

SUGGESTING ADDITIONAL RESOURCES BASED ON USERS TAGS¹

Aleksandar Kotevski, PhD

Partizanska B.B., Faculty of Law Kicevo, Bitola,
aleksandar.kotevski@uklo.edu.mk

Abstract

Critical and important aspect in the modern online learning platforms is selecting the most adequate learning materials based on learners' interests, requirements and their activities. In order to be more efficient, the system needs to be able to suggest additional resources to the users, like links, books, videos and etc. In recent years, using tags in the process of content filtering is very often and is a starting point of almost all e-learning platforms. Through the tagging,

learners can mark or highlight some learning materials and can contribute to organizing and retrieving useful learning materials.

In this paper we propose a method for analyzing user profiles according to their tags in order to predict useful external publications and recommend them. The proposed method uses crawler that reads data from several databases and tries to find out the most adequate publication comparing users' most used tags and publications' tags (publications' keywords, title and the most frequent words in the publications' abstract).

Suggesting useful external resources without some additional actions like searching or browsing in a large dataset of papers is very valuable for users.

Keywords: tags, content filtering, e-learning, crawler, recommendation

INTRODUCTION

An important aspect in the effective e-learning systems is suggesting the most adequate learning materials based on learners' requirements and knowledge

¹ original scientific paper

goals. Today, when a lot of new learning materials are added to the e-learning systems, this aspect becomes especially important because of information overload. Recommending systems in e-learning environments utilize information about learners and learning activities and recommend items such as papers, web pages, courses, lessons and other learning objects that meet the pedagogical characteristics and interests of learners [1]. Tagging represents an action of reflection, where the tagger sums up a series of thoughts into one or more summary tags, each of which stands on its own to describe some aspect of the resource based on the tagger's experiences and beliefs [2]. The main goal of the recommending systems is making predictions using as much as possible user ratings and tags available for a given item.

On the other hand, a lot of systems are using tags in order to allow to the users to express their opinion about some links, materials, stories and etc. In the e-learning systems, users can add tags to some learning materials that are useful for them, to express their opinion about the tagged learning material in order to suggest that to the other users. It means that using tags in e-learning systems allow learners to mark or highlight some learning materials with their own tags.

My previous research was focused on tag-based collaborative filtering and learning style determination in order to suggest useful learning material in adequate format. Also, some papers were focused on the factors that affect the tag-based recommendation.

In this paper we propose a mechanism for recommendation useful external publications by using web crawler that reads data from several databases and tries to find out the most adequate publication comparing users' most used tags and publications' tags (publications' keywords, title and the most frequent words in the publications' abstract).

RELATED WORK

In our previous researches [3,4,5], we have proposed an intelligent e-learning system. It includes the use of an adaptation rules and ontology for knowledge representation and supports the learners by recommending learning materials, online learning activities based on their learning style, knowledge level and the browsing history of other students with similar characteristics, based on the tags entered by the students. In other words, the system uses collaborative filtering based on tags posted from the students.

The authors in [6] proposed a tag-based collaborative filtering approach for recommending personalized items to the users. Based on the distinctive three dimensional relationships among the users, tags and items, they proposed a new

similarity measuring method which generates the neighborhood of users with similar tagging behavior instead of similar implicit ratings. Based on experimental result, the authors show that by using the tagging information, the proposed approach outperforms the standard user and item based collaborative filtering approaches. In [7], the authors proposed a framework for improving recommending systems through exploiting the users tagging activity. They stress social annotation as a new and powerful kind of feedback and as a way to infer knowledge about users. Also, they investigated the role of tags in the definition of the user model and the impact of the tags on the accuracy of the recommendations. The authors in [8] proposed a novel algorithm for tab-based collaborative filtering, which exploits user-contributed tags that are common to multiple domains in order to establish the cross-domain links necessary for successful cross-domain collaborative filtering. The authors introduced a constraint involving tag-based similarities between pairs of users and pairs of items across domains. By using two publicly available collaborative filtering data sets as different domains, the authors experimentally demonstrated that the new algorithm substantially outperforms other state-of-the-art single domain collaborative filtering and cross-domain collaborative filtering approaches. The authors in [9] conclude that similarity based clustering methods might be too rigorous in grouping terms. Within the same paper, they proposed a new method to identify true synonyms in social content systems. Golder and Huberman [10] agree that the polysemy (when a single word has multiple related meanings) and synonymy (when different words have the same meaning) in the tag database both hinder the precision and recall of the tagging systems. In [11], the authors proposed methodology for the analysis of tag-based systems, addressing tag synonymy and homonymy at the same time in a holistic approach. Also, they exploit a tripartite graph to reduce the problem of synonyms and homonyms.

PROPOSED MODEL FOR RECOMMENDATION

The goal of this paper is to propose a mechanism for suggesting external publications based on users' tags. We are going to use web crawler (known as web spider or web robot) – script that searches through the specific URL locations and trying to find adequate matches base on some pre-define template. In the scope of this paper, we are going to use several URL locations – each of them contains a lot of publications from different learning areas. The proposed mechanism starts from a default web locations (which is specified by the administrator) and is going through the all specified URL locations in order to find the most adequate publications by comparing the tags used from the users

and the tags and keywords that are used in the publication. When there are enough matches, the script saves the publications' URL with the percentage of matching.

As soon as the script execution is completed, the system will select top 5 adequate publications (based on the percentage of tags matches).

The proposed crawler uses the follow function:

```
function crawl_site($u){
    global $scrawled_urls;
    $uen=urlencode($u);
    if((array_key_exists($uen,$scrawled_urls)==0 || $scrawled_urls[$uen] <
date("YmdHis",strtotime('-25 seconds', time())))){
        $html = file_get_html($u);
        $scrawled_urls[$uen]=date("YmdHis");
        foreach($html->find("a") as $li){
            $url=perfect_url($li->href,$u);
            $enurl=urlencode($url);
            if($url!=" && substr($url,0,4)!="mail" && substr($url,0,4)!="java" &&
array_key_exists($enurl,$found_urls)==0){
                $found_urls[$enurl]=1;

                $a = $url;
                $b = 'http://journals.ohiolink.edu/ejc/article.cgi?issn=';
                $ss = explode($b,$a);

                if(count($ss)>1){
                    echo "<li><a target='_blank'
href='". $url. "'>". $url. "</a></li>";
                }
            }
        }
    }
}
if(isset($_POST['submit'])){
    $url_1 =
"http://journals.ohiolink.edu/ejc/search.cgi?q=".$_POST['url']."&field=abstract
&op1=and&q2=&field2=&op2=and&q3=&field3=&start_year=&end_year=&
pagesize=10&def_op=AUTO";
}
```

```

$url=$url_1;
if($url==""){
    echo "";
}else{
    $f=fopen("url-crawled.html","a+");

    fwrite($f,"<div><a href='$url'>$url</a> - ".date("Y-m-d
H:i:s")."</div>");

    fclose($f);
    echo "<ul style='word-wrap: break-word;line-height: 25px;'>";
    //if
    (preg_match('/http://journals.ohiolink.edu/ejc/article.cgi?issn=/', $url))
    //{
        crawl_site($url);
    //}
    echo "</ul>";
}
}

```

USE CASE SCENARIO

The proposed recommendation module is a web crawler script that executes every time when some logged user uses some learning material in order to suggest them adequate external publications (publications that are not contains in the current system – they are published on some public repository).

Follow one user-case scenario:

- 1) The user log in and select some learning material
- 2) System is generating list of tags added from the user for the selected learning material
- 3) If no tags added for the selected learning material, the system is going to get the tags added from the user for the similar learning materials to the active learning material
- 4) System is starting the crawler function by matching the tags from the steps 2 or 3 with the tags and keywords mentioned in the external publications
- 5) System is generating a list of matches ordered by the percentage of matches between tags

- 6) System is going to suggest the top 5 publications

CONCLUSION

An important aspect on e-learning system is to select the most adequate learning content and recommend it to the users. There are several parameters that can be consider while filtering materials, but the most common practice is using tags.

In the scope of this paper we proposed a tag-based collaborative filtering algorithm that uses a web crawler in order to select the most adequate external learning materials published on some public repository. The proposed algorithm starts from a default web locations and is going through the all specified URL locations. When there are enough matches, the script saves the publications' URL with the percentage of matching. The proposed algorithm makes decision based on matches on tags used by the logged user and tags and keywords mentioned in the external publication. When the script will finish with process, it generates a list of all external publications order by the percentage of the tag matches.

The proposed mechanism will offer more available adequate learning materials for the users, so the users will be not limited only to the materials from the actual system for e-learning.

This research can be extended with introducing a synonyms to the tags and considering publications ratings too. Also, the crawler script can be extended so it can search for matches in the publications' abstract and content text too. Furthermore, additional parameters can be included in order to optimize the process of crawling.

REFERENCES

- [1] Drachsler, H., Hummel, H., & Koper, R. (2008). Personal recommender systems for learners in lifelong learning networks: the requirements, techniques and model. *International Journal of Learning Technology* 3 (4), 404-423
- [2] Bateman, S., Brooks, C., McCalla, G., & Brusilovsky, P. (2007). Applying collaborative tagging to e-learning. *Proceedings of ACM WWW* 3(4), May 2007
- [3] A.Kotevski, R.Kotevska, Virtual learning group in intelligent e-learning systems, The 2nd International Virtual Conference 2013, (ICTIC 2013), Slovakia

- [4] A.Kotevski, Gj.Mikarovski, Intelligent learning system for High education, ICEST 2012
- [5] A.Kotevski, C.Martinovska Bande, Recommending audio and video materials based on tag-based collaborative filtering, 11th International Conference on Informatics and Information Technologies, CIIT 2014
- [6] Lecture Notes in Computer Science Volume 5589, 2009, pp 666-673 Tag Based Collaborative Filtering for Recommender Systems Huizhi Liang, Yue Xu, Yuefeng Li, Richi Nayak
- [7] Francesca Carmagnola, Federica Cena, Luca Console, Omar Cortassa, Cristina Gena, Anna Goy, Ilaria Torre, Andrea Toso, Fabiana Venero, Tag-based user modeling for social multi-device adaptive guides, User Modeling and User-Adapted Interaction November 2008, Volume 18, Issue 5, pp 497-538
- [8] Yue Shi, Martha Larson, Alan Hanjalic, Tags as Bridges between Domains: Improving Recommendation with Tag-Induced Cross-Domain Collaborative Filtering, User Modeling, Adaption and Personalization Lecture Notes in Computer Science Volume 6787, 2011, pp 305-316
- [9] Maarten Clements, Arjen P. de Vries, and Marcel J. T. Reinders, Detecting synonyms in social tagging systems to improve content retrieval, SIGIR, page 739-740. ACM, (2008)
- [10] Golder, S., and Huberman, B. A. The Structure of Collaborative Tagging Systems. HP Labs technical report, 2005. Available from <http://www.hpl.hp.com/research/idl/papers/tags/>
- [11] Davide Eynard, Luca Mazzola, and Antonina Dattolo, Exploiting tag similarities to discover synonyms and homonyms in folksonomies, DOI: 10.1002/spe.2150