

Revised manuscript accepted at  
*Journal of Informetrics*

## **Deriving Technology Intelligence from Patents: Preposition-based Semantic Analysis**

January 2018

***Sungjoo Lee (Corresponding author)***

Department of Industrial Engineering, Ajou University  
Worldcup-ro 206, Yeongtong-gu, Suwon 16499, Republic of Korea  
TEL: +82 (0)31 219 2419, Fax: +82 (0)31 219 1610, E-mail: [sungjoo@ajou.ac.kr](mailto:sungjoo@ajou.ac.kr)

***Jaehyeong An***

Technology intelligence team, Hyundai Next Generation Vehicle Technology  
Gwanak-ro 1, Gwanak-gu, Seoul 08826, Republic of Korea  
TEL: +82 (0)2 870 8000, Fax: +82 (0)2 870 8888, E-mail: [jaehyeongan@hyundai-ngv.com](mailto:jaehyeongan@hyundai-ngv.com)

***Kyuwoong Kim***

Department of Industrial Engineering, Ajou University  
Worldcup-ro 206, Yeongtong-gu, Suwon 16499, Republic of Korea  
TEL: +82 (0)31 219 2419, Fax: +82 (0)31 219 1610, E-mail: [dongul@ajou.ac.kr](mailto:dongul@ajou.ac.kr)

***Letizia Mortara***

[Centre for Technology Management, Institute for Manufacturing, University of Cambridge](#)  
[Department of Engineering, Alan Reece Building, 17 Charles Babbage Road, Cambridge](#)  
[CB3 0FS - UK](#)

TEL: +44 (0)1223 764831; Fax: +44 (0)1223 464217, E-mail: [lm367@cam.ac.uk](mailto:lm367@cam.ac.uk)

It is confirmed that this paper has not been published nor is it currently being submitted elsewhere. This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korea government (MSIP) (NRF-2013R1A2A2A03016904) and also by the Ajou University research fund (grant number: S-2015-G0001-00333).

# Deriving Technology Intelligence from Patents: Preposition-based Semantic Analysis

## Abstract:

Patents are one of the most reliable sources of technology intelligence, and the true value of patent analysis stems from its capability of describing the content of technology based on the relationships between keywords. To date a number of techniques for analyzing the information contained in patent documents that focus on the relationships between keywords have been suggested. However, a drawback of the existing keyword approaches is that they cannot yet determine the types of relationships between the keywords. This study proposes a novel approach based on preposition semantic analysis network which overcomes the limitations of the existing keywords-based network analysis and demonstrates its potential through an application. A preposition is a word that defines the relationship between two neighboring words, and, in the case of patents, prepositions aid in revealing the relationships between keywords related to technologies. To demonstrate the approach, patents regarding an electric vehicle were employed. 13 prepositions were identified which could be used to define 5 relationships between neighboring technological terms: “inclusion (utilization),” “objective (purpose),” “effect,” “process,” and “likeness.” The proposed approach is expected to improve the usability of keyword-based patent analyses and support more elaborate studies on patent documents.

**Keywords:** technology intelligence; technology search; technology trends; patent analysis; semantic; preposition; text mining; key-words; text mining.

## 1. Introduction

Technology intelligence is defined as “the capture and delivery of technological information as part of the process whereby an organization develops an awareness of technological threats and opportunities” (Kerr et al., 2006, p.75). Understanding the technological landscape and identifying the changes in the landscape could be one of the core [technology intelligence](#) activities considering that companies that fail to adapt to the changing technological environment are more likely not to survive. Patent documents largely consist of technical terms, using which the characteristics of the inventions are clearly explained; if the relationships between the keywords are well identified, the keyword analysis can be a useful tool for revealing the overall technological structure of the invention and, furthermore, the technological landscape. Patents are one of the most reliable sources of information on technologies, and hence, they have been used as an essential tool for analyzing technological developments (Pilkington et al., 2009; Zhang, 2011). Accordingly, patent analysis has been frequently employed in innovation studies (Park et al., 2005), for example for the identification of industry, technology, and competitor trends; investigation of emerging technologies; and the search for potential collaborators (Ding, 2011; Moge, 1991; Narin, 1993; Reitzig, 2004). In particular, the recent advancement in data analysis techniques through text mining has enabled the use of the descriptive parts of patent documents for analysis, extending the scope of patent analysis significantly, receiving considerable attention (Choi et al., 2013; Yoon et al., 2013; Yoon et al., 2015).

To extract meaningful implications from text-mining-based patent analysis, a comprehensive selection of keywords is needed. The keywords chosen for the analysis should represent the content of all the concerned patent documents, and they should be sufficiently specific to distinguish one document from another. The general approach is to use a specific algorithm, such as term-frequency–inverse-document-frequency (TF–IDF) analysis or latent semantic analysis (Li et al., 2009; Tonta and Darvish, 2010), or to use an expert panel with domain knowledge to choose only significant keywords (Jeon et al., 2011; Lee et al., 2015). Once a keyword set is properly defined, further analysis can be conducted based on the set.

Among various types of analyses, this study focuses on the analysis of the relationships between the

keywords; the true value of text-mining-based patent analysis lies in its ability to describe the technological content of a patent based on the relationships between the keywords. The relationship between the keywords may change the meaning of the patent documents. After recognizing the significance of keyword relationships, a subject-action-object (SAO) analysis that considers sentence structure for analyzing keywords was proposed (Moehrle et al., 2005) and applied to patent analysis (Choi et al., 2012; Guo et al., 2016; Park et al., 2013; Wang et al., 2015). The SAO analysis defines the relationship between two keywords (subject and object) using a verb (action). However, despite its contribution and potential, because of the variety of verbs often used in patents, SAO inevitably increases the complexity of the analysis and cannot be used to easily investigate the technological content of a patent document. In addition, the SAO analysis focuses primarily on functional relationships between the components of a technical system (Cascini et al., 2004; Park et al., 2013), while other non-functional relationships can also be investigated to offer valuable technology intelligence from patents.

To overcome the limitations of the existing keywords-based network analysis such as SAO, this study proposes a novel approach called preposition-based semantic analysis and applies it to patent analysis for understanding the relationship between technologies, developing a technology-relation-technology (TRT) network. This method is based on the observation that number of prepositions, i.e. words that define the relationship between two neighboring words, are very limited in the English grammar. Their employment would make it easier to define the relationship between technological terms than using verbs. A preposition can also be used to describe a non-functional relationship—referred to as structural relationship in this study—which can be used to explain the relationships “between the whole and a part” or “between parts of the whole.”

Thus, the research question posed in this study is: how can **semi-automatic** methods for deriving technical insight from large number of patents be improved by using semantic analysis based on the evaluation of how technology relevant terms are linked by prepositions? In response we develop and then test a new method (TRT), comparing the insight it obtains on a particular case-study with that obtained with other state-of-the-art keyword-based methods (e.g., SAO).

The rest of this paper is organized as follows. Section 2 reviews patent-based network analyses and existing approaches for developing a network. In Section 3, based on the background knowledge, a new approach (TRT) proposed in this study is described by employing the case of electric vehicles. Section 4 introduces the method for applying TRT. Finally, Section 5 discusses the limitations of this study and future research directions.

## 2. Background

### 2.1. Patent network analysis

Networks have been widely used as a way to systematically structure an enormous amount of data, visualize the relationships between analysis units, and even quantify such relationships and the effects of one unit on the other (Kho et al., 2013; Otte and Rousseau, 2002)(Borgatti and Cross 2003, Borgatti, Mehra et al. 2009). A network consists of *nodes* and *links*: nodes are a unit of analysis that can be defined at various levels for patent documents, including keywords, papers, and authors (Noyons, 2001). Link represent the relationships between the nodes and are defined using several analysis indicators such as co-citation, co-authorships, and co-words (Chen et al., 2011; Lee and Jeong, 2008; White and McCain, 1997).

Existing studies have stressed on the use of networks to investigate the technological structure, technological changes, and emerging technologies, as a patent-based network helps to extract valuable technological insights from patents. For the circumstances wherein new technologies emerge by combining the knowledge of different fields (Curran and Leker, 2011), a network has been used as a critical tool for facilitating technology intelligence activities. Although various resources, including patents, publications, and project documents, are available for network analysis, patents are one of the most frequently adopted resources as they

provide fruitful information regarding technology innovation, being outcomes of research and development (R&D) activities (Beneito, 2006; Sun et al., 2008; Lee et al., 2010).

Two main categories for developing a network exist. The first category includes a citation-based approach in which knowledge flows are measured via citation relationships and define the relationships between the nodes. For example, focusing on the links, Han and Park (2006) measured inter-industrial knowledge flows using industry networks, whereas Li et al. (2007) tried to understand the knowledge transfer process by establishing country and institution networks. Furthermore, by focusing more on the nodes, Sternitzke et al. (2008) aimed to identify the position of applicants and inventors from inventor and applicant networks, whereas Érdi et al. (2013) predicted emerging technologies using technological networks. The second category includes a keyword-based approach, in which the co-occurrences of keywords are analyzed to define the relationships between nodes, where the nodes can be defined at the keyword, patent, technological, and organizational levels. For example, Yoon and Park (2004) analyzed up-to-date trends of technologies based on the keyword-based patent network. Similarly, by using a more advanced keyword analysis, Yoon and Kim (2011) identified evolving technological trends based on the SAO-based semantic patent network.

## 2.2. Approaches to develop a patent network

Information within patents can be broadly classified into two types: structured and unstructured data. The former comes from the bibliometric part of patent documents (e.g., applicants, inventors, international patent classification, and citations), and is relatively easy to analyze, whereas the latter is derived from the descriptive part of patent documents (e.g., abstracts, claims, and descriptions), which offers more detailed knowledge but is relatively difficult to analyze.

Most existing studies have adopted citation information to develop a patent network, exploiting the structured data within patents (e.g., Érdi et al., 2013; Han and Park, 2006; Li et al., 2007; Sternitzke et al., 2008; Takano et al., 2016). However, a citation-based patent network has several limitations. The first limitation is that in time more patents tend to be cited and there are fewer opportunities for recent patents to be cited by subsequent patents. Hence, the use of citation information is limited for recent patents and, in particular, backward citation information (Breitzman and Thomas, 2015; Yoon and Park, 2004). The second limitation is that not all relevant patents are cited; it is not always possible for an inventor to identify all previous technologies relating to his or her own invention.

Unlike this citation-based analysis, co-classification analysis uses patent classification information, which also exploits patents structured data. If a large number of patents are classified both as A and B, the two classes, A and B are regarded as related. Despite its utility, however, this approach has also a limitation in that the unit of analysis is a patent or a collection of patents and detailed analysis at the content level is not possible.

In contrast, a keyword-based network analysis relies on unstructured data (e.g., Yoon and Park, 2004). In these studies, the links between the nodes are determined based on the co-occurrence of specific keywords. This analysis has been applied to various decision-making processes, by providing keywords extracted from patents, ultimately, facilitating advances in the R&D process. For example, such an analysis has been used to identify vacant technologies, mapping technology evaluation, and supporting the derivation of new technology ideas based on the technological keywords (Lee et al., 2008; Suh and Park, 2009). Although several methods for measuring the similarity between the nodes (No et al., 2015) exist, the simplest approach is the basic co-word analysis, where co-occurrence of words in the same document is taken into account for measuring the degree of relationship between the words. A more advanced approach is the use of the word n-gram language model, which observes n-words that co-occur close together in a patent document (Gerken and Moehrle, 2012) and thus highlights closer relations between words on the content level. Here, the knowledge obtained from such networks is only the *degree* of relationships, which is simplistic compared to the rich information that patent documents contain.

However, if the *type* of relationships between the keywords in unstructured patent data could be identified,

the existing approaches to keyword-based network analysis could be improved greatly. For example, knowing that a patent is about ‘buckle’ and ‘comfort’ and that there is a strong relationship between the two words would be much less informative than knowing that the relationship between ‘buckle’ and ‘comfort’ are linked by a relationship that implies that the first ‘increases’ the latter — i.e. that the buckle increases comfort.

The SAO analysis has been proposed to examine the type of relationship between keywords, and has thus received much attention. In this analysis, the structure of words in a sentence is analyzed to represent the functional relationship between the keywords of an invention (Cascini et al., 2004; Bergmann et al., 2008; Choi et al., 2011). Here, the combination of the object and action (OA) indicates the problem (or function) whereas the subject (S) offers the solution. Thus, SAO structures are also called problem–solution structures (Moehrle and Geritz, 2007; Bergmann et al., 2008). Owing to the ability of the SAO structure to represent technological content in patents in an effective manner, it has been regarded as a breakthrough approach to keyword-based patent analysis and is used for various purposes such as technology monitoring (Gerken et al., 2010), technology opportunity analysis (Yoon and Kim, 2011), technology risk analysis (Bergmann et al., 2008; Park et al., 2011), and merge and acquisition (M&A) strategy development (Moehrle and Geritz, 2007). Network analysis is applied as a part of these analyses, where the relationship between patents (or collections of patents) is measured using the ratio of the SAOs shared by two patents. Table 8 represents the existing approaches used for keyword-based patent analysis with a particular focus on their basic concepts, and advantages and disadvantages in obtaining technology intelligence from patent documents.

**Table 1.** Previous approaches to keyword-based patent analysis

### 3. Proposed method

#### 3.1. Technological-relationship-based keyword networks

We propose a technological-relationship-based keyword network - TRT network - based on a semantic relationship between the technological keywords extracted from the descriptive parts of the patent documents. The TRT network has an advantage over the existing co-occurrence network in that it directly indicates the types of relationships between technological keywords. The TRT network provides detailed information regarding a patent or a collection of patents in a systematic way. Figure 1 illustrates the differences between a co-occurrence network and a TRT network where a pair of technological keywords is linked by semantic relations in a sentence and not by co-occurrence in a sentence or document.

We suggest that such relationships can be identified from the prepositions in the text of patent documents since patterns exist wherein prepositions are used in the texts to link technological keywords. In fact, although in general, prepositions can be interpreted in various ways based on the context, unlike for other documents, patent texts are consistently developed in a manner that explains the characteristics of the invention. Here, the context is limited to the description of the invention principles. These characteristics of patent documents make it possible to derive the type of relationship using prepositions. For example, the following is an excerpt from a patent published by the United States Patent and Trademark Office (USPTO).

“A battery ECU provides an N-th arithmetic processing device with an ID assignment signal. The N-th arithmetic processing device receives the ID assignment signal, and stores an ID that is transmitted by the battery ECU. The battery ECU provides an i-th arithmetic processing device (i is a natural number from N to 2) with a command for providing an ID assignment signal. The N-th arithmetic processing device provides the (i-1)-th arithmetic processing device with the ID assignment signal in response to the command for providing an ID assignment signal. The (i-1)-th

arithmetic processing device receives the ID assignment signal, and stores an ID that is transmitted by the battery ECU. (US 9446679, USPTO)”

As shown in the text, the majority of the nouns in a patent document are related to the technological terms, while the prepositions indicate the relationships between these technical terms. The prepositions are assigned to one of the following relationships: inclusion (utilization), objective, effect, process, and likeness.

**Figure 1.** Technological-relationship-based keyword network structure

### 3.2. Overall process of developing a TRT network

The overall process of developing a TRT network comprises four stages (see Figure 2).

**Figure 2.** Overall development process of a TRT network

In the first stage, relevant patent data is collected, where both unstructured data (e.g., patent abstracts) and structured data (e.g., applicants, application year, and citation information) are included for the technologies of concern.

In the second stage, the data is preprocessed, and the noun–preposition phrases in the abstracts are extracted. OpenNLP and CoreNLP of the R programming package were adopted for the natural language processing (NLP). Here, all noun–preposition phrases are first retrieved without any stop-word lists and the target technological keywords are then used to select the final noun–preposition phrases for further analyses. The basic principle for choosing noun–preposition phrases is that the noun phrase (NP) and preposition phrase (PP) should have the same depth in a sentence tree developed using text parsing. If a PP has several NPs (or nouns), which are linked via “and” or “or” that are at the same depth in their sentence tree and linked via “and” or “or,” multiple noun–preposition phrases are extracted. For example, the sentence, “the production process for electric motors and high voltage batteries are as follows” can be parsed into three components at the depth 1 (see Figure 3). At this depth, one noun-preposition phrase [the production process] – [for electric motors and high-voltage batteries] can be extracted. Then, it is found that the preposition phrase has two noun phrases [electric motors] – [high-voltage batteries] after a preposition [for] and these are linked by a conjunction [and] at depth 2. Thus, the noun-preposition phrase initially extracted from the sentence is broken down into two further sub-phrases; 1) [production process] – [for electric motors]; and 2) [production process] – [for high-voltage batteries]. These sub-phrases are used as an input for the following stage.

**Figure 3.** An example of sentence tree to extract noun-preposition phrases

In the third stage, a set of TRTs are identified. To achieve this, the structures of the technological keyword ( $T_j$ ), preposition ( $P_i$ ), and technological keyword ( $T_k$ ), that is a TPT (technology-preposition-technology), are identified using the noun–preposition phrases in the *preposition relationship* analysis. The first technological keyword is extracted from the noun phrase, while the second technological keyword is extracted from the preposition phrase. For the term in between the two keywords, the preposition used for the preposition phrase is adopted. Thereafter, these preposition relationships are clustered into several groups with similar  $P_i$ s based on their roles of defining the relationships between the two keywords. The relationships are required to be verified. That is, it is examined whether the relationships between the two keywords defined by  $P_i$  indicate the correct relationships in the actual patent documents. Hence, the TPTs can be transformed into the TRTs.

In the final stage, a TRT network is developed. Here, the target technological keywords are chosen based on the relevant TRTs that are identified; the target keyword candidates are positioned both before and after the

prepositions (relationships). That is, after collecting all the candidate keywords (by automatically identifying and extracting all possible keywords in the text that surround the TRT), the final keywords are evaluated and selected by domain experts. Expert knowledge is hence required to construct a valid set of keywords (Jeon et al., 2011; Lee et al., 2015), which is an important process of keyword-based patent analysis. A TRT network can be developed either for each type of relationship or for all types of relationships. More specifically, a *relationship matrix* for relationship type  $k$  is constructed based on the TRT with relationship type  $k$ ; the frequency of TRT in the form of [Ti]–[relationship type  $k$ ]–[Tj] becomes a value for cell  $(i, j)$  of the matrix. These cell values are transformed into binary values to obtain an input for the network; the cell value 1 indicates that  $T_i$  is related to  $T_j$ , whereas the cell value 0 indicates that no relationship between the two keywords exists. Such matrixes and networks can be used to understand the structure of the technology, investigate technological trends, and generate new technological ideas. More detailed procedures for developing a TRT network are described in Figure 4.

**Figure 4.** Detailed procedures of developing a TRT network

Here, one important step that must be completed before developing a TRT network involves defining rules for assigning the prepositions used to link the keywords to the relationships between the keywords. In this study, the prepositions extracted from TPTs can be grouped into several relationships in the following manner. Firstly, all prepositions (for, to, etc..) present in the NP–PP pairs are identified, and are prioritized according to their frequency. Secondly, the prepositions are analyzed and assigned to one of the relationships defined. To do this, the original meaning of the prepositions needs to be referenced and calibrated. Hence, using a training data-set, their actual usage in patent documents is initially evaluated by experts. This approach is only done once and on a small portion of the sample and it defines the generic rules which will be adopted for the larger data-set. For example, the preposition “for” has the following definition of “having a purpose or function” according to the oxford dictionary (<https://en.oxforddictionaries.com/definition/for>). If it is observed that the preposition “for” is actually used to indicate the purpose or the function of the technology in several patent documents randomly chosen (the small dataset subjected to experts judgement), we will consider this preposition useful to represent the relationship between technological terms which relates to the aim/objective/purpose of the use of the technology in the largest dataset where the further analysis is applied. As a result of the initial expert-led approach, a set of prepositions that can be used to define the various relationships which will be listed as a provisional set. On the other hand, if the prepositions are not univocally associated with a type of relationship or are very rare, they will be not employed as their generalizability is not possible. Finally, the prepositions in the provisional set identified above are tested using a test data set to check whether the assignments of prepositions (e.g., for) to the predefined relationships (e.g., purpose) are valid. Here, statistical approaches are used, where the confidence interval of the population proportion (the ratio of using the preposition,  $p_i$ , for the relationship assigned to it) is estimated at the significance level of 0.05. Accordingly, if the min value on the estimate is less than 0.6 (cut-off value), we would regard the assignment is not valid and thus the preposition  $p_i$ , would be dropped. In the following sections, a more detailed explanation is presented with the help of the case of electric vehicles.

### 3.3. Detailed procedures of developing a TRT network

#### 3.3.1 Collection of patent data

For the electric vehicle, which is employed as a case study for explaining the method of developing a TRT network, 2,838 relevant patents registered in the USPTO from 2011 to 2015 were collected. Only published patents were focused on owing to their value as novel inventions, and “WISDOMAIN,” a commercial website (<http://www.wisdomain.com>), was used for this collection. The search string used was “TAC=((electric W/1 car) or (electric W/1 vehicle) or (electric W/1 automobile)) and AD=20110101:20151231 and RD<20160701.” Here TAC indicates “title, abstract and claim”, and thus the search was conducted in the three parts of patent documents.

On the other hand, AD and RD mean “application date” and “registered data.” Therefore, the search was limited to the patents applied between 2011 to 2015 and registered before June 2016. We assumed that all the patents retrieved by the search string were valid for analysis because a quick review of the patents indicated that they were clear enough to be used as an input for further analysis. However, it is essential to consider the recall and precision values and eliminate noisy data to construct a valid database (Egghe, 2008). While extracting the TRTs, the abstracts were considered for analysis. As the abstracts of the patent documents describe the principle and the purpose of invention clearly and concisely, the existing studies on patent analysis have made use of the abstracts of patents (e.g., Chen et al., 2003; Tong et al., 2006; Verhaegen et al., 2009); titles may be too general to provide specific information while claims may have legal terms rather than technical terms. The work by Noh et al. (2015) also indicated that using “abstracts” for text mining could produce the most representative keywords for patent documents. In addition to the abstracts, a set of structured data was collected for further analysis based on the TRT network. The final database for analysis included the following four fields: abstracts, applicants, application year, and cooperative patent classification (CPC).

### 3.3.2 Phrase extraction from patent text

The collected patent texts were broken down into phrases using NLP. As the objective of this study is to identify the phrases that are semantically connected through prepositions, all the pairs of phrases in the form of NP–PP were identified, for which an *NP–PP extractor program* was developed internally using the R programming package. Consequently, 18,949 NP–PP pairs were collected from the 2838 patent abstracts. Table 2 lists the NP–PP pairs.

**Table 2.** Examples of electric vehicle patents NP–PP pairs

### 3.3.3 Deriving technology relations

At this stage, the 18,949 NP–PP pairs were transformed into the TRTs. For the transformation process, the TPT structures were first extracted from each NP–PP pair; one technological keyword was identified from each NP and PP, while a preposition was extracted from the PP. Moreover, using the R programming package, a TRT converter that extracts the TPT structures from the NP–PP structures was internally developed. For example, in Table 1, the NP–PP pair of “A hydraulic system-for a hybrid module” (No 2) was restructured to the TPT of “Hydraulic system–for–Hybrid module.” The number of TPTs identified was the same as the number of NP–PP pairs.

In the next step, the prepositions in the whole set of patent documents were analyzed based on their frequency of occurrence. For this analysis, the semantic meaning of each preposition as well as the NP–PP pairs for each preposition, were investigated. From the 18,949 TPTs on electric vehicles, 59 different prepositions were identified and sorted by their frequency. The most frequently observed preposition was “of” (48.8% of the total prepositions used). By applying the threshold value of 0.25% of the total prepositions (48 observations), the top most frequent 15 prepositions were selected. Then, their role in linking two technical keywords was investigated to define the relationships between the keywords indicated by the prepositions. *To do this, 100 randomly selected TPTs were investigated for each preposition, if its observation frequency is greater than 100 (i.e., of, for, to, in, with, from, on, as, by, and at). For the other prepositions (i.e., during, into, through, within, and via), all the relevant TPTs were examined. We regarded the preposition as useful if its observation is large enough (greater than 0.25% of the total prepositions used) and indicates a single relationship (greater than 70% of the observations).*

In this study, the following five types of relationships were defined from those prepositions: inclusion (utilization), objective, effect, process, and likeness. Thus, the 15 prepositions were assigned to these predefined five relationships. However, one of the preposition (i.e., By) could not be assigned to a single relationship, as it



was found to be used for a variety of relationships (hence its univocal meaning could not be pre-determined). Thus, it was excluded from the provisional set for further analysis. Consequently, out of the 59 prepositions, 14 were regarded as meaningful for defining the relationships between two technological keywords, accounting for 93.3% of the prepositions. Table 3 lists the five types of relationships and their corresponding prepositions.

**Table 3.** Preposition analysis results

The relationships defined by the prepositions required verification; the predefined relationships in the previous step were thus investigated to check whether these predefined relationships aligned with the actual relationships in the whole patent documents. For each preposition, 50 TPTs were randomly selected from the 18,949 TPTs. The definitions in the patent text were then compared with those in Table 4.

**Table 4.** Types of relationships and their corresponding prepositions

For example, the following is an excerpt from the patent text having the TRT of “output shaft-of-transmission:”

“the system includes a gear set having first, second, third and fourth elements. The second element connects with an *output shaft* of the *transmission* and a first output shaft of an internal combustion engine connects with the third element for driving the third element. (US 20140100071, USPTO)”

In this context, we can confirm that also the evaluation of the complete text confirms that the relationship implied by the preposition "of" (inclusion) is accurate as the “output shaft” is part of the “transmission”. The verification results in Table 5 indicated that the average accuracy of this approach is 82.6%; only one preposition “at” had an accuracy lower than 70%, and hence it was eliminated from the initial list of prepositions. Table 5 presents the information regarding the accuracy and error with a 95% confidence level. Hence, a total of 17,702 TRTs having 13 prepositions were defined from the 18,949 TPTs obtained initially. Although only 13 prepositions were adopted for the analysis, they explained 92.1% cases of the preposition presented in our experiment. In addition, the rules developed using the above case can be generalized for other cases, as the meaning of the preposition hardly changes with the context (technology areas).

**Table 5.** Verification analysis results

### 3.3.4 Developing a TRT network

The TRT network is intended to provide more information regarding the relationships between the keywords. The 17,702 TRTs extracted in the previous stage were used to develop the network. In the first step, a set of keywords from the list of technological keywords in the TRT set was selected. If there was a set of predefined keywords of concern, they were used for the analysis. In this study, no predefined keywords were provided and thus the top 10% of the keywords based on their frequencies were chosen, from which the general terms were screened such that only the keywords concerning the technical terms were included. From the 17,702 TRTs, 14,349 keywords were identified. The 1,434 keywords corresponding to the top 10% of the frequencies were investigated, which accounts for 776 general keywords and 658 technological keywords.

Those 658 technological keywords (raw technological terms) were then redefined to consolidate the keywords with same (similar) meanings, which were finally reduced to 108 keywords (target technological terms) by clustering those with similar meanings and standardizing the terms indicating the similar meanings (see Table

6). Therefore, although only 108 target terms were used for analysis, they could cover all the 658 terms identified. During this process, three sources were used as a reference for the keyword merging and standardization. The first source is a set of keywords clustered based on similarities listed in WordNet. However, this approach was not as effective as we expected for clustering technical terms. Thus, a keyword co-occurrence table was developed instead to indicate the relationships between the keywords. As a second source, technical documents related to electric vehicles were reviewed for reference. Finally, a list of keywords having high values for both TF and TF-IDF values were analyzed for identifying core keywords. Out of the 17,702 TRTs, 1179 TRTs contained 108 keywords that had exactly the same or similar meanings. The 1179 TRTs consisted of 721 *inclusion (utilization)* relationships (61.2%), 253 *objective* relationships (21.5%), 126 *effect* relationships (10.7%), 42 *process* relationships (3.6%), and 37 *likeness* relationships (3.1%). These customized sets of TRTs were used for further analysis. Finally, a keyword network based on the TRTs is developed, and is used to visualize the key technologies and their relationships. The R package *igraph* was used to visualize the network, with nodes indicating the technological keywords and arcs representing the relationships between the keywords. By using the 1179 TRTs, five TRT networks for each relationship could be developed.

**Table 6.** A list of technological keywords for an electric vehicle

### 3.3.5 Results of a TRT network

Figure 5 shows an example of a TRT network (inclusion (utilization)) on an electric vehicle. The entire networks for each of the five relationships are provided in the appendix (see Figure A2, A3, and A5). Unlike other keyword-based networks, a TRT network is directional; the keywords pointed to by the arrows indicate the keywords T2, whereas those where the arrows start designate the keywords T1. For this visualization, the R package *igraph* was used. In the figure, the two keywords linked to each other are in a relationship of inclusion (utilization). For instance, *user:input* (T1) and *shaft* (T1) are linked to *motor* (T2) with *inclusion (utilization)* relationships, implying that “motor” (T2) includes (utilizes) “user input” (T1) and “shaft” (T1), which provides an overview of the structure of the electric vehicle technologies.

**Figure 5.** TRT network for inclusion (utilization) relationships: an electric vehicle

The five TRT networks can be used either individually or collectively. Figure 6 illustrates a network that describes all the five relationships. The figure is only a part of the network that focuses on battery-related technologies. The keywords concerning “battery” in an electric vehicle include “(high voltage, voltage, lithium ion, rechargeable, traction, and secondary) + battery,” “battery + (cell, module, pack, system, charger, state, management, and capacity),” and “battery.” These keywords are positioned in front of a TRT structure (the former T) and are called *base keywords* in this study (the square nodes in the figure), as they are the base targets for analysis. The keywords linked to the base keywords, regardless of the types of relationships, are named *relation keywords*, which are positioned in the back of a TRT structure (the latter T). The network is particularly useful for understanding the technological structure, but will have more applications when it is used along with other bibliometric parts of patent information. The following sections describe more applications with explanations.

**Figure 6.** TRT network visualization of electric vehicle battery technology

## 4. Application of the proposed method

This section explains how the TRT-network can be used for creating value in [technology intelligence](#) activities, particularly when used along with other structured data from patent documents. The following three possible

applications are suggested considering the case of an electric vehicle: 1) monitoring technology trend; 2) identifying similar technologies; and 3) discovering technology opportunities. Three types of structured data—application year, CPC, and applicants—are used along with the network. However, various other applications can also be made possible by combining other types of structured data.

#### 4.1. Monitoring technology trends

The trend analysis on the TRT network can be conducted using the year of application. The evolution of technological structures investigated based on the changes in technologies (technological keywords) as well as the changes in the relationships between technologies (the relationships between technological keywords) can be revealed from the trend analysis. This investigation is performed to answer the following questions: how are the new technologies combined with existing technologies to shape the emerging technological fields? How are the existing technologies separated from others to form a new technological field?

Figure 7 represents the trend analysis from 2011 to 2015 while focusing only on battery-related keywords. It shows the ways in which the base keywords (battery, battery cell, battery charger, battery module, battery pack, and battery system) have formed (new) relationships with (new) relation keywords (e.g., electrode, drive system, network, etc.) over the course of the five years.

**Figure 7.** Trend TRT analysis on battery technologies from 2011 to 2015

Figure 8 shows such keywords and relationships in a TRT network at two different points: 2011 and 2015. It is apparent that the *electronic battery* has more sub-technologies with a more complicated technological structure in 2015 than in 2011. For example, new technological features, such as *coolant*, *battery state*, *network*, and *dc power*, started to appear in 2015. These batteries are being used in vehicles in 2015; the keyword, *vehicle*, are linked to all the base keywords but one (i.e., *battery cell*, *battery pack*, *electronic battery*, *battery charger*, *battery system*, and *battery*), where these keywords have the relationship of *objective* with the keyword, *vehicle*.

**Figure 8.** TRT networks on battery technologies in 2011 and 2015

#### 4.2. Identifying similar technologies

The TRT network is useful for identifying similar technologies evaluated by both the content similarity (similarity in keywords) and the structural similarity (similarity in the relationships between keywords). To do this, the target technologies for analysis should be determined; thereafter, their TRT networks are developed. Accordingly, five TRT networks for each technology are constructed. Then, the structural similarity between the two technologies can be measured together with content similarity using the same set of keywords. The use of structural similarity along with content similarity is expected to improve the performance in searching for similar technologies.

The detailed processes involved in measuring the content structural similarity between the technologies are as follows. First, content similarities between the technologies are evaluated using a keyword vector and cosine similarity (No et al., 2015). A general approach for analyzing the content similarity is to develop a *keyword vector* for each technology (a collection of patents) that indicates the frequency of each keyword, and then to measure a cosine similarity value between the technologies. Second, the same approach can be applied to measure the structural similarity. For each TRT network, the cosine similarity between two technologies is measured, as shown in Figure 9. Accordingly, five types of cosine similarities are measured for all technological pairs, which are averaged with equal weights. In this study, the more similar the technologies  $i$  and  $j$  are, the closer the value to 1 the cell  $(i,j)$  in the matrix will have; in the opposite case, the cell  $(i,j)$  in the matrix will approach 0. If required, further analysis can be conducted to visualize the relationships between technologies based on the structural

similarly and the pre-determined cut-off value for the degree of similarity.

### **Figure 9.** Structural similarity analysis process

In this study, the CPC was set at the main group level as a unit of analysis (technology), and a similarity analysis on the top 100 CPCs in terms of patent applications on electric vehicles was performed. For the given 100 CPCs, 4950 technological pairs required analysis. Table 7 lists the 10 CPCs with the highest structural similarity values. The analysis results indicated that the content similarity values were not always proportional with the structural similarity values. Moreover, even when the content similarity values are greater than 0.9, their structural similarity values are less than 0.6. A more accurate similarity analysis can be realized using the structural similarity analysis along with the content similarity analysis.

**Table 7.** The top 10 technological pairs with high structural similarity values

Based on these structural similarity values, a technology network was developed to illustrate groups of similar technologies with a cut-off value of 0.5. The cut-off value could be adjusted based on the purpose of the analysis. For the exclusion of isolated CPCs, 26 CPCs were grouped into five clusters, as shown in Figure 10. The cluster with the four CPCs—B60W-0010/06 (including control of combustion engines), B60W-0010/08 (including control of electric propulsion units), B60W-0020/00 (control systems specially adapted for hybrid vehicles), and Y02T-0010/6286 (control systems for power distribution between ICE and other motor or motors)—had the highest average values for the structural similarity analysis. The result reveals that these are the most similar technologies in an electric vehicle. This similarity analysis will be useful for developing a technology tree or avoiding a technology infringement risk.

### **Figure 10.** Clusters of technologies with similar structures

#### 4.3. Discovering technology opportunities

The TRT network can also be used to discover technology opportunities systematically, for which TRT networks with similar structure but with different areas are compared to derive new technological ideas and possible collaboration partners. In particular, technologies with high structural similarity are first identified and their TRT networks are developed for each technology. If a specific keyword set (TRT) exists that appears only once in the networks despite the overall structural similarity, the set is applied to other networks to check whether such an application would facilitate a new technology opportunity. In addition, if a company requires a partner to pursue the opportunity, the information regarding the candidates of the collaboration can be provided by identifying the applicants of the patents with the TRT. For example, Y02T-0010/6286 in Table 7 has three similar technologies, including B60W-0010/06, B60W-0010/08, and B60W-0020/00. Table 8 gives the list of the TRTs that appears in each technology but not in the target Y02T-0010/6286 technology, along with the companies that possess the capabilities related each of the TRTs. Although the technological keywords were defined based on frequently used general terms in this study, more specific and specialized keywords can be adopted based on the purpose of analysis.

**Table 8.** Idea for developing technology Y02T-0010/6286

#### 4.4. Discussions

##### 4.4.1. Methodological issues

This study proposed a novel approach for developing a keyword network based on the patent data and described how the proposed approach can be used to obtain meaningful technology intelligence. During the process, three issues were raised in obtaining the expected value from the approach.

Firstly, the method of assigning the prepositions to the types of relationship needs to be carefully determined. This study relies significantly on the prepositions “for” defining the relationships between the keywords. Therefore, understanding the accurate meaning of the prepositions as a link between keywords in the patent document is essential for guaranteeing the value of the TRT networks. In this study, five relationships were identified and tested by using the patents of an electric vehicle. However, they are required to be tested in other technological areas to ensure their external validity. In addition, assuming that language may evolve over time, a periodic examination is required to confirm the reliability of the TRT networks.

Secondly, as seen in other keyword-based patent analyses, a careful selection of the technological keywords is a precondition for obtaining significant implications. In this study, the keywords with high frequency were chosen for further analysis, and thus relatively general keywords were selected. However, other approaches such as TF-IDF or user-defined keywords can also be used to provide more specific keywords. Consequently, the involvement of experts in developing a TRT network could increase its utility significantly.

Thirdly, the unit of analysis can be determined according to the purpose of analysis. A technology was defined as a CPC. However, the unit of analysis can be defined either at the patent level or at the collection-of-patent level; more specific idea generation is feasible when this approach is used at the patent level. In particular, for this analysis, the use of the data regarding patent claims enables a more accurate evaluation of the possibilities of patent infringement than that with the conventional keyword-based similarity analysis. Furthermore, the approach could be applicable for searching for prior art patents that are similar to a new invention, which would help in developing an invention in order to avoid the infringement of patents.

#### 4.4.2. Practical issues

In addition to the methodological issues, several practical issues need to be discussed. Firstly, although this study described only three applications of TRT to electric vehicles, the application results provided insightful practical implications to develop R&D strategies. For example, the trend analysis results indicate that battery technology for electric vehicles has evolved towards more advanced element technologies such as coolant, battery state, network and dc power; it is becoming more feasible to introduce such technologies into products considering their recent development. The similarity analysis results showed that 26 technologies, out of 100 electric vehicle technologies (defined by CPC), could be grouped into five clusters based on their content and structural similarities; these clusters are major technology fields for electric vehicles. When considering only content similarity, most technologies may be regarded as similar since they are likely to share common keywords on electric vehicles. However, this limitation can be resolved by considering structural similarity as well as content similarity. Finally, the opportunity analysis results presented that when a company with competitive advantage on technology Y02T-0010/6286 may examine the technological concepts B60W-0010/06, B60W-0010/08, and B60W-0020/00, especially, the concepts listed in Table 8, in developing new technologies. During this process, such companies as Tesla Motors, Ford Global Technologies, GM Global Technology, Toyota, Hyundai Motor, Continental Automotive, Suzuki Motor, and Fuji Heavy Industries can be potential collaboration partners, since they share similar technology interests (similar technologies) but with somewhat different approaches (different TRTs).

Secondly, the technological areas in which a TRT network is expected to be used optimally are worth discussing. The suggested approach can be useful particularly in the areas characterized by a rapid technological change and a wide technological scope for innovation such as information and communications technology. Rapid technological changes make it difficult to monitor emerging technologies; nevertheless, the TRT network summarizes the trends and characteristics of the technologies, thus providing much more detailed technological

information than the conventional keyword networks. In addition, a wide technological scope offers various possibilities with respect to technology convergence. A TRT network introduces a new technological structure from other areas into a focal area, which provides possibilities for new technology creation. The suggested approach is based on a large set of data, and if the data size is sufficiently small to allow the identification of only a few prepositions, meaningful relationships between technologies cannot be drawn from the analysis. Technologies that are in their early stage might be difficult to examine with the suggested approach, if there exist only a few patents for the technologies.

Finally, although this study suggested only three applications for the use of the TRT analysis for technology intelligence, other applications can also be designed. In particular, when the TRT approach is combined with other approaches such as a bibliometric analysis on patents, technology intelligence can be expanded to competitor intelligence. In addition, other applications that can make the best use of the values obtained from the TRT networks are required to be devised.

#### 4.4.3. TRT and SAO

To be able to claim the contribution of the technique suggested in this study, we evaluated the relative strength and weakness of the TRT analysis compared to one of the currently most used/known approaches (i.e., SAO analysis). The affinity of TRT approach proposed in this study with the SAO approach stems from the fact that they are both based on semantic content analysis. Park et al. (2013, p. 2374) explained that

“In the technology field, SAO structures have mainly been used to represent the functionalities of technologies. [...] Because a ‘function’, which can be defined as an action or task that a system is able to perform, can be represented as Action-Object (AO), and a ‘tool’ or ‘method’ which makes a function do can be represented as Subject (S) in a technological sentence, the functionalities of technical systems and their components can be clearly expressed using SAO structures. In particular, S and O represent components of a technical system and A specifies a functional relationship between the components (Cascini et al., 2004)”.

However, whilst the SAO analysis focuses on functional relationships derived through the analysis of the text according to the meaning of verbs, the newly proposed TRT analysis is used to represent non-functional structural relationships based on prepositions. Some structural relationships cannot be easily explained using SAO structures, and valuable information could be missed out if we try to investigate the technological landscape relying only on SAO structures with limited verbs. Instead this work proposes that the TRT analysis could extract meaningful information that cannot be otherwise identified using SAO analysis.

Secondly, because of the wide range of possible verbs that can be used, the traditional SAO analysis can be complex. Although some studies have tried to minimize the complexity of SAO by summarizing key verbs or use only key technological terms to reduce this variety (e.g., Choi et al., 2013; Park et al., 2013; Wang et al., 2017;), the TRT analysis, which uses a limited number of clearly defined prepositions to identify the relationships between technological terms, is likely to reduce complexity further and to be simpler to use, more reliable and consistent than the SAO analysis for analyzing technological structures.

For this comparative analysis, the SAO analysis was conducted using the same set of technological keywords and documents used for TRT. A total of 1,127 SAOs were identified as shown in Table 9, a similar number compared to the TRTs extracted (1,179 TRTs). However, the five most frequently occurring prepositions could explain more than 82.22% of the TRTs, whereas only 53.33% of SAOs could be explained by the five most frequently occurring action-keywords. Moreover, the TRT analysis produced only 25 prepositions while 97 action-keywords were obtained from the SAO analysis. These figures indicate that the time and effort to analyze the relationships between technological keywords can be reduced by using TRT, which is the advantage of TRT in

terms of the efficiency of analysis; the complexity of analysis is decreased with the relationships that are clearly defined with prepositions. What is more important is the type of relationship identified in these two approaches. The results show that the TRT analysis is more effective in exploring structural relationships because most of the action-keywords correspond to two structural relationships, inclusion (utilization) and effect. Few action-keywords in the SAO analysis can describe objective, process, and likeness relationships. In contrast, the SAO analysis may offer more information regarding a specific function, such that the specific effects that are important. Therefore, the two approaches have their pros and cons and could be considered complementary.

**Table 9.** SAO analysis results

## 5. Conclusions

This study proposed a preposition-based keyword network, namely, the TRT network. This network is different from those used in the existing approaches in that the keyword network is developed using the semantic relations between the technological keywords, defined by the prepositions in the phrases with the two keywords. In this approach, the four processes for constructing a TRT network from a collection of patent documents using text mining and network visualization techniques were explained. We also exemplified how the TRT network can be used along with other analyses using the bibliometric patent data to create valuable technology intelligence in practice. Although the case of electric vehicles was employed to develop the approach as well as verify its feasibility and usability, the proposed approach is not limited to a particular technology area.

This study is one of the earliest attempts in which prepositions are used for analyzing patent texts by paying attention to the role of prepositions in defining the relationships between the technological keywords linked by the prepositions which have so far attracted little interest in semantic patent analyses. This study has proved the potential of prepositions in patent analysis. The 13 prepositions capable of defining the relationships between relevant keywords were identified and assigned to five categories of relationships—inclusions (utilization), objective, effect, process, and likeness. A TRT network provided information regarding the types of relationships in an effective manner in addition to the degree of relationships. It offered valuable intelligence with respect to technological structure and trends. Hence, it can be used in supporting various types of decision making, monitoring technology trends, identifying similar technologies, and facilitating new technology creation. In practice, the main contribution of this study is to have provided a feasible semi-automated process for the development of a TRT network, which can reduce the time and cost required to extract technology intelligence and support decision-making.

Despite these meaningful contributions, this study has several limitations. Firstly, only a particular set of patents, that is, patents regarding an electric vehicle, was used for assigning the prepositions to the relationships. More experiments are required for a TRT network approach to have external validity. Moreover, a TRT network is developed based on user-defined keywords. As a result, the degree of domain knowledge of the user will affect the amount and specificity of information the network provides. A more elaborate approach is required to help users with the selection of appropriate technological keywords. Secondly, we restricted our focus only to preposition-based relationships. However, other relationships can be used with other approaches. For example, the suggested approach does not deal with items in a series, as this structure uses no prepositions, although these items may be related to each other. Thus, the proposed approach needs to be complemented by other available approaches such as SAO analysis. We started to compare SAO-based approach with TRT-based approach demonstrating their strength and weaknesses and complementarities. However, more systematic analyses need to be conducted on more cases.





## References

- Beneito, P. (2006). The innovative performance of in-house and contracted R&D in terms of patents and utility models. *Research Policy*, 35(4), 502-517.
- Bergmann, I., Butzke, D., Walter, L., Fuerste, J. P., Moehrle, M. G., & Erdmann, V. A. (2008). Evaluating the risk of patent infringement by means of semantic patent analysis: The case of DNA chips. *R&D Management*, 38(5), 550-562.
- Borgatti, S.P., & Cross, R. (2003). A relational view of information seeking and learning in social networks. *Management Science*, 49(4): 432-445.
- Borgatti, S.P., Mehra, A., Brass, D.J., & Labianca, G. (2009). Network analysis in the social sciences. *Science* 323(5916), 892-895.
- Breitzman, A., & Thomas, P. (2015). The emerging clusters model: A tool for identifying emerging technologies across multiple patent systems. *Research Policy*, 44(1), 195-205.
- Cascini, G., Fantechi, A., & Spinicci, E. (2004). Natural language processing of patents and technical documentation. *Lecture Notes in Computer Science*, 3163, 508–520
- Chen, D. Z., Huang, M. H., Hsieh, H. C., & Lin, C. P. (2011). Identifying missing relevant patent citation links by using bibliographic coupling in LED illuminating technology. *Journal of Informetrics*, 5(3), 400-412.
- Chen, L., Tokuda, N., & Adachi, H. (2003). A patent document retrieval system addressing both semantic and syntactic properties: Association for Computational Linguistics.
- Choi, S., Kim, H., Yoon, J., Kim, K., & Lee, J. Y. (2013). An SAO-based text-mining approach for technology roadmapping using patent information. *R&D Management*, 43(1), 52-74.
- Choi, S., Park, H., Kang, D., Lee, J. Y., & Kim, K. (2012). An SAO-based text mining approach to building a technology tree for technology planning. *Expert Systems with Applications*, 39(13), 11443-11455.
- Choi, S., Yoon, J., Kim, K., Lee, J. Y., & Kim, C. H. (2011). SAO network analysis of patents for technology trends identification: a case study of polymer electrolyte membrane technology in proton exchange membrane fuel cells. *Scientometrics*, 88(3), 863-883.
- Curran, C. S., & Leker, J. (2011). Patent indicators for monitoring convergence—examples from NFF and ICT. *Technological Forecasting and Social Change*, 78(2), 256-273.
- Ding, Y. (2011). Scientific collaboration and endorsement: Network analysis of coauthorship and citation networks. *Journal of informetrics*, 5(1), 187-203.
- Duguet, E., & MacGarvie, M. (2005). How well do patent citations measure flows of technology? Evidence from French innovation surveys. *Economics of Innovation and New Technology*, 14(5), 375-393.
- Egghe, L. (2008). The measures precision, recall, fallout and miss as a function of the number of retrieved documents and their mutual interrelations. *Information Processing & Management*, 44(2), 856-876.
- Érdi, P., Makovi, K., Somogyvári, Z., Strandburg, K., Tobochnik, J., Volf, P., & Zálányi, L. (2013). Prediction of emerging technologies based on analysis of the US patent citation network. *Scientometrics*, 95(1), 225-242.
- Gerken, J. M., & Moehrle, M. G. (2012). A new instrument for technology monitoring: novelty in patents measured by semantic patent analysis. *Scientometrics*, 91(3), 645-670.
- Gerken, J. M., Moehrle, M. G., & Walter, L. (2010). Patents as an information source for product forecasting: Insights from a longitudinal study in the automotive industry. *The R&D Management Conference*, 2010-3.
- Guo, J., Wang, X., Li, Q., & Zhu, D. (2016). Subject–action–object-based morphology analysis for determining the direction of technological change. *Technological Forecasting and Social Change*, 105, 27-40.
- Han, Y. J., & Park, Y. (2006). Patent network analysis of inter-industrial knowledge flows: The case of Korea between traditional and emerging industries. *World Patent Information*, 28(3), 235-247.
- Jeon, J., Lee, C., & Park, Y. (2011). How to use patent information to search potential technology partners in open innovation. *Journal of Intellectual Property Rights*, 16(5), 385-393.

- Kho, J., Cho, K., & Cho, Y. (2013). A Study on Recent Research Trend in Management of Technology Using Keywords Network Analysis. *Journal of Intelligence and Information Systems*, 19(2), 101-123.
- Kim, H. B., Hyeok, Y. J., & Kim, K. S. (2012, June). Semantic SAO network of patents for reusability of inventive knowledge. In *Management of Innovation and Technology (ICMIT), 2012 IEEE International Conference on* (pp. 510-515). IEEE.
- Lee, B., & Jeong, Y. I. (2008). Mapping Korea's national R&D domain of robot technology by using the co-word analysis. *Scientometrics*, 77(1), 3-19.
- Lee, C., Kang, B., & Shin, J. (2015). Novelty-focused patent mapping for technology opportunity analysis. *Technological Forecasting and Social Change*, 90, 355-365.
- Lee, M., Kim, K., & Cho, Y. (2010). A study on the relationship between technology diffusion and new product diffusion. *Technological Forecasting and Social Change*, 77(5), 796-802.
- Lee, S., Lee, S., Seol, H., & Park, Y. (2008). Using patent information for designing new product and technology: keyword based technology roadmapping. *R&D Management*, 38(2), 169-188.
- Li, X., Chen, H., Huang, Z., & Roco, M. C. (2007). Patent citation network in nanotechnology (1976–2004). *Journal of Nanoparticle Research*, 9(3), 337-352.
- Li, Y. R., Wang, L. H., & Hong, C. F. (2009). Extracting the significant-rare keywords for patent analysis. *Expert Systems with Applications*, 36(3), 5200-5204.
- Moehrle, M. G., Walter, L., Geritz, A., & Müller, S. (2005). Patent-based inventor profiles as a basis for human resource decisions in research and development. *R&D Management*, 35(5), 513-524.
- Moehrle, M.G., & Geritz, A. (2007). Developing acquisition strategies based on patent maps. *Management of Technology: New Directions in Technology Management*. Sherif, M.H.; Khalil, T.M. (Eds.). Oxford, Amsterdam: Elsevier, pp. 19-29.
- Mogee, M. E. (1991). Using patent data for technology analysis and planning. *Research-Technology Management*, 34(4), 43-49.
- Narin, F. (1993). Technology indicators and corporate strategy. *Review of Business*, 14(3), 19.
- No, H. J., An, Y., & Park, Y. (2015). A structured approach to explore knowledge flows through technology-based business methods by integrating patent citation analysis and text mining. *Technological Forecasting and Social Change*, 97, 181-192.
- Noh, H., Jo, Y., & Lee, S. (2015). Keyword selection and processing strategy for applying text mining to patent analysis. *Expert Systems with Applications*, 42(9), 4348-4360.
- Noyons, E. (2001). Bibliometric mapping of science in a policy context. *Scientometrics*, 50(1), 83-98.
- Otte, E., & Rousseau, R. (2002). Social network analysis: A powerful strategy, also for the information sciences. *Journal of Information Science*, 28(6), 441-453.
- Park, H., Kim, K., Choi, S., & Yoon, J. (2013). A patent intelligence system for strategic technology planning. *Expert Systems with Applications*, 40(7), 2373-2390.
- Park, H., Ree, J. J., & Kim, K. (2013). Identification of promising patents for technology transfers using TRIZ evolution trends. *Expert Systems with Applications*, 40(2), 736-743.
- Park, H., Yoon, J., & Kim, K. (2011). Identifying patent infringement using SAO based semantic technological similarities. *Scientometrics*, 90(2), 515-529.
- Park, Y., Yoon, B., & Lee, S. (2005). The idiosyncrasy and dynamism of technological innovation across industries: patent citation analysis. *Technology in Society*, 27(4), 471-485.
- Pilkington, A., Lee, L. L., Chan, C. K., & Ramakrishna, S. (2009). Defining key inventors: A comparison of fuel cell and nanotechnology industries. *Technological Forecasting and Social Change*, 76(1), 118-127.
- Reitzig, M. (2004). Improving patent valuations for management purposes—validating new indicators by analyzing application rationales. *Research Policy*, 33(6), 939-957.
- Sternitzke, C., Bartkowski, A., & Schramm, R. (2008). Visualizing patent statistics by means of social network

- analysis tools. *World Patent Information*, 30(2), 115-131.
- Suh, J. H., & Park, S. C. (2009). Service-oriented technology roadmap (SoTRM) using patent map for R&D strategy of service industry. *Expert Systems with Applications*, 36(3), 6754-6772.
- Sun, Y., Lu, Y., Wang, T., Ma, H., & He, G. (2008). Pattern of patent-based environmental technology innovation in China. *Technological Forecasting and Social Change*, 75(7), 1032-1042.
- Takano, Y., Mejia, C., & Kajikawa, Y. (2016). Unconnected component inclusion technique for patent network analysis: Case study of Internet of Things-related technologies. *Journal of Informetrics*, 10(4), 967-980.
- Tong, L., Cong, H., & Lixiang, S. (2006). Automatic classification of patent documents for TRIZ users. *World Patent Information*, 28(1), 6-13.
- Tonta, Y., & Darvish, H. R. (2010). Diffusion of latent semantic analysis as a research tool: A social network analysis approach. *Journal of Informetrics*, 4(2), 166-174.
- Verhaegen, P. A., D'Hondt, J., Vertommen, J., Dewulf, S., & Duflou, J. R. (2009). Relating properties and functions from patents to TRIZ trends. *CIRP Journal of Manufacturing Science and Technology*, 1(3), 126-130.
- Wang, X., Qiu, P., Zhu, D., Mitkova, L., Lei, M., & Porter, A. L. (2015). Identification of technology development trends based on subject-action-object analysis: The case of dye-sensitized solar cells. *Technological Forecasting and Social Change*, 98, 24-46.
- Wang, X., Wang, Z., Huang, Y., Liu, Y., Zhang, J., Heng, X., & Zhu, D. (2017). Identifying R&D partners through Subject-Action-Object semantic analysis in a problem & solution pattern. *Technology Analysis & Strategic Management*, 1-14.
- White, H. D. & McCain, K. W. (1997), Visualization of literatures, *Annual review of information science and technology*, 32, 99-168.
- Yoon, B., & Park, Y. (2004). A text-mining-based patent network: Analytical tool for high-technology trend. *The Journal of High Technology Management Research*, 15(1), 37-50.
- Yoon, J., & Kim, K. (2011). Identifying rapidly evolving technological trends for R&D planning using SAO-based semantic patent networks. *Scientometrics*, 88(1), 213-228.
- Yoon, J., Park, H., & Kim, K. (2013). Identifying technological competition trends for R&D planning using dynamic patent maps: SAO-based content analysis. *Scientometrics*, 94(1), 313-331.
- Yoon, J., Park, H., Seo, W., Lee, J. M., Coh, B. Y., & Kim, J. (2015). Technology opportunity discovery (TOD) from existing technologies and products: A function-based TOD framework. *Technological Forecasting and Social Change*, 100, 153-167.
- Zhang, L. (2011). Identifying key technologies in Saskatchewan, Canada: Evidence from patent information. *World Patent Information*, 33(4), 364-370.

## Tables

**Table 1.** Previous approaches to keyword-based patent analysis

Approaches	Concept	Advantages and disadvantages	References
co-word	Co-occurrence of words in the same document is taken account to measure the degree of relationship between the words	<ul style="list-style-type: none"> <li>• easy to apply</li> <li>• measures the degree of relationships (broader sense)</li> <li>• semantic structure not considered</li> </ul>	Lee et al., 2008
n-gram	N-words that co-occur close together in a patent document is considered to measure the degree of relationship between the words	<ul style="list-style-type: none"> <li>• relatively easy to apply</li> <li>• measures the degree of relationships (closer sense)</li> <li>• semantic structure considered</li> </ul>	Gerken and Moehrle, 2012
SAO	The subject(S)-action(A)-object(O) structure of words in a sentence is analyzed to represent the functional relationship of an invention	<ul style="list-style-type: none"> <li>• complex to apply</li> <li>• measures the type of relationships</li> <li>• semantic structure considered (functional structure focused)</li> </ul>	Park et al.,2013

**Table 2.** Examples of electric vehicle patents NP-PP pairs

No	Noun phrases	Preposition phrases
1	A wireless power receiver apparatus	for powering or charging an electric vehicle
2	A hydraulic system	for a hybrid module
3	The total power consumption	of the load control system
4	A network-controlled charge transfer device	for electric vehicle
5	An electricity trading economy	of where they plug into the power grid
⋮	⋮	⋮
18949	An electrical connector	for recharging an electric vehicle

**Table 3.** Preposition analysis results

Rank	Preposition	Frequency	Ratio
1	Of	9243	48.78%
2	For	2552	13.47%
3	To	1593	8.41%
4	In	1352	7.13%
5	With	839	4.43%
6	From	789	4.16%
7	On	351	1.85%
8	As	321	1.69%
9	By	270	1.42%
10	At	216	1.14%
11	During	96	0.51%
12	Into	96	0.51%
13	Through	89	0.47%
14	Within	80	0.42%
15	Via	51	0.27%
⋮	⋮	⋮	⋮
59	Whether	1	0.01%
Total		18949	100.00%

**Table 4.** Types of relationships and their corresponding prepositions

Relationship	Corresponding	Ratio	Meaning	Examples
--------------	---------------	-------	---------	----------

	prepositions			Tech-keyword (T1)	Tech-keyword (T2)
<b>Inclusion (utilization)</b>	of, in, with, from, on, at, within	67.9%	T2 includes (utilizes) T1	Torque	Motor
<b>Objective</b>	for	13.5%	T1 is an objective for T2	Power supply	Motor
<b>Effect</b>	to	8.4%	T1 has an effect on T2	Electric power	Pump
<b>Process</b>	during, into, through, via	1.8%	T1 is a part of T2 process	Engine speed	Auto-start
<b>Likeness</b>	as	1.7%	T1 is similar to T2	Electric vehicle	Hybrid vehicle

**Table 5.** Verification analysis results

Relations	Preposition	Correct relations	Sample accuracy	Population accuracy (Min, Max)		Acceptability
<b>Inclusion (utilization)</b>	Of	50	1.00	0.86	1.00	O
	From	46	0.92	0.88	0.96	O
	In	49	0.98	0.96	1.00	O
	At	8	0.16	0.11	0.21	X
	On	43	0.86	0.81	0.91	O
	With	50	1.00	0.86	1.00	O
<b>Objective</b>	Within	42	0.84	0.79	0.89	O
<b>Effect</b>	For	44	0.88	0.83	0.93	O
<b>Process</b>	To	37	0.74	0.68	0.80	O
	During	46	0.92	0.88	0.96	O
	Via	45	0.90	0.86	0.94	O
	Into	41	0.82	0.77	0.87	O
<b>Likeness</b>	Through	40	0.80	0.74	0.86	O
	As	37	0.74	0.68	0.80	O
<b>Total</b>		578	0.83	-		O

**Table 6.** A list of technological keywords for an electric vehicle

Keywords (1~108)
ac motor, ac power, accelerator, air conditioner, battery, battery capacity, battery cell, battery charger, battery management system, battery module, battery pack, battery state, battery system, body frame, bracket, charging connector, charging device, clutch, combustion engine, control system, control unit, controller, converter, coolant, cooling fan, cooling system, creep torque, cruise control system, CVT, dc power, dc converter, drive motor, drive shaft, drive system, drive unit, electric machine, electric motor, electric power, electric vehicle, electric vehicle drive system, electrode, embodiment, energy consumption, energy storage device, engine, engine clutch, engine power, engine speed, evse, fuel, fuel cell, fuel efficiency, fuel tank, gear, generator, heat exchanger, high voltage battery, hybrid, hybrid powertrain, hybrid vehicle, hybrid vehicle drive system, hydrogen, interface, inverter, kiosk, lithium ion battery, low voltage battery, motor, motor controller, motor drive system, network, power consumption, power control system, power grid, power module, power source, power supply, powertrain, pump, rechargeable battery, resistor, rotor, seat, secondary battery, shaft, solar energy, state detector, steering system, storage unit, system controller, torque, torque control, traction battery, traction motor, transmission, unit cell, user input, vehicle, vehicle speed, voltage battery, voltage detector, voltage value, wheel, wheel torque, wind energy, winding, wireless communication, wireless power transfer

**Table 7.** The top 10 technological pairs with high structural similarity values

Rank	Technological pairs	Content similarity	Structural similarity
1	B60W-0010/08 & Y02T-0010/6286	0.975	0.571
2	B60W-0010/06 & Y02T-0010/6286	0.978	0.570
3	Y02T-0090/121 & Y02T-0090/128	0.987	0.568
4	B60W-0020/00 & Y02T-0010/6286	0.960	0.553
5	B60W-0020/00 & B60W-0010/06	0.952	0.549
6	B60W-0010/08 & B60W-0010/06	0.962	0.548
7	Y02T-0090/14 & Y02T-0010/7072	0.990	0.544
8	Y02T-0090/121 & Y02T-0010/163	0.978	0.542
9	B60W-0010/08 & B60W-0020/00	0.970	0.541
10	Y02T-0010/7005 & Y02T-0010/705	0.938	0.540

**Table 8.** Idea for developing technology Y02T-0010/6286

Technology	Base keywords	Relation keywords	Companies
B60W-0010/06	motor	motor drive system(inclusion)	Tesla Motors
	vehicle	embodiment(influence)	Ford Global Technologies
	vehicle	hybrid(condition)	GM Global Technology
	⋮	⋮	⋮
B60W-0010/08	hybrid vehicle	power supply(influence)	Toyota
	vehicle	transmission(application)	Hyundai Motor
	electric power	fuel(inclusion)	Continental Automotive
	⋮	⋮	⋮
B60W-0020/00	torque	generator(inclusion)	Suzuki Motor
	engine clutch	hybrid(inclusion)	Hyundai Motor
	electric motor	battery charger(inclusion)	Fuji Heavy Industries
	⋮	⋮	⋮

**Table 9.** SAO analysis results

Rank	Action	Frequency	Ratio
1	include	341	30.26%
2	configure	77	6.83%
3	have	70	6.21%
4	be	60	5.32%
5	provide	53	4.70%
6	connect	41	3.64%
7	mount	36	3.19%
8	comprise	35	3.11%
9	use	26	2.31%
10	operate	20	1.77%
⋮	⋮	⋮	⋮
97	prevent	1	0.09%
	Total	1127	100%

Figures

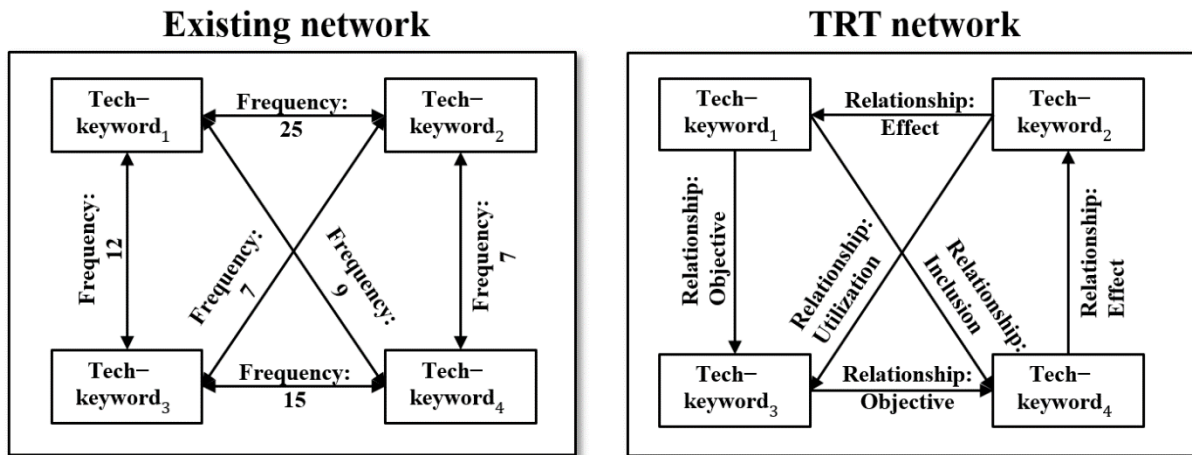


Figure 1. Technological-relationship based keyword network structure

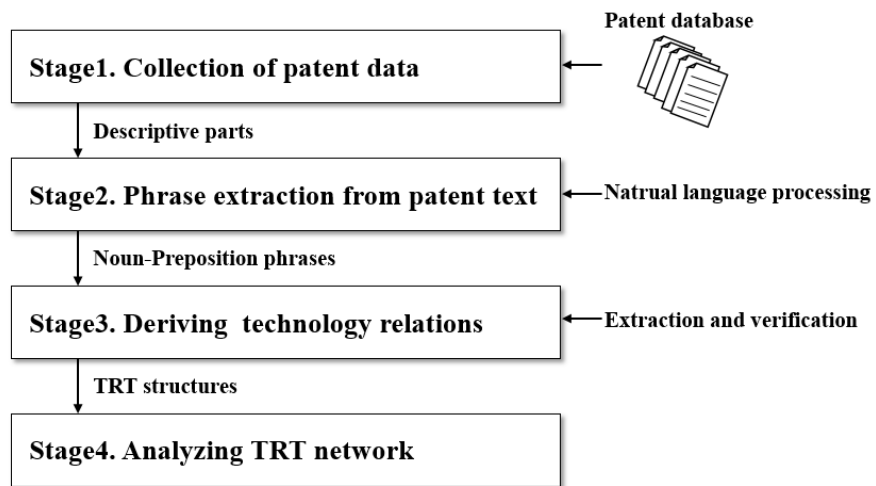


Figure 2. Overall development process of a TRT network

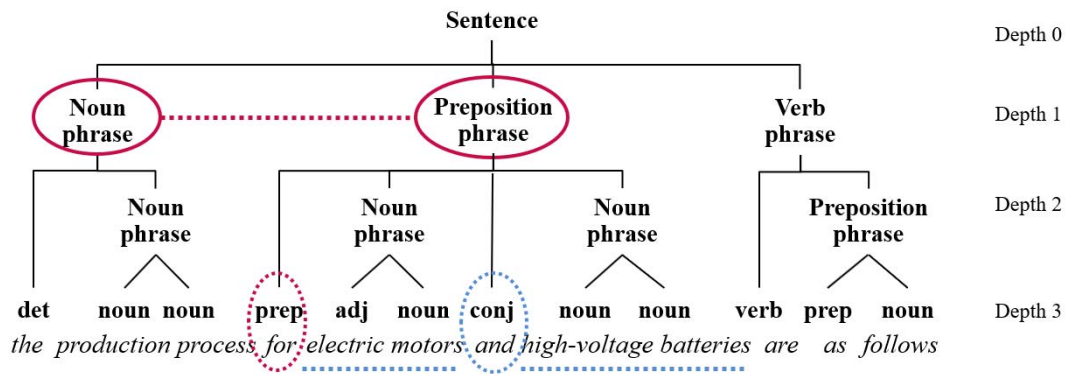


Figure 3. An example of sentence tree to extract noun-preposition phrases

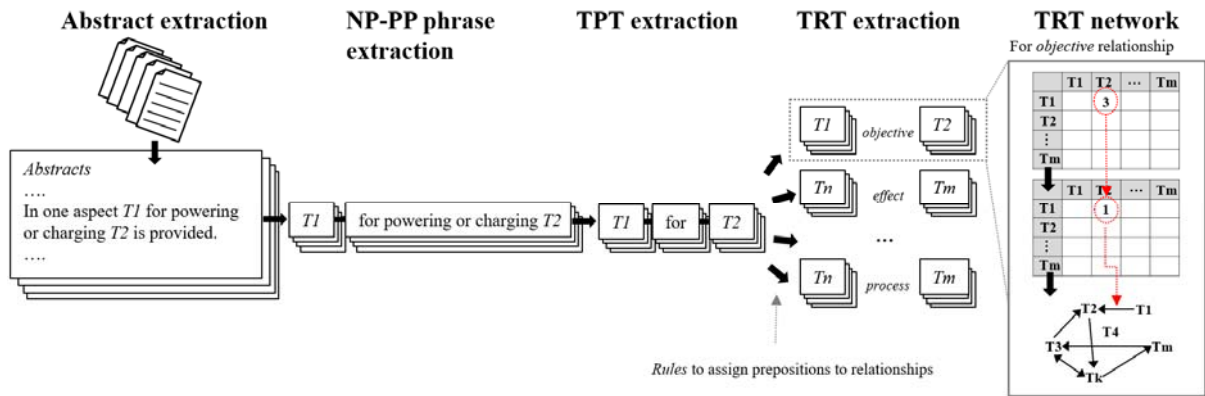


Figure 4. Detailed procedures of developing a TRT network

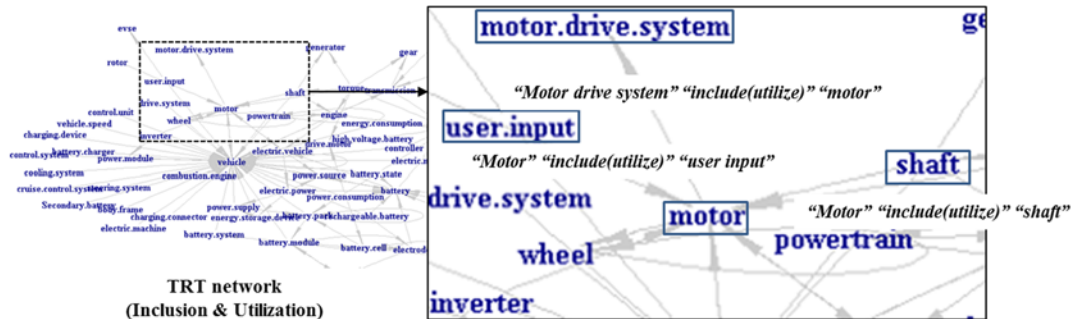


Figure 5. TRT network for inclusion (utilization) relationships: an electric vehicle

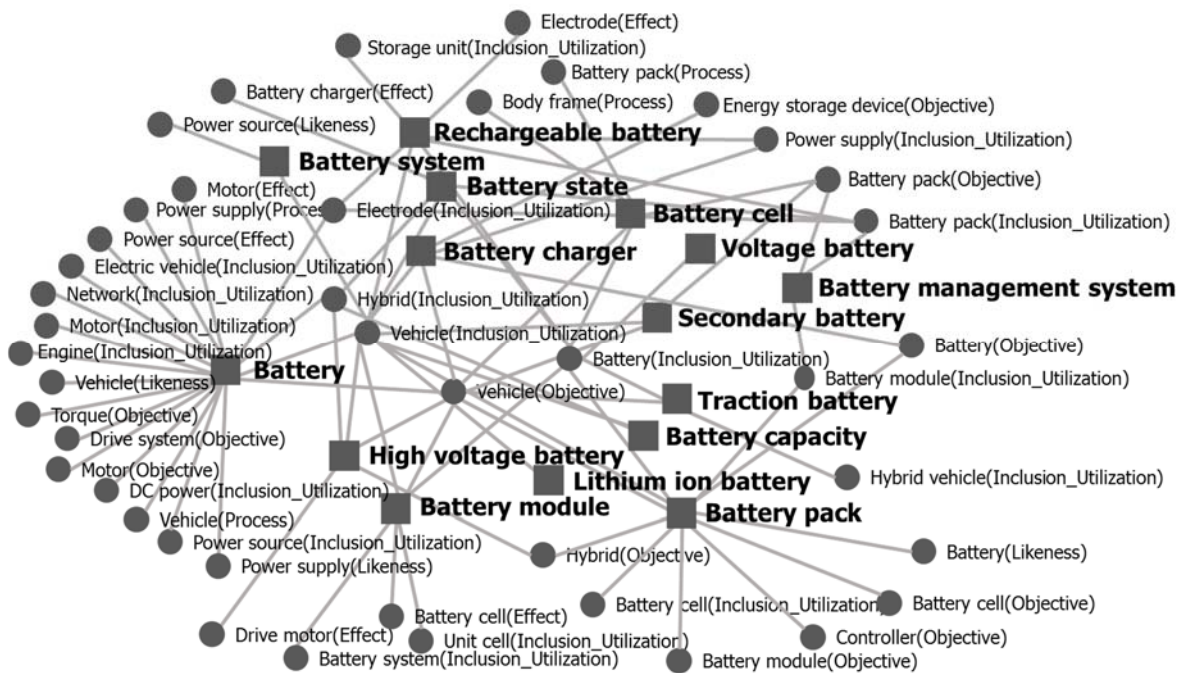


Figure 6. TRT network visualization of electric vehicle battery technology





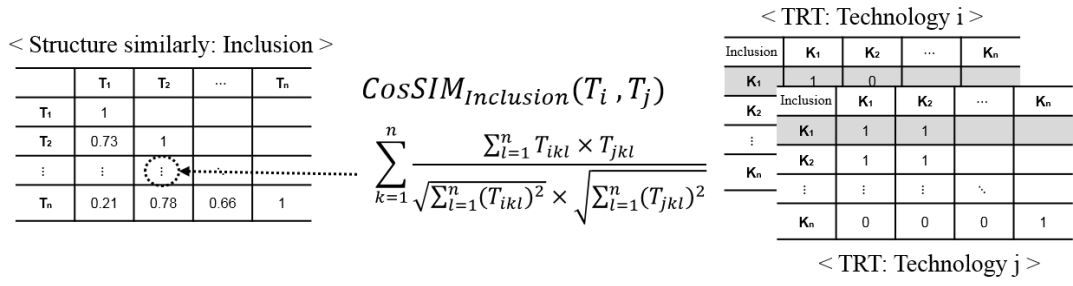


Figure 9. Structural similarity analysis process

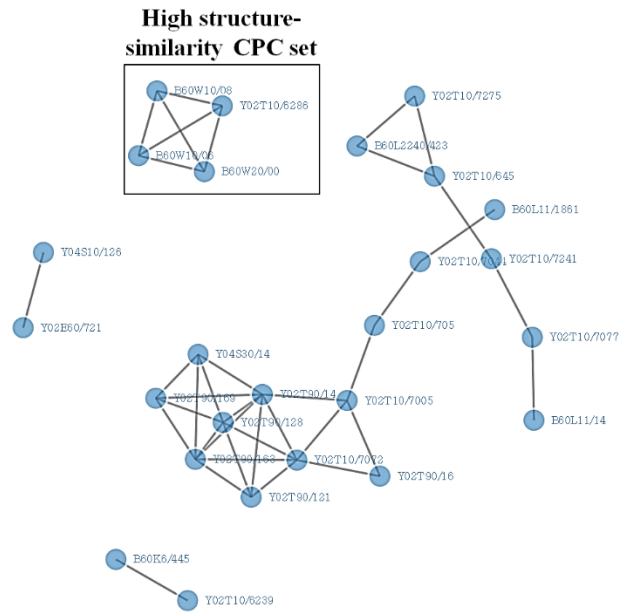


Figure 10. Clusters of technologies with similar structures

## Appendix

Noun_Phrase	Preposition_Phrase	Tech-keywords	Preposition	Tech-keywords
that connects	to an power structure for pow	connects	to	power structure
an power structure	for powering	power structure	for	power structure
the vehicle	to the power structure	vehicle	to	power structure
system , method and apparat	for wirelessly transferring pow	system , method and	for	power
a wireless power receiver app	for powering or charging an	wireless power receiv	for	vehicle
a receive circuit	including a first coil	receive circuit	for	coil
a passive circuit	including a second coil	passive circuit	for	coil
receive power	from a transmit circuit	receive power	from	circuit
the second coil	from the first coil	second coil	from	coil
A hydraulic system	for a hybrid module which is	hydraulic system	for	hybrid module
a parallel arrangement	of a mechanical pump and ar	parallel arrangement	of	pump
oil	to other portion of the hydrat	oil	to	portion
other portion	of the hydraulic system depen	other portion	of	system
condensate	from a charge air cooler	condensate	from	charge air cooler
RPM and airflow	through a charge air cooler	RPM and airflow	through	charge air cooler
condensate	to the engine intake	condensate	to	engine intake
ingestion	of water	ingestion	of	water
wireless receiver	for use with a first electromag	wireless receiver	for	use
use	with a first electromagnetic re	use	with	resonator
The wireless receiver	including a load	wireless receiver	with	load
the drive system	of a vehide using electrical p	drive system	of	vehide
a safety system	for to provide protection	safety system	for	protection
respect	to an object that may becom	respect	to	object
operation	of the first electromagnetic re	operation	of	resonator
The safety system	including a detection substs	safety system	of	detection subsystem

Figure A1. Extraction technology-preposition keywords from Noun-Preposition phrases

TRT_num	Tech-keyword	Relationship	Tech-keyword
1	connects	Effect	power structure
2	power structure	Objective	power structure
3	vehide	Effect	power structure
4	system , method and	Objective	power
5	wireless power transf	Objective	vehide
6	receive circuit	Objective	coil
7	passive circuit	Objective	coil
8	receive power	Inclusion_Utilization	circuit
9	second coil	Inclusion_Utilization	coil
10	hydraulic system	Objective	hybrid module
11	parallel arrangement	Inclusion_Utilization	pump
12	oil	Effect	portion
13	other portion	Inclusion_Utilization	system
14	condensate	Inclusion_Utilization	charge air cooler
15	RPM and airflow	Process	charge air cooler
16	condensate	Effect	engine intake
17	ingestion	Inclusion_Utilization	water
18	wireless power transf	Objective	use
19	use	Inclusion_Utilization	resonator
20	wireless power transf	Inclusion_Utilization	load
21	drive system	Inclusion_Utilization	vehide
22	safety system	Objective	protection
:	:	:	:
17702	accordance	Inclusion_Utilization	method

TRT_num	Tech-keyword	Relationship	Tech-keyword
1	wireless power transf	Objective	vehide
2	drive system	Inclusion_Utilization	vehide
3	battery charger	Objective	vehide
4	power consumption	Inclusion_Utilization	control system
5	drive system	Inclusion_Utilization	vehide
6	drive system	Inclusion_Utilization	vehide
7	power source	Inclusion_Utilization	vehide
8	battery	Process	vehide
9	hybrid powertrain	Objective	vehide
10	hybrid powertrain	Objective	vehide
11	power grid	Likeness	vehide
12	transmission	Objective	hybrid
13	shaft	Inclusion_Utilization	transmission
14	energy storage device	Likeness	vehide
15	engine	Inclusion_Utilization	vehide
16	battery	Process	vehide
17	energy storage device	Inclusion_Utilization	vehide
18	controller	Inclusion_Utilization	vehide
19	battery	Inclusion_Utilization	vehide
20	powertrain	Inclusion_Utilization	vehide
21	electric vehicle	Effect	power grid
22	electric vehicle	Effect	power source
:	:	:	:
1179	power module	Inclusion_Utilization	vehide

Figure A2. Target (electric vehicle technology) TRT Extraction from total TRT



**Table A1.** Electric vehicle patents TPT structure examples

No	T (Technological keyword)	P (Preposition)	T (Technological keyword)
1	Wireless power receiver apparatus	for	Powering electric vehicle Charging electric vehicle
2	Hydraulic system	for	Hybrid module
3	Power consumption	of	Load control system
4	Network-controlled charge transfer device	for	Electric vehicle
5	Electricity trading economy	of	Power grid
⋮	⋮	⋮	⋮
18,949	Electrical connector	for	Recharging electric vehicle

**Table A2.** CPC descriptions for the top 10 technology pairs with high structural similarity values

CPC	CPC descriptions
B60W-0010/06	Conjoint control of vehicle sub-units of different type or different function including control of combustion engines
B60W-0010/08	Conjoint control of vehicle sub-units of different type or different function including control of electric propulsion units, e.g. motors or generators
B60W-0020/00	Control systems specially adapted for hybrid vehicles
Y02T-0010/163	Turbocompound engines
Y02T-0010/6286	Control systems for power distribution between ICE and other motor or motors
Y02T-0010/7005	Batteries
Y02T-0010/705	Controlling vehicles with one battery or one capacitor only
Y02T-0010/7072	Electromobility specific charging systems or methods for batteries, ultracapacitors, supercapacitors or double-layer capacitors
Y02T-0090/121	Enabling technologies or technologies with a potential or indirect contribution to GHG emissions mitigation by conductive energy transmission
Y02T-0090/128	Energy exchange control or determination
Y02T-0090/14	Plug-in electric vehicles