

Automatic Hierarchical Categorization of Research Expertise Using Minimum Information

Gustavo Oliveira de Siqueira, Sérgio Canuto, Marcos André Gonçalves,
and Alberto H.F. Laender^(✉)

Department of Computer Science,
Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil
`{gosiqueira,sergiodaniel,mgoncalv,laender}@dcc.ufmg.br`

Abstract. Throughout the history of science, different knowledge areas have collaborated to overcome major research challenges. The task of associating a researcher with such areas makes a series of tasks feasible such as the organization of digital repositories, expertise recommendation and the formation of research groups for complex problems. In this paper we propose a simple yet effective automatic classification model that is capable of categorizing research expertise according to a hierarchical knowledge area classification scheme. Our proposal relies on discriminative evidence provided by the title of academic works, which is the minimum information capable of relating a researcher to its knowledge area. We also evaluate the use of learning-to-rank as an effective mean to rank experts with minimum information. Our experiments show that using supervised machine learning methods trained with manually labeled information, it is possible to produce effective classification and ranking models.

Keywords: Research expertise categorization · Classification schemes · Supervised classification · Learning-to-rank

1 Introduction

Throughout the evolution of science, scientific problems have become more and more complex over time. Their solution currently requires the combination of multiple expertises for the formation of multidisciplinary research groups to work on those complex problems. One basic premise for this to work is that one may be able to identify the main areas of expertise of scholars/researchers. In fact, the effective and reliable association of a scholar with a knowledge area makes a series of tasks feasible such as: (i) organization of digital repositories according to a knowledge area categorization scheme; (ii) expertise recommendation for specific industrial or scientific problems; and (iii) the formation of research groups for solving very complex problems.

There are currently several sources of information that can be used to identify a researcher's expertise, such as: (i) digital libraries containing information about a researcher's scientific production over time (e.g., DBLP¹ and ACM Digital Library²); (ii) metadata and, in several cases, the full text of an electronic thesis or dissertation (ETD) available in ETD repositories (e.g., NDLTD³); and (iii) curricula vitae of researchers made freely available on the Web or in specific official repositories (e.g., the Brazilian Lattes Platform⁴). However, in most of these sources, the researcher's areas of expertise are not explicitly identified and can only be implicitly inferred from the available content in these repositories. This requires some type of text mining treatment such as unsupervised topic extraction [1, 6, 17, 22], automated supervised classification [18, 21, 23] and learning-to-rank methods [14, 15].

In this paper, we focus on supervised techniques as they have historically produced better results, with the drawback of requiring labeled data. More specifically, we exploit a hierarchical classification⁵ scheme to establish an automatic categorization model to solve the presented problem as discussed by Ribeiro-Neto et al. [18] and Waltinger et al. [23]. We exploit hierarchical classification in order to classify experts in a finer granularity level. However, hierarchical categorization is still a hard research problem faced by the text mining community. For completeness, we also evaluate the problem of ranking experts according to the knowledge areas.

Particularly, we use the knowledge area hierarchical classification scheme proposed by the Brazilian National Council for Scientific and Technological Development (CNPq), which provides a simple mechanism to systematize and characterize information about researchers and research groups. This classification scheme is organized into the following four levels⁶: *major area* (e.g., Earth and Exact Sciences), *area* (e.g., Computer Science), *subarea* (e.g., Theory of Computation) and *specialty* (e.g., Formal Languages and Automata). The third and fourth levels of this classification scheme are not used in this paper due to the fact that a researcher might be associated with more than one subarea or specialty, which would characterize a multi-category classification problem [20]. Table 1 shows an excerpt of the two first levels of the CNPq knowledge area classification scheme, which covers nine major areas including altogether 76 specific areas.

Another important source of information used in this paper is the Lattes Platform. Maintained by CNPq, this platform is an internationally renowned initiative in Brazil [11] that provides a repository of researchers' curricula and research groups, all integrated into a single system. The available curricula present a great amount of information about the researchers that can be used for many purposes. In this paper, we focus on exploiting only the title of a researcher's PhD

¹ <http://dblp.uni-trier.de>.

² <http://dl.acm.org>.

³ <http://www.ndltd.org>.

⁴ <http://lattes.cnpq.br>.

⁵ In this paper we use the terms categorization and classification interchangeably.

⁶ <http://bit.ly/1JM2j1k>.

Table 1. Excerpt of the CNPq knowledge area classification scheme

1.00.00.00-3 Exact and Earth Sciences	5.00.00.00-4 Agrarian Sciences
1.01.00.00-8 Mathematics	5.01.00.00-9 Agronomy
1.02.00.00-2 Probability and Statistics	5.02.00.00-3 Forest Engineering
...	...
1.08.00.00-0 Oceanography	5.07.00.00-6 Food Science and Technology
2.00.00.00-6 Biological Sciences	6.00.00.00-7 Applied and Social Sciences
2.01.00.00-0 General Biology	6.01.00.00-1 Right
2.02.00.00-5 Genetics	6.02.00.00-6 Administration
...	...
2.13.00.00-3 Parasitology	6.13.00.00-4 Tourism
3.00.00.00-9 Engineering	7.00.00.00-0 Humanities
3.01.00.00-3 Civil Engineering	7.01.00.00-4 Philosophy
3.02.00.00-8 Mining Engineering	7.02.00.00-9 Sociology
...	...
3.13.00.00-6 Biomedical Engineering	7.10.00.00-3 Theology
4.00.00.00-1 Health Sciences	8.00.00.00-2 Linguistics, Letters and Arts
4.01.00.00-6 Medicine	8.01.00.00-7 Linguistics
4.02.00.00-0 Dentistry	8.02.00.00-1 Letters
...	8.03.00.00-6 Arts
4.09.00.00-2 Physical Education	9.00.00.00-5 Others

dissertation or Master's thesis found in her Lattes curriculum, since, in the extreme case, it is the only available (and reliable) information about the researcher when considering, for example, metadata from institutions or curricula vitae. The title of a dissertation or thesis is also a specially important source of information when considering new researchers, since sometimes there is little or no other available evidence about their research interest.

Our goal is to test the limits of some of the current state-of-the-art classifiers and learning-to-rank methods to generate categorization and ranking models to categorize researchers according to a knowledge area hierarchical classification scheme using only the title of their academic works. This is not a trivial task, given the difficulty in training machine learning models to obtain satisfactory results using just a small piece of text and, consequently, a reduced set of features. As any given additional information available by the researcher would probably only improve the results, our investigation would provide a lower bound on the results that can be obtained in this difficult scenario.

To summarize, our goal here is to investigate the benefits of applying supervised machine learning techniques to the tasks of categorizing and ranking research expertise using a knowledge area single-label hierarchical classification scheme. Thus, our main contributions in this paper are:

- An investigation on the limits of solving a combination of two hard problems: hierarchical categorization and categorization of very short texts;
- A comparative analysis of three supervised classification techniques applied to solve the aforementioned combined problem;
- An evaluation of the state-of-the-art ranking technique with recently proposed similarity features to solve the task of ranking experts using very short texts.

Our experimental results show an accuracy of up to 75% and 83% when categorizing researchers according to, respectively, the first and second levels of the CNPq knowledge area hierarchical classification scheme using only the title of their academic works and considering a model trained with Support Vector Machines (SVM). In addition, this classifier is more effective for this particular task than those based on Naive Bayes and Random Forests models. Moreover, the precision in the top positions of our ranking models achieve up to 97% and 88% considering the first and second levels of the hierarchy, respectively. These results provide evidence towards the potential benefits of using state-of-the-art feature representations and learning-to-rank techniques on the hard problem of expert search with minimum information.

The remainder of this paper is organized as follows. Section 2 addresses related work. Section 3 describes the dataset used in our experiments. Section 4 presents the methodology applied to the generation of the machine learning models and describes the results of the experiments performed to evaluate them. Finally, Sect. 5 presents our final considerations and provides directions for future work.

2 Related Work

The closest related tasks associated with automatic categorization and research expertise ranking in the literature are automatic expert profile construction [12, 13, 17, 24], automatic categorization of text documents in digital libraries [1, 3, 19, 20, 23] and expert discovery [14, 15].

Most of the previous efforts related to our work address the problem of automatically categorizing academic publications from digital libraries. The most effective techniques exploit the supervised learning paradigm to classify documents according to a set of previously defined knowledge areas, usually structured as a specific taxonomy [20, 23]. Based on a set of training documents, these strategies are capable of achieving effective results using Support Vector Machines (SVM) to address the high sparsity and dimensionality of textual data derived from academic documents. In order to minimize the manual effort to label training documents, some previous works exploit unsupervised and semi-supervised techniques. They use topic models to categorize documents according to automatically generated taxonomies [3], provide alternative topic representations [1] or rely on linguistic patterns for taxonomy learning [19]. Despite related to our work because the categorization process is based on a specific taxonomy, here we focus on exploiting the minimum necessary discriminative information to categorize research expertise instead of classifying individual documents.

The problem of categorizing expertise is also associated with the task of automatic expert profile construction, which uses associations between an expert and her registered documents to model the expertise [12, 13, 17, 24]. Specifically, after collecting all documents related to an expert, some methods [12, 24] classify them using a supervised machine learning approach trained with manually labeled documents from other experts. Alternatively, MacDonald and Ounis [13] model an expert as a set of documents, computing the similarity between her documents and those from a knowledge area. Although automatic methods minimize the manual labor of updating the expert profile, its application in organizational contexts is limited because of the lack of textual documents related to an expert [12].

In addition to classification, the machine learning task of ranking experts has also been recently addressed in the literature [14, 15]. Existing approaches rely on information taken from academic works, their citations and the profile information of experts. In this scenario, the use of learning-to-rank techniques presents an effective strategy to combine these different kinds of information [14]. Moreover, the use of such techniques have also been successfully employed to manipulate location-sensitive information [15].

Unlike previous work, we focus on hierarchically categorizing and ranking research expertise using minimum information. Considering the categorization task, both hierarchical categorization and categorization using only short texts are by themselves hard problems [5] and their combination makes this joint problem even harder. The ranking of experts using limited information is also challenging. In order to alleviate this problem, we apply a recently proposed approach to transform sparse textual features [26]. This approach generates a low-dimensional and informative feature space that is more suitable for the task of ranking academic experts using short texts. We evaluate both tasks, classification and ranking with minimum information, using the multi-area dataset described in the next section.

3 Our Dataset

To train a general model to categorize research expertise according to the CNPq knowledge area hierarchical classification scheme, we used the titles of labeled PhD theses and Master's dissertations found on curricula stored on the Lattes Platform. For this, we collected the curricula (XML versions) of 221,119 researchers holding a PhD degree. The respective excerpts of the collected XML documents including data from a thesis or dissertation were parsed and stored into a CSV file with each row containing the following columns: *title*, *major area* and *area*.

For the sake of completeness, we have removed from our dataset the titles of all theses and dissertations without a major area or area associated to them, thus resulting in a dataset that included the title of 49,508 PhD theses and 150,690 Master's dissertations. We have also cleaned the dataset to remove specific errors, such as incompatible major areas and areas associated to a same

Table 2. Distribution of the number of titles per major knowledge area

Major area	Nr. of Areas	Nr. of Titles
Agrarian Sciences	7	22,773
Biological Sciences	13	33,330
Health Sciences	9	25,983
Exact and Earth Sciences	8	33,085
Humanities	10	33,992
Applied and Social Sciences	13	17,398
Engineering	13	22,766
Linguistics, Letters and Arts	3	12,283

thesis or dissertation. For the purpose of this paper, we have also disregarded the major area *Others* due to its lack of a well delimited area grouping in our dataset. Thus, our final dataset comprises data from a total of 199,610 distinct theses and dissertations. We represent our final dataset using the traditional bag-of-words model [9] with the TF-IDF weighting scheme. Table 2 shows the distribution of the curricula vitae in our dataset according to the eight major areas considered for categorizing the researchers in terms of their main research interests, as well as the number of specific areas within each major area.

4 Experiments

In order to evaluate our proposal, we used distinct supervised machine learning methods to generate specific categorizing models and set up a set of experiments to compare them by means of quality metrics aimed to assess their effectiveness when using minimum information, thus allowing us to identify the most promising learning paradigm in our context.

4.1 Model Generation

As aforementioned, our goal is to investigate the benefits of applying supervised machine learning techniques to the task of categorizing and ranking research expertise using a given knowledge area classification scheme. The basic idea is to use such algorithms to “learn” a *good classification* or *ranking function* based on a set of textual features (bag-of-words). We here evaluate three classification techniques that follow completely different learning paradigms, namely Naive Bayes, Support Vector Machine (SVM) and Random Forests (RF) [9]. We also evaluate the task of ranking research expertise with minimum information. For this, we use the state-of-the-art ranking strategy BROOF-L2R [7] and transform textual features into meta-level features designed to improve the ranking results [26].

Hierarchical Classification Models. Considering the evaluated classification models, the Naive Bayes is the most simple and scalable approach, which applies the Bayes’ theorem to estimate the category of texts from probability estimates of individual words with the “naive” assumption of independence between every pair of words. RF and SVM are two of the most successful classification methods, being considered by many [8,9] as the top-notch supervised algorithms. The RF approach is based on an ensemble of decision trees, which not only makes the strategy highly parallelizable, but also grants effective non-linear capabilities. Unlike RF, SVM is an inherently binary linear classification approach. Particularly, SVM uses a maximum-margin optimization method that tries to find a hyper-plane that best separates training examples (placed in a hyperspace) belonging to two different categories. The limitations of only discriminating between two linearly separable categories can be surpassed by using non-linear kernels to transform the feature space and building one classifier per category, where each category is fitted against all the other ones (one-vs-all).

Our approach for the hierarchical categorization of researchers involves not only training a classifier to discriminate such researchers among the major areas, but also eight more specific classification models to categorize them within the subareas. In other words, we first apply the general model to identify a researcher’s major area (e.g., Exact and Earth Sciences) and, once this is determined, we apply a specific model trained to identify her specific area (e.g., Computer Science)

Learning-to-Rank Approach. In addition to classification, we also exploit the effectiveness of ranking the research expertise in different areas, which can be seen as “queries” in our problem. More specifically, we use the learning-to-rank framework to learn a ranking function from relevant and not relevant items from each area, and then use this function to rank items of unknown relevance.

However, different from the classification task, effective learning-to-rank approaches usually rely on a low-dimensional meta-feature space containing primarily similarity features that explicitly measure the proximity between queries (in our case the knowledge areas) and items [16]. Moreover, the fine-grained features in a high-dimensional feature space of words (the bag-of-words representation) usually used in classification may not be sufficiently expressive for effective learning-to-rank [26].

In order to overcome the challenges related to learning-to-rank using the Lattes categorized data, we propose an effective approach to rank research expertise that first transforms the bag-of-words representation of instances and categories into the recently proposed low-dimensional meta-feature space of similarity features designed for learning-to-rank [26]. In our scenario, these features provide the similarity relationship between an item and a research expertise. Particularly, the research expertise is represented using the centroid of its relevant items, as well as the closest relevant items with respect to a specific item. After this pre-processing transformation step, we use the generated compact

meta-feature space as input to the state of the art learning-to-rank approach BROOF-L2R [7].

4.2 Evaluation, Algorithms and Procedures

The classification models were compared using two standard text categorization metrics: micro averaged F₁ (MicroF₁) and macro averaged F₁ (MacroF₁) [25]. While MicroF₁ measures the classification effectiveness over all decisions (i.e., the pooled contingency tables of all classes), MacroF₁ measures the classification effectiveness for each individual class and averages them.

The ranking results were measured with two widely used ranking evaluation metrics: Precision at position k (P@k) [2] and Normalized Discounted Cumulative Gain (NDCG) [10]. Both measures evaluate the effectiveness of the top-ranked results, which are the most relevant to a human searching for an expert. All experiments were executed using a 5-fold cross-validation procedure (which selects 4/5 of the dataset as training data and the remaining as testing data). The parameters were set via cross-validation on the training set and the effectiveness of distinct algorithms were measured in the test partition. In order to evaluate the classification effectiveness, we used the *scikit-learn*⁷ implementations of RF and Multinomial Naive Bayes, and the LIBLINEAR⁸ implementation of SVM. To evaluate the ranking effectiveness, we used the BROOF-L2R implementation provided by their authors [7]. The free parameters of these classifiers include the cost C for SVM and the number of features N considered in the split of a node on the RF-based approaches. These free parameters were set using a 5-fold cross-validation within the training set. The regularization parameter C of SVM was chosen among 11 values from 2^{-5} to 2^{15} and the parameter N of RF was selected among 10%, 20% and 30% of the number of features. For RF, each tree was grown without pruning, as suggested by Breiman [4]. Considering that the results obtained with 200, 300 and 500 trees were statistically tied (with 95% confidence), we adopted 200 trees due to its lower cost. In all ranking experiments, we adopted the number of iterations of the BROOF-L2R algorithm as 100 iterations, which is also the parameter adopted by its authors [7].

4.3 Experimental Results

Classification Results. Table 3 reports the MicroF₁ (MicF₁) and MacroF₁ (MacF₁) values for the classification of the theses and dissertations in our dataset using the three aforementioned methods. We evaluated our model considering the two upper levels of the CNPq knowledge area classification scheme, *major area* and *area*. In addition, we grouped our results according to the scheme described in Table 1. We would like to emphasize the following aspects of our results.

⁷ <http://scikit-learn.org>.

⁸ <https://www.csie.ntu.edu.tw/~cjlin/liblinear>.

Table 3. Average MacroF₁ and MicroF₁ of the three classification models on each major area

		SVM	NB	RF
Major areas	MicF ₁	75.53 ± 0.25	72.45 ± 0.08	71.04 ± 0.09
	MacF ₁	74.66 ± 0.31	71.23 ± 0.13	69.15 ± 0.16
Agrarian Sciences	MicF ₁	82.84 ± 0.18	80.96 ± 0.16	75.27 ± 0.18
	MacF ₁	73.07 ± 0.29	70.74 ± 0.27	58.59 ± 0.26
Biological Sciences	MicF ₁	62.12 ± 0.63	59.11 ± 0.32	55.68 ± 0.28
	MacF ₁	53.00 ± 0.47	49.39 ± 0.30	44.82 ± 0.28
Health Sciences	MicF ₁	76.47 ± 0.34	71.23 ± 0.20	67.38 ± 0.17
	MacF ₁	65.17 ± 2.02	58.40 ± 0.20	49.19 ± 0.17
Exact and Earth Sciences	MicF ₁	83.47 ± 0.13	81.99 ± 0.12	78.69 ± 0.13
	MacF ₁	76.67 ± 0.13	75.22 ± 0.20	67.98 ± 0.21
Humanities	MicF ₁	68.93 ± 0.29	61.55 ± 0.25	59.74 ± 0.28
	MacF ₁	57.93 ± 0.31	50.35 ± 0.33	45.60 ± 0.31
Applied Social Sciences	MicF ₁	74.88 ± 0.20	68.08 ± 0.20	66.38 ± 0.21
	MacF ₁	51.30 ± 0.95	37.05 ± 0.18	35.52 ± 0.17
Engineering	MicF ₁	67.22 ± 0.11	65.29 ± 0.24	62.41 ± 0.27
	MacF ₁	53.49 ± 0.45	45.56 ± 0.25	41.63 ± 0.25
Linguistics, Letters and Arts	MicF ₁	78.47 ± 0.14	75.14 ± 0.06	73.52 ± 0.13
	MacF ₁	79.70 ± 0.13	74.36 ± 0.08	72.26 ± 0.18

First, the SVM model significantly outperforms all other evaluated models. The primary reason for the effective SVM results is its remarkable capability of learning in high dimensional feature spaces. This is due to the fact that the SVM classifier measures the complexity of hypotheses based on the margin with which it separates data, not on the number of features. The SVM method is also insensitive to the high sparsity of textual data, since it just “adds” the evidence of each word present in a document to classify it. NB also shares the same “additive” nature of SVM, having achieved the second best set of results in our experiments. The method that presented the worst results was RF, which uses complex non-linear patterns extracted by association rules that relate the words of a document to its category. We argue that, due to its complexity, RF generates models that may not generalize well in the case of highly sparse domains as it is the case of short texts.

Second, most of the generated models provide evidence towards the initial hypothesis that it is possible to categorize researchers’ expertise by exploiting the information from the titles of their theses or dissertations. Particularly, the models using only this minimum information achieve up to 83% and 79% on MicroF₁ and MacroF₁, respectively. Moreover, the effectiveness of the results using SVM are superior to 70% in the major areas and most of the areas.

Finally, despite the excellent overall performance, we attribute the low performance (around 51%) in some major areas (e.g., Applied Social Sciences and Engineering) to two different factors. First, the distribution of labeled examples among areas are very imbalanced. Particularly, for some specific areas from Engineering, such as Mining Engineering and Biomedical Engineering, our dataset has less than 10 labeled examples, which makes it difficult to learn effective models for these areas. Second, the high vocabulary overlap among these areas and their fuzzy delimitations can undermine the classification effectiveness.

Ranking Results. We now turn our attention to the classification-related task of ranking according to expertise. Table 4 reports the precision and NDCG values for the top 10 best ranked results using the BROOF-L2R method. Particularly, we evaluate the ranking of theses and dissertations from our dataset according to the expertise of their authors in each area.

Like in the classification task, the overall results show the effectiveness of our ranking strategy. This provides evidence for the benefits of using learning techniques to generate ranking of experts by exploiting only the information from the titles of their theses or dissertations. Particularly, the ranking results considering as queries only the major areas (i.e., the first level of the hierarchy described in Table 1) presented the most effective results for both NDCG@10 and P@10. These results were already expected, since each major area contains thousands of positive training and testing examples (see Table 2). This led to the learning of very effective ranking functions using BROOF-L2R, as well as to plenty of possible positive test examples that can assume the top positions in the ranking.

Coincidentally, our best results occurred on major areas with many positive examples for its areas. Specifically, two major areas (Linguistics, Letters and Arts, and Exact and Earth Sciences) achieved the best results among all major areas due to the fact that both of them have many examples for their specific areas, which led to good ranking functions for each one of them. Therefore, in

Table 4. Average NDCG@10 and P@10 of BROOF-L2R on each major area

	P@10	NDCG@10
Major Areas	0.97 ± 0.03	0.97 ± 0.03
Agrarian Sciences	0.71 ± 0.17	0.72 ± 0.15
Biological Sciences	0.68 ± 0.12	0.68 ± 0.12
Health Sciences	0.73 ± 0.21	0.77 ± 0.20
Exact and Earth Sciences	0.88 ± 0.20	0.89 ± 0.19
Humanities	0.73 ± 0.18	0.75 ± 0.18
Applied Social Sciences	0.53 ± 0.16	0.54 ± 0.15
Engineering	0.59 ± 0.15	0.60 ± 0.14
Linguistics, Letters and Arts	0.77 ± 0.27	0.77 ± 0.25

these cases we obtained a high average ranking effectiveness in all areas (which can be seen as queries) of a major area. Likewise, our worst results occurred on major areas that had only few positive examples (less than ten) in some of their specific areas, which led to poor ranking functions for each one of them. In fact, the worst performing major areas (Engineering and Applied Social Sciences) include specific areas with only three positive examples. Despite these specific cases, the task of ranking specialists using minimum information achieves effective results where there is enough training examples. This provide evidence to our claim that it is possible to effectively rank experts using only short texts.

5 Conclusions and Future Work

To conclude, in this paper we addressed two distinct problems: (i) determining a researcher's expertise area by automatically categorizing the title of her PhD dissertation or Master's thesis according to a hierarchical scheme using an automatic classification model, and (ii) ranking experts with respect to knowledge areas using as well the same piece of information.

The results obtained using supervised classification methods were in general very good, specially given the restriction of using minimum information. We also performed a comparative analysis of three state-of-the-art supervised classification methods to determine the best one for the proposed task, being SVM the one that significantly outperformed the other two. As for the classification task, the state-of-the-art learning-to-rank method using recently proposed ranking features produced excellent results in general, even considering the same minimum information restriction.

As future work, we intend to: (i) expand the study to other datasets using the models learned with the CNPq knowledge area hierarchical classification scheme to categorize researchers not present in the Lattes Platform (Transfer Learning); (ii) explore deeper levels of the CNPq hierarchy and other hierarchical categorization strategies (e.g., fuzzy); and (iii) propose an expert recommendation system based on our results.

Acknowledgements. This work was partially funded by projects InWeb (grant MCT/CNPq 573871/2008-6) and MASWeb (grant FAPEMIG/PRONEX APQ-01400-14), and by the authors' individual grants from CAPES, CNPq and FAPEMIG.

References

1. Aletras, N., Baldwin, T., Lau, J.H., Stevenson, M.: Representing topics labels for exploring digital libraries. In: Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries, pp. 239–248 (2014)
2. Baeza-Yates, R.A., Ribeiro-Neto, B.: Modern Information Retrieval. Addison-Wesley Longman Publishing Co. Inc., Boston (1999)
3. Bakalov, A., McCallum, A., Wallach, H., Mimno, D.: Topic models for taxonomies. In: Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries, pp. 237–240 (2012)

4. Breiman, L.: Random forests. *Mach. Learn.* **45**(1), 5–32 (2001)
5. Chen, M., Jin, X., Shen, D.: Short text classification improved by learning multi-granularity topics. In: Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, vol. 3, pp. 1776–1781 (2011)
6. Chen, Y., Fox, E.A.: Using ACM DL paper metadata as an auxiliary source for building educational collections. In: Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries, pp. 137–140 (2014)
7. de Sá, C.C., Gonçalves, M.A., Sousa, D.X., Salles, T.: Generalized BROOF-L2R: a general framework for learning to rank based on boosting and random forests. In: Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 95–104 (2016)
8. Fernández-Delgado, M., Cernadas, E., Barro, S., Amorim, D.: Do we need hundreds of classifiers to solve real world classification problems? *J. Mach. Learn. Res.* **15**(1), 3133–3181 (2014)
9. Hastie, T., Tibshirani, R., Friedman, J.: *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York (2001)
10. Järvelin, K., Kekäläinen, J.: Cumulated gain-based evaluation of IR techniques. *ACM Trans. Inf. Syst.* **20**(4), 422–446 (2002)
11. Lane, J.: Let's make science metrics more scientific. *Nature* **464**(7288), 488–489 (2010)
12. Li, M., Liu, L., Li, C.-B.: An approach to expert recommendation based on fuzzy linguistic method and fuzzy text classification in knowledge management systems. *Exp. Syst. Appl.* **38**(7), 8586–8596 (2011)
13. Macdonald, C., Ounis, I.: Voting techniques for expert search. *Knowl. Inf. Syst.* **16**(3), 259–280 (2008)
14. Moreira, C., Calado, P., Martins, B.: Learning to rank for expert search in digital libraries of Academic publications. In: Antunes, L., Pinto, H.S. (eds.) *Progress in Artificial Intelligence*, pp. 431–445. Springer, Heidelberg (2011)
15. Niu, W., Liu, Z., Caverlee, J.: On local expert discovery via geo-located crowds, queries, and candidates. *ACM Trans. Spatial Algorithms Syst.* **2**(4), 14:1–14:24 (2016)
16. Qin, T., Liu, T.-Y., Xu, J., Li, H.: Letor: a benchmark collection for research on learning to rank for information retrieval. *Inf. Retr.* **13**(4), 346–374 (2010)
17. Ribeiro, I.S., Santos, R.L.T., Gonçalves, M.A., Laender, A.H.F.: On tag recommendation for expertise profiling: a case study in the scientific domain. In: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, pp. 189–198 (2015)
18. Ribeiro-Neto, B.A., Laender, A.H.F., de Lima, L.R.S.: An experimental study in automatically categorizing medical documents. *JASIST* **52**(5), 391–401 (2001)
19. Sanchez, D., Moreno, A.: Bringing taxonomic structure to large digital libraries. *Int'l. J. Metadata Semant. Ontol.* **2**(2), 112–122 (2007)
20. Seymour, E., Damle, R., Sette, A., Peters, B.: Cost sensitive hierarchical document classification to triage PubMed abstracts for manual curation. *BMC Bioinform.* **12**(1), 482 (2011)
21. Silla Jr., C.N., Freitas, A.A.: A survey of hierarchical classification across different application domains. *Data Mining Knowl. Disc.* **22**(1–2), 31–72 (2011)
22. Srinivasan, V., Fox, E.: Progress towards automated ETD cataloging. In: Proceedings of the 19th International Symposium on Electronic theses, dissertations: Data and dissertations (2016)

23. Waltinger, U., Mehler, A., Lösch, M., Horstmann, W.: Hierarchical classification of OAI metadata using the DDC taxonomy. In: Bernardi, R., Anderson, S., Bjrn, C., Frdrique, G., Zaihrayeu, S. (eds.) Advanced Language Technologies for Digital Libraries, pp. 29–40. Springer, Heidelberg (2011)
24. Yang, K.-W., Huh, S.-Y.: Automatic expert identification using a text categorization technique in knowledge management systems. *Expert Syst. Appl.* **34**(2), 1445–1455 (2008)
25. Yang, Y.: An evaluation of statistical approaches to text categorization. *Inf. Retr. J.* **1**(1–2), 69–90 (1999)
26. Yang, Y., Gopal, S.: Multilabel classification with meta-level features in a learning-to-rank framework. *Mach. Learn.* **88**(1), 47–68 (2012)