A quick approach to get a technology-function matrix for an interested technical topic of patents

Liu, Kuotsan
Associate Professor

Yen, Yunxi
Graduate student of master's degree,
Graduate Institute of Patent
National Taiwan University of Science and Technology

Yen-Hung Kuo
Cloud System Software Institute, Institute for Information Industry

Abstract

A quick approach to get a technology-function matrix, one kind of patent map, is presented in this paper. Two popular technologies, big data and cloud storage in IT, were illustrated. In order to get high quality patents and guidance of R&D, to get a technology-function in advance in short term is important. The approach introduced in this paper is more quickly than traditional method, and is easily update or expand more detail.

Keyword：patent map, patent analysis, cloud storage

1. Introduction

The activities of R&D have become more prosperous and occupied a large percentage of a company's budget. The knowledge produced in technological innovations, especially knowhow and patent, is an important asset for a company.

Knowhow is valuable on its secrecy, on the contrary, patent should open to public, but it is powerful to stop competitors enter claimed scopes based on its exclusive rights. A company owns a big amount of patents is normal in modern industry. Famous companies, for example IBM and Microsoft, have more than 100 thousands patents. That is to say, a company who can't get enough patents may disappear in competition market.

Traditional patent filing procedure of a company is starting at an invention proposal brought up from R&D, after prior arts search to insure its patentability, specifications and claims drafted by patent attorneys, and then file to one or more patent offices. After substantial examination by patent examiners, about 50% to 60% on

the average, of applications could be granted, the other rejected applications mostly lack of novelty, or obviousness based on prior arts. It can be said that 40% to 50% of innovations are unsuccessful, the research fields of R&D which budget invested in had been occupied by prior patents.

In order to get higher benefit of R&D budget, patent data searching or mining is important before R&D. Patent maps are useful to get a macroscopic of view in interested technical fields. This paper will focus on a patent map, specifically, technology-function matrix, which can show the patent density in each technology-function node, and give a good guidance of technical development.

It wastes time to form a technology-function matrix, because it needs to read patent documents and put them into the classified nodes, or baskets. This paper will introduce a new approach to get a matrix more quickly, and illustrating the approach by a popular technical topic in information technology industry, big data and cloud storages patents.

2. Background

There are more than 1.6 millions of patent publications in one year all over the world by statistics of World Intellectual Property Office. Patent mining itself, especially patent bibliometrics and patent map, has become a professional and research field because its complexity and difficulty. The complexity means an invention may file to more than one patent offices, it may be published many times in different languages. A patent family of an invention owns more than one hundred patents is not rare. The difficulty means search by keywords may lose of accuracy because applicants like to use generalized words instead of normal words to broad their claim scope.

2.1 Patent Bibliometrics

Patent bibliometrics is a mathematical and statistical study of patent quantity and quality based on patent documents. This study develops methodology to get useful information from patent publications. It always start at organizing search queries for interested topic, for example, an interested technical field or a competitor, and then get a pool of patents from database, to make some macroscopic analysis by variable method.

In the latest two decades, patent citation analysis to determine strength and value of a patent is prosperous in patent bibliometrics. Researchers developed some indexes to evaluate quality of patent. This paper will use bibliography of patent documents but not focus on patent index.

2.2 Patent map

Patent maps are useful tools to visualizes the distribution of patents, monitor the trend of technological changes, infer the strategy of patent portfolios, by statistical charts or diagrams. New patent maps were developed because of multi-demands and huge patens. This paper proposed a quick approach to get an old patent map, but not a new one.

2.3 Technology- function matrix

A technology-function matrix is one kind of patent map, which is useful to R&D. The innovation activity in R&D for an interested technical topic is usually doing prior arts search first, and then list what kind of technical problem in the prior arts have to be solve, or what kind of function is needed. A new technical means or a solution figured out to a technical problem will be a new invention. It exists more than one technical means for a problem. A patent specification always needs to describe technical fields of the invention, prior arts, the technical problem intend to solve, the technical means solving the problem, the new functions comparing the prior arts, and preferred embodiment. The formality of description complies with the manner of R&D.

A technology-function matrix is a two dimension matrix, which using the functions and the technical means to be its two coordinate axes, and drawing each nodes proportional to the number of patents. A bigger node means higher patent density which is a popular and crowded technical problem and solution. On the contrary, a smaller node means lower patent density which is a neglected problem and solution.

A technology-function matrix is a powerful tool for R&D. Unfortunately, it takes time to read every patent document and decide which node or basket it should drop in. It is normal to get more than one thousand patents in one technical pool of interested topic in modern patent database. People need a quickly methodology to get a technology-function matrix.

3. Methodology and data

The database employed for this study is big data and cloud storage, both has become the spotlight on IT industry. Huge amount of information is keeping challenging the load of computer system, big data needs technology to improve velocity, variety, volume at the data grade of terabyte, petabyte, Exabyte. The demand of big data pushes the digital application to the cloud age. Companies might put their data in cloud storages instead of keep in hand.

The main step of methodology in this paper is using the patent classification system instead of R&D classification. Two major systems, International Patent Classification(IPC) and United States Patent Classification(UPC), are used and published with patent document. We preferred UPC and US patents because of detail classification and rich data in computer software related patents.

Fig.1 shows flow chart of this approach. Let the definitions of patent classification be technical means. For example, UPC707 is “database and file management or data structures”. There are 10 subclasses under UPC707, UPC707/600 is “data warehouse, data mart, online analytical processing (OLAP), decision support systems”, UPC707/607 is “online transactional processing (OLTP) system”, USC707/609 is “file or database maintenance”, etc.

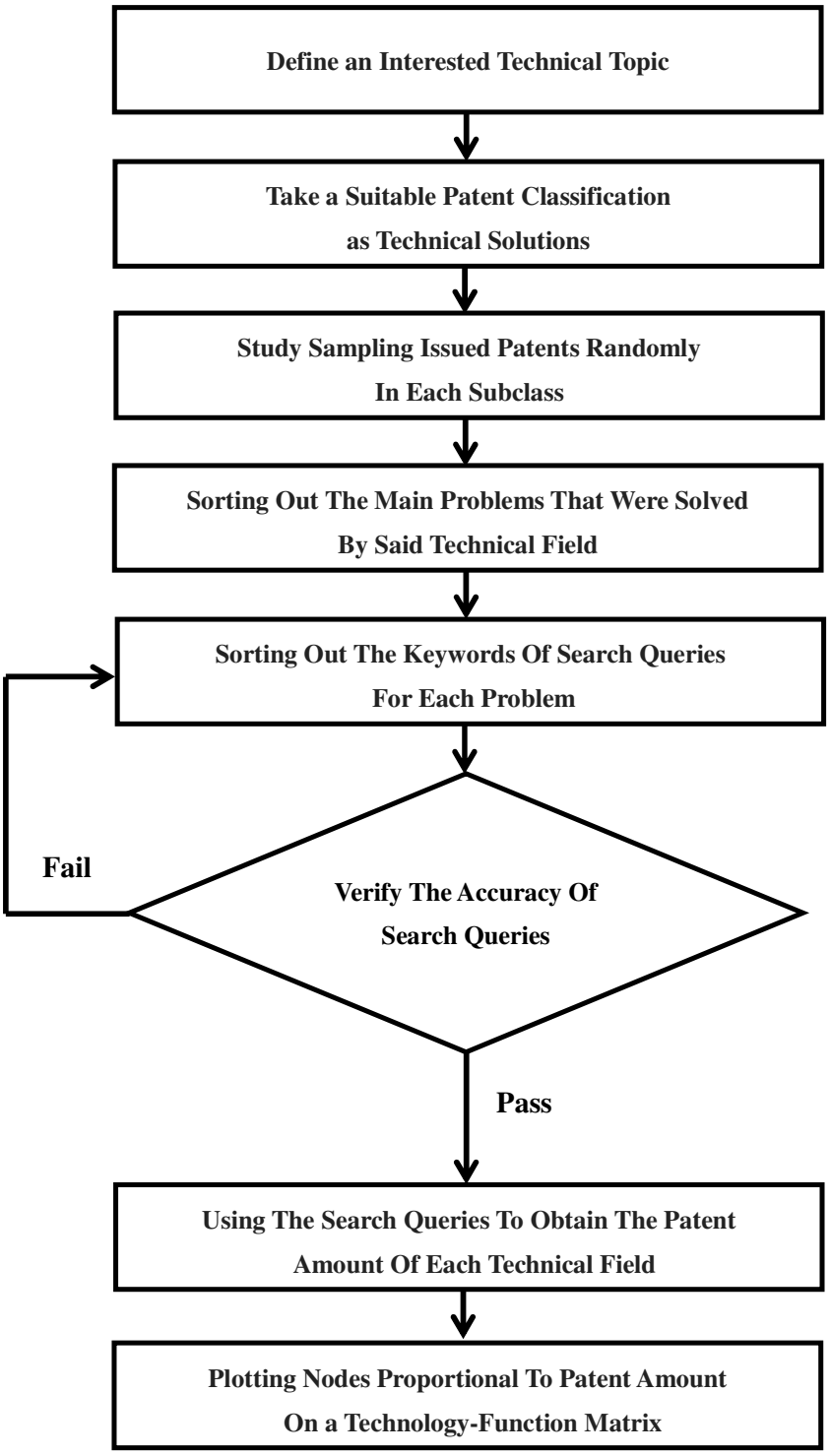

Fig.1 Flow chart to get a technology-function matrix

Each subclass will be a technical means. Patent office has already dropped every patent document to its technical means basket when classified it. Patent examiners are well-educated and experienced. The accuracy and quality of classification can be guaranteed.

One axis of the matrix has already finished by patent classification, the other axis has to be defined by

reading the document. We can take some patents randomly in the interested pool as samples, and list all technical problems stated in these sample documents after reading, and define the search queries for each technical problem. Using each search queries to each subclass, we can get the numbers of patents at each node quickly, instead of labor classification.

4. Illustration

This paper shows two matrixes for two technical topics as illustrations using patent database of United States Patent and Trademark Office(USPTO), one is "database and file management(UPC707)", the other is "Electrical computers and digital processing systems: multicomputer data transferring(UPC709)".

4.1 An overview of UPC707 and UPC709

In order to define an accurate pool of big data, we used three simple search queries to get an overview of UPC707 and UPC 709.

The first search queries S1 is "UPC707 and IPC G06F", G06F is computer software, and limited in the recent 10 years, we can get 26,286 granted patents. S2 limited "large or big" at specification, not only description but also claims, the number reduced to 3186, which is 12.1% of S1. S3 limited further, put "large or big" at title, abstract or claims, the result is 789, which is 3.0% of S1.

Fig.2 is granted patents of UPC 707 in recent 10 years, S2 shows the granted quantity increases in the latest 3 years. The numbers of 2013 may be higher than 2012, because only one third has been published on search date. S3 shows the same tendency with S2 but not so obvious.

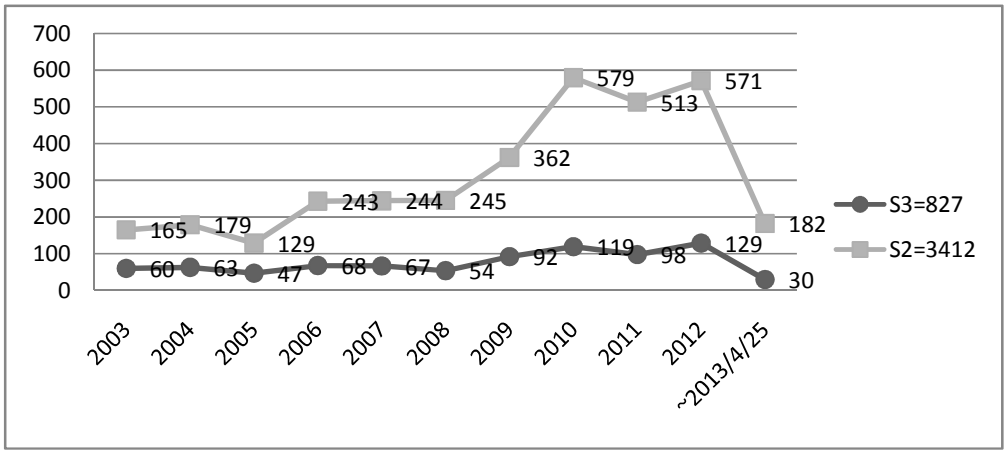

Fig.2 Granted patents of UPC 707 in recent 10 years(latest update:2013.4.25)

Fig.3 is granted patents of UPC 709 in recent 10 years. It shows the same tendency with Fig.2, the number of granted patents double increases at 2010.

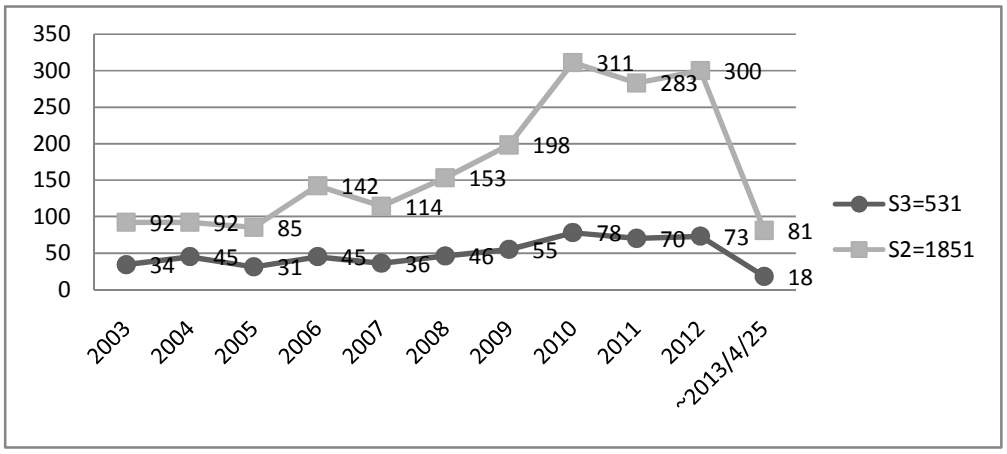

Fig.3 Granted patents of UPC 707 in recent 10 years(latest update:2013.4.25)

Both UPC 707 and 709 show that 2010 is a borderline, which complies with the growth of big data and cloud storage. We limited in recent 3 years to get technology-function matrices, and tried to find guidance of R&D.

4.2 A technology-function matrix of UPC707

There are 10 subclasses under UPC707, Table 1 is their definitions and number of granted, total and in recent 3 years. It shows a popular research field, 83.2% on the average, granted in recent 3 years. The technology-function matrix will limit the pool to the recent 3 years.

Table 1 Definitions and number of patent granted in UPC707

subclass	definition	Total	In recent 3 years (percentage)
600	Data warehouse, data mart, online analytical processing (OLAP), decision support systems	424	333(79%)
607	Online transactional processing(OLTP) system	48	41(85%)
608	Collaborative document database and workflow	86	63(73%)
609	File or database maintenance	2,145	1,766(82%)
687	Data integrity	970	656(68%
705	Database and file access	6,685	5,601(84%)
790	Database design	1,918	1,715(89%)
813	Garbage collection	117	110(94%)
821	File management	472	411(87%)
899	Miscellaneous	29	29(100%)
	Total	12,894	10,725(83.2%)

There are more than 10 thousands of granted patents of UPC707 in recent 3 years. It is difficult to reclassify these patents in short term. We can select 20 patents in each subclass and in the latest one year randomly. After viewing the technical problems or functions in these selected patents, we got the following functions and established their search queries:

(a)Improving data security
((increas* or great* or better or optimiz*) and (data) and (protect* or Secur* Or safe*) not copyright).DSC.
(b)Reducing Costs
(((decreas* or reduc*)and cost*) or "cost down") .DSC.
(c)Reducing Processing Time
((decreas*or reduc*or short*)and ((cost*or wast*) and time)).DSC
(d)Wasted Storage Space Reduction
((storage or stor*) and (cost* or wast* or increas* or decreas* or reduc*or
manag*)) .DSC.
(e)Efficiency Improvement
(quick* or efficien* or fast*) .DSC.

These functions match the trend of big data development. The larger of big data system, the more important of reducing cost and processing time. In order to have enough space to use, wasted storage space reduction is efficient rather than space increase. And when a company transfers their data to cloud server, data security would be the first concern for them.

Applying the search queries to each subclass, we can get the number of patents of each function in each subclass. For example, the number of improving data security is 90 in subclass 600, data warehouse, data mart, online analytical processing(OLAP), decision support systems. Reducing costs is 192, reducing processing time is 9, wasted storage space reductions 297, and efficiency improvement is 236 in the same subclass. One patent may include more than one functions and drop in more than one baskets. After all amounts of patents are found, we can get the technology-function matrix as shown in Fig.4.

Let's check the results one by one problem. The first problem of technological development "improving data security" includes the issues of risk reduction, assessment and protection of large continuous data transmission and the enhancement of system security...etc. The first row indicates different technical means solving the problem, wherein "data base and file access" has high density of patents as 1575. The "data base and file access" were popular in all technical problems except reducing processing time, it owns the highest density comparing to other technical means. Many researchers invested in this technical field and got many patents. The exception "reducing processing time" may be not an urgent problem but might become popular in a few years and can be the next popular research topic, when the parallel computing unsatisfied demands.

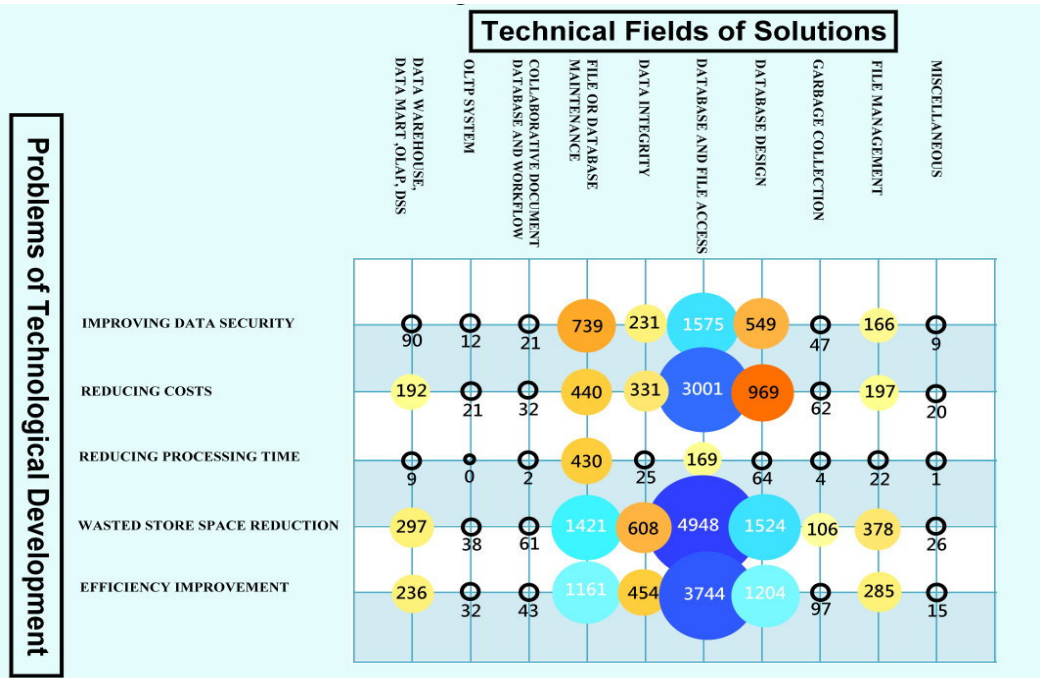

Fig.4 A technology-function matrix of UPC 707 in recent 3 years

(latest update:2013.4.25)

The second row is reducing cost, which includes the issues of reducing the costs of development, design and the labor for management and maintenance of equipment…etc. The highest density is on the database and file access, the same with improving data security, and then the database design. Reducing cost is an important factor to induce users, and is not queer to be a popular function of research. According to the matrix, we can say that not only the costs of processing large scales of data should be considered when an organization is trying to managing and controlling the costs of a system, but the design of database and system structure may also highly affect the implementation costs of the system.

The third row, reducing processing time, includes the issues of data backup, access and retrieval…etc., shows the highest density in file or database maintenance. This row shows lower density comparing to the other four rows. We can learn from the matrix that file or database maintenance technology has the highest occurrence of patent in this problem field which shows that the application of data relevant optimization is the key of development of this field. Moreover, the patent amount of the problem of reducing processing time is relatively few in file or database maintenance technical field but the number of its superordinate concept problem－efficiency improvement has a relatively large amount of patent, which raise the issue of "current database file accessing technology is unable to reduce the processing time directly and effectively". In other words, the applications of reducing processing time are relatively disadvantaged and might get some opportunities to find new or potential direction of research and development.

The fourth row shows the most popular technical development, wasted store space reduction, in UPC707. How to use the store space efficiently is an important problem in huge data increasing. Old data after updated became useless but occupy store space. Specifically speaking, how to use hardware in proper and effective

way to access “big data” might be the priority problems we have to face with nowadays.

The fifth problem of technological development “efficiency improvement” is also an important issue in big data processing. For example, parallel and grid computing, distributed file system and accessing are all within the scope of efficiency improvement problem field. We can also say that the distribution trend of the number of patents is quite similar to the fourth problem of technological development, which means when reducing the wasted of storage space may accompany with the advantage of the efficiency improvement. Therefore, the issue of the combination of these two topics might become a valuable object to research.

4.3 A technology-function matrix of UPC709

There are 20 subclasses under UPC709, Table2 lists definitions of 12 subclasses in it and numbers of patent granted, total and in recent 3 years. The other 8 subclasses is neglected since their percentages in recent 3 years lower than 30%. It shows a popular research field but not so much as UPC707, 40% on the average, granted in recent 3 years.

Table 2 definitions and number of patent granted in UPC709

subclass	definition	Total	In recent 3 years(percentage)
201	Distributed data processing	3,594	1,282(35.7%)
204	Computer conferencing	4,883	2,318(47.5%)
212	Computer-to-computer direct memory accessing	190	76(40.0%)
217	Remote data accessing	3,709	1,475(39.8%)
220	Network computer configuring	2,008	854(42.5%)
223	Computer network managing	8,747	3,840(43.9%)
227	Computer-to computer session/connection establishing	3,708	1,580(42.6%)
230	Computer-to computer protocol implementing	3,580	1,278(35.7%)
238	Computer-to-computer data routing	2,012	671(33.4%)
245	Computer-to-computer data addressing	753	250(33.2%)
246	Computer-to-computer data modifying	976	337(34.5%)
248	Multicomputer synchronizing	377	141(37.4%)
	Total	34,537	14,102(40.8%)

After viewing selected patents in every subclass, we can get the functions and search queries as following:

(a)Enhance the ability of access

((improv* OR increas*) AND (access*)).DSC

(b)Enhance the ability of communicating

((improv* OR increas*) AND communicat*).DSC

(c)Enhance the efficiency of system

((improv* OR efficient*) AND (fast* or time)).DSC

(d)Reduce the associated costs

((decreas* OR reduc*) AND (cost* OR wast*)).DSC

(e)The Problem of Network Controlling

((internet OR network) AND (control*) AND (node*)).DSC

Using the search queries to each subclass, we can easily get the technology-function matrix of UPC709 as shown in Fig.5

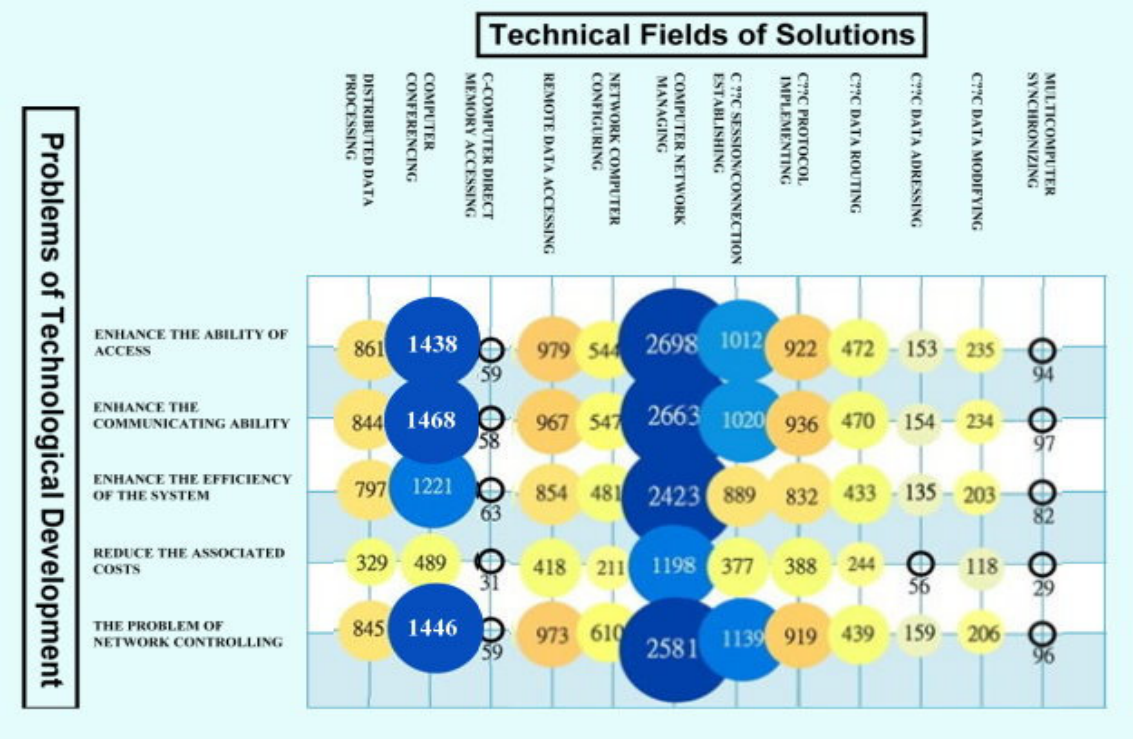

Fig.5 A technology-technical matrix of UPC 709 in recent 3 years

(Latest update: 2013.4.25)

The matrix shows that the most popular technical field is “computer network managing”, which have highest density in all problems.

It is interesting that computer conference get the second rank. Computer conference existed many years ago, but hasn’t become common usage. Researchers may believe that computer conference will prevail in cloud age, after enhancing the ability of access, communicating ability, and efficiency of the system.

Five problems have importance of the same class in every technology. The first problem “enhance the

ability of access" includes the issues of the accessing way and its improvement, which with the second problem "enhance the communicating ability", and the third problem "enhance the efficiency of the system", all are the priority problems have to be solve in cloud age. These three problems are overlapping in some technologies. The fourth problem "Reduce the associated cost" shows lower density comparing with other problems; we can say that this problem was not so favorable as the others.

If researchers try to find a star field in the future, "computer-to computer direct memory accessing" might be a good choice. It is the lowest density on the matrix, people need to make sufficient use of empty space in one set of memories, when increasing storage memory is not economical. "multicomputer synchronizing" also has low density, but it may need more budget to solve some technical problems in hardware, and the rewards are not so valuable.

5. Conclusions

Two illustrations of technical topics both have thousands of patents. It is clear that hardly to get a technology-function matrix by traditional method. The new approach in this paper can get the matrix more quickly and easily.

It will need a long term to finish a matrix even if classify thousands of patents one by one. The longer term it takes, the lower accuracy it is, because of new patent publications in this term. To get a matrix quickly means higher accuracy, and it is easy to update a matrix, just put the search queries to each subclass again.

This approach has another advantage, easy to get more detail matrix. For example, researcher may say Fig.4 is not clear enough in database and file access, highest intensity and thousands of patents of UPC707. We can expand this subclass to one dot subclass in UPC707/705 to get more detail technical means. One dot subclasses under data base and file access includes search engines, query optimization, post processing of search results, preparing data for information retrieval, etc. Applying the same search queries to each one dot subclass, we can quickly get all numbers of patents for all nodes.

The function expansion, y-axis in the matrix, is also convenient. Researchers in R&D may bring up some problems which are not shown in the matrix. We just need to organize the search queries for these problems to get amounts of patents for every technical means, a more detail matrix in function will come into being.

This approach relies on and be limited by patent classification system. UPC is used only by USPTO. If we need a matrix of the same topic showing patent granted by EPO or JPO, change to IPC is necessary, because no UPC in patent documents published by EPO or JPO. It has to choose a suitable classification system and database before working, UPC is more detail than IPC in computer software, that's the two illustrations used.

Patents are not only rich source of information, but also occupy private scope based on exclusive rights. How to retrieve valuable information from huge patent database for R&D is still an important task as patent increasing.

Acknowledgement

This study is conducted under the "Cloud computing systems and software development project(2/3)" of the Institute for Information Industry which is subsidized by the Ministry of Economy Affairs of the Republic of China.

References

Narin, F., & Olivastro, D.(1993), Patent citation cycles. *Library Trends*, 41(4), pp.700-709.

Narin, F.(1994), Patent Bibliometrics. *Scientometrics*, 30(1), pp.147-155.

Narin, F.(1995), Patent as indicators for the evaluation of industrial research output. *Scientometrics*, 34(3),pp.489-496.

Trajtenberg, M., Jaffe, A., & Henderson, R.(1997). University versus corporate patents: A window on the basicness of invention. *Economics of Innovation and New Technology*, 5(1),pp.19-50.

Ernst, H.(1998). Patent portfolios for strategic R&D planning. *Journal of Engineering and Technology Management*, 15(4), 279-308.

Yoon, Byung-Un, Yoon, Chang-Byung, and Part, Yong-Tae(2002), On the development and application of a self-organizing feature map-based patent map. *R&D managemet*, 32(4), pp.291-300.

Ernst, H.(2003). Patent information for strategic technology management. *World Patent Information*, 25(3), 233-242.

Fyfe, Colin(2007), Two Topographic Maps for Data Visualization, *Data Mining and Knowledge Discovery*, 14(2), pp.207-224.

Joo, Si Hyung, Lee, Keun(2010), Samsung's catch-up with Sony: an analysis using US patent data. *Journal of the Asia Pacific Economy*, 15(3), 271-287.

Schneider, Cédric(2011), How important are noncorporate patents? A comparative analysis using patent citations data. *Applied Economics Letters*, 18, pp. 865-871.

Son, Changho, Yongyoon Suh, Jeonghwan Jeon, and Yongtae Park(2012), Development of a GTM-based Patent Map for Identifying Patent Vacuums, *Expert Systems with Applications*, 39(3), pp.2489-2500.

Jeong, Seonkoo, Yoon, Byungun(2013), A Systemic Approach to Exploring an Essential Patent Linking Standard and Patent Maps: Application of Generative Topographic Mapping(GTM), *Engineering Management Journal*, 25(1),pp.48-57.