CSCI8380 (Fall 2013): Paper Review Form

Reviewer Name: Ting Xiao

Paper Name: Enrichment and Ranking of the YouTube Tag Space and Integration with the Linked Data Cloud

Section I. Overview

A. Reader Interest

1. Which category describes this manuscript?
   _X_PRACTICE/APPLICATION/CASE STUDY/EXPERIENCE REPORT
   ___RESEARCH/TECHNOLOGY
   ___SURVEY/TUTORIAL/HOW-TO

B. Content

1. Please explain how this manuscript advances this field of research and/or contributes something new to the literature.

   The authors of the paper has found the use-generated tags sometimes are too meaningless to be used by current retrieve techniques, so they design a framework using Semantic Web technologies to enrich, ranking and integration of web video tags.

   Since the research field is only concentrated on video media, the paper first introduces and analyzes the present study of tagging. According to these main disadvantages of current techniques of tag suggestion, ranking, they designed a new framework to expand the user-generated tag space, to rank and interlink the tags to DBpedia concepts for greater integration with other datasets.

   The result of the experiment shows after using the new framework, tag space become more sufficient. With the help of semantic web, tags are more meaningful, the retrieval and categorization of video is also improved as a result.

C. Presentation

1. Does the introduction state the objectives of the manuscript in terms that encourage the reader to read on?
   _X_Yes
   ___Could be improved
   ___No

2. How would you rate the organization of the manuscript? Is it focused? Is the length appropriate for the topic?
   _X_Satisfactory
   ___Could be improved
   ___Poor
3. Please rate and comment on the readability of this manuscript.
   ___ Easy to read
   X Readable - but requires some effort to understand (*)
   ___ Difficult to read and understand
   ___ Unreadable

   (*) : I need to learn some definition while I was reading. Maybe it is my own problem.

Section II. Evaluation

Please rate the manuscript. Explain your choice.
   ___ Award Quality
   ___ Excellent
   X Good
   ___ Fair
   ___ Poor

   The application is implemented. According to the evaluated result, it can be used to improve the video retrieval by tagging immediately. But I still have some questions about this paper, so I rate it as a good paper.

Section III. Detailed Comments (provide your thoughts/criticism about the ideas in the paper; not only summarize the paper but have a critical look here)

   The whole paper is organized tightly, the implemented system sounds work good. By expanding the user-generated tag set of the video, ranking tags, and relinked to other Linked Open Data, the system implement the semantic tags of the video content, make them more smart and effective on retrieval.
   However, since the standard of filtering the noisy filter is too simple, I think there is some small flaw in the pre-processing step, and could be improved.

Additional Comments:
   1. Provide one aspect that you liked the most in this paper.

       The research field is to make the semantic tag set for video, that attracts me, since it some researchers are concentrated on extracting the semantic content from the video automatically, we can make a relationship for both of them in future, improve the retrieval be more correct.

   2. Provide one aspect that you disliked the most in this paper.

       Basically, this is a good paper, after reading the paper I have tried all the links of URI, however only one is successful, maybe authors have not published their result on the DBpedia. If they can give a real URI example then will be more great.
Section IV. Discussion Points (provide at least 3 discussion topics/questions related to ideas/techniques described in the paper; these will be used for discussions in the class)

1. Noise filter (preliminary filter) standard is too simple: just using two characters and number. For example:
   “number”: "911" is more meaningful for some specific news video.
   “two characters”: AI is just the abbreviation of artificial intelligent.
   Is there any other method to filter noise?

2. User who ranked the result, "2/3" is that mean totally only 5 persons were asked to evaluation the tagging result? If so, would that be a little insufficient?

3. "co-occurrence statistics", how to calculate this value which can influence the final evaluation?