1C04

日本特許約６５万件の全数・全技術分野 サイエンス・リンケージの自動計測

○玉田俊平太（経産研），内藤祐介（人工生命研），児玉文雄（経産研／芝浦工大），
玄場公規（東大／芝浦工大），鈴木　潤（未来工研）

1. The main argument of the study

The authors constructed an original database which contains full text of granted Japanese patent since 1994. Then, from more than 650,000 patent data classified into about 600 categories, our tuned computer program extracted non-patent references and counted the numbers of them per each patent classification sub-class. That number shows the strength of the linkage between science and technology and therefore is called "science linkage." The science linkage among different patent classifications varied significantly. These data shown in figure 1 give us birds'-eye view of the relationship between science and technology, implies the creation process of technology in certain technological area is quite different from the ones in other technological areas, and also shows the possibility of informatics applied to economics.

2. The methodology applied

2.1 Construction of the database

An original patent database and specially designed workstation were constructed for this research. The database contains about 880,000 granted Japanese patents since 1994 and about 2,700,000 publication of unexamined patent applications to Japan Patent Office since 1993. Those data provided in more than 1,100 CD-ROMs were copied to the workstation, sort out by the international patent classifications, inventor's name, address, etc., and put in a large relational database.

2.2 Extraction of cited papers by hand

To improve the computer program which extracts cited references which consist of patent and scientific papers from the full-text of patent application, we need the reference data to score the performance of the program. First, we sampled 1,500 patents from the database and manually picked up cited patents and non-patent references, most of which were academic papers, from the full text of the patent applications and put them into another database. This database is named "teacher database" because this database would be the reference to score the performance of the automatic citation extraction program and tell us whether a modification of the program was successful or not, and whether the performance of the program was good enough compared to human.

2.3 Full automatic extraction of cited papers by computer program

Measuring the number of cited papers within patents and investigating the characteristics of them give us detailed data for further understanding of the process of innovation. At the same time, doing so requires much time and effort. It is practically very difficult to measure the science linkage of more than 600 category of international patent classification by manual labor. Therefore a program that would automatically extract the cited papers and patents from the full-text of the patent was built. The performance of the first version of the program was not satisfactory. Only 40% of the cited papers could be picked out. By tuning up the algorithm and key words, the performance of the program reached 97.8%. Finally, the science linkage of about 650,000 patents granted since 1995 was measured.

3. Results and discussion

Because of filling in the references in the front page of Japanese patent application is not mandatory and quite incomplete, it is necessary to create a program to automate the extraction of cited documents to comprehensively understand the linkage between technologies and sciences in Japan. As a result, a program with very high degree of precision (approximately 98 %) was successfully created. This enabled us to extract cited patents and papers automatically from each and every patent in our database and investigate science linkage comprehensively at the desired level of patent technology classification.

Of nearly 650,000 patents granted and published on Patent Gazettes from 1995 to 1999, we investigated science linkage of technology areas in about 600 categories. Figure 1 shows the average science linkage by subclass, the number of papers cited in patents measured for each of the subclasses (600 categories) divided by the number of patents that belong to that category. The strength of science linkage is quite different from certain technological field to other technological field. The science linkage of "C12N microorganisms or enzymes: compositions thereof" is the strongest and it is about 30 times stronger than the average. Most technological areas which have strong science linkage were biotechnology related.

Table 1 lists the top 20 subclasses by the average science linkage. For Japanese patents, the No. 1 area was "C12N microorganisms or enzymes: compositions thereof" with an average of 14.6, followed by "C07K organic chemistry, peptides" with an average of 12.2. The average for all was 0.5. These results are in line with the trend of science linkage of European patents studied by Michel *et al.* (2001). Specifically, a comparison made at subclass level reveals the top 3 to be the same. Six of the top 10 subclasses in European patents are also ranked in the top 10 subclasses in Japanese patents. These investigation results show that the linkage between technology and science differs due to the nature of each technology area regardless of the place where the technology was invented.

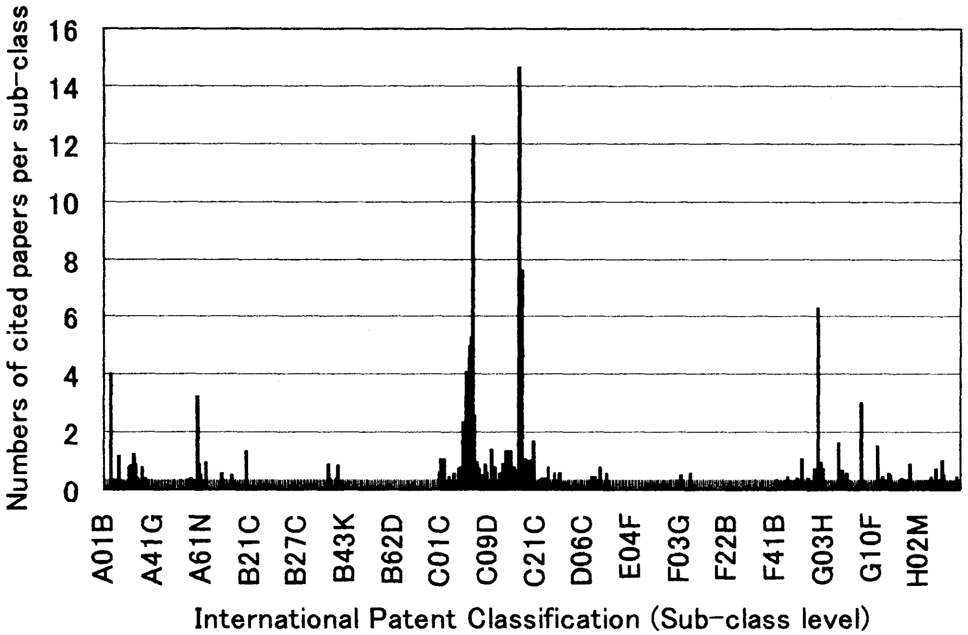
The science linkage indexes between different patent classifications differ significantly. It suggests that the processes of creating new technology differ from technology to technology. The technologies related to bio -technology were by far the closest to science.

Table 1: Top 20 Subclasses by Average Science Linkage

Subclass	Number of Patents	Average Science Linkage
C12N microorganisms or enzymes; compositions thereof; propagation, preservation, and maintenance of microorganisms, mutation or genetic engineering, culture medium	44425	14.6
C07K Peptides	18390	12.3
C12Q A method of measurement or a method of testing microorganisms or enzymes, a composition or a test paper for the method, the method for preparing the composition, and status response control in a microbiological or enzymological method.	5442	7.6
C12P A method for synthesizing the desired chemical material or chemical composition by using fermentation or an enzyme, or the method for separating an optical isomer from a racemic mixture	9617	7.0
G03C Photographic-sensitized material, photography (e.g., motion pictures, X-ray photography, multicolor photography, stereoscopic photography,) auxiliary photographic processing methods.	24018	6.3
C07J Steroids	1373	5.3
C07H Sugars, derivatives thereof, nucleosides, nucleotides, and nucleic acids	2837	5.0
C07D Heterocyclic compounds	24241	4.1
A01H New plants or a treatment to obtain them, propagation of plants by tissue culture	596	4.0
A61K Medical, dental, or cosmetic chemical preparations	23852	3.3
G09C Cipher containing necessary secrets or encryption for other purposes, or decrypting apparatus	233	3.0
C07G A compound having an unknown structure (in organic chemistry)	138	2.7
C07F Non-cyclic, carbon cyclic, or heterocyclic compounds [containing an element other than carbon, hydrogen, a halogen, oxygen, nitrogen, sulfur, selenium, or tellurium (in organic chemistry)]	3651	2.6
C08B Polysaccharides and derivatives thereof (organic high polymer compounds, manufacturing or scientific processing thereof, compositions based thereon)	1155	2.6
C07B General methods (in organic chemistry) or an apparatus thereof	468	2.3
C07C Non-cyclic compounds, carbon cyclic compounds (in organic chemistry)	15291	2.0
C14C Chemical processing of a raw hide, a pelt, or leather	51	1.6
G06E Optical calculating machine	56	1.6
G10L Analysis or synthesis of speech, speech recognition	1761	1.5
C09H Manufacturing method of glue or gelatin	18	1.4

Note: The technology categories enclosed in bold boxes are also ranked in the European top 10.

Fig. 1: Significant difference in science linkage among different technological classifications



References

Albert MB, Avery D, Narin F, McAllister P. 1991. Direct validation of citation counts as indicatours of industrially important patents. Research Policy 20: 251-259

Anderson J, Williams N, Seemungai D, Narin F, Olivastro D. 1996. Human Genetic Technology: Exploring the Links between Scinence and Innovation. Technology Analysis and Strategic Management 8(2): 135-156

Archibugi D. 1992. Patenting as an indicator of technological innovation: a review. Science and Public Policy 19(6)

Michel J, Bettels B. 2001. Patent citation analysis. Scientometrics 51(1): 185-201

Narin F. 1993. Patent Citation Analysis: The Strategic Application of Technology Indicators. Patent World(April): 25-31

Narin F. 1995. Inventive productivity. Research Policy 24: 507-519

Narin F, Hamilton K, Olivastro D. 1997. The increasing linkage between U.S. technology and public science. Research Policy 26: 317-330

Narin F, Olivastro D. 1988. Science Indicators: Their Use in Science Policy and Their Role in Science Studies. DSWO Press: The Netherlands

Narin F, Pinski G, Gee HH. 1976. Structure of the Biomedical Literature. Journal of the American Society for Information Science january-February