The Queries Expansion on Biomedical Terminologies: Toward a Comparison of Knowledge Representation Artifacts for Information Retrieval

Mauricio B Almeida* and Eduardo R Felipe

Federal University of Minas Gerais, Brazil

Submission: May 27, 2021; **Published:** June 02, 2021

*Corresponding author: Mauricio B Almeida, Federal University of Minas Gerais, Brazil

Abstract

We live in a period in which there is a remarkable availability of information. However, we still demand improvements in methods for information retrieval, including those applied to scientific papers. We focus here on artifacts of knowledge representation as a way to provide mechanisms for query expansion, as part of an ongoing project that aims to compare the effectiveness of biomedical terminologies regarding information retrieval, we present a research framework and the partial results obtained thus far.

Keywords: Biomedical ontology; Query expansion; Biomedical terminologies

Introduction

Since academic articles are valorized as instruments for the preservation and transmission of scientific knowledge, information retrieval (IR) is a crucial field. There is a growing interest in themes related to databases, information systems, knowledge bases, and integrated systems for purposes of IR in fruitful areas as Biomedicine [1].

The interest in this topic arose from the observation that artifacts of knowledge representation have large terminological and conceptual scope; namely, they are able to provide highly expressive conceptual connections through terms and relationships [2]. In considering this perspective, such artifacts are also becoming more important for IR, for example, because of the possibility of expanding queries. Expanding queries can be a method for improving information retrieval by finding scientific articles belonging to the same theme as that represented by the artifact [3].

The present work aims to investigate the recall of scientific articles through IR process using two instruments of knowledge representation in the biomedical field: namely, Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT) and Medical Subject Headings (MeSH). To achieve our goal, we have performed:

a)   the construction of a database for the persistence of scientific articles.

b)   the development of algorithms for performing expanded queries with hierarchical and axiomatic relationships of the database.

c)   the submission of queries to the database.

d)   the comparison of statistical results.

Methodology

The main process aims to be realized in stages that demand the following technical and technological aspects summarized in Figure 1. These steps include:

a)   Obtaining terminological artifacts, namely, Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT) and Medical Subject Headings (MeSH).

b)   Selecting scientific articles for analysis, in obstetrics and geriatrics, from the BioMed Central Ltd (BMC) collection.

c)   Extraction of data in PDF format for plain text.

d)   Adjustments in terminological artifacts, to facilitate the computational tasks.

e) Identification of biomedical terms using natural language processing algorithms.

f) Extraction of hierarchical terms as the first strategy in query expansion.

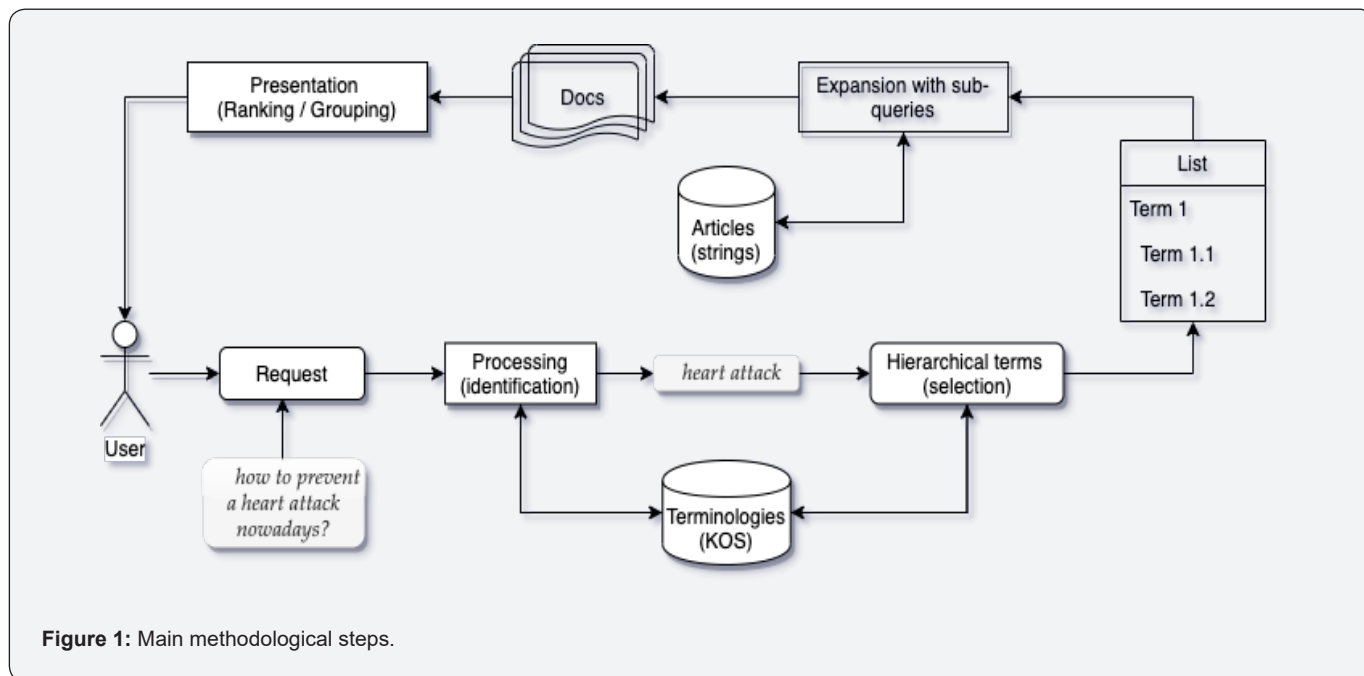g) Statistical analysis of the recall in each artifact.

Figure 1: Main methodological steps.

Results

The work so far has already demonstrated differentials in the experiment that justify the research. In the tests that were carried out, the expanded retrieval by sub-consultations returned a number of specific articles that were not present in a prominent ranking position of the original query, offering the user possibilities to access information that previously could not be found directly.

For example, using the term "heart attack" in the initial request recalled 10 articles. This term in the terminology points to "myocardial infarctions" as a preferred term, which in turn obtained 6 documents (with 1 overlapping). This term consists of the reference for access to the hierarchical structure, therefore obtaining a set of terms that allowed for the recall of a different set of documents. At this moment, the main terms and their revocation are listed (Table 1).

Table 1: Recalled articles starting with term, "heart attack".

| Term	Recall	Overlap	Differentiated articles
heart attack	10	0	10
myocardial infarctions	6	1	5
stemi	6	1	5
shock, cardiogenic	1	0	1
st segment myocardial infarction	10	4	6
myocardial infarct	7	1	6
cardiogenic shock	4	2	2

Final Remarks

We considered an overlap of documents already recovered by another term in order to avoid a duplicate response to the user. The results present possibilities for semantic recovery with the support of terminology, considering the retrieval of articles that would not otherwise be recovered by the original consultation.

IR is based on the ability to recognize words (terms) as a form of representation. Despite using syntax variations such as stemming and lemmatization techniques, among others, it is clear that the semantic connection made possible by terminologies and their structures, allowing for the same concept to be represented by different forms of syntactic expression has the capacity to enrich the IR process.

References

1. Aamodt A, Nygård M (1995) Different roles and mutual dependencies of data, information, and knowledge - An AI perspective on their integration. Data & Knowledge Engineering 16(3): 191-222.

2. Hoppe T, Humm B, Reibold A (2018) Semantic applications. New York, NY: Springer Berlin, Heidelberg.

3. Baeza YR, Ribeiro NB (2011) Modern information retrieval: the concepts and technology behind search (2nd ed: Addison Wesley), New York, USA.
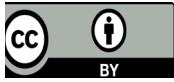
Your next submission with Juniper Publishers
will reach you the below assets

- Quality Editorial service
- Swift Peer Review
- Reprints availability
- E-prints Service
- Manuscript Podcast for convenient understanding
- Global attainment for your research
- Manuscript accessibility in different formats
  (Pdf, E-pub, Full Text, Audio)
- Unceasing customer service

Track the below URL for one-step submission
https://juniperpublishers.com/online-submission.php