

Teaching Universal Decimal Classification (UDC) to Undergraduate Students: A Folksonomy Driven Approach

Tomislav Ivanjko

Department of Information and Communication Sciences,
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, Zagreb, Croatia
tivanjko@ffzg.hr

Summary

Teaching controlled vocabularies and classifications at an introductory level can be challenging in terms of clarifying theoretical notions and motivating the students. In this light, approaches where students are presented with real life data that illustrates key concepts can be very useful. This case study investigated the difference between subject description of a journal article when using uncontrolled vocabulary (free keywords) and controlled vocabulary (UDC notions) as compared to author keywords, title keywords and UDC notions added by the subject specialist. The study was carried out before and after students attended an undergraduate introductory course to the field of controlled vocabularies and classification systems with a special accent on using the Universal Decimal Classification (UDC). The results confirmed two original hypotheses that UDC driven (post-course) tags will show a greater consistency and have higher agreement with the terms added by the subject specialist than free keyword (pre-course) tags. Since the study was carried out in an educational setting one of the goals of the study was to investigate the possible use of the study to illustrate the basic concept differences between uncontrolled and controlled vocabulary in the light of folksonomies and the use of UDC as a pre-coordinated term list.

Key words: folksonomies, UDC, controlled vocabulary, classification

Introduction

With the rise of digital environments, the use and implementation of controlled vocabulary is the ever more important part of everyday tasks performed by information specialists in the LIS field. Whether searching scientific databases, library catalogues or just Google, basic concepts of the use of controlled vocabularies are always present when the retrieval of information is needed. Knowing the advantages and disadvantages of controlled vocabularies as well as proper

approaches to implementation is at the core of every knowledge organization and representation method.

In recent years, a new wave of user participation in creating and describing online resources instigated a new approach in knowledge representation – folksonomies. Folksonomies rely on the process of collaborative tagging, where many users add metadata in the form of keywords to shared content (Golder and Huberman, 2006), and the totality of these added tags on any different platform forms a folksonomy. From the KO aspect, folksonomies represent a "weak" method of knowledge representation because they don't have means to express semantic relations as traditional methods of knowledge representation such as classifications, thesauri or ontologies have (Peters, 2009) so many authors advocate the use of folksonomies as a complementary method of knowledge organization by using power tags extracted from folksonomies along with controlled vocabularies (Kwan and Chan, 2009; Mendes, Quiñonez-Skinner and Skaggs, 2009). In this notion, folksonomies are useful for providing user warrant, i.e. ensuring that the subject description matches the user vocabulary (Mai, 2011).

Considering the significant impact folksonomies had on the field of knowledge organization, they should be included in the syllabus of any course dealing with controlled vocabularies. One of the first attempts in exploring ways to integrate social tagging into teaching students controlled vocabulary was done by librarians at Boston University's Alumni Medical Library (AML) teaching medical students how to use Medical Subject Headings (MeSH) (Maggio et al., 2009). They created in-class exercises to present MeSH and the concept of a controlled vocabulary in a familiar and relevant context for the course's Generation Y student population and provided students the opportunity to actively participate in creating their education. The librarians found that this exercise – including the pre-class activity, intervention, and post-class evaluation – helped clarify the concept of MeSH, and that couching unfamiliar concepts in the context of popular technologies, can lead to more effective teaching (Maggio et al., 2009).

Implementing a case study

Following this initial approach, a small case study was implemented in the course "Classification Systems", offered to students of the Department of Information and Communication Sciences at the Faculty of Humanities and Social Sciences, University of Zagreb, Croatia. The course is offered at the undergraduate level, holding 6 ECTS credits and is intended as an introductory course to the field of controlled vocabularies and classification systems with a special accent on using the Universal Decimal Classification (UDC) as a practical implementation of a classification system widely used by Croatian libraries. The course is offered each year in the summer semester, and in the academic year 2012/13 a total of 30 students enlisted in the course. There were two main reasons this course was selected for the case study: (1) students enrolling in the

course are at the undergraduate level so they have no prior experience with controlled vocabularies in the form of course credits and (2) their age makes them likely Web 2.0 users so it was expected that the use of social tagging (a part of the Web 2.0 suite) could enable students to better understand the use and purpose of controlled vocabularies and classifications in the LIS field.

This case study had 2 main goals, one strictly professional and the other educational. From the professional point of view we wanted to investigate whether using common pre-coordinated vocabulary (UDC notions) and familiarization with the proper use of controlled vocabulary during the course will have significant impact on the description of the same resource. That meant measuring the term similarity of the pre- and post-course exercise in order to investigate and reflect the possible roles and aspects of folksonomies in the field of controlled vocabularies and UDC in particular. The second goal was aimed at exploring the possible use of these results to illustrate the basic concept differences between uncontrolled and controlled vocabulary in the light of folksonomies and the use of UDC to the students.

Methodology and limitations

Since the course is taught during a single semester, there were a total of 12 lectures and exercises planned that addressed different topics within the course topic during the 15-week period. At the beginning of the semester, during the first exercise, as a part of the lecture on the role of abstracts and keywords, the students were presented with three journal articles from the field of Social Sciences. The articles chosen for description were from a scientific journal, so they had abstract and author keywords assigned as well as UDC numbers added to them from a subject specialist, but were stripped from all of them. Students were given an assignment to write a short description about the topic of the articles and tag them. The exercise took place in an online environment of the Moodle based LMS used by the Faculty where an assignment was created, and the students were not aware that this exercise would be used for a case study. Of the three articles, from the study point of view, first two articles were considered only training exercises, with the results based on the third article planning on being used in the study itself. This was done to ensure that students understood the assignment and were able to use the LMS properly to complete it. After that first exercise, the course was taught during a 12-week period covering the topics of controlled vocabularies, classification systems and the practical use of the UDC. During the last exercise session, a follow-up assignment with the same methodology was undertaken on the fore mentioned three articles. This time students were instructed to summarize and tag the articles again, but this time using the knowledge they acquired during the semester and using the UDC notions as tags. Two hypotheses were:

- (1) UDC driven (post-course) tags will show a greater consistency than free keyword (pre-course) tags
- (2) UDC driven (post-course) tags will have higher agreement with the terms added by the subject specialist than free keyword (pre-course) tags

Both hypotheses were based on the presumption that using common vocabulary (UDC notions) and familiarization with the proper use of controlled vocabulary during the course will have significant impact on the results and subject approach. After the assignment was completed, it was found that out of 30 students, a total of 25 students completed both assignments so their tagging results of the required article was included in the further analysis. The idea of the study was not to be comprehensive, but to investigate and illustrate specific aspects of the relationship between controlled vocabularies, UDC and docsonomies within the educational setting. Since this study was carried out in a specific setting of an undergraduate course, the small sample of the students is a limitation of the study.

Results and discussion

First, the two tag sets were analyzed using descriptive statistics as presented in the Table 1.

Table 1: Free tags vs. UDC notions

	# OF STUDENTS	# OF TAGS	AVERAGE	# OF DIFFERENT TAGS	TOP 10 TAGS
FREE TAGS	25	146	5.84	66 (45%)	68 (47%)
UDC NOTIONS	25	156	6.24	43 (27%)	101 (65%)

In the pre-course exercise students assigned 146 tags, with an average of 5.8 tags per student, while the in the post-course exercise students assigned 156 UDC driven tags, with a somewhat higher average of 6.2 tags per student.

Further analysis wanted to examine the consistency vocabulary used in both exercises. It was hypothesised that the post-course tags will show a greater consistency because students are using UDC notions and are familiarized with the proper use of controlled vocabulary. This was confirmed by the analysis where it was shown that the docsonomy (all tags of a concrete document, Peters et al., 2011) for the article described with free tags generated 66 different tags (45%) while the one created post-course, using UDC notions had only 43 different tags (28%). The analysis of tag frequency distribution showed that the top ten tags (with the highest frequency) account for 46% (68) of total tags when students were using free tags, while that percentage has risen to 65% (101) when they

were using the UDC notions. When those values are presented on a graph, we can see that both distributions follow the power law (Figure 1).

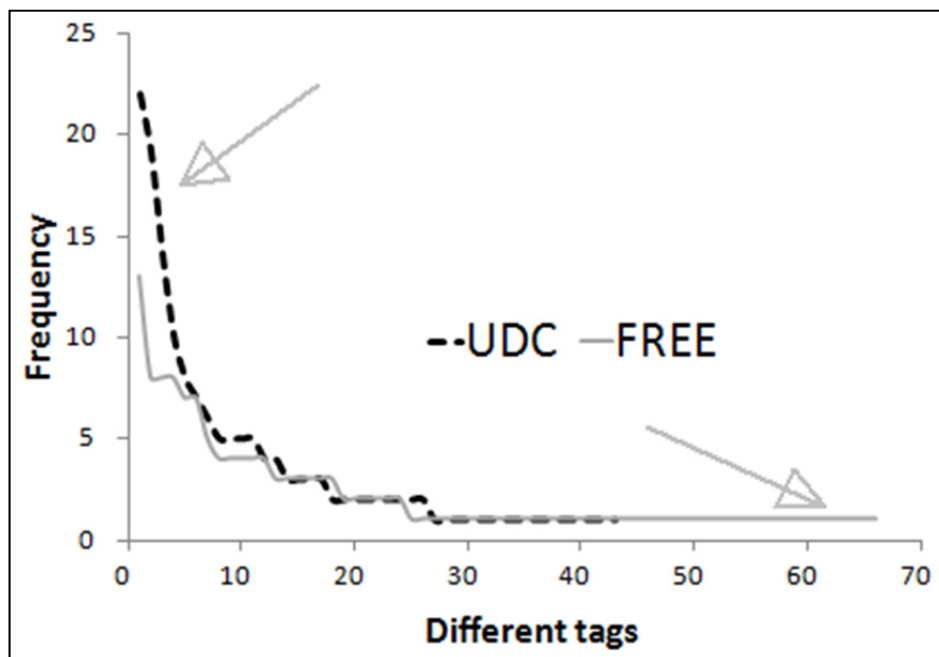


Figure 1: Tag frequency distributions compared

The comparison graph reflects differences where the distribution of free tags produces a longer head and the UDC notion distribution creates a longer tail. Following these results, we can confirm that participants created a more consistent folksonomy when using UDC notions, using fewer tags with higher frequencies. This confirms our hypothesis that the post-course tags will show a greater consistency because students are using pre-coordinated list of UDC notions and are familiarized with the proper use of controlled vocabulary, thus showing UDC driven descriptors more consistent and compact.

Since the article chosen for description was from a scientific journal, it had author keywords assigned as well as UDC number added to it from a subject specialist. From the UDC number added to the article (004:37.018:007) UDC notions associated with the number were extracted (9 terms). Added to these two sets, words from the title were also extracted (8 terms) to be used as a comparison set. The idea was to compare added free keywords (pre-course) and UDC notions (post-course) to see the overlap between those sets. As the user set, only tags with a frequency of 3 and higher were used to ensure basic tag validation (added at least by 3 students). Such criteria yielded 18 free tags and 16 UDC notions. The results are presented in Table 2.

Table 2: Term comparison between sets

	NO. OF INTERSECTING ELEMENTS			
	AUTHOR KW (5)	TITLE KW (8)	SUBJECT SPECIALIST (9)	FREE TAGS/UDC NOTIONS
FREE TAGS (FREQ 3+) (18)	2/5 (40%)	4/8 (50%)	1/9 (11%)	3/16 (19%)
UDC NOTIONS (FREQ 3+) (16)	0/5 (0%)	1/8 (12%)	6/9 (67%)	

As we can see from the analysis, when students had to add free terms to describe the article, the highest overlap was between added terms and title keywords (50%). There was also some overlap between free keywords and author keywords, and only one term was the same as that extracted from the UDC number added by the subject specialist. This illustrates several aspects of the free keyword driven created docsonomy: (1) students presumed that the most important terms are already present in the title of the article so they included them in their description (can be viewed as a copy-paste method); (2) some overlap can be expected between free keywords and author keywords, since both are uncontrolled and (3) there is a low agreement on terms between student free keywords and those assigned by a subject specialist.

When the same analysis was carried out using the UDC driven set of terms added by the students, the results showed highest agreement with the terms added by the subject specialist, where 6 out of 9 terms were present in both sets. This was again in agreement with our second hypotheses that UDC driven (post-course) tags will have the highest agreement with the terms added by the subject specialist. This was underlined with a low agreement between UDC driven tags and author keywords and title keywords.

The last analysis compared added free keywords (pre-course) and UDC notions (post-course) with a frequency higher than 3. It was shown that there were only three common terms that the students used both in pre- and post-course description. This showed how using pre-coordinated terms can yield very different results from using uncontrolled terms.

Conclusion

This case study investigated the difference between subject description of a journal article when using uncontrolled vocabulary (free keywords) and controlled vocabulary (UDC notions). The study was carried out before and after students attended an undergraduate introductory course to the field of controlled vocabularies and classification systems with a special accent on using the Universal Decimal Classification (UDC). The results confirmed two original hypotheses that UDC driven (post-course) tags will show a greater consistency and have higher agreement with the terms added by the subject specialist than free keyword (pre-course) tags. These results followed the line of some earlier

research (Špiranec and Ivanjko, 2013) showing that, from a knowledge organization aspect, users with more knowledge and expertise could create folksonomies of a higher quality.

Since the study was carried out in an educational setting one of the goals of the study was aimed at using these results to illustrate the basic concept differences between uncontrolled and controlled vocabulary in the light of folksonomies and the use of UDC to the students. The methodology and results of this analysis can be used as an illustration of the advantages of using controlled vocabulary in describing resources. It was shown how implementing basic principles of subject analysis and using pre-coordinated terms can yield very different results from using uncontrolled terms. This study illustrated possible weaknesses of folksonomies and the need for complementing them by using controlled vocabularies for enhancing the quality of subject access. This exercise was also useful for illustrating how UDC notions can be used as a pre-controlled vocabulary for subject description. Since teaching controlled vocabularies and classifications at an introductory level can be challenging in terms of clarifying theoretical notions in the field on one hand, and motivating the students on the other, such approach where students are presented with real life data that illustrates key concepts can be very useful.

Further studies in this direction should include larger sample, student evaluation of the approach as well as exploring other ways to implement a case study in the educational process.

References

- Golder, Scott A.; Huberman, Bernardo A. Usage patterns of collaborative tagging systems. // *Journal of Information Science*. 32 (2006), 2; 198-208.
- Maggio et al. A case study: using social tagging to engage students in learning Medical Subject Headings. // *Journal of the Medical Library Association*. 97 (2009), 2; 77-83
- Mai, Jens-Erik. Folksonomies and the New Order: Authority in the Digital Disorder. // *Knowledge Organization*. 38 (2011), 2; 114-122
- Mendes, Luiz H.; Quiñonez-Skinner, Jennie; Skaggs, Danielle. Subjecting the catalog to tagging. // *Library Hi Tech*. 27 (2009), 1; 30-41.
- Peters, Isabella et al. Social tagging & folksonomies: Indexing, retrieving... and beyond. // *Proceedings of the American Society for Information Science and Technology*. 48 (2011), 1; 1-4.
- Peters, Isabella. Folksonomies: indexing and retrieval in Web 2.0. Berlin : De Gruyter Saur, 2009.
- Špiranec, Sonja; Ivanjko, Tomislav. Experts vs. Novices Tagging Behavior: An Exploratory Analysis. // *Procedia - Social and Behavioral Sciences*. 73 (2013); 456-459
- Yi, Kwan; Mai Chan, Lois. Linking folksonomy to Library of Congress subject headings: an exploratory study. // *Journal of Documentation*. 65(2009), 6; 872-900.