www.ietdl.org

Published in IET Software
Received on 1st February 2013
Revised on 10th April 2013
Accepted on 24th April 2013
doi: 10.1049/iet-sen.2013.0015

ISSN 1751-8806

Improved document ranking in ontology-based document search engine using evidential reasoning

Wenhu Tang[1], Long Yan[2], Zhen Yang[2], Qinghua Henry Wu[1,2]

[1]School of Electric Power, South China University of Technology, Guangzhou, People's Republic of China
[2]Department of Electrical Engineering and Electronics, The University of Liverpool, Liverpool L69 3GJ, UK
E-mail: wenhutang@scut.edu.cn

Abstract: This study presents a novel approach to document ranking in an ontology-based document search engine (ODSE) using evidential reasoning (ER). Firstly, a domain ontology model, used for query expansion, and a connection interface to an ODSE are developed. A multiple attribute decision making (MADM) tree model is proposed to organise expanded query terms. Then, an ER algorithm, based on the Dempster–Shafer theory, is used for evidence combination in the MADM tree model. The proposed approach is discussed in a generic frame for document ranking, which is evaluated using document queries in the domain of electrical substation fault diagnosis. The results show that the proposed approach provides a suitable solution to document ranking and the precision at the same recall levels for ODSE searches have been improved significantly with ER embedded, in comparison with a traditional keyword-matching search engine, an ODSE without ER and a non-randomness-based weighting model.

1 Introduction

1.1 Background

In the past few years, a variety of information retrieval (IR) models have been developed for document ranking in document search engines, for example, a probabilistic model, an inference network model and a vector space model (VSM) [1, 2] and so on, out of which VSM is the dominant one. In [3], it summarises that the main function of a search engine is to discover the information in relation to a query input. When a query is submitted, which is typically in the format of several keywords, a search engine retrieves relevant documents according to the query. The result then is usually returned as a list of relevant documents that are ranked in a descending order of their relevance scores concerning the query. In practice, a well-formatted query can explicitly illustrate the information required by a user and thus lead to a high search accuracy. However, the search accuracy of a search engine may be greatly restricted because of unclear and incomplete queries. In order to overcome this problem, query expansion (QE) has been introduced as a viable solution. Generally, QE presents a process of expanding a query input with its related terms. With an expanded query, the documents which do not contain the same keywords of the original query, but are correlative to such inferred terms, can be retrieved. Consequently, the search scale in a search process is suitably broadened and more accurate results may be obtained by retrieving more relevant documents.

Various QE techniques, which are mainly based on the mechanisms of relevance feedback [4] and statistical term co-occurrence [5], have been developed. In most cases, an improved search accuracy can be achieved using these techniques. However, a significant drawback of the above two QE techniques is that the related terms of a query input are obtained by analysing the context of documents stored in a document repository. Thus, the relatedness between related terms and an original query cannot be ensured, if there are not sufficient documents used for analysing before a search process.

As words always have various meanings, a search engine may not distinguish which meaning the user intends to input. As a result, the retrieval accuracy may decrease. The idea of semantics was formed in 1998, aiming to provide meaning or semantics to these data. The core of semantics is 'ontology', which has been applied to deal with the above problem of QE. In theory, an ontology model is 'an explicit specification of a conceptualisation' [6]. Ontologies used for QE can be either a general-purpose ontology, or a domain ontology developed regarding the context of a specific domain. The employment of a general-purpose ontology model for QE can be dated back to the early 1990s. A new ontology-structure-based technique for measuring semantic similarity between terms and concepts within multiple ontologies was presented in [7] for both single-ontology and cross-ontologies. In [8], the WordNet was used as a knowledge model for QE and a set of tests was carried out based on the Text REtrieval Conference document repository. In [9], a semantic IR system was introduced, in which a knowledge base was constructed by annotating documents of a document repository, using an ontology-based semi-automatic annotation mechanism. The test results showed that the semantic IR model achieved high performance compared with that of a keyword-based IR system.

This paper focuses on a domain ontology, which is an ontology model describing a specific domain, in which meanings of involved concepts are defined explicitly with respect to the domain. The advantage of using a well-defined domain ontology for QE, compared with general-purpose ontologies or textual thesauri, is that the meaning of a query term can be restricted within the domain and effectively disambiguated before a QE process, which is not achievable by general-purpose ontologies or published textual thesauri.

1.2 Drawback and solution

Regarding the ontology-based search engines using VSM as mentioned previously, much attention has been paid to the algorithms of discovering the best expansion terms and the number of expansion terms to achieve the best search accuracy. For instance, the related terms, for example, synonyms and hyponyms (subclasses), are added to an original query without considering their hierarchical relationships defined in an ontology. In such a situation, although mutual weights are assigned differently to a pair of terms in an expanded query, the emphasis of the original query could be biased in a search process. On the other hand, various IR models have been used to determine the relevance between terms of an expanded query and a document, as mentioned in Section 1.1. However, the relatedness between an expanded query and a document is normally calculated by the weighted sum of these generated relevance scores. In such a case, the relevance scores generated by the query terms are treated independently during the combination process, which may reduce the search accuracy of an ontology-based document search engine (ODSE) and thus is not a suitable way.

In order to deal with the above problems, an evidential reasoning (ER)-based [10] document ranking approach is adopted in this paper. ER is an algorithm developed for decision making, which can be used to combine multiple probability assignments generated from a range of evidence considering non-linearity and uncertainties [11]. The basic idea of the proposed approach is that expansion terms are considered as auxiliary evidence of an original query term. All terms of an expanded query are organised as a multiple attribute decision making (MADM) tree model of multi-levels, regarding their hierarchical relationships defined in a specific ontology. The Dempster–Shafer (DS) theory [12, 13] is then employed to develop evidence combination rules considering uncertainty and incompleteness. The details of ER are introduced in Sections 3 and 4.

1.3 Paper organisation

The rest of this paper is organised as follows. In Section 2, the calculation of a relevance score between a query term and a document with the VSM and non-randomness-based weighting (NBW) models is explained. Section 3 gives a brief introduction to ontology-based QE used in an ODSE and the methodology designed for organising an expanded query into an MADM tree model. In Section 4, the ER algorithm used for the combination of evidence with the DS theory is illustrated briefly. The experimental work is reported in Section 5, where the search scenario with the proposed ER-based methodology is illustrated and test results are discussed. This paper is summarised in Section 6.

2 Calculation of a relevance score using a classical VSM and an NBW

The concept of VSM was first introduced by Salton [14] in 1971. Briefly, VSM is a mathematical model used to determine the relevance score between a query input and a specific document in an indexed document repository. In a document ranking process using VSM, a document is conceptually represented as a vector. Each keyword extracted from the document is stored as a component of the vector. Meanwhile, a query input is defined as a vector containing query terms, each of which is denoted as a vector component. Then, the weight of a query term in the document vector is calculated by a weight method, for example, the term frequency-inverse document frequency method. Finally, the total relevance score between the query vector and document vector is computed with a cosine function [15].

The NBW model, first proposed by Chou *et al.* [16] in 2011, has been selected as a comparative approach in this research, which is a ranking algorithm for QE based on a term's appearing probability in a single document. The NBW has been proven to be competitive with other four expansion term weighting functions, including BM25, CHI-1, CHI-2 and Kullback-Leibler divergence [16]. It is a suitable comparison algorithm to verify the proposed ER-based ODSE approach. Basically, the NBW model starts with the concept of probability measurement, utilises the concepts of adjustment, and develops an expansion term weighting function through the summation of weights. The weighting function is defined as (1) and the derived weights are further re-weighted under the Rocchio's framework [16]

$$\text{weight}(t_i) = \left(\sum_{d_i \in R} P_{d_i}(t_i) \log_2 \frac{P_{d_i}(t_i)}{P_R(t_i)} \frac{\text{sim}(d_i,\ q)}{\sum_{d_i \in R} \text{sim}(d_i,\ q)} \right) \frac{\log_2 (N/N_{t_i})}{\log_2 (N)} \tag{1}$$

where weight(t_i) is the weight assigned to query term t_i, R is the top-retrieved document set, $P_{d_i}(t_i)$ is the appearance probability of term t_i in d_i, d_i is the document i belonging to R, $P_R(t_i)$ is the appearance probability of t_i in the whole document set, sim(d_i, q) is the value of the similarity comparison between document d_i and the initial query q, N is the number of documents in the whole document repository and N_{t_i} is the number of documents in which term t_i appears.

3 Methodology for organising expanded query with a MADM tree model

3.1 QE with an ontology

A domain ontology named substation ONTology (SONT) has been developed in this research using the Proétgé ontology development software, aiming to expanding an original query for a document search. SONT is programmed regarding the context of power substations only, which is the first ontology model specifically defined regarding the domain of power substations by power system experts. For example, considering the domain and scope of SONT, 'Power System' is defined as 'Thing' at the top ontology level, and a top–down development process is utilised to

doi: 10.1049/iet-sen.2013.0015

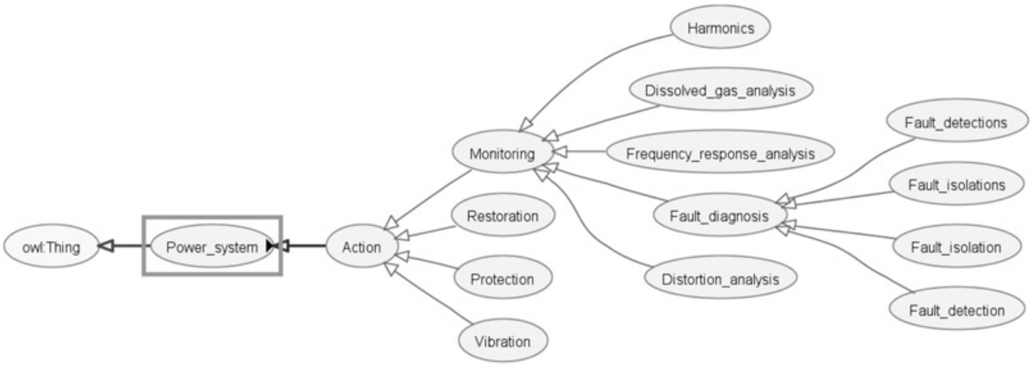

Fig. 1 *Class action and its hierarchy*

define the corresponding classes and the class hierarchy in proétgé. Fig. 1 illustrates one of the second-level classes 'Action' and its hierarchy. It contains four subclasses, namely 'Monitoring', 'Restoration', 'Protection' and 'Vibration'. The following levels include 'Fault diagnosis' and 'Fault detection' etc. Besides, the second-level classes contain nearly all important aspects of power substation, such as 'Attributes', 'Device', 'Other assets', 'Status' and 'Units'. Totally, the current version of SONT has 413 classes (i.e. concepts), 67 properties and 31 579 instances. Briefly, in ontology, a class is a general concept of a group of instances (i.e. individuals). For example, 'Transformer' can be defined as a class, then 'Current Transformer' can be defined as a subclass of 'Transformer'. Their relationship 'subclass' is a property. In the class of 'Current Transformer', numbers of instances can be defined, such as 'Current Transformer A', 'Current Transformer B' and so on. In this study, a query term is extended with its synonyms and hyponyms only. With SONT, a synonym set and hyponym set can be derived as (2) and (3), respectively, given a query as 'Fault diagnosis' for an illustration purpose. The singular and plural of a query term are treated as different terms. This is because of the two forms of the term are practically considered as two separate words in a document search process with the employed document search engines in this study. Thus, in the case that a query term is not expanded with its plural, a number of documents, which do not contain the original query term, but are correlative to its plural, cannot be retrieved during a subsequent document search process. Consequently, the number of relevant documents retrieved by a search engine may be reduced.

$$\{\text{fault diagnoses, condition assessment, condition assessments fault location, fault locations, fault analysis, fault analyses}\} \tag{2}$$

$$\{\text{fault detection, fault detections, fault isolation, fault isolations}\} \tag{3}$$

3.2 Decision making with ER algorithm

In the proposed ER-based approach, a MADM tree model is employed for organising the query terms of an expanded query. The essential idea is to map an expanded query into a hierarchical tree model. In the tree model, the synonyms are located at the same level as the original query and the hyponyms are distributed at a lower level. The relevance scores generated by those expansion terms with the document are designated as auxiliary evidence for evaluating the overall relatedness between the expanded query and the document.

A generalised MADM tree model is illustrated in Fig. 2, where a set of nodes is organised into a hierarchical structure. Two different levels, namely a factor level and an attribute level, are employed. In an evidence, combination process used for calculating the value of 'Overall evaluation', a set of attributes [attribute$_i$($i = 1, \ldots, I$)] that exists in the attribute level are evaluated. Meanwhile, the values of these attributes can either be obtained from users, or generated by several factors located at the factor level. For an instance, the value of attribute$_i$ is determined by factor$_{i,u}$, where $u = 1, \ldots, U$.

As introduced above, an expanded query derived from the query 'Fault diagnosis' can be organised into a tree model by considering hierarchical relationships among involved terms. In this study, the relevance score between the query 'Fault diagnosis' and a specific document, that is, RS_D, can be defined as the value of the 'overall evaluation' node in Fig. 2. The relevance score concerning the document, which are determined by using the query terms of the synonym set (2) and hyponym set (3), are regarded as auxiliary evidence for deciding RS_D. The relevance score between a single query synonym or hyponym and a specific document can be denoted as RS with a subscript, for example, $RS_{\text{faultdiagnoses}}$.

Returning to the MADM tree model, a decision-making process to generate RS_D is depicted as shown in Fig. 3, in which the four hyponyms of 'Fault diagnosis' are located at

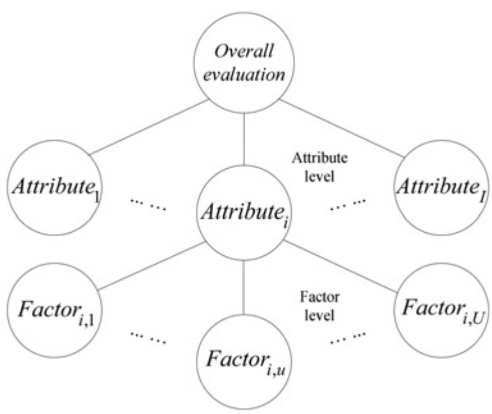

Fig. 2 *MADM tree model*

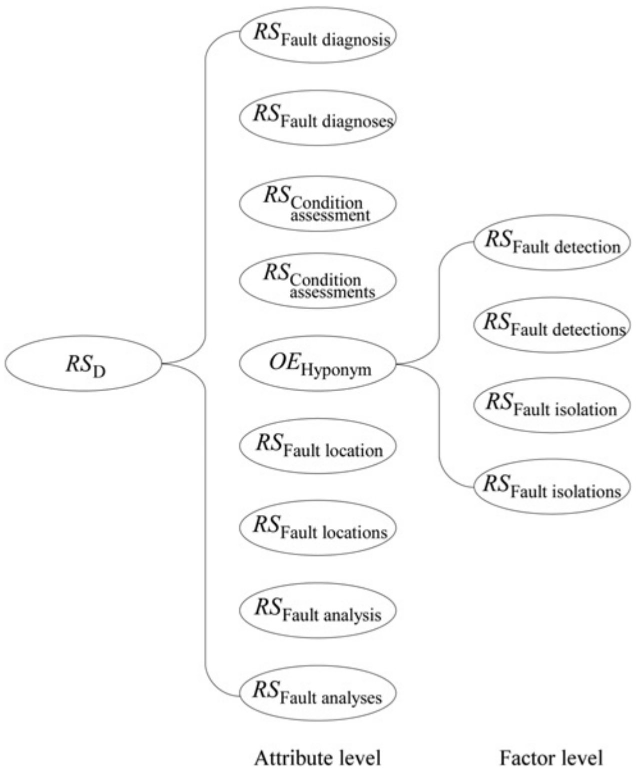

Fig. 3 *Tree model used for the combination of multiple relevance scores*

the factor level. The overall evaluation of the four hyponyms, which can be obtained from their relevance scores regarding the document with ER, is denoted as $\text{OE}_{\text{hyponym}}$. In the attribute level, the relevance scores decided by 'Fault diagnosis' itself and the terms of the synonym set (2) are included, and the value of $\text{OE}_{\text{hyponym}}$. The overall output of this MADM tree model, defined as RS_{D}, is calculated with the ER algorithm as introduced in the following section.

4 Evidential reasoning based on DS theory

4.1 Integrating basic probability assignments using ER

The DS theory is a mathematical theory dedicated to combining evidence with uncertainty. The seminal work of this subject was set by Dempster in 1967 and subsequently expanded by Shafer in 1976 [12, 13, 17]. The main interest of investigating DS in this study is to explore its effect of combining evidence in an MADM tree model. According to the DS theory, the relatedness between an expanded query and one of its relevant documents identified from a document repository can be defined as a hypothesis Ψ. All relatedness values generated between a query and relevant documents are denoted as the sample space Θ. On the other hand, each term of the query can be seen as a piece of evidence supporting the hypothesis Ψ, that is, relatedness, between the query and a document. Meanwhile, the relevance scores are used as the probability assignments of these evidences for calculating a relatedness value considering their relative weights [17].

Supposing that a document repository named D_{repo} comprises N different and independent documents. For the query 'Fault diagnosis', in total W documents of D_{repo} are determined as its relevant documents by a Lucene-based ODSE, after it is expanded with (2) and (3). Moreover, the relevant document set containing the W documents is defined as D_{rele}. Consequently, each of the terms in (2) or (3) and the original query 'Fault diagnosis' generates W different relevance scores regarding the W relevant documents of D_{rele}. A term 'Condition assessment', which is a synonym of 'Fault diagnosis', is selected for an illustration purpose. A set of relevance scores may be obtained with 'Condition assessment' as

$$\left\{\text{RS}_{\text{ca},1},\ \ldots,\ \text{RS}_{\text{ca},w},\ \ldots,\ \text{RS}_{\text{ca},W}\right\} \tag{4}$$

where $\text{RS}_{\text{ca},w}$ is the relevance score between 'Condition assessment' and a relevant document D_w extracted from D_{rele}. The value of $\text{RS}_{\text{ca},w}$ is assigned to a closed interval [0, 1].

Since the relatedness between 'Fault diagnosis' and D_w is denoted as a hypothesis Ψ in the DS theory, $\text{RS}_{\text{ca},w}$ can be considered as a confidence degree of 'Condition assessment' assigned to D_w. In addition, there are W relevant documents of 'Fault diagnosis' existed in D_{rele} and the upper limit of the total belief committed to these documents by 'Condition assessment' is 1. Each relevance score $\text{RS}_{\text{ca},w}$ can then be normalised as

$$\overline{\text{RS}}_{\text{ca},w} = \frac{\text{RS}_{\text{ca},w}}{W} \tag{5}$$

The evidence set located at the attribute level of Fig. 3 can be defined as $e_i (i = 1, \ldots, L_j)$, where L_j is defined as the number of evidence e_i. A basic probability assignment that the evidence e_i supports the overall relatedness between 'Fault diagnosis' and the document D_w then can be expressed as m_i^w. Also, a confidence degree confirmed by evidence e_i, for example, $\overline{\text{RS}}_{\text{ca},w}$, is denoted by $\beta_{D_w}(e_i)$. Therefore the basic probability assignment m_i^w, that is, a belief, is determined by the following equation [18]

$$m_i^w = \lambda_i \beta_{D_w}(e_i) \tag{6}$$

where $\lambda_i = \left[\lambda_1,\ \ldots,\ \lambda_{L_j}\right]^{\text{T}}$ expresses the relative weight assigned to evidence e_i in the evidence set. Obviously, if there exists only one piece of evidence in the attribute level, m_i^w is equal to $\beta_{D_w}(e_i)$. In the next subsection, the algorithm employed for determining the relative weights λ_i is introduced.

4.2 Determining relative weights among attributes

In this section, the standard analytic hierarchy process (AHP) [17, 19] is introduced for generating the relative weights λ_i of evidence mentioned in Section 6.1. Briefly, AHP is a multi-criteria decision support methodology used in management science. In an AHP, the important values of evidence are assigned as a set of grades, which are judged based on the same benchmark. By consulting search engine experts and power system engineers, the mutual importance values between the relevance scores generated by an original query term and a synonym, the synonym and $\text{OE}_{\text{hyponym}}$, the original query and $\text{OE}_{\text{hyponym}}$ of the attribute level in Fig. 3 are defined as 2:1, 2:1 and 3:1, respectively. Meanwhile, the weights are equally assigned to the elements of the factor level of Fig. 3. This is because they are all hyponyms of the original query and thus have equal

doi: 10.1049/iet-sen.2013.0015

influences for determining the overall relatedness between an expanded query derived from the original query and a document. An example of the implementation of AHP technique is illustrated in Section 5.3.

4.3 Implementing DS combination rules for document ranking

Implementing the DS combination rules in ER aims to generate a set of values, each of which represents the overall relatedness between an expanded query and a specific document. For the evidence set e_i as mentioned in Section 4.1, a combined probability assignment that indicates the relatedness between 'Fault diagnosis' and a specific document of $D_{\rm rele}$ can be defined as $m_{e_i}^{\Psi}(\Psi \subseteq D_{\rm rele})$. Furthermore, the remaining belief that is not assigned after commitments to all the documents of $D_{\rm rele}$ is defined as $m_i^{D_{\rm rele}}$, which is $m_i^{D_{\rm rele}} = 1 - \sum_{w=1}^{W} m_i^w$. The recursive formulas used for determining the overall relevance score between 'Fault diagnosis' and D_w is obtained as follows [18]

$$\{D_w\}: m_{e_{i+1}}^w = K_{e_{i+1}}\left(m_{e_i}^w m_{i+1}^w + m_{e_i}^w m_{i+1}^{D_{\rm rele}} + m_{e_i}^{D_{\rm rele}} m_{i+1}^w\right) \quad (7)$$

$$\{D_{\rm rele}\}: m_{e_{i+1}}^{D_{\rm rele}} = K_{e_{i+1}} m_{e_i}^{D_{\rm rele}} m_{i+1}^{D_{\rm rele}}, \quad w = 1, \ldots, W \quad (8)$$

where

$$K_{e_{i+1}} = \left[1 - \sum_{\tau=1}^{W} \sum_{\rho=1,\, \rho \neq \tau}^{W} m_{e_i}^{\tau} m_{i+1}^{\rho}\right]^{-1}, \quad i = 1, \ldots, L_j - 1$$

In the case that a query is composed of more than one keyword, the overall relevance score generated by each of the keywords can be defined as a piece of evidence of the query. Therefore the final relatedness between such a query and a document can then be derived with the above formulas by combining all the available evidence, with equally assigned mutual weights.

5 Test results

5.1 Test schemes

Implementing the ER-based approach to an ODSE aims to improve the search accuracy of a document retrieval process, compared with that achieved by an ODSE using the weighted sum algorithm of VSM. There are four different document search engines, named SE_1, SE_2, SE_3 and SE_4, respectively. The first three have been developed based on the Apache Lucene search engine library and implemented for the tests. SE_1 represents a traditional keyword-matching document search engine with a weighted sum algorithm, which does not employ any QE techniques. SE_2 and SE_3 are two ODSEs with a weighted sum algorithm and with the implementation of the ER algorithm, respectively. SE_4 is the NBW model, which is employed for comparison purposes.

5.2 Experiment platform

As presented in Table 1, in total 136 735 documents have been employed throughout the tests. All the keywords defined in SONT were employed as keywords to search relevant publications in the document repository. Part of the documents is gathered concerning the domain of power substations, including published academic papers, technical reports, emails of power companies and massive maintenance records.

For each of the four search engines, two query sets are utilised for checking practical search performance, which are selected by considering widely used query terms related to a power substation. As illustrated in Table 2, while the first set is composed of ten unique-keyword queries, another ten combined-keyword queries are selected as the second set. It is worth mentioning that the tree model and keywords used in the tests were recommended by power system engineers concerning substation condition monitoring and assessment, which were summarised from technical reports and emails of the power system engineers consulted. It should also be noted that all the documents used in the tests are named differently, so that each of them could be considered as a distinct individual.

5.3 Simple search scenario with SE_3

In order to obtain a better understanding on how to implement the proposed ER-based approach, a document search scenario using SE_3 is presented in this section. Again, the query 'Fault diagnosis' is utilised here for an illustration purpose. In this search process, first of all, the submitted query 'Fault diagnosis' is expanded with its synonyms and hyponyms, as illustrated in (2) and (3), respectively. Then, totally 24 933 documents in the test document set are identified as relevant documents of 'Fault diagnosis' with the Boolean model of Lucene, as introduced in Section 3.2. For each of these relevant documents, relevance scores are generated between the document and terms of the expanded query. Subsequently, the relevance score between the document and the expanded query is calculated with the ER algorithm. In the next step, the documents are listed in a

Table 1 Statistics of the document sets employed in tests

Number of documents	136,735
Unique-keyword queries	10
Combined-keyword queries	10
Average number of documents per document pool	86.3
Average number of relevant documents per document pool	32.5

Table 2 Queries utilised for the tests

Unique-keyword query set	Combined-keyword query set
1. Substation	1. Power system + frequency response analysis
2. Transformer	2. Winding + distortion analysis
3. Coolant	3. Relay + fault location + power system
4. Circuit breaker	4. Harmonics + distortion + power quality
5. Fault diagnosis	5. Transmission line + protection + relay
6. Dissolved gas analysis	6. Temperature + overload + thermal modelling
7. Relay	7. Transformer + dissolved gas analysis
8. Switch	8. Fault analysis + partial discharge
9. Thermal model	9. Transformer + vibration + mechanic
10. Voltage	10. Bus bar + protection

doi: 10.1049/iet-sen.2013.0015

Table 3 Two documents used for ranking

	Document name
D_a	an ontology for fault diagnosis in electrical networks
D_b	automated fault diagnosis at Philips medical systems

Table 4 Relevance scores in the factor level

	Fault detection	Fault detections	Fault isolation	Fault isolations
D_a	0.0000	0.0000	0.0000	0.0000
D_b	0.0202	0.0000	0.0265	0.0000

descending order according to generated overall relevance scores, which are delivered to users as final search results.

The method used for ranking two different documents in this search process is described as follows. As shown in Table 3, two documents, that is, 'An ontology for fault diagnosis in electrical networks' [20] and 'Automated fault diagnosis at Philips medical systems' [21] are randomly selected from the test document set and marked as D_a and D_b, respectively. Their overall relevance scores regarding the query 'Fault diagnosis' are defined as RS_a and RS_b, respectively.

As stated in Section 3, the overall relevance score RS_D between the query 'Fault diagnosis' and a document is determined by the combination of relevance scores of itself, its synonyms and hyponyms. These relevance scores are treated as a set of evidence concerning the generation of the overall relevance score RS_D. Based on the MADM tree model illustrated in Fig. 3, the outputs of the four hyponym branches can be treated as four attributes of the node $OE_{hyponym}$. The relevance scores obtained by the four hyponyms regarding D_a and D_b are shown in Table 4. Then, the normalised relevance scores are calculated as shown in Table 5 using (5).

According to (6), these relevance scores can be further treated as the confidence degrees $\beta_{Dw}(e_i)$. In this study, the relative weights λ_i are assigned as the same value 1/4 for each branch in the factor level of Fig. 3. Therefore the probability assignment m_i^w confirmed by 'Fault detection' is generated as 0.0202/24 933*1/4 = 0.0202/99 732. Hence, when all the probability assignments are obtained for D_a and D_b, the values of $OE_{hyponym}$ are obtained as 0.0000 and 4.6825×10^{-7} with (7) and (8) for D_a and D_b, respectively. Then, the outputs of the evidence located at the attribute level of Fig. 3 are listed in Table 6, when normalised by (5).

In order to integrate the relevance scores of Table 6 into scaled inputs for generating RS_a and RS_b, the relative weights λ_i are derived using AHP discussed Section 4.2. As stated in Section 4.2, the mutual importance values between an original query and a synonym, a synonym and $OE_{hyponym}$, the original query and $OE_{hyponym}$ are defined as 2:1, 2:1 and 3:1, respectively.

With the AHP algorithm, the relative weight set of λ_i is then obtained as $[0.2052, 0.1057, 0.1057, 0.1057, 0.0548, 0.1057, 0.1057, 0.1057, 0.1057]^T$. Finally, RS_a and RS_b are calculated as 3.1123×10^{-6} and 4.3971×10^{-6} with (7) and (8), respectively. Compared with the relevance score 3.1123×10^{-6} of D_a, it shows that D_b has a higher relatedness to the query 'Fault diagnosis' with the relevance score 4.3971×10^{-6}. Consequently, D_b is returned with a higher ranking order in the final search result, that is, a relevant document list, compared with that of D_a. The document ranking scheme illustrated above can be implemented to rank a number of documents in a similar way giving a query.

In the search processes with SE_1 and SE_2, the relevance scores generated between each term of an expanded query and a document is computed with the weighted sum of these relevance scores, as mentioned at the beginning of Section 5. In the next section, a recall and precision curve method, used for verifying the search performance of all the four search engines, is described. The pooling method used in this study is introduced as well.

5.4 Performance evaluation scheme

Typically, recall and precision, which are frequently used for evaluating search engine performance as reported in [16], are considered as two important performance indices employed for evaluating the effectiveness of a document search engine. In [16], the NBW method was proven as the most competitive search engines compared with other four expansion term weighting functions. For each comparison, ten different recall levels have been chosen. The gradient of the precision against recall curve [16] using NBW is smaller than other approaches. In addition, at each recall level, NBW has higher precision. The recall and precision method is a direct way to evaluate the performance of

Table 5 Relevance scores in the factor level after normalisation

	Fault detection	Fault detections	Fault isolation	Fault isolations
D_a	0.0000/24933	0.0000/24933	0.0000/24933	0.0000/24933
D_b	0.0202/24933	0.0000/24933	0.0265/24933	0.0000/24933

Table 6 Relevance scores in the attribute level

	Fault diagnosis	Fault diagnoses	Condition assessment	Condition assessments
D_a	0.3782/24933	0.0000/24933	0.0000/24933	0.0000/24933
D_b	0.5021/24933	0.0202/24933	0.0000/24933	0.0000/24933
$OE_{hyponym}$	Fault location	Fault locations	Fault analysis	Fault analyses
0.0000	0.0000/24933	0.0000/24933	0.0000/24933	0.0000/24933
4.6825×10^{-7}	0.0362/24933	0.0000/24933	0.0000/24933	0.0000/24933

doi: 10.1049/iet-sen.2013.0015

different search engines. Therefore the method of calculating recall and precision is selected in tests for verifying the search performance of SE_1, SE_2, SE_3 and SE_4. Given a query, its relevant documents in a document repository are defined as a set R. A set of relevant documents H is obtained by a document search engine after performing a search process. The recall and precision are defined as follows

$$R_c = \frac{H \cap R}{R} \tag{9}$$

$$P_r = \frac{H \cap R}{H} \tag{10}$$

where R_c is the recall value and P_r expresses the search precision.

In order to obtain the set R for each of the queries in Table 2, the pooling method introduced in [22] has been implemented. For each of the 20 queries shown in Table 2, SE_1, SE_2, SE_3 and SE_4 are utilised to implement a search process using the document repository, respectively. The top 50 ranked documents of each search process then are pooled into a document set for the corresponding query. Typically, such a document set is defined as a document pool. Then, by eliminating the reduplicate items in each document pool, the average amount of documents of the 20 document pools is derived as 86.3, as listed in Table 1.

For each of the 20 document pools, the documents are classified as 'relevant' and 'non-relevant' by power system engineers regarding the corresponding query. As a result, R of a specific query is known as the 'relevant' document set. Moreover, the mean of the document amount of the total 20 queries is obtained as 32.5 as illustrated in Table 1.

Implementing such a pooling method avoids evaluating all the documents in the document repository for a specific query. More significantly, R of all the 20 queries is determined in a logical way and thus may lead to a more accurate test result. Since the method does not guarantee that all the relevant documents of a query can be found and located in the corresponding document pool; therefore a recall value obtained in a search process then is defined as a 'relative recall'. With each of the two query sets of Table 2, the average precisions of SE_1, SE_2, SE_3 and SE_4 are calculated at ten different recall levels separately ranged from 10, 20 to 100%.

5.5 Test results and discussion

Fig. 4 shows the average precision of the ten unique-keyword queries at different recall levels. As can be observed from the figure, SE_1 presents the lowest precision throughout the search processes compared with the other three search engines. This may be because of the following reasons:

1. the synonyms of the original queries are not considered within the search processes of SE_1. Therefore the documents which contain the synonyms, but not holding the original queries, cannot be retrieved by SE_1;
2. the hyponyms of the original queries may have influenced the search performance as well, which means cases without considerations of subclasses in SE_1 may reduce search accuracies;
3. for SE_2 and SE_3, more relevant documents are found by the expanded query and the ER-based document ranking techniques, respectively. At the same recall levels, H may be smaller in the results of SE_2 and SE_3 in tests; however,

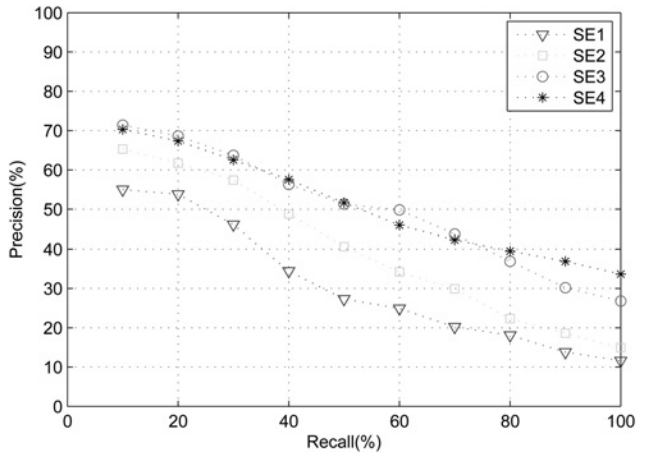

Fig. 4 *Average precision–recall curves with ten unique-keyword queries*

$H \cap R$ may be higher because of the improved ranking of each relevant paper. As a result, the precision values of SE_2 and SE_3 are higher than that of SE_1 with respect to the same recall values, which are also verified in the tests.

In addition, the average precision of SE_3 is higher than those of SE_2 at all the recall levels. This indicates that although the accuracy of a keyword-matching search engine can be improved with a QE technique provided by a domain ontology, the rough-and-tumble organisation of the relevance scores generated by the expansion terms can still restrict the search precision. Therefore the potential of the proposed ER-based approach in improving the search accuracy of an ODSE can be verified with the unique-keyword queries.

Fig. 5 presents the average recall-precision results of the ten combined-keyword queries. Compared with Fig. 4, the precision values of the four search engines have been improved. This means that utilising multi-keyword in one search process can refine the search scale, and thus improve the accuracy of the document search engine outputs. Among the three search engines, SE_3 delivers the best precision at every recall level, while the precision achieved by SE_2 is much higher than that of SE_1. Compared between

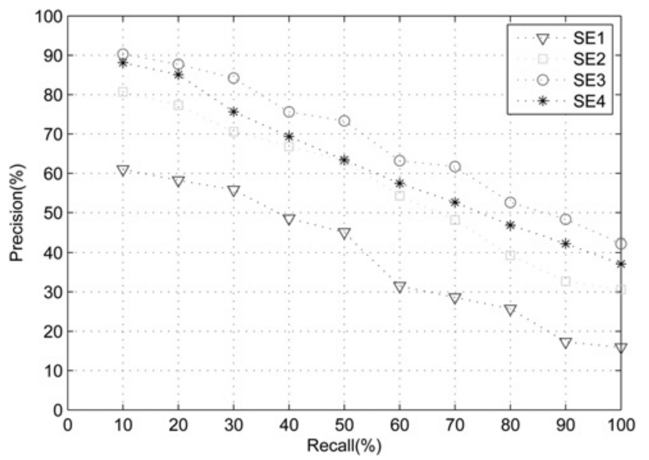

Fig. 5 *Average precision–recall curves with ten combined-keyword queries*

SE_3 and SE_4, SE_3 outperforms SE_4. A possible reason for SE_4 is that some of the retrieved expansion terms are less related to the initial query and in comparison for SE_3 the expanded terms derived by SONT are closely related to the initial query. Therefore for both the unique-keyword query set and the combined-keyword query set the ER-based document ranking approach has demonstrated its capability of improving the search accuracy of an ODSE in terms of precision at the same recall levels.

Finally, the overall evaluation (OE) results of all the 20 queries are shown in Fig. 6. Conventionally, a high precision is more important when it is generated at a low recall level. Therefore as shown in Table 7, at the recall level of 10%, the precision of SE_3 is 80.9%, which is much higher than those of the other three search engines. The above results clearly show that, in an ODSE, the ER algorithm provides a suitable solution for combining multiple relevance scores of an expanded query. Consequently, the search accuracy of an ODSE can be improved with the ER-based document ranking approach. The main theoretical contribution of the proposed approach is the introduction of ER for organising an expanded query into a hierarchical tree model, which can consider the hierarchical relationships among involved query terms. Moreover, the practical contribution is the construction of an ontology model dedicated to power substation fault diagnosis. As mentioned in Section 3.1, the ontology model SONT contains a large number of classes and instances, which cover most important aspects of power substation. For example, monitoring, status, devices etc. That is to say that it can be used as an open platform for expanding SONT. In addition, as reported in the recent research, there are some new approaches on QE. However, most of them are based on the mechanisms of relevance feedback. In [23, 24], the adaptive co-training method and the proximity-based approach were proposed. Both of the two approaches are based on Rocchio's model as NBW and have been proven to be much more competitive than other basic retrieval models. However, as mentioned in Section 1.1, if there are not sufficient documents used for analysis before a search process, the relatedness between related terms and an original query cannot be ensured. In comparison, SONT is a well-defined domain ontology, which functions as WordNet, containing a set of concepts related to power substation and their synonyms and hyponyms. It is a more direct and effective way to realise QE. We believe that an ontology specially designed for substation fault diagnosis is very useful in practice, which is scalable by adding new diagnosis terms and rules in SONT.

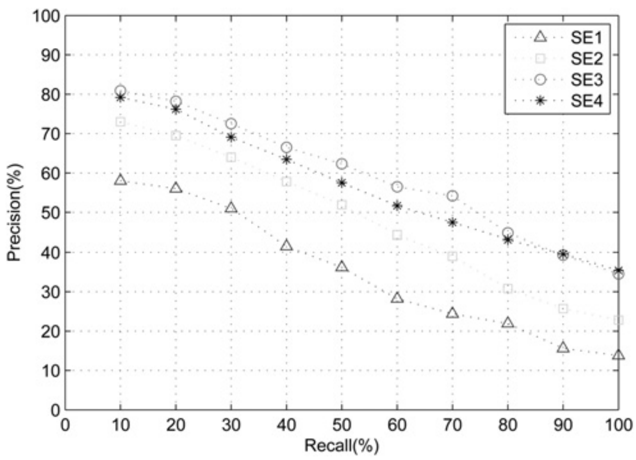

Fig. 6 *Average precision–recall curves with all the 20 queries*

Table 7 Average precisions of four search engines on recall levels of 10 and 20%

Method	10%	20%
SE_1	58.0	56.0
SE_2	73.0	69.5
SE_3	80.9	78.1
SE_4	79.2	76.2

6 Conclusions

In order to tackle the problems existing in traditional ODSEs, an ER-based document ranking approach has been proposed in this paper. It is the first time that the terms of an expanded query are organised into an MADM tree model in the search process of an ODSE. The ER algorithm is proposed for the combination of relevance scores generated between terms of an expanded query and a document. This novel approach used in an ODSE aims to generate the overall relatedness between an expanded query and a document. Practically, it can be implemented with other QE techniques, for example, relevance feedback-based techniques and statistical co-occurrence-based techniques, without considering the way of how an expanded query is generated. The traditional keyword-matching search engine, the NBW method and the two ODSEs with and without the ER algorithm are tested using a number of queries related to IR of substations, respectively. Compared with the keyword-matching search engine and the NBW method, the proposed approach outperforms in terms of both recall and precision. The final results demonstrate that the ER-based approach has provided a viable solution for ranking documents in an ODSE and the search precision of the ODSE has been improved significantly at the same recall level with ER embedded.

Finally, there are still some advanced topics worthy of study in the future. Because of the extra computation procedures of ER, the efficiency of the developed search engine using ER is slightly lower than other search engines without ER embedded. It takes more time than other search engines in a search process. Considering the concept of recall and precision, a perfect search engine has a constant 100% precision as recall increases. However, for a realistic search engine, the curve will always appear descending, which has been reported in many research papers. Figs. 4–6 are employed to find the most competitive search engine. However, they are not designed for efficiency evaluation. This issue will be investigated in our future research.

7 References

1 Wang, L.S.: 'Relevance weighting of multi-term queries for vector space model'. IEEE Symp. Computational Intelligence and Data Mining, 2009 (CIDM'09), 2009, pp. 396–402
2 Manning, C.D., Raghavan, P., Schutze, H.: 'Scoring, term weighting and the vector space model' (Cambridge University Press, 2008, 1st edn.), pp. 109–133
3 Bhatia, M.P.S., Kumar, A.: 'Contextual paradigm for ad hoc retrieval of user-centric web data', *IET Softw.*, 2009, **3**, (4), pp. 264–275
4 Eguchi, K., Ito, H., Kumamoto, A., Kanata, Y.: 'Adaptive and incremental query expansion for cluster-based browsing'. Proc. Sixth Int. Conf. Database Systems for Advanced Application, April 1999, pp. 25–34

doi: 10.1049/iet-sen.2013.0015

5 Chu, W.W., Liu, Z., Mao, W.: 'Textual document indexing and retrieval via knowledge sources and data mining', *Commun. Inst. Inf. Comput. Mach. (CIICM), Taiwan*, 2002, **5**, (2), pp. 135–160
6 Vizcaíno, A., García, F., Caballero, I., Villar, J.C., Piattini, M.: 'Towards an ontology for global software development', *IET Softw.*, 2012, **6**, (3), pp. 214–225
7 Al-Mubaid, H., Nguyen, H.A.: 'Measuring semantic similarity between biomedical concepts within multiple ontologies', *IEEE Trans. Syst. Man Cybern. C, Appl. Rev.*, 2009, **39**, (4), pp. 389–398
8 Bhogal, J., Macfarlane, A., Smith, P.: 'A review of ontology based query expansion', *Inf. Process. Manag.*, 2007, **43**, (4), pp. 866–886
9 Castells, P., Fernandez, M., Vallet, D.: 'An adaptation of the vector-space model for ontology-based information retrieval', *IEEE Trans. Knowl. Data Eng.*, 2007, **19**, (2), pp. 261–272
10 Tang, W.H., Spurgeon, K., Wu, Q.H., Richardson, Z.J.: 'An evidential reasoning approach to transformer condition assessments', *IEEE Trans. Power Deliv.*, 2004, **19**, (4), pp. 1696–1703
11 Shintemirov, A., Tang, W.H., Wu, Q.H.: 'Transformer winding condition assessment using frequency response analysis and evidential reasoning', *IET Electr. Power Appl.*, 2010, **4**, (3), pp. 198–212
12 Reed, S.R., Pétillot, Y.R., Judith, M.: 'Automated approach to classification of mine-like objects in sidescan sonar using highlight and shadow information', *IEE Proc.*, 2004, **151**, (1), pp. 48–56
13 Min, L.H., Chang, C.: 'Application of Dempster–Shafer's theory of evidence for transformer incipient fault diagnosis'. Eighth Int. Conf. Advances in Power System Control, Operation and Management (APSCOM 2009), November 2009, 1–6.
14 Salton, G., (Ed.): 'The SMART retrieval system: experiments in automatic document processing' (Prentice-Hall, Englewood Cliffs, NJ, 1971)
15 Gancarski, A.L., Doucet, A., Henriques, P.R.: 'Attribute grammar-based interactive system to retrieve information from XML documents', *IEE Proc. Softw.*, 2006, **153**, (2), pp. 51–60
16 Chou, S., Cheng, C., Huang, S.: 'A ranking algorithm for query expansion based on the term's appearing probability in the single document', *Online Inf. Rev.*, 2011, **35**, (2), pp. 217–236
17 Awasthi, A., Chauhan, S.S.: 'Using AHP and Dempster–Shafer theory for evaluating sustainable transport solutions', *Environ. Model. Softw.*, 2011, **26**, (6), pp. 787–796
18 Guo, L., Guo, C.X., Tang, W.H., Wu, Q.H.: 'Evidence-based approach to power transmission risk assessment with component failure risk analysis', *IET Gener. Transm. Distrib.*, 2012, **6**, (7), pp. 665–672
19 Farghal, S.A., Kandil, M.S., Elmitwally, A.: 'Quantifying electric power quality via fuzzy modelling and analytic hierarchy processing', *IEE Proc. Gener. Transm. Distrib.*, 2002, **149**, (1), pp. 44–49
20 Bernaras, A., Laresgoit, I., Bartolome, N., Corera, J.: 'An ontology for fault diagnosis in electrical networks'. Proc. Int. Conf. Intelligent Systems Applications to Power Systems, 1996, pp. 199–203
21 Lindhoud, W.M.: 'Automated fault diagnosis at Philips medical systems: a model-based approach'. MSc thesis, Delft University of Technology, Delft, The Netherlands, June 2006
22 Coelho, T., Calado, P., Souza, L., Ribeiro-Neto, B., Muntz, R.: 'Image retrieval using multiple evidence ranking', *IEEE Trans. Knowl. Data Eng.*, 2004, **16**, (4), pp. 408–417
23 Miao, J., Huang, X.J., Ye, Z.: 'Proximity-based Rocchio's model for pseudo relevance feedback'. Proc. 35th Int. ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR'12), 2012, pp. 535–544
24 Huang, X.J., Miao, J., He, B.: 'High performance query expansion using adaptive cotaining', *Inf. Process. Manage.*, 2013, **49**, (2), pp. 441–453