



HAL
open science

Extraction and Modeling of Chinese Patent Information for Technical Advancement Evaluation

Yin-Di Sun, Guo-Zhong Cao, Chang Gao, Wen-Dan Yang, Wei-Pei Han, Kang
Wang

► **To cite this version:**

Yin-Di Sun, Guo-Zhong Cao, Chang Gao, Wen-Dan Yang, Wei-Pei Han, et al.. Extraction and Modeling of Chinese Patent Information for Technical Advancement Evaluation. TRIZ Future conference, Sep 2021, Bolzano, Italy. pp.127-140, 10.1007/978-3-030-86614-3_10 . hal-04067825

HAL Id: hal-04067825

<https://inria.hal.science/hal-04067825>

Submitted on 13 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.




Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

Extraction and Modeling of Chinese Patent Information for Technical Advancement Evaluation

Yin-Di Sun¹ ² [0000-0002-3225-1213], Guo-Zhong CAO^{1,2} [0000-0002-2939-5566], Chang Gao^{1,2},
Wen-Dan Yang^{1,2}, Wei-Pei Han^{1,2}, and Kang Wang^{1,2} [0000-0002-7881-1060]

¹ Hebei University of Technology, Tianjin 300401, China

yindisun@foxmail.com, caoguozhong@hebut.edu.cn

² National Engineering Research Center for Technological Innovation Method and Tool, Tianjin 300401, China

Abstract. In order to achieve the output of high-value patents, enterprises need to consider improving the technical level of the technical solution itself. At this point, it is necessary to compare the technical scheme with the existing patented technology to determine the advanced degree of the technical scheme. The ideal level is the ratio of the useful function to the harmful function in the technical solution, which can be used to describe the current technical level of the technical solution. In this study, idealization level is used to evaluate the gap between technology and existing patented technology. The patent information is different from the technical solution with complete information, and the existing technology described in the patent has a specific form of expression. Therefore, the main task is to evaluate the idealization level of the existing technology. It mainly includes three stages: shallow information analysis, deep information analysis and information reasoning and calculation. Automatic extraction of Chinese patent information is the basis for evaluating the technical advancement. This paper introduces in detail the automatic establishment process of component hierarchy model (CHM) and action and attribute model (AAM) of Chinese patent in the process of deep information analysis. The feasibility of the proposed method is verified by the analysis of shallow and deep information of electric toothbrush.

Keywords: Technical Advancement, Idealization Level, Patent Information Extraction, Chinese Patent.

1 Introduction

As technological innovation plays an increasingly important role in today's knowledge economy, patent strategies play an important role in gaining competitive advantages for innovative entities [1]. Patent layout is an organic combination of corporate industry, market, and legal factors, and patents are organically combined, covering the time, region, technology, and product dimensions related to the company's interests, building a strict and efficient patent protection network, and ultimately forming a favorable structure for the company Patent portfolio. Patent layout is a manifestation of patent offensive and defensive strategies, and a process of patent portfolio with a certain sense

and purpose to achieve the overall strategic goals of an enterprise [2]. Product development and patent strategy are inseparable. Enterprises should expand the scope of patent protection as far as possible to protect the property rights of their products to maintain a certain competitive strength in the field they are good at. From the perspective of patent strategy, it is necessary for companies to evaluate the technical value of their products in real time during the research and development process to determine the design goals of their products during the research and development process.

The technical value of patents is a necessary condition for the realization of patent value. High-value patents must first reach a certain height in technology, and technological value is the basis for measuring the value of patents. Therefore, in the process of patent-oriented portfolio design, we must first take the evaluation criteria of the value of patented technology as design constraints. Technological advancement is an important indicator for evaluating whether technology occupies a leading position. In the process of designing for patentability, previous studies have considered two aspects: On the one hand, whether the design plan can meet the three characteristics of the patent (Practicality, novelty, and creativity) [3, 4]; on the other hand, does the design plan cover comprehensively, in terms of implementation means and effects, whether there are any gaps to continue innovative research [5]. And in the design process, especially in the current era of exponential growth in the output of innovation results, It is increasingly difficult for engineering designers to evaluate the gap between their own research and development results and existing technologies. Therefore, it is necessary to propose a method that can quickly help engineers determine the current level of technology.

The first step in evaluating the state of the art is to understand the existing technology. However, it usually takes a long time for even domain experts to read and understand patent documents. Therefore, automatic extraction technology of patent information should be introduced to assist engineers to extract the technical content of patents. Earlier, Cassini used the semantic processing method of Sao to realize automatic function analysis [6, 7] and TRIZ conflict [8]. With the development of natural language processing (NLP), the depth of automatic extraction of patent information is gradually deepening. Fantoni et al. [9] realized automatic extraction of functions, physical behaviors, and states in English patents through SAO (subject-action-object) tagging, building knowledge base and natural language processing tools. The extracted functions are based on the user's interpretation of physical behaviors; Chiarello et al. [10] proposed a method to automatically extract the advantages and disadvantages of patent technology from patent text.

In recent years, research on the extraction of patent information is still a hot topic. The first task of extracting patent information is to obtain the text content of the patent. The preprocessing of patent texts is the basis for in-depth analysis of patent information, Gaurav Gupta et al. [11] proposed a web crawling algorithm based on data collection, can help users quickly obtain patent data; Naima Vahab et al. [12] use NLP and rule integration methods, help users extract important patent information. Furthermore, patent documents generally have a fixed structure and form, Masayuki Okamoto et al. [13] proposed a patent claim structure extraction method based on information extraction, reduce the workload of analysis. Deeper patent information extraction, keyword

extraction in patents can complete patent classification [14, 15], patent technology trend analysis [16] and other work. Extraction of technical information for patents, often requires deep semantic analysis, including the extraction of subject-behavior-object (SAO) information in patents [17], the extraction of functional information [18] and the use of semantic analysis to graphically express patent technology information [19]. In summary, the acquisition of patent technical information generally requires the acquisition of patent text, extraction of structured information from patents and semantic analysis of patent texts.

In 2020, World Intellectual Property Organization(WIPO)released the World Intellectual Property Indicators 2020 [20]. The National Intellectual Property Administration of the People's Republic of China (CNIPA) received 1.4 million patent applications in 2019. This is more than twice the amount received by the United States Patent and Trademark Office (USPTO), the document showed. China is set to become the top filer of international patent applications; with the highest number of patent applications, it has received in 2019. In addition, China moved up from third position in 2009 to claim the top spot in 2011 and has continued to head the ranking for the past nine years. Several statistics show that China's patent activities are increasingly active and playing an important role in the global innovation landscape.

Therefore, the importance of Chinese patents cannot be ignored, it is of great significance for the research on the automatic extraction of Chinese patent technical information. Simultaneously, Chinese companies also urgently need to use automated means to help them evaluate their technical solutions.

For assessing the strength of an enterprise, the evaluation of patent value is aimed at economic value. including patent intangible asset value[21], patent private value[22] and market value[23], etc., the cost method and income method are commonly used for evaluation [24]. In the current process of evaluating a patent, the patent's legal status, technical level, market conditions, finance, strategy and other aspects are usually used to evaluate the patent [25]. The generally recognized patent value system in China divides the dimensions of patent value into technical value, legal value and economic value. From another point of view, the technological superiority of patents is the basis for the realization of the economic value of patents. Therefore, when innovation entities want to increase the overall value of patents, it is necessary to consider and improve the technical value of its technical solutions.

This article aims to provide a method to help engineering designers evaluate the technological advancement of their products during the design process. On the one hand is the proposal of a way to assess the value of technological advancement; On the other hand, as a basis for evaluation, propose a Chinese patent information extraction method for technical advancement value evaluation. The main structure of this article is as follows: Section 2 introduces the evaluation methods of technological advancement value; Section 3 presents the automatic extraction method of Chinese patent information; Section 4 shows the high-value patent preliminary evaluation in the field of electric toothbrushes and the extraction of a piece of Chinese patent information; The section 5 discusses the outlook for future work

2 Patent Technical Advancement

2.1 Patent Value Index System and Technical Advancement

According to different purposes, the evaluation indexes and classification of patent value system will be different. China Intellectual Property Office developed and revised the index system of patent value analysis, and put forward the patent value degree that represents the value of the patent itself, which can be used as a unified standard for reference. Technology value is the basis of patent value system, and its indexes include technical advancement, technology development trend, application scope, irreplaceability and enforceability. They are defined in Table 1.

Table 1. The primary index definition of technical value of patent

Index of technical value	Definition
Technical advancement	Whether the patented technology is in a leading position at the time of analysis compared to other technologies in the field
Technology development trend	The current development trend of the technical field in which the patented technology is located
Application scope	Application scope of patented technology
Irreplaceability	Whether an alternative solution to the same or similar problem currently exists
Enforceability	Technology reflects the future technology application prospect of patent and the realization degree of technology achievement transformation

In the design and development phase, with the aim of improving the value of patent technology, the technical advancement is the degree to solve the key technical problems and produce better results than the current technology. For example, in the technical effect to improve production efficiency, save costs, improve structure, improve quality, improve productivity, reduce pollution and so on. From the point of engineering design, the technical advancement T_{AD} is the difference between the design level and the existing technology level. The level of technology can be expressed as an idealized level. Thus, the quantification of technological advancement can be expressed in Equation (1), where I represents the idealized level of the design scheme and I_{ET} represents the idealized level of the prior art.

$$T_{AD} = I - I_{ET} \quad (1)$$

In TRIZ, the ideal level is the ratio of useful function to harmful function. According to the prior art recorded in the patent literature, the ideal level formula can be written as Equation (2) by expanding the useful function and harmful function. F_{IM} is the importance of sub-function; P_F is the realization degree of sub-function; C_S represents the cost evaluation value of components; H_S stands for the evaluation value of harmful functions produced by components.

$$I_{ET} = \frac{\sum F_{IM} \cdot P_F}{\sum C_S + \sum H_S} \quad (2)$$

2.2 Evaluation of Highest Ideal Level of Patent Prior Art

It is easy to get F_{IM} , P_F , C_S and H_S when evaluating the idealized level of the enterprise's technical scheme. However, the technical information obtained from patents is often less than that obtained from technical solutions. In essence, the main task of evaluating the value of technological advancement is to evaluate the highest ideal level of the existing technology. A large number of patent documents are often obtained when searching for a certain technology. In order to reduce unnecessary work, it is necessary to analyze patent documents from shallow to deep. As shown in Fig. 1, the analysis of the most patented text is divided into three stages, namely, shallow information analysis, deep information analysis, and information reasoning and calculation. This will be covered in the next three sections.

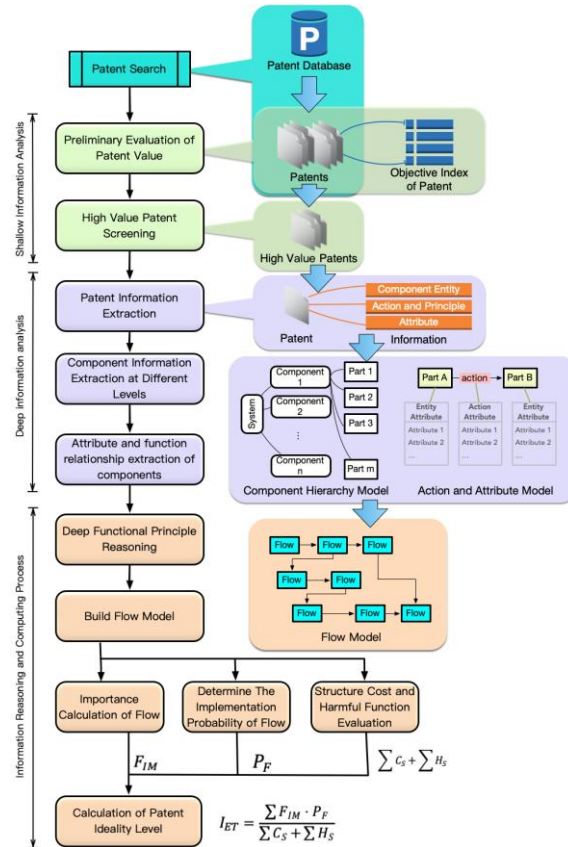


Fig. 1. Evaluation process of existing technology idealization level

2.2.1 Shallow Information Analysis.

Generally, the evaluation methods of patent value can be divided into objective quantitative evaluation and subjective qualitative evaluation based on expert evaluation. The evaluation of technical advancement is to find the best technology with the highest level of idealization. It is not necessary to evaluate all the patents retrieved. Therefore, for a large number of Patent Texts, we can first use the objective index of patent to evaluate and screen out patents with certain value from the perspective of objective index, so as to prepare for deep information processing.

2.2.2 Deep information analysis.

Despite the speed of processing, it is often inaccurate to evaluate the technical advancements of a patent merely by processing the text and using objective indicators to determine the highest level of technology, which requires a certain understanding of the technology described in the patent. The initial idea that we put forward is to discuss the information of a patent at three levels, It includes Component Hierarchy Model (CHM), Action and Attribute Model (AAM) and Flow Model.

CHM is a model describing different levels of information of patent technology. In the patent specification, it is usually stated which parts an assembly includes. Furthermore, it also describes which features are made up of in the part. CHM can help engineers understand the information of entity components in different levels of technology. AAM is a model that describes the action relationship among the entity components, the component attributes and the action attributes in the patent technology. In general, there are component entity information, action and principle information, and attribute information in a patent. Entity component is the carrier of function realization and the concrete structure. In patents, solid components generally appear as proper terms, such as "main shell, suspension magnet motor, etc.". Action is the relationship between physical components. A description of the relationship between two or more entity elements in a patent is usually behavioral information. For example, "Fixed magnets and suspended magnets repel each other."

Attribute is information that describes the characteristics of an entity's components and behavior. It includes: effect, location, time, state, material, shape. Their detailed description is shown in Table 2.

Table 2. Classification and description of attribute information

Attribute	Description
Effect	Specific functions that can be performed by physical components or behaviors, e.g., "cleaning components play the role of cleaning."
Location	The relative position relationship between the component entity and others, e.g., "the main shell is located on the outermost side of the whole unit."
Time	The time when the action occurs or the current action is in a certain state in the working cycle, e.g., "in the initial state of the device, the contact conductor contacts with the conductor sheet."

State	The actual state of a component entity or action, for example: "LED light on, rotating brush head rotating."
Material	Material characteristics used for solid components, e.g., "the main shell is made of waterproof material".
Shape	Shape characteristics of solid components, e.g., "circular particle on handle surface".

In the section 3, the extraction method and process of CHM and AAM in Chinese patent will be described in detail.

2.2.3 Information Reasoning and Computing Process.

The sub-functions of technology are rarely described completely in patents. Even if it is assumed that the sub-functions can be described completely, it is difficult to obtain the degree of realization from patents. Therefore, according to the limited information available in the patent, we can infer from the flow model.

The quantity (or entity) of input and output is called flow. Flow is the basic object to be considered in any design problem. It contains the key physical information in the product model[26]. Flow model can describe the deepest information in patent. The process of constructing flow model requires engineers to infer the AAM with effect library, analyze the principle of each action and the changing physical quantity. The flow model is constructed according to the logic of each action. In this case, the sub-function importance F_{IM} and the sub-function realization degree P_F are equivalent to the flow importance and its possible realization degree.

With the increase of distance, the influence of sub-function on its target or other functions decreases, and the decrease follows the law of exponential decay[27]. Each flow in the flow model is regarded as a node in graph theory.

The shortest distance algorithm can be used to calculate the distance of each flow to the final output flow of the technical system, and the influence degree of each flow can be obtained, that is, the importance degree of the flow F_{IM} . The realization degree of the flow P_F can be evaluated by combining with the simulation technology, and the probability of the flow meeting the predetermined demand can be obtained. The cost $\sum C_s$ can be evaluated according to the materials used and the processing method of the components. In addition, the assembly difficulty between the components should also be considered as the additional cost. The evaluation value $\sum H_s$ of harmful functions generated by components mainly focuses on the harmful effects of components on the internal and external of the technical system.

Current technology should be evaluated in the same way. For the developed technical solutions, it is easier to build the flow model by mastering the complete information of the action principle, and it is also easier to evaluate the cost of the solutions and the harmful effects of the components.

3 Automatic Extraction and Analysis of Chinese Patent Information

In the stage of deep information analysis of patents, it will take a lot of time for engineers to extract the level information and function principle information of patents. At the same time, people will inevitably have some subjective understanding of patent information in the process of patent extraction, which will have a certain impact on the comparison of multiple patents. In the process of prior art evaluation as shown in Fig.1, the automatic extraction of patent information mainly needs to help engineers to complete the establishment of CHM and AAM, to provide the basis for subsequent flow model reasoning.

In Section 1, it is mentioned that Chinese patent information can not be ignored, but there are still big limitations in computer's understanding of Chinese. Therefore, the traditional English patent information extraction methods and algorithms can not be fully applicable to Chinese patents. Therefore, according to the language characteristics of patent documents, this paper proposes a way to extract Chinese patent filing information. In Section 3.1 and Section 3.2, patent information extraction methods of CHM and AAM are introduced.

3.1 Automatic Extraction of CHM

Chinese patents meet certain standards in the process of writing. In a patent specification, the description of a technical solution usually begins with a declaration of the components at each component level. As shown in Table 3, there are two ways to declare the component hierarchy: one is an open style, for example, "the electric toothbrush includes support component, cleaning component, auxiliary drive component and main drive component"; The other is the enclosed. For example, "the electric toothbrush is composed of support component, cleaning component, auxiliary transmission component and main transmission component". The implication of the open writing method is that in addition to the claimed components, the additional components on this basis are still within the protection scope of this patent. Although most people tend to use the open writing method in order to expand the scope of protection when writing patents, all possible situations should be considered in the process of automatic extraction.

Table 3. Component hierarchy writing form in Chinese patent

Writing method	General form	Examples
Open type	... includes ..., ...and ...	The electric toothbrush includes support component, cleaning component, auxiliary drive component and main drive component.
Enclosed	...is composed of..., ...and ...	The electric toothbrush is composed of support assembly, cleaning assembly, auxiliary drive assembly and main drive assembly.

The standard of patent writing can provide some convenience for the extraction of patent information. The specification of a Chinese patent includes five parts: the technical field, the background technology, the content of the invention, the attached drawings and the specific method of implementation. The specific automatic patent extraction process is shown in Fig. 2. The technical details of the patent are generally shown in the content of the invention and the specific implementation method. Firstly, the patent text is segmented by Python to facilitate subsequent processing. According to the component level writing rules of the patent, regular expressions are used to extract components and their corresponding parts in each section respectively. Finally, a visual view is constructed according to the corresponding relationship between Component and Part for engineers' reference.

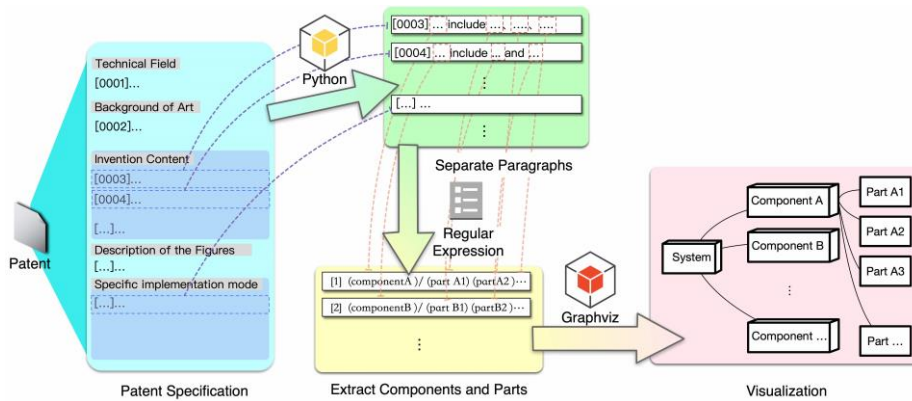


Fig. 2. Automatic building process of CHM

3.2 Automatic Extraction of AAM

The extraction of AAM requires a deep understanding of the statements in the patent. Currently, there are many open-source algorithms for deep semantic analysis of Chinese text, such as JIEBA, SNOWNLP, THU Lexical Analyzer for Chinese (THULAC), Language Technology Platform (LTP), etc.

LTP is an open Chinese natural language processing system developed by social computing and information retrieval research center of Harbin Institute of technology (HIT) [28]. Besides word segmentation (WS) and part of speech (POS) tagging, LTP can also process semantic role labeling (SRL) and semantic dependency parsing (SDP). Among them, SDP can get rid of the constraints of middle and surface syntax and obtain deep semantics directly. This means that in the face of difficult to analyze the extraction of Chinese patent information, we can not only extract simple subject action object (SAO) information, but also understand the actual meaning described in Chinese patent and extract the corresponding attributes of SAO. The label schematic of SDP can be referred to the literature [29]. As shown in Fig. 3, the automatic construction of AAM is divided into two parts:

The first part is to analyze the Chinese semantics of the statements in the Chinese patent through LTP: using WS, POS Tagging and SDP in the LTP to process the Chinese patent statements. The semantic dependency of a single sentence will be obtained.

The second part is to extract the subject element, object element, function and their attribute information by using semantic dependency. In Fig.3, the relationship between subject role, object role, situational role and entity component information and attribute information in LTP is preliminarily marked. In some Chinese sentences, it is not easy to find accurate semantic peripheral roles directly, but also need to use semantic structure relations and semantic attachment markers to help find part, action and attribute.

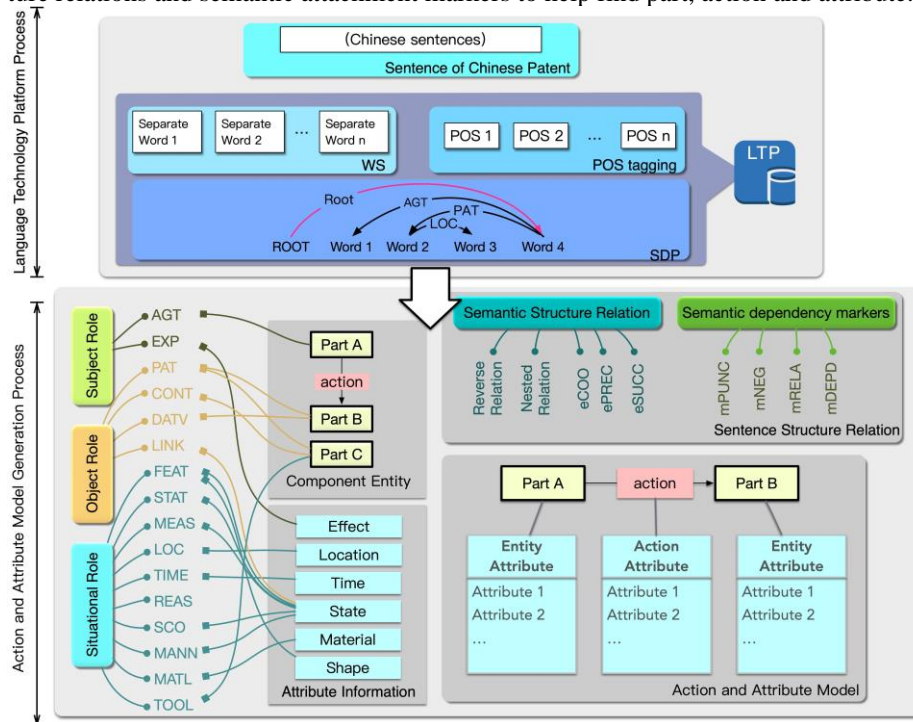


Fig. 3. Automatic building process of AAM

4 Case Study

In this section, the CHM and AAM of a Chinese patent (CN111713846B) are automatically constructed to show the process of deep extraction of patent information. Then, according to the general rules of patent writing shown in Table 3, combined with regular expressions, the component and parts of the technical solution are extracted respectively. Finally, as shown in Fig. 4, CHM is visualized by Graphviz [30] using the corresponding relationship between component and parts.

4.1 CHM Automatic Extraction Process

According to the flow given in Section 3.1, as shown in Figure 1, firstly, Python is used to extract the contents of the invention and the specific implementation methods in the patent.

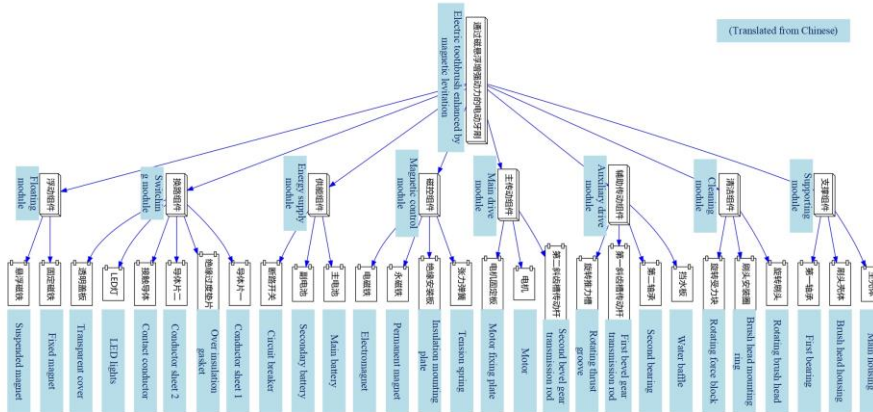


Fig. 4. Visualization of component hierarchy information using Graphviz

4.2 AAM Extraction Process

Due to space constraints, we will not analyze all the sentences in this article. Take a sentence as an example to show the process of building AAM. The processing results of semantic dependency analysis using LTP are shown in Table 4, including WS, POS and SDP. After visualization of the processing results, as shown in Fig.5, there are different semantic dependency labels among words, which represent different semantic meanings. Finally, according to the results of SDP, entity, action and attribute are extracted to construct AAM, as shown in Figure 6.

Table 4. The SDP analysis process using LTP

Steps	Result
Sentence	The pressing block extrudes the permanent magnet towards the conductor by the arc surface.(Translated from Chinese)
WS	['Pressing block', 'uses', 'arc surface', 'make', 'permanent magnet', 'towards', 'conductor', 'direction', 'extrudes', '.'] (Translated from Chinese)
POS	['v', 'v', 'n', 'p', 'n', 'p', 'n', 'n', 'v', 'wp']
SDP	[(1, 2, 'AGT'), (2, 0, 'Root'), (3, 2, 'CONT'), (4, 5, 'mRELA'), (5, 9, 'PAT'), (6, 8, 'mRELA'), (7, 8, 'FEAT'), (8, 9, 'LOC'), (9, 2, 'eSUCC'), (10, 9, 'mPUNC')]

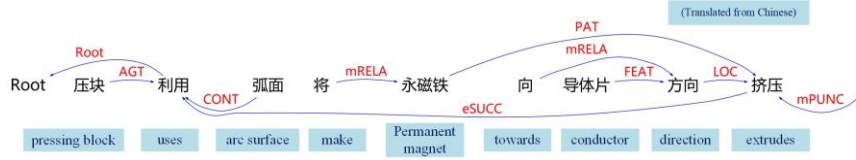


Fig. 5. Semantic dependency relationship of Chinese sentences

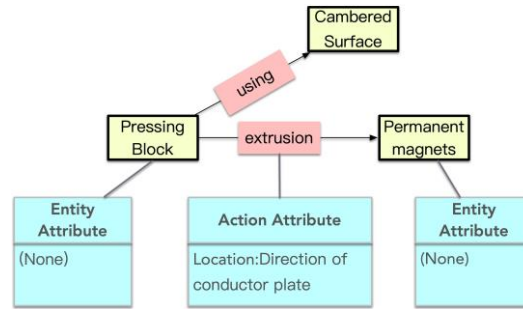


Fig. 6. AAM of single Chinese patent sentence

5 Conclusions and Future Work

The value of a patent comes from the level of patented technology. This means that in order to improve the overall value of patents, it is essential to cultivate their technical value. The ideal level is an important tool to evaluate the gap between the technical scheme and the existing technical level in the process of evaluating the advanced nature of the technical scheme. There are four important indexes, they are: sub-function importance F_{IM} , sub-function realization degree P_F , component cost evaluation value C_S , component harmful function evaluation value H_S . For the evaluation of the enterprise's current technical solutions, the information of these four indicators is easy to obtain. In order to obtain the ideal level index of the existing technology, the evaluation process of the ideal level of patent is divided into shallow information analysis, deep information analysis and information reasoning and calculation process. However, it is unrealistic for Chinese enterprises to only rely on manual extraction of patent technology principle features when facing a large number of patents. Therefore, this research focuses on the automatic extraction of Chinese patents, and tries to build the principle level model of patent technology information by using the extracted information.

As for the evaluation process of technical advancement, there are still many points to be broken in this study in the future:

(1) In the process of shallow patent information analysis, although the use of objective indicators in the patent can quickly evaluate the comprehensive value of the patent. However, if we only use the objective index of patent to evaluate the value of patent technology, it often leads to misleading evaluation results, and many new high-level patent technologies are ignored. There are also some methods based on word frequency statistics to evaluate the novelty degree field of patents, which can also be used as a

way to quickly screen patents. However, due to the different forms of Chinese words and English words, we need to continue to study the statistical methods of Chinese words.

(2) In the process of deep patent information analysis, not all Chinese patents will describe the component-level information according to the standard writing method, or even may not describe the component-level information. Therefore, not every patent can extract CHM. The accuracy of AAM depends on the SDP in LTP. The accuracy of SDP in the current LTP 4.0 is 76.62% (Model: Base2), which can ensure that AAM has a certain accuracy, but some contents need manual calibration. At the same time, the relationship between attribute information and semantic tags of SDP still needs to be further considered. It is possible that one tag may correspond to multiple attributes, and this problem will be solved later.

(3) In the process of information reasoning and calculation, the relationship between AAM and the Flow Model is still not given. At the same time, it is also a big problem to obtain the degree of realization of sub-function P_F . Therefore, the main research content in the future is how the Flow model will be presented in the end.

Acknowledgements. This research is sponsored by the National Innovation Method Fund of China (2019IM020200). We thank colleagues and experts for their help and reviewers for the improvement comments.

References

1. Jell, F., Henkel, J., Wallin, M.W.: Offensive Patent Portfolio Races. *Long Range Plann.* 50, 531–549 (2017).
2. Yu, X., Zhang, B.: Obtaining advantages from technology revolution: A patent roadmap for competition analysis and strategy planning. *Technol. Forecast. Soc. Change.* 145, 273–283 (2019).
3. Kokshagina, O., Le, P., Benoit, M.: Should we manage the process of inventing? Designing for patentability. *Res. Eng. Des.* (2016).
4. Kokshagina, O., Masson, P. Le, Weil, B., Felk, Y.: Innovative field exploration and associated patent portfolio design models. *Res. Interact. Des.* 4, 402–408 (2016).
5. Felk, Y., Masson, P. Le, Weil, B., Cogez, P., Hatchuel, A.: Designing patent portfolio for disruptive innovation—a new methodology based on CK theory. In: Maier, A. (ed.) *Proceedings of the 18th International Conference on Engineering Design (ICED 11)*. pp. 214–255. Design Society, Copenhagen (2011).
6. Cascini, G., Russo, D., Zini, M.: Computer-aided patent analysis: Finding invention peculiarities. *IFIP Int. Fed. Inf. Process.* 250, 167–178 (2007).
7. Cascini, G., Fantechi, A., Spinicci, E.: Natural language processing of patents and technical documentation. In: *International Workshop on Document Analysis Systems*. pp. 508–520. Springer, Berlin (2004).
8. Cascini, G., Russo, D.: Computer-aided analysis of patents and search for TRIZ contradictions. *Int. J. Prod. Dev.* 4, 52–67 (2007).
9. Fantoni, G., Apreda, R., Dell’Orletta, F., Monge, M.: Automatic extraction of function-behaviour-state information from patents. *Adv. Eng. Informatics.* 27, 317–334 (2013).
10. Chiarello, F., Fantoni, G., Bonaccorsi, A.: Product description in terms of advantages and drawbacks: Exploiting patent information in novel ways. In: *Proceedings of the*

- International Conference on Engineering Design, ICED. pp. 101–110. [s.n.], Vancouver (2017).
11. Gupta, G., Kumar, N.: Data acquisition based web scrapping algorithm for extraction of data sets from Patent Portal. (2015).
 12. Vahab, N., Neethu, K. V., Binu, R., Mathew, S., Khamari, L.: An NLP based information extraction system for patents. In: ACM International Conference Proceeding Series (2016).
 13. Okamoto, M., Shan, Z., Orihara, R.: Applying information extraction for patent structure analysis. SIGIR 2017 - Proc. 40th Int. ACM SIGIR Conf. Res. Dev. Inf. Retr. 989–992 (2017).
 14. Hu, J., Li, S., Yao, Y., Yu, L., Yang, G., Hu, J.: Patent keyword extraction algorithm based on distributed representation for patent classification. *Entropy*. 20, (2018).
 15. Huang, Z., Xie, Z.: A patent keywords extraction method using TextRank model with prior public knowledge. *Complex Intell. Syst.* (2021).
 16. Takano, K., Tanaka, M., Sakai, H., Kitajima, R., Ota, T., Tanabe, C., Sakaji, H.: Extraction of Characteristic Terms from Patent Documents for Technical Trend Analysis. *Proc. - 2019 8th Int. Congr. Adv. Appl. Informatics, IIAI-AAI 2019*. 667–672 (2019).
 17. Othman, R., Noordin, M.F., Gusmita, R.H., Sembok, T.M.T., Zulkifli, Z.: SAO extraction on patent discovery system development for Islamic Finance and Banking. *Proc. - 6th Int. Conf. Inf. Commun. Technol. Muslim World, ICT4M 2016*. 59–63 (2017).
 18. Wang, L. yin, Zhao, D.: Cross-domain function analysis and trend study in Chinese construction industry based on patent semantic analysis. *Technol. Forecast. Soc. Change*. 162, 120331 (2021).
 19. Ding, W., Wang, J., Zhu, H.: Using Graph Representations for Semantic Information Extraction from Chinese Patents. *ACM Int. Conf. Proceeding Ser.* (2019).
 20. World Intellectual Property Organization: World Intellectual Property Indicators 2020. WIPO, Geneva (2020).
 21. Iazzolino, G., Migliano, G.: The Valuation of a Patent through the Real Options Approach: A Tutorial. *J. Bus. Valuat. Econ. Loss Anal.* 10, 99–116 (2015).
 22. Danish, M.S., Ranjan, P., Sharma, R.: Valuation of patents in emerging economies: a renewal model-based study of Indian patents. *Technol. Anal. Strateg. Manag.* 32, 457–473 (2020).
 23. Russell, M.: The valuation of pharmaceutical intangibles. *J. Intellect. Cap.* 17, 484–506 (2016).
 24. Ma, S.C., Feng, L., Yin, Y., Wang, J.: Research on petroleum patent valuation based on Value Capture Theory. *World Pat. Inf.* 56, 29–38 (2019).
 25. Ignat, V.: Modern evaluation of patents. In: 7th International Conference on Advanced Concepts in Mechanical Engineering. pp. 1–7. IOP Publishing, Romania (2016).
 26. Cao, G., Tan, R.: FBES model for product conceptual design. *Int. J. Prod. Dev.* 4, 22 (2007).
 27. Cao, G., Tan, R.: Function design principle and Application. Higher Education Press (2016).
 28. Che, W., Li, Z., Liu, T.: LTP: A chinese language technology platform. In: Coling 2010 - 23rd International Conference on Computational Linguistics, Proceedings of the Conference. pp. 13–16 (2010).
 29. Semantic dependency annotations, <http://ltp.ai/docs/appendix.html#id6>, last accessed 2021/04/21.
 30. Graphviz, <http://www.graphviz.org/>, last accessed 2021/04/22.