Knowledge Mapping of Iranian Nano Science and Technology: Text mining Approach

Ehsan Mohammadi

Master of library and Information Science
Email: eh3anmohammadi@gmail.com

Abstract

Conceptual structure of Nano Science and technology is vague for research because of Sky rocketing growth of publications and interdisciplinary nature of the field. This research done with the aim of mapping subfields of Iranian NST. At the first NST publications affiliated by Iranian universities or organizations retrieved from WOS based on Search Algorithm of Georgia Institute of Technology from 1974 to 2007.Then by text mining method terms of title, abstract and keywords extracted and by application of data mining algorithms essential terms of Iranian NST distinguished. Finally based on terms co-occurrence Iranian NST subfields clustered.

Introduction

For developing future intellectual pursuits researchers should Stand on the shoulders of giants. In other words, Researchers brighten the future of their specific fields based on the knowledge their predecessors have generated. A conceptual perception of science structure of any field could be a helpful bridge for getting insight into past and present status of a field to researchers. Some relations and phenomena in science structures are conceptual and understandable for minds; while they are invisible for eyes. Information scientists try to discover these invisible phenomena and visualize them in multidimensional scale through science maps result in better understanding

Why this research?

✓NST is a new appealing research area. Researchers from a range of different fields contribute in development of NST hence NST is considered as an interdisciplinary area.
✓interdisciplinary nature of Nano, ambiguity of the future of nano because of its innovative and dynamic researches content, using nano in scientific community and industry as a luxury word to appealing more.
✓This dissertation aims to apply text mining methods to map the structure of Iranian (NST). In this research text mining, as the method of text processing with natural language processing (NLP) was used in order to discover conceptual relations of Nano science and technology in Iran

Methodology

1. First step was concerned with searching NST publications affiliated by Iranian universities or organizations in WOS based on Search Algorithm of Georgia Institute of Technology from 1974 to 2007.
2. abstracts of all retrieved articles imported into text mining software with the aim of extracting NST terms A program with SQL server was designed for this research. In this phase stop words were excluded.
3. In the third phase we used weighting algorithm for distinguishing appropriate terms and the terms that weighted more than 0.03 selected as related terms (772 terms).
4. In the fourth step, in order to optimize the terms, we created a database from all terms of the existing nano search algorithms and refined our selection. Finally, 96 terms were selected as the appropriate terms.
5. The Fifth step was measuring and calculating co-occurrence of 96 terms together in all publications based on query searching and a square matrix (96*96) was the result of the co-occurrence of the terms.
6. In the sixth step, raw matrix data was normalized. We normalized this 96*96 co-occurrence matrix using the Salton-index (see, for instance, Peters & Van Raan 1993a, 1993b). This normalized matrix was used as the input for a cluster analysis in SPSS.
7. The final step was visualizing the data by hieratical clustering and Multi Dimensional Scaling (MDS).

In this research for allocating the terms to sub-subjects of Iranian NST to fifteen clusters, we created a database of NST terms from previous related researches. Some terms did not categorized in none of exists NST sub-subjects, therefore based on Mogoutov & Kahane research (Mogoutov & Kahane, 2007) we allocated those in a cluster named "new research area".
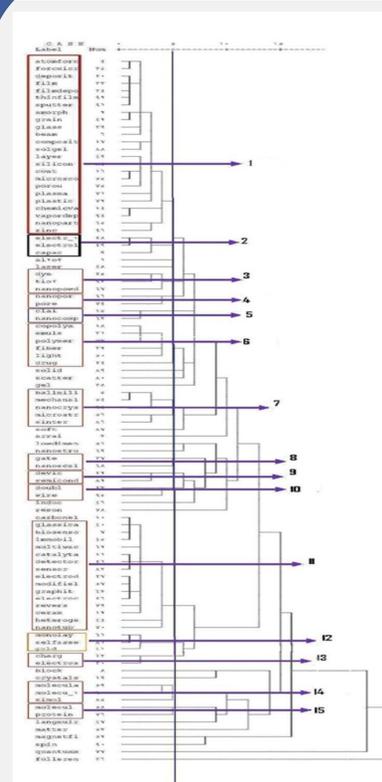
Results and conclusions

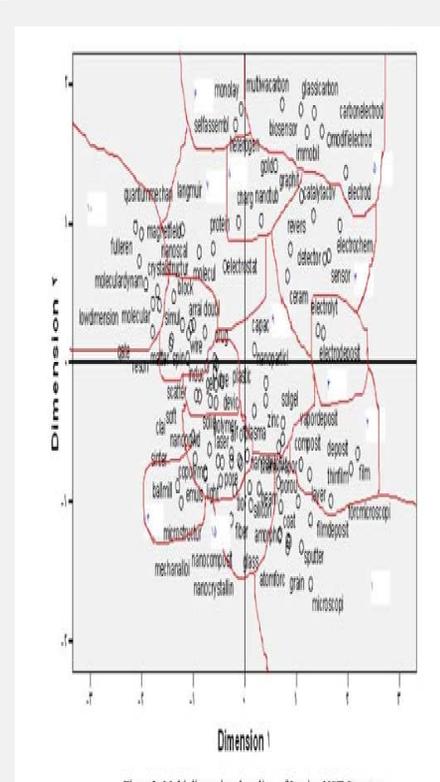Figure1: Hierarchical Clustering of Iranian NST Structure

Figure2: Multi dimensional scaling of Iranian NST Structure

Co-occurrence analysis of the extracted terms showed the terms in relation together in fifteen clusters and some terms that had fewer co-occurrences with the others were not placed in any cluster and left independent and separately. The cluster grouped in three large, medium and small sizes. Cluster number one with 24 terms and Cluster number eleven with 16 terms grouped as the large clusters. Cluster number six with 6 terms and seven with 5 terms classified as medium. Other clusters with converge of two or three terms considered as small clusters.
Cluster number one consisted of physics science terms and Cluster number two included analytical chemistry terms. Cluster number three combined chemistry physics area. Cluster number four encompassed two new terms that categorized as new areas of NST. Cluster number five contained material science and engineering terms. Cluster number six composed of polymer terms. Cluster number seven covered material science and engineering terms. Cluster number eight included two terms that could be applied in all subfields of NST. Cluster number nine and ten consisted of four terms related to physics science Cluster number eleven contained analytical chemistry terms whiles some terms of chemistry physics were seen in this cluster. Clusters number twelve and thirteen terms are categorized as new research areas of NST. Clusters number fourteen consisted of three terms; one of them related to physics and the rests can be applied to all areas of NST subcategories Clusters number fifteen covered two terms related to biochemistry.

Future Work

✓Knowledge Mapping Iranian NST structure by co-citation analysis method.
✓Comparison of Iranian NST structure based on text mining with co-citation analysis approach.
✓Comparison of Iranian NST structure with Global NST structure.

References

1. Mogoutov, A., & Kahane, B. (2007). Data search strategy for science and technology emergence: A scalable and evolutionary query for nanotechnology tracking. Research Policy, 36(6): 893-903.
2. Porter,A.L.,J.Youtie,et al.(2008).Refining search terms for nanotechnology. Journal of Nanoparticle Research. 10(5): 715-728.
3. Zitt, M. Bassecoulard, E (2006) Delineating complex scientific fields by an hybrid lexical-citation method: An application to nanosciences. Information Processing and Management 42: 1513–1531.