

Article

Expert Finding Considering Dynamic Profiles and Trust in Social Networks

Kyoungsoo Bok ¹, Inbae Jeon ², Jongtae Lim ² and Jaesoo Yoo ^{2,*} 

¹ Department of SW Convergence Technology, Wonkwang University, Iksandae 460, Iksan, Jeonbuk 54538, Korea; ksbok@wku.ac.kr

² Department of Information and Communication Engineering, Chungbuk National University, Chungdae-ro 1, Seowon-Gu, Cheongju, Chungbuk 28644, Korea; lovejib@naver.com (I.J.); efzotz@gmail.com (J.L.)

* Correspondence: yjs@chungbuk.ac.kr; Tel.: +82-43-261-3230

Received: 24 September 2019; Accepted: 13 October 2019; Published: 15 October 2019



Abstract: Recently, social network services that express individual opinions and thoughts have been significantly developed. As unreliable information is generated and shared by arbitrary users in social network services, many studies have been conducted to find users who provide reliable and professional information. In this paper, we propose an expert finding scheme to discover users who can answer users' questions professionally in social network services. We use a dynamic profile to extract the user's latest interest through an analysis of the user's recent activity. To improve the accuracy of the expert finding results, we consider the user trust and response quality. We conduct a performance evaluation with the existing schemes through various experiments to verify the superiority of the proposed scheme.

Keywords: social network; expert finding; social activity; trust; dynamic profile

1. Introduction

Due to the complexity and diversity of societies, the information distribution has changed rapidly, and the needs of users have also diversified. As the active participation of users through online services increases, Social Network Services (SNS) are being used as a media platform to deliver information among users based on human networks [1–3]. Recently, SNS have been changed into media to create and reprocess various types of information [4–8]. In SNS, inaccurate or meaningless information can be shared among users, since untrusted users can generate and distribute information indiscriminately [9–12]. Therefore, it is difficult to select the information that users want and to find users who provide the information they want among many users [13–17]. In addition, it is not guaranteed that a user connected to a network relationship is an expert who can provide a high quality response for user questions [18–23].

There have been studies on SNS to find experts who can provide users with high-quality information or answer users' questions [23–28]. The existing expert finding schemes determine users with keywords of interest similar to the user's desired information using the profile of social network users [29,30]. For example, an expert is found by comparing keywords included in a user's question with keywords extracted from the user's profile [15,18–20,31,32]. In [15], an expert finding scheme proposed to determine experts by comparing question keywords and keywords included the user's profile. To improve the accuracy of expert finding, the schemes that analyze the user's recent activities and compare them with question keywords have been proposed [33–35]. Since only the keywords of interest are extracted through user activity analysis, a social network user that creates a lot of information is determined as an expert. This does not guarantee that the user identified is an

expert [36–38]. It is not possible to determine whether a user identified as an expert can answer the question well [39–44]. Therefore, the trust and quality of the identified user should be considered to receive reliable responses to the user's question.

We have previously proposed a basic expert finding scheme [45]. In this paper, we propose an effective scheme to find an expert who can answer users' questions with reliability, as the extended version of the scheme proposed in [45]. The proposed scheme analyzes the user's recent social activities to generate a dynamic profile to express the user's interest. To ensure the latest user's dynamic profile, the user's social activities are periodically collected and analyzed. The generated dynamic profile is compared with question keywords extended through ontology to extend the finding ranges to keywords with similar meanings. Since the user who created and distributed meaningless or false information can be identified as an expert, the untrusted user should be removed, and the expert who can answer the question should be identified. For this, the proposed scheme determines the trust of the social network users and considers the response quality of the users identified as experts.

This paper is organized as follows. In Section 2, related works are described. The proposed scheme that finds experts in social network environments is explained in Section 3. In Section 4, the performance of the proposed scheme is verified through various performance evaluations. Finally, the conclusion of this paper is presented in Section 5.

2. Related Work

To identify users with relevant expertise or experience for a given topic, Zhang proposed a propagation-based scheme that considers both person local information and social relationships between users [19]. To calculate an initial expert score for each user, the user location information is used. The social relationship is used to improve the accuracy of expert finding.

Fu proposed an expertise propagation scheme to find experts using social network analysis [31]. This scheme generated the social networks from email messages and web pages to discover other potential experts. A candidate can acquire extra expertise probability from a trusted expert who has a strong relationship with the candidate. Bozzon proposed an expert finding scheme considering expertise information and routes to reach experts to select top-k experts [34]. To analyze the expertise of a user, profile and social activities were considered.

Duchateau proposed three strategies for finding an acquaintance who is able to answer a given question [15]. First, the proposed scheme integrates profiles of the same user to detect the interests of a user. Each profile includes a set of tags and a set of acquaintances. Three strategies, such as simple matching, clusters, and trees, are developed to rank acquaintances. A simple matching strategy finds experts by comparing user profiles and question words, along with cluster and tree strategies, in order to improve the expert finding performance. The cluster strategy expands queries and profile information with a thesaurus, using synonyms to compare them. The tree strategy constructs semantic inclusion relationships between words as a tree, thereby finding profile information similar to the queried keywords to improve the processing speed.

Vu developed a social search engine called Odin to help users search targeted answers [26]. In this approach, a link-based latent variable model is applied to infer social relationships between users from their social network to form a strength-weighted relationship graph. User expertise is determined by mining social network data, as well as sensor data, from mobile devices. A pagerank-like algorithm considering the relationship strength and the user expertise is used to search for a user that is able to answer the question.

3. The Proposed Expert Finding Scheme

3.1. Architecture

In this paper, we propose a scheme to find experts who can provide reliable responses to users' questions in SNS. The proposed scheme considers the trust and response quality of the user identified

as an expert to improve the accuracy of the expert finding results. The existing schemes use a static profile described by users directly or generate a recent profile through analyzing social activities. In the proposed scheme, not only user created information and shared information, but also additional social activities such as comments, scraps, and recommendations, are analyzed to generate a user's dynamic profile. The user trust is determined by using the user reputation so that an untrusted user is not extracted as an expert. The user who can provide a reliable response to the question is determined by using the dynamic profile and trust. The user who can provide a high-quality response is identified by considering the user's past response quality. The proposed scheme employs a server that finds experts as the main server; a database storage system that manages user profiles; and ontology to solve the issues of inaccuracy for the expert finding results, as well as the time cost due to matching results in real time in existing schemes. The main server manages the database storage system and performs query processing to reduce the time cost of matching results and improve the accuracy by utilizing ontology. The database storage system analyzes recent social activities of users in social networks periodically and updates dynamic profiles to ensure up-to-date and accurate results.

Figure 1 shows the proposed expert finding system architecture. The expert finding system consists of seven modules, including social activity analysis, reputation analysis, response analysis, query analysis, clustering creation, ontology, and result generation. The social activity analysis, reputation analysis, and response analysis determine the dynamic profile, user trust, and user quality. The proposed scheme analyzes recent social activity periodically and stores dynamic profiles to solve the issues in existing schemes, which only produce a finding when they receive a question. Ontology contains the synonyms, hyponyms, and antonyms of words. In addition, a query cluster is configured using ontology, and the finding results are provided to users by matching their question with dynamic profiles in the database storage system.

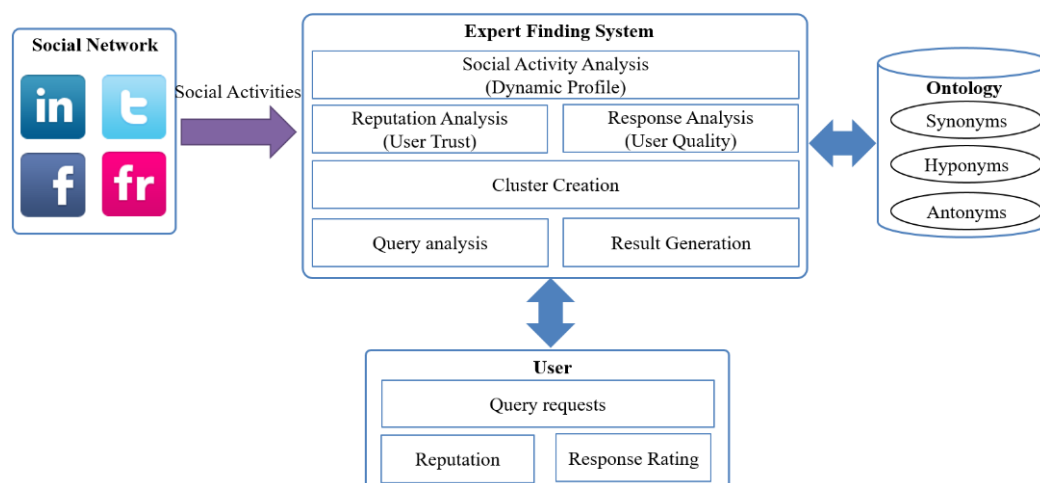


Figure 1. Proposed expert finding system architecture.

Figure 2 shows the process of expert finding of the proposed scheme. The scheme periodically analyzes the social activities, reputation, and response histories to determine the dynamic profile, user trust, and user quality. When a user asks a question in an expert finding system, the server analyzes the question and extracts the keywords from the question. The extracted keywords generate a cluster using ontology to produce expanded results. The candidate experts are generated by comparing the user's dynamic profile with the extracted keywords included in a cluster. The final experts which can provide a reliable response are recommended by using the user trust and user quality. The user evaluates the response quality of the expert and the evaluated content is used for the answer quality of the expert.

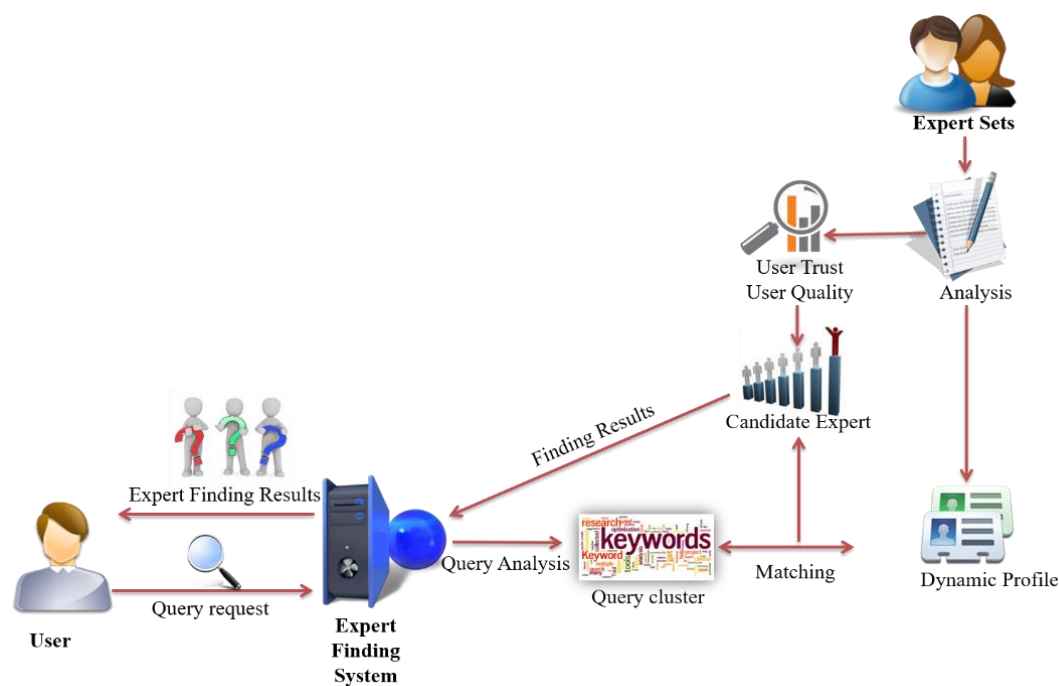


Figure 2. Proposed expert finding process.

3.2. Dynamic Profile

In SNS, users build a human network and write personal information and interests in profiles to find users with similar preferences and expand human networking. In this paper, these profiles are called static profiles. General users do not update their profiles, even if their interests change. The profile that the user directly described is written from a personal perspective and so it is difficult to use it as a way to express the user's expertise. Therefore, the social activities are analyzed to generate a dynamic profile which represents the user's interest in an objective perspective.

Figure 3 shows the process of social activity analysis employed to generate a dynamic profile. The social activity analysis consists of three steps, such as social information collection, social information analysis, and dynamic profile generation. The social information collection step collects the user-related documents and the other user's social activities. The user-related documents are documents created or shared through the user's direct social activities and the other user's social activities are scrap posts, recommendations, comments, and the number of users included by other users in user-related documents. The social information analysis step extracts keywords from the user-related documents and calculates the influence through analyzing the other user's social activities. The more users participate in the user-related documents, the more popular and reliable it is, and the more weight it has. The dynamic profile creation step calculates an expertise score by assigning influence and weight to keywords extracted from the document. The average value of the expertise score and the occurrence frequency of each keyword are calculated, because the same keywords can be extracted from different documents. When the social activity analysis is completed, the user's dynamic profile is generated through the keyword list, expertise score, and occurrence frequency.

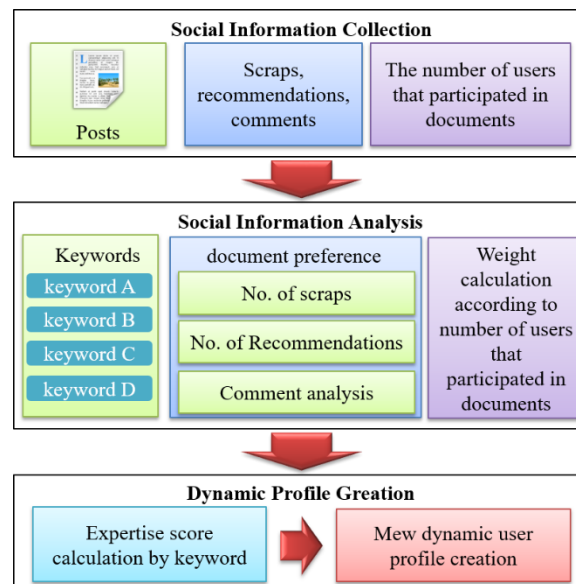


Figure 3. Social activity analysis process.

To generate a user's dynamic profile, the scheme collects documents $D = \{D_1, D_2, \dots, D_n\}$ created or shared over SNS during a recent time interval. In the social information analysis step, the keywords are extracted from each collected document D_i by performing text mining. The influence and weight of each document are also calculated. The keyword importance SE_i is calculated for the extracted keywords by considering the influence and weight. Equation (1) is the expression that is used to calculate the keyword importance SE_i , where W_i and S_i are the weight and influence of document D_i , k is the amount of activity, and b is a constant value. The amount of activity k is the number of documents created by the user. The greater the utilization, the more active the activity is determined as one user and the keyword specialty is increased. The larger the utilization, the greater the number of more active users identified and the more the keyword importance is increased.

$$SE_i = W_i S_i - \frac{b}{k} \quad (1)$$

The more documents are used and recommended by many users, the more popular and reliable documents can be determined. Therefore, the proposed scheme assigns a weight considering the number of users that participated in a corresponding document. Equation (2) calculates the weight W_i of document D_i , where N_i is the number of users that participated in document D_i , while a is a constant value. In Equation (2), if more than a users are interested in a document, the document is regarded as reliable and objective; that is, a high weight is assigned if a specific document is read by many users or creates many comments.

$$W_i = e^{\frac{N_i}{a}} - 1 \quad (2)$$

If users give positive comments to specific documents or make a lot of scraps and recommendations for a specific document, the document can be judged as a document that has a great influence on other users. The influence of document D_i is calculated by the number of scraps, the number of recommendations, and the number of comments. Equation (3) is an expression that calculates the influence of document D_i , where ε , NS_i , NR_i , and NC_i are a constant value, the number of scraps of D_i , the number of recommendations of D_i , and the comment score of D_i , respectively. By analyzing the

meaning of the comment, the comment score is given as 1, 0, or 0.5, depending on whether the meaning of the comment is positive, negative, or neutral. If S_i is larger than the constant ε , 1 is assigned.

$$S_i = \begin{cases} 1, & S_i \geq \varepsilon \\ \frac{NS_i + NC_i + NR_i}{N_i}, & 0 \leq S_i < \varepsilon \end{cases} \quad (3)$$

The keyword expertise K_{ij} of the keyword j extracted from document D_i is assigned the same SE_i assigned to D_i . In order to determine a user's expertise, the expertise score EK_j of keyword j is assigned as an average value of keyword expertise K_{ij} of the same keyword extracted from various documents created or shared by a user. A user profile is generated by using the expertise score EK_j of a keyword extracted from a document created or shared by a user. This profile is periodically updated using recent documents.

Figure 4 shows an example of how to generate a profile through keywords extracted from a document created by a user. Assuming that keywords A, B, C, D, E, F, and G are extracted from documents (Posts 1, 2, and 3) created by a user, a user's dynamic profile is created based on the calculation of a mean of SE_i that represents the expertise EK_j per keyword and the frequency. In the social information analysis step, a keyword is extracted from all documents created by a user, and a keyword importance SE_i per keyword is calculated to generate a dynamic profile. In the keyword sorting step, the same keywords as keywords E and G are found by sorting all keywords to calculate SE_i of the same keyword. In the dynamic profile generation step, an average expertise score of the same keyword is calculated. For example, EK_E of keyword E is 0.645, while EK_G of keyword G is 0.385. The presence of overlapped keywords means that a user is interested in the keyword and uses it frequently, so the dynamic user profile consists of keywords, its expertise score, and its frequency.

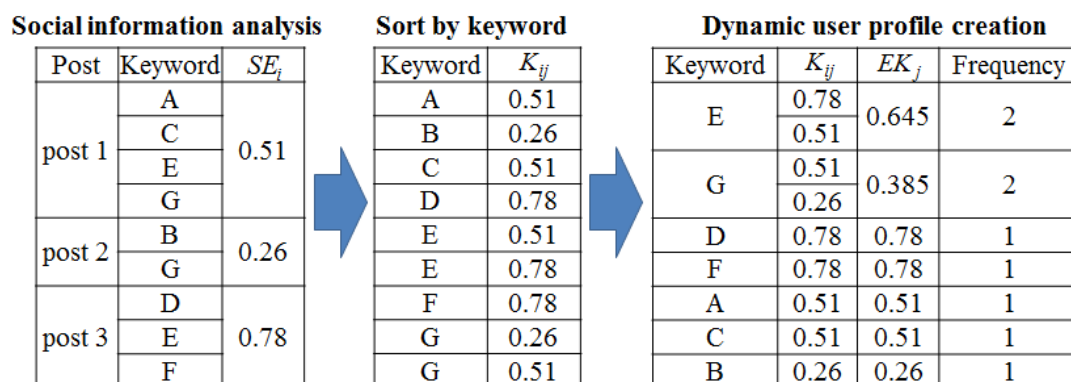


Figure 4. Dynamic user profile creation process.

3.3. User Trust

In SNS, the trust of users who provide information is very important because unreliable information is shared and distributed by malicious users. The existing schemes do not consider the expert trust because they simply analyze the user's profile or recent activities to determine the user's interest. Therefore, even if a user creates or shares a lot of wrong information in a specific field, the user can be selected as an expert. In the proposed scheme, even if the user with low trust performs a lot of activities related to a specific field, the user is not identified as an expert. For this purpose, we assume that each user can evaluate another user and use the evaluated score as a reputation for the user [44].

If user trust is evaluated using the reputation of all users in SNS, this requires many computation costs. Therefore, we use only the reputation of users in a human network path to calculate the user trust. Users give the reputation score to the user, providing the information in the process of sharing or delivering the information. This reputation score is used to calculate trust among social users using the weighted average of ratings [46]. Equation (4) calculates the trust between two users i and j , where

AT_{ij} is the sum of trust of users present in all human network paths located between i and j , and WAT_{ij} is a value that weighs the trust of users directly connected with i .

$$T_{ij} = \frac{WAT_{ij}}{AT_{ij}} \quad (4)$$

Figure 5 shows an example of the process of calculating the user trust using reputation. B and C are users who have first-degree friendships with User A, while D, E, and F have second- or third-degree friendships with User A. To calculate the trust between Users A and F, paths between B and C, which are connected directly with A and F, are used. Path 1 is A-B-F, Path 2 is A-B-D-F, Path 3 is A-C-D-F, and Path 4 is A-C-D-B-F. Therefore, the trust values of the paths are 0.6, 0.833, 0.733, and 0.525, respectively. Based on this, the trust between A and F considering all paths between them is AT_{AF} , which is 2.691, whereas WAT_{AF} is 1.9187 and T_{AF} is 0.713, by using Equation (4).

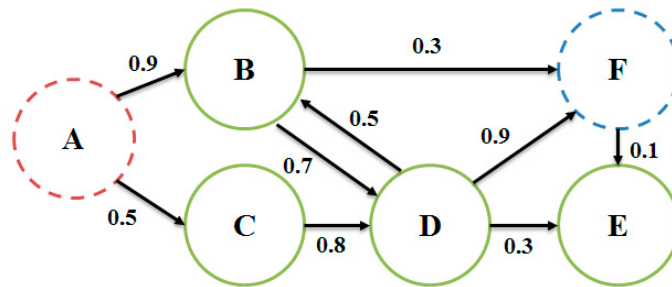


Figure 5. Trust network between users.

3.4. User Quality

Although users with a lot of domain knowledge have been selected as experts, they cannot respond to the user's questions or perform insincere responses. Therefore, experts should provide high-quality responses to users' questions in a short time. This scheme determines the user quality to improve the response rate to the user's question and the quality of the response. To determine the user quality, the response frequency, the response length, the rating for the response, and the response speed are considered. Equation (5) is an expression employed for calculating the user quality of the user j , where OP_{j1} , OP_{j2} , OP_{j3} , and OP_{j4} are the response ratio, the response length ratio, the average rating ratio, and the average speed ratio performed by the user j , respectively. OP_{j1} , OP_{j2} , OP_{j3} , and OP_{j4} are calculated by Equations (6)–(9). In Equation (6), n is the number of questions asked by users and RN_k indicates whether the user j has provided a response to the question, which is 1 if the user has responded to the question k , or 0 if the user has not. In Equations (7)–(9), m is the number of the responses performed by user j . $MAXL$, $MAXR$, and $MAXS$ are the maximum length of the response, the maximum rating of the response, and the maximum speed of the response, respectively. RL_r , RR_r , and RS_r are the user j 's response length, the user j 's response rating, and the user j 's response speed for answering the question, respectively.

$$UAS_j = \frac{OP_{j1} + OP_{j2} + OP_{j3} + OP_{j4}}{4} \quad (5)$$

$$OP_{j1} = \frac{\sum_{k=1}^n RN_k}{n} \quad (6)$$

$$OP_{j2} = \frac{\frac{1}{m} \sum_{r=1}^m RL_k}{MAXL} \quad (7)$$

$$OP_{j3} = \frac{\frac{1}{m} \sum_{r=1}^m RR_k}{MAXR} \quad (8)$$

$$OP_{j4} = \frac{\frac{1}{m} \sum_{r=1}^m RS_k}{MAXS} \quad (9)$$

Table 1 shows users' qualities according to the rating criteria. All users have their own mean value according to the rating criteria. Let us assume that the number of questions, the maximum response length, and the maximum response rating are 500 bytes, and 1.0, respectively, and if these values were exceeded, they would be regarded as the maximum values for the calculation. Since the response speed has to have a higher score as the time to respond decreases, up to a 10 h response time is valid in the calculation. User A has values of 0.7, 0.734, 0.7, and 0.8. Since OP_{A1} , OP_{A2} , OP_{A3} , and OP_{A4} are 0.7, 0.734, 0.7, and 0.8, respectively, UAS_A is 0.7335 using Equation (5). After the UAS_j values of five users were all calculated and compared, it was found that User F had the highest score, as shown in Table 1.

Table 1. Qualities of users.

Feature	User A	User B	User C	User D	User E
# of the response	21	10	7	18	6
Average length	367	560	892	533	471
Average rating	0.7	0.9	0.8	0.8	0.8
Average speed	2	4	7	3	2
Quality	0.7335	0.7083	0.5833	0.7750	0.6855

3.5. Expert Finding

If a user is found to be an expert by using only the keywords included in the user's question, the accuracy can be degraded because the user who can give a professional response cannot be found as an expert. The proposed scheme extracts the keywords included in the question and generates the query cluster including synonyms and dialects using ontology. The candidate experts are selected by comparing keywords included in the query cluster with dynamic profiles. Most users may not answer questions of users without human relations. Therefore, we identified the experts by comparing the user's dynamic profile according to the human network. Equation (10) is an expression used for calculating the similarity between user j and query cluster QC_i of request user j . The user with high similarity through Equation (10) is generated as the final expert finding result. In Equation (10), AER_{ij} denotes an average expertise score which is calculated through Equation (11), TT_{ij} denotes trust between two users i and j , and UAS_j denotes the user quality of the user j . In Equation (11), NC_i denotes the number of keywords that match the query cluster, QEK_k denotes the expertise score of keywords that match the query cluster, $MAXF$ denotes the maximum number of keywords generated, and F_k denotes the keyword frequency.

$$SIM_{ij} = \alpha AER_{ij} + \beta TT_{ij} + \gamma UAS_j \quad (10)$$

$$AER_{ij} = \frac{\sum_{k=1}^n QEK_k \frac{F_k}{MAXF}}{NC_i} \quad (11)$$

Figure 6 shows a process that calculates similarity by comparing a question and a user. The first process creates a cluster in response to a question once a user asks a question and starts a search of users from first-degree friendships. It is assumed that $\alpha = 0.5$, $\beta = 0.3$, and $\gamma = 0.2$, while the query cluster consists of keywords E and D, synonyms of EA and DA, and dialects of EB and DB, to calculate similarity. If the dynamic profile of User F is already created, as shown in Figure 4, keywords E and D are matched with each other while comparing the query cluster and the dynamic profile. A mean expertise score $AER_{ij} = 0.5175$ is calculated considering the number of matched keywords and frequency. The trust score in Figure 5 shows that $TT_{ij} = 0.713$, and User F's quality in Table 1 is $UAS_j = 0.6855$, which produces the similarity of User F, $SIM_{ij} = 0.6098$. The similarities of Users F, H, J, and K, who are matched with a query cluster, are calculated iteratively and outputted in order of high similarity.

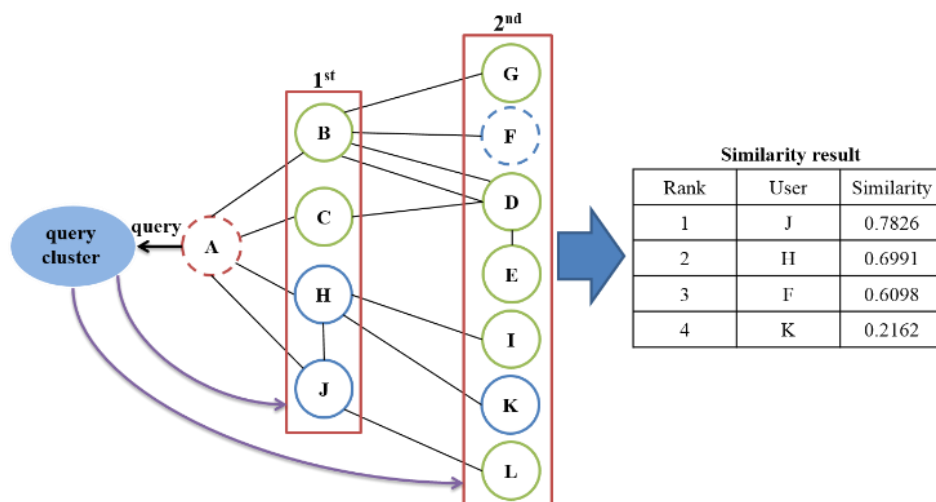


Figure 6. Calculated similarity results for a question and user profiles.

Figure 7 shows the pseudo code employed for the expert finding. The input variables are QC (query cluster), NC (number of matches between the query cluster and keywords), MaxF (maximum number of keywords), F (frequency of keyword), QEK (list of expert scores matched between query cluster and keywords), and UP (set of user profiles). The output variable SS is a set of similarity scores. Once a question is inputted, regardless of whether synonyms or dialects of the question words are found, an extended query cluster of the question is created. A query cluster and a server compare periodically created dynamic user profiles and determine expert ranks in order of high similarity score, which is calculated considering trust scores, and the output is produced.

Expert finding

Input variables:

- Query cluster $QC = \{query\ Q, query\ synonym\ QS, query\ dialect\ QD\}$
- Number of query clusters and keywords that match NC
- Max number of keywords MaxF
- Frequency of keyword F
- List of expert scores matching query clusters and keywords QEK
- Set of user profiles $UP = \{P_1, P_2, \dots, P_n\}$

Output variable:

- Set of similarity scores $SS = \{Sim_1, Sim_2, \dots, Sim_n\}$
-

/* search query extension and cluster creation */

if exist query synonym QS, QD

create $QC \leftarrow \{Q, QS, QD\}$

else if

create $QC \leftarrow Q$

/* similarity calculation */

while there exist UP satisfied QC

calculate similarity score Sim

create $SS \leftarrow similarity\ score\ Sim$

end while

Figure 7. Expert finding algorithm.

4. Performance Evaluation

In some studies, recommendation schemes using artificial neural networks have been conducted. However, most expert finding schemes have analyzed user activities or user-related documents to determine experts. Therefore, we compared the proposed scheme with the simple and cluster matching expert finding scheme proposed in [15], in terms of the finding time, accuracy of the expert finding result, Mean Reciprocal Rank (MRR), and expert finding results using weight, to show the superiority of the proposed scheme. The experimental data allows 100 users to generate up to 100 documents for one month, as shown in Table 2. However, the expert's profile uses the scheme proposed in [34]. AlchemyAPI [47] was used to extract keywords in the generated documents, while SAS Text Miner [48] was employed to analyze comments and the influences of the documents. The performance evaluation was implemented using JAVA to easily interoperate with the results of text mining on an Intel (R) Core (TM) i3-2100 CPU 3.10 GHz, RAM 4.00 GB, and Microsoft Window 7 64-bit operating system.

Table 2. Performance evaluation environment.

Parameter	Value
# of users (persons)	100~1000
# of documents left recently	0~100
# of keywords in a document	2
Trust	0~1
# of types of keywords	100
Response rating	0~1

Since a large number of users in social networks can use services at the same time, the processing speed of a server can be degraded. Therefore, it is necessary to determine how much more effective the proposed scheme is than the existing expert matching scheme in terms of the finding time. In this study, the time taken from the beginning of the finding request to search completion was measured according to the number of users and compared with the existing expert matching scheme to evaluate the performance of the proposed scheme. Figure 8 shows the time taken to generate a finding result when the number of users subscribed to social network changes. The simple matching scheme extracts keywords from the question and compares them with the user's interest connected to the human networks to determine the experts. Since the simple matching scheme only performs a simple keyword comparison, its expert finding time is very fast and even if the number of users increases, the expert finding time rarely increases. In contrast, the cluster matching scheme shows a rapid increase in time cost as the number of users increases, since it extends a cluster of questions and areas of interest in profiles whenever it receives a question. The proposed scheme only creates a cluster for a question while profiles are updated periodically in the server, which is why the proposed scheme achieves about a 43% decrease in finding time compared to the cluster scheme, on average, although the computation increases as the number of users increases.

In the experimental results of the existing expert finding scheme, there was no absolute criteria to rate accuracy. Therefore, in their experiment, accuracy was judged based on how finding results could result in high rankings by deliberately creating information on a particular user expertise. Accordingly, the experimental evaluation of this paper also compares and evaluates the accuracy of the proposed scheme with existing expert finding schemes by checking whether the user with the highest expertise score is placed in the highest rank. Table 3 shows a ranking of the expert finding result. To ensure that the expert is accurately found, the user's dynamic profile was composed of keywords in the query cluster and the expertise score was set to the highest score. The number of SNS users was set to 200. The proposed scheme and query cluster extended the range of the expert finding through ontology. Therefore, we can see that user 29 was accurately ranked as first in the ranking. On the other hand, since the simple matching technique is only based on the keywords in the profile, the user is ranked 11th, showing a low accuracy.

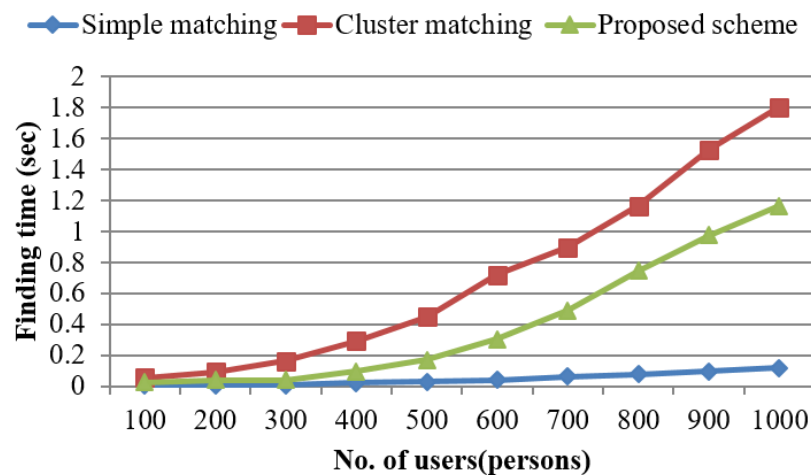


Figure 8. Finding time.

Table 3. Expert finding results.

Rank	Matching Scheme		Cluster Scheme		Proposed Scheme	
	User	Score	User	Score	User	Score
1	90	6	29	0.9375	29	0.9728
2	170	6	43	0.7500	42	0.9397
3	176	6	95	0.7500	185	0.8550
4	39	5	198	0.7500	187	0.7131
5	3	5	1	0.5000	24	0.6987
6	102	5	191	0.4938	40	0.6612
7	105	5	130	0.4938	172	0.6319
8	133	5	86	0.4444	4	0.6065
9	168	5	162	0.3333	77	0.6002
10	13	5	75	0.3320	102	0.5843
11	29	5	37	0.2500	70	0.5768
12	6	4	149	0.2496	110	0.5723
13	36	4	58	0.2480	176	0.5375
14	103	4	192	0.2480	39	0.5362
15	19	4	89	0.2480	175	0.5264

The expert finding result above lacked diversity and objectivity. Accordingly, the accuracy of the finding results was measured by scores according to differences in the comparison of ranks of users who satisfied specific criteria found by finding results and expected ranks through the MRR evaluation method. Figure 9 shows the accuracy of the finding result obtained through the MRR evaluation. In this study, three criteria were set using the MRR evaluation method, and scores were measured according to the difference between ranks of users who satisfied the criteria and corresponding expected ranks. According to the three criteria (a: the number of matched keywords, b: trust, and c: response quality), MRRs of 100 users were measured by assuming that eight users had scores between 0 and 1, the highest and lowest scores, and 92 users followed a normal distribution. The existing schemes had low MRRs, except for “000” and “111,” whereas the proposed scheme had high MRRs of eight users, which indicated accurate finding results.

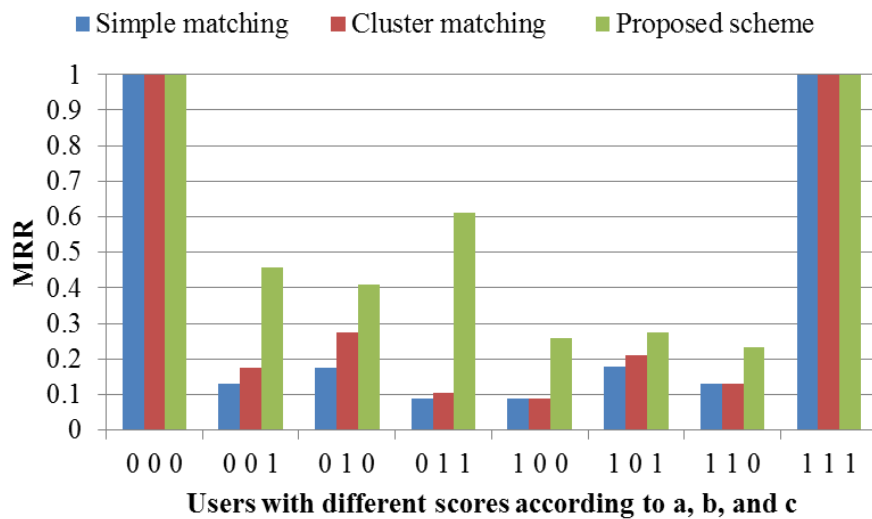


Figure 9. Mean Reciprocal Rank (MRR) evaluation results.

In the proposed scheme, users may have different preferences with respect to expertise, trust, and response qualities. Every weight is distributed appropriately to satisfy all users. However, to produce experimental evaluations according to the weight distribution, a large number of experiments are required from various viewpoints, as well as an analysis of the experimental results, which is why we set this aside as future research. Performance results do not depend on computer parameters. The proposed scheme uses α , β , and γ weights to determine the final expert using Equation (10), so the performance results may vary, depending α , β , and γ . Therefore, we instead discuss how the finding results differ according to the weight distribution and the issues of this method. Figures 10–12 show how the expert finding results differed according to the weight distribution of α , β , and γ in Equation (10). The 100 users previously used in the MRR measurement were assumed again, and the finding results according to the weight distribution were measured. According to the three criteria (a: the number of matched keywords, b: trust, and c: response quality), how often users with scores between the highest and lowest scores (0, 1) appeared in the finding result on average was measured, and their probabilities were calculated.

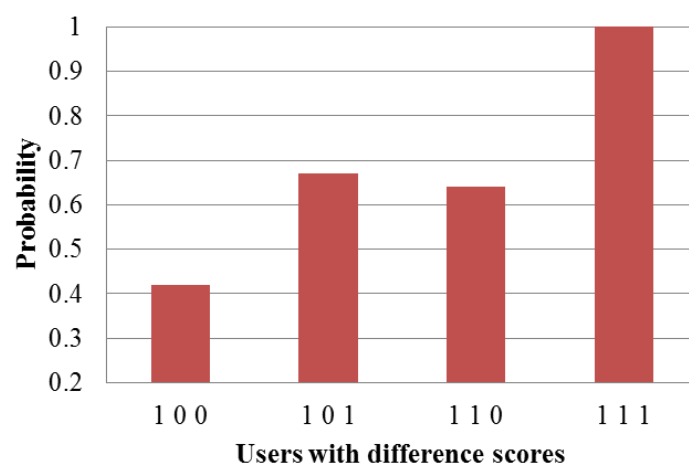


Figure 10. Finding results when $\alpha = 0.8$, $\beta = 0.1$, and $\gamma = 0.1$.

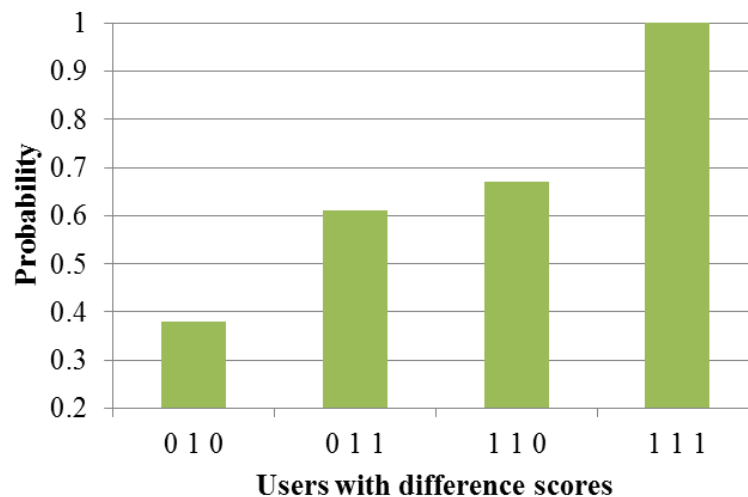


Figure 11. Search results when $\alpha = 0.1$, $\beta = 0.8$, and $\gamma = 0.1$.

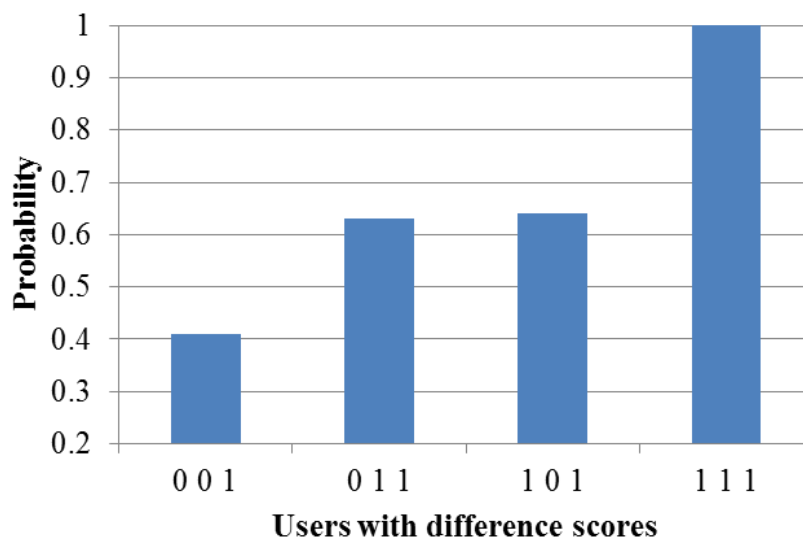


Figure 12. Finding results when $\alpha = 0.1$, $\beta = 0.1$, and $\gamma = 0.8$.

Figure 10 shows that a user with high expertise due to the highest weight $\alpha = 0.8$ had a high probability of being included in the finding result. However, this figure also showed that users with low trust and qualities could be included in the finding results. Therefore, it was also verified that trust and response qualities must be considered with high weight.

Figure 11 shows that a user with high trust due to the highest weight $\beta = 0.8$ had a high probability of being included in the finding result. However, this figure also showed that users with low expertise and qualities could be included in the finding results. Therefore, it was also verified that expertise and response qualities must be considered with high weight.

Figure 12 shows that a user with a high response quality due to the highest weight $\gamma = 0.8$ had a high probability of being included in the finding result. However, this figure also showed that users with low expertise and trust could be included in the finding results. Therefore, it was also verified that expertise and trust must be considered with high weight. Such expert finding results verified that finding results can differ, depending on how weights related to expertise, trust, and response quality are distributed in Equation (10).

5. Conclusions

In this paper, we have proposed an expert finding scheme that can provide high-quality responses to users' questions. The proposed scheme analyzes the user's recent social activities to generate a dynamic profile and improves the accuracy of expert finding using query clustering. In order to ensure the reliability of expert responses, user trust and user quality were considered. The experimental results showed that the average time cost in SNS was reduced by about 43% compared to the existing cluster scheme. The results of the search showed that the experts have a high ranking, and the accuracy of the MRR measurements to ensure the objectivity of the search results was also confirmed. For future research, we will consider the human network characteristics of experts to improve the performance of expert finding. In addition, we will apply the proposed scheme to the actual system.

Author Contributions: Conceptualization, K.B., I.J., J.L., and J.Y.; methodology K.B., I.J., and J.Y.; software, I.J.; validation, K.B. and I.J.; formal analysis, K.B., I.J., and J.L.; data curation, I.J.; writing—original draft preparation, K.B. and I.J.; writing—review and editing, J.Y.

Funding: This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. B0101-15-0266, Development of High Performance Visual BigData Discovery Platform for Large-Scale Realtime Data Analysis); by the Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT (No. NRF-2017M3C4A7069432); and by "Rediscovery of the Past R&D Result" through the Ministry of Trade, Industry and Energy (MOTIE) and the Korea Institute for Advancement of Technology (KIAT) (Grant No.: P0010202, Project: SNS Big data-based personalized employment content recommendation(curation) and digital certificate issuing system).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Han, X.; Wang, L.; Crespi, N.; Park, S.; Cuevas, Á. Alike people, alike interests? Inferring interest similarity in online social networks. *Decis. Support Syst.* **2015**, *69*, 92–106. [\[CrossRef\]](#)
2. Dong, T.; Cheng, N.; Wu, Y.J. A study of the social networking website service in digital content industries: The Facebook case in Taiwan. *Comput. Hum. Behav.* **2014**, *30*, 708–714. [\[CrossRef\]](#)
3. Kwak, H.; Lee, C.; Park, H.; Moon, S. What is Twitter, a social network or a news media? In Proceedings of the International conference on World Wide Web, Raleigh, NC, USA, 26–30 April 2010.
4. Spirin, N.V.; He, J.; Develin, M.; Karahalios, K.G.; Boucher, M. People Search within an Online Social Network: Large Scale Analysis of Facebook Graph Search Query Logs. In Proceedings of the International Conference on Conference on Information and Knowledge Management, Shanghai, China, 3–7 November 2014.
5. Zheng, Y.; Xie, X.; Ma, W. GeoLife: A Collaborative Social Networking Service among User, Location and Trajectory. *IEEE Data Eng. Bull.* **2010**, *33*, 32–39.
6. Sánchez, L.Q.; Díaz-Agudo, B.; Recio-García, J.A. Development of a group recommender application in a Social Network. *Knowl.-Based Syst.* **2014**, *71*, 72–85. [\[CrossRef\]](#)
7. Liang, B.; Liu, Y.; Zhang, M.; Ma, S.; Ru, L.; Zhang, K. Searching for people to follow in social networks. *Expert Syst. Appl.* **2014**, *41*, 7455–7465. [\[CrossRef\]](#)
8. Omidvar, A.; Garakani, M.; Safarpour, H.R. Context based user ranking in forums for expert finding using WordNet dictionary and social network analysis. *Inf. Technol. Manag.* **2014**, *15*, 51–63. [\[CrossRef\]](#)
9. Amato, F.; Castiglione, A.; Mercurio, F.; Mezzanzanica, M.; Moscato, V.; Picariello, A.; Sperli, G. Multimedia story creation on social networks. *Future Gener. Comput. Syst.* **2018**, *86*, 412–420. [\[CrossRef\]](#)
10. Heimbach, I.; Hinz, O. The Impact of Sharing Mechanism Design on Content Sharing in Online Social Networks. *Inf. Syst. Res.* **2018**, *29*, 592–611. [\[CrossRef\]](#)
11. Guidi, B.; Conti, M.; Passarella, A.; Ricci, L. Managing social contents in Decentralized Online Social Networks: A survey. *Online Soc. Netw. Media* **2018**, *7*, 12–29. [\[CrossRef\]](#)
12. Fang, G.; Su, L.; Jiang, D.; Wu, L. Group Recommendation Systems Based on External Social-Trust Networks. *Wirel. Commun. Mob. Comput.* **2018**, *2018*, 1–11. [\[CrossRef\]](#)
13. Campana, M.G.; Delmastro, F. Recommender Systems for Online and Mobile Social Networks: A survey. *Online Soc. Netw. Media* **2017**, *3–4*, 75–97. [\[CrossRef\]](#)

14. Li, M.; Wang, X.; Gao, K.; Zhang, S. A Survey on Information Diffusion in Online Social Networks: Models and Methods. *Information* **2017**, *8*, 118.
15. Duchateau, F. Who can Best Answer a Query in My Social Network? In Proceedings of the International Conference on Data Engineering Workshops, Hannover, Germany, 11–16 April 2011.
16. Lin, C.; Xie, R.; Guan, X.; Li, L.; Li, T. Personalized news recommendation via implicit social experts. *Inf. Sci.* **2014**, *254*, 1–18. [[CrossRef](#)]
17. Yin, Z.; Wang, X.; Wu, X.; Liang, C.; Xu, C. Collaborative Ranking via Learning Social Experts. In Proceedings of the International Conference on Tools with Artificial Intelligence, Limassol, Cyprus, 10–12 November 2014.
18. Ehrlich, K.; Shami, N.S. Searching for expertise. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Florence, Italy, 5–10 April 2008.
19. Zhang, J.; Tang, J.; Li, J. Expert Finding in a Social Network. In Proceedings of the International Conference on Database Systems for Advanced Applications, Bangkok, Thailand, 9–12 April 2007.
20. Joung, Y.; Chen, S.M.; Wu, C.; Chiu, T.H. A Comparative Study of Expert Search Strategies in Online Social Networks. In Proceedings of the International Conference on Advanced Information Networking and Applications, Barcelona, Spain, 25–28 March 2013.
21. Uddin, M.U.; Duong, T.H.; Oh, K.; Jung, J.; Jo, G. Experts Search and Rank with Social Network: An Ontology-Based Approach. *Int. J. Softw. Eng. Knowl. Eng.* **2013**, *23*, 31–50. [[CrossRef](#)]
22. Awuor, F.M.; Wang, C.; Tsai, T. Motivating Content Sharing and Trustworthiness in Mobile Social Networks. *IEEE Access* **2018**, *6*, 28339–28355. [[CrossRef](#)]
23. Jiménez-Díaz, G.; Gómez-Martín, P.P.; Gómez-Martín, M.A.; Sánchez-Ruiz, A.A. Similarity metrics from social network analysis for content recommender systems. *AI Commun.* **2017**, *30*, 223–234. [[CrossRef](#)]
24. Hristoskova, A.; Tsiporkova, E.; Tourw, T.; Buelens, S.; Putman, M.; Turck, F.D. Identifying Experts through a Framework for Knowledge Extraction from Public Online Sources. In Proceedings of the Dutch-Belgian Information Retrieval Workshop, Ghent, Belgium, 23–24 February 2012.
25. Bitton, Y.; Fire, M.; Kagan, D.; Shapira, B.; Rokach, L.; Bar-Ilan, J. Social Network Based Search for Experts. *arXiv* **2012**, arXiv:1212.3540.
26. Vu, T.; Baid, A. Ask, don't search: A social help engine for online social network mobile users. In Proceedings of the IEEE Sarnoff Symposium, Newark, NJ, USA, 21–22 May 2012.
27. Stankovic, M.; Wagner, C.; Jovanovic, J.; Laublet, P. Looking for experts? What can linked data do for you? In Proceedings of the WWW Workshop on Linked Data on the Web, Raleigh, NC, USA, 27 April 2010.
28. Liu, D.; Wang, L.; Zheng, J.; Ning, K.; Zhang, L. Influence Analysis Based Expert Finding Model and Its Applications in Enterprise Social Network. In Proceedings of the International Conference on Services Computing, Santa Clara, CA, USA, 28 June–3 July 2013.
29. Jeng, W.; DesAutels, S.; He, D.; Li, L. Information exchange on an academic social networking site: A multidiscipline comparison on researchgate Q&A. *J. Assoc. Inf. Sci. Technol.* **2017**, *68*, 638–652.
30. Amjad, T.; Daud, A.; Aljohani, N.R. Ranking authors in academic social networks: A survey. *Libr. Hi Tech* **2018**, *36*, 97–128. [[CrossRef](#)]
31. Fu, Y.; Xiang, R.; Liu, Y.; Zhang, M.; Ma, S. Finding experts using social network analysis. In Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, Silicon Valley, CA, USA, 2–5 November 2007.
32. Balog, K.; Azzopardi, L.; Rijke, M.D. Formal models for expert finding in enterprise corpora. In Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Seattle, WA, USA, 6–11 August 2006.
33. Horowitz, D.; Kamvar, S.D. The Anatomy of a Large-Scale Social Search Engine. In Proceedings of the International Conference on World Wide Web, Raleigh, NC, USA, 26–30 April 2010.
34. Bozzon, A.; Brambilla, M.; Ceri, S.; Silvestri, M.; Vesci, G. Choosing the Right Crowd: Expert Finding in Social Networks. In Proceedings of the International Conference on Extending Database Technology, Genoa, Italy, 18–22 March 2013.
35. Alkouz, A.; Luca, E.W.D.; Albayrak, S. Latent Semantic Social Graph Model for Expert Discovery in Facebook. In Proceedings of the International Conference on Innovative Internet Community Services, Deutsche Telekom Laboratories, Berlin, Germany, 15–17 June 2011.
36. Das, R.; Kamruzzaman, J.; Karmakar, G.C. Modelling majority and expert influences on opinion formation in online social networks. *World Wide Web* **2018**, *21*, 663–685. [[CrossRef](#)]

37. Zarrinkalam, F.; Kahani, M.; Bagheri, E. Mining user interests over active topics on social networks. *Inf. Process. Manag.* **2018**, *54*, 339–357. [[CrossRef](#)]
38. Al-garadi, M.A.; Varathan, K.D.; Ravana, S.D.; Ahmed, E.; Shaikh, G.M.; Khan, M.U.S.; Khan, S.U. Analysis of Online Social Network Connections for Identification of Influential Users: Survey and Open Research Issues. *ACM Comput. Surv.* **2018**, *51*, 1–37. [[CrossRef](#)]
39. Tareaf, R.B.; Berger, P.; Hennig, P.; Meinel, C. Malicious Behaviour Identification in Online Social Networks. In Proceedings of the International Conference on Distributed Applications and Interoperable Systems, Madrid, Spain, 18–21 June 2018.
40. Abbas, F.; Rajput, U.; Oh, H. PRISM: PRivacy-Aware Interest Sharing and Matching in Mobile Social Networks. *IEEE Access* **2016**, *4*, 2594–2603. [[CrossRef](#)]
41. Li, X.; Yang, Q.; Lin, X.; Wu, S.; Wittie, M.P. iTrust: Interpersonal trust measurements from social interactions. *IEEE Netw.* **2016**, *30*, 54–58. [[CrossRef](#)]
42. Chang, S.E.; Liu, A.Y.; Shen, W. User trust in social networking services: A comparison of Facebook and LinkedIn. *Comput. Hum. Behav.* **2017**, *69*, 207–217. [[CrossRef](#)]
43. Tian, H.; Liang, P. Improved Recommendations Based on Trust Relationships in Social Networks. *Future Internet* **2017**, *9*, 9. [[CrossRef](#)]
44. Bok, K.; Yun, J.; Kim, Y.; Lim, J.; Yoo, J. User Reputation computation Method Based on Implicit Ratings on Social Media. *KSII Trans. Internet Inf. Syst.* **2017**, *11*, 1570–1594.
45. Bok, K.; Lim, J.; Jeon, I.; Yoo, J. An Expert Search Scheme Using User Activities and Reliabilities in Social Networks. In Proceedings of the International Conference on Big Data and Smart Computing, Jeju, Korea, 9–12 February 2015.
46. Qureshi, B.; Min, G.; Kouvatsos, D. A Distributed Reputation and Trust Management Scheme for Mobile Peer-to-Peer Networks. *Comput. Commun.* **2012**, *35*, 608–618. [[CrossRef](#)]
47. AlchemyAPI. Available online: <https://github.com/AlchemyAPI> (accessed on 21 August 2017).
48. Yu, L.; Zheng, J.; Wu, B.; Wang, B.; Shen, C.; Qian, L.; Zhang, R. BC-PDM: Data mining, social network analysis and text mining system based on cloud computing. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Beijing, China, 12–16 August 2012.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).