Copyright, Patent and Trade Secret on Digital Libraries: Current Issues and Future Trends

Hideyasu Sasaki[1,2]

[1] Ritsumeikan University, Department of Information Science and Engineering
6-4-10 Wakakusa, Kusatsu, Shiga, 525-0045 Japan
[2] Attorney-at-Law, New York State Bar
hsasaki@alumni.uchicago.edu

Abstract. In this paper, we discuss current issues and future trends on intellectual properties of digital libraries by interpreting legal concepts in engineering manner as a reference to Asia-Pacific DL researchers and practitioners. First, we discuss problems on copyright entities in digital libraries and patent objects in their retrieval mechanisms. Second, we formulate the conditions of copyrightability on the multimedia databases as digital libraries and the patentability on the parameter setting components in retrieval mechanisms. Third, we discuss a new direction for protecting numerical parametric information as trade secret embedded in the patentable parameter setting components.

1 Introduction

Digital library is the global information infrastructure in the networked society [1]. The intellectual property protection of digital libraries is a critical issue in the digital library community, which demands frameworks for recouping their investment in database design and system implementation. A digital library, as an information system, consists of digital contents stored in databases and their retrieval mechanisms. Intellectual property law gives incentive to advance appropriate investment in database design and implementation with two conventional types of intellectual property protection: copyright and patent [2,3]. Nevertheless, present legal studies are not satisfactory as the source of technical interpretation of the intellectual properties regarding digital libraries. The intellectual property protection of the digital libraries demands clear and concise frameworks.

In this paper, we would describe the technical and legal issues on digital library as the objects of copyright, patent and trade secret that have not been discussed with sufficient attention at the present. The principal concern of this paper is to present the conditions of copyrightability on the multimedia databases and the patentability on the parameter setting components in retrieval mechanisms with the directions for protecting numerical parametric information in the parameter setting components as trade secret. Our secondary concern is to provide researchers and practitioners in the DL community with legal references on the concepts, issues, trends and frameworks of intellectual property protection regarding digital libraries in engineering manner.

The scope of this paper is restricted within the current standard of laws and cases in transnational transaction and licensing of intellectual properties regarding digital library. Cultural diversity in the Asia-Pacific region is a source of legislative differences

in intellectual property laws, though those countries join international trade agreements for intellectual property rights. We discuss the harmonized IP law standard regarding digital library with which the Asian-Pacific countries are able to keep up with the foregoing countries.

2   Background

In this section, we discuss three issues on the intellectual property protection regarding digital libraries. The first issue is the copyright protection of databases into which the digital contents are stored as form digital libraries. The second issue is the patent protection of the retrieval mechanisms of database systems or digital library systems. The third issue is the trade secret on the numerical parametric values for retrieval operations in the parameter setting components.

2.1   Copyright on Digital Libraries

U.S. Copyright Act [4] defines that a compilation or assembling of individual contents, *i.e.*, preexisting materials or data, is a copyrightable entity as an original work of authorship. Gorman and Ginsburg [5], and Nimmer, et al. [6] state that a compilation is copyrightable as far as it is an "original work of authorship that is fixed in tangible form".

Digital library systems consist of digital contents which are indexed and stored in databases for appropriate retrieval operations and the retrieval mechanisms which are optimized and applied to object domains of those databases. The entire database is copyrightable in the form of a component of "contents-plus-indexes" while static indexes or metadata are fixed to digital contents in a tangible medium of repository, *i.e.*, database. Static indexes or metadata represent a certain kind of categorization of the entire content of each database (See Fig. 1).

The originality on the categorization makes each database copyrightable as is different from the mere collection of its individual contents. What kind of categorization should be original to constitute a copyrightable compilation on the database? The Court of American Dental Association v. Delta Dental Plan Association [7] determined that minimal creativity in compilation sufficed this requirement of originality on databases. Any standard or framework on the requirement is not clear in the technical or engineering meanings. A uniform framework on the categorization regarding indexes or metadata of databases must be formulated in engineering manner.

The European Union has legislated and executed a scheme for protecting a database including its content *per se*, known as the sui generis right of database protection [8,9,10]. That European scheme shares the same issue on the originality regarding the categorization of digital contents in databases.

2.2   Patent on Digital Library Systems

U.S. Patent Act [11] defines that a data-processing process or method is patentable subject matter in the form of a computer-related invention, *i.e.*, a computer program.
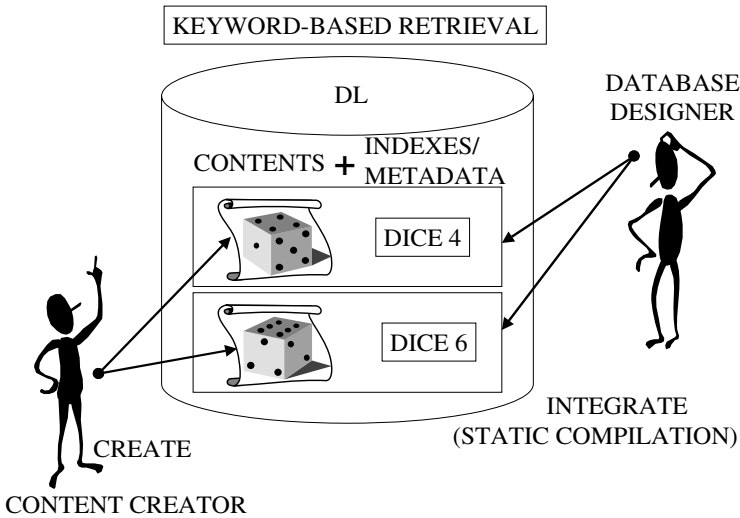
Fig. 1. Formulation for copyrighting digital libraries

The computer program is patentable as far as the "specific machine . . . . . . produce(s) a useful, concrete, and tangible result . . . for transforming . . . " physical data ("physical transformation") [12].

The computer-related inventions often combine means for data-processing, some of which are prior disclosed inventions. A retrieval mechanism in a digital library system consists of a number of "processes", *i.e.*, methods or means for data processing in the form of combination of computer programs. A set of programs focuses on image processing, while another set of programs operates text mining, for example.

Meanwhile, the processes in a retrieval mechanism of a digital library system comprise means or components for parameter setting which is adjusted to retrieve specific kinds of digital contents, for example, images in certain domains. The problem is that which process is to realize technical advancement (nonobviousness) on its combination of the prior arts and is to be specific/enable on its parameter setting. These two issues are emerging problems in the advent of digital library systems. Uniform frameworks on the novel combination and the specific parameter setting must be formulated in engineering manner, respectively.

2.3 Trade Secret in Parameter Setting

Another emerging problem is discussed on the parameter setting of retrieval mechanisms. Patent application on the parameter setting components demands applicants as developers to make public the detailed know-how on the best range of parametric values in practice.

The discovery of those parametric values needs considerable pecuniary investment in research and development. That kind of knowledge should be kept covered in the form of trade secret but not be open in public via patent application. The DL community

demands a framework that determines which parameter setting component should be patentable and kept secret regarding digital library systems.

3   Frameworks for Intellectual Property Protection

In this section, we outline the frameworks for intellectual property protection regarding digital library systems: copyrightable database, patentable retrieval mechanism and embedded trade secret on numerical parametric values for retrieval.

3.1   Digital Library as Copyrightable Entity

Our framework for copyrighting the digital library determines which type of database should be copyrightable in the form of a component of contents-plus-indexes [13,14,15]. The collection of static indexes and individual contents forms a component of contents-plus-indexes. That component identifies the entire content of each database, as is a static and copyrightable compilation. Copyrightable compilation is to be of sufficient creativity, *i.e.*, originality in the form of a component of contents-plus-indexes.

The set of conditions on the original categorization regarding indexes or metadata is formulated as below [14,15]: A categorization regarding indexes or metadata is original only when

1. The type of indexes or metadata accepts discretionary selection in the domain of a problem database; otherwise,
2. The type of taxonomy regarding indexes or metadata accepts discretionary selection in the domain of a problem database.

A typical case of non-original categorization is a photo film album database which has indexes of consecutive numbers. That case does not accept any discretion in the selection of the type of indexes or metadata, or the type of taxonomy. The photo film album database uses its respective film numbers as indexes for its retrieval operations. The taxonomy of the indexes is only based on the consecutive numbering without any discretion in its selection of the type of indexes or taxonomy regarding a digital library.

Meanwhile, the discretionary selection of the type of indexes or metadata, or taxonomy constitutes copyrightable compilation of minimal creativity, *i.e.*, originality on the categorization regarding indexes or metadata. A typical case of discretionary selection of the type of indexes or metadata is the web document encyclopedia as a digital library. Suppose that a database restores pictures of starfish which are manually and numerically numbered by day/hour-chronicle interval that is based on their significant life stages from birth to death. That database is to be an original work of authorship as a copyrightable compilation in the form of a component of contents-plus-indexes. That database of discretionary type of numbering or indexing is an original, *i.e.*, copyrightable database.

3.2   Digital Library System as Patentable Mechanism

Our framework for patenting the retrieval mechanisms of digital library system determines which type of retrieval mechanism should be patentable in the form of a
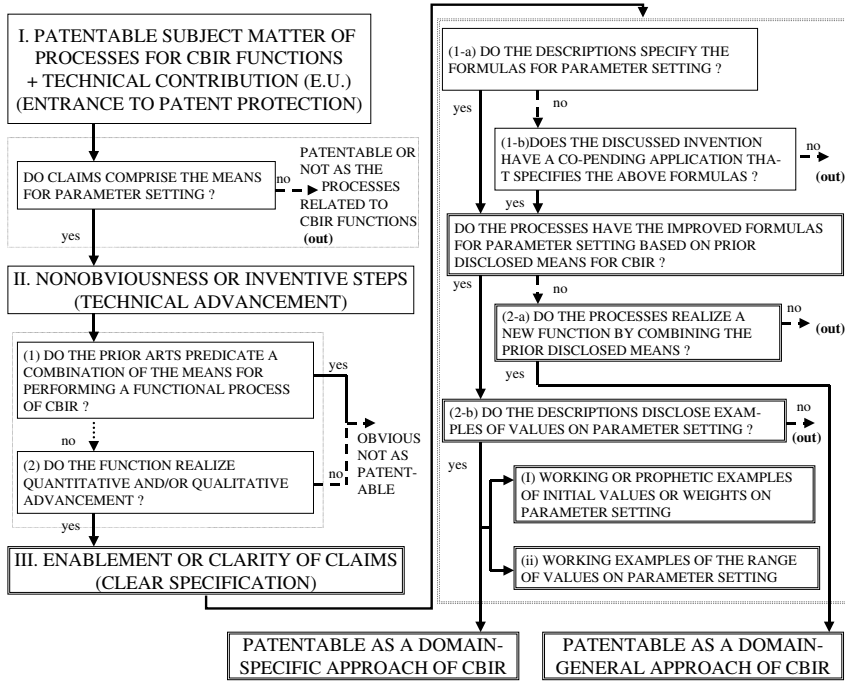
Fig. 2. Formulation for patenting the retrieval processes

component of novel combination of prior disclosed processes and/or a component of specific parameter setting (See Fig. 2) [16,17,18,15]. The frameworks focus on the following three requirements for patentability: "patentable subject matter" (entrance to patent protection), "nonobviousness" (technical advancement) and "enablement" (specification) [19].

The requirement for nonobviousness on the combination of the processes for data-processing as the retrieval mechanism in a digital library system is listed as below [18]:

1. The processes for performing a retrieval mechanism must comprise the combination of prior disclosed means to perform certain mechanism which is not predicated from any combination of the prior arts; in addition,
2. The processes for performing a retrieval mechanism must realize quantitative and/ or qualitative advancement.

Otherwise, the discussed processes are obvious so that they are not patentable as the processes for performing a retrieval mechanism.

First, a combination of prior disclosed means should not be "suggested" from any disclosed means "with the reasonable expectation of success" [20]. Second, its asserted function on the discussed mechanism must be superior to the conventional functions which are realized in the prior disclosed or patented means in the field of the retrieval mechanism of digital library system. On the latter issue, several solutions for performance evaluation are proposed and applicable. Another general strategy is restriction

of the scope of problem claims into a certain narrow field to which no prior arts have been applied. This claiming strategy is known as the local optimization of application scope.

A component for parameter setting realizes thresholding operations in the form of a computer program with a set of ranges of parametric values. In retrieval mechanisms, parametric values determine, as thresholds, which candidate image is similar to an exemplary requested image by computation of similarity of visual features [21,22,23,24]. That parameter setting component is to be a computer-related invention in the form of computer program as far as that parameter setting is sufficiently specified to enable a claimed invention or retrieval mechanism [25].

The requirement for enablement on the parameter setting component of the retrieval mechanism in a digital library system is listed as below [18]:

(1-a) The descriptions of the processes for performing a retrieval mechanism must specify the formulas for parameter setting; otherwise,
(1-b) the disclosed invention of the processes should have its co-pending application that describes the formulas in detail; in addition,
(2-a) the processes must perform a new mechanism by a combination of the prior disclosed means; otherwise,
(2-b) the processes should have improved formulas for parameter setting which is based on the prior disclosed means for performing a retrieval mechanism, and also should give examples of parametric values on parameter setting in descriptions.

For 2-b, the processes must specify the means for parameter setting by "giving a specific example of preparing an" application to enable those skilled in the arts to implement their best mode of the processes without undue experiment [26,27]. U.S. Patent and Trademark Office [25,28] suggested that the processes comprising the means, *i.e.*, the components for parameter setting must disclose at least one of the following examples of parametric values on parameter setting:

(i) Working or prophetic examples of initial values or weights on parameter setting;
(ii) Working examples of the ranges of parametric values on parameter setting.

The "working examples" are parametric values that are confirmed to work at actual laboratory or as prototype testing results. The "prophetic examples" are given without actual work by one skilled in the art.

3.3   A Simulation Example for the Formulated Procedural Diagram

The proposed formulation in Fig. 2 should be clear with its application to an exemplary digital library system. We apply it to "Virage Image Retrieval"(VIR), which was developed in the early 1990s as a typical content-based retrieval of visual objects stored in digital image database systems. VIR is an indexing method for an image search engine with "primitives", which compute similarity of visual features extracted out of typical visual objects, e.g., color, shape and texture of images. VIR evaluates similarity of images with ad hoc weights, *i.e.*, parametric values, which are given to the parameter setting components for correlation-computation, by user-preference. Its claims consist of "function containers" as means-plus-functions for feature extraction and similarity

computation. Its first claim, as described below, constitutes the primitives as the means-plus-functions. Those primitives realize a domain-general approach of CBIR by the formulas on parameter setting.

VIR Claim # 1.
A search engine, comprising: a function container capable of storing primitive functions; . . . a primitive supplying primitive functions . . . . . ., wherein the primitive functions include an analysis function . . . . . . of extracting features from an object . . . .

First in Fig. 2, on its patentable subject matter, its retrieval processes consisting of the formula for parameter setting are to be determined as patentable subject matter in the form of computer programs. Those data-processing processes generate physical transformation on a specific machine, *i.e.*, a computer memory with certain classification results. Second, on its nonobviousness, those data-processing processes are inventive steps that consist of combinations of the prior arts on thresholding functions as implemented in the integration of classification based on similarity computation, visual feature extraction and automatic indexing techniques. Those combinations are not predicated from any conventional keyword-based retrieval technique. Third, on its enablement, VIR's description of preferred embodiments gives its clear specification on the formulas for parameter setting that realizes a domain-general approach of CBIR that was a brand new technology at the time.

VIR Description . . . . . .
For primitives having multiple dimensions, ..., An equation for an exemplary Euclidean metric is as follows.
Primitive design. A primitive encompasses a given feature's representation, extraction, and comparison function. . . . . . . The constraints are as follows: Primitives, in general, map to cognitively relevant image properties of the given domain. The formulation should take advantage of a threshold parameter (when available), ...

The retrieval mechanisms of digital library systems are patentable in the form of components of novel combinations of prior disclosed processes and/or components of specific parameter settings while they are to satisfy the above conditions.

3.4   Embedded Trade Secret in Parameters

It is necessary to prepare a framework that determines how and which part of parameter setting components should take the form of trade secret. The problem is how to interpret the "working examples" of initial values or weights on parameter setting and the ranges of parametric values.

The requirement for patenting parameter setting components as computer-related inventions demands inventors to make public their discovered "working examples" on those parameter values: initial values or ranges. The practice in patent application, nonetheless, does not always force applicants to disclose to examiners complete and perfect evidences on those initial values or ranges of parametric values, but those values as should work in their best mode at the present art.
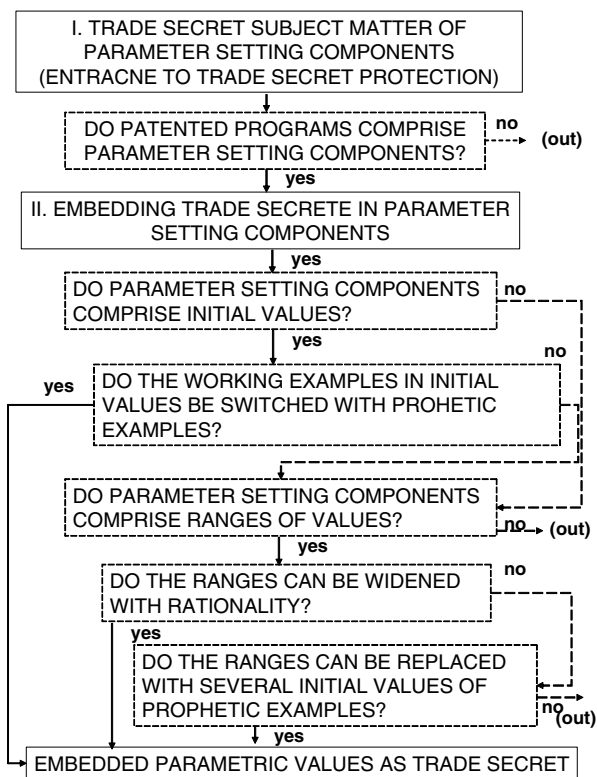
Fig. 3. Formulation for embedding trade secret in parameter setting components

In the reality of application practice, inventors have three choices for embedding trade secrets on their know-how of parametric values in the forms of patentable parameter components:

1. On the initial values, their prophetic examples should be disclosed in patent application, instead of working examples;
2. On the ranges of parametric values, those ranges should be widened as possible at the best but not complete mode;
3. Otherwise, the ranges of parametric values should be replaced with several initial values of prophetic examples.

Fig. 3 shows when a certain set of numerical parametric values in parameter setting components should be embedded as trade secret even those components have been patented as claimed inventions.

4   Conclusions

In this article, we have discussed issues on intellectual property protection regarding digital library systems which consist of indexed multimedia digital contents in

databases and retrieval mechanisms with numerical parametric values. We have presented the frameworks for copyrighting the databases of digital library systems in the forms of components of contents-plus-indexes, patenting the retrieval mechanisms of digital library systems in the forms of combinations of processes and/or components of parameter settings and for embedding the trade secret of numerical parametric values in parameter setting components in retrieval mechanisms.

Acknowledgements

References

1. Borgman, C.L.: From Gutenberg to the Global Information Infrastructure: Access to Information in the Networked World. Digital Libraries and Electronic Publishing. MIT Press, Cambridge, MA (2000)
2. Jakes, J.M., Yoches, E.R.: Legally Speaking: Basic Principles of Patent Protection for Computer Science. Communications of the ACM 32(8), 922–924 (1989)
3. Junghans, C., Levy, A.: Intellectual Property Management: A Guide for Scientists, Engineers, Financiers, and Managers. John Wiley & Sons, Hoboken, NJ (2006)
4. U.S. Copyright Act.: 17 U.S.C. Sec. 101, & 103 (2005)
5. Gorman, R.A., Ginsburg, J.C.: Copyright: Cases and Materials, 6th edn. University casebook series. The Michie Company, Charlottesville, NC (2002)
6. Nimmer, M.B., Marcus, P., Myers, D.A., Nimmer, D.: Cases and Materials on Copyright & Other Aspects of Entertainment Litigation Including Unfair Competition, 7th edn. LexisNexis, Dayton, OH (2006)
7. American Dental Association v. Delta Dental Plan Association.: 126 F.3d 977 (7th Cir. 1997)
8. Reinbothe, J.: The Legal Protection of Non-creative Databases. In: Proc. of the Database Workshop of the International Conference of Electronic Commerce and Intellectual Property. WIPO. Geneva, Switzerland (September 14–16 1999)
9. Samuelson, P.: Legally Speaking: Legal Protection for Database Content. Communications of the ACM 39(12), 17–23 (1996)
10. Aplin, T.: Copyright Law in the Digital Society: The Challenges of Multimedia. Hart Publishing, Oxford, U.K (2005)
11. U.S. Patent Act.: 35 U.S.C. Sec. 101, 103, & 112 (2005)
12. In re Alappat.: 33 F.3d 1526, 31 U.S.P.Q.2d 1545 (en banc) (Fed. Cir. 1994)
13. Sasaki, H., Kiyoki, Y.: A Proposal for Digital Library Protection. In: Proc. of the 3rd ACM/IEEE-CS Joint Conference on Digital Libraries, Houston, TX, May 27–31, p. 392. IEEE Computer Society Press, Los Alamitos (2003)
14. Sasaki, H., Kiyoki, Y.: Copyrighting Digital Libraries from Database Designer Perspective. In: Chen, Z., Chen, H., Miao, Q., Fu, Y., Fox, E., Lim, E.-p. (eds.) ICADL 2004. LNCS, vol. 3334, pp. 11–14. Springer, Heidelberg (2004)
15. Sasaki, H., Kiyoki, Y.: Multimedia Digital Library as Intellectual Property. In: Design and Usability of Digital Libraries: Case Studies in the Asia Pacific, pp. 238–253. Idea Group Press (2005)

16. Sasaki, H., Kiyoki, Y.: Patenting Advanced Search Engines of Multimedia Databases. In: Lesavich, S. (ed.) Proc. of the 3rd International Conference on Law and Technology. International Society of Law and Technology (ISLAT), Cambridge, MA, November 6–7, pp. 34–39. Acta Press, Anaheim, Calgary, Zurich (2002)
17. Sasaki, H., Kiyoki, Y.: Patenting the Processes for Content-based Retrieval in Digital Libraries. In: Lim, E.-p., Foo, S.S.-B., Khoo, C., Chen, H., Fox, E., Urs, S.R., Costantino, T. (eds.) ICADL 2002. LNCS, vol. 2555, pp. 471–482. Springer, Heidelberg (2002)
18. Sasaki, H., Kiyoki, Y.: A Formulation for Patenting Content-based Retrieval Processes in Digital Libraries. Journal of Information Processing and Management 41(1), 57–74 (2005)
19. Merges, R.P., Duffy, J.F.: Patent Law and Policy: Cases and Materials, 3rd edn. LexisNexis, Dayton, OH (2002)
20. In re Dow Chemical Co.: 837 F.2d 469, 473, 5 U.S.P.Q.2d 1529, 1531 (Fed. Cir. 1988)
21. Rui, Y., Huang, T.S., Chang, S.F.: Image Retrieval: Current Techniques, Promising Directions and Open Issues. Journal of Visual Communication and Image Representation 10(4), 39–62 (1999)
22. Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based Image Retrieval at the End of the Early Years. IEEE Trans. on Pattern Analysis and Machine Intelligence 22(12), 1349–1380 (2000)
23. Yoshitaka, A., Ichikawa, T.: A Survey on Content-based Retrieval for Multimedia Databases. IEEE Trans. on Knowledge and Data Engineering 11(1), 81–93 (1999)
24. Deb, S.: Multimedia Systems and Content-based Retrieval. Idea Group Inc., Hershey, PA (2004)
25. U.S. Patent and Trademark Office.: Examination Guidelines for Computer-related Inventions, 61 Fed. Reg. 7478 (Feb. 28, 1996) ("Guidelines") (1996), Available:
http://www.uspto.gov/web/offices/pac/dapp/oppd/patoc.htm
26. Autogiro Co. of America v. United States.: 384 F.2d 391, 155 U.S.P.Q. 697 (Ct. Cl. 1967)
27. Unique Concepts, Inc. v. Brown. 939 F.2d 1558, 19 U.S.P.Q.2d 1500 (Fed. Cir. 1991)
28. U.S. Patent and Trademark Office.: Examination Guidelines for Computer-related Inventions Training Materials Directed to Business, Artificial Intelligence, and Mathematical Processing Applications ("Training Materials") (1996), Available:
http://www.uspto.gov/web/offices/pac/compexam/examcomp.htm