Computing (2019) 101:1489–1511
https://doi.org/10.1007/s00607-019-00704-9

The role of collaborative tagging and ontologies in emerging semantic of web resources

Sara Qassimi[1] · **El Hassan Abdelwahed**[1]

Received: 18 March 2018 / Accepted: 14 January 2019 / Published online: 25 January 2019

Abstract

The social web interactions have extended the sharing and the growth of web resources on the web. The collaborative web services (folksonomies) enable users to assign their freely chosen keywords (tags) to describe web resources. The advent of folksonomy has evolved the role of web users from consumers to contributors of information. Thus, users attribute their descriptive tags to annotate, organize and classify web resources of interests. Folksonomy became popular with the emergence of collaborative tagging. It offers a practical classification of web resources via the attributed tags. Nonetheless, the freely chosen tags weaken the semantic description of web resources. Folksonomy can give rise to a poor classification system based on ambiguous and inconsistent tags. Therefore, it is essential to pertinently describe the semantic of web resources to enhance their classification, findability and discoverability. The proposed approach represents a combined semantic enrichment strategy that explores collaborative tagging towards describing each web resource using different types of descriptive metadata, namely relevant folksonomy tags, content-based main keywords and matching ontology terms. The experimental evaluation has shown relevant results attesting the efficiency of our proposal. The alignment of social tagging with the ontology will not only enhances the classification of web resources but also constructs their semantic clustering. This emergent semantic will establish new challenges to improve the context-aware recommender systems of web resources in different real-world applications (healthcare, social education and cultural heritage).

Keywords Folksonomy · Semantic web · Ontology · Web resource · Emergent semantic · Recommender system

✉ Sara Qassimi
sara.qassimi@ced.uca.ma

El Hassan Abdelwahed
abdelwahed@uca.ac.ma

[1] LISI Laboratory, Faculty of Sciences Semlalia Marrakech, Cadi Ayyad University, Marrakech, Morocco

1 Introduction

The web is regularly extending the growth of its vast repository of web resources. A web resource is any identifiable thing on the web (e.g. images, videos, scientific articles, selling items, etc.). The availability and accessibility of knowledge on the web have influenced the user search behaviour. Web users browse the web by observing and heeding one available web content to another. They usually explore the web without a planned search strategy. Thus, they tend to move on quickly from one web resource to another when their contents are not easily understandable, unintelligible and not directly useful [1]. The lack of a complete indexation or classification of web resources decreases their discoverability and findability. It has called the attention to the importance of extracting pertinent descriptive information from the extended set of shared web resources to enhance their classification. Therefore, it is relevant to describe each web resource with its descriptive "metadata" that express clear and meaningful information by pertinently summarizing its content. In traditional libraries, professional indexers or domain experts use controlled vocabularies to assign terms "experts keywords" which appropriately identify the main topic of a web resource. However, the owners of large sets of various web resources prefer using advanced automatic technologies for the classification process. The consulting of professional indexers requires a costly and intensive task to maintain the classification of the rapid spread-shared web resources. The need for automatic and semi-automatic processes of expressing web resources' main topic has increased throughout this technological century. For instance, the process of extracting the main keywords from a resource's textual content involves text mining techniques like the tools of natural language processing. Although, the expert's terms and the content-based main keywords describing the web resource can be incomprehensible to the users. It has to contain also non-expert annotations, like the users' freely chosen keywords called tags. The provided advantages of social annotation services (folksonomy) enable users to order, locate and re-find their web resources by themselves. The generated folks' tags collaboratively classify the shared web resources. Folksonomy defines the process of using users' tags for the classification of different types of web resources. It is known also as collaborative tagging, social classification and social indexing. For example, CiteULike users employ freely chosen tags to share and classify their reference lists. Rather than including only annotations of experts, the use of non-expert or novice users annotations leads to more comprehensive folksonomies [2]. Furthermore, the recent semantic web researchers believe that collaborative tagging is more reliable knowledge sources than free texts [3]. The popularity of tagging has been introduced by famous web-based systems such as Flickr, CiteULike, YouTube, del.icio.us and Instagram. The web users attribute tags to annotate various types of resources, including images, videos and audios. It is an effective technique that expresses the wisdom of the crowd [4]. Different aspects of folksonomy have been explored in information retrieval [5], social network analysis [6], data mining [7], recommendation systems [8–12], and others.

Regardless of its popularity, folksonomy lacks semantics [13]. The tags "folks' keywords" are derived from an uncontrolled and unsupervised vocabulary. The social tagging brings up inconsistent and ambiguous tags. The attributed irrelevant tags lead to misapprehend web resources. Regardless of the misspelling, synonymy and poly-

semy of tags, and their infrequency and uncommonness, abbreviations also reduce and weaken the description of web resources. For instance, the abbreviation "Ca" has several significations like calcium and cancer. The word "plethora" means a large amount of something but expresses also an excess of a bodily fluid or blood in medicine. The lack of semantics paves the way to irrelevant annotations weakening the web resources' semantic description and therefore their classification.

This article aims to pertinently describe the semantic of web resources by using collaborative tagging and ontologies. The purpose is to enhance the descriptive annotations of web resources by solving folksonomy's weaknesses. Indeed, relevant annotations "metadata" will not only improve the semantic description of web resources but will also enhance their clustering and organization. The proposed approach combines semantic annotation strategies towards increasing the comprehension of web resources. It stands on constructing an emergent semantic of web resources by efficiently gathering their relevant descriptive metadata. This paper explores the advantages of folksonomy and ontology to extract relevant web resources' descriptors "metadata", namely relevant folksonomy tags, content-based main keyword and matching ontology terms.

The rest of the paper is organized as follows: Sect. 2 presents the motivating applications and the purpose of this paper within the overall research challenges. The related work is reviewed in Sect. 3. Section 4 depicts the proposed approach of the combined emergent semantic strategy to extract pertinent descriptive metadata of web resources. The experimental evaluation is described in Sect. 5. Section 6 presents different alternatives and perspectives of comparing the semantic similarity of web resources. Finally, the conclusion and future directions are delineated in Sect. 7.

2 Research challenges and motivating applications

The main challenges of our research study are conducted within a global project deployed on three levels (see Fig. 1). One of the main motivations stands on constructing an emergent semantic of web resources (see Fig. 1, Level I). It consists of combining semantic annotation strategies by investigating collaborative tagging [13]. The purpose is to enrich the description of web resources with a combined semantic annotation. Instead of annotating the resources with ontology's terms, we are aiming to investigate the extent to which the collaborative tagging can enhance the resources' description, comprehension and categorization. The extraction of a descriptive semantic for each web resource will emerge from different types of descriptive metadata, namely the relevant tags from the folksonomy, the extracted content-based main keywords and the matching terms from a domain ontology. To illustrate this approach, we consider a healthcare scenario. Social media is a powerful tool for raising awareness and advocacy regarding public health issues [14]. Patients can benefit from using social media services through networking, exchanging relevant information and receiving medical support. Healthcare leaders are aware of the importance of sharing and spreading knowledge through social interactions. Physicians participate in online communities to communicate and interact with their colleagues and patients [15]. However, social media tools, like folksonomy, present potential risks to patients

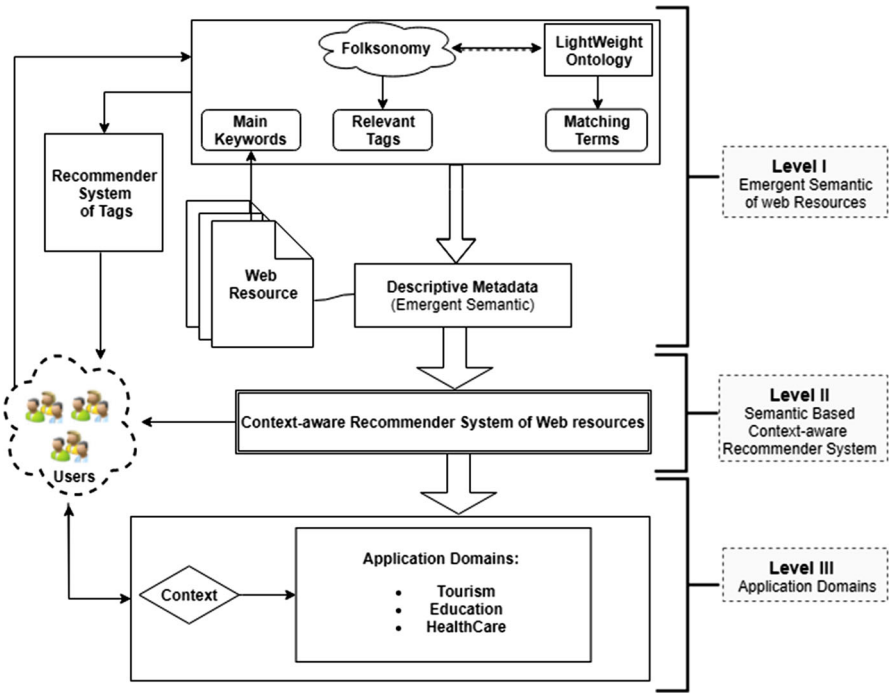

Fig. 1 General architecture of the overall research study

and healthcare professionals regarding the distribution of poor-quality information [16]. The unsupervised nature of folksonomy tags may reduce their effectiveness of describing interesting resources, thereby hindering the task of resources' classification and indexing and users searching. Therefore, it would be convenient to increase web resource description by applying the proposed combined semantic annotation method that uses not only relevant folksonomy tags but also content-based main keywords and matching ontology terms. The descriptive semantic emerges from the wisdom of the healthcare professionals (Ontologies) and the folks' interactions (Folksonomies) describing health-related resources. The emergent semantic of web resources will enhance their organization and clustering using semantic similarities. Consequently, it will increase the chances of discovering and finding interesting resources that users might not have come across yet through their searching. This semantic relatedness of web resources will improve the information filtering system, like recommender system, to assist users in selecting relevant resources that best meet their needs and preferences. The emergent semantic (see Fig. 1, Level I) will be used to enhance the context-aware recommender system (CARS) of web resources (see Fig. 1, Level II). A recommender system is a leading tool and technique available for users to speed up the information seeking by retrieving the most relevant items from the large information sets. The recommender systems usually employ the collaborative filtering (CF), content-based (CB) and hybrid-based recommendations methods [9]. The CF analyzes the behaviours of users (e.g. rating, tagging and liking items) to filter items of users

with similar preference patterns. The CB filtering approach focuses on the content of items (e.g. its keywords, features and characteristics) to suggest similar items matching the user's previously preferred items. The hybrid-based recommender system combines the two or more filtering recommendations approaches. The use of the context awareness in recommender systems filters items based on contextual information provided by the application domain. The context is any useful information that has an impact on the users' interactions with the system [17]. The contextual information can be static (e.g. the user's date and place of birth, gender and ethnicity) or dynamic (e.g. location, time, the user's family status and his activities). The context information may precisely affect the recommendations. For example, in the touristic domain, a user will be interested in visiting a particular site depending not only on his preferences but also on the weather, the timing, the proximity, and even the year's season. In healthcare domain, recommendations based on user's preferences might contradict the user's health conditions. The system should not recommend nearby candy stores for a diabetic person who likes sugary foods. The recommender system's computational process incorporates the contextual information in the definition of features characterizing the item (or, resource) and the user profiles. For example, the contextual features can be the common location (longitude and latitude data) of both the available touristic places (static contextual information) and the user (dynamic contextual information). The contextual filtering strategy defines the contextual information as features joined to the emergent semantic describing each item to enhance its significance. Therefore, it will reduce the searching task of the item's filtering by discarding a part of available items matching the user's profile. The selection of the closet items to the user's preferences is measured by computing user-user, item-item and item-user similarities, since the items and users profiles have the same dimensional features' space (e.g. the user's profile is described as a vector of his contextual information, the attributed tags exposing his preferences for certain items; the item's profile is described as a vector of its contextual information and its emergent descriptive semantic (metadata: relevant folksonomy tags, main keywords and matching terms)). The emergent semantic of resources (or, items) can lead to construct and explore clusters of semantically related items annotated by a particular user, then extract his used tags describing them in order to maintain the specificity of the user profile vector corresponding to each domain. The CARS has a great impact on facilitating the process of decision making in many real-world applications. The use of semantic-based CARS will be deployed in education, tourism and healthcare application domains (see Fig. 1, Level III).

In tourism, establishing a semantic-based context-aware recommender system will enhance the valorization of the cultural heritage by suggesting historical places that suit the visitor's interest. For example, the CARS filters items by considering the similarity of visitors (based on their same age, ethnicity, gender and the same assigned tags describing a visited place), the similarity of historical places (based on their similar descriptive annotations) and the geographic proximity (based on the contextual location information). In education, the collaborative tagging is an adequate metacognitive strategy that successfully engages learners in the learning process [18]. Folksonomy tags add semantics, comprehensible for learners, describing open educational resources (OER) (freely accessible and openly licensed texts, medias, e-books, online videos, tutorials, reading reports, etc.). The intake of using collaborative tag-

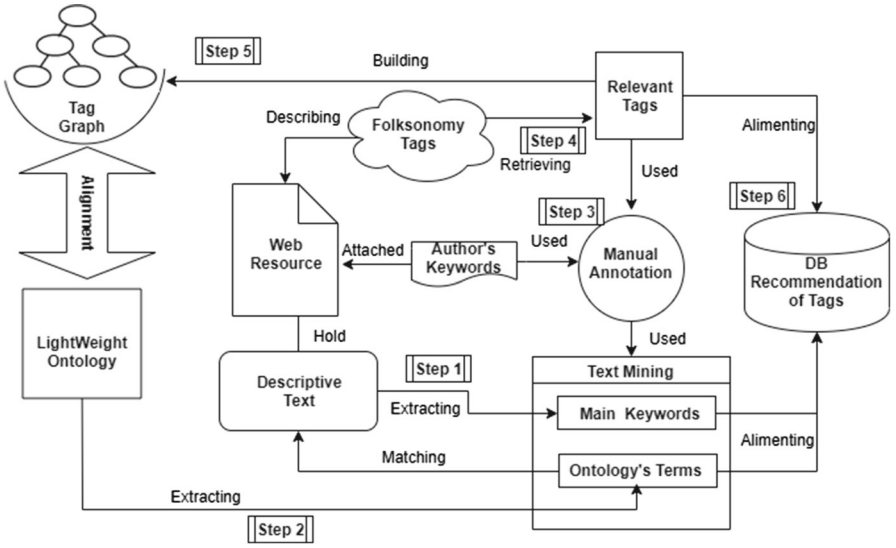

Fig. 2 Combined emergent semantic annotation approach

ging to construct the emergent semantic of educational resources will advance their recommendations. The generated folksonomy will enhance the closeness between the user (the tags' provider) and items (described by the user's tags). In healthcare, the healthcare CARS recommend resources about symptoms and therapies enhancing the awareness and providing useful guidelines to the appropriate end users (the patient, his family and close friends). For health professionals, the health recommender system is a decision support system. The organization of patient's electronic health record (EHR) annotated with relevant descriptive metadata (extracted content-based keywords, assigned tags by physicians, matching medical terms) will aid healthcare professionals in decision-making. The semantic annotations describing patients' EHRs will cluster patients having the same health matters. For example, the emergent semantic (Level I) based CARS (Level II) will detect fitting similarities between patients and their archived EHRs, then generate meaningful recommendations for a diabetic patient's case to prevent complications in diabetes mellitus. This paper mainly focuses on the first step of the proposed architecture (see Fig. 1, Level I and Fig. 2).

3 Related work

The process of annotating web resources is performed with different main shortcuts descriptors "metadata" depending on whether the main topic originate from text contents (keywords), controlled vocabularies (terms) or collaborative tagging systems (tags) [19].

Each web resource usually holds a rich text content. Data mining algorithms can do the extraction of information to retrieve the resource's relevant keywords. The advantage of using content-based annotation enables an automatic keyword extraction

process independent of human involvement. The content-based annotation strategy relies on key phrases or keywords extraction methods that derive main keywords from the web resource's text content. The existent online RESTFul APIs "semantic annotators" analyze a text to identify its relevant sequences of words and link them to pertinent Wikipedia pages. Though, they are unable to outperform keyword extractors [20]. The automatic keyword extraction approach "extractive summary" is classified into four categories, namely, simple statistical, linguistics, machine learning and hybrid approaches [21]. The keywords extraction method is improved by considering machine learning models that combine several features. For instance, the two competing methods: the hybrid genetic algorithm GenEx [22] "Genitor and Extractor" and KEA [23] "Keyphrase Extraction Algorithm" that generates and filters candidates based on their weights of features. More attention has been given to KEA for its open availability and simplicity of use [24,25]. The keywords extraction methods have achieved impressive results but require training data. The unsupervised extraction techniques use heuristic filtering to compensate the lack of training data by using complex analysis like shallow parsing (deep analytics) or statistical-based methods based on an independent domain like KP-Miner [26]. The main disadvantage of content-based annotation method is the limitation consistency of the resulting keywords based only on the description given by the web resources' authors. Even though it offers certain flexibility without a controlled vocabulary, it lacks semantics (e.g. unclustered synonyms).

An expert has an advanced and a high level of knowledge about a particular domain [27]. The terms assigned by professional indexers construct a controlled vocabulary (e.i. ontology and thesauri) depicting a strong knowledge representation by expressing semantic relations [28]. The controlled vocabulary-based annotation method is called term assignment or subject indexing method. The term assignment method uses a controlled vocabulary to select terms that best match the resources' descriptive. The controlled vocabulary-based annotation process tries to find mappings between the web resource's candidate terms and the concept's terms in the controlled vocabulary. It expresses the web resource's descriptive metadata extracted from knowledge-based concepts. Consequently, the classification of web resources can make use of semantic relationships in the ontology to accomplish enhanced categorization, like exploring the relationship among broader or more specific concepts. The term assignment method has been applied in different areas of knowledge organization and retrieval. The Gene Ontology (GO) provides the logical structure of the biological terms and their relationships. The bioinformatics initiative maintains the GO annotations relating a specific gene product to a specific ontology term [29]. The authors in [30] used a physician annotated corpus to identify, extract and rank medical terms from each electronic health record (EHR) notes of patients. The semantic-based recommender system HealthRecSys [10] provides relevant education health websites to complement the selected health videos. The algorithm selects candidate terms from diabetes-related videos' textual content and cross-match them with Bio-Ontology terms. Recent automatic identification of the resources' terms methods are based on large web knowledge repositories Wikipedia, either by constructing Wikipedia Hierarchical Ontology (WHO) [31] or based on probabilistic model based on DBpedia hierarchical model [32]. Other works [33,34] relied on semantic technology to build a classification and indexing system of web resources (respectively, sports images and building information modeling (BIM)

resources). They used ontology theory to semantically describe web resources, then facilitate their retrieval and searching process. However, users can only employ the provided concepts' terms to describe their web resources. The use of terms extracted from the controlled vocabulary to annotate web resources can generate misapprehension and incomprehension for non-expert and novice users.

The social tagging has the advantage of producing a large scale of tags. The purpose of collaborative tagging approach is to generate tags matching the human understanding of the web resource "abstractive summary". The authors in [35] consider the large numbers of users' generated tags on social tagging systems to produce a social classification of web resources. Social tags are helpful to identify the users' preferences and the resources' characteristics. The exploration of tags' information and their interaction dynamically adjusts the recommendations [36]. However, the collaborative tagging suffers from the inconsistency of tags: polysemous and synonymous tags [37]. A hybrid approach [38] exploited social annotations to describe resources by relating tags to concepts from WordNet and Wikipedia. This strategy associates tags with conceptual entities to improve web resources' classification. Another alternative to address tags' inconsistency problem is to use automatic tags' suggestions [19]. Thus, tag recommendations limit the redundancy and the ambiguousness of tags. The recommender system of tags controls the wide variety of tags and requires less cognitive effort to assign them. The authors in [39] came up with a method based on user tagging status to improve the quality of tag recommendations. However, they investigated the archived tagging behaviours of users without considering the new user status. Most of tag recommendations' techniques use the strategy of finding similar tagged resources, then ranking the selected collection of tags. This strategy restricts the suggestion only on pre-existing tags. Similar approaches have been adherent by combining multi-features "tag frequency, co-occurrence and document similarity" [40]. Almost none of the research of the tagging field have explored term assignment and keyword extraction methods to support failures of tagging methods.

Inside out this analysis, there are three approaches of assigning descriptive annotations (see Table 2). They address the descriptive semantic of web resources with different methodologies "keyword extraction, term assignment and social tagging" (see Table 1).

The current controlled vocabulary-based approaches employed background knowledge in the form of a hierarchical ontologies [10,31,32] or based on expert annotation corpus (thesaurus) [30] to improve the performance of text mining algorithms for extracting resources' terms. However, maintaining and enriching an ontology within the rapid growth of shared web resources is expensive in term of time spending and professional indexers services expenses. Besides, web resources might have insufficient or absent textual content or inaccessible representative data [35]. Insufficient available resources' descriptive data overburdens the automatic text mining tasks. In the folksonomy, multi-authors (folks) are producing collections of tags which represent the textual descriptive annotations of the large set of web resources. Moreover, the semantic web researchers have focused their discussions on social involvements, rather than coping with the extraction of knowledge from free texts [41].

Compared to this related works, we propose a combined annotation method that semantically enriches the description of web resources by exploring collaborative

Table 1 The approaches of assigning descriptive annotations

Approach	Advantages	Disadvantages	Method
Content-based	Automated process; Not involving human; Avoid cold start	Lack of semantics; Computationally intensive; Limited notion;	Keyword extraction
Controlled vocabulary-based	Expert terms; Semantic web; Main topic	Costly process; Difficult scalability; Uncommon language; Time spending	Term assignment
Folksonomy-based	Large scalability; Social indexing; Wisdom of the crowds; Common folks words; Diversity	Polysemy and synonymy; Lack of semantics; Cold start problem; Ill-formed words; Uncleaned tags	Social tagging

Table 2 Comparison of related works

Works	Approach			Annotation		
	Controlled vocabulary-based	Content based	Folksonomy based	Term	Keyword	Tag
[21]		✓			✓	
[26]		✓			✓	
[24]		✓			✓	
[25]		✓			✓	
[29]	✓			✓		
[10]	✓			✓		
[30]	✓			✓		
[33]	✓			✓		
[34]	✓			✓		
[31]	✓			✓		
[32]	✓			✓		
[35]			✓			✓
[36]			✓			✓
[38]	✓		✓	✓		✓
[39]			✓			✓
[40]			✓			✓

tagging (integrating human cognition) and bridging between the advantages of the discussed approaches.

4 Proposed approach: a combined emergent semantic annotation

The proposed approach retrieves relevant tags from the folksonomy, extracts main keywords from the resource's text content with a reference of controlled vocabulary's

matching terms. The approach describes a combined semantic annotation of describing web resource's content. Extracting keywords from web resource's text content could be inconsistent. For instance, two authors might publish similar web resources described with different main keywords. Consequently, it is relevant to extract their set of matching terms using a controlled vocabulary represented by a lightweight ontology. The steps of the proposed methodology (see Fig. 2) are as follows.

4.1 Content-based main keywords and extracted ontology terms

The process of extracting main keywords aims to describe the main topic of a web resource. The automatic keyword extraction process is handled by machine learning methods as a supervised learning problem which needs a training dataset and classifiers. It has been extensively addressed using the open software KEA [23] which uses supervised machine learning method based on naive Bayes classifiers. KEA is used either to automatically extract keywords or key phrases from free text (content-based main keywords) or from a controlled vocabulary (matching terms). It has encouraged several researchers [42,43] to adapt or extend KEA to perform the extraction of keywords from text content. Therefore, our approach considers an extension of the KEA's classifier to extract content-based main keywords and ontology's terms. The proposed approach explores folksonomy tags to build a model that learns the extraction strategy from the manually assigned annotations.

Step 1 The act of extracting main keywords consists of two stages [42]. The first stage involves generating candidates keywords by using stop words and tokenizing text into sentences then extracting candidates (one or more words). The extracted candidates are reduced to their roots by applying a stemmer (e.g. Lovins stemmer [44]). The second stage is about filtering candidates keyword that involves generating features for each candidate. The commonly used features are: The frequency of each candidate (TFxIDF score combines the word's frequency with the inverse document's frequency to select relevant frequent keyword); the occurrence (a candidate appears at least more than two times); The type of a candidate (noun phrase, not exceed trigrams); The positioning of the candidate in the text content (beginning and end). In the filtering stage, several features are computed for each candidate as inputs for the machine learning model to obtain the probability of being the main keyword indeed.

Step 2 The extraction of the set of terms matching the text content of a web resource relies on matching each candidate term to the descriptive of the ontology' concepts [42]. It is operated by generating candidate terms from text content using techniques of normalization: collecting words that match the length of the longest term in the vocabulary, lowercasing, removing stopwords and stemming. Then, each candidate term is ranked based on their semantic relatedness computed by comparing its relatedness to all other candidates terms. The more a candidate is related to others, the more is significant. The filtering stage avoids disambiguation during the mapping. The use of a machine learning technique computes the probability "score" for each candidate keyword and candidate term of being respectively a content-based main keyword and

a matching ontology term. The final set of main keywords and matching ontology terms are selected by setting a threshold (a limit number of the top ranked candidates).

Step 3 The main keywords and matching terms extraction strategies have many supervised extraction systems based on the KEA, like Maui [42]. However, the exploration of folksonomy tags has not previously been used in the extraction strategy. The Multi-purpose Automatic topic Indexing keyword extraction system (Maui) is KEA's reincarnation that uses Wikipedia as a reference. Maui uses a supervised algorithm based on bagging decision trees classifier to rank candidates. The extraction strategy is learned from the manual annotation that uses the keywords assigned by the resources' authors. The novelty of our proposed approach stands on exploring relevant extracted tags from the folksonomy: the manual annotation is created not only with the prerequisite authors' keywords but also with the relevant folksonomy tags that additionally aliment the training data. The higher the size of the training data is the more accurate the performance of the classifier becomes. However, the approach considers only relevant tags among the amount of generated folksonomy tags. The use of both relevant tags and authors' keywords in the manual annotation will improve the classifier's accuracy, and consequently will enhance the extraction strategy of obtaining more accurate content-based main keywords and matching ontology terms.

4.2 Retrieving relevant folksonomy tags

The folksonomy tags are not only describing web resources but also summarizing their content "abstractive summary" by expressing the users' understanding. None of the standard algorithms has achieved yet the abstractive summary done by humans [21]. Thus, the tags reflect users' opinions, attract readers and invite them to bring their own tags. Besides, the keywords of the resources' authors are often not sufficiently expressive for ordinary users. However, the folksonomy lacks semantics.

Step 4 Tag processing is required in order to handle low quality of the generated folks' tags. The use of a spell checker tool and a blacklist of forbidden words will eliminate personal, misspelled and multi-word tags (e.g. "BreastCancer" and "Breast-Cancer"). The folksonomy suffers from inconsistent tags due to its uncontrolled vocabulary. Though, applying a stemmer will reduce words' variation to their stems (e.g."Infectious" and "Infection" are reduced to their root word "Infect"). The consistency of each tag can be assessed by finding it in a thesaurus, or it has to be used by at least two distinct users depending on the size of the community. To better solve the quality degradation of folksonomy, different tags quality measurements are possible by applying guidelines, rules and regulation [45]. The more experts assign a term as a quality tag, the more it is assumed to be relevant. Nonetheless, more comprehensive folksonomies emerge from non-expert or novice users' tags than from experts' tags only [46] . Therefore, the proposed approach considers the extraction of tags which are frequently used and understood by many users of the community. A community of users $U = \{u_h\}$ annotate a set of web resources $R = \{r_k\}$ with a set of tags $T = \{t_i\}$. Where, $1 \leqslant h \leqslant l$; $1 \leqslant k \leqslant m$; $1 \leqslant i \leqslant n$ and l, m and n are finite numbers.

We consider a resource $r_k \in R$ described by a set of tags from T. The extraction of relevant tags describing this resource r_k is computed by considering the degree of frequency of each tag t_i (1), denoted by $DF(r_k,t_i)$.

$$DF(r_k, t_i) = \sqrt{FT(r_k, t_i)^2 + FU(r_k, t_i)^2} \tag{1}$$

where $FT(r_k,t_i)$ is the Frequency of the tag t_i annotating the resource r_k (2);

$FU(r_k,t_i)$ is the Frequency of users who use the tag t_i to annotate the resource r_k (3).

$$FT(r_k, t_i) = \frac{Number\ of\ times\ the\ tag\ t_i\ is\ used\ to\ describe\ the\ resource\ r_k}{Number\ of\ tags\ used\ to\ describe\ the\ resource\ r_k} \tag{2}$$

$$FU(r_k, t_i) = \frac{Number\ of\ users\ who\ use\ the\ tag\ t_i\ to\ annotate\ the\ resource\ r_k}{Number\ of\ users\ who\ annotate\ the\ resource\ r_k} \tag{3}$$

The relevant tags are those with higher degree of frequency.

Step 5 The purpose of constructing a hierarchical graph of tags is to highlight the differences between tags having the same meaning (synonymous tags). It constructs taxonomic relationships (broader, narrower) among tags. The hierarchy of tags is built based on the inclusion index [3]. $I_i(t_i,t_j)$ measures the inclusion of the tag t_i regarding the tag t_j (4). For example, "$I_1(t_1,t_4) > I_4(t_4,t_1)$" scales how general the tag t_1 is compared to another tag t_4 (i.e. the tag t_1 is broader than tag t_4). Consequently, each tag t_i has its inclusion score $S_i(t_i)$ that identifies how strongly the tag t_i is related to other tags (5).

For $t_i , t_j \in T , t_i \neq t_j$

$$I_i(t_i, t_j) = \frac{Number\ of\ resources\ described\ by\ both\ tags\ t_i\ and\ t_j}{Number\ of\ resources\ described\ with\ the\ tag\ t_j} \tag{4}$$

$$S_i(t_i) = \sum_{j=1}^{n} I_i(t_i, t_j) \tag{5}$$

Implicit relationships also play an essential role in enhancing the organization of web resources. Such as defining tags' community clustered into groups of semantically close tags [47]. The association between tags, resources and users will enhance the precision of detecting relevant tags and their semantic relationships (Fig. 3). The generated folks' tags semantic graph is considered as a undirected graph whose nodes represent the tags linked together by edges $W(t_i,t_j)$. The weight $W(t_i,t_j)$ identifies the semantic relationships among tags (6). It scales how strongly two tags t_i and t_j are semantically related regarding their commonly usage by distinct users $W_u(t_i,t_j)$ (7) and their joint assignment to describe web resources $W_r(t_i,t_j)$ (8).

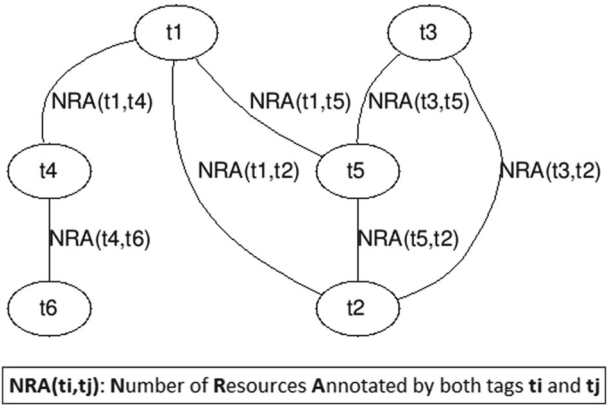

Fig. 3 Joint tagged resources driven tags' graph

$$W(t_i, t_j) = \sqrt{W_r(t_i, t_j)^2 + W_u(t_i, t_j)^2} \tag{6}$$

$$W_u(t_i, t_j) = \frac{Number\ of\ users\ who\ use\ both\ tags\ t_i\ and\ t_j}{Number\ of\ users\ in\ U} \tag{7}$$

$$W_r(t_i, t_j) = \frac{Number\ of\ resources\ described\ by\ both\ tags\ t_i\ and\ t_j}{Number\ of\ resources\ tagged\ with\ tags\ in\ T} \tag{8}$$

Therefore, the emergent folks' tags semantic graph (see Fig. 4) is beneficial to describe the relationship among web resources annotated with connected tags. For instance, the recommender system of tags will take advantage of the emergent folks' tags semantic graph to recommend semantically close tags. It allows a graph-based reasoning about the relationships between tags attributed to describe different resources. The reasoning of the folks' tags semantic graph can be extended by projecting tags on the ontology's concepts descriptive. On the other hand, the ontology can benefit from the emergent semantic graph of folks' tags by adding new terms (relevant tags) that clearly describe related contents. The folksonomy and ontology alignment will enhance the ontology's concept descriptive with additional information provided not only from new frequently used tags but also from their semantic relationship. The enrichment of the ontology's concepts is done due to mapping relevant tags to the matching concept's attributes guided by the formalism of the Simple Knowledge Organization System SKOS [37]. The Vocabulary SKOS [48] is a common data model formulated on Resource Descriptive Framework. Its aim is to describe ontology's concepts and their semantic relationships (broad, narrow and related).

Step 6 The preference of using a tag depends on the user's motivation. There are two types of users involved in tagging: the categorizers who employ their mental models and personal preferences; the describers who summarize the resource's content using mostly synonyms [49]. The users' interest might change gradually with the passage of time and so for the significance of the used tags. Consequently, the relevance and significance of the generated tags are related to the closeness to the current period of time

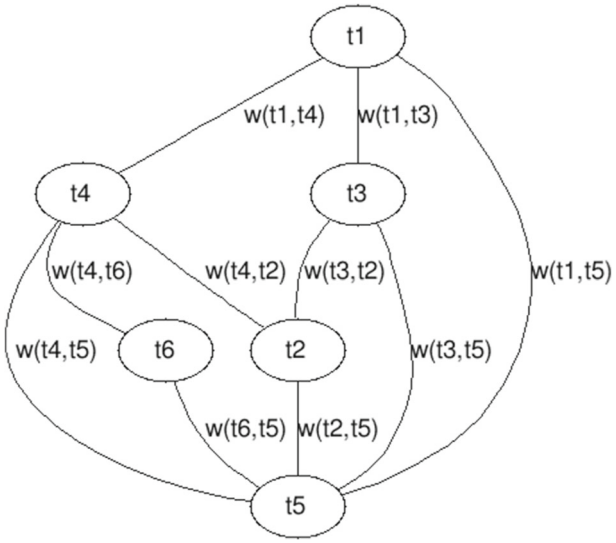

Fig. 4 Folks' tags semantic graph

[50]. The influencing factors in the user tagging behavior have a direct impact on the folksonomy tags' quality. Besides, the significant variations of tag usages describing a web resource are induced because of the lack of guidelines. In a matter of fact, our study proposes the recommendation of tags to enhance the quality of the generated folksonomy and improve the web resources' attributed tags. The tag recommendations can enhance the convergence of the folksonomy to a common vocabulary constructed with more reliable descriptive and heterogeneous tags. Accordingly, it can alleviate the drawbacks of folksonomy mentioned before (synonymy and polysemy of tags). The recommender system of tags incentivizes users to annotate a large number of resources. The suggested tags will reduce the users' cognitive load dealing with choosing the appropriate tags to describe a resource. The previously assigned web resources' tags will influence users' choices of assigning new descriptive tags [51]. However, little attention is given to new web resources, the suggestion of tags relies only on the previously assigned tags to the same or similar web resources. Consequently, it will be pertinent to recommend the main keywords and matching ontology terms of the new never-tagged resource in the cold start. Therefore, the database of the recommender system of tags will be alimented with relevant folksonomy tags also with the extracted main keywords and their matching ontology terms. The recommender system of tags will narrow the gap between the uncontrolled nature of tags and the conceptual terms of the ontology.

5 Evaluation and results

In order to evaluate the performance of the proposed approach, we collected 550 random bio-medical articles "web resources" (Figs. 5, 6) described with their authors'

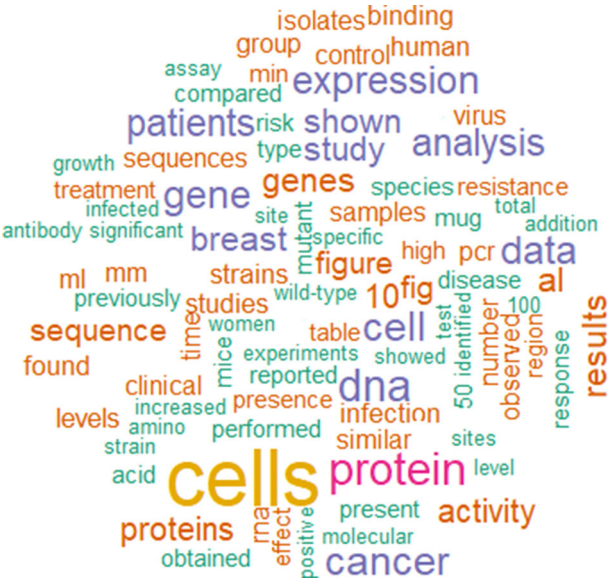

Fig. 5 Words cloud of the corpus using the statistical software R

keywords and annotated by tags from the "folksonomy" CiteULike [52]. We used the Medical Subject Heading (MeSH) as the controlled vocabulary "lightweight ontology". MeSH terms [53], managed by the U.S National Library of Medicine, describe bio-medical research items. We set 548 bio-medical articles for training and 2 articles for testing, namely Article A [54] and Article B [55]. We compared the python implementation of the Rapid Automatic Keyword Extraction RAKE [56] against the Multi-purpose automatic topic indexing Maui [42]. The comparison of their performances is measured using the standard information retrieval measures [57]: Precision P is the percentage of correct annotations "keywords or terms" among those extracted (9); Recall R is the percentage of correctly extracted annotations "keywords or terms" among all correct ones (10); F-Measure F is the combination of both P and R (11). The manually assigned annotations are the correct annotations.

$$P = \frac{Number\ of\ correct\ extracted\ annotations}{Number\ of\ all\ extracted\ annotations} \tag{9}$$

$$R = \frac{Number\ of\ correct\ extracted\ annotations}{Number\ of\ all\ correct\ annotations} \tag{10}$$

$$F = \frac{2 \times P \times R}{P + R} \tag{11}$$

For better performance, we consider ensemble machine learning algorithms to extract content-based main keywords and matching ontology terms. The multi-classifiers aim to enhance the precision of the model's prediction. They train multiple models by using the same learning algorithm, where a set of weak learners (e.g. One level decision tree "J48") are combined to obtain a strong learner (e.g. AdaBoostM1).

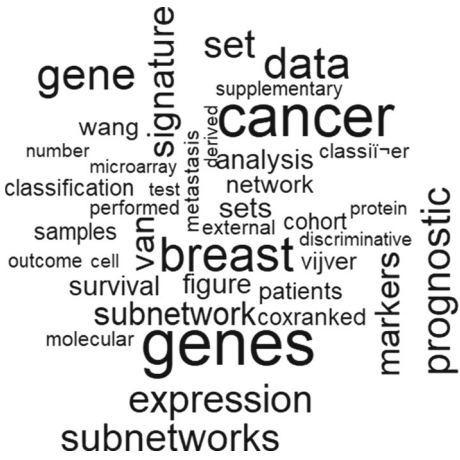

Fig. 6 Words cloud of articles A and B using the statistical software R

Table 3 Comparing performances of keyword extracting tools (main keywords)

Manual annotations	RAKE			Maui Bag			Maui Boost		
	P	R	F	P	R	F	P	R	F
Authors' keywords	6.25	10.0	7.69	12.5	16.67	14.29	12.5	20	15.38
Tags	–	–	–	81.25	8.22	14.92	37.5	3.93	7.12
Authors' keywords + tags	6.25	0.34	0.64	75	6.68	12.26	56.25	5.14	9.42
Authors' keywords + Relevant tags	6.25	2.17	3.22	**81.25**	**35.25**	**49.17**	50	22.98	31.49

We trained Maui by using two ensemble machine learning classifiers to rank candidates: Maui based on the bagging decision trees classifier (Maui Bag); And Maui based on the boosting classifier called AdaBoostM1 using classification trees as single classifiers (Maui Boost).

The highest measures' values of precision P, recall R, and F-measure F are highlighted in bold (see Tables 3, 4). The extraction of 8 main keywords from the two bio-medical testing articles is performed using RAKE, Maui Bag and Maui Boost. The highest measures' values are achieved using the manual annotation of "authors' keywords with relevant tags" with Maui based on the bagging classifier (Maui Bag) (see Table 3). In the cold start, the use of the manual annotation of "authors' keywords" to train the boosting classifier (AbaBoostM1) of (Maui Boost) provides better performances. The accuracy of extracting main keywords is improved by training Maui on manually chosen relevant tags added to authors' keywords, which builds a model that learns the keyword extraction strategy based on bagging decision trees classifier. Whereas, RAKE shows limited accuracy due to the lack of normalization that excludes valid candidates.

The SKOS version of the MeSH terms [58] is used as the lightweight ontology. The highest measures' results are for the fourth category of manual annotation of "authors' keywords with relevant tags" by using Maui Bag (see Table 4). The term assignment

Table 4 Comparing performances of keyword extracting tools (MeSH Terms)

Manual annotations	Maui Bag			Maui Boost		
	P	R	F	P	R	F
Authors' keywords	10	16.67	12.5	10	18.33	12.94
Tags	35	4.9	8.95	25	2.97	5.31
Authors' keywords + tags	40	5.32	9.4	10	1.2	2.14
Authors' keywords + relevant tags	**45**	**26.55**	**33.4**	10	5.75	7.3

model matches each candidate term against the ontology MeSH terms. It extracts 10 MeSH terms for each testing bio-medical article.

The evaluation proves the relevancy of exploring relevant folksonomy tags to aliment the manual annotations. By gathering two types of manually assigned keywords "authors' keywords and relevant tags", we notice a better performance of both: extracting MeSH terms and content-based main keywords. These results demonstrate the effectiveness of our proposal that combines semantic annotation strategies towards pertinently describing a web resource.

Therefore, we consider that each web resource is represented by a vector of a set of attributes (12). The vector's attributes are represented with the couple metadata and its computed score. We delineate the definition of a web resource's description:

$$\begin{aligned} Description\ Web\ Resource &= \{(metadata, score)\} \\ metadata &= \begin{cases} Relevant\ Tags \\ Main\ Keywords \\ Ontology\ Terms \end{cases} \end{aligned} \tag{12}$$

A web resource's metadata are the relevant folksonomy tags, extracted content-based main keywords, and matching ontology terms retrieved from the lightweight ontology.

Description Article A ={(cancer, 0.936); (breast, 1.409); (breast cancer, 0.488); (risk factors, 0.437); (sequence Analysis DNA, 0.199); (signature, 0.0003); (microarray, 0.0003); (human, 0.0003); (breast neoplasms, 0.870); (neoplasms, 0.854); (computational Biology, 0.544); (systems biology, 0.496); (gene expression, 0.309); (classification, 0.293); (network, 0.344); (gene expression profiles, 1.105); (lighting, 0.20)}

Description Article B ={(cancer, 1.004); (breast, 1.342); (breast cancer, 0.565); (signature, 0.804); (microarray, 0.421); (prognosis, 0.351); (human, 0.012); (gene expression, 0.818); (oncogenes, 0.304); (neoplasms, 0.288); (carcinogens, 0.304); (breast neoplasms, 0.860); (survival, 0.345); (hospitals urban, 0.274); (survival analysis, 0.391); (prognostic gene, 0.325); (menopause, 0.287); (classification, 0.003)}

6 Semantic similarity perspectives and alternatives

The emergent descriptive semantic of a web resource is presented as a Vector Space Model. The similarity measurement between web resources is computed based on the

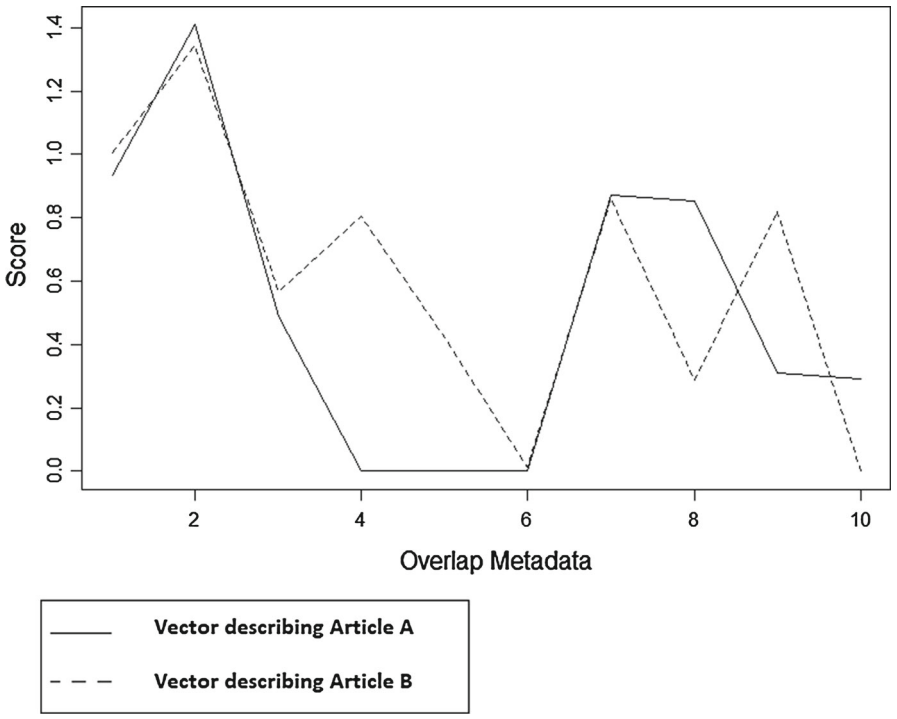

Fig. 7 Vectors describing the two articles A and B based on the relevant tags, main keywords, matching mesh terms

similarity of their descriptive vectors. A relevant clustering of web resources can be calculated with the assumptions of similarities theory by comparing their descriptive vectors.

The semantic similarity between the two vectors describing the two web resources "Article A and Article B" is related to the analysis of the score of their overlap metadata (see Fig. 7). The similarity comparison (see Table 5) of the two vectors describing the corresponding web resources is based on their descriptive metadata using either their content-based main keywords, or extracted Mesh terms deriving from the ontology, or on both of them added to relevant tags. The measure of similarity between the two vectors is computed by applying extensively used similarity measures (see Table 5), namely, Cosine similarity, Euclidean, Manhattan and Jaccard similarity. For distance similarity measures, the more the distance is small, the higher is the degree of similarity between the web resources' descriptive vectors. For cosine similarity, the number of common attributes is divided by the total number of possible attributes. Whereas in Jaccard Similarity, the number of common attributes is divided by the number of attributes that exist in at least one of the two resources' vectors. In practice, it is easier to calculate the cosine of the angle between the vectors, instead of the angle itself. If the cosine value is close to zero, it means that the web resources' vectors are orthogonal and dissimilar.

Table 5 Comparing the similarity of the two vectors

Similarity measures	Main keywords-based	Ontology terms-based	Metadata-based
Cosine similarity	0.992	0.982	0.855
Euclidean distance	0.136	0.3247	1.226
Manhattan distance	0.187	0.388	2.824
Jaccard similarity	0.5	0.23	0.666

The more the similarity distance measures' value is big and the cosine similarity value is small comparing the similarity of the two vectors, the more the descriptive of these two vectors brings up consistent meaning (i.e. avoid mistakenly grouping two distinct web resources into a cluster). Comparing the similarity measures' results of our case study, we notice that the relevant value of similarity measures are obtained based on web resources' vectors described with the three types of metadata "relevant tags, main keywords and extracted MeSH terms". These results demonstrate the effectiveness of considering a combined annotation approach to pertinently describe web resources. This emergent semantic of web resources will properly help in their clustering and organization.

However, the web resources' descriptive metadata hold uncertainty provided by the folksonomy. The imprecision of the emergent semantic describing the web resources has an effect on their clustering. For instance, we cannot absolutely point out certainty that two web resources are strongly related based on their semantic descriptive. Therefore, the semantic similarity can be computed using the fuzzy logic approach that manages the uncertainty. It helps to evaluate the similarity of web resources' descriptive vectors based on degrees of truth rather than considering unambiguously true or false boolean logic. The web resources' comparison perspective will focus on the soft computing techniques, mainly fuzzy based semantic similarity of web resources. The choice of the use of fuzzy logic based similarity measurement relies on the uncertainty vagueness and impreciseness of tags describing web resources. Fuzzy logic assimilates the human way of thinking and judgments. The web resources will not just be objectively similar or not but instead will contain four level of similarities. The construction of the fuzzy rules statements is based on fuzzy inference system described as a collection fuzzy if-then rules that perform logical operations on fuzzy sets (see Table 6). The inputs are the overlap (co-occurrence) of the descriptive metadata and their score. The output is the similarity of the two web resources' vectors. We used the Matlab Fuzzy Logic Toolbox based on the triangular membership function to illustrate those fuzzy rules (see Fig. 8). For instance, if the overlap of metadata and their score are high, then the degree of the similarity between the two vectors is very high (i.e. the percentage of similarity is between 80 and 100%). For this case, the two compared web resources are highly similar to each other.

The emergent semantic similarity of web resources illustrates the intensity of similarity among web resources. Therefore, the organization of web resources is achieved based on their expressed descriptive semantics. It accommodates the Linked Open Data (LOD) initiative that encourages the organization of shared web resources by

Table 6 Fuzzy rules

Overlap	Score	Similarity	%
High	High	Very high	80–100
Medium	High	High	60–80
Medium	Medium	Medium	40–60
Medium	Low	Low	20–40
Low	Low	Very low	0–20

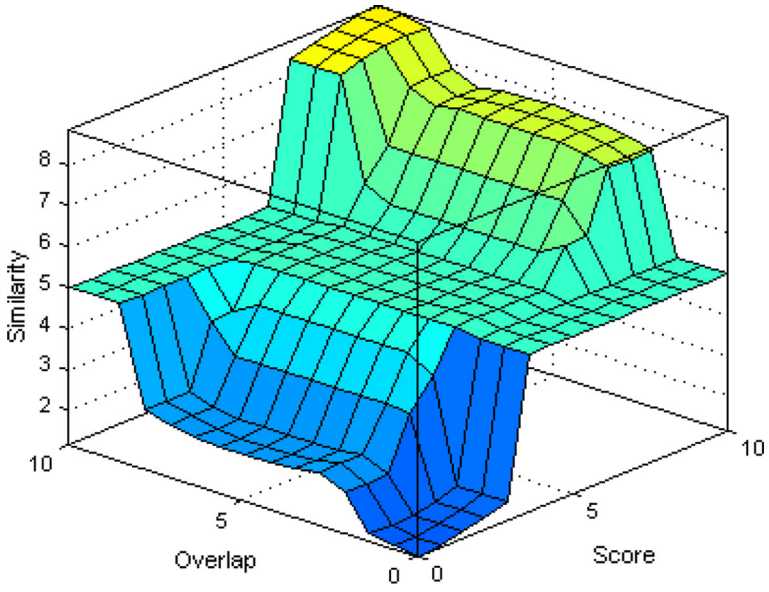

Fig. 8 Fuzzy surface view

expressing their semantics and interlinking. The effectiveness of the recommender systems is investigated by exploiting the Linked Open Data [59]. Our goal aims to explore the semantic relatedness of web resources in order to improve the recommendation process. Indeed, an effective classification and clustering of web resources will enhance the semantic-based context-aware recommender system by suggesting similar items fitting users' preferences.

7 Conclusion and future works

The oncoming of the collaborative social web has raised an extended set of web resources. It has called the attention to the importance of extracting only web resources' relevant descriptive information. Indeed, to achieve an optimal organization of the growing shared web resources, it is essential to pertinently retrieve their relevant semantic descriptors. This paper presents a combined semantic annotation approach to pertinently describe web resources by overcoming folksonomy's weaknesses. Each

web resource is described with its semantic descriptors "metadata" namely, relevant folksonomy tags, content-based main keywords and extracted matching ontology terms. Moreover, the proposal incorporates a recommender system of tags that aims to improve folksonomy's quality by solving the cold start problem of tagging and guiding generation of new tags. The tag recommendations will raise up the users' understanding, promote their contribution and enhances the description of the resources. The experimental evaluation has shown relevant results attesting the effectiveness of our approach. Future perspectives will focus on capturing, describing and exploring the context that arises from the application domains (healthcare, education, tourism). We aim to investigate the potential of using the LOD to increase semantics relatedness of web resources. Our future challenge will focus on the development of a semantic-based context-aware recommender system of web resources to address the needs of a community of users in a specific domain of interest (health community of practices, social learning, open university and the valorization of cultural heritage). The recommendations of relevant resources will feed users' needs, increase their interests and improve their interactions.

References

1. Baker M (2013) Every page is page one. XML Press. Laguna Hills. ISBN 978-1937434281
2. Kang J-H, Lerman K (2011) Leveraging user diversity to harvest knowledge on the social web. In: Proceedings of the IEEE third international conference on social computing (SocialCom)
3. Lau Raymond YK, Leon Zhao J, Wenping Z, Yi C, Ngai Eric WT (2015) Learning contect-sensitive domain ontologies from folksonomies: a cognitively motivated method. Inf J Comput 27:561–578
4. Daglas S, Kakali C, Kakavoulis D, Koumaki M, Papatheodorou C (2012) A methodology for folksonomy evaluation. In: Zaphiris P, Buchanan G, Rasmussen E, Loizides F (eds) Theory and practice of digital libraries. Lecture notes in computer science, vol 7489. Springer, Berlin
5. Kumar KPK, Srivastava A, Geethakumari G (2016) A psychometric analysis of information propagation in online social networks using latent trait theory. Computing 98:583. https://doi.org/10.1007/s00607-015-0472-7
6. Feicheng M, Yating L (2014) Utilising social network analysis to study the characteristics and functions of the co-occurrence network of online tags. Online Inf Rev 38(2):232–247
7. Khan Minhas MF, Abbasi RA, Aljohani NR, Albeshri AA, Mushtaq M (2015) Intweems: a framework for incremental clustering of tweet streams. In: Proceedings of the 17th international conference on information integration and web-based applications and services, iiWAS 15. ACM, New York, NY, USA, pp 87:1–87:4
8. Godoy D, Corbellini A (2016) Folksonomy-based recommender systems: a state-of-the-art review. Int J Intell Syst 31(4):314–346. https://doi.org/10.1002/int.21753
9. Abbas A, Zhang L, Khan SU (2015) A survey on context-aware recommender systems based on computational intelligence techniques. Computing 97(7):667–690
10. Sanchez Bocanegra CL, Sevillano Ramos JL, Rizo C, Civit A, Fernandez-Luque L (2017) HealthRecSys: a semantic content-based recommender system to complement health videos. BMC Med Inform Decis Mak 17:63. https://doi.org/10.1186/s12911-017-0431-7
11. Klašnja-Milićević A, Ivanović M, Vesin B et al (2017) Enhancing e-learning systems with personalized recommendation based on collaborative tagging techniques. Appl Intell. https://doi.org/10.1007/s10489-017-1051-8
12. Bao J, Zheng Y, Wilkie D et al (2015) Recommendations in location-based social networks: a survey. Geoinformatica 19:525. https://doi.org/10.1007/s10707-014-0220-8
13. Qassimi S, Abdelwahed EH, Hafidi M, Lamrani R (2017) Towards an emergent semantic of web resources using collaborative tagging. In: Ouhammou Y, Ivanovic M, Abelló A, Bellatreche L (eds)

Model and data engineering. MEDI 2017. Lecture notes in computer science, vol 10563. Springer, Cham
14. Farnan JM, Snyder SL, Worster BK et al (2013) Online medical professionalism: patient and public relationships: policy statement from the American college of physicians and the federation of state medical boards. Ann Intern Med 158(8):620–627
15. Househ M (2013) The use of social media in healthcare: organizational, clinical, and patient perspectives. Stud Health Technol Inform 183:244–248
16. Ventola CL (2014) Social media and health care professionals: benefits, risks, and best practices. Pharm Ther 39(7):491–499
17. Villegas NM, Sánchez C, Díaz-Cely J, Tamura G (2018) Characterizing context-aware recommender systems: a systematic literature review. Knowl Based Syst 140:173–200. https://doi.org/10.1016/j.knosys.2017.11.003
18. Cao Y, Kovachev D, Klamma R, Jarke M, Lau RW (2015) Tagging diversity in personal learning environments. J Comput Educ 2(1):93–121
19. Klašnja-Milićević A, Vesin B, Ivanović M, Budimac Z, Jain LC (2017) Folksonomy and tag-based recommender systems in e-learning environments. In: E-learning systems. Intelligent systems reference library, vol 112. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-41163-7_7
20. Jean-Louis L, Zouaq A, Gagnon M, Ensan F (2014) An assessment of online semantic annotators for the keyword extraction task. In: Pham DN, Park SB (eds) PRICAI 2014: trends in artificial intelligence. PRICAI 2014. Lecture Notes in Computer Science, vol 8862. Springer, Cham, pp 548–560. https://doi.org/10.1007/978-3-319-13560-1_44
21. Thomas J R, Bharti SK, Babu KS (2016) Automatic keyword extraction for text summarization in e-newspapers. In: Proceedings of the international conference on informatics and analytics, pp 86-93. ACM
22. Turney PD (1999) Learning to extract keyphrases from text. Technical report ERB-1057, National Research Council Canada, Institute for Information technology
23. Witten IH, Paynter GW, Frank E, Gutwin C, Nevill-Manning CG (1999) Kea: practical automatic keyphrase extraction. In Proceedings of the ACM conference on digital libraries, Berkeley, CA, US. ACM Press, New York, NY, pp 254–255
24. Sarkar K (2013) A hybrid approach to extract keyphrases from medical documents. Int J Comput Appl 63(18):14–19. https://doi.org/10.5120/10565-5528
25. Krapivin M, Autayeu M, Marchese M, Blanzieri E, Segata N (2010) Improving machine learning approaches for keyphrases extraction from scientific documents with natural language knowledge. In: Proceedings of the joint JCDL/ICADL international digital libraries conference. Gold Coast, Australia, pp 102–111
26. El-Beltagy SR, Rafea A (2009) Kp-miner: a keyphrase extraction system for English and Arabic documents. Inf Syst 34:132–144
27. Marinho LB, Nanopoulos A, Schmidt-Thieme L, Jäschke R, Hotho A, Stumme G (2011) Social tagging recommender systems. In: Ricci F, Rokach L, Shapira B, Kantor PB (eds) Recommender systems handbook. Springer, Boston, MA, pp 615–644. https://doi.org/10.1007/978-0-387-85820-3_19
28. Špiraneca S, Ivanjkob T (2013) Experts vs. novices tagging behavior: an exploratory analysis. Procedia Soc Behav Sci 73:456–459
29. Consortium GO et al (2017) Expansion of the gene ontology knowledgebase and resources. Nucl Acids Res 45(D1):D331–D338
30. Chen J, Zheng J, Yu H (2016) Finding important terms for patients in their electronic health records: a learning-to-rank approach using expert annotations. JMIR Med Inform 4(4):e40. https://doi.org/10.2196/medinform.6373
31. Hassan MM, Karray F, Kamel MS (2012) Automatic document topic identification using wikipedia hierarchical ontology. In: Proceedings of the eleventh IEEE international conference on information science, signal processing and their applications, pp 237–242
32. Allahyari M, Kochut K (2016) Semantic tagging using topic models exploiting wikipedia category network. In: Proceedings of the 10th international conference on semantic computing
33. Osman T, Thakker D, Schaefer G (2014) Utilising semantic technologies for intelligent indexing and retrieval of digital images. Computing 96(7):651–668
34. Gao G, Liu Y-S, Lin P, Wang M, Gu M, Yong J-H (2017) BIMTag: concept-based automatic semantic annotation of online BIM product resources. Adv Eng Inform 31:48–61

35. Zubiaga A, Fresno V, Martinez R, Garcia-Plaza AP (2013) Harnessing folksonomies to produce a social classification of resources. IEEE Trans Knowl Data Eng 25(8):1801–1813
36. Xie Q, Xiong F, Han T et al (2018) Interactive resource recommendation algorithm based on tag information. World Wide Web. https://doi.org/10.1007/s11280-018-0532-y
37. Qassimi S, Abdelwahed EH, Hafidi M, Lamrani R (2016) Enrichment of ontology by exploiting collaborative tagging systems: a contextual semantic approach. In: Third international conference on systems of collaboration (SysCo). IEEE Conference Publications, pp 1–6
38. Tommasel A, Godoy D (2015) Semantic grounding of social annotations for enhancing resource classification in folksonomies. J Intell Inf Syst 44(3):415–446. https://doi.org/10.1007/s10844-014-0339-y
39. Yu H, Zhou B, Deng M et al (2017) Tag recommendation method in folksonomy based on user tagging status. J Intell Inf Syst. https://doi.org/10.1007/s10844-017-0468-1
40. Belém FM, Martins EF, Almeida JM, Goncalves MA (2014) Personalized and object-centered tag recommendation methods for web 2.0 applications. Inf Process Manag 50(4):524–553
41. Fang Q, Xu Ch, Jitao S, Shamim Hossain M, Ghoneim A (2016) Folksonomy-based visual ontology construction and its applications. IEEE Trans Multimed 18(4):702–713
42. Maui—multi-purpose automatic topic indexing, Homepage. http://www.medelyan.com/software. Accessed 16 Mar 2018
43. Duwairi R, Hedaya M (2016) Automatic keyphrase extraction for arabic news documents based on kea system. J Intell Fuzzy Syst 30(4):2101–2110
44. Lovins JB (1968) Development of a stemming algorithm. Mech Transl Comput Linguist 11(1–2):11–31
45. Jabeen F, Khusro S (2015) Quality-protected folksonomy maintenance approaches: a brief survey. Knowl Eng Rev 30(5):521–544. https://doi.org/10.1017/S0269888915000120
46. Kang J, Lerman K (2011) Leveraging user diversity to harvest knowledge on the social web.In: Privacy, Security, Risk and trust (PASSAT) and 2011 IEEE 3rd international conference on social computing (SocialCom), pp 215–222
47. Papadopoulos S, Vakali A, Kompatsiaris Y (2011) Community detection in collaborative tagging systems. Community-built databases. Springer, Berlin, pp 107–131
48. SKOS simple knowledge organization system. https://www.w3.org/TR/skos-reference/. Accessed 16 Mar 2018
49. Nandipati A (2011) Assessment of metadata associated with geotag pictures. Masters thesis, University of Muenster
50. Zhang L, Tang J, Zhang M (2012) Integrating temporal usage pattern into personalized tag prediction. In: Sheng QZ, Wang G, Jensen CS, Xu G (eds) Web technologies and applications. LNCS 7235. Springer, Berlin, pp 354–365
51. Fu W-T, Kannampallil T, Kang R, He J (2010) Semantic imitation in social tagging. ACM Trans Comput Hum Interact 17(3):1–37
52. citeulike homepage. http://www.citeulike.org/. Accessed 16 Mar 2018
53. US National Library of Medicine National Institutes of Health: Medical Subject Headings (MeSH). https://www.nlm.nih.gov/mesh. Accessed 16 Mar 2018
54. Chuang H-Y et al (2007) Network-based classification of breast cancer metastasis. Mol Syst Biol 3:140. https://doi.org/10.1038/msb4100180
55. Naderi A, Teschendorff AE, Barbosa-Morais NL, Pinder SE, Green AR, Powe DG, Robertson JF, Aparicio S, Ellis IO, Brenton JD, Caldas C (2007) A gene-expression signature to predict survival in breast cancer across independent data sets. Oncogene 26:1507–1516. https://doi.org/10.1038/sj.onc.1209920
56. RAKE Homepage. https://hackage.haskell.org/package/rake. Accessed 16 Mar 2018
57. van Rijsbergen CJ (1979) Information retrieval. Butterworths, London
58. Vrije Universiteit Amsterdam, MeSH terms Homepage. http://libguides.vu.nl/PMroadmap/MeSH. Accessed 16 Mar 2018
59. Musto C, Basile P, Lops P, de Gemmis M, Semeraro G (2017) Introducing linked open data in graph-based recommender systems. Inf Process Manag 53(2):405–435