

A DYNAMIC LICENSE AGREEMENT SYSTEM FOR REUSE OF WEB CONTENTS

Michiko Yasukawa

Department of Social Informatics, Graduate School of Informatics, Kyoto University.

michi@db.soc.i.kyoto-u.ac.jp

Abstract In the point of view of copyright, license agreements are required for proper reuse of web contents. In this paper, a new license agreement model and a licensing system for reuse of web contents are proposed. With this model and system, both providers and users of web contents can find the point of compromise to make license agreements dynamically. The system responds to users' requests and generates license agreements according to the directives, which are registered by providers. In order to adjust the directives, providers can search requested conditions and agreed conditions of users. On the other hand, when users search for available contents for reuse, three independent conditions of search are combined: 1) substance of content, 2) allowed use conditions, and 3) license fee. In this paper, the prototype licensing system is also described.

Keywords: Copyright, Distribution, Contracts, Electronic commerce, Ranking

1. Introduction

Recent research and development has enabled practical information sharing systems, such as distance learning systems and desktop conferencing systems. In order to share valuable documents in these systems, it is helpful to use contents on the web as a mass of material. However, the system, which enables a large variety of reuse of web contents, has not yet been developed. Here reuse of web contents is "using web contents included in a web document as a material of another web document". Here a web document is called a web page, and "web content" is a single file, which includes image, text, sound, or video content.

Most of the current ECMS (Electronic Copyright Management System) charges for end-use of web contents and prevent their illegal reuse. Digital watermarks can be applied as a technique that prevents illegal reuse of web contents. PictureMarc of Digimarc [dig] is one of the digital watermark

The original version of this chapter was revised: The copyright line was incorrect. This has been corrected. The Erratum to this chapter is available at DOI: [10.1007/978-0-387-35658-7_21](https://doi.org/10.1007/978-0-387-35658-7_21)

R. Meersman et al. (eds.), *Semantic Issues in E-Commerce Systems*

© IFIP International Federation for Information Processing 2003

techniques. MarcSpider of Digimarc chases a watermarked image and reports the location of the image. It enables detection of the pirated copies and unauthorized use of image contents. Superdistribution [M⁺90] [M⁺97] was proposed as the model of collecting fees for end-use. Superdistribution does not prevent reproduction of web contents. The system based on Superdistribution collects fees for end-use from subscribers and pays them to the author automatically. Although these techniques are indispensable for copyright management of web contents, variety of reuse of web contents is not lawfully realized with these techniques only. Except for special circumstances where copyright is restricted, reuse of web contents basically requires a license agreement between a licensor and a licensee [Smi97] [Sam98]. Here a licensor is a person who owns copyright or other intellectual property of web contents and provides their web contents lawfully. This person is called a “provider” here. A licensee is a person that intends to reuse web contents. This person is called a “user” here. Namely, a provider can grant a license for reuse of web contents to a user.

Reuse of web contents is shown in Figure 1. It is characterized by merits on the web. Hence, a license agreement for reuse of web contents is different from that of traditional copyrighted works. A license agreement that matches the characteristics of reuse of web contents is required.

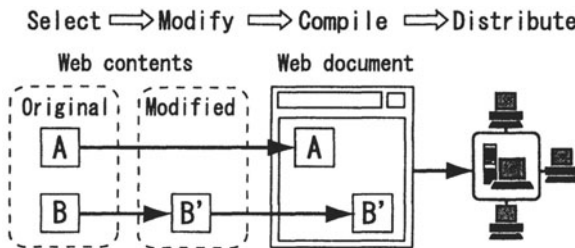


Figure 1. Reuse of Web contents

In this paper, the features and requirements of the license agreement are analyzed (Section 2), and a new license agreement model for reuse of web contents is proposed (Section 3). In the proposed model, both providers and users represent their preferences so that they can make license agreements dynamically on the web. The prototype system based on the proposed model is described in Section 4.

2. Analysis of license agreement for reuse of web contents

2.1 Comparison between traditional copyrighted works and web contents

Traditionally, professional artists have created copyrighted works. License agreements are made by the artists with particular companies. Royalties are paid to the artist according to the retail price of copyrighted works. The retail price depends on the market price. License agreements of traditional copyrighted works are made on the basis of specific and limited use of copyrighted works. Furthermore, publishing of copyrighted works takes much time and cost. On the other hand, reuse of web contents is characterized as followings:

1 Swiftness

Web contents are received on the web and web documents are distributed on the web. Hence the user makes web documents and distributes them much more quickly than by using traditional copyrighted works.

2 Frequency

At the latest estimation, the number of people online is 377.65 million [nua]. It means that there are many providers and many users online. Hence, opportunity to provide and reuse web contents happens frequently.

3 Diversity

The processes of reuse are original and creative actions of users. Reuse of web contents does not mean sales of reproductions of contents. Firstly, a user picks out valuable web contents from enormous collections of web contents. Secondly, the user adds modifications to the content if necessary. Thirdly, the users compile a web document including selected or modified web contents. In addition, subscribers may be children, students, office workers, and/or disabled persons. Hence, reuse of web contents is so diverse that providers cannot imagine how their contents will be reused.

4 Specific economic structure

Web contents and web documents are distributed on the web without the physical distribution network. Thus, physical distribution costs equal zero. In addition, royalties on web contents are not fixed. Although one of the publishing companies has fixed a new royalty rate on electronic books recently, a large majority of the companies are not ready to set royalty rates [wir]. Hence, reuse of web contents is in a specific and unprecedented economic structure, which does not have a fixed market price.

2.2 Current license agreement models

Current models for license agreement are categorized in two types: one-to-one model and one-way model. The characteristics and problems of each model are followings.

1 One-to-one model

This is a mutual individual contract model between a provider (licenser) and a user (licensee) via e-mail or other personal communication means. In this model, a user needs to look up the contact address of a provider. Once a contact is established, the user can start negotiation with the provider. One example of this model is a service called “Media Commerce” of Digimarc Corporation. It enables users to be informed of the provider’s contact address by decoding watermark in images. Since the license agreement process is not quick and easy for both providers and users, one-to-one model is not suitable for reuse of web contents that is frequent and requires swiftness.

2 One-way model

This is a contract organized by providers(licensors). In this model, providers can register their copyrighted works and their offered use conditions are stored in a database. Users search the database to make agreements with providers. When offered use conditions by providers are acceptable and license fees for them are reasonable, users make license agreements with the providers. One example of this model is the prototype system of Copymart [Kit97] [cop]. This model is suitable for contracts of traditional copyrighted works accompanied by the fixed license fee rate in the market. However, in this model, it is hard for providers to register preferable combination of use conditions accompanied by a license fee, because reuse of web contents is so diverse and its economic structure is specific and unprecedented.

2.3 Requirements for license agreements of web contents

The problems of current models reveal the following requirements of the licensing system for reuse of web contents.

1 Mutual license agreement

Because of the diversity and the specific economic structure of reuse of web contents, providers cannot exactly imagine how users use web contents. If providers make one-way contracts and present their demands unilaterally, users will be unwilling to agree with providers. and license agreements will not be made successfully. Providers(licensors)

and users(licensees) need to interact in order to know each other's demands.

2 Reduction of providers' difficulties

Because of the diversity and the frequency of reuse of web contents, licensing process of the provider will become burdensome. If the licensing process is so complicated and difficult that the system requires times and effort to providers, they will be unwilling to grant license. Efficiency is required when they grant use conditions accompanied by license fees.

3 Reduction of users' difficulties

Because of the swiftness of reuse of web contents, licensing process should be quick and easy. If licensing process takes much time, users' reuse of web contents is disturbed. However, search conditions for preferable web contents are complicated since the search conditions are related to substance of content, allowed use conditions, and license fee. Efficiency is required in searching for web contents for reuse.

3. The licensing model

In order to meet the above mentioned requirements, the proposed licensing model has the three characteristics, such as 1) dynamic license agreement via a licensing system, 2) directives by providers, 3) combining search conditions. These characteristics are detailed in followings.

3.1 Dynamic license agreement

The first step of the process of the dynamic license agreement is preference registration of users. This is different from current copyright management systems based on the one-way model. In one-way model where providers register their preference to the licensing system, users don't present their preference to providers. In the proposed licensing system, both of the user's preference and the provider's preference are registered to the licensing system and are represented to both users and providers. In licensing processes, users and providers can search registered users' preference. Such registration and search are the characteristics that one-to-one model and one-way model don't have. These characteristics bring mutual understanding between providers and users on the web. In other words, it can be said that this is the agreement formation between many users and many providers. The users and the providers pay attention to the trend of the registered preference of the whole providers and the whole users.

The licensing process between users and the licensing system is shown in Figure 2. A user registers his preference as search conditions in the process of the content search. When users can find preferable web contents, he makes license agreements with the providers of the web contents. User's agreements are registered in the licensing system and the user can reuse the contents successfully. When users cannot find preferable contents, the licensing process ends in failure. Like this, users know whether web contents are reusable or not without direct contact with providers.

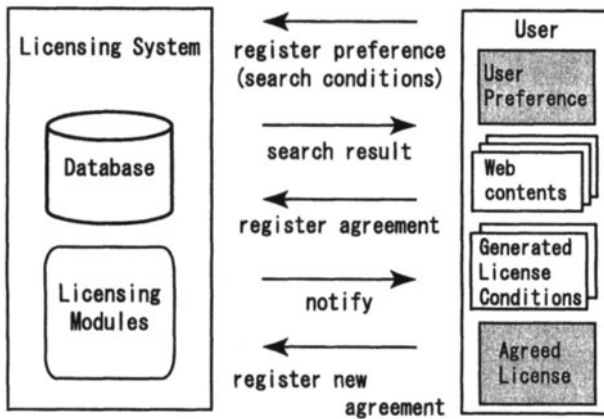


Figure 2. Agreement Process of Users

Since agreed license can be destroyed, user can destroy agreed licenses at anytime and can make a license agreement newly. Namely, when a user finds a more preferable web content, the user can make a new license agreement with the provider of the web content. An old license becomes annulled when a new license is granted. In order to know registration of new preferable contents, preferable use conditions or preferable license fee is registered to the licensing system as notify conditions. When the remarkable registration of web contents occurs, the licensing system notifies it to users according to the notify conditions.

The licensing process between providers and the licensing system is shown in Figure 3. Providers search for registered users' preference in order to grasp the trend of users. Then, they register their own preference. This avoids the unilateral representation of use conditions and license fees. After the registration of preference, the providers can adjust registered preference if necessary. The providers ask the licensing system to notify the trend of users' preference. The providers register remarkable user's trend as a notify condition, such as preferred use conditions, preferred license fee, agreed use conditions, and agreed license fee.

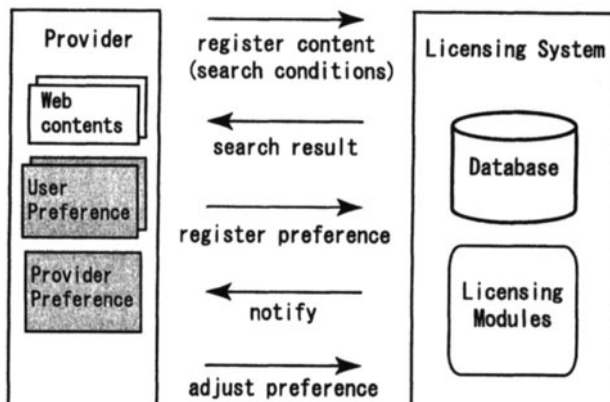


Figure 3. Agreement Process of Providers

3.2 License Generate Directive

In order to present various reuse of web contents, use conditions are defined abundantly and classified in several groups. Each use condition is categorized in a group. To give an explicit example, here groups are defined as three groups: 1) a group of use conditions for modification of web contents, 2) a group of use conditions for compilation of a web document, and 3) a group of use conditions for distribution of a web document.

When reuse of web contents is diverse and frequent, consideration of the combination of acceptable use conditions accompanied by license fees is complicated and troublesome. It can be obstructive for providers to grant licenses. In current systems that manage licensing process, efficiency in granting the licenses is not considered enough.

In the prototype of Copymart system, the providers are required to register combinations of use conditions accompanied by license fees. If this method is introduced to the licensing system for reuse of web contents, the providers are required to fix registration as shown in Table 1.

In Table 1, $X_1, \dots, X_i, Y_1, \dots, Y_j,$ and Z_1, \dots, Z_k represent a group of use conditions respectively. When a new condition X_{i+1} is added, providers are required to consider “agree or disagree” and license fees to all new combinations of use conditions, which are total $j \times k$. In addition, license fee is difficult to define because there is no market price of license fee for reuse of web contents (See, 2.1).

In the proposed model, a license fee rate is considered as a replacement of a fixed license fee in the proposed model. License fees are computed according to income of a user and shown to the user. Income of a user is profit on a web

Table 1. License Conditions

Use Condition			Agree	License
Modify	Compile	Distribute	Type	Fee
X_1	Y_1	Z_1	A	L_{111}
X_1	Y_1	Z_2	D	–
X_1	Y_1	Z_3	D	–
X_1	Y_1	Z_4	A	L_{114}
X_1	Y_1	Z_5	D	–
...				
X_i	Y_j	Z_{k-2}	D	–
X_i	Y_j	Z_{k-1}	D	–
X_i	Y_j	Z_k	A	L_{ijk}

A=Agree,D=Disagree

document, such as an advertisement charge from a sponsor and subscription fees from subscribers.

In addition, License Generate Directive (LGD) is introduced in proposed licensing model. LGD represents preference of the providers for licensing process. LGDs are registered with the system to generate use conditions and license fees in request to users' request. In this licensing process, the providers are required to register a default license fee rate and three agree types for each condition. The agree types are, agree, disagree, and compromise. When a provider selects compromise for a condition, he fixes "Degree of Compromise" (DC) for the condition. DC is the numerical value that describes how much degree of compromise will be made for the condition in the licensing process. At the same time, DC represents an extra rate for the license fee rate for granting the use condition. When a provider makes a compromise, a license fee rate is computed by adding DC to the default license fee rate.

This licensing procedure is defined as follows. A provider assigns the agreement types A_i to each use condition U_i . A value of A_i is T (True: agree), F (False: disagree), or C (Compromise). When A_i equals to C , the provider assigns a value to d_j (Degree of Compromise). On the other hand, a user assigns Request types R_i to each use condition U_i . A value of R_i is T (True: request) or F (False: not request). When the provider assigns n to the default license rate, the license rate r is computed for each j that is $R_j = T$:

$$r = n + \sum d_j$$

In the case $A_j = T$, d_j is assigned to zero. If j that fills $R_j = T$ and $A_j = F$ exists, neither use conditions nor license fee rate is generated. Nor license is granted to user. Let I be income of the user on a web document. When license is granted, license fee L is computed:

$$L = rI$$

For example, conditions fixed by the provider are shown in Table 2. In the Table, X_1, \dots, X_i represents use conditions for modification of web contents. Likewise, let Y_1, \dots, Y_j and Z_1, \dots, Z_k be use conditions for compilation of a web document and use conditions for distribution of a web document respectively. In this method, when a new condition for modification $X_i + 1$ is added, the providers need to consider agreement type is agree, disagree, and compromise to $X_i + 1$, and fix d_j (Degree of Compromise) if compromise is selected. Complexity does not arise when a new use condition is added for the diverse reuse of web contents. Providers are only required to consider agreement type and DC for $X_i + 1$, which are independent j and k . In addition, license fees are computed automatically. It saves time and labor of providers when they grant licenses.

Table 2. License Generation Directive

Modify	X_1	X_2	...	X_i
Agree Type	A	D	...	u_{X_i}
Compile	Y_1	Y_2	...	Y_j
Agree Type	D	u_{Y_2}	...	A
Distribute	Z_1	Z_2	...	Z_k
Agree Type	A	u_{Z_2}	...	D

Default License Fee Rate= n
 X, Y, Z =Use Condition
 A =Agree, D =Disagree
 u =Degree of Compromise

Although this method brings efficient licensing process, specific combinations of use conditions are not explained with LGD. For example, in Table 2, if a provider wants to grant a license in the case of “ X_1 and Y_1 ” and “ X_2 and Y_2 ”, whereas he does not want to grant a license in the case of “ X_1 and X_2 ” and “ X_2 and Y_1 ”, it cannot be described with LGD. To describe this special case, “Special Conditions” are provided. In addition, limitation of a license fee and limitation of total DC are also described in Special Condition. Thus, providers

can avoid exceedingly low license fees and a disagreeable reuse that has a large DC.

3.3 Content search for license agreement

The users search for web contents in the viewpoint of 1) substance of contents, 2) allowed use conditions, and 3) the license fee. Substance of contents is scored by the metric of term-frequency statistics that has been applied in current information retrieval. On the other hand, allowed use conditions and license fee are new metrics of contents search for reuse of web contents. These metrics are specific requirements to the contents search for license agreement.

If Boolean model is applied to this search, search result is mixed more preferable contents with less preferable contents. User repeats the relevance feedback to obtain the most preferable contents. When many contents are reused or opportunities to reuse web contents occur frequently, the search is burdensome for users because search conditions are complicated for users to adjust.

In this model, vector space model, which is one of the techniques of information retrieval, is adopted. The search result is presented by user's preferable order. Hence, users easily get useful contents even if their preference is diverse. With this search, user can move swiftly to the process of license agreement since this search is efficient for frequent and diverse reuse of web contents.

Scores of three independent search conditions are score of substance of content (S_c), score of allowed use conditions (S_u), and score of license fee (S_f). The final score that represent usefulness of content (F) is calculated by S_c , S_u , and S_f . These scores are calculated as follows.

The first score is S_c . The score S_c is assigned a score of result of general information retrieval systems on the Web. A user selects one retrieval system and inputs keywords to them. Then the system returns the URLs of contents and each score of them as a search result. Scores of contents are assigned to the value of S_{c_k} , where $1 \leq k \leq N$ and N is the total of retrieved contents. If each score of contents is not given, the score for S_c is simply calculated by the ranking in a search result. When the ranking is high, the score for S_c is given a large value. The score S_c is computed by the following equation. Let r_k be the ranking for each content, $S_{c_k} = (N - r_k + 1)/N$ where $1 \leq k \leq N$ and N is the total of retrieved contents.

The second score is use condition score S_{u_k} . It is computed on the basis of LGD (License Generate Directive; See 3.3) and user's importance for each use condition. A user decides importance for each use condition subjectively.

Although many use conditions exist, allowed use conditions that a user may take are limited. For example, when a provider of a web content permits the reuse by children, a user that is a 30 year-old office worker is not permitted to reuse the web content. When a user needs to add a specific modification to

contents, the user should obtain contents that are acceptable to modify. In order to properly reuse, a user assigns a range to use conditions and then he searches contents of which use conditions are within the range.

The range C assigned by a user is defined as a Boolean expression with c_i , where c_i is a use condition. use condition is such as “use for scientific web document”, “use in teaching materials”, or “use for profit”. For example, a user assigns an expression (“use for artistic document” or “use for document about sports”) and (“distribute among students”)) for C .

Since importance for each c_i is different from every user or every reuse, the user subjectively assigns weight w_i ($w_i > 0$) for each c_i . When user thinks that c_1 is more important than c_2 , he assigns larger value for w_1 than the value of w_2 .

Let S_{u_k} be the final score for use conditions of content k , where $1 \leq k \leq M$ and M is the total of contents accompanying allowed use conditions within C . Let S_{k_i} be each score for use condition c_i of content k . Let a_{k_i} be an agree type for use condition c_i of content k . The provider assigns an agree type (agree, disagree, or compromise) to a_{k_i} . Then, S_{u_k} is computed as follows.

- 1 If a_{k_i} is agree or compromise, the score S_{k_i} is w_i , otherwise S_{k_i} is zero.
- 2 If condition c_i is combined with c_j ($j \neq i$) by the operator OR, the score of the combination is computed as $Maximum(i, j)$.
- 3 If condition c_i is combined with c_j ($j \neq i$) by the operator AND, the score of the combination is computed as $Multiply(i, j)$.

The third score S_f is license fee score. The score for S_f is computed based on LGD (License Generate Directive; See 3.3) and license fee range defined by user. At the search time, user defines minimum value (f_{min}) and maximum value (f_{max}) of a reasonable license fee for him.

Let S_{f_k} be license fee score for content k and f_k be license fee for content k , where $1 \leq k \leq M$ and M is the total of contents accompanying. Then, S_{f_k} is computed as follows.

$$\begin{aligned} S_{f_k} &= (f_k - f_{min}) / (f_{max} - f_{min}) && \text{(If } f_{min} \leq f_k \leq f_{max} \text{)} \\ S_{f_k} &= 0 && \text{(Otherwise)} \end{aligned}$$

The final score F is computed on the basis of S_c , S_u , and S_f . The representation of documents by using vectors in Euclidean space suggests “numerical similarity metrics” for documents. Document search intends to calculate

similarity score that does not rely on vectors' norms, which mean the length of documents. However, in the search for license agreement, vectors' norms are significant since the vector's norm is expressing user's satisfaction. Hence, the similarity based on the inner products is introduced to the search. The inner product $x \cdot y$ is computed as follows.

$$\sum_{i=1}^t x_i \cdot y_i$$

Suppose vector $q = [W_c, W_u, W_f]^T$ as a query vector. Suppose W_c , W_u , and W_f as users importance for S_c , S_u , and S_f respectively. Suppose vector $d_k = [S_{c_k}, S_{u_k}, S_{f_k}]^T$ as a score vector. Then, final score of F_k for content k is calculated by inner products of q and d_k .

$$F_k = (W_c \times S_{c_k}) + (W_u \times S_{u_k}) + (W_f \times S_{f_k})$$

Because both use conditions selected by user and use conditions permitted by provider are not unique, allowed use conditions can be multiple. For all possible allowed use conditions for each web content, possible score S_u and S_f are computed. Then, the best value in computed scores is assigned to final score F_k for content k .

4. Prototype of the licensing system

Prototype licensing system is based on a central server that is accessed via a Web interface. The whole functionality can be accessed from a standard Web browser. The Web-based interface is generated using server-side scripts (PHP [PHP]). Licensing system maintains license information (user's preference, provider's preference, and agreed license information). License information is stored in a relational database (PostgreSQL [Pos]). Server script looks up stored information in database and generates an HTML page in response to a client's request.

Figure 4 shows the architecture of the prototype system. The licensing module is composed of six kinds of sub module. They are 1) search module, 2) registration module, 3) notification module, 4) update module, 5) license generate module, and 6) confirmation module.

With the search module, providers search preference of users and agreement of users. With the registration module, providers register the preference and notification conditions. The notification module is activated periodically and detects users' preference and agreement registered to the relational database. When the notify module discovers that newly registered preference or agreement

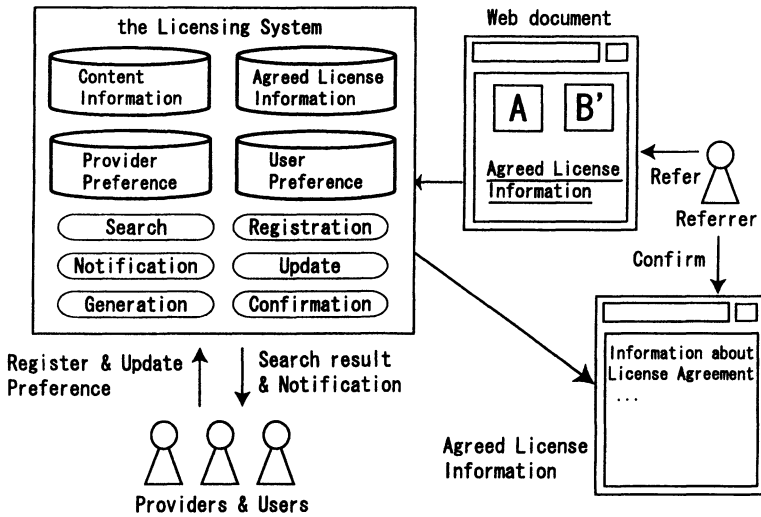


Figure 4. Architecture of Prototype System

of users meets the notify conditions registered by a provider, the system notifies the provider of the preference or the agreement. The provider adjusts the LGD if necessary. With the update module, adjusted LGD gets reflected in LGD registered to database.

On the other hand, a user can search web contents in three independent search conditions, such as substance of content, allowed use conditions, and license fee, with search module. At that time, the user registers the preference. Subsequently, agreed license information is registered to the system if the user selects web contents and agrees on the generated conditions of the web contents. The user can register notification conditions to know that newly registered web contents or adjusted LGD matches his preference. Thus, the user can update the web document with more preferable web contents.

Agreed license information is generated every URL of the Web documents. A referrer that is a person or a software agent and intends to confirm that the web document is lawful investigates agreed license information.

5. Concluding Remarks

In this paper, a new licensing model for reuse of web contents was proposed. The proposed model meets requirements of a license agreement for reuse of web contents, such as 1) two-way communications between providers and users, 2) reduction in the difficulties of generating use conditions and license fees, and 3) reduction in the difficulties of complicated searches of reusable contents.

A licensing system based on the proposed model enables both providers and users to present their preferences. Hence, enabling mutual understanding between users and providers to dynamically make license agreements for reuse of web contents.

In the proposed model, providers utilize License Generate Directives (LGDs), the principal technique. Providers register LGD to the system to make it generate preferable use conditions and preferable license fees. Agreed license conditions are registered to the system as one of the license information, which consists of provider's preference (including LGD), user's preference, and agreed license information. On the basis of aforementioned agreed license information, the system generates a written contract in an electronic form.

Developing practical versions of the licensing system is a future project. To complete the licensing system, 1) advanced content search for reuse of web contents, 2) user support for lawful reuse, such as modification of web contents and compilation of web documents, and 3) a remuneration and authentication system based on agreed license information are required.

Acknowledgments

The author is grateful to Professor Yahiko Kambayashi of Kyoto University for his helpful suggestions and stimulating discussions about this research theme. The author would like to thank Professor Takashi Oseto of Ritsumeikan University for his professional suggestions on the background of legal knowledge. The author would like to give thanks to Dr. Hiroshi Hoshino and Dr. Atsushi Yamada of Advanced Software Technology and Mechatronics Research Institute of Kyoto, for giving helpful advise. The author would like to express his gratitude to Professor Eiji Okubo of Ritsumeikan University, who encouraged the author, and gave thoughtful suggestions.

References

- [cop] Copymart homepage. <http://www.copymart.gr.jp/>.
- [dig] Digimarc corporation. <http://www.digimarc.com/>.
- [Kit97] Z. Kitagawa. Technological developments in electronic copyright management systems in japan for the digital environment, 1997.
- [M⁺90] R. Mori et al. Superdistribution: The concept and the architecture. *Trans. IEICE*, 73(7):1133–1146, 1990.
- [M⁺97] R. Mori et al. Superdistribution: An electronic infrastructure for the economy of the futer. *Trans. JPSJ*, 38(7):1465–1472, 1997.
- [nua] How many online? http://www.nua.ie/surveys/how_many_online/index.html.

- [PHP] Php: Hypertext preprocessor. <http://www.php.net/>.
- [Pos] Postgresql. <http://www.postgresql.org/>.
- [Sam98] P. Samuelson. *The future of the information society and the role of copyright in it*, 1998.
- [Smi97] G. Smith. *Internet law and regulation*, 2nd ed., Sweet & Maxwell, 1997.
- [wir] Random house sets royalty rate.
<http://www.wired.com/news/culture/0,1284,40143,00.html>.

Session II: Security Issues