Original

An ontology evolution method based on folksonomy

Shufeng Wang*, Wen Wang, Yanbin Zhuang, Xianju Fei

School of Computer Information Engineer, Changzhou Institute of Technology, Changzhou, Jiangsu, P.R. China

Abstract

Folksonomies are increasingly adopted in web systems. These "social taxonomies", which emerge from collaborative tagging, contrast with the formalism and the systematic creation process applied to ontologies. However, they can play complementary roles, as the knowledge systematically formalized in ontologies by a restricted group can be enriched by the implicit knowledge collaboratively produced by a much wider group. Existing initiatives that involve folksonomies and ontologies are often unidirectional, i.e., ontologies improve tag operations or tags are used to automatically create ontologies. We propose a new fusion approach in which the semantics travels in both directions from folksonomies to ontologies and vice versa. The result of this fusion is our Folksonomized Ontology (FO). In this paper, we present our 3E Steps technique (Extraction, Enrichment, and Evolution), which explores the latent semantics of a given folksonomy (expressed in a FO) to support ontology review and enhancement. It was implemented and tested in a visual review/enhancement tool.

1. Introduction

A growing number of web systems offer services for content storage, indexing, and sharing. Most of these systems use tag-based social networks to organize and index the stored content. Their users associate free-form tags with each resource, without a central vocabulary. The term folksonomy has been used to characterize the product which emerges from this tagging in a social environment.

In order to analyze, index, and classify their content, web systems compare tags attached to resources. Instead of considering the semantics of each tag in the comparison, tag-based systems usually rely on string matching approaches. While ontologies are increasingly adopted to enrich tag semantics, one common problem with the proposals to associate tags with formal ontologies concerns their unidirectionality, i.e., ontologies improve tag semantics, or the implicit/potential semantics of folksonomies is extracted to produce ontologies.

Differently from traditional techniques, we proposed a fusion approach, called folksonomized ontology (FO), which goes beyond this unidirectional perspective. In one direction, the ontologies are "folksonomized", i.e., the latent semantics from the folksonomic tissue is extracted and fused to ontologies. In the other direction the knowledge systematically organized and formalized in ontologies gives structure to the folksonomic se-

mantics, enhancing operations involving tags, e.g., content indexation and discovery. The folksonomic data fused to an ontology will tune it up to contextualize inferences over the repository.

This paper focus on a technique we propose to support ontology review and enhancement, which we call 3E Steps: Extraction, Enrichment and Evolution. In order to validate our proposal, we developed a tool to extract folksonomic data from Flickr and Delicious and to integrate them into the WordNet ontology. The data is further used in a visual tool that supports ontology review and enhancement.

2. Relate work

Many approaches to automatically or semi-automatically develop ontologies were proposed. Some of them aim at discovering relations and building ontologies from a given corpus of texts (Maedche & Staab, 2000; Maedche et al., 2001; Toledo-Alvarado et al., 2012). Alternative approaches adopt folksonomic data (Specia & Motta., 2007; Van Damme et al., 2007), instead of texts. In either case, the ontologies are built from scratch. These approaches contrast of our proposal, which does not build new ontologies, but departs from existing ones and takes advantage of their structure to build a new entity, which is an enriched (folksonomized) ontology.

Van Damme et al. (2007) employ folksonomic data and lexical/sematic resources, like Leo Dictionary, Wordnet, Google

---

*Corresponding author.
E-mail address: wangsf@czu.cn (S. Wang).

and Wikipedia, to build and to maintain ontologies. They aggregate sets of tags, mapping them to ontology concepts. The relations of those ontologies are mapped back to the folksonomy, in order to produce a social ontology. One important aspect of their proposal is the mechanism to validate the ontology, in which the community that produced the folksonomy validates the results by accepting or discarding the proposed concepts. They employ stemming algorithms to clean the folksonomic data, specially plural nouns and conjugated verbs. This operation occurs in the pre-process phase of the extraction step —as in 3E steps— aiming to improve the quality of the data, grouping the tags that have strong relations.

Specia and Motta (2007) proposed a technique to map clusters of tags to ontology concepts, making explicit the semantics of the tag space. They depart from a set of tags, creating clusters of high-related tags, using co-occurrence information. The relations between these clusters are aligned with external resources like Wikipedia, Google and ontology bases to produce an ontology. Those resources were used to improve the folksonomic data, mainly making explicit the semantics of the tags. In the step of pre-processing, they group morphologically similar tags using Levenshtein similarity metric (Levenshtein, 1966). As the authors point out, this technique could find minor variations (cat and cats, and san francisco, sanfrancisco and san.francisco are the examples given by them). But it is important to mention that the use of this technique could led to undesired results. For example, the distance between the pair of words range and orange is the same as the distance between the pair orange and oranges. While the latter have much semantic relatedness, the former does not. For this reason, we used stemming algorithms to group tags.

Angeletou (2008.) proposed a tool, called FLOR, to perform semantic enrichment on folksonomic data. The first enrichment step is connecting tags to semantic entities and then connecting those entities directly to resources managed by the system. Connecting tags to semantic entities is divided into three sub-steps: *(i)* lexical processing, in which the decision of which tags are meaningful is made, and the normalization step, selecting lexical representations for each tag; *(ii)* semantic expansion, in which the tags are processed in order to disambiguate their meanings, using WordNet; *(iii)* semantic enrichment, where the tags are finally mapped to ontology concepts — ontologies that were found by querying web repositories. But, differently of our approach, the FLOR tool aim to annotate resources annotated by tags with semantic entities. We focused on mapping folksonomic data on concepts of a single ontology, enriching it in the process.

Cattuto et al. (2008) calculated several measures of tag relatedness, based in their co-occurences, mapping them to WordNet synsets (sets of synonyms). They do not group related tags; each individual tag of the folksonomy is associated with a concept in the WordNet ontology. Synsets are sets of synonyms that play an equivalent role of concepts in ontologies. The similarity of the related synsets are then transferred to the respective tags. The step Enrichment of our approach have similar objectives, but are not based only in the co-ocurrences of the tags, but in the topology of the target ontology and the relations between its concepts as well.

Cantador et al. (2011) proposed a mechanism to filter and classify tags, producing a graph of clusters. Then, they mapped these clusters to knowledge bases, like WordNet and Wikipedia, aiming to discover the corresponding semantic entities. Different from previous approaches, in order to map clusters to ontologies, they predefine a set of possible categories and relation types among tag sets, using direct association or natural language processing heuristics. They build categories and them classify tags into them. Our approach does not classify tags in predefined categories, but map them to ontologies concepts. In this way, we can build more malleable entities, apt to easy expansion, in order to accommodate new tags and relations.

Like many of the related work, Tesconi et al. (2008) used external resources, namely Wikipedia, and ontologies like WordNet and YAGO (Suchanek et al., 2007). Their objective was disambiguate tags, "semantifying" them. They developed an algorithm to disambiguate tags, grouping them by sense. Its tagsets are finally linked to Wikipedia categories and ontology concepts, producing social ontologies. Therefore, they cannot take into account tags that does not have a direct map to an external resource (e.g., Wikipedia concept). As folksonomies allow users to create tags as they wish, it is likely that new terms would emerge.

Bang et al. (2008) proposed the concept of "structurable tags", in which tags can be linked through relations, allowing basic inference operations. They expanded the folksonomic model, allowing users to create two types of relations between tags: inclusion and synonymy. These types of relations support the transformation of folksonomic data into semantically richer models. Due to the synonymy relation, the system can group tags with the same meaning. On the other hand, the inclusion relation led to an hierarchical organization, as a simplified ontology. In the 3E steps, the users are not forced to change their natural use of folksonomies. The extra effort of creating relations between the tags is the responsibility of the system, in an automated way.

Heymann and Garcia-Molina (2006) proposed an algorithm to build a graph departing from folksonomic data. It first aggregates tags in tag vectors, in which the *vtl [om]* corresponds to the number of times that the tag *tl* annotates the object *om*. In the resulting unweighted graph, the vertexes will be the tags, and there will be an edge for each pair of tags whose relatedness is above a threshold. The resulting graph, without weights and maintaining just the relevant edges, contains a "latent hierarchical taxonomy". It is captured by an algorithm that builds a subsumption hierarchy, derived from the centrality of each node in the graph.

As ontologies become bigger and more complex, the process of evolving them requires an increasing effort. As pointed out by Ding and Foo (2002), almost all techniques to evolve ontologies require manual intervention. According to Stumme et al. (2000), even though the ontology evolving process requires human experts, in order to address the increasing complexity of modern and large ontologies, tools to support the evolving process are necessary. This observation motivated our work to propose a technique and a related tool to support ontology evolvement.

3. Ontology and folksonomy

3.1. Ontology

Ontology is defined as "an explicit specification of a conceptualization" (Gruber, 1993). An ontology, as a shared conceptualization, expresses a consensus among people, conducting to a consensus among machines. There are several strategies to look for a consensus, e.g., a selected group of representatives and/or specialists designs an ontology incorporating a consensual perspective of a given domain (Levenshtein, 1966); tools extract latent semantics from a body of digital artifacts produced by many people (a statistical consensus), automatically deriving it to an ontology (Cantador et al., 2011).

An ontology is represented as a tuple (G_0, RT, F_{RT}), where $G_0 = <C, E_R>$ is a directed graph with vertex set C formed by concepts and arc set E_R representing relations between concepts. RT is a set of relation types between concepts. F_{RT} is a function $F_{RT}: E_R \rightarrow RT$, which associates a type with each relation (arc). See more details of this model and its relationship with other models (e.g., folksonomies and social ontologies) in Alves and Santanchè (2012), García-Cuéllar et al. (2013), and Rafe et al. (2013).

3.2. Folksonomy

The term folksonomy —combining the words "folk" and "taxonomy" (Vander, 2007)— has been used to characterize the product which emerges from this tagging in a social environment.

A folksonomy can represent a perspective of a wider group, but the semantics extracted from the implicit relations among tags are rather simple. An ontology is usually built by a more restrict group, but has the richness of an engineered product. There are initiatives towards exploring the interplay between folksonomies and ontologies. However, one common problem concerns their unidirectionality, i.e., in one direction there are proposals to use ontologies to improve tags' semantics, in the other direction there are proposals to extract the implicit/potential semantics of folksonomies in order to produce ontologies.

In folksonomy-based systems, users can attach a set of tags to resources. These tags are not tied to any centralized vocabulary, so the users are free to create and combine tags. Some strengths of folksonomies are their easiness of use and the fact that they reflect the vocabulary of their users (Mathes, 2004). In a first glimpse, tagging can transmit the wrong idea of a poor classification system. However, thanks to its simplicity, users are producing millions of correlated tags. It is a shift from classical approaches —in which a restricted group of people formalize a set of concepts and relations—into a social approach —in which the concepts and their relations emerge from collective tagging (Shirky, 2005). In order to perform a systematic folksonomy analysis, to subsidize the extraction of its potential semantics, researchers are proposing models to represent its key aspects. Gruber (2007) models a folksonomy departing from its basic "tagging" element, defined as the following relation: Tagging (object, tag, tagger, source), in which object is the described resource, tag is the tag itself —a string containing a word or combined words—, tagger is the tag's author, and source is the folksonomy system, which allows to record the tag provenance (e.g., Delicious, Flickr, etc.).

4. The ontology revolution based folksonomy

4.1. Folksonomized ontology

We define a folksonomized ontology as an ontology aligned with terms of a folksonomy and enriched with their contextual data. By contextual data we mean data which emerges from a statistical analysis of a folksonomy, e.g. tag frequency, co-occurrence and information content. The ontology to be folksonomized can be of any kind, which meets the abstract model presented in the beginning of section 3.1. The choice of the domain covered by the ontology and the folksonomy have direct impacts in the results. The results will be as good as the overlap between their domains.

In one direction, the FO, which is aligned with tags, drives richer semantic-based matching, categorization and tag suggestion. In the other direction, contextual data is used to review and improve the ontology. Figure 1 schematizes the roles played by an ontology and a folksonomy in a folksonomized ontology building.

A FO is defined as a tuple (G, RT, F_1, F_2, F_3), where $G = <V, E>$ is a directed graph with vertex set V formed by ontology concepts and arc set E representing relations between these concepts, RT is a set of relation types between concepts, F_1 is a
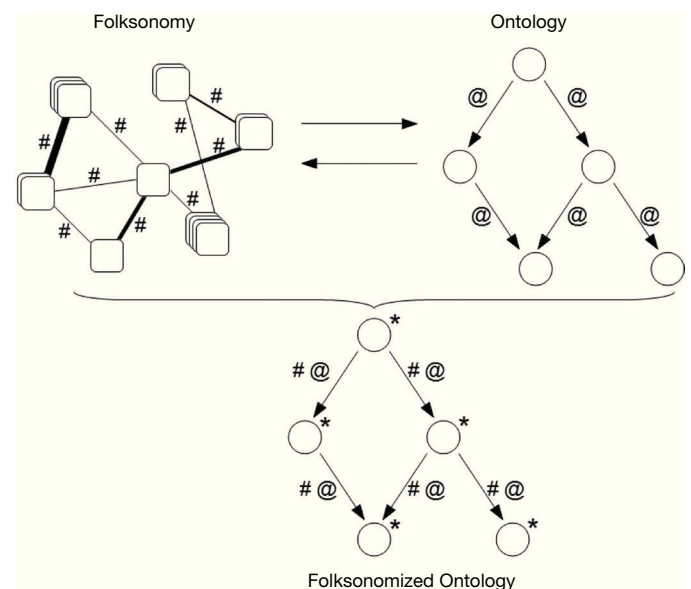

Fig. 1. Folksonomized ontology. #: weight; *: infortmation content; @: typed relationship.

weighting function $F_1 : E \rightarrow N$, where the weight of the relation is derived from the total of co-occurrences between tags represented by the respective concepts, the function $F_2 : E \rightarrow RT$ defines the type of the relation as in ontologies, and the function $F_3 : V \rightarrow R$ associates the information content *(ic)* related to each concept, calculated by $ic(c) = -\log p(c)$, where $p(c)$ is the probability of a given concept *c*. This *ic* value also derives from a statistical analysis of the folksonomy and will substantiate computations of semantic similarity between the concepts using.

4.2. The step on ontology revolution

In the following sections, we will describe our technique, illustrated in Figure 2, involving three steps: Extraction — the folksonomy data are mined in order to collect the metadata used in the next step; Enrichment — the latent semantics from the folksonomic tissue is extracted and fused with ontologies, and it comprises the map and fuse phases; and Evolution — the ontology managers could analyze the FO data and visualize the cases in which the collaborative knowledge indicates that the ontology needs to be reviewed and/or enhanced. The first two steps (gray boxes) are divided in phases illustrated inside the gray boxes.

4.2.1. Extraction

This step is organized in two phases: collect and pre-process/aggregate. The collect phase involves accessing external systems in order to retrieve tag data from primary sources. The pre-process/aggregate phase cleans the data and aggregates tags according to their meaning in tagsets.

Step 1: collected tag data

Web-based content portals offer web service interfaces to access their data (APIs). Due to the heterogeneity in the APIs, the tag data collecting module was designed to be customizable and it was tested in Delicious and Flickr systems. It access these web services to select and retrieve tags and their metadata, which are stored in a database.

In order to better obtain the emergent properties of the semantics extracted from folksonomies, this module was designed to afford large datasets. They are stored as triples of resources, users and tags, including their relations. Statistical data were computed and stored during data collection, avoiding extra post-processing work. These data co-occurrence between tags and frequency (used to calculate probability and information content) feeds subsequent phases that compute the values addressed by the functions F_1 and F_3 of a FO, previously in the formal model. The updating process is incremental, i.e., it collects and stores just the differences of a previous execution.

Step 2: pre-process tag data

In the pre-processing phase, unusual tags were eliminated to improve the quality of the tag set. We classify as unusual those tags with low number of occurrences, i.e., $\leq$ a constant LO. The value of LO varies according to the size of the data set and the domain.

After this pre-processing phase, we aggregated tags that refer to the same term. For instance, the tags tip and tips are tightly connected and represent the same term. The grouping algorithm is divided in two steps: marks analysis (the algorithm groups tags differing only by special characters, e.g., " ", "-", ".", etc.), and morphological analysis — it groups tags by morphological relatedness.
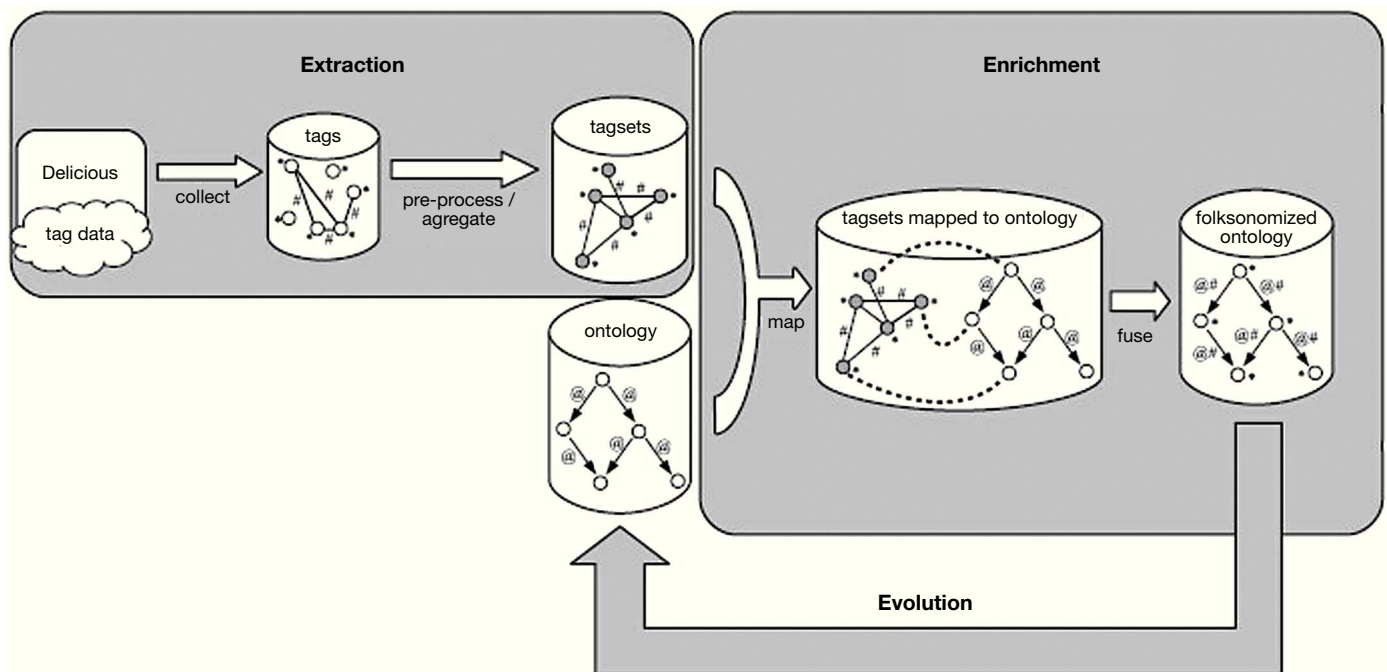

Fig. 2. 3E steps technique.

In order to represent multiple-word tags, users resort to different strategies, e.g., concatenating words with or without separating signs. By analyzing the similarity of tags without the special characters, we group tags like search-engine, search engine, and search engine. These tags are clearly very close to each other and represent different user approaches for using multiple-word tags. So, all special characters of tags were removed, allowing to group tags that became equal without punctuation.

The morphological analysis and grouping go beyond spelling comparisons, considering morphological variations as singular and plural tags, or tags of different verb tenses. The algorithm retrieves morphological variations of tags from the WordNet ontology, grouping them together.

4.2.2. Enrichment

This step is organized in two phases: map and fuse. The map phase involves mapping tagsets produced in the previous phase to concepts of an ontology. The fuse phase involves fusing the ontology to the folksonomic data to produce the folksonomized ontology.

Step 3: mapping tag into ontology concept

The map phase is not a simple task, due to the lack of semantic information related to the tagsets. The tagsets cannot be directly mapped based on their words, since the same word can have multiple meanings in the ontology. In WordNet, for instance, a word can have multiple senses, called synsets, which are differentiated through identifiers combining the original word plus two affixes.

The first one is a character that describes the synset type (namely noun, verb, adjective, or adverb) and the second one is a sequential number to differentiate each meaning. For instance, the synset dog.n.01 represents a noun and it is one of the synsets for the word dog.

To find out which concept (or which synset in WordNet) corresponds of each tagset, we developed a technique that encompasses the relation of concepts (WordNet synsets) and tag co-occurrences, divided in three steps: *(i)* tagset key election — a tag of each tagset is chosen to represent it; *(ii)* co-occurrence selection — the co-occurrence values of the tags are selected; and *(iii)* tagset key mapping — finally, each tagset is mapped to an ontology concept.

The algorithm illustrated in Figure 3 map group keys to synsets.

Step 4: fuse ontology into folksonomy

The fusion phase combine the data from the previous phase to produce a single unified Folksonomized Ontology. The resulting graph presented in Figure 2 is expanded in Figure 1. It departs from the ontology and enriches it with the statistical data obtained in the previous phases. Therefore, concepts (which were mapped to tagsets) are enriched with information content (asterisks) and their relations are enriched with co-occurrence rates (sharps).

Resuming the formal model presented in section 3.1, the graph G, the RT set and the function F_2 are derived from the preexisting ontology. The functions F_1 and F_3 represent the enrichment computed from folksonomies.

4.2.3. Evolution
Step 5: review and enhancement

During the implementation of the two previous steps we observed that our approach to relate folksonomies and ontologies produced a rich data set, which can support ontology review and enhancement:

- A popular tagset without a respective concept in the ontology. It can indicate a candidate to a new concept to be added in the ontology.
- A strong relation between two tagsets that has no correspondent relation between the respective concepts. It can indicate some important relation not represented in the ontology.
- Tagsets embed rich information about relations among tags and concepts. A tagset aggregates many tags around a meaning. Its internal network of relations and the connections they have with the concepts in the ontology are rich sources for the analysis of how words are related to the meaning of concepts.

Therefore, the two previous steps were incorporated in our 3E Steps technique, in which the third step is the evolution of the ontology. This is also an important step, because it leads to a symbiotic cycle, in which folksonomized ontologies help to tune up the underlying ontologies which, in turn, will improve the results of the folksonomized ontology itself.

5. Visual review/enhancement tool

In order to support ontology evolvement, we developed a tool to visually explore the interplay between the latent semantics of the folksonomy system and a given ontology. It is important to point out that our tool is built upon the idea of offering more information to the managers of the ontology. It is designed to explore data and to suggest changes, but does not apply any automatic modification and does not offer support for ontology editing.

```
Algorithm to map group keys to synsets
Input:  G: set of groups key(tags)
Input: minmap: minimum co-occurrence mapping
Output: S: set of group keys(tags) mapped to synsets
    1: S←{}
    2: while exist t in G| fit(t) do
    3: t ←choose(G)
    4: cooc_list ← getcooc(t)
    5: list←{}
    6: for all synset s in synsets(t) do
    7:   for all elememt e in cooc_list do
    8:     include(s,sim(s,synmap(e),coocval(t,e))) in list
    9:   endfor
   10: endfor
   11: S[t] ←max(list)
   12: remove t from G
   13: end while
```
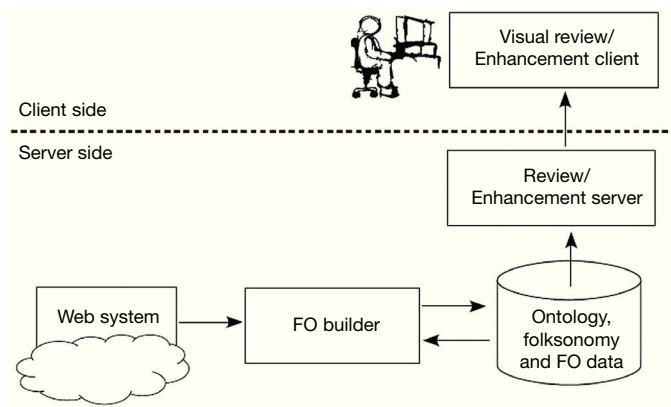
Fig. 3. Mapping algorithm.

Fig. 4. Architectural diagram of the review enhancement tool. FO: folksonomized ontology.

5.1. Implementation aspects

Figure 4 presents an architectural diagram of the tool. These are the main modules:

Web system. Web-based repositories aimed at sharing content, links or metadata (e.g., Delicious and Flickr) which use tag-based classification mechanisms.

FO Builder, responsible for collecting folksonomy data from web systems. It implements the two first steps of the 3E Steps technique (see Fig. 4). The module is implemented in Python and uses SQLite to manage the database. It is designed to be extensible, i.e., it allows developers to extend it to work with different kinds of web systems. The default implementation works with Flickr and Delicious. The Delicious extension adopts a third party library, the Delicious API. The module retrieves and stores the data in an incremental way, i.e., only the difference from the previous processing is stored, saving processing resources. The data is further filtered, cleaned, and homogenized as described in the 3E Steps technique. Finally, the module maps tagsets to ontology concepts and fuse them to produce a FO.

Review/Enhancement Server. The interactive module (responsible for visually presenting the data to the user) is designed to run over the browser. It is organized in two sub-modules, a server module implemented in Python and a client module implemented in HTML + JavaScript. The server sub-module is responsible for the following operations: *(i)* it builds the HTML + JavaScript module from a template and dispatches to the client (web browser); and *(ii)* it interacts with the database retrieving and filtering relevant information for the client.

Visual Review/Enhancement Client. This module uses a third-party library, JavaScript InfoVis Toolkit, to provide the interactive visualizations. It is responsible for visually presenting data to the end-user. This tool is detailed in the next sub-section.
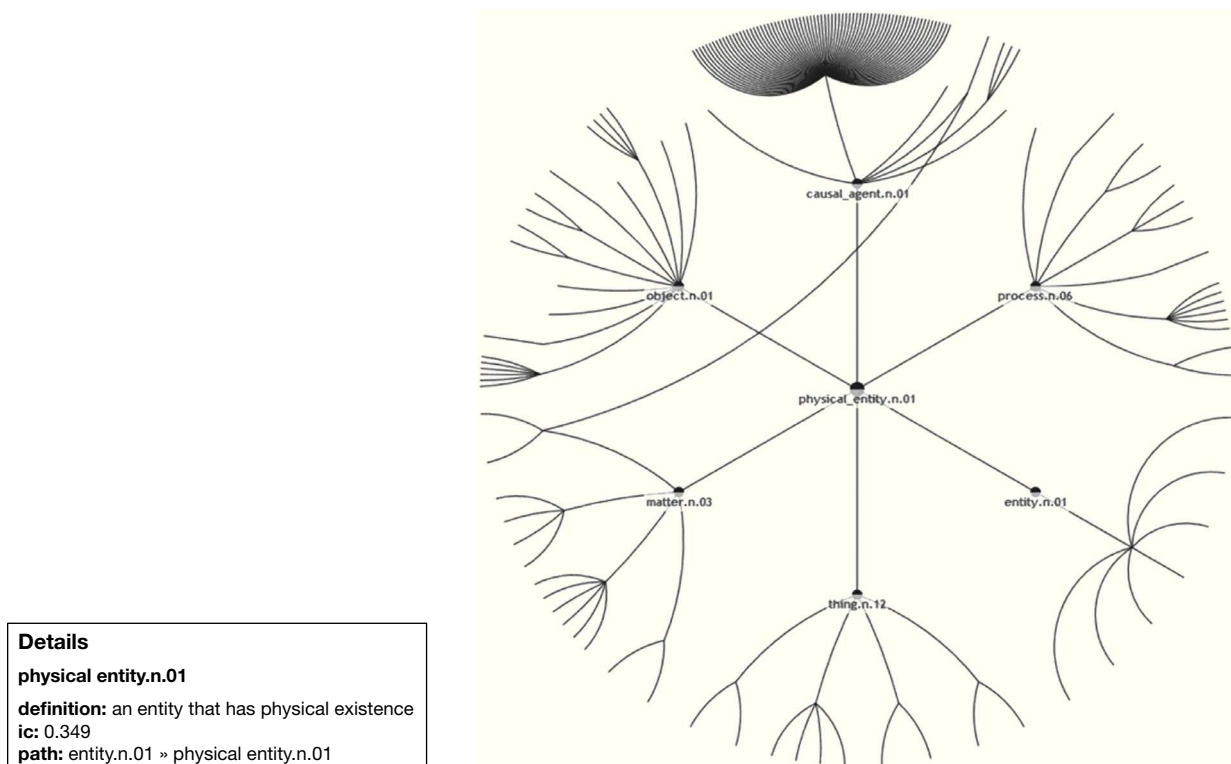

Fig. 5. Folksonomized ontology (FO) visualization.

5.2. Visual Review/Enhancement

Figure 5 shows a screenshot of the main screen of the Visual Review/Enhancement Client, illustrated in Figure 4. It is organized in two main areas: control panel (displayed in the bottom side) and an interactive graph area (displayed in the top side). The control panel can switch among three possible sub-panels: navigation, details and history.

In Figure 5, the Details Panel is selected. In the interactive graph area, a segment of an FO is displayed, centered in the physical entity node. The tool generates an interactive graphical representation of the segment. In this representation, the user can drag the nodes, zoom, and pan the visualization. When a node is clicked, the information associated with it is showed in the Details Panel. The data can be explored in two modalities: relations and concepts.

5.3. Analyzing relations

The goal of this modality is to analyze the relations among concepts in the ontology and confront them with relations captured from the folksonomy. Figure 6 shows a typical graph presented for analysis. It derives from our abstract model presented in sub-section 3.1. There are three parameters that control the details of the visualization: more nodes, virtual nodes, and edge.

As mentioned before, the Details panel shows the correspondent information of the selected node. It shows co-occurrence and *ic* values, representing the functions of the formal model F_1 and F_3, respectively.

There are two approaches for navigation: overview or compare. In the overview approach, the user can freely navigate in the entire FO tree, by using the interactive focus provided by the hyperbolic tree. In the compare approach, the user will analyze a given pair of concepts, the path of relations that connects both and the relations with near concepts. Therefore, the first step for an analysis in this approach is to select that pair of concepts.

When the user selects the compare approach to navigate, it is possible to manually assign two concepts of the FO or enter in the assisted mode, in which the tool finds two candidates to review. The basic principle of the assisted mode is to look for relevant discrepancies among data coming from the ontology and from the folksonomy. The current version can find two concepts in the FO that have a weak similarity in the originating ontology, but have a strong connection in the originating folksonomy.

In the assisted mode, the tool runs the following process to automatically two candidates for analysis:
- A set of candidate concepts is selected to be tested. The size of this set can be configured.
- The tool creates an analysis tree with a branch of the original FO for each selected candidate. This branch includes all concepts (and the respective relations) that can be reached departing from the candidate concept, in a given customizable depth.
- The next step involves finding two nodes with low similarity (considering the path coming from the FO) and high co-occurrence (considering the data extracted from the folksonomy).
- For each concept in a given branch, the tool tests if its co-occurrence with the candidate concept is higher than a threshold. If so, the distance of both is tested.
- If the two conditions —higher co-occurrence and long distance (low similarity)— are satisfied, the pair of concepts is selected to be analyzed in the visual tool.

5.4. Practical examples of relations analysis

In this sub-section, we will show practical examples of the tool, illustrating its support for ontology enhancement and the improvements achieved by the use of FOs.

Figure 7 shows a visualization generated by the tool. The pair of concepts to be analyzed in the visualization is (bible, christian), and the common ancestor —entity— is highlight-
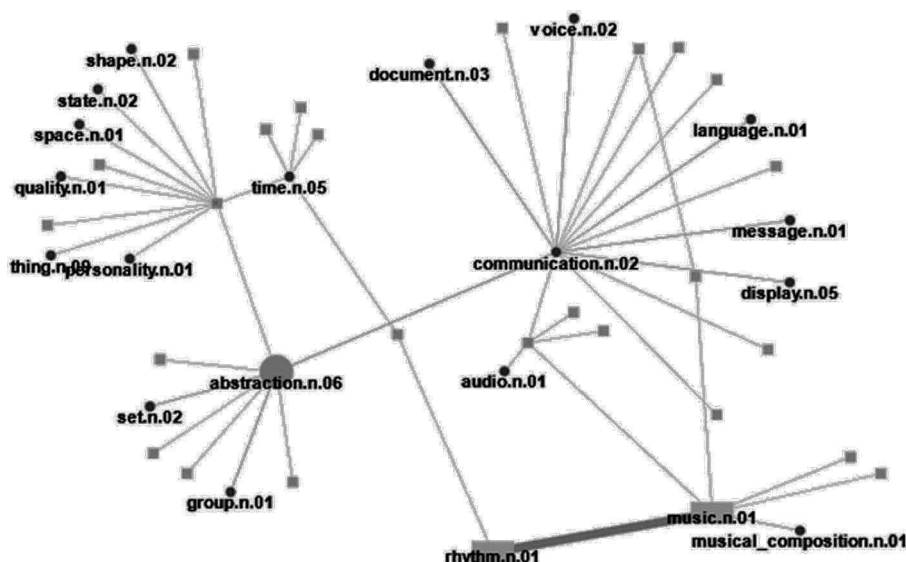

Fig. 6. Graph to visualize and to analyze relations.
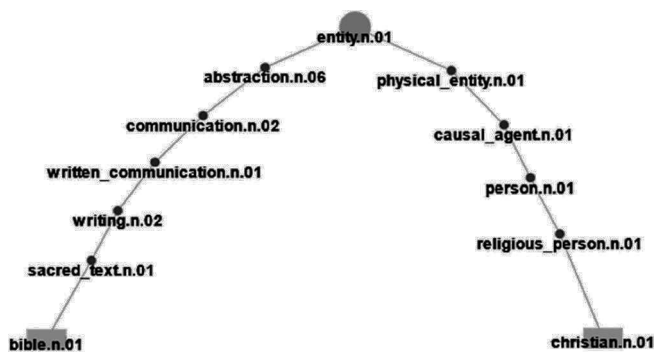
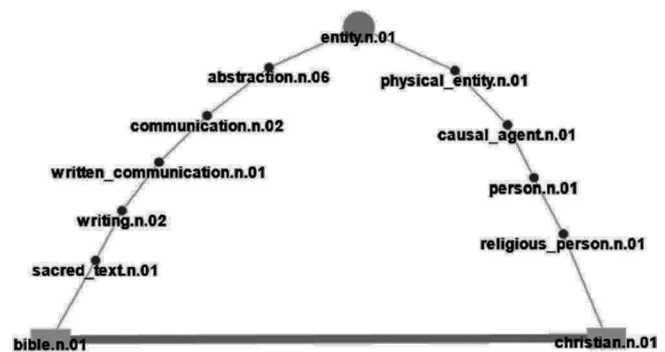Fig. 7. Default visualization.

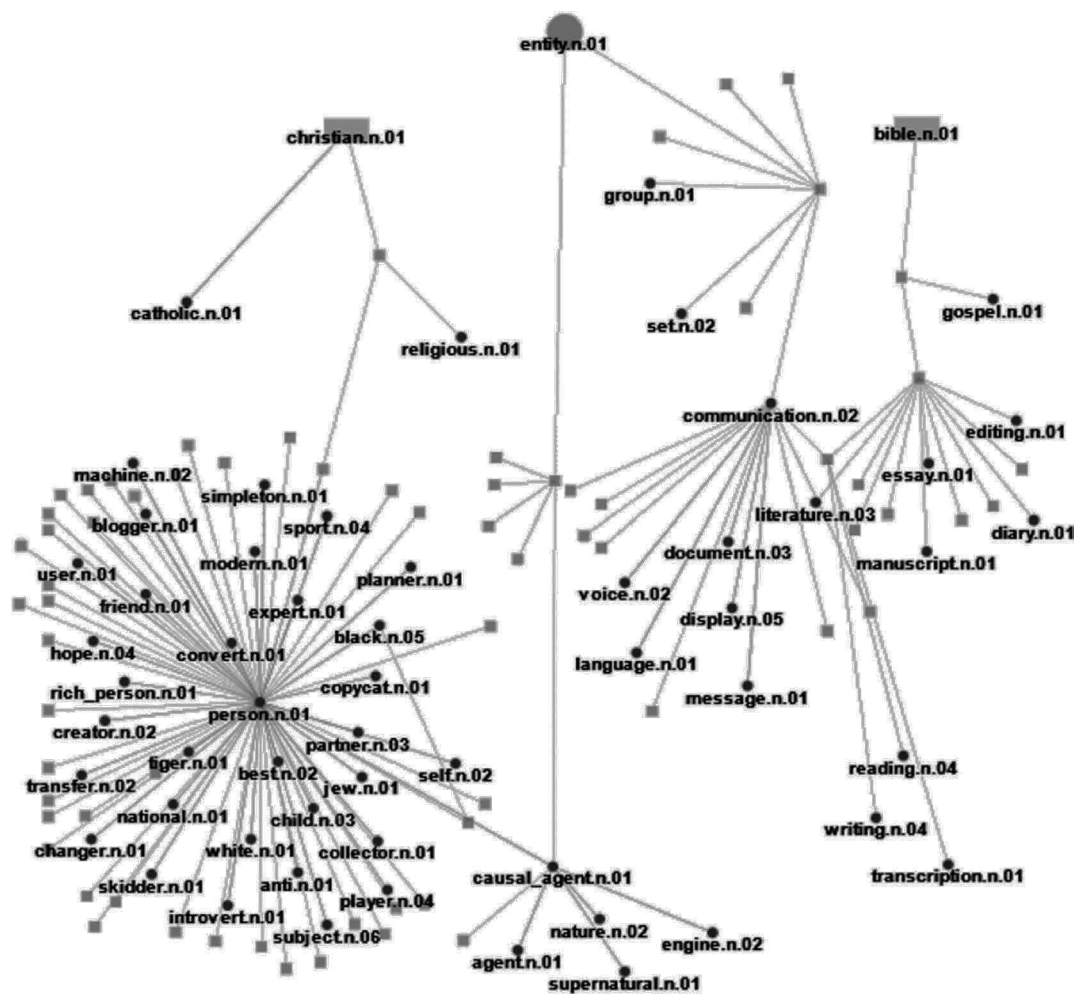Fig. 8. Visualization with more nodes, virtual nodes, and edge.

Fig. 9. Visualization with more nodes and virtual nodes.

ed as well. This example shows a scenario in which the tool can be used to improve the ontology. The concepts bible and christian are separated by a long path and their common ancestor, entity, is the most general concept —the root— in the ontology. Any ontology based approach to compare the terms will return a low similarity, due to the long path and

the generic common ancestor, which has zero of information content.

When the parameter edge is activated, the tool draws a strong edge between bible and christian, as they have a strong co-occurrence in the originating folksonomy (Fig. 8). This edge does not mean that both nodes should have a direct link, but just empha-

sizes that something must be reviewed or improved in the ontology, considering the observations of the folksonomy.

In this scenario, the ontology managers, facing the task of enhancing the ontology, can use the tool to find and visualize how and where the ontology could be improved. If they need more nodes in the visualization, they can use the respective parameter to increase the number of nodes.

Figure 9 shows the resulting visualization when More Nodes and Virtual Nodes parameters are selected. The distinction between regular and virtual nodes makes explicit the contrast between concepts shared by the ontology and the folksonomy, and concepts of the ontology not present in the folksonomy. This is an useful synthesis to analyze the popular use of concepts present in the ontology.

Besides the review/enhancement process, the visualization tool can be used to inspect the improvements of the FOs, when confronted with traditional ontologies.

In Figure 10, the tool is focusing on the pair of nodes (graphics, inspiration). This pair has a high relation in the folksonomy, but they are separated by a great distance in the ontology. Our practical experiments showed that when statistical data (information content and co-occurrences) embedded in the FOs are explored in the similarity algorithm, they achieve better results than ontologies alone. This case shows that the FO is an entity that can support improvements in the operations over its data, because it can use more than one source of semantics folksonomies and ontologies.

The following example shows another improvement of the FO in the opposite direction. In Figure 11, the tool is focusing on the pair of nodes (war, conflict). Conflict is a concept coming from the originating ontology, which has no corresponding tagset in the folksonomy. Conflict is a virtual node, inserted to keep the topology of the ontology. Conflict is a virtual node, and in the traditional folksonomies it would not be considered. Since the FO considers virtual nodes in the comparison algorithm, even if they do not appear in the folksonomy, it will achieve better results in this kind of scenarios when compared with traditional folksonomies.

5.5. Inside concepts

As mentioned in sub-section 4.2, the second approach to review/enhance ontologies allows the user to inspect inside a tagset and its relation with other tags in the folksonomy.

As described in section 3.1, in order to relate tags to concepts of an ontology, the tag is not evaluated alone. There is a network of relations among a tag and other tags, which have high co-occurrence with it. This network is essential to provide a context to the tag. It is the basis to relate a tagset to a given concept.

A graphical presentation of a tag and its cooccurrences is therefore a rich source of information. Figure 12 shows that presentation. The intensity evaluation of relations among tags considers also transitive relations among tags.
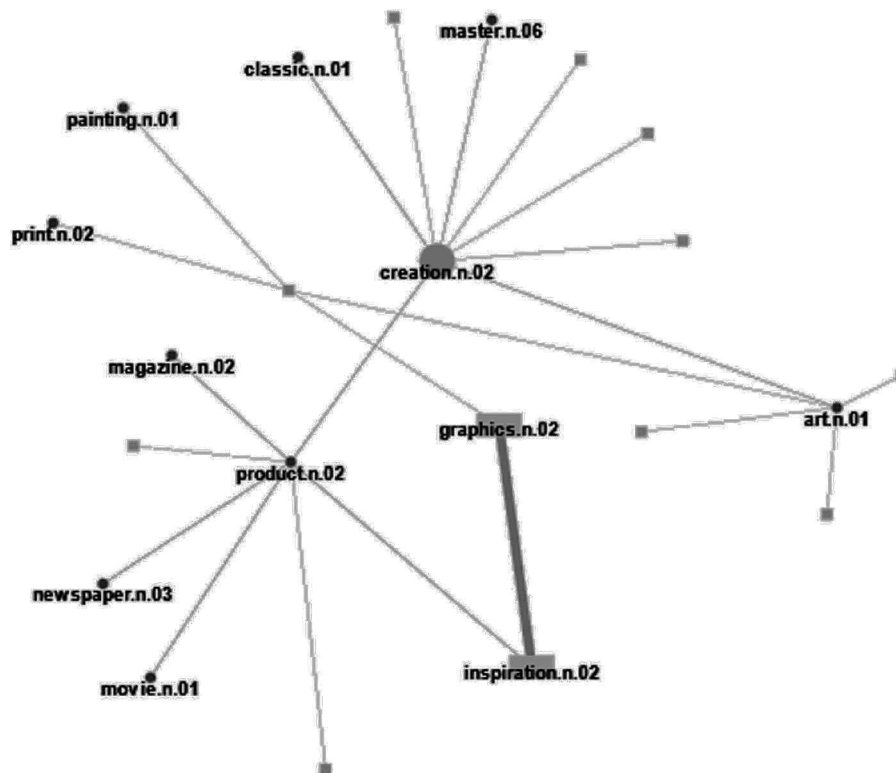

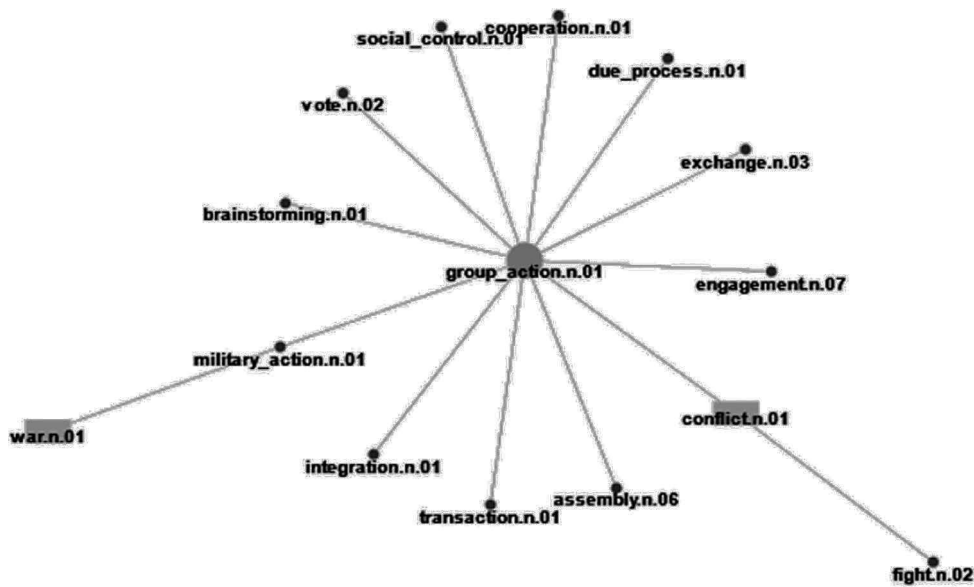Fig. 10. Visualization improvement of folksonomized ontologies.

Fig. 11. Visualization improvement of virtual nodes.

6. Conclusions

This paper presented our 3E Steps technique to review and enhance ontologies and our approach to build and use a folksonomized ontology (FO) in this context. A FO is a hybrid entity fusing folksonomies and ontologies. It is a symbiotic combination, taking advantage of both semantic organizations. Ontologies provide a formal semantic basis, which is contextualized by folksonomic data, improving operations over tags based on ontologies. Conversely, the FOs were used as tools to analyze the ontology and to support the process of ontology evolvement, showing the discrepancies between the emergent knowledge of a community and the formal representation of this knowledge in the ontology.

In this paper, we described the 3E Steps: Extraction, Enrichment, and Evolution. Extraction is the step where the semantic information is collected from the folksonomies and processed. In the Enrichment step, we combine the two entities, building a third one, with the best of both worlds. Finally, Evolution is the step where the folksonomized ontology is used to support the review and enhancement in the original ontology, closing the circle.

In our point of view, the work presented here opens an interesting field of applying latent semantics, socially produced by wide communities, to improve engineered ontologies. Related work addressing ontologies and folksonomies does not explore the full potential of this interaction, due to their unidirectionality. Our fusion approach explores the symbiotic complementarity of ontologies and folksonomies.

Future work include: *(i)* to expand the folksonomized model to include other relations (besides the generalization); *(ii)* to run tests in specialized contexts applying domain ontologies; and *(iii)* to expand our tool that allows the visualization of the individual tags inside a cluster, improving the observation of the interrelation between the tags.
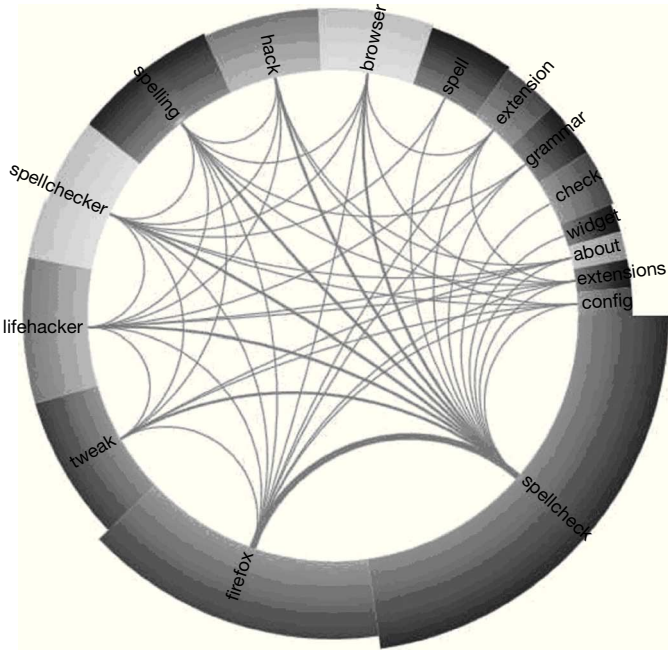
Fig. 12. Diagram to inspect a tagset.

References

Alves, H., & Santanchè, A. (2012). *Formal aspects of social ontologies and folksonomized ontologies*. In 4th International Workshop on Semantic Web Information Management.

Angeletou, S. (2008). *Semantic enrichment of folksonomy tagspaces* (pp. 889-894). In: Proceedings of the 7th International Conference on The Semantic Web.

Bang, B.H.K., Dané, E., & Grandbastien, M. (2008). *Merging semantic and participative approaches for organising teachers' documents* (pp. 4959-4966). In Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications: AACE 2008.

Cantador, I., Konstas, I., & Jose, J.M. (2011). Categorising social tags to improve folksonomy-based recommendations. *Web Semantics: Science, Services and Agents on the World Wide Web, 9,* 1-15.

Cattuto, C., Benz, D., Hotho, A., & Stumme, G. (2008). *Semantic grounding of tag relatedness in social bookmarking systems* (vol. 5318 of LNAI, pp. 615-631). In The Semantic Web ISWC 2008.

Ding, Y., & Foo, S. (2002). Ontology research and development part 2 — A review of ontology mapping and evolving. *Journal of Information Science, 28,* 375-388.

García-Cuéllar, M.M., Ochoa-Tello, L., Atrián-Salazar, M.L., Palacio-Mejía, L.S., Hernández-Ávila, J.E., & González-González, E.L. (2013). Web Tools 2.0 for Health Promotion in Mexico. *Journal of Applied Research and Technology, 11,* 708-713.

Gruber, T. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition, 5,* 199-220. doi: 10.1006/knac.1993.1008

Gruber, T. (2007). Ontology of Folksonomy: A mash-up of apples and oranges. *International Journal on Semantic Web & Information Systems, 3,* 1-11.

Heymann, P., & Garcia-Molina, H. (2006). *Collaborative creation of communal hierarchical taxonomies in social tagging systems*. In Tech. Rep. 2006-10, Computer Science Department, Standford University.

Levenshtein, V. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady, 10,* 707-710.

Maedche, A., & Staab, S. (2000). *Semi-automatic engineering of ontologies from text* (pp. 5-7). In Proceedings of the 12th Internal Conference on Software and Knowledge Engineering.

Maedche, A., Maedche, E., & Volz, R. (2001). *The ontology extraction maintenance framework text-to-onto*. In Proceedings of the ICDM'01 Workshop on Integrating Data Mining and Knowledge Management.

Mathes, A. (2004). *Folksonomies — cooperative classification and communication through shared metadata*. Retrieved on April, 2014 from http://www.adammathes.com/academic/computermediated-communication/folksonomies.html

Rafe, V., & Goodarzi, M.H. (2013). A novel web-based human advisor fuzzy expert system. *Journal of Applied Research and Technology, 11,* 161-167.

Shirky, C. (2005). *Ontology is overrated: categories, links, and tags.* Retrieved on May, 2014 from http://www.shirky.com/writings/ontology overrated.html

Specia, L., & Motta, E. (2007). *Integrating folksonomies with the semantic web* (vol. 4519 of LNCS, pp. 624-639). In: Proceedings of the European Semantic Web Conference (ESWC2007).

Stumme, G., Studer, R., & Sure, Y. (2000). Towards an order-theoretical foundation for maintaining and merging ontologies. In Bodendorfand, F., & Grauer, M. (Eds.), *Verbundtagung Wirtschaftsinformatik* (pp. 136-149). Shakerand Aachen.

Suchanek, F.M., Kasneci, G., & Weikum, G. (2007). *Yago: a core of semantic knowledge* (pp. 697-706). In Proceedings of the 16th International Conference on World Wide Web. WWW '07.

Tesconi, M., Ronzano, F., Marchetti, A., & Minutoli, S. (2008). *Semantify del.icio.us: Automatically Turn your Tags into Senses.* In First Workshop on Social Data on the Web (SDoW2008). Retrieved from http://CEUR-WS.org/Vol-405/paper1.pdf

Toledo-Alvarado, J.I., Guzmán-Arenas, A., & Martínez-Luna, G.L. (2012). Automatic building of an ontology from a corpus of text documents using data mining tools. *Journal of Applied Research and Technology, 10,* 398-404.

Van Damme, C., Hepp, M., & Siorpaes, K. (2007). *Folks Ontology: An Integrated Approach for Turning Folksonomies into Ontologies* (pp. 57-70). In Proceedings of the ESWC Workshop "Bridging the Gap between Semantic Web and Web 2.0".

Van der Wal, T. (2007). *Folksonomy.* Retrieved on April, 2014 from http://vanderwal.net/folksonomy.html