

Expertization Level Ranking for Query Transformation

A. Abdul Azeez Khan* and P. Sheik Abdul Khader

Department of Computer Applications, B. S. Abdur Rahman University, Chennai - 600048, Tamil Nadu, India; abdulazeekhan@bsauniv.ac.in, psakhader@bsauniv.ac.in

Abstract

Objectives: To develop an approach to find one's expertization level in a given field. **Methods/Statistical Analysis:** The search engines were utilized to extract the expert's data available in the internet. The results generated by the search engines were downloaded in the database. Intern these results were used as an input to the relevance and expertise mapping process. We have used Relevance Algorithm and Expertise Mapping Algorithm to process the data generated by the search engines. These process yields one's expertization level on a particular area. **Findings:** E-learning systems were automated to identify an expert by using two methods self-classification and document-based-relevance. These methods assume that relevance of one's specified keywords or documents to the query is positively related to their expertise. In reality, one can be identified as an expert in a specific domain by the contribution made on that particular domain. The expert expertization data available in websites could be utilized by the e-learning systems to evaluate the expert's. Hence it is proposed to ensure the expertise level of an expert using a Dynamic Expertization Estimating System (DEES) in an e-learning environment. In this approach, search engines were utilized as an agent to extract the expert's data available in the internet and weightage were given to the expert according to their contribution made by the expert towards the given expertise area. The results retrieved by applying this mechanism yielded data with high accuracy levels to ensure the expertization level of an expert. **Application/Improvements:** Connecting the Dynamic Expertization Estimating System (DEES) with social media to fetch more data pertaining to the expert expertization and produce accurate expertise level for each expert.

Keywords: E-learning System, Expert Finding, Knowledge Management, Knowledge Sharing, Knowledge Capture, Tacit Knowledge

1. Introduction

Recent years have witnessed the enormous growth of e-learning system in the education and corporate sectors. The advancement in e-learning system has begun to reshape the learning approaches from traditional learning methodology to smart learning solutions. The e-learning system plays a significant role on knowledge transformations, which includes interchanging information, opinions, experiences and perceptions. The success rate of an e-learning system can be determined by the factors: such as expert's expertization, quality of content, experts and learner connectivity, adaptability and ease of use. A major portion of these factors can be

achieved by any e-learning system on fine-tuning the expert finding approaches.

Existing expert finding techniques often rely on the following indicators to find one's expertise area and the level of expertization: Self-classification and document-based relevance. Self-Classification basically includes the expert's expertization or skill sets manually entered by the expert and document-based relevance is evaluated based on relevant document available in the system. Most automated expert finding techniques rely on document-based relevance to measure the expertise level of an expert to assign a user query. This technique assumes that the relevance of one's authored documents to the query is positively related to their expertise level on the query¹.

* Author for correspondence

These perceptions clearly depict that, the expert's expertise levels were only measured with the data available in the system. Therefore the existing expert finding approaches cannot be considered as an accurate evaluation on expert's expertization. Since the existing expert finding techniques doesn't use the internet as a potential factor to find the updated information about the expert's expertization.

To find the expertization level of an expert using internet, it is required to have the expert's expertise keywords along with the expert identity data such as expert name, co-authors, email id etc. The expert given expertise keywords along with expert identity data have to be cross verified with the data available in the internet to ensure the contribution made by the expert on the given expertise area, the verification has to be made in various resources in the internet such as websites, blogs and journals etc. But most of the websites and journals require authenticity to access the resources available in their database. These constraints create a barrier to the researchers on utilizing the internet as an important data source to evaluate an expert's expertise level.

To overcome from this problem, we propose an approach to ensure the expertise level of an expert by utilizing the search engines as an agent to extract the expert's data available in the internet and weight according to their contribution made towards the given expertise area. The search engines are used as a data retrieving agents, since they can access Meta data or data of any websites without any authenticity of the accessing website. The expert's keywords have to be searched along with the expert name to extract the data of the expert related to the given keyword. This action will provide a bulk of results linked to the expert contribution towards the expertise keywords. The results produced by the search engines would contain many junk data. So the entire result of the search engine cannot be considered as a valuable data input to measure expert's expertise level. Hence it is required to remove the junk data and extract the records related to the expert and the given keyword. In addition, these results have to be processed with constraint parameters to find the exact record of an expert from the extracted records. The proposed data extraction technique gives a valuable input data with reasonable accuracy to measure the expertise level of an expert towards the given expertise keyword. This method can be considered as the best solution to extract the updated expertization level of an expert towards the expertise area.

Expert finding is comparatively a new area of research with ample scope. In recent times, a large number of researchers have put their significant efforts in this area. The automatic search for knowledgeable people in the scope of specific user communities, with basis on documents describing people's activities, is an information retrieval problem that has been getting increasing attention. Usually referred to as expert finding, the task involves taking a short user query as input, denoting a topic of expertise, and returning a list of people sorted by their level of expertise in what concerns the query topic. Several effective approaches for finding experts have been proposed, exploring different retrieval models and different sources of evidence for estimating expertise. However, the current state-of-the-art is still lacking in principled approaches for combining the multiple sources of evidence that can be used to estimate expertise².

The task of expert finding is intended to identifying people with relevant expertise skill or experience for a given topic. The topic of finding subject experts for problem solving is an important issue in an e-learning environment. In an e-learning environment, there is no direct approach to finding the superior individuals. The current methods of analyzing the discussions or considering the learner require a large amount of data and contain limitations. In this work, concept maps are used to define the experts in an e-learning environment³.

Authors in^{4,5} have surveyed the most important concepts and representative previous works in the expert finding. The two of the most popular and well-performing types of methods are the profile-centric and the document-centric approaches^{6,7}. It is emphasized that profile-centric approaches build an expert profile as a pseudo document, by aggregating text segments relevant to the expert⁸. These profiles are later indexed and used to support the search for experts on a topic. Document-centric approaches are typically based on traditional document retrieval techniques, using the documents directly.

The self-classification and document-based relevance are widely used method to find the experts. Self-disclosure requires expert candidates to explicitly declare their expertise in their posted profiles. Expert recommendation systems that employ this approach include yellowpages.com, guru.com, 88 owls.com, and other opt-in directory listings of experts. The manual process is time-consuming and expertise profiles are also unlikely to remain current as each user's expertise is continuously expanding⁹.

The difficulty with this approach was that it required manual entry to keep the expert profile updated. Studies have revealed that even if efforts are frequently put into updating the expert profile, they do not provide efficient and exact outcome. The reasons for this insufficiency are that the people who are entrusted the assignment of maintaining the expert profile do not have enough knowledge of the experts. Also, manual information entry of every individual in an e-learning system was extremely laborious and expensive. Thus, the manual entry of expert profiles into an e-learning system did not prove to be a feasible strategy.

Documents that were written or reviewed by an expert candidate are also a useful expertise indicator. There are a number of expert finding systems that use text mining techniques to automatically capture the authors' expertise. E-learning suggests that efficient e-learning system should make accessible not only content but also authors who are the foundation of the provided content. The major component of the learning process is the result of the learner's collaborations and communications with subject experts. The collaborations in online group-based learning provide better opportunities to develop skills and knowledge compared with individual courses. The hybrid approach was carried to effectively find experts for the category of the target question in question-answering websites. Different from the conventional approaches that only consider user profile or user authority, but their study considered user subject relevance, user reputation and authority of a category in finding experts. In review with the previous studies, it is been identified that there is no solid approach has been carried out to connect the expert finding system with internet resources, were the system can verify expert expertise skill set entered manually by the experts and update the expert profile automatically on expansion of the expert skill sets^{10,11}. The authorities and staff of the Education should pay special attention to better and more effective use of ICT to enjoy more of the positive and useful applications of these tools in educating and solving their problems¹². In the last century to the present, the matter of management of intangible assets as part of the vital resources is considered. Among these assets, knowledge is a most important among other intangible assets¹³. The people all over the world are in need of gaining knowledge which is achieved by means of searching information through web services. Even though search results obtained from web services are relevant, the user is trying to obtain the best information providers¹⁴.

2. Methodology

The tacit knowledge of an expert can be externalized in an e-learning environment by using the approach of capturing and sharing of knowledge with the right person at the right time. Currently e-learning systems were automated to identify an expert in a specific domain by using two methods 1. Self classification and 2. Document based relevance. These two methods assume that the relevance of one's specified keywords or authored documents to the query is positively related to their expertise level on the query. In reality, one can be identified as an expert in a specific domain by the contribution made by the expert on that particular domain. The expert expertization data available in various websites could be utilized by the e-learning systems to evaluate the expert's expertization level in addition to self classification and document based relevance.

The problem identified in this process is that, the existing expert finding techniques doesn't have the provision to update the data of an expert using the internet resources. As the expert expertization gets updated day by day, it is required to every expert finding technique to keep track on expert data available in the internet. The expert given expertise keywords has to be verified with the data available in the internet to ensure the contribution made by the expert on the given expertise area, the verification has to be made in various resources in the internet such as websites, blogs and journals etc. But most of the websites and journals require authenticity to access the resources available in their database. These constraints create a barrier to the researchers on utilizing the internet as an important factor to evaluate the expert's expertise level with the updated information available in various websites.

In this paper, we propose an approach to ensure the expertise level of an expert using a Dynamic Expertization Estimating System (DEES) in an e-learning environment. In this approach, search engines were utilized as an agent to extract the expert's data available in the internet and weightage were given to the expert according to their contribution made by the expert towards the given expertise area. The search engines are used as a data retrieving agents, since they can access Meta data or data of any websites without any accessing authenticity of the website. The expert's keywords have to be searched along with the expert name to extract the data of the expert related to the given keyword. This action will provide a

bulk of results linked to the expert contribution towards the expertise keywords. The results produced by the search engines would contain many junk data. So the entire result of the search engine cannot be considered as a valuable data input to measure expert’s expertise level. Hence it is required to remove the junk data and extract the records related to the expert and given keyword. In addition these results have to be processed by having constraint parameters to find the exact record of the expert from the extracted result. The proposed data extraction technique gives a valuable input data with reasonable accuracy to measure the expertise level of an expert towards the given expertise keyword.

The diagram in Figure 1 shows the framework of the dynamic expertization estimating system (DEES) in an e-learning environment to automatically estimate the updated expertization skill level of an expert through internet resources. This mechanism uses a single entry point into the system with pre-connected expert supporting parameters and expert’s expertise keywords list.

2.1 Expert Detail Segment

In this system, the expert detail segment act as a single point of input section to inflow the expert details such as expert Name, expert supporting parameters list and expertise keywords list. The expert name is the first and initial content get entered into the system along with

supporting parameter list and expertise keyword list. The supporting parameter list contains the primary detail of the expert such as

- Expert Name (EN).
- Expert Current Designation (ECD).
- Expert Previous Designation (EPD).
- Current Email ID of the expert (CEID).
- Current Employer (CE).
- Previous Employer (PE).
- Co-Authors (COA).
- Current Organization Website (CWEB).
- Previous Organization Website (PWEB).

The supporting parameter list is used by the dynamic expertization estimating system for the identification of expert record. The expertization keyword list is a segment of data pertaining to the expert expertization on a specific domain.

These combinations of data were supplied as an input to the search engine in the format of expert name combined with one of the expertization keyword (i.e., Expert Name + Keyword). The result yielded by the search engine would be downloaded in the database. The same process will be continued to all the expertization keywords present in the expert’s expertization list to retrieve the related data of the expert available in various websites, blogs etc. The results received from the search engine are not much accurate to define one’s contribution

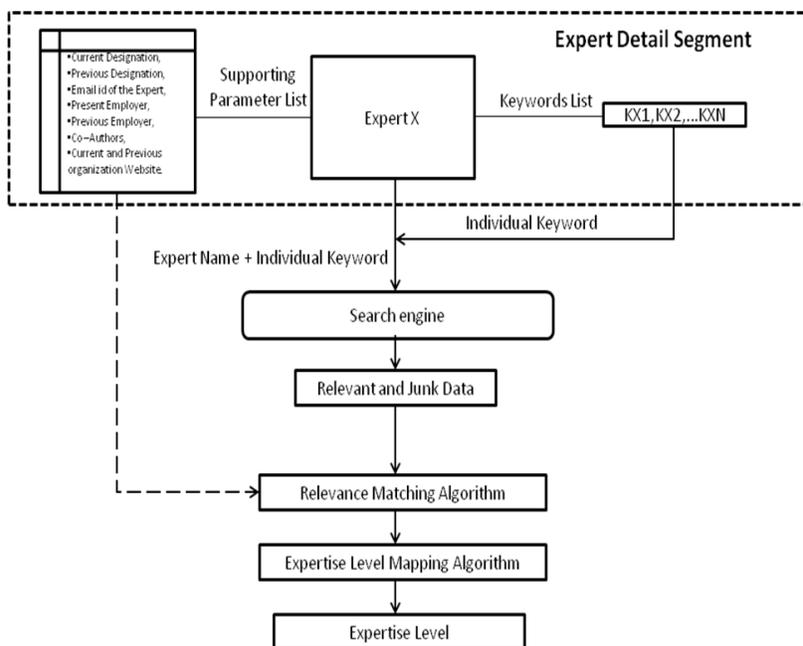


Figure 1 Dynamic expertization estimating system.

towards the specialized field or keyword. Since the result may have many junk data not related to the searched expert or keyword. Hence, it is required to process the search engine results for extracting the actual records of the expert from the bulk extracted data. This could be made possible on processing the extracted data with expert's parameter list, such as comparing each and every output record with the expert parameters list. The highest match could be considered as the actual record of the expert from the multiple records. The same process has to be continued to retrieve the actual records of the expert and keyword.

3. Implementation

The proposed mechanism has been tested with a semi prototype implementation of the Dynamic Expertization Estimating System (DEES) (Figure 1). The partial implementations of the system were designed cautiously to use the information available in the Internet on identifying the expertization level of an expert. The prototype implementation of the Dynamic Expertization Estimating System (DEES) (Figure 1.) is distributed into three combinations of processes:

- Expertise Detail Segment.
- Search Agent Section.
- Relevance and Expertise Mapping process.

The Expertise Detail Segment is a feeding process used by the Dynamic Expertization Estimating System (DEES) to collect the information from the expert and store it in the database. The feeding information such as Experts Expertization keywords (Knowledge Engineering, Networking, Ontology etc) and Supporting Parameter List (Expert Name (EN), Expert Current Designation (ECD), Email id of the expert (EID), Expert Previous Designation (EPD), Current Employer (CE), Previous Employer (PE), Co-Authors (COA), Current Organization Website (CWEB), Previous Organization Website (PWEB)). These parameters were highly utilized in the system to cross verify an expert's expertization level or contribution on a given field or area.

The search agent section is a data fetching process, which is used to retrieve the data of an expert available in the internet. We have used Google search engine as an agent to retrieve the data of an expert from the internet sources. The expert names along with specialized keywords were used together to fetch the relevant data

of the expert from the internet. The expert name and specialized keywords were supplied in the below mention format.

```
Search 1:      <Expert Name>, < Expert Keyword 1>
Search 2:      <Expert Name>, < Expert Keyword 2>
...
...
...
Search n:      <Expert Name>, < Expert Keyword n>
```

The results generated by the search engines were downloaded in the database. Intern these results were used as an input to the relevance and expertise mapping process. We have used Relevance Algorithm and Expertise Mapping Algorithm to process the data generated by the search engines. These process yields one's expertization level on a particular area or field.

3.1 Relevance Matching Algorithm

The algorithm generated for the expertization relevance matching is described below:

Step 1: Expert Name and the specialized keywords are supplied as an input to the search engine.

(Expert Name + Keyword → Search Engine).

Step 2: The results are stored in the database. (Search Engine output → Database).

Step 3: Each record is process with NLP parser.

Step 3: The database records are compared with expert supporting parameter list.

(Expert Name (EN), Expert Current Designation (ECD), Email ID of the expert (EID), Expert Previous Designation (EPD), Current Employer (CE), Previous Employer (PE), Co-Authors (COA), Current Organization Website (CWEB), Previous Organization Website (PWEB))

Step 4: The Record (R) belongs to Expert (Ex).

n = number of records;

I = 1.....n;

For (R=1; I<=n; I++)

{

IF (R = EN AND R= EID) {

R∈ EX (R belong to Expert);

} ELSEIF (R=EN AND R=CWEB) {

R∈ EX (R belong to Expert);

} ELSEIF (R=EN AND R=COA) {

R∈ EX (R belong to Expert);

} ELSEIF (R=EN AND R=ECD) {

R∈ EX (R belong to Expert);

```

} ELSEIF (R=EN AND R=EPD) {
R ∈ EX (R belong to Expert);
} ELSEIF (R=EN AND R=CE) {
R ∈ EX (R belong to Expert);
} ELSEIF (R=EN AND R=PF) {
R ∈ EX (R belong to Expert);
} ELSEIF (R=EN AND R=PWEB) {
R ∈ EX (R belong to Expert);
} ELSEIF (R=EID) {
R ∈ EX (R belong to Expert);
} ELSE {
R ∉ EX (R Not belong to Expert);
}
}

```

The algorithm generated for relevance matching indicates that the Expert Name and Expertise keywords are supplied as an input to the search engine. The results retrieved from the search engine are processed with NLP parser. The cleaned data records is then compared with expert supporting parameter list (*Expert Name (EN), Expert Current Designation (ECD), Email id of the expert (EID), Expert Previous Designation (EPD), Current Employer (CE), Previous Employer (PE), Co-Authors (COA), Current Organization Website (CWEB), Previous Organization Website (PWEB)*) to ensure the number of records belong to the expert.

3.2 Expertise Level Mapping Algorithm

The algorithm generated for the expertization level mapping is described below:

Step 1: Let TR as Total Number of Records Retrieved using Search Engine.

- Let AR as Records actually belong to Expert.
- Let EMP as Expertise Mean Point.
- Let TKR as Target Keyword Relevance.
- Let PKR as Partial Keyword Relevance.
- Let EL as Expertise Level

Step 2: $EMP = TR - AR$.

Step 3: $n =$ number of records;

```

I = 1.....n;
TKR=0;
PKR=0;
For (AR=1; I<=n; I++)
{
IF (KEYWORD ∈ AR) THEN
{

```

```

TKR = TKR+1;
}
ELSEIF (PARTIAL KEYWORD ∈ AR) {
PKR =PKR + 1;
}
ELSE
{
TKR =0;
PKR=0;
}
}

```

Step 4: $EL = (TKR * 0.9 + PKR * 0.1)$.

The algorithm generated for Expertise Level Mapping indicates that Expertise Mean Point is generated by subtracting the number of records actually belongs to Expert from the Total Number of Records Retrieved using Search Engine. The Target Keyword Relevance and Partial Keyword Relevance are calculated by comparing the expert keyword with the Records actually belongs to Expert. The Expertise Level is then generated by adding the Target Keyword Relevance by multiplying with 0.9 and Partial Keyword Relevance with multiplying 0.1.

4. Results

A functionality audit of the prototype system was carried out with three experts using six specialized keywords. Nearly 100 results were downloaded for each specialized keyword of an expert. In an average 600 search engine results were collected for each individual expert. Altogether for four experts and six specialized keywords, we were downloaded 1799 results with the help of search engine. The three expert name and specialized keywords are given in the Table. 1.

Table 1. Expert name and Keywords

Expert Name	Expertise Keywords
Dr. Jaya	<ul style="list-style-type: none"> • Knowledge Engineering
Dr. P. Sheik Abdul Khader	<ul style="list-style-type: none"> • Ontology • Networks
Dr. Munir Ahmed Rabbani	<ul style="list-style-type: none"> • Tacit • RFID • Information Retrieval

The Figure 2 shows few sample results from the results downloaded for specialized keywords of an expert.

The results downloaded for each specific keyword of an expert were cross verified with the relevance and expertise level mapping algorithm. This process yielded the result as given in the Table 2.

The reveals that, the experts utilized are Dr. Jaya, Dr. P. Sheik Abdul Khader and Dr. Munir Ahmed Rabbani. The specialized keywords used such as Knowledge Engineering, Ontology, Networks, Tacit, RFID and Information Retrieval. The data downloaded for each expert on each particular keyword were 100 nos. The actual expert data field shows the exact data pertaining to the expert from the downloaded data. The expert data percentage is the average

point retrieved by calculating the downloaded data divided by actual expert data and the result multiplied with 100. The target keyword relevance field is the data, which show case the exact number of search engine result contains the specialized keyword of the expert. The partial keyword relevance field is the data, which depict the number of search engine results were partially matched with the expert's specialized keywords. The expertise level mapping is calculated on multiplying 0.9 with target keyword relevance data and 0.1 with partial keyword relevance data. Finally both the results were added to obtain the expert expertise level on a specific keyword. The Table 3 depict the obtained results for each individual expert on each individual specialized keywords.

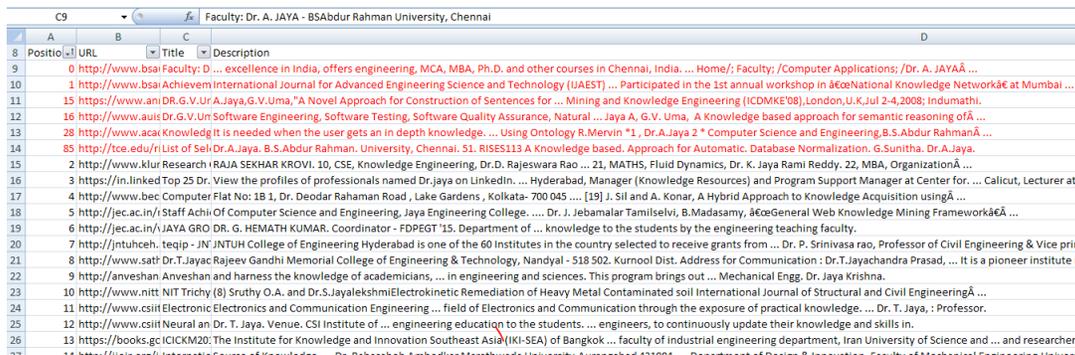


Figure 2. Expert extracted data.

Table 2. Expert expertise relevance and mapping

Expert Name	Expertise Keyword	Data Downloaded	Actual Expert Data	Expert Data Percentage	Target Keyword Relevance	Partial Keyword Relevance	Expert Expertise Level
Dr. Jaya	Knowledge Engineering	100	7	7	1	5	1.4
	Ontology	100	6	6	5	0	4.5
	Networks	100	2	2	1	0	0.9
	Tacit	100	1	1	1	0	0.9
	RFID	100	1	1	1	0	0.9
	Information Retrieval	100	5	5	3	4	3.1
Dr. P.	Knowledge Engineering	100	42	42	0	32	3.2
Sheik Abdul Khader	Ontology	100	21	21	16	0	14.4
	Networks	100	49	49	43	0	38.7
	Tacit	100	9	9	2	0	1.8
	RFID	100	22	22	17	0	15.3
	Information Retrieval	100	32	32	16	16	16
Dr. Munir Ahmed Rabbani	Knowledge Engineering	100	6	6	0	5	0.5
	Ontology	100	8	8	8	0	7.2
	Networks	100	1	1	1	0	0.9
	Tacit	99	17	17.1717172	16	0	14.4
	RFID	100	11	11	11	0	9.9
Information Retrieval	100	13	13	7	5	6.8	

Table 3. Expert expertise mapping

Keywords	Dr.Jaya	Dr.Munir Ahmed Rabbani	Dr.P.Sheik Abdul Khader
Information Retrieval	3.1	6.8	16
Knowledge Engineering	1.4	0.5	3.2
Networks	0.9	0.9	38.7
Ontology	4.5	7.2	14.4
RFID	0.9	9.9	15.3
Tacit	0.9	14.4	1.8
Grand Total	11.7	39.7	89.4

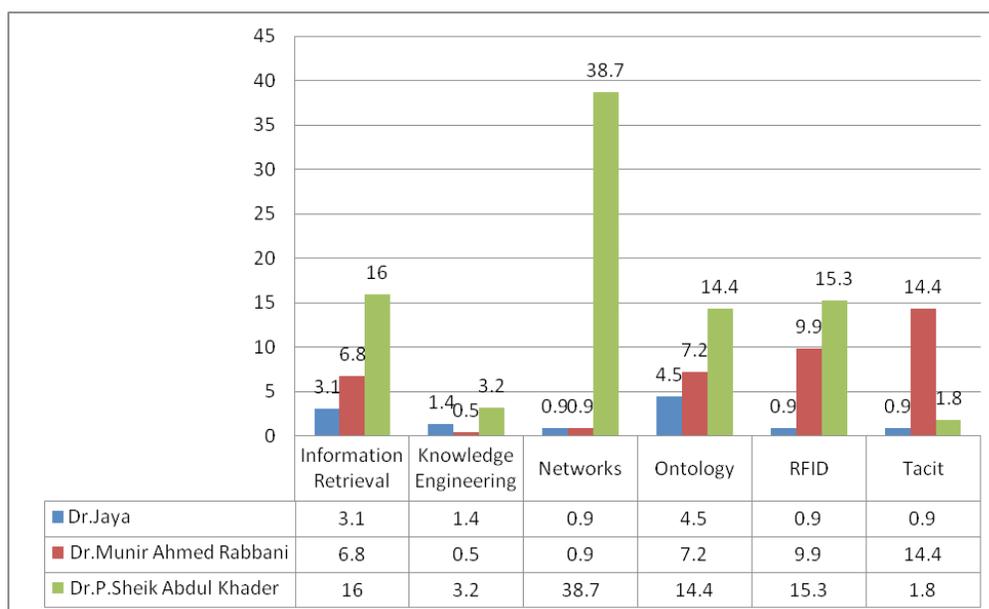


Chart 1. Expert extracted data.

The above chart presents each expert expertization level on specific field or area. Indeed these datasets are considered as actual or updated contribution of an expert on a specialized field compared with existing conventional expert finding technique.

5. Conclusion and Future Direction

The tacit knowledge of an expert can be externalized in an e-learning environment by using the approach of capturing and sharing of knowledge with the right person at the right time. Currently e-learning systems were automated to identify an expert in a specific domain by using two methods 1. Self classification and 2. Document based relevance. These two methods assume that the relevance of one’s specified keywords

or authored documents to the query is positively related to their expertise level on the query. In reality, one can be identified as an expert in a specific domain by the contribution made by the expert on that particular domain. The expert expertization data available in various websites could be utilized by the e-learning systems to evaluate the expert’s expertization level in addition to self classification and document based relevance. Hence it is proposed to ensure the expertise level of an expert using a Dynamic Expertization Estimating System (DEES) in an e-learning environment. In this approach, search engines were utilized as an agent to extract the expert’s data available in the internet and weightage were given to the expert according to their contribution made by the expert towards the given expertise area. The results retrieved by applying this mechanism yielded data with high accuracy levels to ensure the expertization level of

an expert. Future work could include connecting the Dynamic Expertization Estimating System (DEES) with social media to fetch more data pertaining to the expert expertization and produce accurate expertise level for each expert. This future extension may bring new expert connection with other similar experts and drive the expert finding approach to a greater level.

6. References

1. Wang G, Alan A. ExpertRank: A topic-aware expert finding algorithm for online knowledge communities. *Decision Support Systems*. 2013; 54(3):1442–51.
2. Moreira M, Catarina C, Martins B, Calado P. Using rank aggregation for expert search in academic digital libraries. *Ar Xivpreprint*. Portugal; 2015. p. 1–13.
3. Kardan K, Hendijanifard K. Expert finding system in e-learning. Using concept maps. *International Journal of Information and Communication Technology*. 2011; 3(1):73–81.
4. Macdonald C, Ounis I. Voting techniques for expert search. *Knowledge and Information Systems*. 2008; 16(3):259–80.
5. Serdyukov P. Search for Expertise: Going beyond direct evidence [PhD thesis]. University of Twente; 2009. p. 1–158.
6. Soboroff I, Vries APD, Craswell N. Overview of the TREC-2006 enterprise track. *Proceedings of the 15th Text Retrieval Conference*; UK. 2007. p. 1–20.
7. Craswell N, Vries APD, Soboroff I. Overview of the TREC-2005 enterprise track. *Proceedings of the 14th Text Retrieval Conference*; USA. 2006. p. 1–7.
8. Balog K, Azzopardi L, Rijke MD. Formal models for expert finding in enterprise corpora. *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*; USA. 2006. p. 43–50.
9. Wang G, Alan A. ExpertRank: A topic-aware expert finding algorithm for online knowledge communities. *Decision Support Systems*. 2013; 54(3):1442–51.
10. Mirsaeedghazi T, Mahmood K, Ahmad K, Neda E. A framework for learner preparation and support in academic e-learning environment. *International Journal of Information and Communication Technology*. 2011; 3(1):33–41.
11. Liu L, Duen-Ren D. Integrating expert profile, reputation and link analysis for expert finding in question-answering websites. *Information Processing and Management*. 2013; 49(1):312–29.
12. Venkataramanan M, Prema TS, Nandini SV. Knowledge management through distance education. *Indian Journal of Science and Technology*. 2011 Mar; 4(3):1–3. Nikabadi MS, Dehghan M, Farmanian-Arani M. The effect of knowledge management strategies on performance of new product development in knowledge-based companies. *Indian Journal of Science and Technology*. 2015 Apr; 8(S7):1–15.
13. Monisha S, Saranya R, Vijay K. Gauzy knowledge sharing in conspiring environment using text mining. *Indian Journal of Science and Technology*. 2016 Jul; 9(27):1–5.