Semantic-Based Technology Trend Analysis

Chao Yang[1,2], Donghua Zhu[1]

[1.] School of Management and Economics,
Beijing Institute of Technology,
Beijing 100081, China
chao.yang@uts.edu.au, zhudh111@bit.edu.cn

Guangquan Zhang[2]

[2.] Centre for Quantum Computation and Intelligent Systems,
University of Technology Sydney,
NSW 2007, Australia
guangquan.zhang@uts.edu.au

Abstract—**Technology trend analysis offers a flexible instrument to understand both opportunity and competition for emerging technologies. Semantic information is used in Science, Technology & Innovation (ST&I) records which makes the technology trend analysis more challenging. This paper proposes a semantic-based approach for technology trend analysis through emphasizing Subject-Action-Object (SAO) structure; It also applies the trend analysis approach to extract technology information and identify and predict the trend of technology development more effectively. An empirical study on Graphene is completed to demonstrate the proposed trend analysis approach.**

Keywords: Semantic Analysis, technology trend analysis, Technology roadmapping, Text Mining

I.   INTRODUCTION

Today's worldwide economy highly depends on technological innovation. Not only engineers but also scientists, public organizations, and society focus on the newly emerging science & technology (NEST) [1]. However, it's not easy to catch the pathway of technology development, since checking the novelty of a technical solution requires big efforts even for skilled researchers [2].

Keyword-based extracting/searching and analysis approach has enjoyed success in assisting information retrieval and data mining [3]. However, current keyword-based searching approach ignores the role that verb plays in the analysis of technology documents [4], and generally tends to overstate the qualitative rather than incorporating quantitative and objective information [5]. To solve these problems vast amounts of semantic information are needed for deeper and more meaningful understanding of technology invention information [3, 6]. In this paper, we propose a semantic-based technology trend analysis approach, and apply it to distinguish technology developments. The semantic information is extracted from ST&I records and is expressed in SAO structure. The main contributions of this paper are: 1) we construct a comprehensive system for extracting and defining SAO structure of Science, Technology & Innovation (ST&I) records; 2) we introduce SAO analysis into technology trend analysis and technology roadmapping (TRM).

The rest of this paper is organized as follows: In Section 2, we summarize key contributions from literature on SAO related semantic analysis and technology roadmapping. Section 3 elaborates on our methodology for using Semantic-based approach to extract technology information and identify the development trend. In Section 4, we present a case study of patents related to Graphene technology to demonstrate the effectiveness of the SAO-based semantic analysis. Finally, we draw conclusions in Section 5.

II.   LITERATURE REVIEW

A. SAO related semantic techniques

SAO is a semantic structure extracted from text corpus and has been studied by many researchers. Bergmann, et al. [7] fulfilled the extraction of semantic structures via a natural language processing implemented in the software Knowledgist[TM] 2.5. Choi, et al. [4] presented a method that formulated an SAO network, and applied actor network theory to analyze technology implications. Moehrle, et al. [8] transferred patent text into a SAO and created similarity matrices to produce inventor competence map. Verbitsky [9] defined Semantic TRIZ and extracted from a sentence its semantic items such as subject, action, and object. There are also some semantic techniques which are similar with SAO, including: Resource Description Framework (RDF), Semantic Relation and Problem & Solution (P&S) Patterns.

The structure of a RDF is a triple, each consisting of a subject, a predicate and an object. The assertion of a RDF triple says that some relationship, indicated by the predicate, holds between the resources denoted by the subject and object. The predicate denotes a property, that is, a resource that can be thought of as a binary relation [10]. RDF imposes structure that provides the ability for resource description communities to define semantics [11].

Hearst [12] identified a set of lexico-syntactic patterns and extracted some relations between nouns like: A such as B. Shinmori, et al. [13] proposed a framework to represent the structure of Japanese patent claims, and found three models: does, and, be characterized by. Soo, et al. [14] defined eight types of regular expression to extract the semantic information which is expressed as a triple relation structure, like part-of relations. Liu and Singh [3] extended WordNet's repertoire of semantic relations from the triplet of synonym to twenty semantic relations.

Kim, et al. [15] argued that by extracting a problem and an associated solution in a domain, they can identify a manifestation of technology. Such analysis is derived from the problem-solution pattern between technological concepts. We can map the "subject/object" to the "problem", while transferring the whole SAO model to the "solution" with its "action" or "function" [16].

B. Technology roadmapping

The pathways for commercialization of new technologies are typically complex and fraught with challenges [17]. TRM provides a structured and graphical means for exploring and communicating the relationships between evolving and developing markets, products and technologies over time [18]. However, TRM generally tends to overstate the qualitative and expert-dependent knowledge [5]. In order to improve TRM applicability, keyword-based product-technology maps are proposed from which objective and quantitative information can be derived [5]. Even combined with keyword-based quantitative information, TRM still requires a lot of expert knowledge to identify the NEST and describe the development of technology, which leads to subjectivity. Zhang, et al. [19] used P&S patterns with semantic TRIZ methods to take the place of the topical factors. Chen, et al. [20] proposed a time series processing module with piecewise linear representation to identify trend signal of technologies.

III. SEMANTIC-BASED TECHNOLOGY TREND ANALYSIS METHODOLOGY

We attempt to analyse technology trend through a SAO semantic structure. SAO can not only express the meaning of "what technology", it also has the potential to significantly describe "what to do", "how to do" and show the relationship between technologies. We construct a framework for Semantic-based technology trend Analysis (shown as Fig. 1).

A. Part 1: Semantic information extraction

We extract semantic information (subject-action-object) from ST&I records. The SAO structure referred in this paper is not limited in the form of Subject-Action-Object. Like SAO structure, SA and AO have the equivalent potential in the expression of technology. The stepwise procedure of Semantic information extraction is shown as Table 1.

Step 1. Defining the principle of Semantic information extraction

Different from the conventional named entity recognition in Natural Language Processing, the Semantic information extraction emphasizes the guideline of extraction destination on what and how to extract SAO-based technology information. There are two principles: 1) The Semantic information should has a direct relationship with the specific technology field. Not all of SAO are necessary for our patent analysis. 2) The SAO-document co-occurrence matrix cannot be too sparse. Because that SAO is more complex than term, there will be many literally different SAOs which have the same meaning.

Step 2. Pre-processing the corpus

Before the extraction of Semantic information, we need to preprocess raw ST&I records. The pretreatment includes a set of natural language processing such as tokenization, sentence splitting, morphological analysis and POS tagging, which lay the foundation for further semantic analysis.

Step 3. Annotating core terms/phrases

The purpose of this step is to obtain the core terms/phrases of specific technology field, and then locate SAO-based semantic information. Core terms/phrases compose the Subject and Object of SAO. We introduce term clumping to implement this function. Term clumping is a series of steps to identify and consolidate topical terms in a collection of ST&I records [21].

Step 4. Extracting SAO structure

Although ST&I records (e.g. patent) has complex grammatical structures and is extremely difficult to understand, there are some common patterns in text writing. For example, the expression of patent records written by patent analyst and applicant is unique and fixed. Based on the in-depth analysis of ST&I records, we design a set of algorithms and rules for SAO Extraction (shown as Table 2).
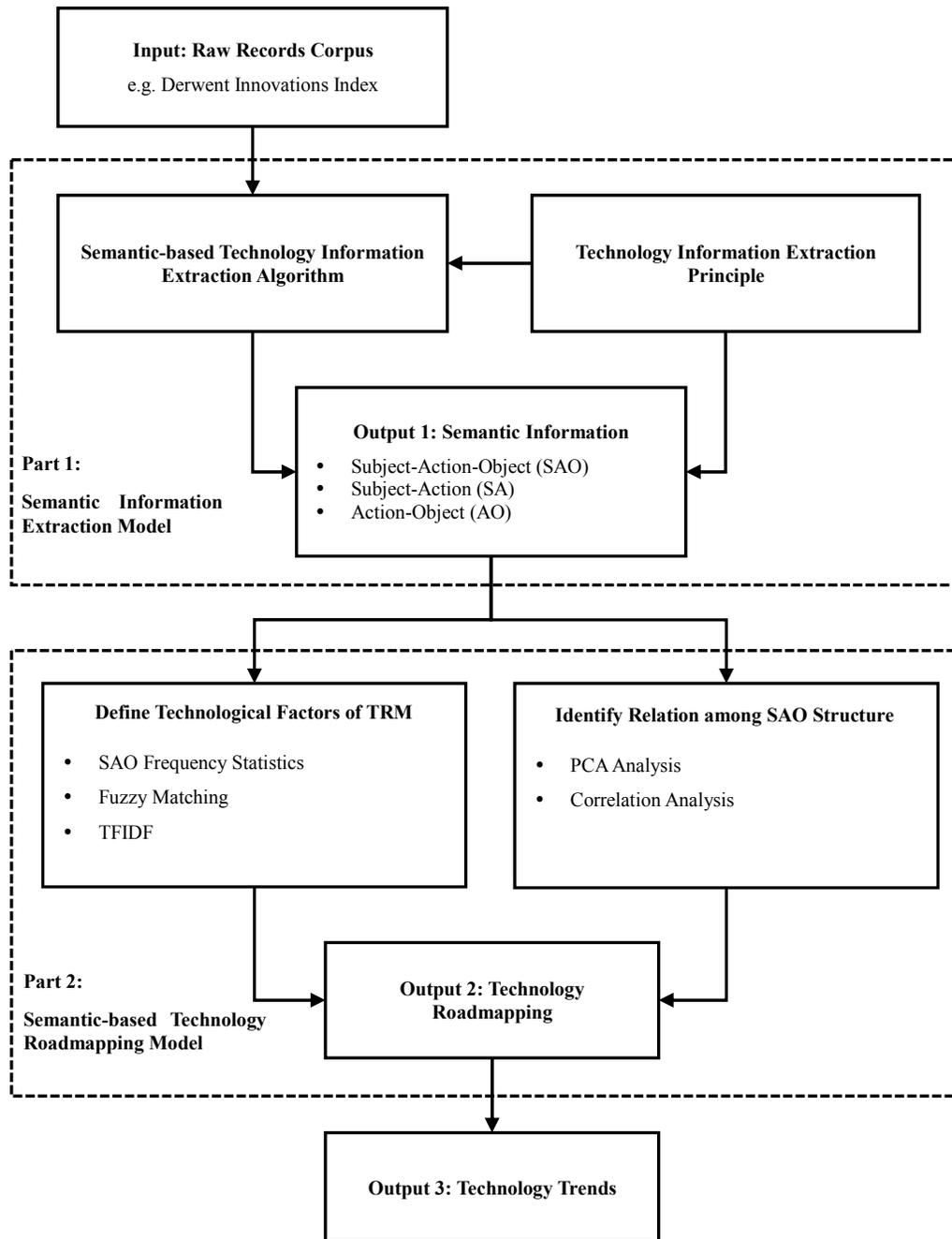
Figure 1.   Framework of Semantic-based technology trend Analysis

Table 1.   Stepwise procedure of Semantic information extraction

Procedure	Explanation
Patents with which we begin	Derwent Innovations Index Patent Records
Step 1. Defining the principle of Semantic information extraction	Specific principle and destination for semantic information extraction in technology documents
Step 2. Preprocessing the corpus	Tokenization, POS Tagging
Step 3. Annotating core terms/phrases	Identify core terms/phrases based on the term clumping
Step 4. Extracting SAO structure	Extract SAO-based semantic information based on the extraction algorithms and core terms

Table 2.   SAO structure extraction rules

Number	Extraction Rule	Example
1	Extracting the SAO of Various Tense	A do B
2	Extracting the SAO of Passive Voice	A be done by B
3	Extracting the SAO of non-finite verb form	A to do B; A doing B; A done by B

B. Part 2: Semantic-based technology roadmapping

There are two steps to implement Semantic-based technology roadmapping:

1) Define technological factors of TRM

Technological Factor is a set of items posed in the map and expressed with SAO structure in this paper. We use a range of methods to get technological factors: 1) We use the fuzzy matching function to combine terms with similar information based on stemming; 2) Based on the SAO frequency statistics and TFIDF (Term Frequency-Inverse Document Frequency), we can identify key and emerging technology. Then we get the initial technological factors.

2) Identify relationship of SAO structure

There are two primary relationships between SAO-based technology information in TRM: temporal relation and correlative relation. Temporal relation indicates the order in which the specific technology occurs and shows a trend of technology development. We use cross-correlation map to show temporal relation among SAO based on application year. Correlative relation can show the association among SAO structures. We generate a factor map via the PCA analysis of software VantagePoint based on "SAO co-occurrence". Through PCA, we can identify the correlative relation and further identify the technological factors.

Based on the identification of technological factors and relations among SAO structure, we can finally get technology roadmapping with the help of literature review and expert and identify the trend of technology development.

IV. A CASE STUDY

In the case study, we chose Derwent Innovations Index (DII) as our patent data source. The search strategy is a simple one: all patents from 1963 to 2014 (November) whose title contained the keyword "graphene" were downloaded. The search strategy resulted in a total of 7413 patent family records spanning 30 countries, 1803 institutional affiliations, 7299 inventors.

Based on semantic information extraction model and experts' aid, we got 934 SAO structures from 7413 DII patents. We consolidate SAO structure with fuzzy matching and screen statistical results by TFIDF. At last the samples of SAO-based initial technological factors in TRM are shown in Table 3.

Table 3.   Sample of technological factors in TRM

Top 1-5	Top 101-105
prepare graphene	forming dielectric layer
form graphene	grow graphene layer
obtain graphene oxide	introducing hydrogen
dispersing graphene oxide	laminating transfer sheet
form graphene layer	manufacture graphene sheet
Top 201-205	**Top 251-255**
improves stability	obtain carboxylic graphene oxide
has high yield	obtain doped graphene
method enables economical manufacture	Monolayer Graphene Using Electrostatic Substrate Engineering
obtain lithium	obtain fiber graphene
obtain multilayer graphene	obtain fluorinated graphene

We further identify the relation among SAO by principal component analysis (shown as Fig. 2). Every note in this figure includes several SAO structures, just like the note "doping nitrogen" shows. Depending on the technological factors and relations among these factors we can build the TRM for Graphene technology (shown as Fig. 3).

Based on the Fig. 2 and Fig. 3, we summarized four trends: 1) Trend 1 is "boron-doped graphene", 2009-2014. 2) Trend 2 is "prepare three-dimensional porous graphene", 2009-2014. 3) Trend 3 is "(Commercially) preparing high-quality graphene", 2009-2014. 4) Trend 4 is "obtain graphene aerogel", 2010-2013.

V. CONCLUSIONS AND FURTHER STUDY

This paper proposed a semantic-based technology trend analysis approach and constructed a comprehensive system for extracting and defining SAO structure of Science, Technology & Innovation records. Compared with keyword-based method, semantic-based analysis emphasizes "SAO structure" and can express a deeper, more specific and more detailed technology information, and is very conducive to the subsequent analysis and interpretation.

Further studies to extend our approach can take several directions. We will continue to try different ways to combine similar SAO without affecting original semantic information. We also attempt to implement technology roadmapping automatically and reduce the influence of artificial factor.
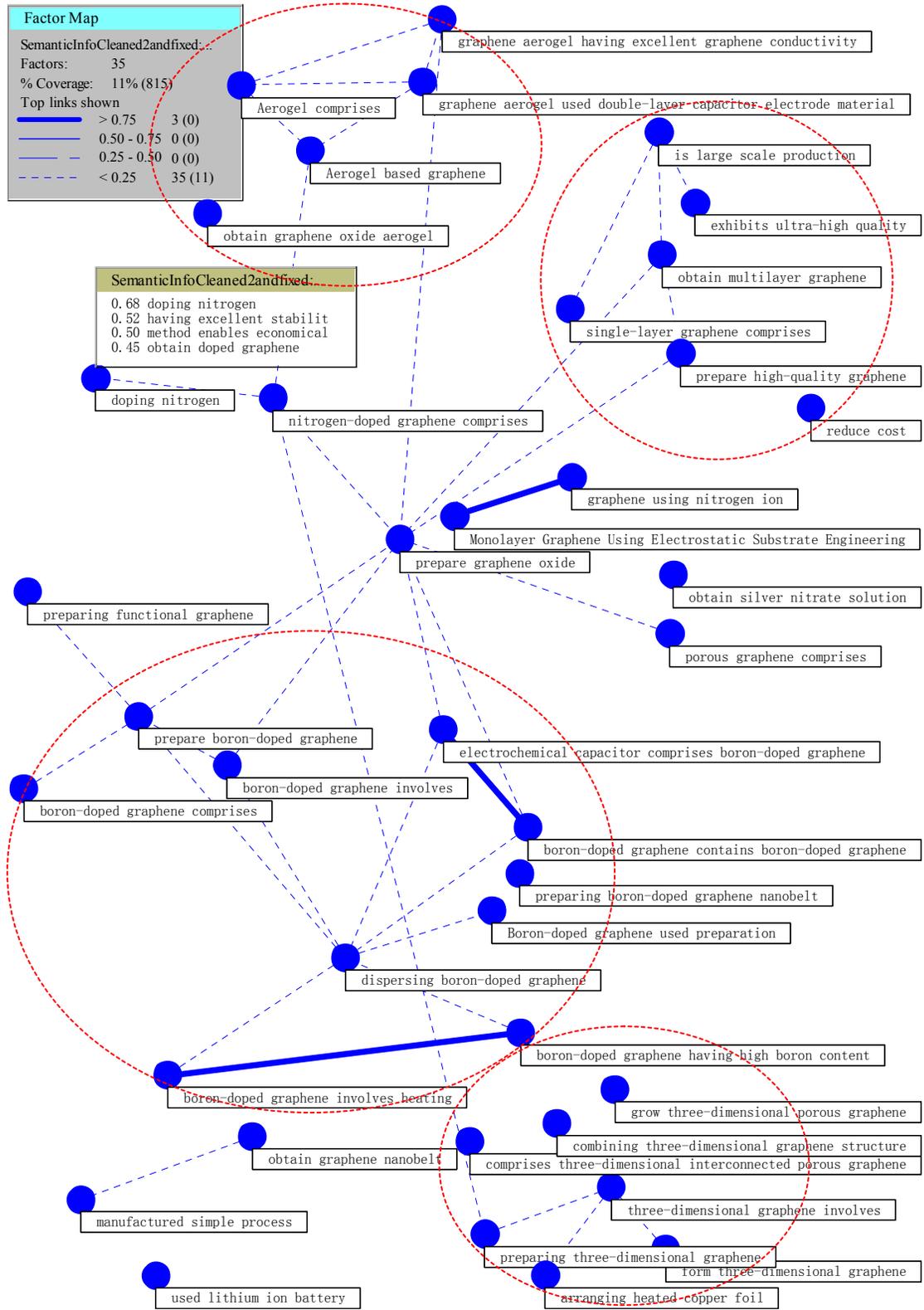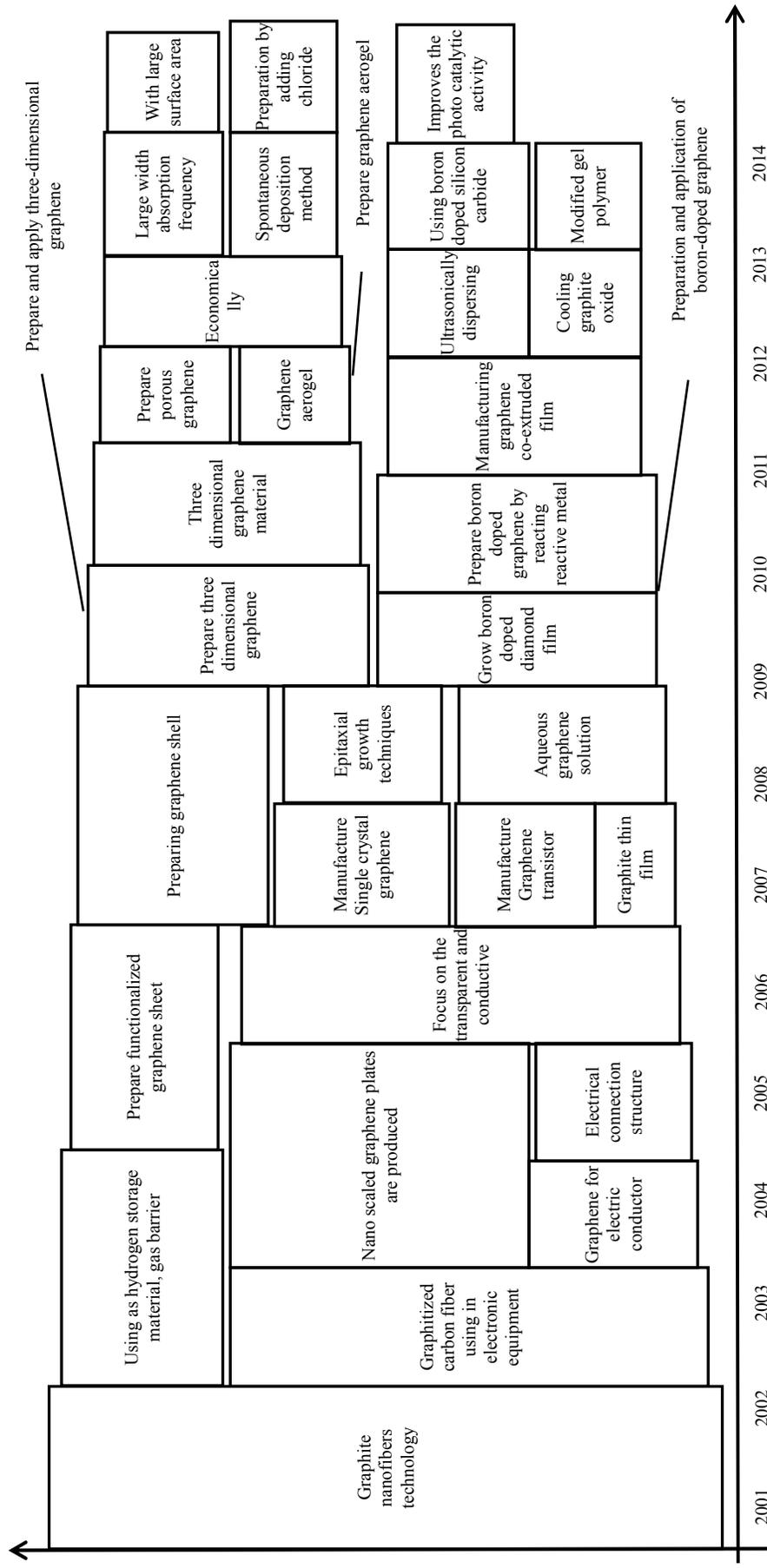
Figure 2. Factor mapping of SAO

Factor Map

SemanticInfoCleaned2andfixed:
Factors: 35
% Coverage: 11% (815)
Top links shown
> 0.75   3 (0)
0.50 - 0.75   0 (0)
0.25 - 0.50   0 (0)
< 0.25   35 (11)

graphene aerogel having excellent graphene conductivity

Aerogel comprises

graphene aerogel used double-layer capacitor electrode material

Aerogel based graphene

is large scale production

exhibits ultra-high quality

obtain multilayer graphene

obtain graphene oxide aerogel

single-layer graphene comprises

prepare high-quality graphene

SemanticInfoCleaned2andfixed:

0.68 doping nitrogen
0.52 having excellent stabilit
0.50 method enables economical
0.45 obtain doped graphene

doping nitrogen

nitrogen-doped graphene comprises

reduce cost

graphene using nitrogen ion

Monolayer Graphene Using Electrostatic Substrate Engineering

prepare graphene oxide

preparing functional graphene

obtain silver nitrate solution

porous graphene comprises

prepare boron-doped graphene

electrochemical capacitor comprises boron-doped graphene

boron-doped graphene involves

boron-doped graphene comprises

boron-doped graphene contains boron-doped graphene

preparing boron-doped graphene nanobelt

Boron-doped graphene used preparation

dispersing boron-doped graphene

boron-doped graphene having high boron content

grow three-dimensional porous graphene

boron-doped graphene involves heating

combining three-dimensional graphene structure

comprises three-dimensional interconnected porous graphene

obtain graphene nanobelt

three-dimensional graphene involves

manufactured simple process

preparing three-dimensional graphene

form three-dimensional graphene

used lithium ion battery

arranging heated copper foil

Figure 3. Technology roadmapping for graphene technology

REFERENCES

[1] A. L. Porter and S. W. Cunningham, *Tech mining: Exploiting new technologies for competitive advantage* vol. 29. New York: John Wiley & Sons, 2004.

[2] G. Cascini, A. Fantechi, and E. Spinicci, "Natural language processing of patents and technical documentation," in *Document Analysis Systems VI.* vol. 3163, S. Marinai and A. Dengel, Eds., ed Berlin: Springer Berlin Heidelberg, 2004, pp. 508-520.

[3] H. Liu and P. Singh, "ConceptNet - a practical commonsense reasoning tool-kit," *Bt Technology Journal,* vol. 22, pp. 211-226, Oct 2004.

[4] S. Choi, J. Yoon, K. Kim, J. Y. Lee, and C. H. Kim, "SAO network analysis of patents for technology trends identification: a case study of polymer electrolyte membrane technology in proton exchange membrane fuel cells," *Scientometrics,* vol. 88, pp. 863-883, Sep 2011.

[5] S. Lee, S. Lee, H. Seol, and Y. Park, "Using patent information for designing new product and technology: keyword based technology roadmapping," *R&D Management,* vol. 38, pp. 169-188, 2008.

[6] M. Al-hassan, H. Lu, and J. Lu, "A semantic enhanced hybrid recommendation approach: A case study of e-Government tourism service recommendation system," *Decision Support Systems,* vol. 72, pp. 97-109, / 2015.

[7] I. Bergmann, D. Butzke, L. Walter, J. P. Fuerste, M. G. Moehrle, and V. A. Erdmann, "Evaluating the risk of patent infringement by means of semantic patent analysis: the case of DNA chips," *R&D Management,* vol. 38, pp. 550-562, 2008.

[8] M. G. Moehrle, L. Walter, A. Geritz, and S. Muller, "Patent-based inventor profiles as a basis for human resource decisions in research and development," *R & D Management,* vol. 35, pp. 513-524, Nov 2005.

[9] M. Verbitsky. (2004, 5 January). *Semantic TRIZ.* Available: http://www.triz-journal.com/archives/2004/

[10] R. Cyganiak, D. Wood, and M. Lanthaler, "RDF 1.1 concepts and abstract syntax," 2014.

[11] E. Miller, "An introduction to the resource description framework," *Bulletin of the American Society for Information Science and Technology,* vol. 25, pp. 15-19, 1998.

[12] M. A. Hearst, "Automatic acquisition of hyponyms from large text corpora," presented at the 14th Conference on Computational Linguistics - Volume 2, Nantes, France, 1992.

[13] A. Shinmori, M. Okumura, Y. Marukawa, and M. Iwayama, "Patent claim processing for readability: structure analysis and term explanation," presented at the ACL-2003 Workshop on Patent Corpus Processing - Volume 20, Sapporo, Japan, 2003.

[14] V.-W. Soo, S.-Y. Lin, S.-Y. Yang, S.-N. Lin, and S.-L. Cheng, "A cooperative multi-agent platform for invention based on patent document analysis and ontology," *Expert Systems with Applications,* vol. 31, pp. 766-775, 11// 2006.

[15] Y. Kim, Y. Tian, Y. Jeong, R. Jihee, and S.-H. Myaeng, "Automatic discovery of technology trends from patent text," presented at the 2009 ACM Symposium on Applied Computing, Honolulu, Hawaii, 2009.

[16] Y. Zhang, X. Zhou, A. L. Porter, and J. M. Vicente Gomila, "How to combine term clumping and technology roadmapping for newly emerging science & technology competitive intelligence: "problem & solution" pattern based semantic TRIZ tool and case study," *Scientometrics,* vol. 101, pp. 1375-1389, 2014.

[17] S. K. Arora, J. Youtie, P. Shapira, L. Gao, and T. Ma, "Entry strategies in an emerging technology: a pilot web-based study of graphene firms," *Scientometrics,* vol. 95, pp. 1189-1207, 2013.

[18] R. Phaal, C. J. P. Farrukh, and D. R. Probert, "Technology roadmapping - A planning framework for evolution and revolution," *Technological Forecasting and Social Change,* vol. 71, pp. 5-26, Jan-Feb 2004.

[19] Y. Zhang, X. Zhou, A. L. Porter, J. M. V. Gomila, and A. Yan, "Triple Helix innovation in China's dye-sensitized solar cell industry: hybrid methods with semantic TRIZ and technology roadmapping," *Scientometrics,* vol. 99, pp. 55-75, 2013.

[20] H. Chen, G. Zhang, D. Zhu, and J. Lu, "A patent time series processing component for technology intelligence by trend identification functionality," *Neural Computing and Applications,* vol. 26, pp. 345-353, / 2015.

[21] Y. Zhang, A. L. Porter, Z. Hu, Y. Guo, and N. C. Newman, ""Term clumping" for technical intelligence: a case study on dye-sensitized solar cells," *Technological Forecasting and Social Change,* vol. 85, pp. 26-39, 2014.