Towards Ontological Structures Extraction from Folksonomies: An Efficient Fuzzy Clustering Approach

Marouf Zahia, EEDIS Laboratory, Djillali Liabes University of Sidi Bel Abbes, Sidi Bel Abbes, Algeria

Benslimane Sidi Mohamed, EEDIS Laboratory, Djillali Liabes University of Sidi Bel Abbes, Sidi Bel Abbes, Algeria

ABSTRACT

Folksonomies are one of the technologies of Web 2.0 that permit users to annotate resources on the Web. In this paper, the authors propose an integrated approach to extract ontological structures from unstructured and semi-structured resources. Our proposal overcomes limitations of existing approaches. It gives a formal, simple, and efficient solution to the tag clustering and disambiguation problem. Moreover, their approach doesn’t need any ontology as an upper guide during the generation process. The generated ontology can be used to enhance various tasks such as ontology evolution and enrichment.

Keywords: Collaborative Tagging, Folksonomies, Fuzzy Clustering, Ontologies, Similarity Measure

INTRODUCTION

Folksonomies (Vander, 2007) have recently emerged as a powerful way to label and organize large collections of data. These systems allow users to use any keywords or tags relevant to the content to annotate their favorite resources on the web. In this paper, we propose an approach to creating a hierarchy that captures the hierarchical semantic structure of folksonomies. Our approach combines the advantages of folksonomies and ontologies, and overcomes the ontology learning difficulties and costs. The main contributions of the present work are:

• A new representation of the folksonomy based on a tag-tag binary matrix.
• A new similarity measure that calculates relatedness between tags.
• A new generality measure that determines the abstractness level of tags.
• We address the issue of ambiguity in folksonomy based services to return more results that would be missed by traditional search mechanisms.
• A new fuzzy clustering algorithm that doesn’t need determining previously the number of cluster, but deduces it when performing the clustering process.

DOI: 10.4018/ijiit.2014100103
Finally, we conduct comprehensive experiments to evaluate our approach by using real dataset from social bookmarking systems.

The rest of the paper is structured as follows. In section 2, we provide an overview of related work. In section 3, we outline the proposed approach. Section 4 introduces an experimental methodology to evaluate the approach. Finally, section 5 concludes the paper and points directions for future work.

RELATED WORK

Existing approaches to infer tag relationships can broadly be assigned to one of the following classes:

- Clustering approaches: These approaches identify the semantics of tags by clustering tags according to some relations among them (Mika, 2007; Hamasaki, 2007; Begelman, 2006; Kennedy, 2007; Heymann, 2006; Benz, 2010; Marouf, 2013). These clustering approaches measure tags similarity based only on one dimension (user or resource), and most of them do not deal with ambiguity problem or do not give a formal solution to it. Furthermore, the majority of these approaches do not make the hierarchical relations explicit between tags.

- Association rule mining approaches: This class of approaches applies association rule mining techniques in the ontology learning process (Schmitz, 2006; Jäschke, 2008; Trabelsi, 2012). These approaches output a hierarchical representation of tags described by a tri-ordered set of triples called triadic concepts in Formal Concept Analysis FCA (Lehmann, 1995), where each triple consists of a set of users, a set of tags, and a set of resources. However, the relationships between tags in different hierarchical levels are not defined semantically, and there is no strategy to deal with ambiguous tags.

- Semantic based approaches: These approaches aim at associating semantic entities to tags as a way to formally define their meaning (Angeletou, 2008; Cantador, 2008; Garcia, 2009; Djuana, 2012). The lack of ontologies that well match the tags in folksonomies is one of the major obstacles applying these approaches.

- Hybrid approaches: these approaches integrate multiple techniques from those cited previously (Giannakidou, 2008; Specia, 2007; Lin, 2009; Schmitz, 2006). These approaches are however limited as they do not present any disambiguation activity.

Our approach address all these limitations as it deals with all the modes of the folksonomy (user, tag, resource) when calculating relatedness between tags. It gives a formal, simple, and efficient solution to the ambiguity problem, and it represents tags in a hierarchy with broader or narrower ones. Moreover, our approach doesn’t need any ontology as an upper guide during the generation process, so it is language independent.

THE PROPOSED APPROACH

In this section, we describe an extension to our approach that aims to extract ontological structures from folksonomies (Marouf, 2013). The new approach has the same steps as the old one, but it overcomes some limitations that are discussed one by one in the following sections. The overall process is depicted in the Figure 01.

Cleansing of Tags

Since actors can choose any keyword for categorizing their content, they are applying their own spelling and tagging rules (e.g. singular or plural nouns, conjugated verbs). Consequently, tags are polluted and need to be cleansed. For this purpose, we proceed as described in the algorithm 1. We start by deleting all infrequent tags, stop words, and meaningless tags (line 3). The algorithm verifies if the tag is meaningful by checking it existence in Wordnet (line 5).
Related Content

Fuzzy Critical Path Method Based on a New Approach of Ranking Fuzzy Numbers using Centroid of Centroids
[www.igi-global.com/article/fuzzy-critical-path-method-based/77860?camid=4v1a](http://www.igi-global.com/article/fuzzy-critical-path-method-based/77860?camid=4v1a)

Knowledge-Based Systems
[www.igi-global.com/chapter/knowledge-based-systems/10363?camid=4v1a](http://www.igi-global.com/chapter/knowledge-based-systems/10363?camid=4v1a)
Supply Chain Model for Expiring Items Following Ramp-Type Demand With Stochastic Lead Time Under Crisp and Fuzzy Environment

The Application of Swarm Intelligence to Collective Robots
[www.igi-global.com/chapter/application-swarm-intelligence-collective-robots/4676?camid=4v1a](http://www.igi-global.com/chapter/application-swarm-intelligence-collective-robots/4676?camid=4v1a)