Web Scraping Model based on Transclusion in e-Learning Environment

Ja-Ryoung Choi¹, Soon-Bum Lim¹

¹Dept. Of Multimedia Science, Sookmyung Women’s University, Cheongpa-ro 47-gil 100, Yongsan-gu, Seoul, 140-742, Korea
{j2arlove, sblim}@sookmyung.ac.kr

Abstract. In this paper, we suggest a Web Scraping Model based on Transclusion, one of models for collecting contents on the web, in order to reuse web contents in e-Learning environment. Recently, most of instructors create learning materials using web contents. Several studies have suggested focusing on the design of an interface for gathering web contents. However, not much consideration has been given to copyright issues. Since web contents are in digital format, there is a problem with the illegal content copying on the web. In order to solve this problem, our model classified scrap-notes into three different types based on copyright restrictions: Trans-quotation, Trans-reference and Trans-annotation. We expect that our model can solve copyright issues when instructors make learning materials with recorded information of web contents.

Keywords: Transclusion, Web Contents, Web Scrap, e-Learning.

1 Introduction

Nowadays, we can use various web contents with the development of digital devices and web technology. We can re-create digital contents brought from the web. Especially, we need to use web contents to support e-learning. When instructors make learning materials, they collect various web contents to reuse for citation. Since web is an open platform, people use web contents without permission and illegal content reproduction issues arise. Hence, we need a new scraping model that can collect web contents to make learning materials.

In this paper, we suggest a Web Scraping Model that can be used in making scrap-notes by collecting various web contents for learning material. We create three different types of scrap-note based on copyright restrictions and recorded information. Therefore, when a scrap-note is created, information is recorded on XML file to connect to the original web content.
2 Related Works

Previous researches on e-learning system have made many attempts to study LMS (Learning Management System) and provide customized learning content. One of researches related to ours is about reusing learning contents. ReLOAMS[1] (Reusable Learning Objects Authoring and Management System) is designed to reuse learning contents by adding small modules. TMDC[2] (Teaching-Material Design Center) is an ontology-based system for finding teaching templates and learning objects by reusing customized e-materials. Our research is designed to reuse web contents for e-learning.

Ongoing collecting techniques include clipping, transclusion, curation and many more but our work is strongly related to transclusion. Transclusion is proposed by Xanadu project[3]. Xanadu project is an existing collecting technique for connecting documents and files to the web and it is first suggested by Ted Nelson. Transclusion that is based on unbreakable links for connecting to original sources is a technique for the inclusion of existing content into a new document. Writers actually do not make a copy of an existing content, but reference to the virtually stored original content.

Similar study is a TED(Transclusions in Enterprice Documents)[4]. TED system is composed of TED Inclusion Processor and TED Viewer and designed as XML-Based Transclusions.

3 Requirement Analysis

Instructors construct methodical class by using ADDIE model, which is an instructional design framework as phases of Analysis, Design, Development, Implementation, and Evaluation[5]. In this process, it is important to create proper learning contents for a good learning environment. Thus, instructors usually spend a lot of time making their learning materials. Since instructors want to provide their students with diverse learning contents, they use variety of web contents. However, in order to use web contents properly, instructors must satisfy copyright conditions such as resource reproduction and distribution. Therefore, we implemented a Web Scraping Model and classified existing transclusion techniques into three types based on copyright restrictions to make it easy for instructors to create learning material.

![ADDIE Model](image)

Fig. 1 ADDIE Model[5]
4 Web Scraping Model

When instructors make their own learning materials, most of their work is to find a variety of contents on the web. At this time, instructors collect contents using web scraping. In this paper, they can create scrap-notes following copyright restriction for using web contents. We present scrap-notes based on transclusion technique which is connected to the original content. Being connected to the original content is very important due to copyright restrictions. Therefore, we define a XML file to record information of original sources.

There are three types of scrap-notes based on transclusion as follows:

- **Trans-quotation**. In cases where both source reproduction and distribution are allowed, Trans-quotation includes a direct quote from its original content. If an instructor wants to directly quote an original content, Trans-quotation is displayed as itself. Trans-quotation needs some information such as URL address, title of scrap-note, date of publication, and file size.

- **Trans-reference**. In cases where source distribution is allowed but not source reproduction, Trans-reference includes an indirect quotation by instructor from its original source. Thus, Trans-reference needs some information such as URL address, title of scrap-note, description of scrap-note, date of publication, and file size.

- **Trans-annotation**. In cases where neither source reproduction nor distribution is allowed, Trans-annotation separates content from the original source. Since direct quotation is not allowed for this type, instructor should include the link address in his learning material. So, it needs some information such as URL address, and title of scrap-note.

5 Conclusion

In this paper, we suggested a Web Scraping Model to create learning material for e-Learning environment. We analyzed requirements for creating learning materials and defined a Web Scraping Model. First, we classified transclusions into Trans-quotation, Trans-reference, and Trans-annotation according to copyright restrictions. Second, we design a XML file that records the information to connect to the original sources. In the future, we plan to expand our study to interact between scrap-notes and original web contents for adapting to dynamic changes.

Acknowledgments. This research is supported by Ministry of Culture, Sports and Tourism(MCST) and Korea Creative Content Agency(KOCCA) in the Culture Technology(CT) Research & Development Program 2014.
References