The contribution of syntactic-semantic approach to the search for complementary literatures for scientific or technical discovery.

José M. Vicente-Gomila

Universitat Politècnica de Valencia

Introduction

Using Tech Mining and its power to elucidate information in new ways, trigger new ideas (1) and resolve problems (2), stimulate scientific discovery (3) and assess scientific discovery (4) (5) is not new, however it is a very thorough and time consuming task which involves reading hundreds of documents to identify linkages from a first concept A to a second B and from the second B to a third concept C. By means of the usage of syntactic and semantic software, the author tries to compare results from known cases done with Tech Mining to the results of a combination of syntactics-semantics (S-S) and Tech Mining (TM). With S-S the author tries to identify B concepts linked to A and C respectively (complementarity) by means of the semantic relationship and then, in parallel, to look for non-interaction either by citation as in the known literature and from excluding presence of A concepts in C and from C concepts in A by analyzing the full texts.

TRIZ based syntactic-semantic indexing takes the TRIZ methodology concept of a system which delivers a function to an object, and applies it to substantives and its actions with their related direct complements and other grammatical structures. Such linking relationships can make logic sense and may link disparate literatures in the sense of knowledge discovery in the published related literature. The present paper tries to show the advantages of syntactic linking to relate initially disparate literatures and so easing the potential discovery of new knowledge. Moreover by finding links with human like logic sense, the author tries to show that less expert judgment is needed to ascertain new knowledge in comparison with term based literature discovery.
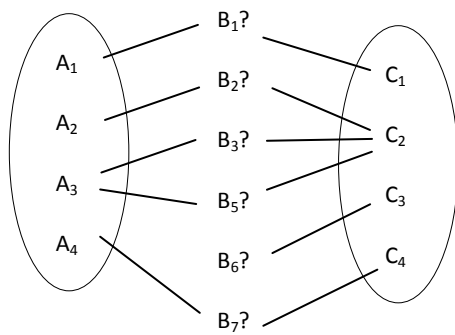
Classic knowledge discovery
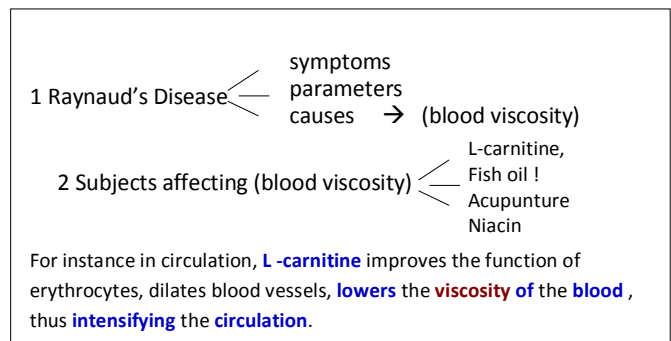
S-S knowledge discovery

In the classic approach to knowledge discovery B terms are identified either by frequency (6), expert judgment (7) or by text similarity (8). In the S-S approach, starting from one problem or one disease, one can find causes or subjects acting on the disease (Raynaud's as in the classic)

 Data and methods

To demonstrate the advantages of using S-S TRIZ , the author has mined 499 full text papers from Elsevier Science Direct – query terms: 'Meniere disease' in title + abstract + keywords last date 06-08-2012.

Once several causes are extracted for the term -system- 'meniere disease', (Eustachian tube blockage, imbalance of fluid in the inner ear, etc.) the next step is to find candidate systems which interact or effect the - object 'causes' - . By seeing the actions one can easily select those with logic meaning, being more probable to have meaning also to an expert. The logic chains obtained by S-S TRIZ, avoid many relationships of terms without significance and brings

Also to demonstrate the possibilities in the industrial domain, the author used the already mined database of US patents, European patents, PCT patents and Japan patents, selecting only applications and not granted to avoid duplications and grouped by patent families –with same priority date- from 1971 to present, to increase accuracy of count. With this, the author will use the approach to other technical areas as energy storage systems as a way to discover possible inventions.

(1) Porter A. , Cunningham, S. Tech Mining; Wiley Interscience (2005)
(2) Kostoff , R. Literature –related discovery and innovation – Update (2011)
(3) Swanson D, Smalheiser, N ; An interactive system for finding complementary literatures: a stimulus to scientific discovery; Artificial intelligence 91 (1997)
(4) Smalheiser, N, Swanson D; Using Arrowsmith: a computer assisted approach to formulating and assessing scientific hypotheses
(5) Swanson D; Running esophageal acid reflux, and atrial fibrillation: a chain of events linked by evidence from separate medical literatures; Medical Hypotheses (2008) 71
(6) Swanson D., Smalheiser, N.; An interactive system for finding complementary literatures: a stimulus to scientific discovery; Artificial intelligence 91 (1997)
(7) Kostoff, R. ; Systematic acceleration of radical discovery and innovation in science and technology; Technological forecasting & Socail change 73 (2006)
(8) Gordon, M., Dumais, S.; Using latent semantic indexing for literature based discovery; Journal of the American society for information science 49 (1998)