SIMT: A Semantic Interest Modeling Toolkit

MOHAMED AMINE CHATTI, University of Duisburg-Essen, Germany
FANGZHENG JI, University of Duisburg-Essen, Germany
MOUADH GUESMI, University of Duisburg-Essen, Germany
ARHAM MUSLIM, University of Duisburg-Essen, Germany
RAVI KUMAR SINGH, RWTH Aachen University

Interest modeling is a crucial task to achieve personalized services, such as recommendation. Applying interest modeling on textual data is often associated with semantic-related problems. In this paper, we focus on semantic interest modeling and present SIMT as a toolkit that harnesses the semantic information to effectively generate interest models and compute their similarities. SIMT follows a mixed-method approach that combines unsupervised keyword extraction algorithms, knowledge bases, and word embedding techniques to address the semantic issues in the interest modeling process. The SIMT approach has proven effective at the generation and similarity computation of interest models, outperforming standard interest modeling approaches.

Additional Key Words and Phrases: Interest modeling, Keyword extraction, Semantic similarity, Interest embedding

ACM Reference Format:
Mohamed Amine Chatti, Fangzheng Ji, Mouadh Guesmi, Arham Muslim, and Ravi Kumar Singh. 2021. SIMT: A Semantic Interest Modeling Toolkit. In *UMAP '21: ACM UMAP, June 21–25, 2021, Utrecht, the Netherlands.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

The rise of social media not only supports users to communicate with each other but also provides them opportunities for knowledge generation, collaborative learning, and exchanging resources [24]. This exponential growth of user-generated (big) data results in the problem of "information explosion" and introduces the challenge of extracting relevant information for the user. In order to address this issue, there is a need to understand the interests of a user, which can be used to automate and personalize content-based services, e.g. recommend news of interest [1], find other like-minded users [52], or predict the future interests [4]. Therefore, user interest modeling based on heterogeneous sources has become an increasingly important task in recent personalized systems [56].

Interest modeling can be seen as the process of constructing a model to represent individual user's interests based on their long-term and/or short-term information. The widespread usage of social media and digital publications attracted researchers to focus on *generating user interest model* based on textual content containing keywords/keyphrases, which can be self-annotated by the user or automatically extracted using keyword extraction algorithms. The generated interest model can be used to cluster users or perform recommendations based on the *similarity of interest models*. However, there are many challenges in both the interest model generation process and similarity computation that needs to be addressed. For example, based on the following two interest models I_a and I_b:

Manuscript submitted to ACM

I_a = { *educational technology, massive open online course, MOOC, personalized learning, personal learning environments, learning analytics, open learning analytics* }

I_b = { *learning environment, technology enhanced learning, E-learning, elearning, knowledge management, dataset* }

It can be seen that both models contain similar interests represented in form of acronyms (e.g. *MOOC* and *massive open online course*), synonyms (e.g. *technology enhanced learning* and *elearning*) and lexical variants (e.g. *elearning* and *E-learning*). There also exists overgeneration problem, e.g. the keyphrases *open learning analytics* and *learning analytics* represent the same interest *learning analytics*. Additionally, the keyword extraction algorithms can generate some irrelevant keywords (e.g. *dataset*) which might not describe the user's interest. Concerning the similarity of interest models, it is clear that both interest models are semantically similar as they contain acronyms, synonyms, and lexical variants. However, due to the lack of semantic knowledge, traditional similarity methods (e.g. Jaccard or cosine similarity) will fail to understand the user interest model and will identify the two interest models as different, which might influence the accuracy of a recommender system.

In this paper, we address these semantic problems in the interest modeling task and present a Semantic Interest Modeling Toolkit (SIMT) for the effective generation and similarity computation of interest models, based on semantic information. Inspired by the fact that an interest is often a well-defined concept (article) in Wikipedia, we introduce a novel method of interest modeling by leveraging Wikipedia as a knowledge-base to add semantic information in the generation of interest models process. Moreover, we propose new similarity measures that leverage Wikipedia and word embedding techniques to address the semantic issues in the computation of similarity between interest models. We further conduct experiments to compare our approach with traditional keyphrase extraction and similarity computation approaches in the interest modeling task.

The rest of the paper is organized as follows. We present related research work in the area of interest modeling in Section 2. Section 3 describes our proposed methodology and the architecture of SIMT. In Section 4, we go through concrete examples and experiments to evaluate the effectiveness of our approach for the generation and similarity computation of interest models. Finally, we summarize our results in Section 5 and provide perspectives for future work in the field of semantic interest modeling.

2 RELATED WORK

In the user modeling literature, many recent research works focus on inferring and analyzing user's interests. In this section, we first discuss the methods of generating user's interest models. Then, we review how similarity between user models is computed.

2.1 Interest Model Generation

Different text mining techniques have been used to generate user interest models. These include text classification [27, 38, 47], named-entity recognition [30, 39], and keyphrase extraction, which is widely used to generate interest models from text-based and social media data. In this work we focus on keyphrase extraction methods. Keyphrase/Keyword extraction is the task of automatically identifying a set of phrases/words that can best represent the content of a document/Web Page [5, 18]. The existing approaches can be categorized based on the type of the method (e.g. supervised, unsupervised) used to generate interest keyphrases/keywords [44].

In supervised approaches, classification algorithms are commonly used to allocate users to predefined interest classes based on their data. For example, Raghuram et al. [40] proposed a supervised approach that categorizes Twitter users based on three features (tweet-based, user-based, and time-series based) into six interest categories. Even though

supervised approaches are relatively simple and easy to apply, they are domain-dependent and limited in identifying only predefined interests which were used to train the prediction model.

Unsupervised approaches, on the other hand, are not dependent on any predefined prediction model. Thus, they can generate a more diverse set of user interests. Unsupervised keyword extraction approaches can be further divided into statistical-based and graph-based approaches. Among the statistical approaches, Latent Dirichlet Allocation (LDA) is one of the most commonly used unsupervised topic modeling techniques used to model user's interests. For instance, Mehrotra et al. [29] and Pu et al. [39] merged related tweets of a user in a single document and used LDA to generate interest models. However, LDA raises questions about the reliability and validity of such probabilistic approach in comparison to alternative methods [17, 19, 23, 39]. There are many other well-known statistical-based approaches available in the literature. For instance, Term Frequency–Inverse Document Frequency (TF-IDF) compares the frequency of a term in a document with regards to the whole collection [46]. Rapid Automatic Keyword Extraction (Rake) generates keywords based on features like word frequency, degree, and a ration between both measures [42]. Recently, a new statistical approach called YAKE! has been proposed which extracts keywords by devising and combining a number of features in order to describe the nature of each term [11].

For graph-based approaches, TextRank is one of the most famous methods [32]. It assumes that words are represented as vertices and uses a ranking algorithm similar to Google's PageRank [35] to construct a graph based on co-occurrence relationships in order to extract relevant keywords. SingleRank is another example of graph-based approaches based on TextRank where a classification algorithm is applied to define the value of words and those with the highest value are considered as keywords [50]. In other research works, the graph was constructed based on different features. For example, TopicalPageRank [28] and TopicRank [10] construct the graph through topics. Similarly, one of the most recent approaches called MultipartitieRank uses topics to build the graph and additionally applies TextRank to classify the candidate keywords [9]. PositionRank is another approach that uses the position information of the words' occurrence to build the graph [15].

Many studies in the literature compare selected keyword extraction approaches to investigate algorithms' performance in different scenarios. For instance, Wu et al. [54] inferred interests from user tweets using two standard keyword extraction algorithms, namely TF-IDF and TextRank. Their evaluation showed that TextRank performs better than TF-IDF when extracting top-5 keywords but TF-IDF outperforms TextRank when determining top-10 keywords. Similarly, Vu and Perez [49] explored the complementary relation between TF-IDF and TextRank in ranking interest candidates. The authors found out that TF-IDF and TextRank are both suitable for extracting user interests from tweets. Moreover, the combination of TF-IDF and TextRank consistently yields the highest user positive feedback.

Both supervised and unsupervised approaches are used for keyphrase/keyword extraction tasks. According to Caragea et al. [12], the supervised approaches usually have stronger modeling ability yet achieve higher accuracy than the unsupervised methods. On the other hand, unsupervised approaches have the ability to train the model without the need for labeled data and can be applied in various domains. However, both approaches do not consider the semantic issues during the keyword extraction phase.

In order to address this issue, some research works incorporated knowledge bases such as DBPedia [37], Freebase [57], Linked Open Data (LOD) cloud [2], YAGO [45], and Wikipedia to infer user interest models. In this work we focus on Wikipedia as knowledge base. Mihalcea and Csomai [31] use extracted N-grams from the document and check for their presence in a controlled vocabulary generated by taking all the titles of the Wikipedia articles. Jean-Louis et al. [21] use Wikipedia categories and their instances for generating features for candidate generation. They use N-grams to get candidate keywords and then use a classification algorithm to check the probability that the candidate is in

the Wikipedia category. Michelson and Macskassy [30] present a simple non-machine learning approach to discover Twitter users' topics of interest by examining the entities they mention in their tweets. Their approach leverages Wikipedia as a knowledge base to disambiguate and categorize the entities in the tweets, followed by developing a "topic profile" which characterizes users' topics of interest. However, the noisy and ambiguous nature of Twitter makes it challenging to find the entities within the tweets. Besel et al. [7] infer user interests by extracting named entities from user's followees using the English Wikipedia as knowledge base. They compare the profiles created with the followee-based approach against tweet-based profiles and find that the followee-based approach can compete with the state of the art. Narducci et al. [34] aim at providing more transparent and serendipitous user models. They compare two techniques, keyword-based and encyclopedic-based extracted from Wikipedia, for representing user preferences extracted from social networks. Their results indicate that using an encyclopedic-based representation better reflects user preferences, and helps to introduce new interesting topics. Tommaso et al. [48] extract preferences from messages on content sharing platforms, or induced from their list of followees, which represent an interest rather than a social relation between peers. In addition, user interests are matched with Wikipedia articles describing them.

Our approach for interest model generation goes beyond existing works by mixing unsupervised keyword extraction algorithms and Wikipedia as a knowledge base to generate semantically-enriched domain-independent user interest models.

2.2 Interest Model Similarity

Since in our work interest models are represented in textual format, we leverage the techniques for measuring the similarity between texts or words to measure the similarity between interest models. A basic metric called Jaccard similarity coefficient can be used to compare interest models. It measures the similarity between two sets of data by calculating the number of common interests in two interest models divided by the sum of all distinct interests in those two models. Although it is easy to interpret, it can not capture the semantic information and may provide erroneous results.

Bag-of-Words (BoW) is one of the most commonly used vector space models to represent text or document when computing similarity [43]. It can be either term-frequency based representation or binary representation [3]. However, this approach does not consider the problem that the semantically similar words can be lexical dissimilar and, thus, will not perform well.

To the best of our knowledge, there are no research works that focus on finding semantic similarity between interest models. However, semantic similarity has been introduced and applied by many researchers in the NLP community through (a) knowledge-based algorithms and (b) corpus-based algorithms. Knowledge-based similarity algorithms focus on identifying the degree of similarity between words using information derived from semantic networks [16]. Researchers have come up with six algorithms with WordNet, which is a large lexical database for English. Three of them are based on information content [22, 26, 41]. The other three measures are based on path length [16, 25, 55]. Another well known method called Wiki-Link Measure (WLM), proposed by Witten and Milne [53], uses the Wikipedia's hyperlink structure rather than its category or text content to compute the similarity between words/concepts. The basic idea of this measure is that two Wikipedia articles are considered to be topically related if there are many Wikipedia articles that link to both [45]. The semantic relatedness measure $sr(a, b)$ between two Wikipedia articles a and b is given by:

$$sr(a,b) = \frac{log(max(|A|,|B|) - log(|A \bigcap B|))}{log(|W|) - log(min(|A|,|B|))} \quad (1)$$

Where A and B are the sets of Wikipedia articles that link to a and b respectively and W is the set of all articles in Wikipedia. However, these methods are used to determine the similarity between words/concepts but cannot be used to determine similarity between interest models directly.

Corpus-based algorithms compute similarity between words by using large corpora. Thereby, semantic representations of the words are needed. Two of the most famous models are Latent Semantic Analysis (LSA) and Explicit Semantic Analysis (ESA). However, these methods require high performance computing power. Word embedding is another corpus-based semantic representation technology which has been widely used recently in various NLP applications. It was first proposed by Bengio et al. [6] for statistical language modeling. An important contribution was later provided by Mikolov et al. [33] who proposed a Continuous Bag of Words (CBOW) and Skip-gram (SG) model, also known as Word2Vec which significantly speeds up the training process. Other works encountered in the literature include GloVe [36], which generated the vector based on co-occurrence matrix, and FastText [8], which was based on Skip-gram model and each word is represented as a bag of character n-grams.

In our work, we apply word embedding methods to represent the generated interest models. Also, since we are generating the interest model using Wikipedia, we propose another semantic method by modifying the BoW score schema with Wiki-Link Measure. Based on these two methods, we can use the cosine measure to calculate the similarity between interest models.

3 SEMANTIC INTEREST MODELING TOOLKIT

The Semantic Interest Modeling Toolkit (SIMT) aims at using the semantic information to effectively generate interest models and compute their similarities. SIMT consists of two main components, namely "Interest Model Generation" and "Semantic Interest Model Similarity", as depicted in Figure 1. These components are developed as RESTful APIs allowing them to be easily used by any application that requires semantic user interest modeling. In the following sections, we present in detail both components in terms of approach, example, and evaluation.

3.1 Interest Model Generation

3.1.1 Approach. This component is responsible for generating a user's interest model. It contains two sub-components: 'Keyword Extractor' and 'Semantic Enrichment'. The 'Keyword Extractor' sub-component is responsible for extracting candidate interest keywords from the user-generated textual content (posts/publications) using various unsupervised keyword extraction algorithms including TextRank, SingleRank, TopicRank, TopicalPageRank, PositionRank, MultipartitieRak, Rake, and YAKE!. These algorithms are employed from three different open-source python libraries for keyword extraction, namely python keyphrase extraction (pke)[1], yake[2], and rake-nltk[3]. The 'Semantic Enrichment' sub-component leverages semantic information from Wikipedia to generate the user's interest model based on the candidate interest keywords generated from the 'Keyword Extractor' sub-component. The interest model generation process is depicted in Figure 2 and its steps are described below:

[1]https://github.com/boudinfl/pke
[2]https://pypi.org/project/yake/
[3]https://pypi.org/project/rake-nltk/

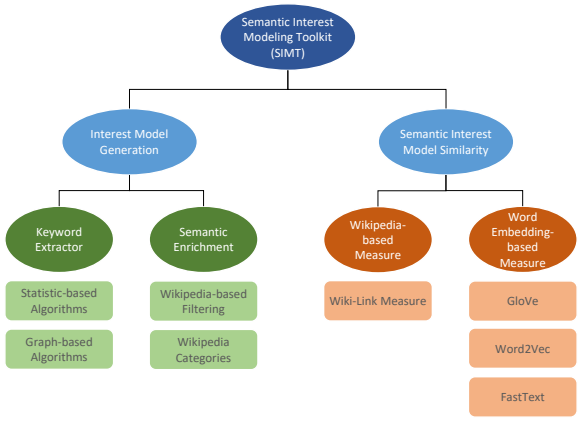

Fig. 1. SIMT abstract architecture

Keyword-based Interest Model

(1) Use different preprocessing techniques, such as data cleaning, transformation, reduction, and integration to prepare the data (textual content) for keyword extraction
(2) Apply an unsupervised keyword extraction algorithm on the preprocessed data to get candidate interest keywords
(3) Merge extracted keywords and assign them a weight based on their occurrences. Afterwards, normalize the weights to generate a keyword-based interest model of the user

Wiki-based Interest Model

(4) Leverage a lexical knowledge base (Wikipedia) to map the candidate interest keywords to their linked Wikipedia articles. That is, if there exists a Wikipedia article for the keyword, then the article's title is added to the Wiki-based interest model, otherwise, the keyword is discarded
(5) Identify redirected keywords, which do not have their own Wikipedia articles but they are linked to some other articles. Then, merge the weights of all the keywords redirecting to the same Wikipedia article in order to improve the Wiki-based interest model by reducing redundancy
(6) Normalize the weights of the Wiki-based interest model

Wiki Category-based Interest Model

(7) Select the top three interests in the Wiki-based interest model
(8) Generate a Wiki category-based interest model of the user using the categories of the selected top three interests from the Wiki-based interest model

3.1.2 Example. In order to clarify the interest model generation process, we present three interest models (i.e. keyword-based, wiki-based, and wiki category-based) for Prof. C (name omitted for blind review) based on his research publications. The first step in generating the interest model is to have a corpus on which the interests of Prof. C can be modeled. Therefore, his research publications from the last 5 years are gathered to create a publication dataset corpus.

Keyword-based Interest Model. The publication dataset corpus is used to generate the keyword-based interest model for Prof. C by following steps 1-3 of the interest model generation process. SingleRank is used as a keyword extraction

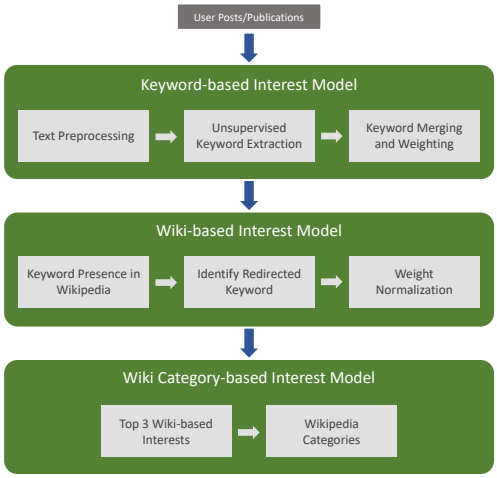

Fig. 2. Interest model generation process

algorithm since it outperformed all other keyword extraction algorithms (see Section 4). Figure 3 shows the top 15 interests of Prof. C in which the size of the interest represents its weight.

Fig. 3. Top 15 interests in the keyword-based interest model of Prof. C generated using SingleRank

Fig. 4. Top 15 interests in the Wiki-based interest model of Prof. C

Wiki-based Interest Model. Step 4-6 of the interest model generation process are followed to map the keyword-based interest model of Prof. C to their related Wikipedia articles. Followed by the identification of redirected keywords and merging of the semantically similar keyword weights. Finally, the weights of all the Wiki-based interest keywords are normalized to generate the Wiki-based interest model. The top 15 Wiki-based interests of Prof. C are shown in Figure 4. By comparing the keyword-based interest model (Figure 3) and the Wiki-based interest model (Figure 4), it can be observed that:

- Synonym interests are merged: *learning* and *learning process* are merged to *Learning*, *peer assessment* and *peer assessment process* are combined under *Peer assessment*
- Acronym interests are reduced: *moocs* is replaced with *Massive open online course*
- Less interesting keywords are removed: keywords having no linked Wikipedia articles are removed, e.g. *educational data sets*

Wiki Category-based Interest Model. Each article in Wikipedia belongs to a set of categories. For instance, *Personalized learning* belongs to *Pedagogy*, *Educational practices* and *Educational psychology*. Using the categories of the top three Wiki-based interests of Prof. C (i.e. *Learning*, *Massive open online course*, and *Learning Analytics*), the Wiki category-based interest model is generated, as shown in Figure 5.

Fig. 5. Wiki category-based interest model of Prof. C

3.2 Semantic Interest Model Similarity

3.2.1 Approach. This component is responsible for calculating semantic similarity scores between two interest models. The first step in calculating the similarity is to generate a vector representation of both models. Afterwards, the similarity is calculated by applying cosine similarity to the two interest model vectors. SIMT computes the semantic similarity of interest models using two different approaches, namely 'Wikipedia-based Measure' and 'Word Embedding-based Measure'. The detailed steps taken by both approaches to compute the semantic similarity between two interest models are discussed below:

Wikipedia-based Measure. This approach uses the Bag-of-Words (BoW) technique together with the Wiki-Link measure (see Section 2.2) to first generate vector representations of two interest models and then compute the cosine similarity between the two vectors, as shown in Figure 6. The steps to compute the semantic similarity between two interest models using the 'Wikipedia-based Measure' are described below:

(1) Generate an interest vector space containing all the interests from both interest models
(2) Represent each interest model with an interest vector and assign a score of '1' to all the keywords in the vector space which are also present in the interest model. For each keyword w in the vector space which is not present in the interest model, use the Wiki-Link measure (see Equation 1) to calculate the semantic relatedness (sr) between w and all the other keywords in the interest model having a score of '1'. Then, take an average of all the calculated semantic relatedness results and assign it as the score of w in the interest vector
(3) Calculate the cosine similarity between the two interest vectors

Word Embedding-based Measure. This approach uses word embedding techniques to generate *interest embeddings* (i.e. word embedding representations of interest) and *interest model embeddings* (i.e. word embedding representations of interest models). We use GloVe [36], as it is trained on Wikipedia data. The steps to compute the semantic similarity between two interest models using the 'Word Embedding-based Measure' are shown in Figure 7 and described below:

(1) Load a GloVe pre-trained word embedding model

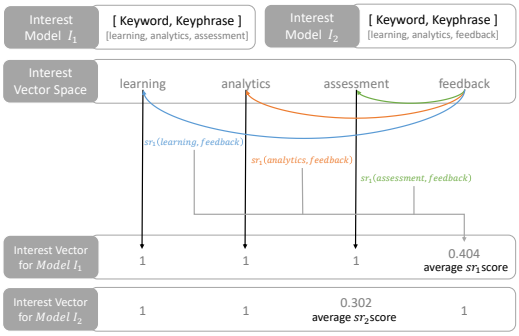

Fig. 6. Wikipedia-based Measure

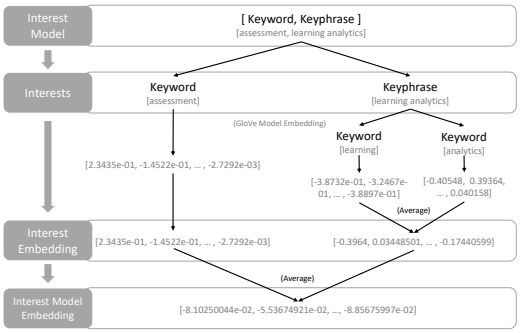

Fig. 7. Word Embedding-based Measure

(2) Collect all interests from an interest model. If an interest is a keyphrase, split it into keywords
(3) Use GloVe to create an interest embedding for each interest. If an interest is a keyphrase, take the average of the interest embeddings of each keyword in the keyphrase
(4) Compute the interest model embedding as the average value of all the interest embeddings
(5) Calculate the cosine similarity between the two interest model embeddings

4 EVALUATION

We performed interest modeling based on two different data sources, namely paper abstracts and tweets. We carried out offline evaluations as well as user studies to gauge the effectiveness of the "Interest Model Generation" and the "Semantic Interest Model Similarity" components in SIMT.

4.1 Interest Model Generation

4.1.1 Keyword Extraction Performance. We first evaluated the performance of various keyword extraction algorithms to select the best performing one for each data source, based on the Precision, Recall, and F-measure metrics. We chose the exact match as a way to compute these metrics. The performance measures for the different keyword extraction algorithms were benchmarked based on the Inspec publication dataset [20]. The results of the computation are summarized in Table 1, which indicates that SingleRank outperforms all other selected algorithms when extracting the top 10 and top 15 keywords. For the top 5 keywords, PositionRank performs better than the others and for the top 20 keywords, TopicalPageRank holds a slight lead. To evaluate the performance of the different algorithms on tweets, we relied on the evaluation results reported in [14] which revealed that YAKE! outformed all other algorithms in terms of Precision, Recall, and F-measure.

4.1.2 Offline Evaluation. We evaluated the performance of the three interest model generation methods: keyword-based interest model, wiki-based interest model, and wiki category-based interest model. We selected the top 15 interests extracted with the keyword-based and wiki-based methods. To avoid extracting too much wiki category-based interests, we used only the top 5 wiki-based interests to generate the wiki category-based interests. As result, for a given user, we have three sets of interests:

(1) *Keyword-based interests* extracted using the best performing keyword extraction algorithms. We use SingleRank for paper abstracts and YAKE! for tweets.

Table 1. Keyword extraction algorithms performance measures for Inspec dataset

Algorithm	K=5			K=10			K=15			K=20		
	P	R	F	P	R	F	P	R	F	P	R	F
TextRank	18.15	7.10	9.79	16.15	9.58	11.51	14.88	10.15	11.48	14.61	10.40	11.52
SingleRank	30.96	13.60	17.99	**26.95**	**22.04**	**23.02**	**23.57**	**27.01**	**24.03**	21.33	**30.27**	24.02
TopicRank	26.97	11.52	15.38	21.86	17.31	18.41	19.53	21.24	19.51	18.62	23.90	20.13
TopicalPageRank	30.36	13.37	17.67	26.31	21.44	22.47	23.34	26.43	23.69	**21.52**	29.88	**24.04**
PositionRank	**32.12**	**13.82**	**18.38**	25.45	20.79	21.77	22.79	25.80	23.15	21.14	29.20	23.56
MultipartitieRank	28.60	12.11	16.20	21.99	17.83	18.70	19.76	22.75	20.20	19.00	26.68	21.32
Rake	20.02	9.13	11.87	21.54	18.27	18.75	18.42	21.57	18.97	19.22	28.21	21.89
YAKE!	24.80	11.14	14.59	20.32	17.70	17.88	17.86	22.78	18.96	15.74	26.39	18.72

P - precision, R - recall, F - F-measure, K - number of keywords

(2) *Wiki-based interests* extracted using Wikipedia's article titles based on the terms in the keyword-based model.
(3) *Wiki Category-based interests* extracted using categories of Wikipedia articles based on the top 5 interest terms in the Wiki-based interest model.

To evaluate the results in these interest sets, we need to compare the extracted interests with some ground truth (i.e. known interests of some specific users extracted from their publications and/or tweets). Since the ground truth is difficult to obtain for a random user, we use two approaches in our evaluation, namely "using declared interests" and "using human feedback", as described in the following.

One way to obtain the ground truth of a user is by using the interests declared by the users themselves. To evaluate the extracted interests from paper abstracts, we selected researchers from a research training group at the University of X (name omitted for blind review). We collected their interests manually from different sources on the Web (e.g. Google scholar, LinkedIn, ReserchGate, personal web pages). For each one of them, the research publications from the last 5 years were gathered to generate their interest models. Then we conducted an empirical comparison between these author's interests and the extracted interests using the three methods outlined above. We show some of the results in Table 2.

Similarly, to evaluate the extracted interests from tweets, we selected some users who have indicated their interests in their 'bio' section on Twitter. Specifically, we looked for users whose bio contains a phrase such as "like <interest terms>", "love <interest terms>" or "interested in <interest terms>", or similar phrases. For each user, the posted tweets from the last 6 months were collected to generate their interest models. Then, we checked whether the extracted interests match the interests declared by the users themselves or not. Table 3 shows some examples of the results.

We performed interest modeling for 13 researchers and 10 twitter users. We quantitatively evaluated the keyword-based, wiki-based, and wiki category-based interest modeling results using the Precision and Recall measures, by counting the number of the extracted interests that match the declared interests. The results summarized in Table 4 show that the wiki-based method significantly improved the results of the keyword-based one, both in terms of Precision and Recall. This is mainly because some of the unrelated keywords are filtered and reduced by Wikipedia. The wiki category-based interest model did not perform well due to the fact that more abstract categories are associated with the wiki-based interests. Moreover, the accuracy of wiki category-based interests highly relies on the accuracy of the first 5 wiki-based interests. Unsurprisingly, all three methods achieved better results from publication abstracts compared with the results from tweets, mainly due to the noisy data in tweets. Moreover, some users post a lot of popular topics that they might not necessarily be interested in.

4.1.3 User Study. The goal of this user study was to identify which one of the three generated interest models best represents users' interests. We asked 10 participants to evaluate the quality of their extracted interests. Four participants

Table 2. Examples of group researchers' top interests

Researchers and their declared interests	Top interests extracted by the different methods		
	Keyword-based	Wiki-based	Wiki Category-based
Researcher 1 (names omitted for blind review): Social Computing, Web Information Systems, Data Science, Learning Technologies, Visual Analytics, Learning Analytics, Knowledge management	Learning, Learning process, Educational data sets, Knowldege management, La, Learning analytics, mooc, Recommender systems, Open learning analytics, Personal learning environment, Peer assessment process, Massive open online courses, Blended mooc environment, Peer assessment	Learning, Analytics, Massive open online course, Peer assessment, Learning analytics, Knowledge management, Data,Personalized learning, Learning environment, Recommender system, La	Cognitive science Developmental psychology, Educational psychology, Learning, Educational technology, Higher education, Open educational resources
Researcher 2: Information Retrieval, Retrieval models, User-oriented retrieval methods, Social media retrieval, Data Mining, Natural language processing	Information retrieval, IR, digital libraries, Prediction, evaluation, Social media, markov models Natural language processing, Web search, experimental results, Query level resource weighting	Information retrieval, IR, Digital library, Natural language processing, Prediction, Evaluation, Recommender system, Social media, Markov model, Web search engine	Information retrieval, Natural language processing, Digital libraries, Library science, Artificial intelligence, Computational linguistics, Speech recognition
Researcher 3: Social network analysis, Community support, Collaborative learning, Interactive and collaborative media for learning and knowledge construction, Analysis, modelling, and intelligent support of online learning	PBL, group, learning process, Discussion forum, Social media, PBL process, Online courses, Small group collaboration Group work, visual search, Learning environment, Learning analytics, Large online course Massive open online courses, Social network analysis	Group, Learning, Social media, Educational technology, Massive open online course, Social network analysis, Group work, Learning analytics, Learning environment, Visual search, Internet forum	Natural language processing, Data mining, Computational linguistics, Information retrieval, Artificial intelligence, Computational linguistics, Speech processing
Researcher 4: Human-Computer Interaction, Information Visualization, Intelligent User Interfaces, Semantic Data, Recommender System	Recommender system, User, users, group, Interactive, model, Information visualization, Collaborative filtering, User experience, user interface, Group recommender system, Latent factors, task models recommendation quality	Recommender system User, Model, Information Visualization, Group, Interactivity, User experience, Collaborative filtering, User Interface, Explanation	Information systems, Recommender systems, Information visualization, Computational science, Computer graphics, Infographics, Scientific modeling

were PhD students, whose interests were extracted from their publication data. The rest of them were Twitter users whose interests were extracted from tweets. For each participant, the keyword-based, wiki-based, and wiki category-based interest models were built and the top 15 interests were shown to him or her in form of a word cloud chart. To prevent bias in judgment, the three interest models were anonymized, i.e., the participants were not informed which interest model is generated by which method. Inspired by the evaluation conducted in [34], we evaluated for each interest model the *accuracy* as the overlapping between the extracted and the actual interests as well as the *serendipity* as the presence of unexpected but still relevant interests, which can be important in recommender systems in order to suggest surprisingly interesting items that users might not have otherwise discovered . The users provided their feedback on accuracy and serendipity of the generated interest models using a 5-points discrete rating scale. The results are summarized in Table 5.

Two main results can be observed. The wiki-based interest model best reflects the actual user's interests. The highest serendipity was provided by the wiki category-based interest models, as the wiki category-based method is the only one that enriches the model with new interests. These results confirm that leveraging Wikipedia in the interest modeling task leads to more accurate and serendipitous interests.

4.2 Semantic Interest Model Similarity

4.2.1 Offline Evaluation. We selected Prof. C and Prof. H from the University of X to compare the semantic similarity between them based on their publication dataset corpus. We first generate the three interest models (i.e. keyword-based,

Table 3. Examples of Twitter users' top interests

Twitter users and their declared interests	Top interests extracted by the different methods		
	Keyword-based	Wiki-based	Wiki Category-based
@GrahamDumpleton: C/Python developer. Interested in Apache, WSGI, Python web hosting, Jupyter/JupyterHub, Kubernetes, OpenShift, Docker and Platform as a Service (PaaS).	Jupyter notebook, Python, Open source, red hat, wsgi, kubernetes apache https, Python web, Deploying jupyter, Python version, Docker image, wsgi application, Wsgi server, anaconda python, Plain kubernetes	Project Jupyter, Python, Open source, Web Server Gateway Interface, Kubernetes, Red Hat, Apache HTTP Server	Collaborative software, Computer law, Data publishing, Free culture movement, Free simulation software, Cloud infrastructure, Containerization software, Linux Containerization
@francisjigo2: Tech enthusiast. Interested in data science, machine learning, web development and product engineering.	machine learning, data science, web development, problem data, data analysis, learning algorithms, top machine, science competition, expires june, matt przybyla, data structures, collect data, data website, python data, tensorflow keras	Machine learning, Data science, Web development, Data analysis, Science fair, Data structure	Machine learning, Cybernetics, Learning, Information science, Computer occupations, Data analysis, Web development, Scientific method, Science education
@Ze_Judge: Interested in Cartoons,Anime,Manga, Video Games,Gravity Falls enthusiast/Critical Role has taken over my life/toss a coin to your wizard.	animation, game, pokemon, video game, animal crossing, work, draw, cat, gravity falls, Smash Bros, time, finally finished, zelda game, years ago, flash spark	Animation, Game, Pokémon, Video game, Animal Crossing, Work, Draw, Cat, Gravity Falls, Time, Super Smash Bros	Animation,Cartooning, Film and video technology, Games, Leisure activities, Japanese brands, Mass media franchises, Nintendo franchises, Pokémon, Digital media, Video games, Metaphysical theories, Metaphysics, Time
@AJStein_de: Agricultural economist. Interested in food, nutrition, health, technology, sustainability, development, trade. PhD. Personal account.	food amp, health, amp crops, food system, amp production, crops amp, food production, ghg amp, production amp, higher amp, global food, food, food supply, food crops, crop production	Health, Food system, Food industry, Food, Food security, Crop, Agriculture	Health,Personal life, Agricultural economics, Food and the environment, Food industry, Farms, Agriculture, Mass production, Foods, Cuisine, Food security, Sustainable food system, Climate change

Table 4. Interest modeling performance results

Interest Model	Publications		Tweets		Average	
	Precision	Recall	Precision	Recall	Precision	Recall
Keyword-based interests	0.662	0.759	0.519	0.559	0.591	0.659
Wiki-based interests	**0.746**	**0.769**	**0.650**	**0.589**	**0.698**	**0.679**
Wiki Category-based interests	0.578	0.651	0.545	0.556	0.562	0.604

Table 5. Interest models accuracy and serendipity

Interest Model	Accuracy				Serendipity			
	Avg. Score	Min Score	Max Score	Std. dev	Avg. Score	Min Score	Max Score	Std. dev
Keyword-based	2.8	1	4	0.98	2	1	4	0.78
Wiki-based	**3.3**	1	5	1.2	1.7	1	3	0.64
Wiki Category-based	2.3	1	4	1.19	**2.9**	1	5	1.04

wiki-based, and wiki category-based) for both researchers using the "Interest Model Generation" component. The top 5 interests are selected in the first two interest models.

- Interest models for Prof. C
  - Keyword-based interests: *moocs*, *learning process*, *massive open online courses*, *learning*, *learning analytics*
  - Wiki-based interests: *Learning*, *Massive open online course*, *Learning analytics*, *Analytics*, *Peer assessment*
  - Wiki Category-based interests: *Educational psychology*, *Intelligence*, *Cognitive science*, *Systems science*, *Learning*, *Neuropsychological assessment*, *Developmental psychology*, *Higher education*, *Free software*, *Educational technology*, *Open educational resources*, *E-learning*, *Statistics of education*, *Types of analytics*
- Interest models for Prof. H

- Keyword-based interests: *groups*, *online courses*, *knowledge exchange*, *social network analysis*, *learning process*
- Wiki-based interests: *Group*, *Educational technology*, *Knowledge transfer*, *Social network analysis*, *Learning*
- Wiki Category-based interests: *Education terminology*, *Technology in society*, *Educational technology*, *E-learning*, *Educational psychology*, *Knowledge transfer*, *Information society*

Based on the generated interest models, we use the 'Wikipedia-based Measure (WLM)' and the 'Word Embedding-based Measure (WEM)' from the "Semantic Interest Model Similarity" component to compute the similarity between the interest models of the two researchers. These measures are then compared to two traditional similarity measures, namely 'Jaccard Similarity Measure (JSM)' and 'Bag of Words-based Measure (BWM)'. The results are summarized in Table 6.

Table 6. Similarity between Prof. C and Prof. H based on different interest models and similarity measures

Interest Model	Similarity Measure	Similarity Score
Keyword-based	JSM	11.11%
	BWM	20.00%
	WEM	57.52%
Wiki-based	JSM	11.11%
	BWM	20.00%
	WLM	74.35%
	WEM	67.13%
Wiki category-based	JSM	16.66%
	BWM	30.00%
	WLM	58.22%
	WEM	51.47%

By observing the interest models of the two researchers, it can be seen that they have high semantic similarity. The traditional JSM and BWM approaches failed to capture this semantic information and gave low similarities. Using the WLM and WEM semantic measures enabled to significantly increase the similarity between the two interest models. Moreover, we found that the similarity results between the wiki category-based interest models were less than the wiki-based ones. This is due to the different and more abstract categories associated with each Wikipedia-based interest. Furthermore, WLM resulted in a higher similarity compared to WEM. However, in our experiment, computing the similarity based on WLM run much slower than the WEM approach, mainly due to the many Wikipedia API requests that had to be performed.

4.2.2 User Study. Evaluating similarity measures offline is a tough task since it is difficult to obtain the ground truth of the similarity between two given interest models. Thus, we conducted a user study to compare the performance of our proposed two semantic similarity measures WLM and WEM with the two traditional similarity measures JSM and BWM.

We used the pairwise comparison approach to assess the performance of each pair of the four measures WLM, WEM, JSM and BWM. This pairwise comparison method has been previously applied in the literature to compare different similarity metrics [13, 51]. Specifically, given an interest model (we call it the *base* interest model) and two measures M1 and M2 (a *measure pair*), we use measure M1 and M2 to choose two *candidate* interest models that are the most similar to the base interest model. Two candidate interest models together with their base interest model are called an *interest model set*. For each measure pair, we sample a number of interest model sets. We conduct a user study to obtain the ground truth from human judgement by asking the evaluators to judge if the candidate interest models are similar to the base interest model. A measure in a measure pair obtains "1" if it aligns with the human judgment on an interest model set, "0" otherwise.

In order to obtain human judgments, we performed a user study with five of the evaluators from the previous study. We use the 23 user's wiki-based interest models generated in the previous study as the pool of base interest models. For each evaluator, we randomly picked 10 models from the base interest model pool. For a given measure pair, each measure selects the most similar interest model to a base model as the candidate interest model. Accordingly, we obtain 10 interest model sets for each measure pair (60 in total). Each evaluator was presented with multiple interest model sets. Then he or she was asked to choose between the two candidate interest models the one which is more similar to the base interest model. If an evaluator cannot make a decision, he or she can choose the option to select either of them. The number of votes for the 6 metric pairs are listed in Table 7. The results revealed that our proposed WEM significantly outperforms the traditional JSM and BWM (35 vs. 7 and 36 vs. 8, respectively). Similarly, the WLM was also significantly better than JSM and BWM. However, there was no difference between WEM and WLM. To conclude, the study showed that our proposed semantic similarity measures align best with human judgements, and significantly outperform traditional similarity measures in estimating the similarity of interest models.

Table 7. Comparison of the 6 metric pairs

Metrics	WEM vs. WLM	WEM vs. BWM	WEM vs. JSM	WLM vs. BWM	WLM vs. JSM	BWM vs. JSM
number of votes	30 vs. 32	35 vs.7	36 vs.8	31 vs. 11	32 vs. 10	10 vs. 8

5 CONCLUSION AND FUTURE WORK

In this paper, we highlighted semantic-related problems in the interest modeling task. We presented the Semantic Interest Modeling Toolkit (SIMT) that harnesses the semantic information based on Wikipedia and word embedding techniques for the effective generation and similarity computation of interest models. The evaluation results show that our approach can deal with the semantic problems in interest modeling and is effective to improve the quality of the generated interest models and the accuracy of the similarity between two interest models. As future work, we will use these approaches as initial steps to provide transparent, scrutable, and explainable user interest models as well as visually explainable, interest-based recommendations that take the semantic information into account.

REFERENCES

[1] Fabian Abel, Qi Gao, Geert-Jan Houben, and Ke Tao. 2013. Twitter-based user modeling for news recommendations. In *Twenty-Third International Joint Conference on Artificial Intelligence*.

[2] F. Abel, C. Hauff, G. Houben, and K. Tao. 2012. Leveraging User Modeling on the Social Web with Linked Data. In *ICWE*.

[3] Sunil Aryal, Kai Ming Ting, Takashi Washio, and Gholamreza Haffari. 2019. A new simple and effective measure for bag-of-word inter-document similarity measurement. *arXiv preprint arXiv:1902.03402* (2019).

[4] Hongyun Bao, Qiudan Li, Stephen Shaoyi Liao, Shuangyong Song, and Heng Gao. 2013. A new temporal and social PMF-based method to predict users' interests in micro-blogging. *Decision Support Systems* 55, 3 (2013), 698–709.

[5] Slobodan Beliga, Ana Meštrović, and Sanda Martinčić-Ipšić. 2015. An overview of graph-based keyword extraction methods and approaches. *Journal of information and organizational sciences* 39, 1 (2015), 1–20.

[6] Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Jauvin. 2003. A neural probabilistic language model. *Journal of machine learning research* 3, Feb (2003), 1137–1155.

[7] Christoph Besel, Jörg Schlötterer, and Michael Granitzer. 2016. On the quality of semantic interest profiles for onine social network consumers. *ACM Sigapp Applied Computing Review* 16 (2016), 5–14.

[8] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching Word Vectors with Subword Information. *Transactions of the Association for Computational Linguistics* 5 (2017), 135–146.

[9] Florian Boudin. 2018. Unsupervised keyphrase extraction with multipartite graphs. *arXiv preprint arXiv:1803.08721* (2018).

[10] Adrien Bougouin, Florian Boudin, and Béatrice Daille. 2013. Topicrank: Graph-based topic ranking for keyphrase extraction.

[11] Ricardo Campos, Vítor Mangaravite, Arian Pasquali, Alípio Mário Jorge, Célia Nunes, and Adam Jatowt. 2018. YAKE! collection-independent automatic keyword extractor. In *European Conference on Information Retrieval*. Springer, 806–810.

[12] Cornelia Caragea, Florin Adrian Bulgarov, Andreea Godea, and Sujatha Das Gollapalli. 2014. Citation-enhanced keyphrase extraction from research papers: A supervised approach. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 1435–1446.

[13] Anjie Fang, Craig Macdonald, Iadh Ounis, and Philip Habel. 2016. Using word embedding to evaluate the coherence of topics from Twitter data. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. 1057–1060.

[14] André Filipe Neves Farinha. 2018. *Extracting keywords from tweets*. Ph.D. Dissertation.

[15] Corina Florescu and Cornelia Caragea. 2017. Positionrank: An unsupervised approach to keyphrase extraction from scholarly documents. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 1105–1115.

[16] Wael H Gomaa and Aly A Fahmy. 2013. A survey of text similarity approaches. *International Journal of Computer Applications* 68, 13 (2013), 13–18.

[17] E Dario Gutiérrez, Ekaterina Shutova, Patricia Lichtenstein, Gerard de Melo, and Luca Gilardi. 2016. Detecting cross-cultural differences using a multilingual topic model. *Transactions of the Association for Computational Linguistics* 4 (2016), 47–60.

[18] Kazi Saidul Hasan and Vincent Ng. 2014. Automatic keyphrase extraction: A survey of the state of the art. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 1262–1273.

[19] Tobias Hecking and Loet Leydesdorff. 2019. Can topic models be used in research evaluations? Reproducibility, validity, and reliability when compared with semantic maps. *Research Evaluation* 28, 3 (2019), 263–272.

[20] Anette Hulth. 2003. Improved automatic keyword extraction given more linguistic knowledge. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*. Association for Computational Linguistics, 216–223.

[21] Ludovic Jean-Louis, Michel Gagnon, and Eric Charton. 2013. A knowledge-base oriented approach for automatic keyword extraction. *Computación y Sistemas* 17, 2 (2013), 187–196.

[22] Jay J Jiang and David W Conrath. 1997. Semantic similarity based on corpus statistics and lexical taxonomy. *arXiv preprint cmp-lg/9709008* (1997).

[23] Coriane Nana Jipmo, Gianluca Quercini, and Nacéra Bennacer. 2017. FRISK: A Multilingual Approach to Find twitteR InterestS via wiKipedia. In *Advanced Data Mining and Applications*, Gao Cong, Wen-Chih Peng, Wei Emma Zhang, Chengliang Li, and Aixin Sun (Eds.). Springer International Publishing, Cham, 243–256.

[24] Andreas U Kuswara and Debbie Richards. 2011. Realising the potential of Web 2.0 for collaborative learning using affordances. *J. UCS* 17, 2 (2011), 311–331.

[25] C Leacock and M Chodorow. 1998. Combining local context and WordNet sense similarity for word sense identification. WordNet, An Electronic Lexical Database. *The MIT Press* (1998).

[26] Dekang Lin. 1998. Extracting collocations from text corpora. In *First workshop on computational terminology*. Citeseer, 57–63.

[27] Kangmiao Liu, Wei Chen, Jiajun Bu, Chun Chen, and Lijun Zhang. 2007. User modeling for recommendation in blogspace. In *2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology-Workshops*. IEEE, 79–82.

[28] Zhiyuan Liu, Wenyi Huang, Yabin Zheng, and Maosong Sun. 2010. Automatic keyphrase extraction via topic decomposition. In *Proceedings of the 2010 conference on empirical methods in natural language processing*. Association for Computational Linguistics, 366–376.

[29] Rishabh Mehrotra, Scott Sanner, Wray Buntine, and Lexing Xie. 2013. Improving lda topic models for microblogs via tweet pooling and automatic labeling. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*. ACM, 889–892.

[30] Matthew Michelson and Sofus A Macskassy. 2010. Discovering users' topics of interest on twitter: a first look. In *Proceedings of the fourth workshop on Analytics for noisy unstructured text data*. ACM, 73–80.

[31] Rada Mihalcea and Andras Csomai. 2007. Wikify! Linking documents to encyclopedic knowledge. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*. 233–242.

[32] Rada Mihalcea and Paul Tarau. 2004. Textrank: Bringing order into text. In *Proceedings of the 2004 conference on empirical methods in natural language processing*. 404–411.

[33] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. 3111–3119.

[34] F. Narducci, C. Musto, G. Semeraro, P. Lops, and M. Degemmis. 2013. Leveraging Encyclopedic Knowledge for Transparent and Serendipitous User Profiles. In *UMAP*.

[35] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. *The PageRank citation ranking: Bringing order to the web*. Technical Report. Stanford InfoLab.

[36] Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. 1532–1543.

[37] Guangyuan Piao and John G Breslin. 2016. User modeling on Twitter with WordNet Synsets and DBpedia concepts for personalized recommendations. In *proceedings of the 25th ACM international on conference on information and knowledge management*. 2057–2060.

[38] Bayu Yudha Pratama and Riyanarto Sarno. 2015. Personality classification based on Twitter text using Naive Bayes, KNN and SVM. In *2015 International Conference on Data and Software Engineering (ICoDSE)*. IEEE, 170–174.

[39] Xiao Pu, Mohamed Amine Chatti, Ulrik Schroeder, et al. 2016. Wiki-lda: A mixed-method approach for effective interest mining on twitter data. In *Proceedings Of The 8Th International Conference On Computer Supported Education, Vol 1 (Csedu)*. Scitepress, 426–433.

[40] Mandyam Annasamy Raghuram, K Akshay, and K Chandrasekaran. 2016. Efficient user profiling in twitter social network using traditional classifiers. In *Intelligent systems technologies and applications*. Springer, 399–411.

[41] Philip Resnik. 1995. Using information content to evaluate semantic similarity in a taxonomy. *arXiv preprint cmp-lg/9511007* (1995).

[42] Stuart Rose, Dave Engel, Nick Cramer, and Wendy Cowley. 2010. Automatic keyword extraction from individual documents. *Text mining: applications and theory* 1 (2010), 1–20.

[43] Gerard Salton and Michael J. McGill. 1986. *Introduction to Modern Information Retrieval*. McGraw-Hill, Inc., USA.

[44] Nacéra Bennacer Seghouani, Coriane Nana Jipmo, and Gianluca Quercini. 2019. Determining the interests of social media users: two approaches. *Information Retrieval Journal* 22, 1-2 (2019), 129–158.

[45] Wei Shen, Jianyong Wang, Ping Luo, and Min Wang. 2013. Linking named entities in tweets with knowledge base via user interest modeling. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. 68–76.

[46] Karen Sparck Jones. 1972. A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation* 28, 1 (1972), 11–21.

[47] Mia Stern, Joseph Beck, and Beverly Park Woolf. 1999. Naive Bayes classifiers for user modeling. *Center for Knowledge Communication, Computer Science Department, University of Massachusetts* (1999).

[48] Giorgia Di Tommaso, Stefano Faralli, Giovanni Stilo, and Paola Velardi. 2018. Wiki-MID: A Very Large Multi-domain Interests Dataset of Twitter Users with Mappings to Wikipedia. In *International Semantic Web Conference*.

[49] Thuy Vu and Victor Perez. 2013. Interest mining from user tweets. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*. ACM, 1869–1872.

[50] Xiaojun Wan and Jianguo Xiao. 2008. Single Document Keyphrase Extraction Using Neighborhood Knowledge.. In *AAAI*, Vol. 8. 855–860.

[51] Xi Wang, Anjie Fang, Iadh Ounis, and Craig Macdonald. 2019. Evaluating Similarity Metrics for Latent Twitter Topics. In *European Conference on Information Retrieval*. Springer, 787–794.

[52] Xufei Wang, Huan Liu, and Wei Fan. 2011. Connecting users with similar interests via tag network inference. In *Proceedings of the 20th ACM international conference on Information and knowledge management*. ACM, 1019–1024.

[53] I.H. Witten and D. Milne. 2008. An effective, low-cost measure of semantic relatedness obtained from Wikipedia links. In *Proceeding of AAAI Workshop on Wikipedia and Artificial Intelligence: an Evolving Synergy, AAAI Press, Chicago, USA*. 25–30. https://www.aaai.org/Papers/Workshops/2008/WS-08-15/WS08-15-005.pdf

[54] Wei Wu, Bin Zhang, and Mari Ostendorf. 2010. Automatic generation of personalized annotation tags for twitter users. In *Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics*. 689–692.

[55] Zhibiao Wu and Martha Palmer. 1994. Verbs semantics and lexical selection. In *Proceedings of the 32nd annual meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 133–138.

[56] Renfeng Yang, Wenbo Xie, and Duanbing Chen. 2018. A Three-layer Model on Users' Interests Mining. *J. Inf. Sci.* 44, 1 (Feb. 2018), 136–144. https://doi.org/10.1177/0165551517743645

[57] Xiao Yu, Hao Ma, Bo-June Hsu, and Jiawei Han. 2014. On building entity recommender systems using user click log and freebase knowledge. In *Proceedings of the 7th ACM international conference on Web search and data mining*. 263–272.