G&I Genomics & Informatics

pISSN 1598-866X eISSN 2234-0742
Genomics Inform 2012;10(4):271-274
http://dx.doi.org/10.5808/GI.2012.10.4.271

RESEARCH COMMUNICATION

Preliminary Study of Bioinformatics Patents and Their Classifications Registered in the KIPRIS Database

Hyun-Seok Park*

Department of Computer Science, Ewha Womans University, Seoul 120-750, Korea

Whereas a vast amount of new information on bioinformatics is made available to the public through patents, only a small set of patents are cited in academic papers. A detailed analysis of registered bioinformatics patents, using the existing patent search system, can provide valuable information links between science and technology. However, it is extremely difficult to select keywords to capture bioinformatics patents, reflecting the convergence of several underlying technologies. No single word or even several words are sufficient to identify such patents. The analysis of patent subclasses can provide valuable information. In this paper, I did a preliminary study of the current status of bioinformatics patents and their International Patent Classification (IPC) groups registered in the Korea Intellectual Property Rights Information Service (KIPRIS) database.

Keywords: bioinformatics patents database, International Patent Classification (IPC), Korea Intellectual Property Rights Information Service (KIPRIS), US Patent Classification (USPC)

Introduction

There are many ways that research gaps between science and technology can be identified. Bibliographic citation of both academic literature and patent documents is one of the solutions. Unfortunately, patent documents are considerably less accessible [1], compared to academic papers. Whereas vast amount of new information is made available to the public through patents, only a small set of patents are cited in academic papers [2].

Thus, the essential problem arising from citing patents related to bioinformatics is to devise a classification methodology that identifies the relevant patent and only those patents. The suitability of the methodology depends on the purpose of the analysis. However, no single methodology is ideal for identifying patents for a convergence technology, such as bioinformatics. By its nature, it is at the intersection of a number of patent classes.

Accordingly, this paper tentatively employs several methodologies simultaneously to analyze the current status of bioinformatics and to find relevant subclasses, using the Korea Intellectual Property Rights Information Service (KIPRIS, http://www.kipris.or.kr/) patent search system. Initially, a multiple-word search is undertaken of these patents' titles and abstracts. The selected lists are then tabulated, based on International Patent Classification (IPC) subclasses together with some known bioinformatics subclasses. These were then reviewed manually to establish a revised list of bioinformatics patents. These steps have been repeated several times to establish the list of relevant IPC subclasses that appear most frequently in KIPRIS bioinformatics patents.

Outline of Patent Search Methodologies

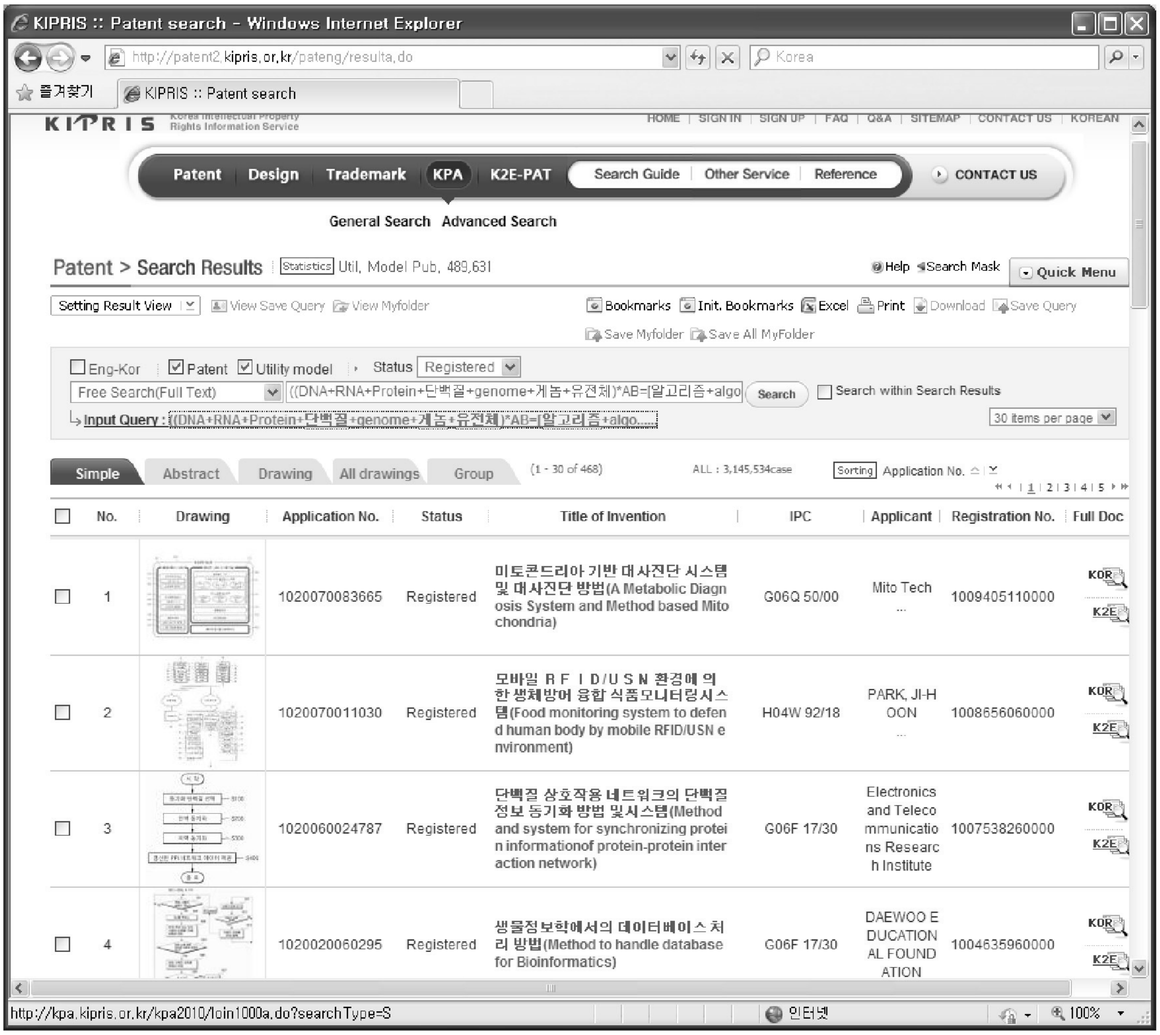
KIPRIS is an internet-based patent document search service that is available to the public free of charge. Korea Institute of Patent Information has been providing KIPRIS since 1996 on behalf of Korean Intellectual Property Office (KIPO). It covers publications of Korean Intellectual Property Rights (IPR) applications, legal status information, and trial information. Using the KIPRIS system, the following simple query (documents, containing and the keyword DNA, RNA, and protein, and containing the keyword 'algorithm, database, visualization' in the abstract),

(DNA + RNA + protein + genome) * AB = [algorithm + database + visualization],

would retrieve 468 registered patents from the KIPRIS database, as in Fig. 1.

Received October 26, 2012; Revised November 8, 2012; Accepted November 13, 2012

***Corresponding author:** Tel: +82-2-3277-2831, Fax: +82-2-3277-2306, E-mail: neo@ewha.ac.kr



http://patent2.kipris.or.kr/pateng/resulta.do

Korea

Patent > Search Results | Statistics Util. Model Pub. 489,631

□ Eng-Kor ☑ Patent ☑ Utility model Status Registered

Free Search(Full Text) ((DNA+RNA+Protein+단백질+genome+게놈+유전체)*AB=[알고리즘+algo

□ Search within Search Results

Input Query : ((DNA+RNA+Protein+단백질+genome+게놈+유전체)*AB=[알고리즘+algo.....

30 items per page

(1 - 30 of 468) ALL : 3,145,534case Sorting Application No.

1 | 2 | 3 | 4 | 5

□	No.	Drawing	Application No.	Status	Title of Invention	IPC	Applicant	Registration No.	Full Doc
□	1		1020070083665	Registered	미토콘드리아 기반 대사진단 시스템 및 대사진단 방법(A Metabolic Diagnosis System and Method based Mitochondria)	G06Q 50/00	Mito Tech ...	1009405110000	KOR K2E
□	2		1020070011030	Registered	모바일 R F I D/U S N 환경에 의한 생체방어 융합 식품모니터링시스템(Food monitoring system to defend human body by mobile RFID/USN environment)	H04W 92/18	PARK, JI-HOON ...	1008656060000	KOR K2E
□	3		1020060024787	Registered	단백질 상호작용 네트워크의 단백질 정보 동기화 방법 및시스템(Method and system for synchronizing protein informationof protein-protein interaction network)	G06F 17/30	Electronics and Telecommunications Research Institute	1007538260000	KOR K2E
□	4		1020020060295	Registered	생물정보학에서의 데이터베이스 처리 방법(Method to handle database for Bioinformatics)	G06F 17/30	DAEWOO EDUCATIONAL FOUNDATION	1004635960000	KOR K2E

http://kpa.kipris.or.kr/kpa2010/loin1000a.do?searchType=S

Fig. 1. Query result of (DNA + RNA + protein + genome) * AB = [algorithm + database + visualization].

The number of patents can be further reduced by manual screening. Likewise, a common methodology is to undertake keyword searches of the patent title, abstract, or claim, as it is not practical to review the complete claims, specifications, and drawings of related patents.

However, this approach of a pure keyword search has the inherent problem of both including irrelevant patents along with bioinformatics patents and of potentially missing relevant bioinformatics patents. For example, the number of patent documents that have the keyword 'bioinformatics' in the KIPRIS database is 446, as of November 18, 2012. However, the majority of them can not be classified as bioinformatics patents. On the other hand, there is no warranty that actual bioinformatics patents would at least have the keyword 'bioinformatics' in their documents. Nonetheless, this approach illustrates the converging nature of the technologies comprising bioinformatics and represents a good starting point for examining the cross-classification of bioinformatics patents.

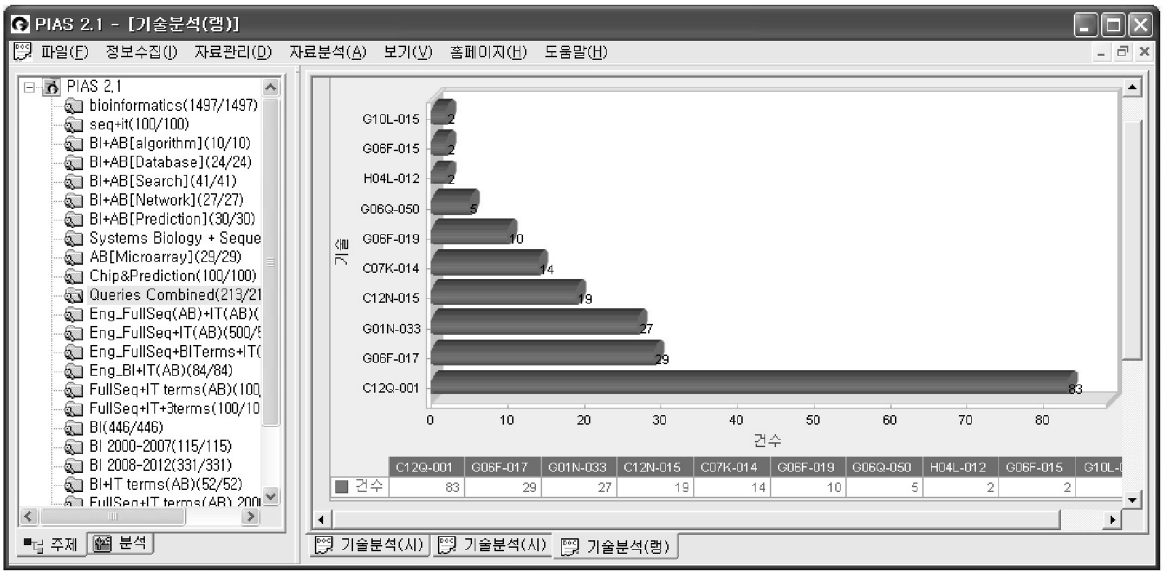
An alternative approach to expand a patent search beyond pure keywords is to include classification information and to use a definition of bioinformatics patents based on several patent subclasses in which bioinformatics patents are expected to be clustered. A good starting point to choose relevant IPC (http://www.wipo.int/classifications/ipc/) bioinformatics classes would be to reference preselected bioinformatics classes. For example, the US Patent Classification (USPC) bioinformatics patent subclasses (382/125, 702/19, 702/20, 702/21, 703/11, and 703/12) suggested in [3, 4] could be used, and they have been mapped into IPC groups G06K 9/00, G01N 33/48, G01N 33/50, G01N 31/00, and G06G 7/58, as in Table 1, by using the USPC-to-IPC reverse concordance system (http://www.uspto.gov/patents/resources/classification/index.jsp).

Then, queries using various keywords combined with the groups in Table 1 were initially performed, using the Patent

Table 1. Mapping US Patent Subclasses Described in Rasmussen [4] to IPC subclasses and groups

USPC class	USPC subclass	USPC description	IPC subclass	IPC group	IPC description
382	125	Image analysis and relates in particular to 'DNA or RNA pattern reading'	G06K	9/00	Methods or arrangements for reading or recognizing printed or written characters or for recognizing patterns, e.g., fingerprints
702	19	Data processing: measuring, calibrating or testing—biological or biochemical	G01N	33/48	Biological material, e.g., blood, urine; hemocytometers
702	20	As for 702/19 but relating specifically to 'gene sequence determination'	G01N	33/50	Chemical analysis of biological material, e.g., blood, urine; testing involving biospecific ligand binding methods; immunological testing
702	21	As for 702/19 but relating specifically to 'cell contour shape or size analysis'	G01N	31/00	Investigating or analyzing non-biological materials by the use of the chemical methods specified in the subgroups
703	11	Data processing: simulating nonelectrical device or system—biological or biochemical	G06G	7/58	Analog computers for specific processes, systems, or devices, e.g., simulators for chemical processes
703	12	Data processing: simulating nonelectrical device or system—chemical	G06G	7/58	Analog computers for specific processes, systems, or devices, e.g., simulators for chemical processes

The US Patent & Trademark Office has developed a concordance between its USPC system and the IPC. The United States Patent Classification To International Patent Classification Concordance is intended only as a guide or tool providing information that allows the user to quickly access the correct, general classification area of the IPC. However, this approach has a limitation, due to the differences in the philosophies of the USPC and IPC systems, and Table 1 does not complete the one-to-one relationship.
IPC, International Patent Classification; USPC, US Patent Classification.

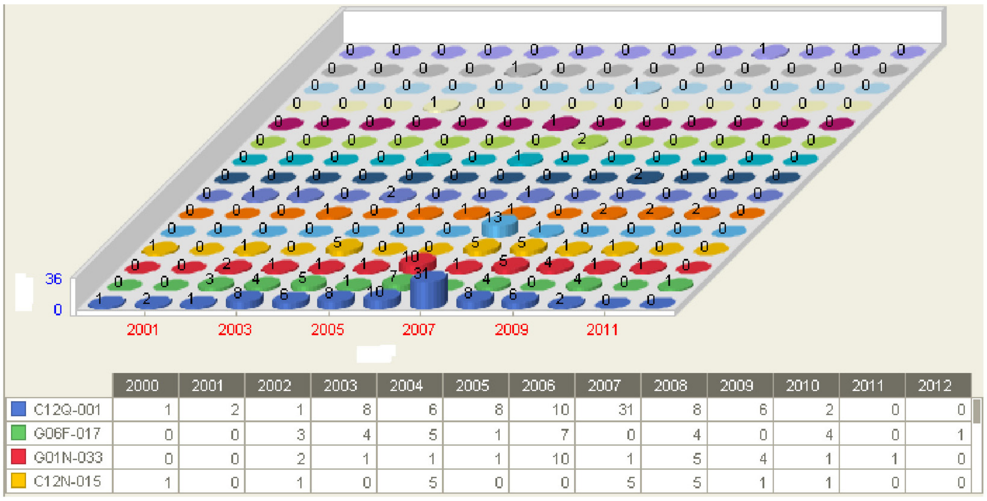


- PIAS 2.1
  - bioinformatics(1497/1497)
  - seq+it(100/100)
  - BI+AB[algorithm](10/10)
  - BI+AB[Database](24/24)
  - BI+AB[Search](41/41)
  - BI+AB[Network](27/27)
  - BI+AB[Prediction](30/30)
  - Systems Biology + Seque
  - AB[Microarray](29/29)
  - Chip&Prediction(100/100)
  - Queries Combined(213/21
  - Eng_FullSeq(AB)+IT(AB)(
  - Eng_FullSeq+IT(AB)(500/5
  - Eng_FullSeq+BITerms+IT(
  - Eng_BI+IT(AB)(84/84)
  - FullSeq+IT terms(AB)(100
  - FullSeq+IT+3terms(100/10
  - BI(446/446)
  - BI 2000-2007(115/115)
  - BI 2008-2012(331/331)
  - BI+IT terms(AB)(52/52)
  - FullSeq+IT terms(AB) 200

	C12Q-001	G06F-017	G01N-033	C12N-015	C07K-014	G06F-019	G06Q-050	H04L-012	G06F-015	G10L-0
■ 건수	83	29	27	19	14	10	5	2	2	

Fig. 2. Combining the results of various queries, using Patent Information Analysis System (PIAS, http://www.kipris.or.kr/kor/add_service/service/service_pias_1.jsp): combining biological keywords (dna? + rna? + "nucleic acid" + protein? + polynucleotid? + nucleotid? + deoxyribonuc? + ribonucleo? + "amino acid" + genom? + proteom? + SNP + "Single Nucleotide Polymorphism" + (metabolic ^2 pathway?) + (signal? ^2 pathway?) + gene + oligonucleotid? + microarray + (DNA ^2 chip?)) with informatics keywords (computer? + software? + program? + algorithm? + data? + protocol? + internet? + interfac? + proteomics + bioinformatics + ((text? + data?) ^2 (Mining? + retrieval?)) + mining? + ontology? + clustering? + statistic? + automat? + visualiz?) in various ways.

Information Analysis System (PIAS, http://www.kipris.or.kr/kor/add_service/service/service_pias_1.jsp) system, as in Fig. 2. New subclasses or groups linked to bioinformatics can be identified by using various combinations of query terms, as is listed in the left window of Fig. 2. These were then reviewed manually to establish a revised list of



	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
🟦 C12Q-001	1	2	1	8	6	8	10	31	8	6	2	0	0
🟩 G06F-017	0	0	3	4	5	1	7	0	4	0	4	0	1
🟥 G01N-033	0	0	2	1	1	1	10	1	5	4	1	1	0
🟨 C12N-015	1	0	1	0	5	0	0	5	5	1	1	0	0

Fig. 3. Yearly distribution of patent groups in Fig. 2.

bioinformatics patents. These steps were repeated several times to extend the list of relevant IPC groups that appear most frequently in the KIPRIS database. The query terms can be dependent on the characteristics of the intellectual property associated with patents, such as algorithms, data contents, network analysis, and user interfaces.

The most frequently occurring groups based on the query term containing the keyword 'bioinformatics' in addition to the combination of biological sequence keywords in the abstract was C07K 14/47 (peptides having more than 20 amino acids from mammals), followed by C07K 16/08, C12N 15/11, C12N 15/31, C12Q 1/68, C12N 15/10, G01N 33/68, G06F 17/30, and G06F 19/00. The most frequently occurring groups containing informatics keywords such as visualization, algorithm, software, and database, together with some biological sequences was C12N 15/00 (mutation or genetic engineering; DNA or RNA concerning genetic engineering), followed by C12N 15/63, C12Q 1/68, G01N 33/68, G06F 17/30, and G06F 19/00. The most frequently occurring groups related to database contents or networks was G06F 17/30 (digital computing or data processing equipment or methods, specially adapted for specific functions; information retrieval, database structures therefor).

Combining the results of various queries, I finally selected 213 patents that are most related to bioinformatics. The main groups with the highest frequency were C12Q 001 (measuring or testing processes involving enzymes or micro-organisms) and G06F 017 (digital computing or data processing equipment or methods, specially adapted for specific functions), followed by G01N 033, C12N 015, C07K 014, G06F 019, G06Q 050, H04L 012, G06F 015, and G10L 015. Fig. 3 shows the yearly distribution of these groups.

Conclusion

By identifying the relevant subclasses of bioinformatics patents, the research gaps between science and technology can be identified. In this paper, I did a preliminary study of the current status of bioinformatics patents registered in the KIPRIS database. I identified some of the subclasses and groups related to bioinformatics patents, but the work presented here should be regarded as preliminary, and a more detailed analysis using text-mining techniques will follow in the near future.

Acknowledgments

I would like to thank patent attorney, Kwang-Ho Kim at Ewha Womans University, for valuable comments. This work was partially supported by G-SMATT.

References

1. Eisinger D, Wächter T, Bundschus M, Wieneke U, Schroeder M. Analysis of MeSH and IPC as a prerequisite for guided patent search. *Bio-Ontologies* 2012 Oct 4 [Epub]. http://bio-ontologies.knowledgeblog.org/346.
2. Glänzel W, Meyer M. Patents cited in the scientific literature: an exploratory study of 'reverse' citation relations. *Scientometrics* 2003;58:415-428.
3. Patel P. 'UK performance in biotechnology-related innovation: an analysis of patent data'. Final report prepared for the assessment unit of the UK Department of Trade and Industry. Brighton: SPRU, University of Sussex, 2003.
4. Rasmussen B. Commercialisation processes in bioinformatics: analysis of bioinformatics patents. CSES Working Paper No. 26. Melbourne: Centre for Strategic Economic Studies, Victoria University, 2005.