IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received December 12, 2018, accepted January 6, 2019, date of publication January 24, 2019, date of current version February 14, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2894679

A Taxonomy and Survey of Semantic Approaches for Query Expansion

MUHAMMAD AHSAN RAZA[1,2], RAHMAH MOKHTAR[1], NORAZIAH AHMAD[1], MARUF PASHA[2], AND UROOJ PASHA[3]

[1]Faculty of Computer Systems and Software Engineering, Universiti Malaysia Pahang, Kuantan 26300, Malaysia
[2]Department of Information Technology, Bahauddin Zakariya University, Multan 60000, Pakistan
[3]Institute of Management Sciences, Bahauddin Zakariya University, Multan 60000, Pakistan

Corresponding author: Muhammad Ahsan Raza (ahsan_0136@yahoo.com)

ABSTRACT Conventional approaches to query expansion (QE) rely on the integration of an unstructured corpus and probabilistic rules for the extraction of candidate expansion terms. These methods do not consider search query semantics, thereby resulting in ineffective retrieval of information. The semantic approaches for QE overcome this limitation, whereby a search query is expanded with meaningful terms that accord with user information needs. This paper surveys recent approaches to semantic QE that employ different models and strategies and leverages various knowledge structures. We organize these approaches into a taxonomy that includes linguistic methods, ontology-based methods, and mixed-mode methods. We also discuss the strengths and limitations of each type of semantic QE method. In addition, we evaluate various semantic QE approaches in terms of knowledge structure utilization, corpus collection, baseline model adaptation, and retrieval performance. Finally, future directions in exploiting personalized social information and multiple ontologies for semantic QE are suggested.

INDEX TERMS Information retrieval, morphological expansion, ontology, semantic query expansion.

I. INTRODUCTION

The main objective of an information retrieval (IR) system is to retrieve documents that are relevant to a user's intentions from a large information space. Such systems calculate the similarity between a search query and documents, and retrieve a list of documents that are arranged in descending order of similarity. The retrieved list of documents is sometimes large and contains many irrelevant documents, especially when searching the Web. The main issue that is encountered in the retrieval of documents that are not related to user needs is the vocabulary mismatch problem: the terms that the author has used to describe a concept in the document differ among users. Furnas *et al.* [1] conclude that the likelihood of searchers using the same word for a concept is less than 20%. The main reasons for such a low percentage include the use of words that have similar meanings (synonyms) and the use of words that represent the same term (polysemy).

This critical issue of vocabulary mismatch is further aggravated by short queries, which are becoming increasingly common in web search. Most of the prevalent web search queries contain no more than two or three words [2]; thus, the likelihood of encountering the severe issues of synonymy and polysemy is very low. Addressing the problem of vocabulary mismatch is essential for such short queries and important for effective information retrieval.

Query expansion (QE) is one of the most effective techniques for dealing with term scarcity, which is prevalent in web search queries. Several semantic QE approaches, including linguistic and ontology-based approaches, for addressing the vocabulary mismatch problem have been proposed in recent years. In this paper, we will discuss semantic QE approaches, which have substantial advantages over manual and statistical QE techniques [3] as they expand each search query with meaningful concepts that are captured from a knowledge structure (manually or automatically constructed) to represent the searcher's intent.

Existing semantic QE approaches can be categorized as linguistic, ontology-based and mixed-mode (hybrid). In linguistic approaches, word senses are derived from a thesaurus (a lexicon that includes synonyms, hyponyms and other possible word sense relationships among concepts) based on the original search query terms, which serve as expansion terms. The source of the expansion terms is an ontology (a knowledge structure that describes the concepts, properties and relations between concepts) for QE approaches of the second category. Ontology-based QE approaches exploit the generalization, specialization or other relationships of ontology to extract meaningful words for expanding a query. A mixed-mode approach combines features of linguistic and ontology-based approaches; thus, expansion terms may be obtained from multiple knowledge structures.

Although semantic QE approaches are effective in improving retrieval performance, extensive review work for these approaches is not available. One notable effort is made by [4], which only reviews ontology-oriented semantic QE approaches, along with QE case studies. However, the focus of the paper is to compare ontological QE with relevance feedback QE. Wu *et al.* [5] present a synthesized model for ontology-based query expansion and has surveyed several ontology-based QE paradigms. The paper classifies QE approaches from only the ontology class; however, these techniques have not been systematically compared with one another.

The remainder of the paper is organized as follows: Section 2 discusses semantic QE and the contribution of knowledge structures in realizing semantics for QE is given in Section 3. Semantic QE approaches are classified in Section 4 and various approaches for semantic QE are reviewed. Section 5 presents a brief analysis of semantic QE approaches and evaluates them in terms of key features. Future directions for semantic QE are discussed in Section 6. Finally, the last section presents our conclusions.

II. SEMANTIC QUERY EXPANSION

The basic strategy of QE is to compute expansion terms that are relevant to the user intent and add them to the initial search query to optimize the IR system interpretation of the search query. Traditional QE approaches such as global analysis [6] and local analysis [7] utilize a corpus to expand the search query. The candidate expansion terms are extracted after statistical analysis of the corpus contents. According to [4], these corpus-driven statistical approaches perform well when a large corpus is available and the corpus content is relevant to the domain aspects of the search query. In contrast, semantic QE approaches do not have such limitations as they are based on corpus-independent knowledge structures (e.g., a lexical thesaurus or ontology).

Semantic QE softens the IR interpretation of the search query using the concepts of a knowledge structure. The expansion terms are obtained based on a similarity measure between the initial search query terms and the concepts of the knowledge structure, and are used to expand the search query with meaningful terms (concepts) that are closer to the user intent. One major advantage of semantic QE is that the knowledge structure is always available in the expansion mechanism, in contrast to the QE methods, such as relevance feedback [7], where the expansion process depends on the initial search results.

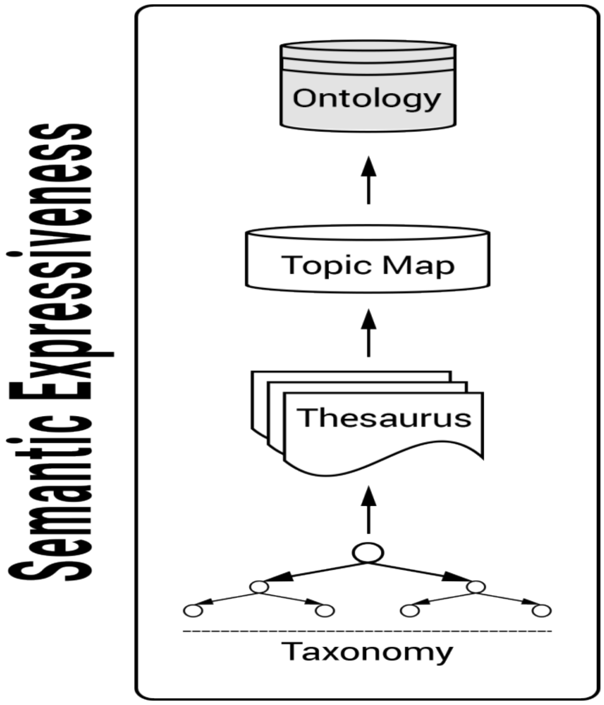

FIGURE 1. Types of knowledge structures.

The strategy of using a thesaurus or ontology for QE can be traced at least to the work of [8], where search query terms were manually disambiguated and used to retrieve the related concepts from an ontology. The main objective of the research was to demonstrate the usefulness of a knowledge structure in the QE process. Knowledge structures represent imperative sources for semantic QE. A knowledge structure is constituted by the relationships among concepts, from which the context (semantics) of a concept can be used to generate meaningful expansion terms. Efthimiadis [9] work classified corpus-independent knowledge structures into three categories: (1) domain-specific thesauri, which represent synonym or other related relationships between concepts of a domain; (2) general thesauri, such as WordNet [10], which are not limited to a specified domain; and (3) dictionaries such as the Collins dictionary [11]. According to strength of semantic expressiveness, [12] has identified four types of knowledge structures: taxonomies, thesauri, topic maps and ontologies. Figure 1 presents these types of knowledge structures in increasing order of semantic expressiveness. Ontologies are at the top of the model, as they have the highest level of semantic power, followed by topic maps, which represent association relations among topics. Thesauri, which show less semantic expressiveness than topic maps, can be described as structured vocabularies that represent the semantic relations of a certain domain, while taxonomies, which have the lowest level of semantic expressiveness, represent only hierarchical relations among the concepts of a domain.

III. IMPORTANCE OF KNOWLEDGE STRUCTURE FOR QUERY EXPANSION

As discussed previously, a knowledge structure represents a potential source of semantically related terms. A user query can be expanded with more comprehensible terms using knowledge-structure based QE. Biswas *et al.* [13] report that the gap between the search query and document space can be better filled with expansion terms that are generated from an ontology. Now, we discuss characteristics of knowledge structures that appear to be important for effective semantic QE.

A. SEMANTIC VOCABULARY

A knowledge structure can be domain-specific (describing the classification of a specified domain) or general (for instance, Cyc and EuroWordNet). Concepts, relations between concepts and the properties of concepts constitute the vocabulary of the knowledge structure, thereby leveraging the selection of semantically rich terms for QE. However, according to [14] and [15], the performance of QE is highly dependent on the quality of the vocabulary, for instance, its accuracy, completeness and up-to-date representation of knowledge.

For example, Dey *et al.* [16] expanded the search query using the vocabulary of two ontologies: plant and wine ontologies. The terms that have the shortest semantic distance between search query terms and ontology terms are selected. Results on the Google search engine demonstrated a 41.5% increase in precision for queries that were expanded with plant ontology terms and a 22% increase in precision with wine ontology terms.

B. TAXONOMY RELATIONS

Simple taxonomy relations, such as hypernym (Has-A) and hyponym (Is-A), enable the upward and downward traversal of the taxonomy for general categories and subcategories, respectively. The rationale behind using such relations in QE is to obtain more general concepts or specified concepts for search query terms from the knowledge structure, thereby expanding the query with closely related concepts. However, the selection of an appropriate hierarchical distance (for example, two or more levels from the original concept) for obtaining candidate expansion concepts from a knowledge structure remains a crucial problem [17]. According to [18], a conceptual distance of less than five levels is considered appropriate by users.

Consider as an example the work of [19], where each query is expanded with subcategories in the taxonomy. Improvement in the search results over short-length documents was demonstrated. Subcategories are also reported as satisfactory candidates for QE by [20] and [21], which demonstrated improvement in recall and precision measures. In performing experiments using the ProQuest thesaurus, [22] evaluated the usefulness of various thesaurus relations (such as synonym, descendent and ancestor) in QE. The experimental results demonstrated an increase in the recall measure for queries that were expanded with ancestor terms, but at the cost of a significant decline in the precision measure.

C. NON-TAXONOMIC RELATIONS

Rather than relying on hierarchical relations (e.g., is-a relations) in selecting expansion terms, a different type of QE approaches focus on non-taxonomic relations of the knowledge structure, such as synonym, troponymy, antonymy, part-of, semantic-role, dependence, typical-location, cause and telic-role relations [23]–[27]. Navigli and Velardi [28] found that expanding queries with non-taxonomic relations yields promising results.

According to [8], neighbor (same-level) taxonomy relations, such as synonym, antonym and related-concept relations, are effective in disambiguating query term senses and can be easily obtained from knowledge resources (for instance, dictionaries, thesauri or WordNet). Related-concept relations are also found to be very effective in obtaining expansion terms from an ontology by [29]. Word senses that are described within WordNet have been widely used by many researchers to successfully disambiguate the initial search query terms [30], [31]. Navigli and Velardi [32] suggest that words that are extracted from gloss relations in WordNet are better candidates for QE than hypernyms and hyponyms. They used gloss words to identify the semantic domain of search query terms and exploited the common nodes in the semantic domain to expand the search query. The research demonstrated improvement over baseline unexpanded queries.

IV. CLASSIFICATION OF SEMANTIC QE APPROACHES

The previous section describes how knowledge structures can be effectively used for the semantic expansion of search queries. According to the level of semantic expressiveness of the knowledge structures (see Section I), the semantic QE approaches can be divided into three main categories: linguistic approaches, ontology-based approaches and mixed-mode approaches. Each category can be further divided into subcategories according to key features. Figure 2 presents the general classification of these approaches, each of which is discussed in detail in the following subsections.

A. LINGUISTIC APPROACHES

Recent QE approaches focus on term dependency information, such as term morphology or related terms, in natural language to expand the initial search query. These are named linguistic QE approaches as they exploit natural language properties to generate the expansion terms; thus, they can be more effective in the disambiguation of word sense [33].

1) MORPHOLOGICAL EXPANSION

Morphological expansion approaches leverage morphological forms of query words (for instance, the stem, part of speech, adjectives, and progressive form of a word) to generate the expansion features. Bilotti *et al.* [34] expanded

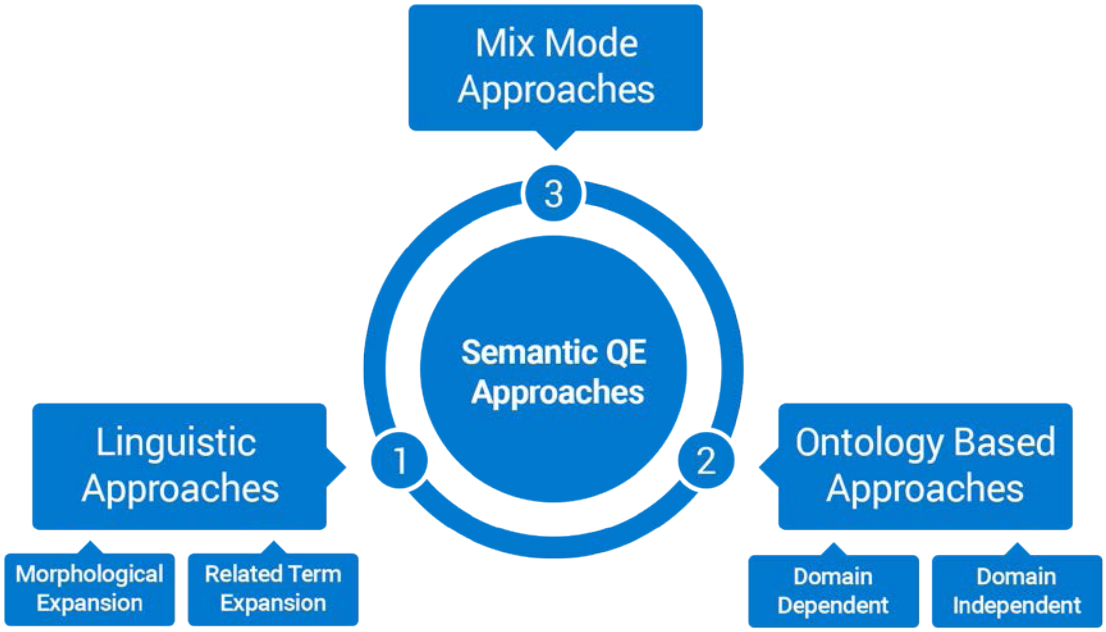

FIGURE 2. Classification of semantic QE approaches.

queries with morphological variation in a question-answering system and concluded that morphological expansion results in a high recall ratio. Experiments that were carried out by [35] using various languages corpora also demonstrated that expanding queries with morphological variants of query terms (extracted automatically from the documents) yields satisfactory retrieval performance.

Word stemming involves finding the root of a word via affix stripping; this information can be added to the initial search query for efficient information retrieval [36]. Many researchers have proposed stemmers, such as Paice/Husk [37], Krovetz [38] and Porter [39], which have been employed in most QE strategies and yielded substantial improvement in the recall and precision measures [40]–[44]. The benefit of using stemmers in QE has been analyzed by [45], in which the authors performed experiments using two stemmers, namely, Krovetz and Porter, and the MeSH thesaurus. QE experiments on the Porter stemmer yielded more accurate results in terms of the MAP measure, while the Krovetz stemmer performed well in terms of the R-Prec measure.

Identifying the parts of speech (POS) of query terms is another morphological approach for QE. POS taggers, for instance, Brill's tagger [46] and the log-linear tagger [47], are used to identify the morphological categories (e.g., noun, verb, or adjective) of words. Several researchers have exploited POS tagging in their QE approaches. Jayanthi *et al.* [48] proposed a semantic QE approach for personalized search in which a POS tagger is used to extract candidate expansion phrases using WordNet. Lu *et al.* [49] also assigned POS tags to query terms and extracted synonyms from WordNet for each term. To minimize the inclusion of inappropriate synonyms, the approach selects only synonyms that have the same POS tag. Experimental results demonstrated significant improvement in the precision and recall ratios.

2) RELATED-TERM EXPANSION

Related-term expansion approaches take advantage of synonymy and other semantically related word relationships to augment the search query. Knowledge structures such as dictionaries and thesauri are useful sources for extracting these word relationships based on the context of the search query. According to [50], the LDOCE dictionary and Roget's thesaurus are among the most commonly used knowledge structures for obtaining the senses of a word. Mandala *et al.* [51] investigated three types of thesauri, namely, hand-crafted, co-occurrence and head-modifier thesauri, for QE and reported increases in recall performance. In recent work by [52], QE is performed by using the multidisciplinary EuroVoc thesaurus. First, the search query is mapped with thesaurus terms. Then, parent, child, parallel and related relations for matched terms are used to expand the original query. QE using thesaurus relations yielded improved results for queries that previously provided no result.

WordNet is another well-known lexical database for deriving candidate word senses for QE. Both dictionary and thesaurus features have been combined into WordNet and Stevenson [53] argues that such a combination yields more comprehensive knowledge than the individual feature sets. George Miller and his team made a notable effort in developing an English language lexicon, namely, WordNet [54] which classifies words into four senses (nouns, adjectives, verbs and adverbs). Further, words that have the same meaning are grouped into sets, which are named synsets, whereby each synset has semantic relations with others (for instance, hyponym and meronym relations). Most research toward QE focused on extracting lexical relations from WordNet to add meaningful terms to search queries [30], [55]–[57]. In the following, we describe more sophisticated research for QE using WordNet.

The effect of using synsets from WordNet for QE has been evaluated by Song *et al.* [58]. They identified key phrases of the search query and matched them with the WordNet vocabulary to obtain the synsets. The most similar synsets are extracted based on search query terms and used to extend the query. Experimental settings that used the TREC corpus yielded improvements in the precision measure for the expanded queries. Nakade *et al.* [59] proposed an alternative semantic QE approach for obtaining relevant tweets using thesauri (from thesaurus.com) and a corpus of 35000 tweets. The approach uses thesaurus.com thesauri to calculate synonyms for original search topics and reformulate an expanded query expression by adding search topics and synonyms using parenthesis and OR operators. According to an evaluation of the queries, the overall retrieval performance was improved. Indeed, the method appears to be useful for searching for relevant tweets; however, it does not cover the expansion of query terms such as the name of a place.

B. ONTOLOGY-BASED APPROACHES

In these query expansion approaches, concepts from ontology are added to initial queries prior to the submission of the queries on the IR system. Concepts that represent the entities in a domain are semantically interconnected to form a knowledge structure, namely, an ontology. Thus, an ontology describes the traits of a domain. Weber [60] has classified ontologies into two main categories: those that describe the traits of a domain, which are named domain-dependent ontologies, and those that cover general traits of the real world (i.e., multiple domains), which are named domain-independent ontologies. Both categories attract attention from the IR community, especially in the field of QE, for extending search queries with contextual information [50]. In the following, we will discuss several works that utilize ontologies for query expansion.

1) QUERY EXPANSION USING DOMAIN-DEPENDENT ONTOLOGIES

Many domain-dependent ontologies exist, for instance, in the fields of agriculture, medicine, business, law, sport, the academy, annals and food. The DARPA Agent Mark-up Language (DAML) aims at maintaining a large repository of such ontologies (http://www.daml.org/ontologies/, accessed 20November 2018). A substantial amount of work on QE has been conducted on deriving contextual information from domain ontologies (e.g., [61]–[64]). This section discusses recent QE methods that employ domain-dependent ontologies to generate candidate expansion terms. The experimental results of these methods (as will be presented) demonstrate that domain-dependent ontology-based QE has a positive effect on the overall performance of IR systems.

Alromima *et al.* [65] implemented a QE mechanism that is based on a domain-specific ontology. The authors used Protégé to build an Arabic ontology and the SPARQL language to extract the candidate expansion terms from the ontology. In addition, a vector-space model based IR system was developed for retrieving the results from the Arabic dataset. The experimental results demonstrated improvements in both the precision and recall measures compared to simple keyword-based retrieval. Another example is [66], in which authors proposed a semantic QE approach that uses an ontology in the medical domain. The method constructed a hepatitis ontology using emerging natural languages and the MeSH knowledge base. This ontology was used to extract the expansion term set using semantic relationships including synonym, hypernym and related-word relationships. The overall performance of the system prototype was improved in terms of the precision and average precision ratios.

2) QUERY EXPANSION USING DOMAIN-INDEPENDENT ONTOLOGIES

Ontologies that are independent of any domain contain general-purpose vocabulary and relationships such as subtypes, related concepts and instances. These ontologies aim at covering the knowledge of multiple domains; thus, they can be a useful resource for semantic applications. Substantial efforts have been made to build these knowledge structures, such as OpenCyc [67], YAGO [68], Freebase [69], DBpedia [70] and UNIpedia [71]. As presented in the following, the QE paradigm has recently focused on utilizing domain-independent ontologies to generate candidate expansion terms.

A semantic query enrichment model was presented by Aggarwal and Buitelaar [72]. They used Wikipedia and DBpedia ontologies to obtain the candidate expansion concepts. Semantic relatedness among candidate concepts was computed via the explicit semantic analysis (ESA) method. Finally, the K best concepts in terms of high ESA score were selected to extend the search query. Experimental results on the CHiC dataset demonstrated high precision for the proposed QE model compared to manual relevance QE. Xiong and Callan [73] investigated the Freebase ontology in their QE paradigm. The QE method identifies topics (that correspond to real-world entities) that are related to query terms from Freebase. Topic relations within the knowledge graph of Freebase were examined to identify candidate expansion terms. The expansion terms were added to queries that were selected from the Bing search engine log. Using the Indri search engine and the ClueWeb09 corpus, the QE model was 20% more effective than the baseline QE strategy. Another QE model for image retrieval based on a domain-independent knowledge base was proposed by [74]. Semantically related expansion features were computed using comprehensive knowledge of the Cyc ontology. Overall, the performance of proposed system was significantly improved over the Bing search engine.

Despite successful attempts and improved results, several studies have demonstrated that QE using a domain-independent ontology does not always yield improved retrieval efficiency. The work in [4] and [8] identified two problems in query expansion with a domain-independent ontology: First, a general ontology includes ambiguous terms. Second, such an ontology is too general to capture the specialized properties of a domain. Accordingly, Hersh *et al.* [75] demonstrated that QE with a general-purpose ontology does not necessarily yield fruitful results.

C. MIXED-MODE APPROACHES

Mixed-mode QE approaches take the advantage of both linguistic and semantic QE techniques. Many researchers have examined the effects of combining lexicon (linguistic) and ontology (semantic) techniques in QE mechanisms (e.g., [76], [77]). According to [50], utilizing both lexicon and ontology techniques in the generation of expansion terms is beneficial, as the former provide knowledge on related terms and the latter provide context for each term.

One of the approaches in this category was investigated in [78], in which linguistic and semantic information were integrated in QE. The methodology used WordNet to capture the linguistic information, while semantic knowledge was acquired from the ResearchCyc (domain-independent) ontology. The contextual knowledge from both sources is used to generate the final extended query. Prototype results of the proposed methodology demonstrated improvement for web queries in terms of the relevance percentage. Recently, Devi and Gandhi [79] proposed a semantic query expansion model for the sports domain. The main steps of the proposed QE algorithm are as follows: (1) parse the search query, (2) extract synonyms from WordNet, (3) expand the query with words from a sports ontology, and (4) retrieve the results using the Google search API. The approach yielded noticeable improvement in terms of the precision and recall performance.

V. EVALUATION OF SEMANTIC QUERY EXPANSION APPROACHES

In this section, we assess the strengths and limitations of the major categories of semantic QE approaches. Then, we evaluate and compare various semantic QE approaches in terms of several criteria.

A. ASSESSMENT OF SEMANTIC QE APPROACHES

Table 1 lists distinctive advantages and disadvantages for the main classes of semantic QE approaches. Linguistic approaches are valuable for sense disambiguation of each query keyword; thus, they can be useful in the processing of natural language queries. Moreover, with a focus on term-based sense disambiguation, these techniques yield high values for the retrieval recall ratio. However, linguistic analysis is less effective in satisfying user information needs as it lacks semantic analysis and accurate domain identification of user search queries.

TABLE 1. Advantages and disadvantages of the main semantic QE approaches.

Semantic QE approaches	Advantages	Disadvantages
Linguistic approaches	• Effective processing of queries that are written in natural language • Provide terms that are similar to the query terms • Effective in word sense disambiguation • Higher recall ratio	• Lack semantic knowledge (e.g., semantic relationships) • Do not consider the domain of the query • Low precision ratio
Ontology-based approaches	• Provide contextual knowledge that is similar to the user query • Effective in search query disambiguation • User-centric results • More precise results	• Difficult to construct a comprehensive ontology • A domain-specific ontology is preferred over a domain-independent ontology • Require exact matching between a query term and an ontology concept
Mixed-Mode approaches	• Exploit the strengths of both linguistic and semantic approaches • Disambiguate word senses and the search query	• Number of expansion terms increases • Complex and increased number of operations

Recently, exploiting ontology (representing high quality semantics of a domain) in procedure of query expansion appears to be more successful in discovering the contextual data for search query. Ontology-based (whether domain-dependent or domain-independent) QE methods realize the search query semantic and are found to be more efficient in achieving more precise results. A shortcoming of this category of semantic QE is the construction and maintenance of ontology structure, since the poor ontology may degrade the query expansion performance. Furthermore, ontology-based QE approaches rely on exact matching between the query and the ontology vocabulary, which is a complex task for user queries that are written in natural language.

Rather than relying on either linguistic properties or ontology knowledge, an appealing class of semantic QE approaches (namely, mixed-mode QE approaches) exploits both in the procedure for query expansion. Linguistic analysis is useful in user-query sense disambiguation, while the use of an ontology enables the identification of query semantics. Mixed-mode semantic QE approaches are more efficient in the extraction of quality expansion features; however, they require more computations and may result in the generation of a large volume of expansion features (poor features may be included).

TABLE 2. Analysis of semantic approaches for query expansion in terms of key features.

Article	Semantic QE Strategy	Knowledge Structure	Experimental Setup		Baseline Model	Efficiency Measures*
			IR Model	Corpus		
[45]	Morphological Expansion	NLM stopword list, SMART stopword list, MeSH thesaurus	TF-IDF model, BM25 model	Cystic Fibrosis	No Expansion, Rocchio model	MAP, R-P, Dr
[49]	Morphological Expansion	WordNet thesaurus	Similarity score model	Java source code	No Expansion, Conquer model	Dr, P, R-P
[81]	Related Terms Expansion	Rogets thesaurus, WordNet thesaurus	Atire search engine	TREC ad-hoc retrieval tracks dataset	No Expansion, Rocchio model	MAP
[82]	Related Terms Expansion	WordNet thesaurus	Spectral based model	Corpus of 20 documents	No Expansion	MAP
[83]	Related Terms Expansion	WordNet thesaurus	Indri search engine	FIRE 2016 Microblog dataset	Semi-automatic expansion model	P, R, MAP
[65]	Domain dependent Ontology	Arabic Ontology	Vector space model	Arabic corpus	No Expansion, Stem based expansion	P, R
[84]	Domain dependent Ontology	Medical Ontology	Lucene model	Orthopaedic clinical reports	No Expansion	P, R, MAP, F-measure, R-P
[85]	Domain independent Ontology	DBpedia ontology	TF-IDF model	TREC AP88-90 dataset	No Expansion	P
[73]	Domain independent Ontology	Freebase ontology	Indri search engine	ClueWeb09 dataset	No Expansion, PRF model, Sequential model,	R, MAP, NDCG, ERR
[74]	Domain independent Ontology	Cyc ontology	Bing search engine	Image collection of search engine	No Expansion, Bing expansion	P, AP
[86]	Mix mode expansion	Wikipedia knowledge, DBpedia ontology	Terrier model	TREC2011 Microblog, TREC Robust 2004 dataset	No Expansion, PRF model	P, MAP
[87]	Mix mode expansion	Geographical Taxonomy, WordNet thesaurus	LSA package with R software	Web documents	No Expansion, Rocchio model	P. MAP

*The meanings of acronyms are: Dr = Number of relevant Documents Retrieved, P = Precision, R = Recall, R-P = Recall-Precision, AP = Average Precision, MAP = Mean Average Precision, ERR = Expected Reciprocal Rank, NDCG = Normalized Discounted Cumulative Gain

According to the advantages and disadvantages of each semantic QE category, as summarized in Table 1, every technique is valuable in certain circumstances. However, the selection of best semantic approach for QE requires consideration of following aspects:

– Term sense identification or query context identificatio;
– Ontology availability and accuracy;
– Query interface: natural language or keyword-based;
– Short or long querie;
– Retrieval requirements: precision or recal;
– Computational efficiency requirements.

B. EVALUATION OF SEMANTIC QE APPROACHES USING KEY FACETS

To evaluate major approaches for semantic QE, we have considered five operational facets: semantic QE strategy, knowledge structure, experimental setup, baseline model and efficiency measures. These facets have been adopted in a wide range of semantic QE methods. Table 2 lists recently published semantic QE approaches on the basis of these operational facets.

1) SEMANTIC QE STRATEGY

This facet identifies the class of semantic strategy for QE (as outlined in Section 3). As each semantic QE strategy uses a distinct approach to facilitate effective result retrieval, we list two or three published research systems for each type of semantic QE.

2) KNOWLEDGE STRUCTURE

All semantic QE strategies rely on knowledge structures for the extraction of additional conceptual data. Most of the semantic QE approaches exploit ontologies (domain-dependent or independent) as a knowledge source in their expansion procedure.

3) EXPERIMENTAL SETUP

In the experimental setting, the researchers have exploited a corpus and an IR system in the overall procedure of semantic QE. This facet reports various types of corpora that are used as test beds in the IR model to demonstrate the improvement of semantic QE performance. Among several combinations of corpora and IR models in the relevant literature, most of the research works utilize conventional search engines and various TREC datasets to evaluate the retrieval performance.

4) BASELINE MODEL

To assess the performance of a proposed approach, a reasonable strategy is to compare it with a similar model (called a baseline model). According to the results that have been published in the literature, the best baseline model is the 'no expansion' model (a run of the system without expanding the user query), while the pseudo-relevance feedback (PRF) model [7] and Rocchio model [80] are less commonly used as baseline models.

5) EFFICIENCY MEASURES

This facet investigates the metrics that are adopted by the research community to assess the efficiency of their semantic approach to expanding user queries. Among various efficiency metrics, precision, mean average precision and recall are the most widely accepted and used measures in semantic QE systems.

We summarize various semantic approaches for QE using key facets, rather than simply concentrating on identifying related expansion terms using the Wordnet thesaurus (i.e., ontologies). The use of this thesaurus has continued, in combination with other knowledge structures, in sophisticated techniques of semantic QE. In addition, TREC datasets have received a substantial amount of attention in current practice on semantic QE, mainly because they are the most typical source for evaluating the retrieval performance. Furthermore, most semantic QE works adopt the no expansion model (baseline) and variants of the precision metric to evaluate the improvement in system performance.

VI. FUTURE TRENDS FOR SEMANTIC QE

With the aim of facilitating information retrieval, semantic QE techniques acquire and restrict the number of expanded terms using knowledge structures. In this section, we discuss two important directions in semantic QE: personalized social semantic QE and multiple-ontology-based semantic QE.

A. PERSONALIZED SOCIAL SEMANTIC QE

Benefits of extracting semantics from an ontology have been realized by many QE systems. However, the conventional semantic QE techniques do not exploit the individual user search context (i.e., user profile or search history), which is necessary for identifying the correct context of a user query. Identifying the search context of a user query is important for two reasons: (1) the same search query by different users may convey dissimilar information (e.g., the user query 'apple' may refer to the fruit or the mobile device) and (2) each user's requirements may change over time. For instance, at one time, user may search for the apple fruit, while later, the user may search for mobile devices of the Apple brand. Thus, we must consider the user context (namely, personalize information) in future semantic QE to optimize the retrieval performance.

In additional to the user profile, social bookmarking systems (e.g., del.icio.us) and social networking sites, such as Twitter, Facebook and LinkedIn, can be useful sources for gathering personalize information on users. Recently, studies [88]–[91] have demonstrated the benefits of adding personalized social information to semantic QE. For instance, Zhu *et al.* [92] devised a novel QE framework for searching for high-quality tweets. The model combines semantic facets (i.e., nouns and verbs are extracted from tweets) and social behavior (e.g., number of followers and number of likes) to expand the user query. According to an experimental evaluation using Twitter data from TREC 2015 and several performance metrics, this approach outperforms classical personalize QE approaches. A similar approach of using personalized information in semantic QE is described in [93], whereby semantic classes are extracted from ontologies to classify images. Then, the system integrates ontological data and personalized data (captured from user click-through data) to retrieve the user-relevant images. The results demonstrate that the combination of semantic and personalized features yields better retrieval performance.

B. MULTIPLE-ONTOLOGY-BASED SEMANTIC QE

A recent trend in semantic QE is to utilize multiple ontologies for effective retrieval performance. The use of multiple ontologies facilitates the procurement of distinct semantic knowledge; thus, the multiple ontology-based QE technique performs well in obtaining valuable candidate expansion terms and outperforms single-ontology-based QE.

Several semantic QE strategies that use multiple ontologies have been proposed in recent literature [77], [94]. Zingla *et al.* [95] focus on integrating two knowledge structures, namely, Wikipedia and DBpedia, for calculating the candidate expansion terms. Seven QE methods are implemented and the QE method that uses both Wikipedia and DBpedia ontology outperforms other methods in terms of mean precision.

VII. CONCLUSIONS

Semantic QE is an advanced approach that explicitly considers the knowledge structure and improves the information retrieval performance. The objective of semantic QE is to enhance the QE procedure by identifying strategies that more effectively exploit semantic relationships within the knowledge structure (i.e., ontology). Currently, many semantic approaches for QE are available to cater to user information requirements. The present work classifies and discusses a spectrum of these methods, which have reported remarkable improvements in overall retrieval performance.

A common approach is linguistic semantic QE, which exploits natural language characteristics (e.g., morphological forms of words and synonyms of words) to generate terms as candidates for expansion. Of the linguistic techniques, the related-term expansion approach is more effective in the disambiguation of query terms since it uses thesauri or dictionaries to extract the senses for the user query. Methods in this category of QE realize high recall values, but at the cost of imprecise document retrieval (i.e., low precision). More advanced semantic QE approaches rely on an ontology rather than a thesaurus to obtain semantically related terms for expansion. Ontologies (either domain-dependent or domain-independent) are found to be most useful resources for describing traits (e.g., concepts, relationships among concepts, and constraints) in practice; thus, they can provide satisfactory expansion features. The most recent ontology-based QE effort focuses on the use of a domain-specific ontology (instead of a general ontology) and obtains results that are of not only high precision but also increased recall. However, the retrieval performance most highly depends on the quality of the ontology knowledge.

A recent trend to semantic QE is the development of mixed-mode strategies that combine linguistic and ontological knowledge to identify optimized expansion features. These methods yielded improved results on experimental benchmarks (i.e., TREC evaluation) and were found to be more robust in terms of performance on short and long user queries and the availability of knowledge structures.

In summary, a wide range of semantic QE approaches are available for effectively extending user queries and realizing improvements in average retrieval performance. This survey may act as a preliminary guide to modern approaches for semantic QE, as it bridges the literature deficits and gaps. In the future, the use of personalized social data and multiple ontologies in semantic QE methods must be further invested to maximize the retrieval performance.

REFERENCES

[1] G. W. Furnas, T. K. Landauer, L. M. Gomez, and S. T. Dumais, "The vocabulary problem in human-system communication," *Commun. ACM*, vol. 30, no. 11, pp. 964–971, 1987.

[2] E. Gabrilovich *et al.*, "Classifying search queries using the Web as a source of knowledge," *ACM Trans. Web*, vol. 3, no. 2, pp. 1–28, 2009.

[3] M. A. Raza, R. Mokhtar, and N. Ahmad, "A survey of statistical approaches for query expansion," *Knowl. Inf. Syst.*, pp. 1–25, Sep. 2018.

[4] J. Bhogal, A. MacFarlane, and P. Smith, "A review of ontology based query expansion," *Inf. Process. Manage.*, vol. 43, no. 4, pp. 866–886, 2007.

[5] J. Wu, I. Ilyas, and G. Weddell, "A study of ontology-based query expansion," Cheriton School Comput. Sci., Univ. Waterloo, Waterloo, ON, USA, Tech. Rep. CS-2011-04, 2011.

[6] Y. Jing and W. B. Croft, "An association thesaurus for information retrieval," presented at the 4th Int. Conf. Comput. Assist. Inf. Retr. Recherche d'Informations Assistee Ordinateur (RIAO), 1994.

[7] C. Buckley, G. Salton, J. Allan, and A. Singhal, "Automatic query expansion using SMART: TREC 3," presented at the 3rd Text Retr. Conf. (TREC), 1995.

[8] E. M. Voorhees, "Query expansion using lexical-semantic relations," presented at the 17th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., Dublin, Ireland, 1994.

[9] E. N. Efthimiadis, "Query expansion," *Annu. Rev. Inf. Sci. Technol.*, vol. 31, no. 5, pp. 121–187, 1996.

[10] G. A. Miller, "WordNet: A lexical database for English," *Commun. ACM*, vol. 38, no. 11, pp. 39–41, 1995.

[11] W. T. McLeod, *The New Collins Concise Dictionary of the English Language*. London, U.K.: Collins, 1988.

[12] M. Ullrich, A. Maier, and J. Angele, "Taxonomie, thesaurus, topic map, ontologie–ein vergleich V. 1.4," ontoprise GmbH, Karlsruhe, Germany, White Paper V.1.4, 2004.

[13] G. Biswas, J. Bezdek, and R. L. Oakman, "A knowledge-based approach to online document retrieval system design," in *Proc. ACM SIGART Int. Symp. Methodol. Intell. Syst.*, 1986, pp. 112–120.

[14] W. K. Young and H. K. Jin, "A model of knowledge based information retrieval with hierarchical concept graph," *J. Documentation*, vol. 46, no. 2, pp. 113–136, 1990.

[15] S. Jones, "A thesaurus data model for an intelligent retrieval system," *J. Inf. Sci.*, vol. 19, no. 3, pp. 167–178, 1993.

[16] L. Dey, S. Singh, R. Rai, and S. Gupta, "Ontology aided query expansion for retrieving relevant texts," in *Proc. Int. Atlantic Web Intell. Conf.*, 2005, pp. 126–132.

[17] A. M. Collins and E. F. Loftus, "A spreading-activation theory of semantic processing," *Psychol. Rev.*, vol. 82, no. 6, p. 407, 1975.

[18] S. Jones, M. Gatford, S. Robertson, M. Hancock-Beaulieu, J. Secker, and S. Walker, "Interactive thesaurus navigation: Intelligence rules OK?" *J. Amer. Soc. Inf. Sci.*, vol. 46, pp. 52–59, Jan. 1995.

[19] D. L. McGuinness, *Ontological Issues for Knowledge-Enhanced Search* (Frontiers in Artificial Intelligence and Applications). Washington, DC, USA: IOS-Press, 1998.

[20] J. Greenberg, "Optimal query expansion (QE) processing methods with semantically encoded structured thesauri terminology," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 52, no. 6, pp. 487–498, 2001.

[21] K. Järvelin, J. Kekäläinen, and T. Niemi, "Expansion Tool: Concept-based query expansion and construction," *Inf. Retr.*, vol. 4, pp. 231–255, Sep. 2001.

[22] J. Greenberg, "Automatic query expansion via lexical–semantic relationships," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 52, no. 5, pp. 402–415, 2001.

[23] N. V. Loukachevitch and B. V. Dobrov, "Development of ontologies with minimal set of conceptual relations," in *Proc. LREC*, 2004, pp. 1889–1892.

[24] R. Kolikipogu, B. P. Rani, and V. Kakulapati, "Reformulation of Telugu Web query using word semantic relationships," in *Proc. Int. Conf. Adv. Comput., Commun. Inform.*, 2012, pp. 774–780.

[25] L. Ramadier, M. Zarrouk, M. Lafourcade, and A. Micheau, "Spreading relation annotations in a lexical semantic network applied to radiology," in *Computational Linguistics and Intelligent Text Processing CICLing* (Lecture Notes in Computer Science), vol. 8403, A. Gelbukh, Ed. Berlin, Germany: Springer, 2014.

[26] M. Zarrouk, M. Lafourcade, and A. Joubert, "About inferences in a crowd-sourced lexical-semantic network," in *Proc. EACL*, 2014, p. 174.

[27] F. Ingrosso and A. Polguère, "How terms meet in small-world lexical networks: The case of chemistry terminology," in *Proc. Terminology Artif. Intell. (TIA)*, 2015, pp. 167–171.

[28] R. Navigli and P. Velardi, "A knowledge-based approach to ontology learning and semantic annotation," in *Proc. CAiSE Workshops*, vol. 3, pp. 189–203, Jun. 2004.

[29] A. Sihvonen and P. Vakkari, "Subject knowledge improves interactive query expansion assisted by a thesaurus," *J. Documentation*, vol. 60, no. 6, pp. 673–690, 2004.

[30] S. Liu, F. Liu, C. Yu, and W. Meng, "An effective approach to document retrieval via utilizing WordNet and recognizing phrases," in *Proc. 27th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2004, pp. 266–272.

[31] D. I. Moldovan and R. Mihalcea, "Using WordNet and lexical operators to improve Internet searches," *IEEE Internet Comput.*, vol. 4, no. 1, pp. 34–43, Jan. 2000.

[32] R. Navigli and P. Velardi, "An analysis of ontology-based query expansion strategies," in *Proc. 14th Eur. Conf. Mach. Learn., Workshop Adapt. Text Extraction Mining*, Cavtat, Croatia, 2003, pp. 42–49.

[33] C. Carpineto and G. Romano, "A survey of automatic query expansion in information retrieval," *ACM Comput. Surv. (CSUR)*, vol. 44, no. 1, p. 1, 2012.

[34] M. W. Bilotti, B. Katz, and J. Lin, "What works better for question answering: Stemming or morphological query expansion," in *Proc. Inf. Retr. Question Answering (IR4QA) Workshop SIGIR*, 2004, pp. 1–3.

[35] F. Moreau, V. Claveau, and P. Sébillot, "Automatic morphological query expansion using analogy-based machine learning," in *Proc. Eur. Conf. Inf. Retr.*, 2007, pp. 222–233.

[36] C. Moral, A. de Antonio, R. Imbert, and J. Ramírez, "A survey of stemming algorithms in information retrieval," *Inf. Res., Int. Electron. J.*, vol. 19, no. 1, p. n1, 2014.

[37] C. D. Paice, "Another stemmer," *SIGIR Forum*, vol. 24, no. 3, pp. 56–61, 1990.

[38] R. Krovetz, "Viewing morphology as an inference process," in *Proc. 16th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 1993, pp. 191–202.

[39] M. F. Porter, "An algorithm for suffix stripping," in *Readings in Information Retrieval*, J. K. Sparck and W. Peter, Eds. Burlington, MA, USA: Morgan Kaufmann, 1997, pp. 313–316.

[40] Y. Ogawa, H. Mano, M. Narita, and S. Honma, "Structuring and expanding queries in the probabilistic model," in *Proc. TREC*, 2000, pp. 541–548.
[41] L. Gan and H. Hong, "Improving query expansion for information retrieval using Wikipedia," *Int. J. Database Theory Appl.*, vol. 8, no. 3, pp. 27–40, 2015.
[42] B. He and I. Ounis, "Studying query expansion effectiveness," presented at the 31th Eur. Conf. IR Res. Adv. Inf. Retr., Toulouse, France, 2009.
[43] B. Hammo, A. Sleit, and M. El-Haj, "Enhancing retrieval effectiveness of diacritisized Arabic passages using stemmer and thesaurus," in *Proc. 19th Midwest Artif. Intell. Cogn. Sci. Conf. (MAICS)*, 2008, pp. 189–196.
[44] S. Abdou, P. Ruch, and J. Savoy, "Evaluation of stemming, query expansion and manual indexing approaches for the genomic task," *Cell*, vol. 501, p. 105, Nov. 2005.
[45] A. R. Rivas, E. L. Iglesias, and L. Borrajo, "Study of query expansion techniques and their application in the biomedical information retrieval," *Sci. World J.*, vol. 2014, Mar. 2014, Art. no. 132158. [Online]. Available: https://www.hindawi.com/journals/tswj/2014/132158/
[46] E. Brill, "A simple rule-based part of speech tagger," in *Proc. Workshop Speech Natural Lang.*, 1992, pp. 112–116.
[47] K. Toutanova, D. Klein, C. D. Manning, and Y. Singer, "Feature-rich part-of-speech tagging with a cyclic dependency network," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics Hum. Lang. Technol.*, vol. 1, 2003, pp. 173–180.
[48] J. Jayanthi, K. S. Jayakumar, and B. Akalya, "Personalized query expansion based on phrases semantic similarity," in *Proc. 3rd Int. Conf. Electron. Comput. Technol. (ICECT)*, Apr. 2011, pp. 273–277.
[49] M. Lu, X. Sun, S. Wang, D. Lo, and Y. Duan, "Query expansion via WordNet for effective code search," in *Proc. IEEE 22nd Int. Conf. Softw. Anal., Evol. Reengineering (SANER)*, Mar. 2015, pp. 545–549.
[50] V. C. Storey, A. Burton-Jones, V. Sugumaran, and S. Purao, "CONQUER: A methodology for context-aware query processing on the World Wide Web," *Inf. Syst. Res.*, vol. 19, no. 1, pp. 3–25, 2008.
[51] R. Mandala, T. Tokunaga, and H. Tanaka, "Combining multiple evidence from different types of thesaurus for query expansion," presented at the 22nd Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., Berkeley, CA, USA, 1999.
[52] S. Kapidakis, A. Mastora, and M. Peponakis, "Query expansion of zero-hit subject searches: Using a thesaurus in conjunction with NLP techniques," in *Proc. Int. Conf. Theory Pract. Digit. Libraries*, 2012, pp. 433–438.
[53] M. Stevenson, *Word Sense Disambiguation: The Case for Combinations of Knowledge Sources*. Stanford, CA, USA: Csli Publications, 2003.
[54] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller, "Introduction to WordNet: An on-line lexical database," *Int. J. Lexicography*, vol. 3, no. 4, pp. 235–244, 1990.
[55] Z. Gong, C. W. Cheang, and U. L. Hou, "Web query expansion by WordNet," presented at the 16th Int. Conf. Database Expert Syst. Appl., Copenhagen, Denmark, 2005.
[56] B. Sam and A. Shetty, "A hybrid framework to refine queries using ontology," *Indian J. Sci. Technol.*, vol. 8, no. 24, p. 1, 2015.
[57] C. H. Leung, Y. Li, A. Milani, and V. Franzoni, "Collective evolutionary concept distance based query expansion for effective Web document retrieval," in *Proc. Int. Conf. Comput. Sci. Appl.*, 2013, pp. 657–672.
[58] M. Song, I.-Y. Song, X. Hu, and R. B. Allen, "Integration of association rules and ontologies for semantic query expansion," *Data Knowl. Eng.*, vol. 63, pp. 63–75, Oct. 2007.
[59] V. Nakade, A. Musaev, and T. Atkison, "Preliminary research on thesaurus-based query expansion for Twitter data extraction," presented at the ACMSE Conf., Richmond, VA, USA, 2018.
[60] R. A. Weber, "Ontological issues in accounting information systems," in *Researching Accounting As an Information System Discipline*, V. Arnold and S. G. Sutton, Eds. Sarasota, FL, USA: Amer. Accounting Assoc., 2002, p. 21.
[61] H. Wang, Q. Zhang, and J. Yuan, "Semantically enhanced medical information retrieval system: A tensor factorization based approach," *IEEE Access*, vol. 5, pp. 7584–7593, 2017.
[62] M. Kamel, A. Lee, and E. Powers, "Ontology-aided vs. Keyword-based Web searches: A comparative user study," in *Proc. Large Scale Semantic Access Content (Text, Image, Video, Sound)*, 2007, pp. 639–652.
[63] M. Sodanil, P. Phonarin, and N. Porrawatpreyakorn, "An ontology-based query expansion for an agricultural expert retrieval system," presented at the Int. Conf. Inf. Integr. Web-Based Appl. Services, Vienna, Austria, 2013.
[64] L. Guo, X. Su, L. Zhang, G. Huang, X. Gao, and Z. Ding, "Query expansion based on semantic related network," in *Proc. Pacific Rim Int. Conf. Artif. Intell.*, 2018, pp. 19–28.
[65] M. W. Alromima, I. F. Moawad, R. Elgohary, and M. Aref, "Ontology-based query expansion for arabic text retrieval," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 8, pp. 223–230, 2016.
[66] C. Yunzhi, L. Huijuan, L. Shapiro, R. S. Travillian, and L. Lanjuan, "An approach to semantic query expansion system based on Hepatitis ontology," *J. Biol. Res.-Thessaloniki*, vol. 23, no. 1, p. 11, 2016.
[67] C. Matuszek, J. Cabral, M. J. Witbrock, and J. DeOliveira, "An introduction to the syntax and content of cyc," in *Proc. AAAI Spring Symp., Formalizing Compiling Background Knowl. Appl. Knowl. Represent. Question Answering*, 2006, pp. 44–49.
[68] F. M. Suchanek, G. Kasneci, and G. Weikum, "Yago: A core of semantic knowledge," presented at the 16th Int. Conf. World Wide Web, Banff, AB, Canada, 2007.
[69] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor, "Freebase: A collaboratively created graph database for structuring human knowledge," presented at the ACM SIGMOD Int. Conf. Manage. Data, Vancouver, BC, Canada, 2008.
[70] C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, and S. Hellmann, "DBpedia—A crystallization point for the Web of data," *J. Web Semantics*, vol. 7, pp. 154–165, Sep. 2009.
[71] M. Kalender, J. Dang, and S. Uskudarli, "Unipedia: A unified ontological knowledge platform for semantic content tagging and search," in *Proc. IEEE 4th Int. Conf. Semantic Comput. (ICSC)*, Sep. 2010, pp. 293–298.
[72] N. Aggarwal and P. Buitelaar, "Query expansion using Wikipedia and Dbpedia," presented at the CLEF Online Working Notes/Labs/Workshop, 2012.
[73] C. Xiong and J. Callan, "Query expansion with freebase," presented at the Int. Conf. Theory Inf. Retr., Boston, MA, USA, 2015.
[74] Y. He, Y. Li, J. Lei, and C. H. C. Leung, "A framework of query expansion for image retrieval based on knowledge base and concept similarity," *Neurocomputing*, vol. 204, pp. 26–32, Sep. 2016.
[75] W. Hersh, S. Price, and L. Donohoe, "Assessing thesaurus-based query expansion using the UMLS metathesaurus," in *Proc. AMIA Symp.*, 2000, p. 344.
[76] M.-H. Hsu, M.-F. Tsai, and H.-H. Chen, "Query expansion with conceptnet and WordNet: An intrinsic comparison," in *Proc. Asia Inf. Retr. Symp.*, 2006, pp. 1–13.
[77] S. K. Ray, S. Singh, and B. P. Joshi, "Exploring multiple ontologies and WordNet framework to expand query for question answering system," in *Proc. 1st Int. Conf. Intell. Hum. Comput. Interact.*, New Delhi, India, 2009, pp. 296–305.
[78] J. Conesa, V. C. Storey, and V. Sugumaran, "Improving Web-query processing through semantic knowledge," *Data Knowl. Eng.*, vol. 66, pp. 18–34, Jul. 2008.
[79] M. U. Devi and G. M. Gandhi, "Wordnet and ontology based query expansion for semantic information retrieval in sports domain," *J. Comput. Sci.*, vol. 11, no. 2, pp. 361–371, 2015.
[80] J. J. Rocchio, "Relevance feedback in information retrieval," *The Smart Retrieval System—Experiments in Automatic Document Processing*, G. Salton, Eds. Englewood Cliffs, NJ, USA: Prentice-Hall, 1971, pp. 313–323.
[81] R. Crimp and A. Trotman, "Automatic term reweighting for query expansion," presented at the 22nd Australas. Document Comput. Symp., Brisbane, QLD, Australia, 2017.
[82] M. Y. Dahab, S. Alnofaie, and M. Kamel, "A tutorial on information retrieval using query expansion," in *Intelligent Natural Language Processing: Trends and Applications*, K. Shaalan, A. Hassanien, and F. Tolba, Eds. Cham, Switzerland: Springer, 2018, pp. 761–776.
[83] W. Li, D. Ganguly, and G. J. F. Jones, "Using wordnet for query expansion: ADAPT FIRE 2016 microblog track," in *Proc. FIRE*, 2016, pp. 62–65.
[84] G. Aicha, B. Fatiha, B. Ghalem, and M. Farid, "Query expansion using medical information extraction for improving information retrieval in french medical domain," *Int. J. Intell. Inf. Technol.*, vol. 14, no. 3, pp. 1–17, 2018.
[85] S. Dahir, A. El Qadi, and H. Bennis, "Enriching user queries using Dbpedia features and relevance feedback," *Procedia Comput. Sci.*, vol. 127, pp. 499–504, Jan. 2018.
[86] M. A. Zingla, C. Latiri, P. Mulhem, C. Berrut, and Y. Slimani, "Hybrid query expansion model for text and microblog information retrieval," *Inf. Retr. J.*, vol. 21, no. 4, pp. 337–367, 2018.
[87] O. El Midaoui, B. El Ghali, A. El Qadi, and M. D. Rahmani, "Geographical query reformulation using a geographical taxonomy and WordNet," *Procedia Comput. Sci.*, vol. 127, pp. 489–498, Jan. 2018.

[88] C. Biancalana, F. Gasparetti, A. Micarelli, and G. Sansonetti, "Social semantic query expansion," *ACM Trans. Intell. Syst. Technol.*, vol. 4, no. 4, p. 60, 2013.

[89] D. Zhou, X. Wu, W. Zhao, S. Lawless, and J. Liu, "Query expansion with enriched user profiles for personalized search utilizing folksonomy data," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 7, pp. 1536–1548, Jul. 2017.

[90] S. Aarti and S. Anu, "Web semantics for personalized information retrieval," in *Information Retrieval and Management: Concepts, Methodologies, Tools, and Applications*. Hershey, PA, USA: IGI Global, 2018, pp. 795–810.

[91] W. Cui *et al.*, "Extended search method based on a semantic hashtag graph combining social and conceptual information," in *Proc. World Wide Web*, Jul. 2018, pp. 1–22.

[92] X. Zhu, J. Huang, B. Zhou, A. Li, and Y. Jia, "Real-time personalized twitter search based on semantic expansion and quality model," *Neurocomputing*, vol. 254, pp. 13–21, Sep. 2017.

[93] G. Deepak and J. S. Priyadarshini, "Personalized and enhanced hybridized semantic algorithm for Web image retrieval incorporating ontology classification, strategic query expansion, and content-based analysis," *Comput. Elect. Eng.*, vol. 72, pp. 14–25, Nov. 2018.

[94] R. Anand and A. Kotov, "An empirical comparison of statistical term association graphs with Dbpedia and conceptnet for query expansion," presented at the 7th Forum Inf. Retr. Eval., Gandhinagar, India, 2015.

[95] M. A. Zingla, L. Chiraz, and Y. Slimani, "Short query expansion for microblog retrieval," *Procedia Comput. Sci.*, vol. 96, pp. 225–234, Jan. 2016.

MUHAMMAD AHSAN RAZA received the M.S. degree in computer science from the Department of Computer Science, Bahauddin Zakariya University, Multan, Pakistan. He is currently pursuing the Ph.D. degree in computer science with the Department of Computer Systems and Software Engineering, Universiti Malaysia Pahang, Kuantan, Malaysia. He is also a Lecturer with the Department of Information Technology, Bahauddin Zakariya University. His main research interests include semantic web, information retrieval, and natural language processing.

RAHMAH MOKHTAR received a degree from The National University of Malaysia (UKM). She is currently a Senior Lecturer with the Faculty of Computer Systems and Software Engineering, Universiti Malaysia Pahang, Kuantan, Malaysia. She has actively participated in a number of research and commercialize projects in the domain of computer science. Her current research interests include information retrieval, human–computer interaction, and data processing and manipulation, such as big data and data mining related to artificial intelligence.

NORAZIAH AHMAD received the B.Sc. degree from Universiti Putra Malaysia and the Ph.D. degree from Universiti Malaysia Terengganu. She holds a Certificate in teaching and learning from Northern Illinois University, USA. She is currently an Associate Professor with the Faculty of Computer Systems and Software Engineering, Universiti Malaysia Pahang, Kuantan, Malaysia. Her current research interests include distributed databases, data grid and data mining, computational intelligence, cloud computing, and big data.

MARUF PASHA received the M.S. degree in IT from NUST, Pakistan, and the Ph.D. degree from the University of South Brittany, France. He is currently the Head of the Department of Information Technology, Bahauddin Zakariya University, Multan. His professional career consists of activities ranging from R&D, research fellowships, conferences' chair, and reviewer for several international journals and conferences. He has published a number of articles at national and international conferences and in journals. His research interests include semantic Web, big data, and the IoT systems' processing and manipulation, such as big data and data mining related to artificial intelligence.

UROOJ PASHA received the master's degree in industrial logistic and computer science, and the Ph.D. degree from the Molde University College, Norway, in 2015. She is currently an Assistant Professor with the Institute of Management Sciences, Bahauddin Zakariya University, Multan. She has published a number of articles in international journals.