

# An Integrated Approach to Extracting Ontological Structures from Folksonomies

Author: Huairen Lin, lin@t.usyd.edu.au

Supervisor: A/Prof. Joseph Davis

School of Information Technologies



## 1. Aim

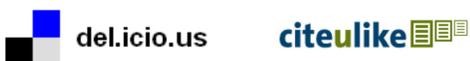
- To integrate folksonomy and ontology for better support of navigation and search on the World Wide Web

## 2. Main Contributions

- The word-formations of folksonomies are discussed.
- A novel integrated computational and crowdsourcing framework is proposed for extracting ontological structure from folksonomies.
- The algorithm of association rule mining has been leveraged in simple low support association rule mining (sLARM) for analyzing a large subset of a folksonomy. Experiments show the promising effectiveness and efficiency of sLARM.
- Unlike traditional ontology building method, our approach involves knowledge from actual community users. Experiments shows that compound tags and jargon tags from users are properly incorporated into the ontological structures as well as standard tags.
- The proposed integrated framework shows impressive results in two applications: multi-dimensional view and cataloguing & indexing

## 3. Introduction

- Collaborative tagging systems (CTS) such as delicious.com, citeulike.org have recently emerged as one of the rapidly growing web 2.0 applications.



- The flat and non-hierarchical structure with unsupervised vocabularies in CTS leads to low search precision and poor resource navigation and retrieval.
- This drawback has created the need for ontological structures which provide shared vocabularies and semantic relations for translating and integrating the different sources.

## Folksonomy

- Describes the flat, social classification structure derived from the practice and method of collaboratively creating and managing tags to annotate and categorize content [1]

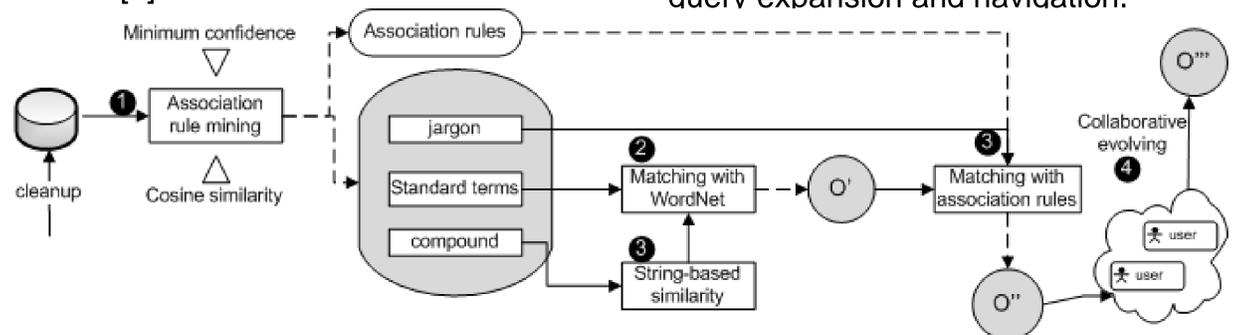


Fig.1. The extraction process for extracting ontological structures from folksonomies

## 4. System Architecture

- Low support association rule mining is applied to generate association rules representing relations between correlated tags.
- WordNet is implemented as an upper ontology providing the semantic is-a relation, which is called as hypernyms in WordNet. Standard tags have been connected to each other via semantic relations from WordNet.
- A series of similarity filters are employed to interpret the compound tags before matching them with WordNet.
- Jargon tags are incorporated into the previously built ontological structure by matching tags using association rules and similarity coefficient.
- A human-aid semantic search assist based on the extracted ontology is introduced to channel users' efforts for ontology evolution.

Rules	Support	Confidence	Cosine	Accept?
folksonomy → tagging	1.59%	0.82	0.722	Y
macroeconomics → economics	0.09%	0.96	0.2671	Y
cyber-ethnography → ethnography	0.06%	1.00	0.2872	Y
asc → collaborator	0.03%	1.00	0.172	N
final → social	0.04%	0.90	0.1679	N
seeking → information	0.03%	0.85	0.1605	N

Table 1. Rules generated with 0.02% support, 80% confidence from Citeulike.org

```

|-science (762)
|-----anthropology (111)
|-----ethnography (128)
|-----biology (256)
|-----genetics (154)
|-----evolutionary-genomics (41)
|-----evolutionary-proteomics (22)
|-----genomics (250)
|-----proteomics (127)
|-----neurobiology (41)
|-----neuroscience (199)
|-----neurophysiology (24)
|-----sociobiology (26)
|-----system_biology (6)
|-----sysbio (74)
|-----cryptography (25)
|-----economics (259)
|-----macroeconomics (21)

```

Fig.2. A fragment of ontological structure in science

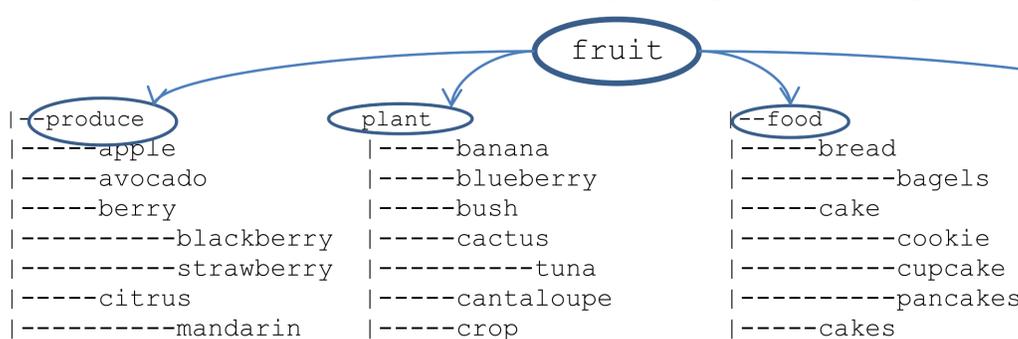


Fig. 3. A fragment output of "fruit" ontological structure, extracted from Flickr dataset

## Ontology

- Describes a specific domain with a set of terms and provides explicit semantic relation between them.
- Improves search with ontology based query expansion and navigation.

## 5. Experiments and Results

- We have conducted experiments and produced useful results based on two folksonomies datasets from Flickr.com and Citeulike.org.
- Table 1 shows the effect of the three thresholds. Low support value helps to preserve rarely occurred pairs while cosine similarity acts as a guard to exclude rules consisting of tag pairs not highly related.
- The evaluation shows that our extracted ontological structure reflects the fruit domain knowledge well and organizes the related resources into several navigable dimensions (Fig.3).
- It also shows a conceptual framework for cataloguing and indexing the resources (Fig.2.)

References:

- [1] Pink, Daniel H. (December 11, 2005). "Folksonomy". New York Times.



The University of Sydney