Software Program Similarity based on Cases

Seong-Hoon Lee¹ and Dong-Woo Lee*²

¹Division of Information and Communication, Baekseok University, 115, Anseo-dong, Cheonan, Choongnam, Korea; shlee@bu.ac.kr
²Department of Computer Information, Woosong University, 17-2, Jayang-dong, Dong-gu, Daejeon, Korea; dwlee@wsu.ac.kr

Abstract

Objectives: There have been occurred many quantitative and qualitative changes in computer software industry. Some of the changes are negative aspects. We studied an intellectual property rights among negative aspects. Methods/Statistical Analysis: The subject of this study is actual cases to occur in real situation. We choose a GIS program. Fake names in this paper are used for related companies. The company A’s AA program and company B’s BB program have the four similarities and plagiarism such as import/export, space calculation, storage manager and display. Therefore, we choose appraisal standards on the following programs the four objections in this study. Findings: We analyzed and conducted in the procedure from setting of appraisal items to evaluation of appraisal. Firstly, five items were set in the step of determining items to analyze similarity of the two programs. As the step of adding relative significance on appraisal items on the previous step, significance of each item is added in our principles. Degree of copying on the items is calculated by our principles. As the last step of calculating overall degree of copying on the two programs are calculated by our principles in this paper. The overall copying degree showed very low value of 5.17 Percentage. This means that the copying rate is very low or none had taken place. Improvements/Applications: We introduce the principles needed to calculate the degree of the copying in programs. And we also showed a procedure to analyze similarity and copying.

Keywords: Application, Cases, Computer Software, Similarity, Software

1. Introduction

Computer software or simply software is a part of computer systems. It consists of encoded information and computer instructions, in contrast to the physical hardware from which the system is built. Software is instructions (computer programs) and data structure and documents. Instructions when executed provide desired function and performance. Data structure enables the programs to adequately manipulate information. Software is documents. Documents describe the operation and use of the programs.

Computer software can be divided into Application software and System software. Application software is software that uses the computer system to perform special functions or provide entertainment functions beyond the basic operation of the computer itself. Software configuration is as shown in Figure 1.

Figure 1. Software configuration.

The software license gives the user the right to use the software in the licensed environment, and in the case of free software licenses, also grants other rights such as the right to make copies. Proprietary software can be divided into two types. Freeware includes the category of ‘free trial’ software or ‘freemium’ software. This is sometimes only true for a limited period of time or with limited functionality. Software available for a fee, often inaccurately termed ‘commercial software’, which can only be legally used on purchase of a license. Open source
software, on the other hand, comes with a free software license, granting the recipient the rights to modify and redistribute the software.

There have been occurred many quantitative and qualitative changes in computer software industry in KOREA by the multiple support policies. These policies have been done by the government. Some of the changes are the positive aspects, the other of changes are the negative aspects. One of the negative aspects is the issue on related with the intellectual property rights, and there are many disputes and civil suit or criminal suits on related with the intellectual property rights. Therefore, improvement of systems and systemization is important for relief of fair disputes on intellectual property rights in the current convergence society. In this paper, we studied on plagiarism of the computer software. We selected a GIS program on actual cases. And we inferred a degree of similarity and plagiarism.

2. Context of Study

2.1 Procedure of study
The actual cases in intellectual property rights of computer program are the theme of this study. The assumed names in this paper are used for related companies. The assumed name A is a company, and other assumed name is B. Here, company A raises objection to the central district court that BB program had copied their company’s AA program. The programs of the two companies, company A’s AA program and company B’s BB program, have the four similarities. The four parts of similarities are Import/Export, Space calculation, Storage manager, Display. In this paper, we choose appraisal standards to calculate the copying rate on two programs the four objections. We conducted in the procedure of Figure 2. Insert figure 2.
Appraisal on similarity and plagiarism of this program was requested by the plaintiff(company A) on the defendant(company B) to see how much the program of the plaintiff’s program was cited and plagiarized as the analyzed subjects are limited to the four items written above.

- Import/Export part - Space calculation part
- Storage management part - Display part

When context of appraisal item is determined to use open software, degree of copying is “0.” This study progresses the order of figure 2.

2.2 Determination of Appraisal Items
Firstly, to analyze similarity and plagiarism of the two programs, we set the following five items such as Table 1 in the step of determining items.

<table>
<thead>
<tr>
<th>Appraisal Items</th>
<th>Comparison objects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source program</td>
<td>Correspondence of code Program composition method</td>
</tr>
<tr>
<td>Storage management form</td>
<td>Index structure Data structure</td>
</tr>
<tr>
<td>Space calculation form</td>
<td>Space calculation open software form</td>
</tr>
<tr>
<td>Screen composition</td>
<td>Composed menus Composed Icon form</td>
</tr>
<tr>
<td>Etc</td>
<td>Correspondence of file name</td>
</tr>
</tbody>
</table>

2.3 Significance Setting by of Appraisal Items
As the step of adding relative significance on appraisal items on the previous step, the following significance was added in the following principles.

- 1) Source program: To judge similarity and plagiarism of programs, comparison and appraisal in source code level is most clear. If context of source code is the same, it can be said this is clearly plagiarized. If plagiarism is determined in source code level, comparison and analysis on other analysis items is not required. Therefore, 50% weight was added in total degree of importance. Source code items are specified as follows.

  - Code correspondence: correspondence on source codes on analyzed subject

  Most clear element in deciding copying or plagiarism Significance: 40% added
• Program composition method: similarity of programming technique
  Cases changing parts have occurred as appraisal methods have been known
  General tendency of programming method when other person has programmed by same context
  Therefore, it is judged that there is similarity if programming method is corresponding
  Significance: 10% added

• Storage management form: Comparison, analysis on management is important on saving, managing space and non-space data that are related to GIS on the appraisal subjects of this paper. Therefore, 15% significance was added on this as the specific context is as follows
  • Index structure
    Comparison of index structure for space and non-space access
    Significance: 5% added
    • Correspondence of data structure (10)
  • Correspondence of record fields
    Significance: 10% added

• Space calculation: Calculation on context related to space data(Union, Intersection, Difference, Join) are compared, analyzed to correspond to open software. Therefore, form of using open software in this appraisal was compared to add 15% significance

• Screen composition: Context displayed on screen for users are compared, analyzed to add 15% significance in this part. Specific screen compositions are as follows.
  • Menu composition

Comparison of menu composition provided by program
  Decision of similarity by degree of correspondence of menu composition
  Significance: 10% added
  • Icon composition and form
  Criterion to judge program similarity by correspondence of icon composition and form
  Significance: 5% added
  • Etc.: Investigation, analysis on similarity and plagiarism of appraisal subjects
  • Investigation, analysis context
  Correspondence of program file name
  • Significance: 5% added
  The discussed appraisal items and significance are put together as following Table 2.

### 2.4 Calculation of Degree of Copying in Each Appraisal Item

Degree of copying on the appraisal items determined above are calculated by the following principles. Degree of copying on corresponding items was set as “0” for appraisal standards used before the opened software.

Degree of copied source codes
  • Corresponding context are extracted in line units of the two programs to calculate
  • Ex) A company program line number: 100, Number of corresponding lines in company B program : 10
  Degree of copying = 10%

Degree of copying on storage management form

<table>
<thead>
<tr>
<th>Appraisal Items</th>
<th>Specific Content</th>
<th>Significance (%)</th>
<th>Sum of Significance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source program</td>
<td>Program composition method</td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Correspondence of codes</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>Storage Management Form</td>
<td>Index structure</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Correspondence of data structure</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Space Calculation</td>
<td>Used form of open software</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Screen Composition</td>
<td>Composed menus</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Icon composition form</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Etc.</td>
<td>Correspondence of files</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
Software Program Similarity based on Cases

- Index structure for GIS related space and non-space data access
  - Ex) A company index structure: space=R-tree, non-space=B-tree
  - B company index structure: space=R-tree, non-tree=B-tree
  Degree of copying: 100% Degree of copying when using opened software: 0%
- For file: calculates correspondence of record structure
  - Ex) Number of fields: 20
  Number of corresponding fields: 5
  Degree of copying: 25%
  Degree of copying on space calculation
- Calculates correspondence of usage form on calculation
  - Ex) A company: 4 calculations, B company: 3 calculations,
  Number of correspondence in used form: 2 calculations
  Degree of copying: 50%
  Degree of copying when using opened software: 0%
  Degree of copying on screen composition
- Calculates correspondence of main menus and attached menus
  - Ex) Menu items: 10, number of similar menus: 2
  Degree of copying: 20%
- Calculates correspondence of used icons
  - Ex) Number of icons: 10, number of similar icons: 2
  Degree of copying: 20%

Other
- Correspondence of file names
- Investigation on correspondence of file names on 576 programs provided by the plaintiff(company A) on appraisal parts(4 parts) on the program files of the defendant(company B)
  - Ex) Number of files: 20, number of corresponding files: 10
  Degree of copying: 50%
  Degree of copying on each items discussed above are calculated as shown in Table 3.

2.5 Calculation on Overall Degree of Copying
The last step, calculating of degree of copying of total appraisal items, based on context of clause 2.3 and clause 2.4 discussed above, was performed by the following principles.
- Degree of copying context of the original version to the copied version on appraisal items are added
- Significance of appraisal items are reflected in calculation
  \[ \text{overall degree of copying} = \sum_{i=1}^{n} (\text{significance of appraisal item} \times \text{degree of copying in each item}) \] (1)
  \((n : \text{Number of appraisal items})\)

  Finally, the value given by equation (1) shows the degree of copying. Overall copying degree showed very low value of 5.17 percentage (%). This copying or plagiarism of the program is very low or none had taken place.

3. Conclusion
There have been occurred many quantitative and qualitative changes in computer software industry in KOREA by the multiple support policies. These policies have been done by the government. Some of the changes are the positive aspects, the other of changes are the negative aspects. One of the negative aspects is the issue on related with the intellectual property rights, and there are many disputes and civil suit or criminal suits on related with the intellectual property rights. In this study, we deal with a program for copying rate in intellectual property rights. We define the procedure for calculating the copying rate for two programs. Lastly, based on this, we calculate the overall degree of copying and plagiarism.

4. References