ABSTRACT - Users who fail to properly understand the rights management information (RMI) on the site that provides the public domain works or who fail to get the correct information may be subject to copyright infringement[3,4]. For this reason, sites that provide public domain works should integrate, manage, and update the latest information to provide accurate and up-to-date RMI information for public domain works. However, since images with the same rights may be circulated in different formats according to various image formats and sizes, it is important to handle data redundancy. In this paper, we try to solve the problem of redundant data on the RMI managed by the intelligent information based public domain images search system using the image feature detection technology.

**Keywords:** Public Domain Works, Rights Management Information(RMI), Web Crawling

1. **INTRODUCTION**

Recently, various agencies around the world have been operating websites for public works freely available according to the type of copyright in order to guarantee correct use of the work.

In Korea, there are 'Gong-u Madang' and 'Gong-Gong Nuri' as typical examples. Table 1 shows that the usage status is gradually increasing[1].

| Table 1. Current Status of Public-Domain Works in South Korea |
|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| linguistic        | 84,759            | 651,009           | 710,526           | 1,005,727         | 1,322,159         | 1,535,085         | 1,875,073         |
| music             | 131,622           | 31,139            | 61,947            | 278,319           | 269,367           | 193,983           | 155,637           |
| art               | 147,658           | 147,175           | 257,794           | 1,074,854         | 3,295,706         | 1,345,680         | 1,129,819         |
| construct         | 25,509            | 22,471            | 19,800            | 10,946            | 35                | 1,945             | 0                 |
| photo             | 187,936           | 54,920            | 135,735           | 1,056,905         | 3,174,003         | 1,659,809         | 3,950,762         |
| video             | 0                 | 0                 | 0                 | 9,597             | 165,936           | 181,957           | 208,832           |
| computer program  | 0                 | 0                 | 0                 | 0                 | 1,142             | 843               | 6,106             |
| etc               | 0                 | 0                 | 0                 | 0                 | 0                 | 211,327           | 0                 |
| total             | 577,484           | 906,714           | 1,184,902         | 3,436,348         | 8,228,348         | 4,919,302         | 7,537,556         |

Reference: Korea Copyright Commission 'Copyright Statistics, Vol. 6, No. 7, July'

Such public works have the advantage of being used relatively freely if they are in the scope of the usage
permission, but they may lead to copyright infringement if they exceed the scope of the usage permission[2]. In order to solve this problem, each public domain works site is designed to prevent the copyright infringement by creating a mark indicating the license conditions for the copyrighted works, but it may cause confusion to the users due to different signs among sites, which emphasizes the importance of integrated management of RMI [3,4].

However, even in the integrated management, images having the same rights are diversely distributed according to various image formats and sizes, so it is important to handle data redundancy[5]. In this paper, we propose a method for redundant data processing by using image feature point detection technology to manage and update RMI that can be used in intelligent information based image search system.

2. Duplicate data processing using image feature points for integrating and updating RMI of public domain images

2.1. Intelligent information-based image search engine

The search engine for public domain images is shown in Figure 1. In the training part for tagging images, images and metadata are collected using the crawling module. Using collected data, the system will learn tags that match each image in Deep Learning module. When data that a search engine wants to show to a user is added later, it produces metadata information such as tags and color information according to results that are learned from Deep Learning module and saves them in database. Saved data can be retrieved by search engine according to various search conditions, such as specific tag or image color, object, etc.

![Figure 1. Configuration of Intelligent Public Domain Image Search System](image)

2.2. RMI Integration and Update Algorithm

Figure 2 shows the RMI integration and update algorithm proposed in this paper. First, the data of the public domain works is extracted from the targeted public domain site using the web crawler technology. The RMI of the work is verified from the extracted data, and the copyrighted work and the work which is managed at the national level are excluded. The selected data in this way are first stored as a new data set database, not as the database of search engine, and is integrated and updated by matching with the data
of the integrated public domain works RMI database of the search engine.

![Figure 2. Algorithm Integrating and Updating RMI of Public Domain Images](image)

### 2.3. Data Duplication Processing Scheme

In the process of matching for updating and integrating data, it is necessary to extract and match the hash value of the image. However, since the sizes of the images per site are different, the corresponding hash values become different, which makes the hash difficult to be set as the identifier. Therefore, the feature information of the image is selected as the identifier. We use a method to generate a code that combines the average of the RGB values of the width and height of the image [6]. Using the feature information, the matched data is integrated and updated according to the algorithm of Fig. 3. If there is no matching data in the database of the temporary database B in the database of the search engine A, the data is added to the database of A judged as the newly registered data. If there is data and the meta information such as RMI is changed, the update operation is performed. During this matching process, the database in B may have a number of matches to the database in A. In this case, the date of registration and date of modification of the image shall be the selection criteria, and if the date is the same, the data with the higher number of hits shall be selected and updated.
3. Conclusion

In this paper, we have solved the problem of redundant data about the RMI managed by the intelligent information based public domain images system using the image feature detection technology. This enables a search engine that needs to provide accurate and up-to-date RMI for a public domain work to resolve possible redundancy problem of images with the same rights circulated in other forms when updating RMI, thereby maintaining consistent and up-to-date information.

In the future, it will be necessary to improve the reliability and accuracy of the proposed algorithm which processes the duplicated data using the feature detection technology by checking the matching method.

ACKNOWLEDGMENT

This work was supported by Ministry of Culture, Sports and Tourism(MCST) and Korea Copyright Commission in 2018(2017-SHARE-9500).

REFERENCES

3. Do Doo Hyung., Method of Use for., and Infringement on, the Author’s Property Right for Joint Work, 29, No 1, 185-205, JUDICIAL PRECEDENT REVIEW.