Research on Technology Development Trend Based on SAO Network

——Taking AIDS treatment technology as an example

INTRODUCTION

The advent of the era of big data has made people pay more and more attention to information and data. How to obtain useful information from massive data becomes the key. The biomedical literature stores a large amount of medical information, and mining the literature data can reveal the trend law implicit in the literature information, which helps researchers grasp the overall development status of the disease and help researchers make reasonable scientific and technological decisions. With the rapid growth of biomedical literature, Knowledge Discovery in Text (KDT) has become an effective means of discovering and acquiring information in the literature. Text knowledge discovery extracts previously unknown, understandable, and ultimately usable knowledge from large amounts of textual data, and uses this knowledge to better organize information for reference. Information extraction is one of the core technologies of KDT. Its purpose is to automatically extract information such as entities, entity attributes and semantic relations between entities as the basic knowledge unit of knowledge discovery.

The traditional text mining methods use key words/phrases as the basic semantic unit of the analysis. This method is not conducive to identifying the “subliminal semantics” and “the relationship between the themes”. With the increase in the amount of medical data, this problem has become more prominent. The SAO triad is a kind of knowledge representation in the form of "subject-predicate-object" to represent the knowledge units and their semantic relations in the literature. It has the advantages of rich semantic representation, simple structure and mature technology. Through the text mining operations such as clustering, classification, reconstruction and dimensional reduction in the subject-predicate-object in the SAO triad, combined with visual analysis tools, the domain knowledge topics and important concepts can be revealed quickly, clearly and intuitively. And their relationship.

This paper intends to use AIDS treatment technology as an example. Based on the "AIDS/HIV" keyword in the PubMed database for retrieval and identification technology, the SAO semantic mining method is used to identify the "key problem-solution" pair and obtain the disease. Treatment technology. Then, based on the SAO semantic structure and keyword co-occurrence, the evolutionary relationship between disease treatment techniques is identified, and the analysis and evolution path identification of disease treatment techniques are realized. On this basis, the SAO network is constructed, and the relationship strength between nodes is calculated. The SAO network analysis is carried out. The development trend of emerging technology is analyzed with five indicators: in&out degree, key action , node "degree distribution" evolution and network center deviation.

FRAMEWORK

1. SAO structure recognition

This paper selects the PubMed literature database as a literature search library. Document search and data downloading are carried out according to the established search strategy. Then we use the Semrep tool to extract the SAO triple structure. The workflow of Semrep is shown in Figure 1.

2. SAO network construction

A list of S-A-O node pairs is constructed based on the obtained SAO structure to represent nodes and edges in the SAO network. The node pair list has four fields: Subject Node, Object Node, Action Edge, and

Number of Relation. The network is directed. The calculation formula of the relationship strength between the Subject node and the Object node is as follows:

$$RS_{S\to O} = \sum_{A\in Actions\ of\ S\to O} RS_{S-A-O}$$

3. Technology development trend analysis

In order to analyze the development trend of technology, this paper designs a dynamic development analysis method for SAO networks. Network dynamic development analysis is intended to assess and identify changes in network structure, predicting technological developments. The method is based on the actor network theory, including: SAO network in-out degree analysis, key Action analysis, structural competitiveness analysis and SAO network dynamic development analysis.

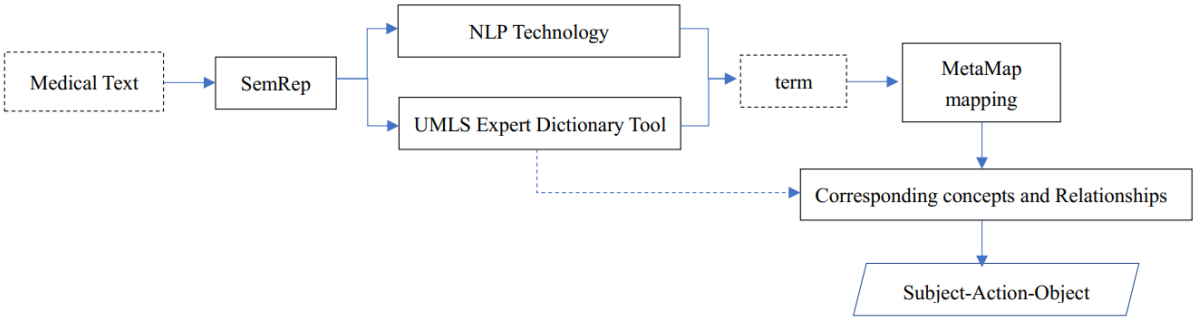

Figure 1 Semrep extracts SAO structure flow

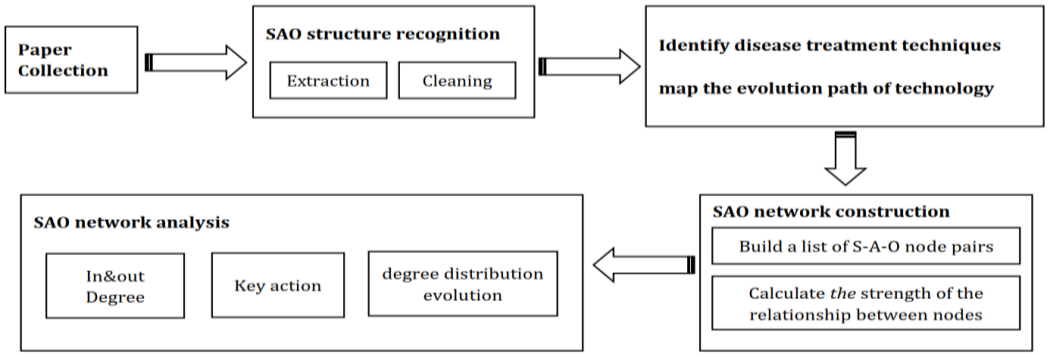

Figure 2 Framework

REFERENCE

[1] Petersen A M, Rotolo D, Leydesdorff L. A triple helix model of medical innovation: Supply, demand, and technological capabilities in terms of Medical Subject Headings[J]. Research Policy, 2016, 45(3):666-681.

[2] NCBI.UMLS® Reference Manual[EB/OL].[2017-0531].http://www.ncbi.nlm.nih.gov/books/NBK9676/.

[3] Zhang Y , Chen H , Lu J , et al. Detecting and predicting the topic change of Knowledge-based Systems: A topic-based bibliometric analysis from 1991 to 2016[J]. Knowledge-Based Systems, 2017:S0950705117303271.

[4] ARNOLD P,R AHM E.Semrep:a repository for semantic mapping[EB/OL].[2017-09-08].https://dbs.uni-leipzig.de/en/publication/title/semrep_a_repository_for_semantic_mapping.

[5] NCBI.Introduction to MeSH[EB/OL].[2017-05-31].https://www.ncbi.nlm.nih.gov/mesh