Methods of Information in Medicine

J. J. Cimino

Columbia University,
New York, NY, USA

Editorial

The Concepts of Language and the Language of Concepts

The field of medical informatics often seems obscure to other health care professionals, who don't understand why we have not yet solved what seem to be straightforward problems, such as the development of ubiquitous electronic patient records. An analogous situation occurs, when considering terminological work: the development of controlled terminologies for use in medical systems seems to be a straightforward problem as well. In fact, many researchers and developers express surprise when they learn that there is a problem at all – that there are not already ubiquitous standards for representing health care data. Those of us who labor towards a solution, find that it is not sufficient to simply generate as many terms as we can think of and cram them into some appealing classification. Instead, we must return to reasoning based on first principles – indeed, we often need to go back and create first principles where none have been developed. From these, we can construct a principled, conceptual framework for the language of health care by which we can start to solve the "terminology problem".

As we hypothesize and test these principles, we must be careful to avoid reinventing the wheel and – instead – to build, wherever possible, upon the works of others. This turns out to be challenging, because researchers often work in isolation, with little chance to interact. Due to this isolation and because computer-based terminological work is a new field, there is no common language by which we can communicate our ideas through the literature. Although we are often talking about the same thing, we do so in confusing dialects, with seemingly interchangeable phrases, such as "terminology", "vocabulary", "nomenclature", "taxonomy", "dictionary", "language", "attributes", "properties", and so on. Clearly, to solve the "terminology problem" for medical informatics, we must solve our own internal "terminology of terminology" problem first.

One way to accomplish this is through working conferences. The Canon group, an informal aggregation of terminological researchers, found that publications, electronic mail and telephone conversations were insufficient for establishing meaningful collaboration. The group found that a face-to-face meeting, initially intended for developing a research product, was put to much better use as a way to begin to speak each other's language. In fact, this need for inter-person communication is sometimes crucial for establishing a common language among collaborators.

It is in this spirit that the International Medical Informatics Association's Working Group on Natural Language, Classification and Concept Representation (IMIA WG6) convened a meeting last year. We designed an agenda that covered a broad set of issues in terminological work, and brought together the leaders of the field to discuss them. Former WG6 chairman Jean-Raoul Scherrer, who recapitulated the previous work of WG6 and identified current problem areas, set the stage. The Natural Language Processing session explored the current state of the art in the conversion of narrative text into coded form. The Clinical Classifications session reviewed efforts to build sophisticated standard terminologies that embody many of the "desiderata" implied by emerging principles. The Cognitive Evaluations session explored many of the relevant aspects that arise when we try to represent human thought in coded form and then expect humans to reason using coded data. Finally, the Terminology Models session discussed a variety of approaches for representing terms and codes that appear to address the issues raised in the other three sessions.

The papers presented in this issue of *Methods* are each a contribution to the literature in their own right. However, when taken in aggregate, the collection offers powerful evidence that the "add more terms until we are done" approach to terminological work is no longer sufficient. For example, while it is convenient to treat terms, concepts and the objects they represent as interchangeable, several of the papers presented go to great lengths to differentiate between them, generally through some variant of the triangular arrangement originally proposed by Ogden and Richards. To the conference attendees, at least, such distinctions were not seen as hair-splitting, but rather fundamental to the principles we are developing and putting into practice. The chance to come together and discuss these issues was invaluable for all participants. For those who could not attend, the papers which remain as artifacts of the meeting convey much of the depth of thought and breadth of discussions that took place. After reading these papers, perhaps the "terminology problem" can be appreciated for being a vital organ, perhaps the heart, of medical informatics.

Address of the author:
James J. Cimino, M.D.,
Atchley Pavilion, Room 1310,
Columbia-Presbyterian Medical Center,
New York, NY 1032, USA
E-mail: James.Cimino@columbia.edu