Exploiting Social Annotations to Generate Resource Descriptions in a Distributed Environment: Cooperative Multi-Agent Simulation on Query-Based Sampling

Zakaria Saoud, Samir Kechid, Mahmoud Saoud, Antoine Doucet

Rev Socionetwork Strat (2017) X:XX-XX

Exploiting Social Annotations to Generate Resource Descriptions in a Distributed Environment: Cooperative Multi-agent Simulation on Query-based Sampling

Zakaria SAOUD[1)], Samir KECHID[1)],
Mahmoud SAOUD[2)] and Antoine DOUCET[3)]

1) Computer Sciences Department, University of Sciences and Technologies Houari Boumediene BP 32 EL ALIA Bab Ezzouar, 16111, Algiers, Algeria
2) Department of Mathematics, Ecole Normale Supérieure BP 92 Kouba; 16050 Alger, Algeria
3) Université de La Rochelle, L3i Laboratory, Avenue Michel Crépeau, 17042 La Rochelle, France

zakaria.saoud@live.fr
skechid@usthb.dz
saoud_m@yahoo.fr
antoine.doucet@univ-lr.fr

Received: Date Month 20XX / Accepted: Date Month 20XX

Abstract. In distributed information retrieval, resource descriptions play a principal role in facilitating the task of other processes such as the resource selection process and the merging process. The previous approach for acquiring resource descriptions was based on different techniques to improve the retrieval process, but they have many limitations. In this paper, we describe a new approach for acquiring precise resource descriptions, based on social annotations available in the social bookmarking service.

Keywords: Resource description, social network, information retrieval.

1 Introduction

Distributed information retrieval systems aim to optimize the search process across multiple distributed sources (resources) or databases. A major problem of this type of search is how to choose the right resource for a given query, to find the relevant documents and to satisfy the user's needs. Hence, many techniques of resource selection have been developed [1] [2] [3] [4]. The resource selection methods can return suitable resources containing the relevant documents for a given query, but only if each individual resource description can precisely represent the contents of each resource [5]. Our goal is to study the possibility of using social annotations available in social bookmarking services in order to obtain accurate resource descriptions. We propose a new approach for acquiring precise resource descriptions. To achieve this goal, we use the set of tags and social information available on the social bookmarking service. The acquiring of resource descriptions in our approach is realized through two steps: (1) The creation of the basic resource descriptions, using the set of tags of their documents, (2) The updating of the resource descriptions in each search session, using an improved version of the QBS algorithm. The rest of the paper is organized as follows: Section 2 provides an overview of the related work. Section 3 presents our method for acquiring resource descriptions. Section 4 describes the experimental testbeds and the evaluation metrics. Section 5 describes the experimental results. Finally, we conclude the paper in Section 6.

2 RELATED WORKS

Resource selection module is responsible for the selection of resources with a high probability of containing relevant documents for a given query. Several resources selection approaches have been proposed in the literature [1] [2] [3] [4]. The previous resource selection approaches perform better when the resource descriptions are well made. Resource description presents the information and the content of each resource in the distributed environment. The used method for the construction of resource descriptors depends on the used algorithm in the resource selection process [6]. Different techniques have been developed to acquire precise resource descriptions. The resource description can be done manually by a specialist [7], who chooses the most representative keywords and terms of the resource description, or automatically by indexing algorithms [5][8].

3 Social-based resource description

Social tagging systems are a kind of social network, that allows users to add, edit, and share bookmarks of web documents. Users can annotate their bookmarks with a set of keywords, called tags. The collection of a user's tags constitutes their personomy, and the collection of all users' personomy constitutes the folksonomy. The folksonomy can be considered as a tripartite (undirected) hypergraph $G = (V, E)$, which involve documents, users, and tag, where $V = U \cup T \cup R$ is the set of nodes and $E = \{\{u, t, r\} | (u, t, r) \in G\}$ is the set of hyperedges which connect documents, users and tags [9][10]. Figure 1 shows the tripartite graph structure of the folksonomy.

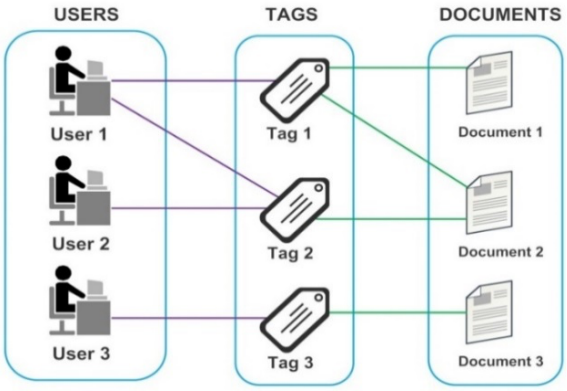

Figure 1. Tripartite graph structure of the folksonomy.

Our method of making the resource descriptions combines the properties of manual construction methods, which require cooperation from resource providers, and the properties of automatic constructing methods, by using the set of tags in a social network. The positive point of using tags and social annotations is the small quantity of these tags compared to set of terms available in the documents and web pages. This helps us to make the construction process and the updating process of resource description, faster, unlike the QBS algorithm, which is expensive in terms of time and computation [11]. The social annotations provide accurate information, which helps us to get accurate resource descriptions. In our approach, we define an improved version of the QBS algorithm. We use it to update the resource descriptions in each search session, unlike the original QBS algorithm which was used directly to construct the resource descriptions. In our updating process, we use the entire running query to retrieve the top k documents returned by the resource, rather than using each query term separately. That allows reducing the number of iterations and the complexity of our new algorithm, compared to original QBS algorithm.

3.1 Resource description construction

The majority of approaches of resource descriptions construction, are based on the term frequency of the documents contained in the resources. The large number of these documents and their large amount of information constitutes a big obstacle in front of distributed information retrieval systems. That is because of the vast time required in the indexation process and the large number of terms, which can affect the accuracy of the resource description. This reflects negatively on the performance of the distributed information retrieval system. Hence we decided to find a suitable alternative to the set of document terms. This alternative should be more precise to describe the documents, to construct a precise resource description, and should be slimmer in quantity to accelerate the updating process. These properties can be provided easily by social annotations and tags. The resource description of a resource S, is made by the set of tags, which used to annotate the resource documents, as follows:

$$\text{description}(s) = \{ (\text{tag}_i, \text{tf}_S(\text{tag}_i)) | i \in [1..N] \}$$

Where:

N: is the number of used tags in the resource s.

$\text{tf}_S(\text{tag}_i)$: is the Resource-based tag frequency of the tag tag_i, which is defined as follows:

$$\text{tf}_S(\text{tag}_i) = \sum_{d_n \in S} \text{tf}_{d_n}(\text{tag}_i) \quad (1)$$

$\text{tf}_{d_n}(\text{tag}_i)$: is the Document-based tag frequency. It represents how many times the document d_n was tagged by the tag tag_i.

Table 1. A fragment of the used tags in the description of the resource $S1$.

Used tags	Documents tagged by this tag	Document-based tag frequency in each document	Resource-based tag frequency
medias	d_1 = *http://www.projectcensored.org/* d_2 = *http://www.lostremote.com/*	$\text{tf}_{d_1}(medias) = 2$ $\text{tf}_{d_2}(medias) = 1$	$\text{tf}_{S1}(medias) = 3$
socialism	d_1 = *http://www.labourstart.org/* d_2 = *http://www.greenleft.org.au/*	$\text{tf}_{d_1}(socialism) = 1$ $\text{tf}_{d_2}(socialism) = 5$	$\text{tf}_{S1}(socialism) = 6$
nytimes	d_1 = *http://www.thomaslfriedman.com/* d_2 = *http://www.feedroom.com/*	$\text{tf}_{d_1}(nytimes) = 3$ $\text{tf}_{d_2}(nytimes) = 2$	$\text{tf}_{S1}(nytimes) = 5$

Example 1. The description of the resource $S1$ which contains a set of documents of the category "News", will be constructed from the set of tags used to annotate these documents. The frequency of each tag in this resource description, represents its Resource-based tag frequency. This is calculated by the sum of its Document_based

tag frequency in each document in the resource $S1$. Table1 shows a fragment of the used tags in the description of the resource $S1$.

3.2 Resource description updating

A part of our updating technique was inspired from the QBS approach developed by Callan et al. [8]. Rather than using the set of terms, we use the set of tags and we combine the documents obtained with the running queries, with the former resource description. We prefer to use the set of tags in this step because the basic resource descriptions in our approach are created by the set of tags. This provides better features than the set of terms as mentioned in the previous section. Many studies have proved that the set of tags can be more useful than the set of terms in many tasks such as: documents classification [12] [13], information retrieval [14], and recommender system [15]. Our resource description will be updated automatically, in each search session. When the user runs a query, we use it to select the k top documents. Then we use the most frequent tags of these documents to update the content of former resource description. The updating process implemented through a new algorithm, inspired from the QBS algorithm. Our new algorithm is summarized as follows:

1- Select a query and run it.
2- Select to k top documents returned by the resource S.
3- Extract the tags and their frequency from the k top returned documents. We note $\mathrm{tf_k(tag_i)}$: the frequency of the tag $\mathrm{tag_i}$ in the k top returned documents.
4- Use the most frequent tags $\mathrm{Top\_tags}$ in the k top returned documents, to update the resource description content.
5- If the queries are not expired: Select other query and go to 1.

The final result of this algorithm presents the updated resource description: $\mathrm{updated(s)}$. To find the k top returned documents in the third step of the algorithm, we rank the documents according to their similarity with the query. This similarity is calculated by the following formula:

$$\mathrm{sim(d_n, q)} = \sum_{\mathrm{t_i} \in \mathrm{q}} \mathrm{tf(t_i)} \quad (2)$$

Where:
$\mathrm{t_i}$: is the term i of the query q.
$\mathrm{tf(t_i)}$: is the frequency of the term $\mathrm{t_i}$ in the document $\mathrm{d_n}$, which is calculated as follows:

$$\mathrm{tf}(t_i) = \begin{cases} \mathrm{tf}_{d_n}(t_i) \text{ if } t_i \in \mathrm{tags}_{d_n} \text{ and } t_i \notin \mathrm{terms}_{d_n} \\ \mathrm{freq}_{d_n}(t_i) \text{ if } t_i \in \mathrm{terms}_{d_n} \text{ and } t_i \notin \mathrm{tags}_{d_n} \\ \mathrm{tf}_{d_n}(t_i) * \mathrm{freq}_{d_n}(t_i) \text{ if } t_i \in \mathrm{tags}_{d_n} \text{ and } t_i \in \mathrm{terms}_{d_n} \end{cases} \quad (3)$$

With:

tags_{d_n}: is the set of tags used to annotate the document d_n.

terms_{d_n}: is the set of terms of the document d_n.

$\mathrm{freq}_{d_n}(t_i)$: is the frequency of the term t_i in the document d_n. (t_i is a document's term)

$\mathrm{tf}_{d_n}(t_i)$: is the document-based tag frequency of the term t_i (t_i is a document's tag).

Equation 3 have been used to calculate the term frequency, in order to give more importance to documents that contain the query terms and that have been tagged with the query terms, at the same time. To update the resource description content in the fourth steps of the algorithm, we use the following proposed formula:

$$\mathrm{updated}(s) = \bigcup_{\mathrm{tag}_i \in \mathrm{description}(s)} (\mathrm{tag}_i, \overline{\mathrm{tf}_S}(\mathrm{tag}_i)) \quad (4)$$

Where:

$\overline{\mathrm{tf}_S}(\mathrm{tag}_i)$: is the new tag frequency value, which is equal to:

$$\overline{\mathrm{tf}_S}(\mathrm{tag}_i) = \begin{cases} \mathrm{tf}_S(\mathrm{tag}_i) + \mathrm{tf}_k(\mathrm{tag}_i) \text{ if } \mathrm{tag}_i \in \mathrm{Top\_tags} \\ \mathrm{tf}_S(\mathrm{tag}_i) \text{ if } \mathrm{tag}_i \notin \mathrm{Top\_tags} \end{cases} \quad (5)$$

With:

$\mathrm{tf}_k(\mathrm{tag}_i)$: is the sum of document-based tag frequency of the tag tag_i in the k returned documents.

$\mathrm{tf}_S(\mathrm{tag}_i)$: is the Resource-based tag frequency of the tag tag_i.

The list of most frequent tags Top_tags is defined as follows:

$$\mathrm{Top\_tags} = \{ (\mathrm{tag}_i, \mathrm{tf}_k(\mathrm{tag}_i)) \mid \mathrm{tf}_k(\mathrm{tag}_i) \geq \mathrm{avgtags}_D \text{ and } i \in [1..L] \}$$

Where:

L: is the number of tags in the k top returned documents.

$\mathrm{avgtags}_D$: represents the average frequency of all tags in the k top returned documents. We calculate this average frequency using the following proposed formula:

$$\mathrm{avgtags}_D = \frac{\sum_{i=1}^{L} \mathrm{tf}_k(\mathrm{tag}_i)}{\mathrm{tags\_number}_D}$$

With $tags\_number_D$: is the number of tags in the k top returned documents.

4 Experimental Testbeds and Evaluation Metrics

To evaluate the proposed approach, we preferred the use of "CABS120k08" dataset which is constructed by Noll and Meinel [12]. For measuring the resource description quality, many measures have been proposed [16][17]. These measures consist of comparing the content of the estimated resource description with the content of the actual resource description (the complete resource description). However, the previous measures were proposed to evaluate the resource descriptions construction approach in uncooperative environment, where each resource returns only ranked lists of documents without any other information about the other documents. For this reason, we are not allowed to evaluate our approach which requires a cooperative environment. In our approach, the resource descriptions are constructed from the set of tags of the wholes documents of these resources. These resources are considered, in this case, as cooperative resources, which can provide other information about their entire documents. Hence, we have decided to evaluate our approach using the resource selection method: CORI [1] in order to test the performance of our approach in the improving of resource selection results. This decision came as part of a previous hypothesis posed by Callan and Connell [5] which states to compare the resource selection results to judge the performance of the resource descriptions construction method. That is because the quality of resource descriptions can affect the resource selection process, negatively or positively.

Table 2. A fragment of the resource descriptions in each algorithm.

Resource description construction algorithm	**Resource descriptions**		
	$S1$	$S2$	$S3$
Algorithm 1	*{hollywood, movies, film production, games, cinema, entertainment….}*	*{**sale**, clothes, job, Smartphones, banks, animals, computer, magazine, foods…}*	*{Horseback, **horses**, business, sport, shopping, blankets, **sale,** equestrian… }*
Algorithm 2	*{ punk , cinema , music, film, tv, videos, music, television, action…}*	*{ fashion , clothes, job, accessories , **sale** , e-commerce, ebay…..}*	*{sport, shopping, training, fitness , cycling , workout , health, , …. }*

Example 2. We have one query and three resources which represent three different categories: (1) The resource $S1$ which is constructed from a set of documents of the category “Arts”, (2) The resource S2 which is constructed from a set of documents of the category “Business”, (3) The resource S3, which is constructed from a set of documents of the category “Sports”. Table 2 shows a fragment of the descriptions of these resources which are created by two different algorithms: algorithm 1 and

algorithm 2. We want to find the relevant resource of the query "horses for sale" using a resource selection method. We suppose that the resource selection method will select the resource which contains the largest number of the query terms.

For the resource descriptions which are created by the algorithm 1, the selected resource for our query "horses for sale", will be the resource $S3$. That is because its description contains more query terms (two query terms) than the other resource descriptions. The resource $S3$ is more relevant to the query, because it contains four relevant documents. This means that the selection process was successful. However, for the resource descriptions which are created by the algorithm 2, the resource selection algorithm will select for the same query, the resource $S2$ rather than the resource $S3$. That is because the description of the resource $S2$ contains more query terms (one query term) than the other resources descriptions (0 query terms). Hence, we can see that the resource selection algorithm performs better when the resource descriptions are well made.

To evaluate the performance of CORI approach in each resource descriptions construction approach, we use the recall metric R_n [18]. This metric provides a comparison between a given method of resource selection (CORI in our case) and a baseline resource selection method. This allows us to measure the percentage of relevant documents contained in the n top-ranked resources by the given method. In our evaluation, we preferred the use of the recall metric $\mathrm{R_n}$ because is more appropriate for our needs than the other metrics. For example, Callan and Connell [5] has used this metric in the evaluation of resource selection algorithm in order to evaluate the performance of the QBS algorithm, which is similar to our case. In our case, we use the R_n metric to test the performance of the resource selection method in order to evaluate the quality of each resource descriptions construction approach. This metric is calculated as follows:

$$\mathrm{R_n} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{E_i}}{\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{B_i}} \quad (6)$$

Where:

E : is the collection selection ranking (the estimated ranking). In our case, this ranking is calculated by CORI algorithm.

B : is the baseline ranking of the resources (the ideal ranking). In our case, the baseline ranking is calculated according to the number of relevant documents in each resource.

$\mathrm{E_i}$: is the ideal score of the resource i in the list E.

$\mathrm{B_i}$: is the ideal score of the resource i in the list B.

n : is the number of top selected resources.

In our experiments, the ideal score of the resource is the number of its relevant documents.

Example 3. We would like to calculate the recall R_4 for the 3 top-ranked resources, for the query "$satellite\ images$". According to table 3, the recall R_4 will be calculated as follows:

$$R_5 = \frac{Ideal(S1) + Ideal(S15) + Ideal(S3) +}{Ideal(S15) + Ideal(S4) + Ideal(S1)} = \frac{2 + 16 + 0}{16 + 5 + 2} = 0.78$$

From the previous results, we can see that the ideal ranking does not include the resource $S3$, which makes its ideal score $Ideal(S3) = 0$. Table 3 shows the ideal scores of the 3 top-ranked resources, with their estimated ideal and ranking.

Table 3. The ideal and the estimated ranking of the resources for the query $\mathbf{q}$.

The ideal ranking		**The estimated ranking (CORI)**	
The resource	***The resource score***	***The resource***	***The resource score***
S15	16	S1	0.86
S4	5	S15	0.79
S1	2	S3	0.64

To evaluate our approach in a distributed environment, we have divided the "CABS120k08" dataset into 16 resources. Each resource of our divided resources represents one of the categories of the Open Directory Project DMOZ. The DMOZ groups the similar websites (web documents) into smaller categories, based on a hierarchical ontology scheme.

To construct the basic resource descriptions in our approach, we use the set of tags of all documents in each resource, as defined in section 3.1. To update these basic resource descriptions, we have used 60 running queries which will also be used to construct the resource descriptions in the QBS approach. For each query term, we download the top five relevant documents. These queries are collected from the "CABS120k08" dataset which contains “2,617,326” search queries. Once completed, each resource description in our approach will be updated from a sample of 300 documents. That is because the previous experiments of Callan and Connell [5], have shown that a sample of 300 documents can provide reasonably accurate resource descriptions at a relatively low cost. For the QBS algorithm, the resource descriptions will be constructed from a sample of 600 documents. This was chosen as the stopping criteria of this algorithm in our experiments. That is because this algorithm will download the top five relevant documents for each query term. Between the 60 used queries, we have 35 queries with 3 terms, 19 queries with 2 terms and 6 queries with 1 term. For the resource selection evaluation, we have chosen the most frequent queries

in all resources. Representing 3629 queries in our dataset, allow us to exploit the largest possible number of resources. In our case, a query is frequent when it was used in more than three resources.

5 EXPERIMENTAL RESULTS

Our experimental method was based on comparing the effectiveness of the resource selection algorithm CORI when using our approach and QBS approach. For each defined number of selected resources, we run a set of queries. Then, we compute the average recall related to our approach, and the average recall related to QBS approach. The average recall represents the average value of recall values of the whole queries.

Table 4 shows the average recall values for each resource descriptions construction approach. From table 4 we can observe that our proposed approach performs better than the QBS approach. The obtained results confirm us that the use of tags in resource description constructions, improves the resource selection process in the distributed information retrieval system. Compared to QBS approach, our resource description construction approach contributes to obtain more precise resource descriptions, through an updating process in each search session. The resource selection approach CORI performs better when the resource descriptions are constructed from social annotations. The accurate information provided by social annotation, help the resource selection algorithm to choose the most relevant resources, which contain more relevant documents for a given query. The resource description in the QBS approach is constructed from the set of terms of the returned documents. This set of terms can be inadequate and inaccurate for describing the resource content, which impedes the resource selection algorithm to select the relevant resources for a given query.

The experimental results also demonstrate that in terms of average recall values, the difference between the two approaches decreases when the resource selection algorithm selects more resources. From table 4, we can see that the difference between the two approaches in terms of average recall values is equal to 0.119, when the resource selection algorithm selects four resources. This difference is equal to 0.069 when the resource selection algorithm selects seven resources.

Table 4. Average recall values in each resource descriptions construction approach.

Resource descriptions construction approach	Number of selected resources				
	$n = 4$	$n = 5$	$n = 6$	$n = 7$	$n \geq 8$
QBS	0.352	0.420	0.476	0.554	0.661
Our approach	0.471	0.533	0.559	0.623	0.745

6 Conclusion and future work

This paper describes a new approach for acquiring accurate resource descriptions. Our approach combines the properties of manual construction methods and the properties of automatic construction methods of resource descriptions. We tried to exploit the cognitive value of social annotations combined with the content of resource documents. This allowed us to construct a rich resource description. We also defined a new updating technique, which consists of modifying the tags' frequency in the resource descriptions, in order to give more importance to the relevant tags in each search task. The obtained results show that our proposed approach performs better than the QBS approach and can improve the performance of the resource selection process in the distributed information retrieval system. In our future work, we aim to evaluate the performance of our proposed approach with other resource selection algorithms and test its' effectiveness in the result merging process.

Acknowledgements

This paper is a revised and expanded version of the paper entitled "Acquiring resource descriptions using social annotations", presented at the Fifth ASE International Conference on Big Data, 2015, Taiwan [19].

References

1. Callan J., Lu Z., Croft B. Searching distributed collection with inference networks. In the Eighteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Seattle, WA, ACM-SIGIR'95, p. 21-28, 1995

2. Balakrishnan, R., & Kambhampati, S.: SourceRank: relevance and trust assessment for deep web sources based on inter-source agreement. In Proceedings of the 20th international conference on World wide web (pp. 227-236). ACM, 2011
3. Jaime Arguello, Jamie Callan, and Fernando Diaz.: Classification-based resource selection. In Proceedings of CIKM. 1277–1286, 2009
4. Puppin, D., Silvestri, F., Perego, R., & Baeza-Yates, R.: Tuning the capacity of search engines: Load-driven routing and incremental caching to reduce and balance the load. ACM Transactions on Information Systems (TOIS), 28(2), 5, 2010
5. Callan, J., & Connell, M.: Query-based sampling of text databases.ACM Transactions on Information Systems (TOIS), 19(2), 97-130, 2001
6. Arguello, J.: Federated Search for Heterogeneous Environments(Doctoral dissertation, Yahoo! Research, 2011
7. Chakravarthy, A. S., & Haase, K. B.: NetSerf: using semantic knowledge to find Internet information archives. In Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 4-11). ACM, 1995
8. Callan, J., Connell, M., & Du, A.: Automatic discovery of language models for text databases. In ACM SIGMOD Record (Vol. 28, No. 2, pp. 479-490). ACM, 1999
9. Hotho, A., Jäschke, R., Schmitz, C., & Stumme, G.: Information retrieval in folksonomies: Search and ranking (pp. 411-426). Springer Berlin Heidelberg, 2006
10. Mika, P.: Ontologies are us: A unified model of social networks and semantics. In The Semantic Web–ISWC 2005 (pp. 522-536). Springer Berlin Heidelberg, 2005
11. Baillie, M., Carman, M. J., & Crestani, F.: A topic-based measure of resource description quality for distributed information retrieval. In Advances in Information Retrieval (pp. 485-496). Springer Berlin Heidelberg, 2009
12. Noll, M. G., & Meinel, C.: The metadata triumvirate: Social annotations, anchor texts and search queries. In Web Intelligence and Intelligent Agent Technology, 2008. WI-IAT'08. IEEE/WIC/ACM International Conference on (Vol. 1, pp. 640-647). IEEE, 2008
13. Aliakbary, S., Abolhassani, H., Rahmani, H., & Nobakht, B.: Web page classification using social tags. In Computational Science and Engineering, 2009. CSE'09. International Conference on (Vol. 4, pp. 588-593). IEEE, 2009
14. Heymann, P., Koutrika, G., & Garcia-Molina, H.: Can social bookmarking improve web search?. In Proceedings of the 2008 International Conference on Web Search and Data Mining (pp. 195-206). ACM, 2008
15. Vatturi, P. K., Geyer, W., Dugan, C., Muller, M., & Brownholtz, B. (2008, October). Tag-based filtering for personalized bookmark recommendations. InProceedings of the 17th ACM conference on Information and knowledge management (pp. 1395-1396). ACM, 2008
16. Baillie, M., Azzopardi, L., & Crestani, F.: Towards better measures: Evaluation of estimated resource description quality for distributed IR. In Proceedings of the 1st international conference on Scalable information systems (p. 41). ACM, 2006
17. Ipeirotis, P. G., & Gravano, L.: When one sample is not enough: improving text database selection using shrinkage. In Proceedings of the 2004 ACM SIGMOD international conference on Management of data (pp. 767-778). ACM, 2004
18. Powell, A. L., & French, J. C.: Comparing the performance of collection selection algorithms. ACM Transactions on Information Systems (TOIS), 21(4), 412-456, 2003
19. Saoud, Z., & Kechid, S.: Acquiring resource descriptions using social annotations. In Proceedings of the ASE BigData & SocialInformatics 2015 (p. 23). ACM, 2015