A Researcher Expertise Search System using Ontology-Based Data Mining

Ravikarn Punnarut and Gridaphat Sriharee

Department of Computer and Information Science
King Mongkut's University of Technology North Bangkok
1518 Pibulsongkram, Bangsur, Bangkok 10800, Thailand

Email: [ravikarn_pu | gridaphat] @kmutnb.ac.th

Abstract

This paper proposes an approach to discover the expertise of researchers using data mining with skill classification ontology. The skill classification ontology is an information model containing skills of doing research in the area of computer and information science. A methodology to build the ontology is presented. The expertise search system is developed, which uses the skill classification ontology, researcher profiles and research profiles in the retrieving process. These profiles and ontology are expressed by OWL. Also, the matching and ranking processes are proposed and these follow semantic-based matching. We explored the evaluation of the retrieving process and the result shows that the proposed approach enables the expertise search system to be efficient regarding accuracy.

Keywords: Expertise Search, Ontology, Matching, Ranking, Data Mining.

1 Introduction

Expertise is the embodiment of knowledge and skills within individuals (Crowder et al., 2002). An individual may have different levels of expertise about different topics, and the expertise distinguishes experts from less experienced people and novices. Expertise search describes the process of seeking to find the people who might have the desired knowledge and skills. The seeking requires a range of information relating to levels of knowledge or experience possessed. There are many techniques available to obtain the knowledge related to expertise. For example, Ehrlich and Shami (2008) analyse the tools that people use to search for an expert such as questioning through dialog with an expert, personal networks and directories. Some systems use the profiles that indicate their expertise; such profiles may be obtained from different information sources such as curriculum vitae, publications, blogs, web sites and research project details. Combining such different data together may require gluing of vocabulary (Aleman-Meza et al., 2006). The expertise search is thus a mechanism to support expert search.

Currently, there are a number of expertise search systems and that are implemented for particular systems or online communities. In the former case, the system may have a limited data set and searching is by keyword/field matching according to the used technologies (e.g. the databases or the profiles). For expertise search in an online community, correlation of the data and the analysis of them are important. Therefore, it is hard to locate particular expertise by searching in an online community because of the vast amount of data. The expert search system may also provide an initial analysis of information by assigning scores based on degree of expertise. Also, ranking relevant expertise is focused.

Data mining is the process of discovering knowledge and their associations from a large amount of the data. The data mining with ontology is using ontology to represent the results (Nigro et al., 2007). The system is able to determine the knowledge at different concept levels (Han, 1995). The most representative applications relate to many research areas such as Medicine, Biology and Spatial Data, etc. Using ontology, an analysis of information relevancy is implemented by considering on the semantic relations of terms defined in ontology. Also, the degree of the relevancy is defined. Applying data mining, the expertise search system develops some algorithms to extract association rules or knowledge from a large collection of data.

In this paper, we propose an approach of using ontology to determine the expertise of the researcher and for retrieving process. We developed an expertise search system that used skill classification ontology to analyse the expertise of the researcher. Here, skill represents the expertise of the researcher and is analysed based on the research conducted by that researcher. We propose some contributions with detailed descriptions as follows.

(i) A process to build the skill classification ontology. The skill classification ontology is an information model that contains terms related to various types of expertise in the area of computer and information science.

(ii) A methodology of data mining to determine expertise of the researcher using the skill classification ontology. Probability value of the determination is defined.

(iii) The matching and ranking methodology in retrieving the relevant researchers who may have competency matched to the desired expertise. The matching follows semantic-based matching according to the skill classification ontology. Also, the evaluation is presented.

The rest of the paper is organised as follows. Section 2 discusses some related works. Section 3 presents the development of the expertise search system and describes the data set used in this work. Section 4 describes the skill classification ontology and the methodology to create it. Section 5 presents the determination of expertise of the researchers. Section 6 presents the matching and ranking process in retrieving the researchers according to the desired expertise. Section 7 illustrates the developed expertise search system and the evaluation is discussed. Section 8 is a conclusion.

2 Related Work

There are various approaches related to expertise search. Zhang et al. (2007) utilise an online community to find the people who may have expertise for answering a particular question. They analyse the experts by considering interactions of the people in questioning and answering the questions. Sim and Crowder (2004) use existing organisationally heterogeneous information sources to locate the experts. They provide an expertise model that defines the relationship of different information sources to locate the experts. Another approach of locating expertise is the work of McDonald and Ackerman (2000). They propose making referrals of information to assist the people to find the experts. Ackerman et al. (2003) present an approach that allows everyone to contribute their competencies by labelling their expertise with relative concepts that describe the expertise. Macdonald and Ounis (2008) propose an approach that uses a range of documents in searching for expertise. The degree of the document is defined and used in the ranking process. Tang et al. (2007) propose an expertise search system that analyses information from a web community. They use ontology to determine the correlation between information collected from different sources.

Our work is close to that of Tang et al. (2007) by which the ontology is used to determine the expertise of the researchers. However, our work focuses on the conceptual level of the expertise according to the ontology. We address the ranking of relevant results – similar to the work of Macdonald and Ounis (2008) – in which the number of profiles related to the person is one of the criteria for ranking. In contrast to the works mentioned above, we use ontology for expertise analysis and the matching and ranking relevant results are addressed according to ontology and subsumption relationship (Brachman, 1983) of the concepts that represent the expertise.

3 The Process and Data Set

3.1 The Overview of the Process

Figure 1 represents the process of the development of the expertise search system. There are various steps involved with details as follows.

(i) Collecting the data set. The data set consists of the description of the research papers and research projects (described in Section 3.2).

(ii) Creating skill classification ontology. The skill classification ontology consists of CCS ontology and support ontology. The CCS ontology follows the ACM category (ACM, 1998). The support ontology is an extension model of the CCS ontology. The skill classification ontology is used for retrieving expertise.

(iii) Expertise analysis of the researcher. We use the skill classification ontology to analyse the related expertise of the researcher (described in Section 5).

(iv) Creating researcher and research profile. In this step, we provided a generator to create the profiles that are expressed by OWL (OWL, 2004).

(v) Developing the system and including the matching and ranking process for retrieving expertise.

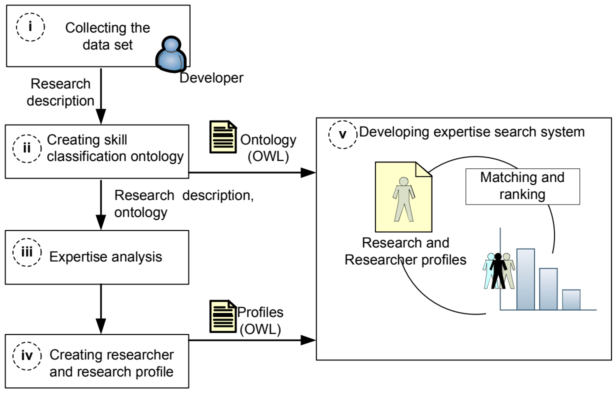

Figure 1: The process

3.2 Data Set

We collect description for use in the system from two sources: the proceedings of the JCSSE and NCSEC conferences and the research report of NRCT. The JCSSE (International Joint Conference on Computer Science and Software Engineering) and NCSEC (National Computer Science and Engineering Conference) are well-known conferences for national researcher collaboration in the areas of computing, computer engineering and information science. The proceedings of JCSSE are available online at http://jcsse.cp.eng.chula.ac.th and the proceedings of NCSEC are available in hard copy. The NRCT (National Research Council of Thailand) provides information on conducted research that is funded by the council. The information includes project description and researcher description, and both are provided in terms of XML documents. The project description is a description of research projects that are conducted and reported in a particular year and the researcher description is a description of the researchers that have registered with the NRCT.

4 Creating Skill Classification Ontology

4.1 The Process

Currently, there is no ontology available that represents expertise. Therefore, we built a skill classification ontology. Figure 2 depicts the building process of the skill classification ontology. We extracted terms from the titles of the conducted researches of the data set (a). The extracted terms are considered in matching (b) with XML-based ACM category (ACM, 1998) (see Figure 3). The mismatched terms are the terms that are not matched to ACM category, but such terms can be considered to be defined into the support ontology since these indicate some skills. We consider the terms to be defined into the support ontology using FOLDOC computer dictionary (FOLDOC, 1993) (e). The support ontology and CCS ontology are created and these are expressed by OWL. The CCS ontology consists of terms defined in the ACM category. The skill classification ontology is used for determining expertise of the researcher, matching and ranking in the expertise search system.

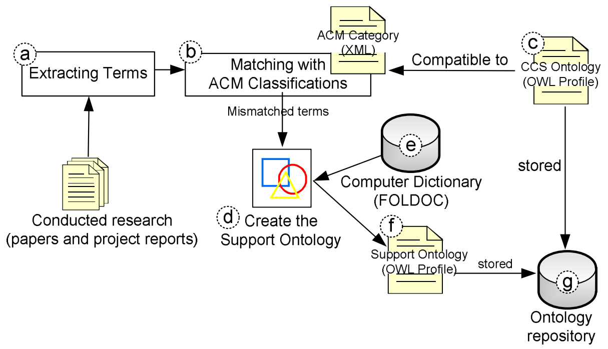

Figure 2: Building skill classification ontology

4.2 CCS Ontology

The ACM Computing Classification System (ACM, 1998) is adopted to create the CCS ontology. It provides indexing for computing publications so that the users can search through ACM's digital library. The ACM Computing Classification System represents a four-level tree that has three coded levels of terms. The reader can see the CCS model at http://www.acm.org/class/. Figure 3 is an example of an ACM category and their sub-categories and related terms.

We formalised the ACM category. For example (see Figure 3), the category H.2.4 is classified into sub-categories: *H.2.4.1 Concurrency*, *H.2.4.2 Distributed databases*, *H.2.4.3 Multimedia Database* and so on. The formalised terms are defined into the CCS ontology.

4.3 Support Ontology

The support ontology contains skill terms related to expertise, and it is an extension model of the CCS ontology. We conducted an experiment to extract terms from the title of the research paper and project. Of the results explored, 1,272 research titles are used for matching to the ACM category (ACM, 1998). Table 1 represents an example of the matched terms and extension terms.

The associations of the terms defined in the support ontology and the CCS ontology are defined with OWL properties such as `owl:equivalentClass` and `rdfs:subClassOf`. Figure 4 represents an example of the association between the term *Authentication* in the support ontology and in the CCS ontology.

```
<node id="acmccs98" label="ACMCCS98">
 <isComposedBy>
<node id="A." label="General
Literature">...</node> …
<node id="H." label="Information Systems">
<isComposedBy>
<node id="H.0" label="GENERAL"/>
<node id="H.1" label="MODELS AND PRINCIPLES">
<node id="H.2" label="DATABASE MANAGEMENT">
  <isRelatedTo><node id="E.5"/></isRelatedTo>
  <isComposedBy>
   <node id="H.2.0" label="General"> … </node>
    …
   <node id="H.2.4" label="Systems">
   <isComposedBy>
     <node label="Concurrency" />
     <node label="Distributed databases" />
     <node label="Multimedia databases" />
     <node label="Object-oriented databases" />
     <node label="Parallel databases" />
     <node label="Query processing" />
     <node label="Relational databases" />
     <node label="Rule-based databases" />
     <node label="Textual databases" />
     <node label="Transaction processing" />
   </isComposedBy>
   </node>…
</node>
```

Figure 3: Example of ACM category

CCS Categories	Ex. of Matched terms	Ex. of extension terms (defined in support ontology)
A. General Literature	dictionaries	-
B. Hardware	microcomputers, circuits, channel	clock gating, VHDL, OFDM
C. Computer Systems Organization	Security, asynchronous, transfer mode, protocols	Satellite, time division multiple access, radio wave
D. Software	enhancement measurement,prototyping	UML, XML
E. Data	data encryption standard (DES), compression, encoding	data migration, data reduction, video coder, decision tree
F. Theory of Computation	Mathematical, neural networks, routing	mathematical models
G. Mathematics of Computing	Statistic, time series analysis, wavelets	brownian motion, location estimation, sequential random
H. Information Systems	information services, library automation, GIS	knowledge base, sound synthesis, data fusion
I. Computing Methodologies	three dimensional, artificial intelligence, fuzzy	message passing interface (MPI), hidden markov model
J. Computer Applications	sciences, financial, manufacturing	Hazard and operability, quality control
K. Computing Milieux	CAI, software development, policy	electronic learning, watermarking, open source

Table 1: Matching terms with CCS

The association is defined with equivalent property where the concepts *Fingerprint, Watermarking, Electronic Signature* and *Face Detection* are defined as subconcepts (i.e. specific concepts) of *Authentication*. Combining the support ontology and the CCS ontology forms the skill classification ontology.

The skill classification ontology contains 1,644 terms (unduplicated concepts) in which 420 terms are defined in the support ontology and the rest are defined in the CCS ontology.

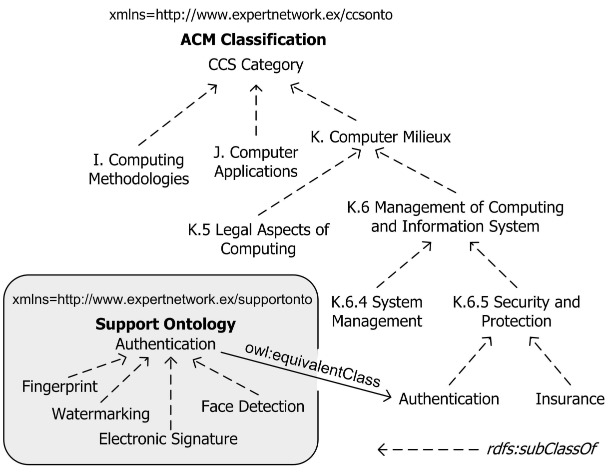

Figure 4: Part of skill classification ontology

5 The Expertise Determination

Using skill classification ontology, the system is able to determine the expertise of the researcher. Also, the system defines a score to represent the degree of expertise possessed by an individual researcher. Figure 5 depicts the research entitled "*Biochemistry Multimedia Learning on Web*" and the extracted terms: *Multimedia*, *Learning* and *Web* to be analysed. The extracted terms from the research title are denoted as follows.

$$\mathcal{T} = \{t_1, t_2, \ldots, t_n\} \qquad \text{...(a)}$$

Where $\mathcal{T}$ represents a set of extracted terms (t_1 to t_n) of a particular research title.

The detailed description of the determination in each step is as follows.

Step 1: Matching skill. This step aims to discover the skill categories related to the terms of the research. In this example (see Figure 5), there are skill categories H.2.4.3 and H.5.1 matched to the term *Multimedia*; I.2.6, I.2.6.1, I.2.6.7, K.3.1.0, and K.3.1.3 matched to the term *Learning*; and H.3.5.2 and H.5.3.7 matched to the term *Web*. The matched terms of each extracted term (from formula (a)) are the set of terms defined in the skill classification ontology. This is denoted as follows.

$$t_i = \{t_{i1}, t_{i2}, \ldots, t_{in}\} \qquad \text{...(b)}$$

Where t_{i1} to t_{in} are the matched terms of the extracted terms t_i (from formula (a)).

Step 2: Computing probability for relevant skill categories. This step aims to define the probability for matched skill categories. This is denoted as follows.

$$Prob(t_i) = 1 / N \qquad \text{...(c)}$$

Where N is the number of matched categories relating to the extracted term t_i.

In this example, the term *Web* relates to two skill categories such as *H.3.5.2 Online Information Services* and *H.5.3.7 Group and Organization Interfaces*. Hence, the probability of each matched category is defined as 0.5. This means the term *Web* may relate to category H.3.5.2 or category H.5.3.7 with equivalent probability. The probability represents the degree of the matched category of the extracted term from the research title.

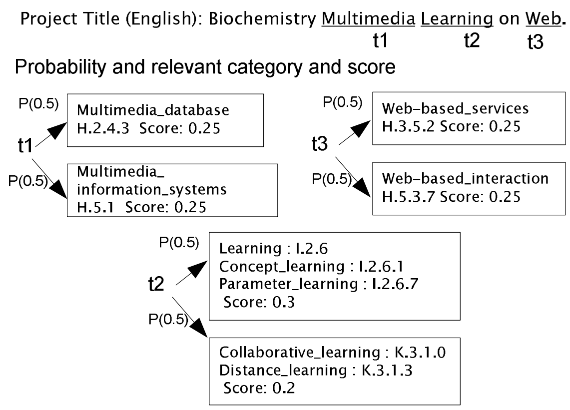

Figure 5: Example of expertise and score

Step 3: Computing score for relevant expertise. This step aims to assign a score for each relevant expertise, and the score varies with respect to the number of relevant skills of the categories. The score of a relevant expertise can be computed by multiplying the probability of the category by the match-term ratio. The match-term ratio is the number of matched terms of each category compared with the number of matched terms related to the extracted term. This is represented by the formula below.

$$Score_C = \frac{Prob(t_i) \times M_C}{M} \qquad \text{...(d)}$$

Where M_c is the number of matched terms for particular category (c) that is related to the extracted term t_i, M is the number of matched terms (t_{i1} to t_{in} in formula (b)) of term t_i and $Prob(t_i)$ computed from formula (c).

For example (see Figure 5), the term *Learning* has the defined probability 0.5 and 5 relevant skill terms, of which 3 terms related to I.2.6 and 2 terms related to K.3.1; thus, the computed scores are 0.3 (i.e. 0.5 x (3/5)) and 0.2 (i.e. 0.5 x (2/5)), respectively. The computed scores are assigned to each matched term in particular category. The matched term and its score represent the defined expertise judged by the system and these are considered in matching and ranking later. Note that the computed score in this step does not represent any significance regarding the degree of expertise. In this work, the number of the conducted research is considered regarding the degree of expertise (see Section 6.2.1).

The obtained expertise and its relevant score are defined into the research profile. In case the conducted research has co-researchers, the system defines the determined expertise and scores to all of them.

Figure 6 depicts an example of a researcher profile with the linkages to the research profile (see Figure 7) conducted (indicated by <projectMember>) by two researchers. Both profiles are expressed by OWL (OWL, 2004). In this work, we built an OWL profiles generator to create the research and researcher profiles, and the generator is developed with integrating Jena API (Jena, 2001).

```
<position rdf:datatype=" xsd:string">
      รองศาสตราจารย์ ระดับ 9</position>
<highestEducation rdf:datatype="xsd:string">
      Doctoral Degree </highestEducation>
<workOrg rdf:datatype="xsd:string">
   มหาวิทยาลัยนเรศวร คณะวิทยาศาสตร์การแพทย์</workOrg>
<conductProject rdf:resource="#P11000264"/>
<know rdf:resource="#R11000058"/>
</Researcher>
```

Figure 6: An example of the researcher profile

6 Matching and Ranking

6.1 Expertise Matching

We define two modes for retrieving: exact match mode and flexible match mode. These matching are defined based on subsumption relationship (Brachman, 1983) with details as follows.

(i) *Exact match mode.* Given the query with the desired expertise E (i.e. *Query*(E)), the system retrieves the researcher who may have the skill T (i.e. *Retrieve*(T)) where T is equivalent term of E . This is denoted as follows.

$$Query(E) = Retrieve(T) \text{ where } T \equiv E \quad ...(e)$$

(ii) *Flexible match mode.* Given the query with the desired expertise E, the system retrieves the researcher who may have the expertise T where T is a specialised concept of E. This is denoted as follows.

$$Query(E) = Retrieve(T) \text{ where } T \sqsubseteq E \quad ...(f)$$

In the case where there is no specialised concept of E, the system will perform a generalised match in which the generalised concept T corresponding to E will be retrieved. This is denoted as follows.

$$Query(E) = Retrieve(T) \text{ where } E \sqsubseteq T \quad ...(g)$$

In the case where the query is specified with multiple terms (expertise), the matching is considered for each particular term. For multiple terms query, the researcher profile satisfies the query if the set of retrieved skill matches to the set of expertise specified in the query. This is denoted as follows.

$$Query(Æ) = Retrieve(\tau) \quad ...(h)$$

Where Æ is a set of expertise specified in the query and τ is a set of retrieved expertise.

```
<Researcher rdf:ID="R42040150">
  <researcherNameThai rdf:datatype="xsd:string">
     ดาวัลย์ ฉิมภู่</researcherNameThai>
  <researcherNameEng rdf:datatype="xsd:string" >
    Dawan  Shimbhu,Mrs. </researcherNameEng>
<ResearchProject rdf:ID="P11000264">
  <projectNameThai rdf:datatype="xsd:string">
     การสร้างสื่อการเรียนวิชาชีวเคมีผ่านระบบเครือข่ายคอมพิวเตอร์
  </projectNameThai>
  <projectNameEng rdf:datatype="xsd:string">
       Biochemistry Multimedia Learning on Web
  </projectNameEng>
  <group rdf:datatype="xsd:string" >NRCT</group>
  <keyWord rdf:datatype="xsd:string">
       multimedia, learning, web</keyWord>
  <hasSkill rdf:parseType="Resource">
     <skillCCS rdf:resource="#H.2.4.3"/>
       <score rdf:datatype="xsd:float">0.25</score>
  </hasSkill>
  <hasSkill rdf:parseType="Resource">
     <skillCCS rdf:resource="#H.5.1"/>
       <score rdf:datatype="xsd:float">0.25</score>
  </hasSkill>
   …More defined skill ...
  <ProjectMember rdf:resource="#R42040150"/>
  <ProjectMember rdf:resource="#R11000058"/>
</ResearchProject>
```

Figure 7: An example of the research profile

Figure 8 represents an example of the matching when the query is specified with the term *H.2 Database_management*. The expertise of the researcher B is exact match to the query and the researcher A can be retrieved regarding specialised match. In case the query is specified with the term *H.2.4.3 Multimedia_database*, the researcher A is an exact match to the query and the researcher B is a generalised match to the query.

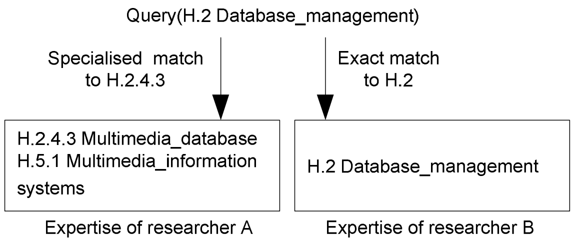

Figure 8: An example of matching

6.2 Ranking the Retrieved Researchers

We break down ranking into two steps: ranking according to *rank-score* and ranking according to *path-rank*. The former relates to the score defined in formula (d) (see Section 5) and the latter relates to the hierarchy of the skill classification ontology.

6.2.1 Rank-Score

Ranking can be considered from rank-score denoted as follows.

$$RankScore\ (E) = Maximum(Score\ (T_1) x N_1, \ldots, Score\ (T_n) x N_n) \qquad \ldots(i)$$

Where N_1 to N_n is the number of papers related to the retrieved skill T_1 to T_n that are the matched terms of E.

Note that the scores that are considered the maximum score are obtained from formula (d) (see Section 5).

Figure 9 depicts an example of retrieved results in flexible match mode when the category I.2.8 is specified in the query. The researchers A, B and C are retrieved since their expertise are specialised match to the category I.2.8.

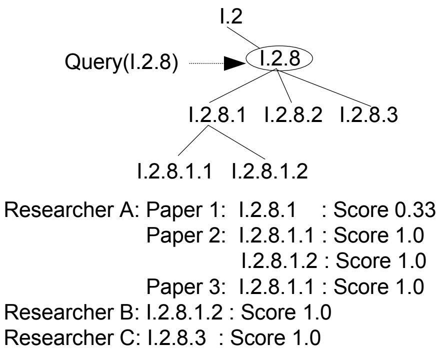

Figure 9: An example of matching

With regard to formula (i), the rank-score related to the matched expertise is computed as follows.

Researcher A: $RankScore\ (I.2.8) = Maximum(0.33_{I.2.8.1} x1, 1_{I.2.8.1.1} x2, 1_{I.2.8.1.1} x2, 1_{I.2.8.1.2} x1) = 2$

Researcher B: $RankScore\ (I.2.8) = Maximum(1_{I.2.8.1.2} x1) = 1$

Researcher C: $RankScore\ (I.2.8) = Maximum(1_{I.2.8.3} x1) = 1$

In this example, the researcher A is ranked in a higher order than the researchers B and C because of the number of the papers. The researchers B and C are in the same rank. According to the query with multiple terms, ranking can be considered from the summation of the rank-score of the set of retrieved expertise that matches to the query.

6.2.2 Path-Rank

Since rank-score may not be able to distinguish between different types of researcher expertise, path-rank is needed. We define the path-rank as the distance between the desired expertise and the matched term according to the hierarchy of skill classification ontology.

We focused retrieving based on user preference. The user may specify the query with the generic term or specific term (narrower semantics) according to terms defined in the ontology. With generic term, it represents that the user may have no knowledge of the expertise in the deep level or may not expect the narrower expertise in querying. In contrast, the query with the specific term represents that the user searches with particular requirement. In this work, the system ranks the results by giving the significance to the shortest distance between terms specified in the query and the retrieved expertise. This is denoted as follows.

$$PathRankScore(E) = PathRank(E \rightarrow T) \quad \ldots(k)$$

Where the path-rank is a value according to the distance from E to T in the hierarchy of skill classification ontology.

Note that path-rank is implemented in two retrieving modes: exact match mode, and flexible match mode with specialised match only.

In the previous step (Section 6.2.1), the researchers B and C are retrieved in equivalent rank because their rank-score are the same i.e. 1.0. Regarding the path-rank, the distance from the node I.2.8 to I.2.8.3 is shorter than the distance from the node I.2.8 to I.2.8.1.2. Thus, the system ranks the researcher C in a higher order than the researcher B. The long distance represents a poor rank. The system hence ranks the results by considering the ascending order of the path-rank. The query with multiple terms can be considered according to the summation of path-rank scores of the set of retrieved expertise that matches to the query.

7 The Development and Evaluation

We performed the evaluation by giving a set of expertise terms corresponding to the skill classification ontology and evaluated the results by investigating the retrieved results. The results are retrieved according exact and specialised match. The evaluation is conducted (by human) with regard to the following cases.

- (i) The system is able to retrieve the researchers who may have expertise matched to both the term specified in the query and the keyword specified in the research.
- (ii) The system is able to retrieve the researchers who may have expertise matched to the query but the keywords indicated in the research are not relevant to the query. However, the research paper is relevant to the desired expertise specified in the queries. In this case, the author may have specified keywords that are not appropriate to the researches.
- (iii) The system retrieves irrelevant results. This means the results are not relevant to the desired expertise.

The evaluation described above assesses the efficiency of the system and the proposed approach. For example, evaluation results represented by cases (i) and (ii) represent positive outcomes whereas case (iii) represents a negative outcome.

The data set for retrieving contains 1,046 research profiles (OWL) with the defined expertise (by system) and specified keywords. There are 1,693 researcher profiles that are used in the system and that have the linkage to the research profiles. Within the data set used, 245 research papers have their keywords (from ACM) specified by the authors and these keywords can be considered as the authors' description of their own skills. We also specified the keywords (from skill classification ontology) to 801 researches (papers/projects) that are not specified keywords before. Such these keywords are judged by the developer whether the results are relevant or not relevant to the test query.

Table 2 depicts the results of the evaluation. The evaluation can be summarised in terms of precision and recall. The recall value is computed from the ratio between the number of retrieved objects and the number

Query(E)	Retrieved objects (1)	Relevant to author keyword and query (+) (2)	Irrelevant to author keyword/ no keywords but relevant to query (+) (3)	Irrelevant to query (-) (4)	Precision = (2)+(3)/ (1)
Authentication	21 (31)	3 (5)	11 (15)	7 (11)	0.67 (0.65)
Cellular architecture	57 (92)	14 (27)	33 (51)	10 (14)	0.82 (0.85)
Cryptographic controls	7 (22)	1 (1)	4 (16)	2 (5)	0.71 (0.77)
Data encryption	15 (26)	2 (2)	8 (15)	5 (9)	0.67 (0.65)
Data mining	54 (72)	25 (28)	16 (15)	13 (29)	0.76 (0.60)
Data warehouse and repository	6 (18)	2 (5)	3 (12)	1 (1)	0.83 (0.94)
Network protocol	38 (52)	2 (4)	21 (28)	15 (20)	0.61 (0.62)
Programming techniques	23 (33)	0 (0)	10 (19)	13 (14)	0.43 (0.58)
Scene Analysis	30 (54)	1 (2)	16 (31)	13 (21)	0.57 (0.61)
Semiconductor Memories	4 (5)	0 (0)	4 (5)	0 (0)	1.00 (1.00)
Sound and Music Computing	20 (29)	1 (3)	11 (21)	8 (15)	0.60 (0.83)
Fuzzy set	25 (39)	4 (9)	10 (16)	11 (14)	0.56 (0.64)
Average					0.69 (0.73)

Table 2: Matching terms with CCS

of relevant objects in the data set. Note that the number specified without parenthesis represents the number of papers/projects while the number within parenthesis represents the number of researchers. The second column (indicated by (1)) represents the number of the retrieved objects according to the query in the first column. The third to the fifth columns represent the results corresponding to the cases (i) - (iii) mentioned above, and the final column represents the precision value.

From Table 2, the system returned the recall value that is 1.0 which means the number of relevant objects in the data set is equal to the number of retrieved objects. Also, the precision value is 0.69 and 0.73 regarding the number of retrieved papers/projects and the number of the researchers respectively.

Figure 10 and 11 depict the examples of the user interface for the developed expertise search system. The user can specify the desired expertise through the provided ontology browser implemented by AJAX technology. The system is implemented by JSF framework integrating with Jena API version 2.1 (Jena, 2001) for querying.

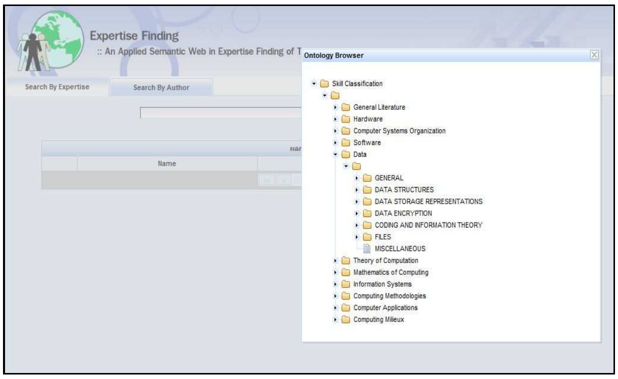

Figure 10: User interface for query

Figure 11: User interface expertise search system

8 Conclusion

In this work, we proposed a methodology for building a skill classification ontology by extracting text from research titles. With the skill classification ontology, the system is able to determine the related expertise of the researcher, and that expertise is analysed in regard to both broad skills and deeper skills. Matching and ranking processes are presented and these are conducted according to a semantic-based approach. Currently, the system supports the query with single term and two terms. From our experimentation, the proposed matching and ranking algorithm is practical and enables efficient search. We also evaluated the system using others set of the test query and the system returned approximately precision value 0.70.

Our work is focused on the research papers. However, the proposed approach may be applied to use with other information that share topics of interest of the people. It is possible to consider the different degree of expertise according to the positions of conducting research. For example, the leader of the conducted project or being the first author and second author respectively. In our system, the number of conducted researches is the most significant factor that yields well-justification of the degree of expertise. We did not focus on satisfaction of the users but rather focused on efficiency of the search system regarding accuracy. However, the evaluation on satisfaction can be implemented later.

References

Ackerman, M.S., Wulf, V., and Pipek, V. (2003): Sharing Expertise: Beyond Knowledge Management. *MIT Press*, Cambridge, MA.

ACM (1998). ACM Computing Classification System. Available at http://www.acm.org/class/

Aleman-Meza, B., Bojars, U., Boley, H., Breslin, J.G., Mochol, M., Nixon, L.J., Polleres, A. and Zhdanova, A.V (2007): Combining RDF Vocabularies for Expert Finding. *The Semantic Web: Research and Applications*, Springer Berlin / Heidelberg, Volume 4519.

Brachman, R. (1983): What IS-A Is and Isn't: An Analysis of Taxonomic Links in Semantic Networks. *IEEE Computer*, 16(10): 30-36.

Crowder, R., Hughes, G., and Hall, W. (2002): An Agent Based Approach to Finding Expertise. The 4th International Conference on Practical Aspects of Knowledge Management, LNAI 2569, Vienna, Austria, December 2-3: pp.179–188.

FOLDOC (1993): Free Online Dictionary of Computing. Available at http://foldoc.org/

Han, J. (1995): Mining Knowledge at Multiple Concept Levels. Proceedings of the 4th International Conference on Information and Knowledge Management (pp. 19-24). New York.

Jena (2001): Jena A Semantic Web Framework for Java. Available at http://jena.sourceforge.net/

Ehrlich, K., Shami, N.S. (2008). Searching for Expertise. Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems, Florence, Italy.

McDonald, D.W., Ackerman, M.S. (2000): Expertise Recommender: A Flexible Recommendation System and Architecture. CSCW 2000, Philadelphia, PA, December 2-6.

Macdonald, C., Ounis, I. (2008): Searching for Expertise: Experiments with the Voting Model. *The Computer Journal on Expertise Profiling*, March 6.

Nigro, H.O., Elizabeth, S., Cisaro, G., Xodo, D.H. (2007): Data Mining with Ontologies: Implementations, Findings, and frameworks, Information Science Reference, ISBN: 978-1-59904-618.

NRCT, http://www.riclib.nrct.go.th/

OWL (2004): OWL Web Ontology Language Guide. Available at http://www.w3.org/TR/owl-guide/

Sim, Y., Crowder, R. (2004): Evaluation of an Approach to Expertise Finding. The 5th International Conference on Practical Aspects of Knowledge Management, vol. 3336, Vienna, Austria, December 2-3: pp. 141-152.

Tang, J., Zhang, J., Zhang, D., Yao, L. and Zhu, C. (2007): ArnetMiner: An Expertise Oriented Search System for Web Community. Semantic Web Challenge. In Proceedings of the 6th International Conference of Semantic Web (ISWC’2007).

Zhang, J., Ackerman, M.S., Adamic, L. (2007): Expertise Networks in Online Communities: Structure and Algorithms. The 16th international conference on World Wide Web, Banff, Alberta, Canada, May 8-12.