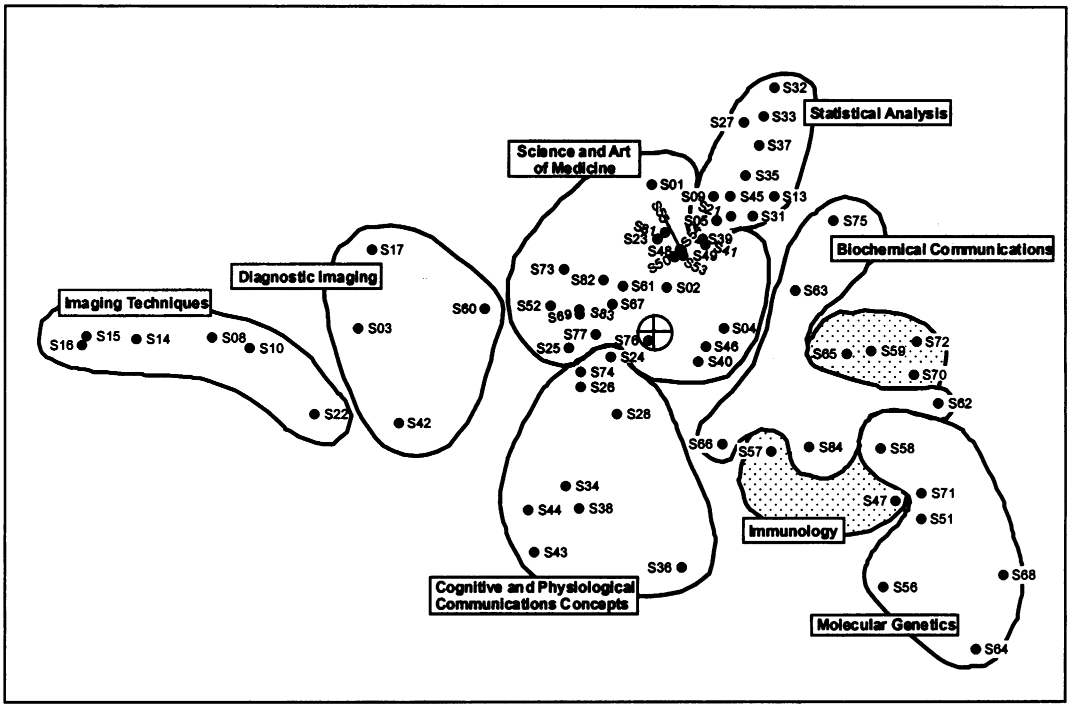

Figure 7. MEDLINE Multidimensional Scaling Display with Cluster Boundaries

Set	Query Concept
S01	Information Science
S02	Data Collection
S03	Diagnostic Imaging
S04	Technology, Medical
S05	Probability
S08	Image Enhancement (6)
S09	Analysis of Variance
S10	Magnetic Resonance Imaging
S13	Regression Analysis (2)
S14	Radiographic Image Enhancement
S15	Tomography, X-Ray (7)
S16	Tomography, X-Ray Computed
S17	Ultrasonography
S21	Models, Statistical
S22	Image Processing, Computer-Assisted
S23	Predictive Value of Tests
S24	Mathematics Computing (3)
S25	Forecasting
S26	Algorithms
S27	Multivariate Analysis (1)
S28	Computer Simulation
S31	Statistical Distributions
S32	Logistic Models
S33	Chi-Square Distribution
S34	Software
S35	Linear Models
S36	Databases
S37	Odds Ratio
S38	Terminology
S39	Statistics, Nonparametric
S40	Mathematics
S41	Data Interpretation, Statistical
S42	Signal Processing, Computer-Assisted
S43	Communication
S44	Artificial Intelligence
S45	Confidence Intervals
S46	Chemical Actions and Uses
S47	Amino Acids, Peptides, and Proteins
S48	Environment and Public Health
S49	Public Health
S50	Health Care Quality, Access, and Evaluation
S51	Proteins
S52	Diagnosis
S53	Quality of Health Care
S54	Health Care Evaluation Mechanisms
S55	Epidemiologic Methods
S56	Biochemical Phenomena, Metabolism, and Nutrition
S57	Cells
S58	Genetics
S59	Immunologic and Biological Factors
S60	Diagnostic Techniques and Procedures
S61	Symptoms and General Pathology
S62	Enzymes, Coenzymes, and Enzyme Inhibitors
S63	Organic Chemicals
S64	Biochemical Phenomena (4)
S65	Biological Phenomena, Cell Phenomena, and Immunology
S66	Specialty Chemicals and Products
S67	Neoplasms
S68	Enzymes
S69	Therapeutics
S70	Immunologic Factors (5)
S71	Nucleic Acids, Nucleotides, and Nucleosides
S72	Biological Factors (5)
S73	Surgical Procedures, Operative
S74	Musculoskeletal, Neural, and Ocular Physiology
S75	Heterocyclic Compounds
S76	Physiological Processes
S77	Health Services Administration
S81	Epidemiologic Studies
S82	Cardiovascular Diseases
S83	Health Care Facilities, Manpower, and Services
S84	Neurotransmitters and Neurotransmitter Agents

Table 1. Retrieval Query Sets and Concepts