ABBREVIATION IN MEDICAL TERMINOLOGY

O.Tsugan , *student ЛС-709,*
Terletska I.M., *EL adviser*

Text normalization, the process of identifying variants and bringing them to a common form, is an important aspect of successful information retrieval from medical documents such as health records, clinical notes, radiology reports and discharge summaries. In the medical domain a significant part of the general problem of text normalization is abbreviation. Numerous abbreviations are used throughout medical texts and identifying their meaning is critical to understanding the document.

A problem is presented by the fact that abbreviations are highly ambiguous with respect to their meaning. The Unified Medical Language System (UMLS) is a database containing biomedical information and tools developed at the National Library of Medicine. Using the UMLS as an example, "RA" can have such meanings as "rheumatoid arthritis", "renal artery", "right atrium", "refractory anemia", "right arm" and other terms. It has been estimated that about 33% of the abbreviations in the UMLS are ambiguous. In addition to problems associated with text interpretation, abbreviations constitute a major source of errors in a system that automatically generates lexicons for medical natural language processing (NLP).

When processing documents to identify those that contain a specific term, it would be desirable to identify all the documents that also use an abbreviation for the specific term. For example, if searching for documents containing the term "rheumatoid arthritis," it would be desirable to retrieve all those documents that use the abbreviation "RA" in the sense of "rheumatoid arthritis." At the same time, it is desirable not to identify documents that use the same abbreviation, but with a sense different from that of "rheumatoid arthritis." Continuing with the above example, it would be desirable that the search not identify those documents where "RA" means "right atrium."

One way to take context into account is to encode the type of discourse in which the abbreviation occurs. Discourse can, for example, be defined as the type of medical document and the medical specialty. As a more particular example, "RA" in a cardiology report can be normalized to "right atrium," while "RA" in a rheumatology note can be normalized to "rheumatoid arthritis." Unfortunately, this method of using the global context to resolve abbreviation ambiguity suffers from a number of drawbacks that limit its use in automatic document processing applications.