International Journal of Innovative Computing, Information and Control
Volume **15**, Number **4**, August **2019**

 ISSN 1349-4198
pp. 1591–1599

RESEARCH ON AUTOMATIC COMPUTATIONAL MODEL OF TECHNOLOGY LIFE CYCLE BASED ON DOMAIN ONTOLOGY

Yao Liu[1], Qian Zhang[2] and Yi Huang[1]

[1]Institute of Scientific and Technical Information of China
No. 15, Fuxing Road, Beijing 100038, P. R. China
liuy@istic.ac.cn

[2]School of Software and Microelectronics
Peking University
No. 5, Yiheyuan Road, Beijing 100080, P. R. China

Received August 2018; revised December 2018

Abstract. *With low-level automation and inexact technical classification, traditional technology life cycle computational model cannot generate dynamic technology life cycle. After summarizing the existing technology life cycle construction methods, this paper proposes a new evaluation model of technology life cycle based on domain ontology. First, this model constructs a domain ontology with domain vocabularies and natural language processing methods which could effectively define and extract technical words. Next, through article data, a three-layer ontology structure can be built. Then, the technical words of each article are labeled by semantic tagging, and the relationship between technical words has also been redefined. By this means, we can conveniently evaluate technology life cycle by statistics and analysis of the development trend of articles. At last, this paper achieves an automatic computational model of dynamic technology life cycle. The empirical study on AI and graphene fields shows the effectiveness of the model.*

Keywords: Technology life cycle, Domain ontology, Domain technology, Calculation of trend

1. **Introduction.** Proposed by Little [1], the technology life cycle (TLC) can measure technological changes through the competitive impact and integration in products or process. With the rapid development of science and technology, life cycles of technology are getting shorter over time. Researchers often face enormous amounts of data and information resources for analysis and evaluation. How to effectively use these resources to evaluate the life cycle of technology has become a problem to be solved urgently. The traditional TLC evaluation method often requires lots of manpower and time. In fact, before researchers calculate the TLC of some field, it is necessary to find out all the related technologies in a research field and to determine the correlation of these technologies by experience, and need to repeat the complex and time-consuming preparation process if research areas or technology points are changed. Therefore, this paper proposes an automatic computational model based on domain ontology, simplifies the preliminary work of the TLC evaluation, and finds related technologies in the field by semantic relations extracted from science and technology corpus.

2. **Literature Review.** One of the mainstream theories of TLC is "four phase theory", which believes that the life cycle of technology, like other things in nature, follows the rules of germination, growth, maturity and decline. Little and Foster are the representative

DOI: 10.24507/ijicic.15.04.1591

scholars of this theory [1,2]. TLC evaluation methods based on technology resources can be divided into two categories: qualitative and quantitative methods.

The quantitative method is based on a large amount of data, using data characteristics and mathematical methods to determine the life cycle of the technology. The quantitative method can also be divided into two categories: supervised [3,4] and unsupervised [5,6]. However, most quantitative methods need to rely on a large number of data, and no matter what model it takes, the results of the final output have some uninterpretability. Makovetskaya and Bernadsky [7] analyzed quantitative relationships between papers, patents, and standards to determine the stage of TLC.

The qualitative evaluation method is basically a kind of data statistics, and researcher makes an evaluation on the stage of TLC in accordance with his own experience and subjective understanding. Zhao et al. [8] used Thomson Data Analyzer to perform data mining of patents on genetically modified crop research and analyzed the development trend by relating patent growth to industry background, such as policy implications. The qualitative method is highly dependent on the experience of experts or patent analysts, which leads to high human cost and low efficiency.

In addition, the current TLC evaluation methods, after the complex pre-processing of corpus, can only be applied to a specific field. For example, [9,10] are typical TLC evaluations for a specific field. At present, there are few studies on the improvement of corpus mining in the process of TLC evaluation. Therefore, this paper proposes to use ontology for corpus mining, and on this basis, to establish a TLC evaluation model.

3. **Automatic Computational Model of TLC Based on Domain Ontology.** This paper constructs an automatic TLC computational model based on domain ontology, aiming to automatically evaluate the life cycle of related technologies in the field of core technology.

As shown in Figure 1, the TLC evaluation process based on domain ontology can be divided into five steps:

1) Determine technology scope and obtain data resources;
2) Construct an ontology;
3) Obtain domain technical words;
4) Calculate the growth trend based on journal and conference papers;
5) Determine TLC stages.

Ontology construction is the basic part of the model, and determines the scope of technical words for subsequent process. In the construction process, we parse the text

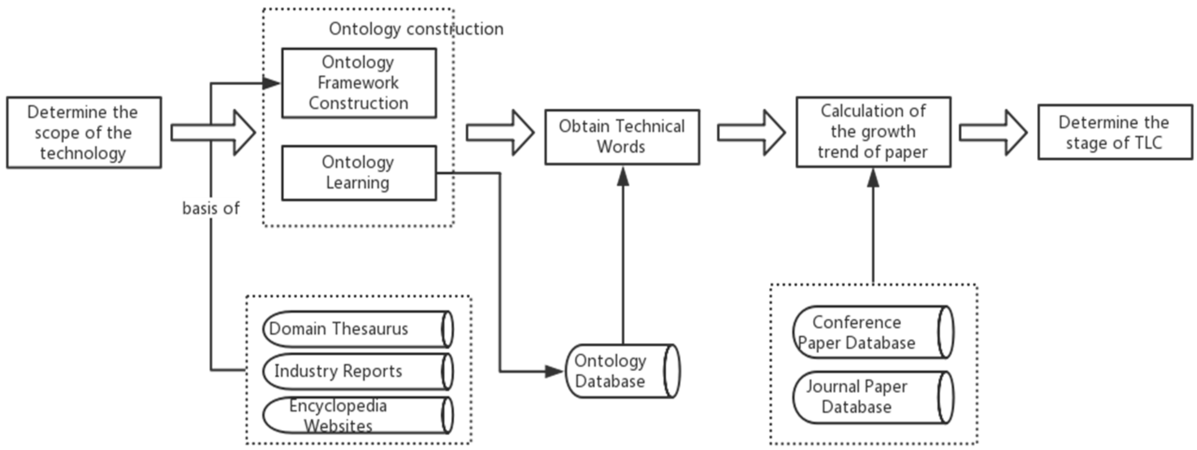

Figure 1. TLC evaluation flow chart based on domain ontology

from the domain database, and preprocess the text, such as segmentation, part-of-speech tagging, stop word elimination, and text label extracting. Then, the core ontology framework is established by referring to thesaurus, international standards, IPC classifications, encyclopedia, catalogues and so on. After that, we can acquire the structure of ontology by using tags extracted from text parsing. With the completed conceptual structure, the ontology can be promoted to evolve by semantic indexing of remaining texts. The evolution process is ontology learning.

The obtainment of technical words is to calculate the similarity between technical concepts in the evolved domain ontology. Through calculation, we can get many technical words related to the core technical word. The life cycle of domain technologies will be calculated based on these proper technical words in the next phase. Then, the growth trend of scientific papers will be calculated with the paper databases. The characteristics of growth trend in different life cycle stages will be the basis for the division. In the following, this article will elaborate from these four aspects: ontology construction, domain technical words obtaining, basis for the division of TLC and trends analysis method.

3.1. **Ontology construction.** Ontology construction can be divided into two parts: ontology framework construction and ontology learning. This paper establishes the ontology core framework by referring to ontology concept description system in [11,12] building methods and other structured content.

First, it is necessary to analyze and collect scientific data such as domain topic word list, book catalogues, standards, industry report and literature to extract technology words related to concepts. The topic in word list is the concept node of the ontology, and according to the analysis of the domain content, the core concept and the vertical relationship between the concepts are determined, and the synonyms are added to the concept synonym. After that, according to the existing structural data, we can analyze the horizontal relationship between concepts and determine the attribute words to construct the attribute word list.

After establishing the ontology framework, ontology learning is conducted with the paper data in the database and build a three-level conceptual structure [13]. In essence, the process of ontology learning is to match three-tuple of concept word, attribute and attribute value in corpus, and establish relations between concept, attribute and value. If the match is successful, the attribute word becomes the attribute of the concept, and then the result and data information is stored in database. If the match fails, the three tuples are discarded.

3.2. **Obtainment of domain technical words.** Domain technology can be divided into two categories. One is that technologies are similar to the core technology in the same ontology, such as machine learning (core technology) and deep learning. The other is the technology directly related to the core technology, which can be understood as the relationship between concepts and attributes in the same ontology, such as machine learning and FP-Tree. Therefore, we first take the core technical words as the center to obtain all the attribute words under the same concept, and then calculate the similar concept words as a supplement to obtain the domain technical words.

The similarity between ontology concepts is mainly affected by the ontology structure. Therefore, the acquisition of related technologies is mainly determined by the structural similarity of the concept in ontology. The formula for calculating the similarity of ontology [14] is as follows:

$$sim(C_1, C_2) = Dist(C_1, C_2) \tag{1}$$

$Dist(C_1, C_2)$: The semantic distance of the relationship path between ontology concepts C_1 and C_2, namely, the sum of the distance between concepts C_1, C_2 and LCA (lowest common ancestor).

According to the above formula, a part of the technical concept with the highest similarity with the core technology concept is selected, and all the attribute words corresponding to the core concept are added to form the list of domain technical words. By calculating the life cycle of all technical words in the list, we can directly show the development of technology in a certain field.

3.3. **Basis for the division of TLC.** In this paper, the selection of TLC characteristics mainly focuses on the performance of technology in scientific literature. This paper is based on the characteristics of different TLC stages in the scientific papers [15,16]. In embryonic period, emerging technology is the focus of discussion in field conference. This feature will be reflected in the conference papers. With the continuous development of technology, many relevant articles will be cited in the journal articles, the number of journal articles will increase while the conference papers will be gradually reduced. At the same time, there are more theoretical papers. After the technology theory matures, the focus will be transferred to applied research. Therefore, the number of applied papers is increasing in the late stage of technology development. Based on the above assumptions, we evaluate the characteristics of each stage TLC as follows.

Embryonic period: There are many papers in the conference and a few papers in the journal. These papers are mostly theoretical ones.

Growth period: Both conference papers and journal papers have grown rapidly, and applied literature has begun to appear, but the proportion of theoretical papers is larger.

Maturity period: Conference papers have been reduced, journal papers have continued to grow, applied literature has increased, and theoretical literature has decreased.

Aging period: The number of conference papers is almost zero, journal articles are reduced, and both theoretical and applied literature is reduced.

Table 1 and Table 2 describe the characteristics of the papers in the various stages of the TLC in the form of formulas.

Table 1. Trend division of different types of scientific papers

	True (Increase)	False (Decrease)
Trend of conference papers	A1	A2
Trend of journal papers	B1	B2

Therefore, according to the list of domain technology words, the number of journals and conference papers in a specific time period can be counted, and the growth trend of journals and conference papers can be evaluated, and the TLC can be evaluated by combining the technology's performance characteristics in different life cycles.

3.4. **Trends analysis method.** We can divide papers into different time intervals according to development trend, so we can use the linear regression method to fit the line equation, so that the slope of the line segment can be calculated.

According to the above hypothesis, there is a certain correlation between peroid and the number of papers in the TLC, and this correlation is linear. Therefore, we can construct a linear regression equation as follows:

$$\hat{y} = a + bx \tag{2}$$

Among them, $\hat{y}$ is the annual number of journals and conference papers, x is the year, a is the regression constant, and b is the regression coefficient, which is the slope of the

Table 2. The characteristics of each stage of TLC

Stage of TLC	Feature	Description
Embryonic	A1+B1 AND $k_{meetingArticle} - k_{journalArticle} > 0$	The conference and journal papers showed a growth trend. The growth rate of conference papers is faster than that of journal articles.
	A1+B2	Conference papers show an increasing trend and journal papers show a downward trend.
Growth	A1+B1 AND $k_{meetingArticle} > R_{meeting}$, $k_{journalArticle} > R_{journal}$	The conference and journal papers showed a growth trend. Meetings and journals are both growing fast.
Maturity	A1+B1 AND $k_{meetingArticle} < R_{meeting}$, $k_{journalArticle} < R_{journal}$	The conference and journal papers showed a growth trend. Meetings and journals are both growing slowly.
	A2+B1	Conference papers show a downward trend and journal papers show an increasing trend.
Aging	A2+B2	The conference and journal papers showed a downward trend.
Note: k is the slope of curve, R is the threshold from experience, and same as the average.		

straight line. If the slope is greater than zero, the function monotonically increases and the trend is upward; otherwise, the function monotonically decreases and the trend is downward.

4. **Experiment.** Considering characteristics of scientific papers at various stages of the TLC, it can be found that conference papers play an important role in defining the life cycle, since computer science and its related fields pay more attention to the publication of conference papers. Therefore, this paper selects papers in artificial intelligence (AI) field as the experimental data. In addition, graphene as an emerging material has a wide range of applications. Therefore, grapheme field is also chosen to conduct the experiment. There are many technical points involved in these two fields; however, many studies [17-19] in these two fields only focused on one or several specific technical points. In this part, the experiments were carried out from the perspective of the overall development of domain technology areas by using the proposed model.

The experimental data source is WANFANG DATA database, covering a large-scale database of journals, dissertations, conference proceedings, patents, standards, etc. The data of journal and conference papers mainly come from the China Conference Paper Database (CCPD) from 1976 to 2016. Ontology construction refers to the AI industry reports, such as *Hype Cycle for Artificial Intelligence* [20] released by Gartner Inc., and the *Artificial-Intelligence-Terminology*[1] published by Synced, so does graphene.

Figure 2 and Figure 3 show the results based on "machine learning" and "hidden Markov model" (HMM) as core technology word. As shown in the figures, in the field of artificial intelligence, if "machine learning" is used as the core technical word for retrieval, the results show that "machine learning" is still in the embryonic stage. "Deep learning" and

[1]https://github.com/jiqizhixin/artificial-intelligence-terminology.

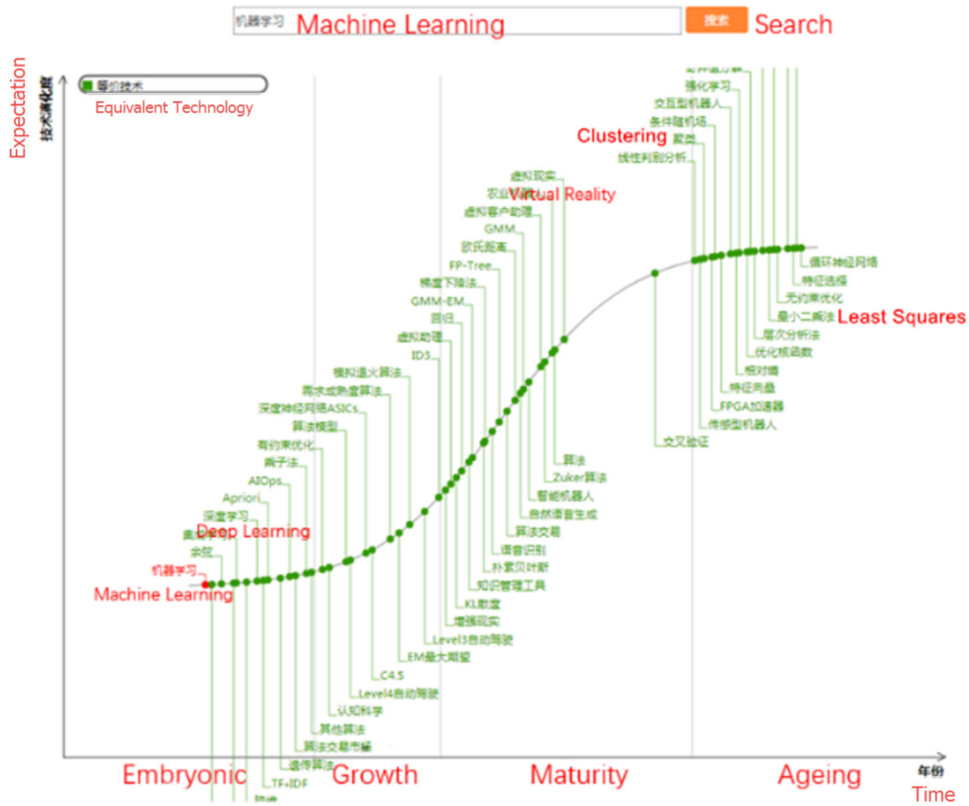

FIGURE 2. Result based on "machine learning"

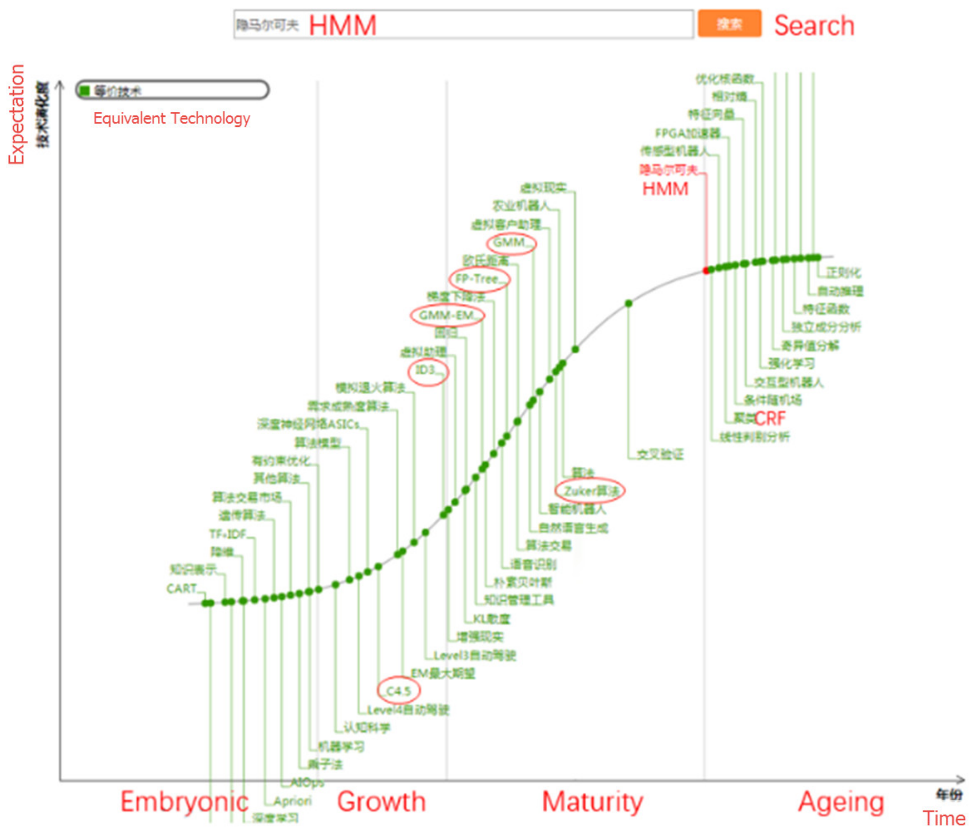

FIGURE 3. Result based on "HMM"

"integrated learning" are also in the embryonic stage because of the similarity to "machine learning". Although the "virtual reality" is also in the field of artificial intelligence, it is at the turning point. This result is consistent with Gartner's report [20]. Among the machine learning algorithms, "clustering" and "least squares" are at aging stage, which indicates that these algorithms are relatively mature. The above results are consistent with common sense in the field of artificial intelligence. If HMM is used as the core technical word for retrieval, the related technical words of the TLC curve of artificial intelligence are different, and words related to mean algorithm are obviously more, as shown in Figure 3.

Figures 4 and Figure 5 are the results based on "graphite tape stripping machine" and "graphite plate" as core technology words. As shown in the figures, in the field of graphene, if "graphite tape stripping machine" is used as the core technical word for retrieval, the result shows that there are many technical words related to "stripping", such as "stripping machine" and "micromechanical stripping". If "graphite plate" is used as the core technical word for retrieval, the result shows that there is no word related to "stripping", because "graphite plate" is a kind of compound which is made mainly through the rolling process, not the stripping process.

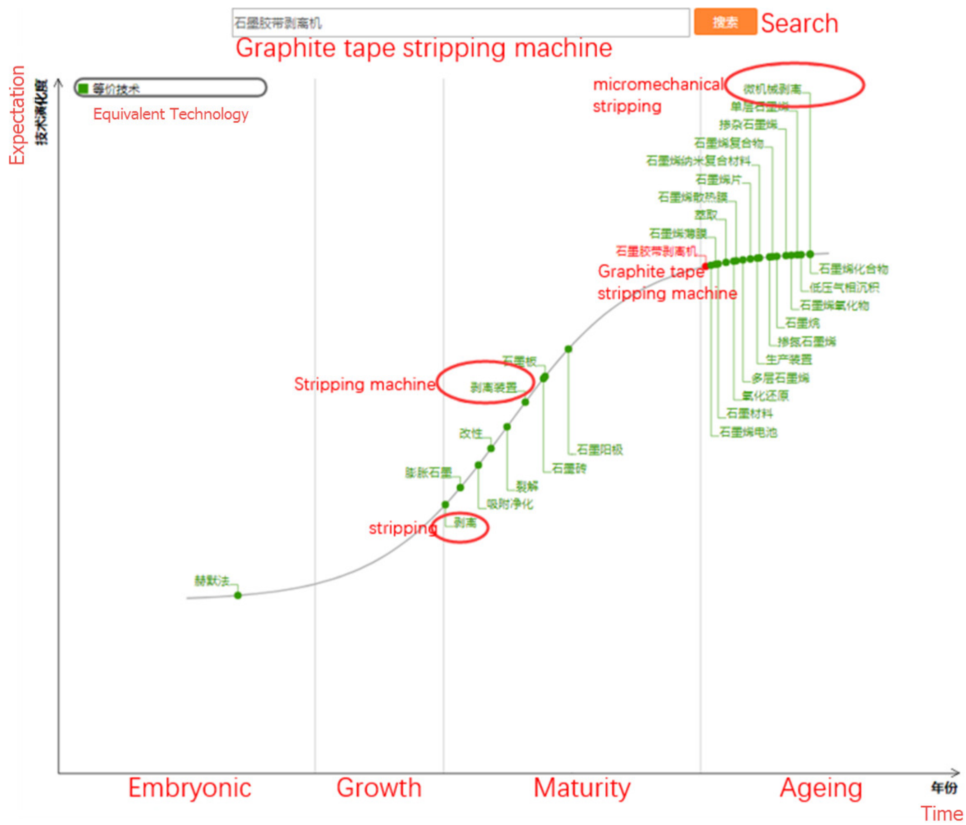

FIGURE 4. Result based on "graphite tape stripping machine"

The above results prove that the proposed model can make the evolution of domain technology ontology by extracting the semantic relationship between scientific papers and calculating the similarity of the ontology concepts. Thus, it can quickly locate a specific technology point in a certain field, and calculate the life cycle of the core technology and its related technology points in a unified way. In the past studies on TLC, the proposed research methods only evaluate the life cycle of core technology in certain field, but neglect the development status of related technologies in the same field. This model improves

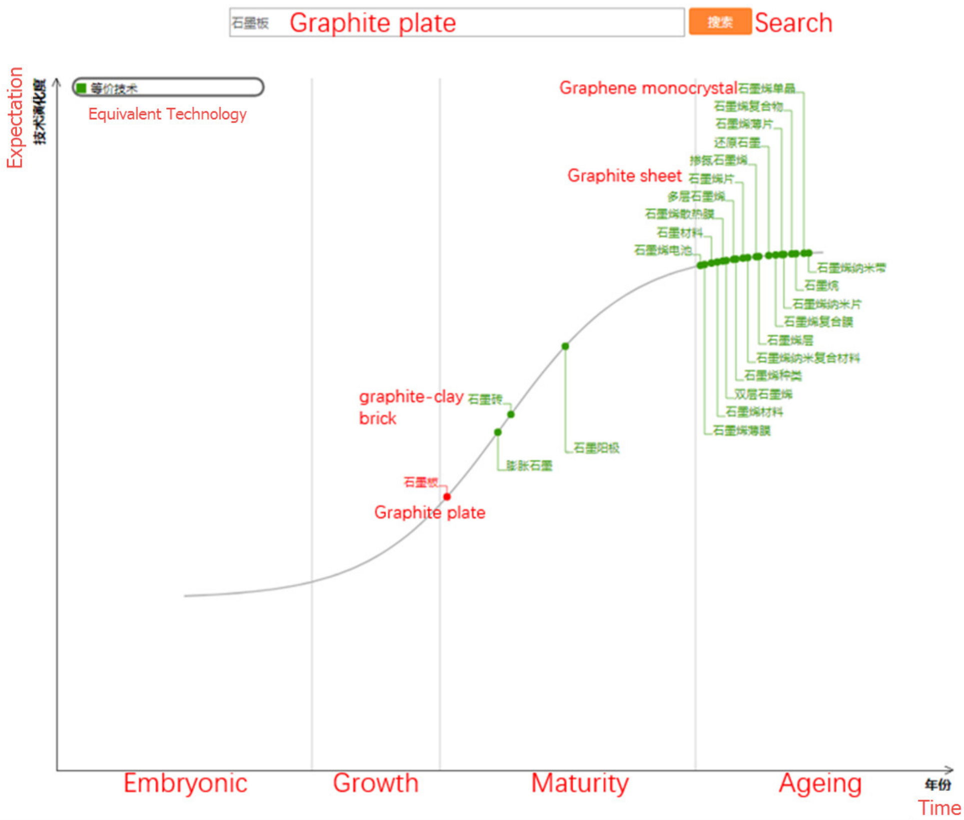

FIGURE 5. Result based on "graphite plate"

this point, which is helpful for researchers to compare and evaluate the development of different technology points in the same field from the overall point of view, and refine the granularity of domain TLC evaluation.

5. **Conclusions.** On the basis of previous research, relying on the professional semantic resource processing platform, this paper proposes an ontology-based method for automatic evaluation of TLC, which improves the efficiency of TLC computation. Compared with the existing methods, this method owns the following advantages. 1) In the ontology construction, the efficiency and accuracy of domain technical words acquisition are improved by combining core words, ontology structure, sematic tagging and so on. 2) Realize the dynamic evaluation of domain TLC. In the process of building domain ontology, we use the corpus to get three-level conceptual ontology structure and establish the relationship between concepts. By analyzing the concept and attribute of the core technical words (query words), we can quickly get the technology domain word list. And multi-angle and multi-level analysis of a certain field of technology can be performed through the evaluation of each technology in the list.

In the future, we intend to proceed along two lines in parallel. 1) Add more relevant features. According to the experimental results, the development trend of scientific papers has a certain influence on the evaluation of the TLC, and it is difficult to accurately determine the TLC by limited number of features. 2) Time series analysis for more periods. Trend analysis of a single time period is difficult to grasp the feature of technology development cycle as a whole. Therefore, it can be combined with more time periods to analyze the time series of the TLC.

REFERENCES

[1] A. D. Little, *The Strategic Management of Technology*, Arthur D. Little, 1981.

[2] R. N. Foster, *Innovation: The Attacker's Advantage*, Summit Books, 1988.

[3] J. Kim et al., Design of TOD model for information analysis and future prediction, *International Conference on U- and E-Service, Science and Technology*, Springer, Berlin, Heidelberg, 2011.

[4] L. Gao et al., Technology life cycle analysis method based on patent documents, *Technological Forecasting and Social Change*, vol.80, no.3, pp.398-407, 2013.

[5] T. Modis, Strengths and weakness of S-curves, *Technological Forecasting and Social Change*, vol.74, no.6, pp.866-872, 2007.

[6] H. Zhong and H. Deng, Patent portfolio analysis based on technology life cycle, *Library and Information Service*, vol.56, no.18, pp.87-92, 2012.

[7] O. Makovetskaya and V. Bernadsky, Scientometric indicators for identification of technology system life cycle phase, *Scientometrics*, vol.30, no.1, pp.105-116, 1994.

[8] X. Zhao, Y. Wang and R. Hu, Patent layout and status analysis of genetically modified crop technology, *Journal of Intelligence*, vol.33, no.9, pp.51-55, 2014.

[9] S. P. Jun, An empirical study of users' hype cycle based on search traffic: The case study on hybrid cars, *Scientometrics*, vol.91, no.1, pp.81-99, 2012.

[10] H. M. Jarvenpaa and S. J. Makinen, An empirical study of the existence of the hype cycle: A case of DVD technology, *IEEE International Engineering Management Conference*, 2008.

[11] Y. Liu et al., Research on automatic construction of Chinese traditional medicine ontology concept's description architecture, *Data Analysis and Knowledge Discovery*, vol.24, no.5, pp.21-26, 2008.

[12] Y. Liu et al., Research on semantic methods of library resource organization interaction based on content and form, *Information Theory and Practice*, no.10, pp.105-107, 2010.

[13] Y. Liu et al., Study on text segmentation based on domain ontology, *Computer Science*, 2018.

[14] W. Li and Y. Zhao, Semantic similarity between concepts algorithm based on ontology structure, *Computer Engineering*, vol.36, no.23, pp.4-6, 2010.

[15] M. Taylor and A. Taylor, The technology life cycle: Conceptualization and managerial implications, *International Journal of Production Economics*, vol.140, no.1, pp.541-553, 2012.

[16] X. Wang et al., The overview of technology life cycle analysis method based on factual database, *Digital Library Forum*, 2013.

[17] J. Sha et al., Study on the trends of global graphene tech-innovation based on patent analysis, *Materials Review*, vol.27, no.15, pp.108-112, 2013.

[18] C. Chen and H. Wu, Visual analysis on graphene patent information of China, *Modern Information*, vol.34, no.3, pp.120-124, 2014.

[19] L. Ge, Research on technology life cycle based on patent information analysis – Taking method for manufacturing graphene of China as an example, *Technology Intelligence Engineering*, vol.1, no.4, pp.58-64, 2015.

[20] K. Brant and T. Austin, *Hype Cycle for Artificial Intelligence, 2017*, Gartner, 2017.