IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received May 16, 2017, accepted June 11, 2017, date of publication June 29, 2017, date of current version July 24, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2718038

Merged Ontology and SVM-Based Information Extraction and Recommendation System for Social Robots

FARMAN ALI[1], DAEHAN KWAK[2], PERVEZ KHAN[3], SHAKER HASSAN A. EI-SAPPAGH[4], S. M. RIAZUL ISLAM[5], (Member, IEEE), DAEYOUNG PARK[1], AND KYUNG-SUP KWAK[1], (Member, IEEE)

[1]Department of Information and Communication Engineering, Inha University, Incheon 22212, South Korea
[2]Department of Computer Science, Kean University, Union, NJ 07083 USA
[3]Department of Electronics Engineering, Incheon National University, Incheon 406-772, South Korea
[4]Department of Information Systems, Minia University, Minya 11432, Egypt
[5]Department of Computer Science and Engineering, Sejong University, Seoul 143-747, South Korea

Corresponding author: Kyung-Sup Kwak (kskwak@inha.ac.kr)

This work was supported by the National Research Foundation of Korea-Grant funded by the Korean Government (Ministry of Science, ICT and Future Planning) under Grant NRF-2017R1A2B2012337.

ABSTRACT The recent technology of human voice capture and interpretation has spawned the *social robot* to convey information and to provide recommendations. This technology helps people obtain information about a particular topic after giving an oral query to a humanoid robot. However, most of the search engines are keyword-matching mechanism-based, and the existing full-text query search engines are inadequate at retrieving relevant information from various oral queries. With only predefined words and sentence-based recommendations, a social robot may not suggest the correct items, if items retrieved along with the information are not predefined. In addition, the available conventional ontology-based systems cannot extract precise data from webpages to show the correct results. In this regard, we propose a merged ontology and support vector machine (SVM)-based information extraction and recommendation system. In the proposed system, when a humanoid robot receives an oral query from a disabled user, the oral query changes into a full-text query, the system mines the full-text query to extract the disabled user's needs, and then converts the query into the correct format for a search engine. The proposed system downloads a collection of information about items (city features, diabetes drugs, and hotel features). The SVM identifies the relevant information on the item and removes anything irrelevant. Merged ontology-based sentiment analysis is then employed to find the polarity of the item for recommendation. The system suggests items with a positive polarity term to the disabled user. The intelligent model and merged ontology were designed by employing Java and Protégé Web Ontology Language 2 software, respectively. Experimentation results show that the proposed system is highly productive when analyzing retrieved information, and provides accurate recommendations.

INDEX TERMS Ontology, full-text-query mining, recommendation system, social robotics, information extraction.

I. INTRODUCTION

The use of humanoid robots is increasing rapidly in the application domains of healthcare, advertisements, and recommendations. They perform the role of conversational partner for patients in homes, receptionists in hotels to assist guests, or as advertisers in a shopping mall to promote products. The recent technology of human voice capture and interpretation has allowed the social robot to be employed as an interface to convey information and to provide recommendations. These technologies assist humans to extract information from the Internet by asking (issuing an oral query to) the social robot (e.g., find diabetes drugs for weight loss, find the best hotel in the city, or find the cheapest restaurant). Nevertheless, the constant increase in Internet data (webpages) creates problems for social robots trying to extract precise information for users. An oral command or query by disabled users is a familiar form of search-engine query to obtain relevant data from dynamic webpages. However, this oral command is not enough to acquire precise information about an explicit topic, because most search engines are based on keyword-matching mechanisms [1], [2].

Recommendation systems provide modified suggestions for different things, such as hotels, restaurants, movies, or books. Such recommendations are based on the information available in the system, and the system suggests items similar to favorable items mentioned in an oral query by users [3], [4]. However, the available recommendation systems are based on predefined words and sentences. A social robot may not recommend correct items if the items retrieved from the Internet are not predefined. This reduces the performance efficiency of information extraction (IE) and the recommendation system (RS). Sentiment analysis–based recommendations are needed in a robot platform to help a social robot examine the correctness of items by determining their popularity on the Internet. In the proposed sentiment analysis, reviews or information about items are retrieved, the retrieved data are mined to extract meaningful information and sentiments about the items' features, and the extracted sentiments are then matched with disabled users' needs. The proposed system presents the results in the form of polarity (positive, neutral, and negative) along with item names, links to webpages, and extracted sentiment information.

At present, regular ontologies are employed to extract data from imprecise information. A regular ontology addresses a predefined set of data, and is unable to retrieve needed data from the Internet. Consequently, a regular ontology can be combined with fuzzy logic to extract correct data. The integration of these approaches has been used in various systems to provide correct results from input queries [5]. However, Internet archived data are growing rapidly because of the sharing of information between various systems. The existing ontology-based systems can extract appropriate documents to a limited extent. Therefore, a support vector machine (SVM) with merged ontology is considered effective technology for extracting facts from the extensive, vague data on the Internet.

To overcome these problems, we propose a merged ontology and SVM-based information extraction and recommendation system for social robots. The proposed system assists disabled users to efficiently access current Internet information. The overall process contains four phases: disabled user interaction with the robot, oral query reformulation, pre-processing of retrieved documents, and SVM- and ontology-based precise IE. The reformulation phase of the oral query is accomplished in four steps, as follows.

- The system converts the oral command to a full-text query (FTQ).
- Morphological analysis of the FTQ identifies a lemma of the words.
- Filter out spaces, commas, and then tag parts of speech (POS).
- Extract nouns, adjectives, and adverbs.
- Convert the filtered data into a proper format for a search query.

In the pre-processing of retrieved documents phase, the system retrieves and tokenizes the information and reviews related items (e.g., city features, diabetes drugs, and hotel features) and then extracts the item features along with the information. The SVM is used to find feature reviews of items and texts that have sentiment words for the purpose of removing irrelevant information. The SVM classifies the retrieved sentences with sentiment words as relevant, and classifies other sentences as irrelevant. The merged ontology is then applied to extract the item with sentiment words from refined information to identify the polarity of the items for recommendation. In addition, description logic (DL) queries are employed to easily extract polarity terms about items with sentiment information. The rest of this research is structured as follows. Section II explains related works. Section III briefly demonstrates the overall process of the proposed system. The experimental results are illustrated in Section IV. Finally, Section V concludes our work.

II. RELATED WORK

The use of social robots and IE from the Internet is a hot topic in the field of recommendation and information engineering research. The main advantage of social robots is to help humans in daily life. At present, people are employing social robots in public for different purposes, including education, recommendations, and healthcare activities. For recommendations, the social robot is unable to provide precise information because of the system limitations. Most of these systems are based on predefined words and sentences. However, based on these predefined words and sentences, a social robot cannot extract efficient information from the hazy environment of the Internet. Therefore, extensive methodological work in text mining is needed to automatically get the FTQ from the social robot, re-treat the FTQ, and then retrieve the information based on the query. A variety of research has been done on this issue, and several ideas on text mining for social robots have been proposed [3], [4].

Unified communications and rich web interactions are employed to help a Neel robot connect with humans in an interpersonal manner in a shopping mall. The rich web interaction enables the robot to make links with users and help customers to get deals and offers. However, this system may not recommend a precise item because it trusts the data updated by sellers. Therefore, this system needs preference data. A coupon-giving robot system was developed for advertising in a shopping mall [3]. In this system, two tests show the efficiency of a recommendation: varying conversation schemas, and the existence of a robot. The presence of a robot strongly attracts customers. However, there is a limitation to rules for recommendations, and a category of words is used for speech recognition, which affects the accuracy of results. A text mining–based recommendation system was developed for human–robot interaction [7]. In this system, the robot communicates with the human to get the oral query, and then obtains the information employed in an external corpora.

This system recommends a movie according to a user's oral query. The idea of robot rules–based recommendation for retail shops was proposed [4]. This system tracks the customer's location and recommends precise products that are near them. The system needs semantic rules to recommend products. Human–robot interaction was examined by using various social activities in a hotel [8]. In this system, one robot detects and greets the customers, and another converses with body gestures, so the guest obtains information by talking. A cloud robotics service is employed in a smart city for emergency management operations [9]. This system uses small unmanned aerial vehicles (UAVs) as agents connected to the network. This network infrastructure allows the UAVs to benefit from storage resources and to control open data from common knowledge. A NAO robot and a smart pen are employed to improve social communications with dementia patients [10]. The NAO robot constantly monitors patient activity and assists the patient in daily life situations. In this system, patients can communicate with the NAO robot by using speech functionality. The NAO robot and multiple cameras are also employed for colored object recognition [11]. This system allows the NAO robot to receive oral commands from users to find a needed colored object. Fuzzy logic is employed to recognize the color based on the user's perception and multiple cameras to improve the quality of the recognitions.

At present, the increase in Internet data makes the recommendation debate more challenging. Most recommendation robot systems are based on predefined words and sentences, which may not allow the robot to extract precise information. Therefore, a full-text query can overcome this issue. The FTQ employs an existing search engine to obtain data about a particular topic or item. However, search engines use a keyword-matching mechanism, and are incapable of extracting the aim of the query from data on their servers [1]. To overcome this problem, an intelligent effort is needed to re-treat the user query and to obtain the required information from the intensively blurred environment of the Internet. A full-text search query is employed to extract information from the Internet [12]. This system showed that an FTQ search performs well when keywords of the query are used in the extracted documents. Its primary concern is precision, not recall. However, there are some limitations in the system; the existing search engines employ keyword-matching mechanisms. Therefore, this approach is inadequate at extracting appropriate data from the heterogeneous sources of Internet information. Currently, sentiment analysis–based recommendation has become a hot topic in research. There are two approaches to sentiment analysis: sentiment classification and feature-level sentiment analysis. The text is categorized (positive, neutral, and negative) in sentiment classification approaches, and the information is intended for manually characterizing the sentiment words [13]. In feature-level sentiment analysis, the features are described to extract sentiment words from text [14]. The idea of summarization and sentiment analysis was presented [15], which defines the feature sentiment (negative or positive) of a product by employing a lexicon-based method.

Recently, an ontology has been employed in the area of recommendations, IE, and sentiment analysis. IE is used to transfer natural language text into structured information (Daniel et al., 2008). This transformation is obtained by identifying relevant concepts, relationships, and instances. However, natural language has ambiguity (single words can have many meanings). Therefore, the process of IE is a difficult task. Ontology-based IE overcomes this difficulty by organizing the domain knowledge through an ontology. An ontology is a shared conceptualization of a specific domain through concepts, instances, and relationships, which is in a human-understandable and machine-readable format [1], [16]. A regular ontology is applied to find features in movie reviews [17]. This ontology is suitable for the extraction of a planned set of data. Nevertheless, Internet data are imprecise in structure. Consequently, a regular ontology is inadequate for describing the fuzzy terms of features (e.g., city {clean, average, and dusty}). A regular ontology with fuzzy logic works remarkably well when the input is uncertain. Opinion categorization of online item reviews was suggested to measure the linguistic hedges on sentiment labels [18]. The system automatically extracts opinion phrases from reviews and categorizes the reviews in terms of positive and negative. Moreover, the sentiment scores are stored in linguistic variables and presented in table format. An ontology is employed to store all variables with an opinion score, and to provide a knowledgebase platform for the classification of feature polarity. Opinion mining based on an ontology was suggested to categorize and examine online reviews [19]. The system shows that parts of speech can be different, which leads to ambiguous analysis, and decreases the accuracy of the review classifier. The system employs an SVM as an opinion analyzer to compute the precise measure of words for sentiment analysis.

Most of the existing research has its limitations in recommendations, IE, and sentiment analysis. Mostly, the recommendation and IE systems are based on predefined words and a regular ontology, respectively. It is a fact that predefined words–based systems are unable to recommend the correct item, and a regular ontology cannot extract the anticipated result from a blurred resource of data. To the best of our knowledge, the proposed merged ontology and SVM-based recommendation and IE is a first effort to automatically retrieve information and extract the meaning of the data for disabled users. This system can extract information related to hotels, the city, and diabetes drugs. In the proposed system, a NAO robot communicates with disabled user to get the oral query; the oral query is then mined to extract the disabled user's needs, and it is then converted to an appropriate form for an information search. Additionally, the proposed merged ontology provides fuzzy logic and semantic web rule language (SWRL)-based semantic knowledge for feature polarity computation. This merged ontology contains a medical ontology, a city ontology, and a hotel ontology, which easily provide information for recommendations.

III. MERGED ONTOLOGY AND SVM-BASED RECOMMENDATION SYSTEM

In this section, the merged ontology and SVM-based recommendation system is presented along with the architecture. This system assists a robot to extract exact information for disabled users from various websites. In the architecture, the mentioned robot (NAO) is already available. NAO is a humanoid robot, which was made by Aldebaran Robotics. Some other robots have been developed to interact with people in different contexts, or to assist disabled users in indoor and outdoor environments. However, limited work has been done to facilitate disabled people in terms of precise information extraction from the intensively vague environment of the Internet. According to World Health Organization (WHO) estimates, some important facts regarding people with impairments, disabilities, and disease are as follows [20], [21].

- The visually disabled number 314 million, and 45 million of that total are completely blind. Most are elderly.
- Every year, 250,000 to 500,000 people suffer a spinal cord injury due to a traffic accident, from falls, from bomb blasts, etc. The symptoms of this injury are loss of motor control in the legs, arms, and body, which increases the number of disabled and disabled people.
- According to a World Bank 2012 estimate, more than one billion people facing some form of disability from disease (15% of the world's population). Among these, 110 million to 190 million children have major difficulties moving from one place to another.
- About 90% of the people with impairment or disabilities from disease are unable to globally extract exact information from websites according to their needs.

These disabled users are unable to access Internet information efficiently. Therefore, a robot with a smart information extraction system is needed to communicate with these disabled users through voice, sign language, or specific gestures to extract precise information according to their wishes. The architecture of the proposed system is shown in Fig. 1. The internal process of the proposed architecture contains the following segments.

- The disabled user's interaction with the robot.
- Full-text query reformulation (FTQR).
- Query storage and allocation to the web crawler and application programing interfaces (APIs).
- Pre-processing of cloud text data.
- Merged ontology and SVM-based precise information extraction.
- Output of the FTQ.

A. THE DISABLED USER'S INTERACTION WITH THE ROBOT

The purpose of a robot implementation is to consistently assist disabled users in terms of information retrieval from websites. Interestingly, the NAO robot is fitted with various languages, which allows it to communicate with many people. In this system, the microphones and camera of the robot are employed for voice recognition and movement detection, respectively. The microphone is automated to sense the commands from disabled users that start or stop application of the information search engine. However, this robot becomes idle when it receives a command because of several applications and ontologies used in the robot. In our case, we give a command to NAO to start or stop the information search with specific semantic knowledge. When the NAO robot receives a command, it will call the ontology-based information search engine and will wait for an oral query from the disabled user. These queries are: "NAO, start/stop the search engine with hotel knowledge" or "NAO, start/stop the search engine with medical knowledge" or "NAO, start/stop the search engine with city knowledge." After the success of a start command, the robot receives an oral query from disabled users, which requires it to find the precise information in websites. This oral query is converted to a text query and then reformulated for the keywords-based search engine. After the conversion of the user query, it will be stored for further processing. The task of this robot is to communicate with disabled users, find the optimal information, and then respond to the disabled or disabled user. Examples of dialogue and commands are in the following form:

Disabled User (DU): "NAO, start the search engine with semantic knowledge."

N: "OK, the search engine has been started with semantic knowledge; would you like to search for something?"

DU: "Yes" (optional)

DU: "I would like to visit Seoul for three days at the end of September. I am a diabetes patient, so I need to find a hotel with the best healthcare and a centrally located restaurant in Seoul. I will also need dietary food and diabetes drugs for weight loss"

(This is a full-text query that needs to be pre-processed.)

N: "OK, I am searching."

N: "Information is retrieved; can I read it?"

DU: "Yes"

N: "Hilton Hotel in Seoul; Victoza drug for weight loss; and Branson Travel Agency for hotel booking."

N: "Seoul is expensive and very crowded."

N: "Website links are retrieved; would you like to visit them?"

DU: "No; find the polarity."

N: "OK."

N: The polarity of the diabetes drug is positive; the polarity of the Hilton Hotel is positive; and the polarity of Seoul is negative."

When a humanoid robot identifies an oral query from disabled user (Task 1 in Fig. 2), it activates the information retrieval search engine and waits for the FTQ (tasks 2 and 3 in Fig. 2). In this search query, the users specify their need for an information search about diabetes drugs, hotel reservations, or the city, accordingly. After getting the query,

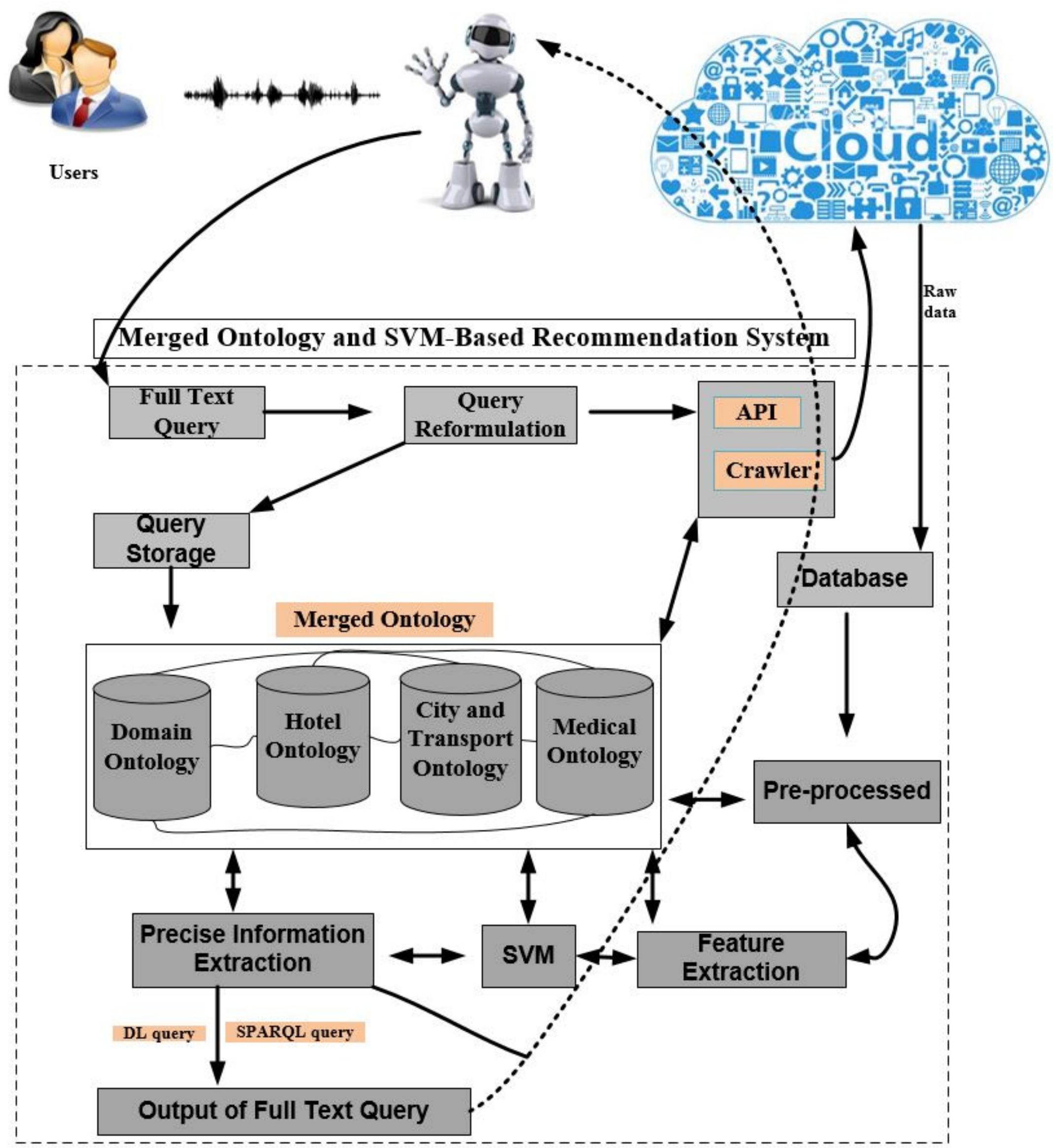


FIGURE 1. Architecture of the proposed system.

NAO assigns it to the FTQR module for reformulation and then submits it to the proposed system (tasks 4 and 5 in Fig. 2). The proposed system employs a web crawler (www) and e-commerce sites (hotels.com, tripadvisor.com, and drugs.com), which offer APIs to download user comments for proper drugs, hotels, or cities. The reformulated query is placed in query storage, and assigned to a merged ontology for further usage. The crawler of the proposed system also retrieves the feature information and reviews of the hotel and city (e.g., parks, hospitals, rooms, and so on). After reviews are downloaded, they are pre-processed to split the review sentences, remove stop words, and tag POS. The advanced classifier SVM is employed to classify the reviews and to remove irrelevant reviews. This classifier increases the precision rate, which uses a linearly separable hyperplane. Feature extraction is based on a merged ontology. Those features that are declared in the merged ontology will be extracted along with features from unstructured reviews. The merged ontology is then used to find sentiment words, which can be employed for feature polarity of the hotel, the city, and the diabetes drugs. The merged ontology is a combination of four ontologies. These ontologies are a hotel ontology, a city and transport ontology, a medical ontology, and a domain ontology, which capture information about hotels, cities, and diabetes drugs. This merged ontology combines all that captured information with polarity scores for precise information extraction. The response to the person is based on the results of the DL queries. Finally, the NAO robot reads the output of these queries, along with precise extracted information, to the disabled user (Task 6 in Fig. 2).

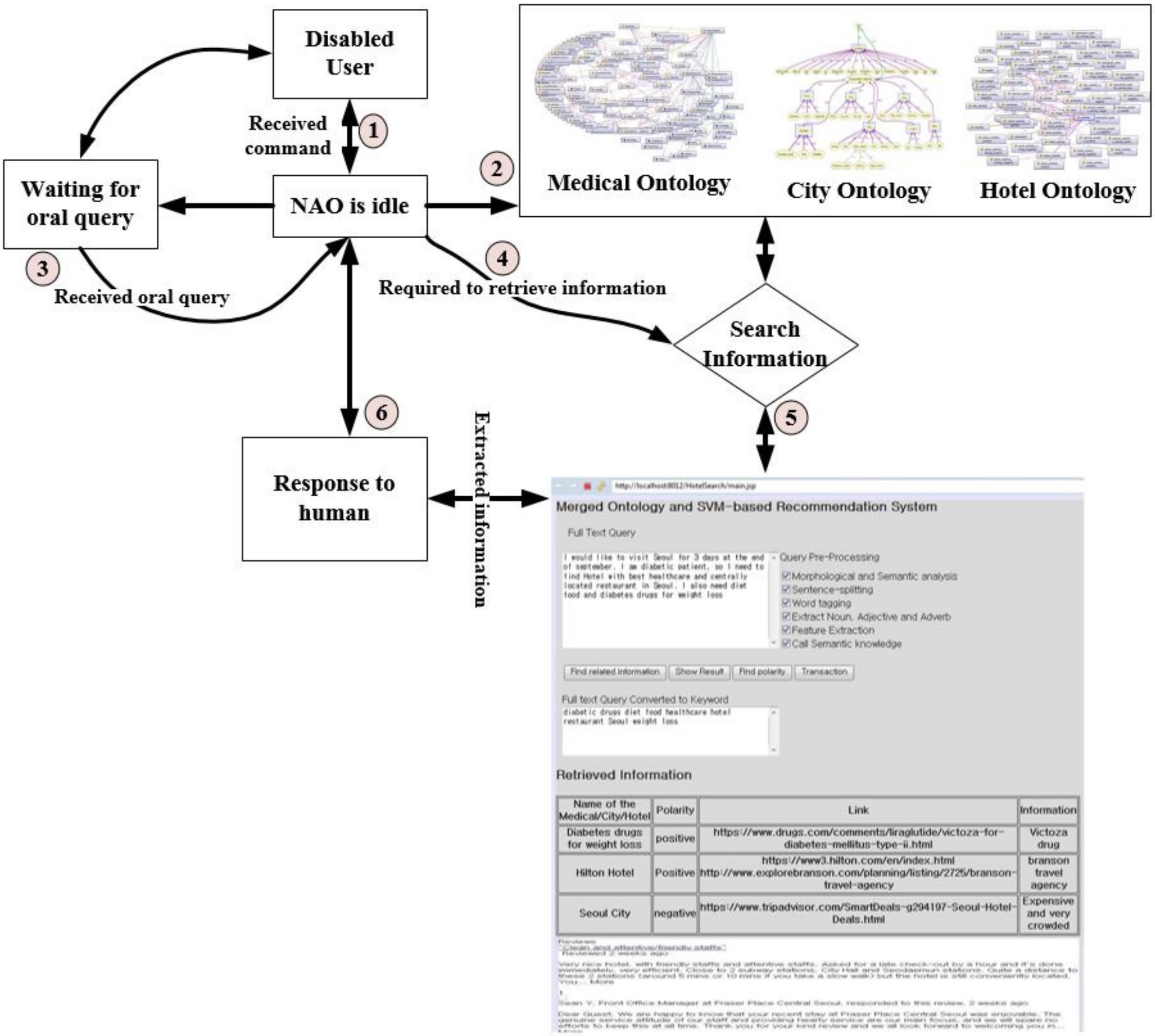


FIGURE 2. State transition diagram of information extraction by the NAO robot.

B. FULL-TEXT QUERY AND REFORMULATION AND QUERY STORAGE

The Google search engine (GSE) is based on keywords. Therefore, a full-text user query is reformulated to produce a more proper and exact query for information extraction. This reformulation method contains four steps, which are performed in a series. The first step is morphological analysis. It is a core part of natural language processing (NLP), which finds different forms of words by employing a lexicon. The lemmatization process with morphological and semantic analysis is applied to the FTQ to identify the lemma of query words. For example, the disabled user query is "I would like to visit Seoul for three days at the end of September. I am a diabetes patient, so I need to find a hotel with the best healthcare and a centrally located restaurant in Seoul. I will also need dietary food and diabetes drugs for weight loss." Every word of the input query is converted into basic forms, which are their lemma. These lemmas are employed as words in the new query. After lemmatization, the semantic analysis uses WordNet to resolve the uncertainties in the meaning of these words. WordNet chooses the suitable meaning for keywords in the user query. The second step is sentence splitting. This is a procedure to separate a composite sentence into small tokens. Disabled user query may have commas and white spaces. The splitting task filters out commas and spaces from the query. After using the splitting function, the proposed system acquires the query in the form of tokens: I, would, like…weight, loss.

The output of the splitting procedure is then stored in array format for further processing. The third step is word tagging. This is the process of POS tagging, which tags each word of the disabled user query, as shown Fig. 3. The tagging and

I/PRP would/MD like/VB to/TO visit/VB Seoul/NNP for/IN 3/CD days/NNS at/IN the/DT end/NN of/IN September/NNP ./.

I/PRP am/VBP diabetic/JJ patient/NN ,/, so/RB I/PRP need/VBP to/TO find/VB hotel/NN with/IN best/JJS healthcare/NN and/CC centrally/RB located/VBN restaurant/NN in/IN Seoul/NNP ./.

I/PRP will/MD also/RB need/VB diet/NN food/NN and/CC diabetes/NN drugs/NNS for/IN weight/NN loss/NN

FIGURE 3. POS tagging of the input query.

parsing process is very complex, because words may alter their meaning when in different positions in a sentence. Nevertheless, generalized architecture text engineering (GATE) can achieve this tagging appropriately [22]. The fourth step is query enrichment. This process removes the pronouns, articles (a, an, the), and prepositions to create an appropriate search query from the FTQ. The output from query enrichment is central, restaurant, hotel, healthcare center, restaurant, Seoul, dietary food, diabetes drugs, weight loss. This algorithm identifies nouns and verbs from the created query array and categorizes them grammatically to make a novel query for information extraction about diabetes drugs, hotels, and the city. The output with query enrichment words has a distinctive meaning in WordNet. Therefore, semantic analysis negotiates with the merged ontology to recreate the query and randomize the query words, such as central hotel Seoul, diabetes drugs, weight loss, and dietary food restaurant. These reformulated queries are also kept in query storage to deliver additional data about disabled user needs.

C. WEB CRAWLER AND APPLICATION PROGRAMING INTERFACES

Precise information extraction from various websites is always a tough task in the domain of information engineering. A web crawler browses the webpages automatically. Our proposed system uses a focused web crawler that searches the webpages and then indexes webpages that are relevant to a particular domain. The reformulated queries are allocated to web crawlers and APIs to search for specific hotels, restaurants, cities, and diabetes drugs, and retrieve the consumer reviews. GSE offers a list of general webpages and the rank-list results according to a given query. The results of the user queries are downloaded in HTML format from their URLs, avoiding Word and PDF files. Each URL characterizes its own HTML text as a superclass. In an ontology-based web crawler, we describe target URL forms for links to track, which holds those webpages that contain information about cities, diabetes drugs, and hotels. This web crawler uses rich site summary (RSS) to supervise HTML pages and update information frequently. Mostly, URLs comprise inappropriate links for retrieving data about a city, diabetes drugs, and hotels. Therefore, the proposed web crawler is limited to following only relevant links. The crawler usually checks many intermediary webpages, but chooses preferential webpages to retrieve information, along with customer reviews. The system also extracts metadata, such as the scores of reviewers, and stores them in a database.

TABLE 1. Example of retrieved text.

I tried Victoza in May 2016. Today, I have lost 45 pounds. The side effects are minimal, but they can be controlled mostly by diet. I joined a free tour and learned about Seoul. Seoul is clean, at least in the places where we visited, but expensive and very crowded. The rooms of the Hilton Hotel are clean but small, and it is between good and cheap restaurants.

D. PRE-PROCESSING OF RETRIEVED DOCUMENTS AND FEATURE EXTRACTION

A collection of documents is obtained from the database. These documents are pre-processed to remove unnecessary content (such as tags, names, and dates). The text of these documents is tagged to identify words and POS employing NLP. The tagging process is explained in Section III-B. After POS tagging, the sentences of the retrieved documents are reduced to increase the accuracy of search features and eliminate determiners (the, a, an), prepositions (e.g., on, in, of), and stop words [23]. These sentences are divided to obtain complete clauses, which comprise nouns and verb phrases. In this step, we are assured that every sentence of the document is a complete passage with a noun and a verb phrase. In the mentioned sentences in Table 1, a comma and connective words are employed to distinguish the sentence. For example, the first sentence "I tried Victoza in May 2016" is a complete sentence. Therefore, it does not need the splitting process. Another sentence "The side effects are minimal, but it can be controlled mostly by diet" has verbs and conjunctions. Consequently, it needs splitting to find complete clauses [24]. Similarly, the overall retrieved text will go through the splitting process to identify complete clauses for further processing. After the splitting process, feature extraction from complete clauses will be achieved by picking noun phrases. A merged ontology defines the concepts of city, hotel, and diabetes drugs features and their relationships with those concepts. The information from item feature extraction and categorization is imported into the merged ontology. It identifies item features in sentences, such as Victoza, park, restaurant, and hotel. The extracted features are checked with the classes of the merged ontology. The matched features will be examined to compute polarity; otherwise, they will be removed.

E. TEXT CLASSIFICATION BASED ON THE SVM

An SVM is a supervised learning technique used for classification [25]. It finds the best possible boundary to separate the relevant samples from irrelevant samples. The main aim of an SVM is to find a maximum margin hyperplane behind the training process to resolve the classification task. Chung developed a well-known and standard library for SVM, called LIBSVM [26]. LIBSVM is merged software for regression, distribution inferences, and support vector classification. It categorizes multi-classes using two steps. First, the SVM makes a training data set model, and then employs that model to achieve facts for a testing data set. After pre-processing retrieved documents, the following is used for text categorization [27]:

$$\emptyset : (D_S \times C_S) \rightarrow \{R, IR\} \tag{1}$$

In the above expression, D_S and C_S refer to sets of documents and categories, respectively. If $\emptyset : (D_S \times C_S) = R$, then D_{S_i} is a relevant member of C_{S_j}, and D_{S_i} is an irrelevant member of C_{S_j} if $\emptyset : (D_S \times C_S) = IR$. The machine learning algorithms are employed to categorize the text or review sentences. The SVM function gives a value of (+1) for the most relevant data points, and a value of (-1) otherwise. Moreover, the training data are mapped by using a non-linear mapping function as follows:

$$\emptyset : R^N \rightarrow R^F \tag{2}$$

Here, R^F is a feature space, and its training data are shown by non-linear mapping, R^N. Therefore, the equation that splits the data set is as follows:

$$\begin{aligned} \mathcal{L}(w, b, \alpha, \xi, \gamma) &= \frac{||w||^2}{2} + c\sum_{i=1}^{n} \xi_i - \sum_{i=1}^{n} \alpha_i \left|y_i (wx_i + b) - 1 + \xi_i\right| \\ &\quad - \sum_{i=1}^{n} \gamma_i \xi_i \end{aligned} \tag{3}$$

In the above equation, $\alpha_i \geq 0, \gamma \geq 0, i = 1, 2, \ldots .n$ are Lagrange multipliers. If $y_i (wx_i + b)$ is 1, the support vector control by the parameter is $y_i (wx_i + b) \in \{0, 1\}$ [27]. The symbol ξ denotes misclassification errors. This support vector categorizes the decision boundary, which is a subset of the training vectors. The decision function for non-zero optimal solutions can be expressed as follows:

$$f(x) = sign\left\{\sum_{i=1}^{n} \alpha_i y_i (x_i \cdot x) + b\right\} \tag{4}$$

where b is the subset of points x_i, and the solution to $\alpha_i |y_i (wx_i + b) - 1|$ will lie for any non-zero α_i next to the hyperplane. The duality of a classical Lagrangian allows solving the primal problem. Therefore, the dual optimization problem of equation 3 is illustrated as follows:

$$max\left[\sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n} \alpha_i \alpha_j y_i y_j k\left(x_i, x_j\right)\right] \tag{5}$$

In the above expression, $\mathrm{k}\left(\mathrm{x_i}, \mathrm{x_j}\right) = (\emptyset(\mathrm{x_i}) \cdot \emptyset(\mathrm{x_j}))$ is the kernel function (KF) based on the dot product. It performs non-linear mapping into the feature space. When the data are nonlinearly separable, the KF grants more decision functions. The KF employed Gaussian radial-basis function (RBF) and polynomial function are as follows:

$$\mathrm{k}\left(\mathrm{x_i}, \mathrm{x_j}\right) = \left(\left(\mathrm{x_i}.\mathrm{x_j}\right) + 1\right)^{\mathrm{d}} \tag{6}$$

$$\mathrm{k}\left(\mathrm{x_i}, \mathrm{x_j}\right) = \exp\left\{\frac{||\mathrm{x_i} - \mathrm{x_j}||^2}{2\sigma^2}\right\} \tag{7}$$

The classifier function from equations 4 and 5 can be demonstrated as follows:

$$\mathrm{f}(\mathrm{x}) = \mathrm{sign}\left\{\sum_{\mathrm{i}=1}^{\mathrm{n}} \alpha_{\mathrm{i}} \mathrm{y_i} \mathrm{k}(\mathrm{x_i} \cdot \mathrm{x}) + \mathrm{b}\right\} \tag{8}$$

where b is the solution to $\alpha_{\mathrm{i}}\left[\mathrm{y_i}\sum_{\mathrm{i}=1}^{\mathrm{n}} \alpha_{\mathrm{i}} \mathrm{y_i} \mathrm{k}(\mathrm{x_i} \cdot \mathrm{x}) + \mathrm{b} - 1\right] = 0$ for non-zero α_{i}.

The LIBSVM is based on RBF kernels. It is employed to build models that predict information for a testing data set. After the classification of text or reviews, a merged ontology and fuzzy logic are employed to extract precise information and evaluate the item polarity, respectively.

F. MERGED ONTOLOGY-BASED POLARITY PREDICTION FOR RECOMMENDATION

An ontology is employed to share information about a specific domain between people and systems. Our proposed ontologies focus on the concepts of the domain, individuals, their properties, and relationships. These are designed using the Protégé Web Ontology Language (OWL). At first, these ontologies were made as a classic ontology, and a fuzzy OWL plug-in was then employed to convert these classic ontologies into a fuzzy ontology [28]. The concepts, instances, and axioms of a classic ontology and a fuzzy ontology are the same. However, the concept values of a classic ontology are blurry terms, which are unable to handle uncertainty. A fuzzy ontology contains ambiguous information using fuzzy concepts to handle any type of condition about opinion word uncertainty. Therefore, this approach needs fuzzy ontology–based semantic knowledge to classify the user reviews related to diabetes drugs, the city, and hotels. Useful information on the city, diabetes drugs, and hotels is accumulated and delivered to a merged ontology for recommendation and information extraction. Mathematically, the ontology can be expressed as follows:

$$\widetilde{\mathrm{Ontology}} = (\mathrm{C}, \mathrm{P}, \mathrm{R}, \mathrm{V}, \mathrm{V_c}) \tag{9}$$

where C, P, R, V, and $\mathrm{V_c}$ stand for concepts, properties, relationships of classes, values, and constraint values of properties, respectively [4]. Fuzzy set theory defines imprecise boundaries, such as positive, negative, and neutral. A fuzzy set, F, over the universe of set A is characterized by its membership function, μ_{F}, which shows an element A in the interval [0, 1] [29], [30]:

$$\mu_{\mathrm{F}}(\mathrm{A}) : \mathrm{A}' \rightarrow [0, 1] \tag{10}$$

Here, A' belongs to A, and μ_{F} shows the membership degree by which $\mathrm{A}' \in \mathrm{A}$. If $\mu_{\mathrm{F}}(\mathrm{A}) = 1$, then A' will be considered a

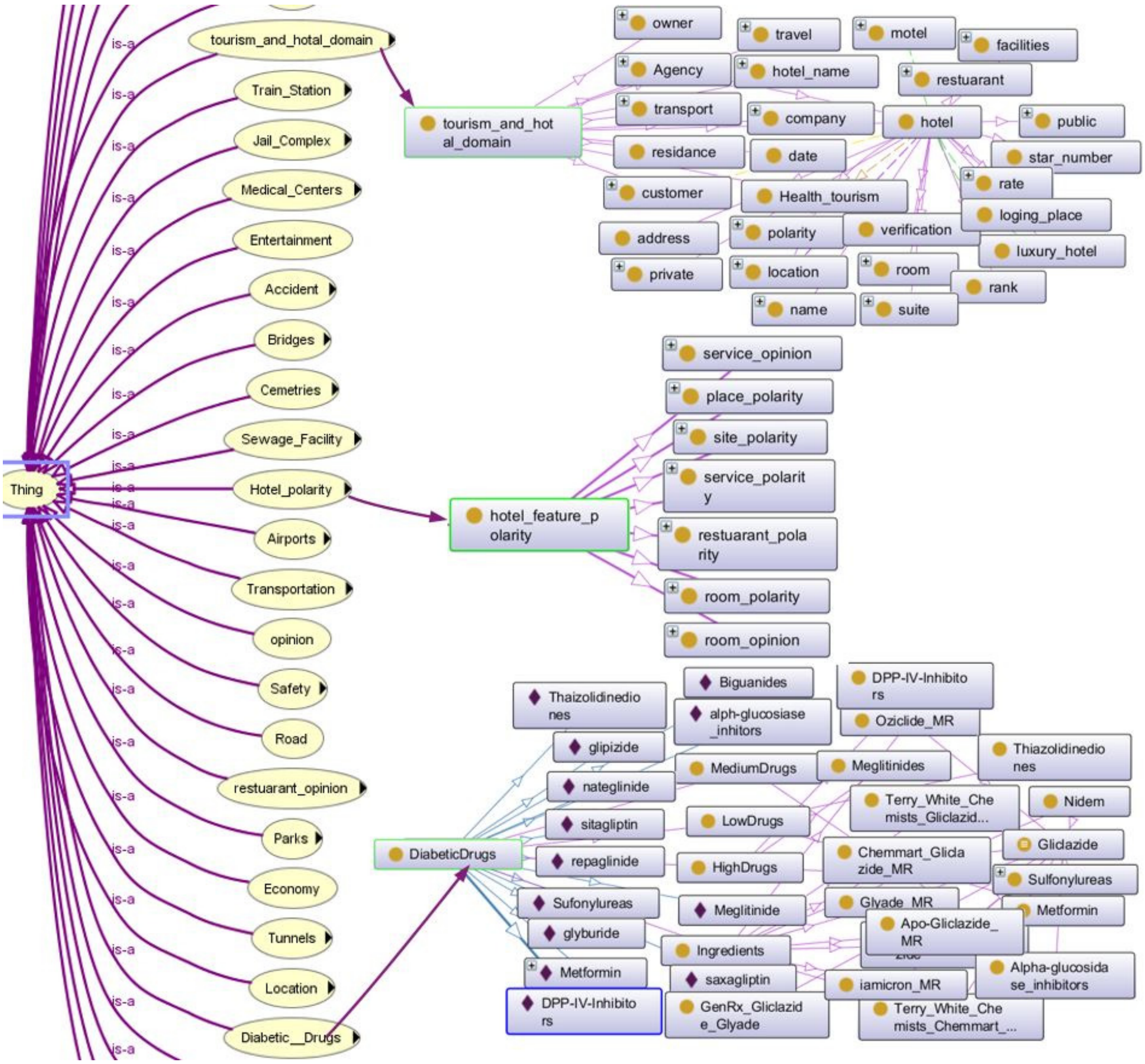


FIGURE 4. Merged ontology–inferred model generated by Protégé OWL.

full member of set A. Otherwise, it will be considered a partial member of set A [1]. A fuzzy ontology exchanges information among comment classifications, feature extractions, and polarity identification. This semantic knowledge defines the concepts of polarity generation of the hotel and city features and diabetes drugs. Protégé OWLViz and OntoGraph plugins are used to automatically generate a graphical representation of all features, as shown in Fig. 4. At first, the information about these features is employed to find lower-class polarity and then high-class polarity. For example, if the polarity of the city is strong negative, then subclasses (parks, hospitals, roads, and environment) can be used to find the cause of that city polarity.

The merged ontology is a combination of two or more sources for the ontology [31]. The hotel ontology, medical ontology, and city ontology have the information and concepts for the same area. Therefore, the merged ontology delivers a combined view of the proposed ontologies and helps the disabled user to efficiently extract precise information. The merged ontology keeps the properties and concepts of the features and the polarity. The composition of ontologies is a complex job. Therefore, we employed a Protégé OWL plugin called Merge Ontology. This plugin imports novel ontologies, and links the concepts of these ontologies using domain axioms [1].

The hotel ontology provides semantic knowledge to extract information from the Internet related to the hotel and its features. It also helps determine the polarity of a hotel feature. In the hotel ontology, we assert intervals and a fuzzy datatype of polarity terms for feature text classification. For example, fuzzy data types for restaurant are good, average, and bad, and the interval for a strong positive term is [0.75-1].

In addition, this semantic knowledge also defines the relationship between the hotel and its features more specifically. These features are room, luggage room, restaurant, WiFi, concierge service, air conditioning, in-room breakfast, swimming pool, business room, healthcare, and night club. These features have their own polarity words and values. For example, room has good and clean sentiment words with a polarity value p=1; medium, normal, and average are sentiment words with a polarity value p=0, and small and dusty are sentiment words with a polarity value p=−1. Furthermore, the initial values of these sentiment words are verified by using SentiWordNet [32]. The fuzzy OWL plug-in is employed to allocate the polarity value for each fuzzy sentiment variable. It regulates the polarity interval employing min-max normalization and maps it into the range [0, 1] [8]. The intervals for each result are the following: strong positive (SP) is [0.75-1], positive (P) is [0.5-0.75], neutral (Neu) is [0.5], negative (N) is [0.25-0.5], and strong negative (SN) is [0.0-0.25] [33]. Our system results are based on these intervals, which predict polarity by employing the values of input opinion words. These polarity results illustrate that the extracted information is relevant to the input query of disabled users.

The city ontology contains observed information such as foods is-a part of restaurant; hospital is-a feature of medical centers. We gathered all information, such as bus stations, medical centers, tunnels, speed, train stations, vehicles, sewage facilities, cemeteries, parks, bridges, jails, airports, streets, and traffic, and delivered them to the city ontology. This gathered information (or classes) illustrates the concept of city feature knowledge. Data and object properties describe the relationships of the classes linked to the basic data types. The interval of membership variables is presented by the fuzzy data type. Opinion mining employs this city ontology-based semantic knowledge and calculates pairs of city feature polarity.

The developed medical ontology expresses well-known medical terminology in order to find the polarity of drugs and assign drugs automatically, according to the disabled user query. Besides medical terminology, annotations (explanations) and components are included in the ontology in order to make it easily understandable in the proposed system. Common names are used in the ontology for identification. Diabetes drugs are divided into two categories: insulin and oral medications. This system retrieves both categories of information. In addition, the diabetes drugs class in the ontology is divided into four subclasses: High-Drugs, Medium-Drugs, Low-Drugs, and components. The High-Drugs class comprises medications that increase insulin production in the body, such as meglitinides and sulfonylureas. The Low-Drugs class comprises drugs that decrease the amount of blood sugar produced by the liver, such as metformin and thiazolidinedione. The Medium-Drugs class contains medications that prevent the breakdown of food into sugar, such as alpha-glucosidase inhibitors. The classes of polarity drugs have different types of variables, such as Very-Low, Low, Medium, High, and Very-High for each review classification of diabetes drug. This classification needs rules, and the SWRL rules plugin is employed to edit instructions for information and review classifications. Now, the proposed system can easily inform the disabled user by extracting the information, along with linguistic terms of features and polarity, using a DL query. After retrieving the exact information and positive sentiment score of the input query, the user will automatically be informed by NAO for further processing.

IV. EXPERIMENTS AND RESULTS

Our experimental environment was an Intel Core i7-2600 CPU with 8 GB RAM running Windows 7. We employed Apache Tomcat 7.0 as a web server, Java, and the NAOqi SDK for the NAO connection with Java. Protégé OWL was used to develop merged ontology-based semantic knowledge and connect it with the web crawler to fetch the results. The evaluation process contains two phases: merged ontology evaluation and measurement of overall system efficiency.

A. MERGED ONTOLOGY EVALUATION

To confirm the effectiveness of the proposed merged ontology-based recommendation system, we employed some well-known ways to evaluate the ontology, such as Protégé description logic queries, ontology expert evaluation, and user-level analysis. Protégé OWL provides a user-friendly plugin to extract individual items from semantic knowledge by using a query. These plugins are DL and the SPARQL queries. We used a DL query that extracts individual items by merging the concepts, variables, and properties of the proposed merged ontology. The DL queries are defined in the Manchester OWL syntax. The Protégé 5.0 package reasoners, such as Pellet, FACT++, and Hermit, must be run before any query execution. These reasoners get the inference result from the ontology. It also concludes the data on the basis of prescribed SWRL rules in the merged ontology. SWRL rules are defined as follows.

Rule 1: DiabetesDrugs(?Thaizolidinediones), WeightGain(?B), LiverDisease(?C), SwellingLegs(?D), LowBloodSuger(?E), OpinionOf(?Thaizolidinediones, SP), Opinion Of(?B, SN), OpinionOf(?C, SN), OpinionOf(?D, SN), OpinionOf(?E, SN) -> PolarityIs(?Thaizolidinediones, SP)

Rule 2: DiabetesDrugs(?Metformin), KidneyDisease(?B), StomachDisease(?C), LowBloodSuger(?D), Dizziness(?E), OpinionOf(?Metformin, SP), OpinionOf(?B, SN), OpinionOf(?C, SN), OpinionOf(?D, SN), OpinionOf(?E, SN) -> PolarityIs(?Metformin, SP)

The Rule 1 illustrates that the polarity of Thaizolidinediones (diabetes drugs) will be 'strong positive' if the opinion of all features is 'strong negative' and the opinion of Thaizolidinediones is 'strong positive'. The Rule 2 explains that if the opinion of Metformin (diabetes drugs) is 'strong positive' and the opinion of all side effects is 'strong negative', the polarity of suggested diabetes drugs will be counted as 'strong positive'.

The proposed system employed FACT++ and the Pellet utility to evaluate the ontology. FACT++ and Pellet analyze

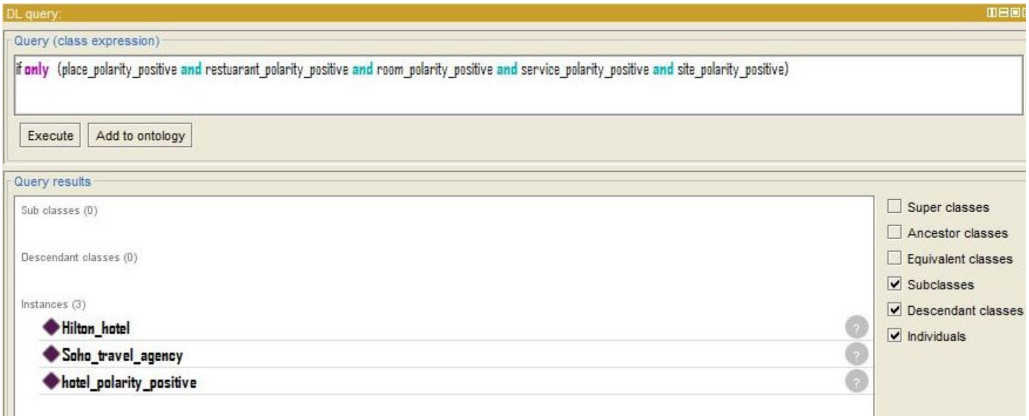


FIGURE 5. Output of DL query 1.

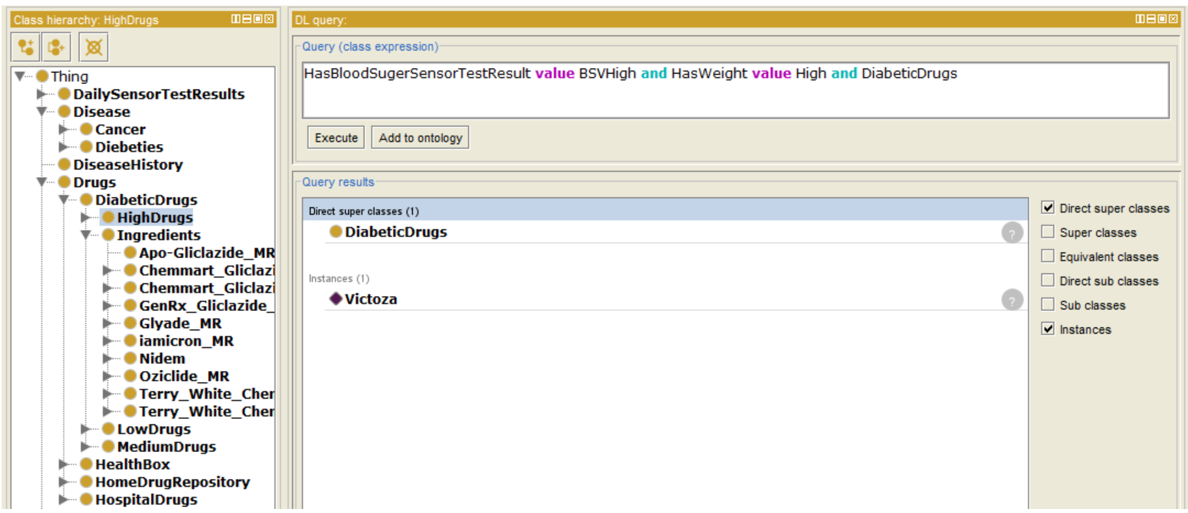


FIGURE 6. Output of DL query 2.

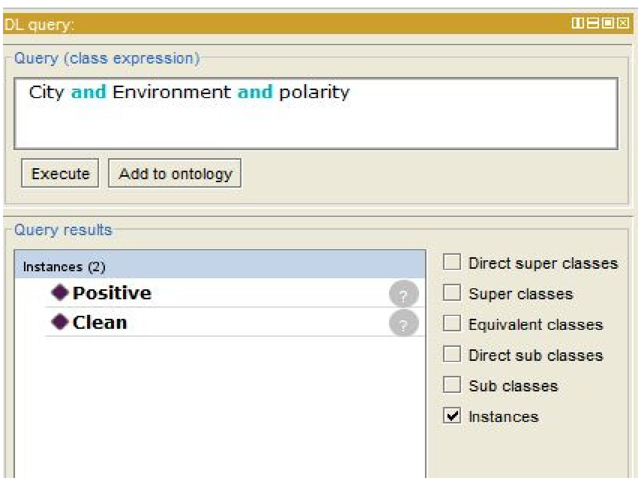


FIGURE 7. Output of DL query 3.

the ontology to retrieve the needed results. The proposed merged ontology has a lot of blurred terms, because it contains four different ontologies along with intensively vague information. Therefore, the FACT++ reasoner converted fuzzy reasoning into a classical format, and then analyzed the performance of the merged ontology. FACT ++ is based on the Jena API and produces reasoning from a conventional ontology. We designed various queries to check the performance effectiveness of the merged ontology and to extract the instances (results), according to disabled user desires, as follows.

TABLE 2. Performance evaluation of information extraction.

Information extraction	Relevant	Irrelevant
Extracted	TP	FP
Not extracted	FN	TN

Syntax of DL query 1: *If only (place polarity positive and room polarity positive and restaurant polarity positive and site polarity positive and service polarity positive).*

Explanation: In DL query 1, the domain user needs to obtain the sentiment result (polarity) of the hotel, where the

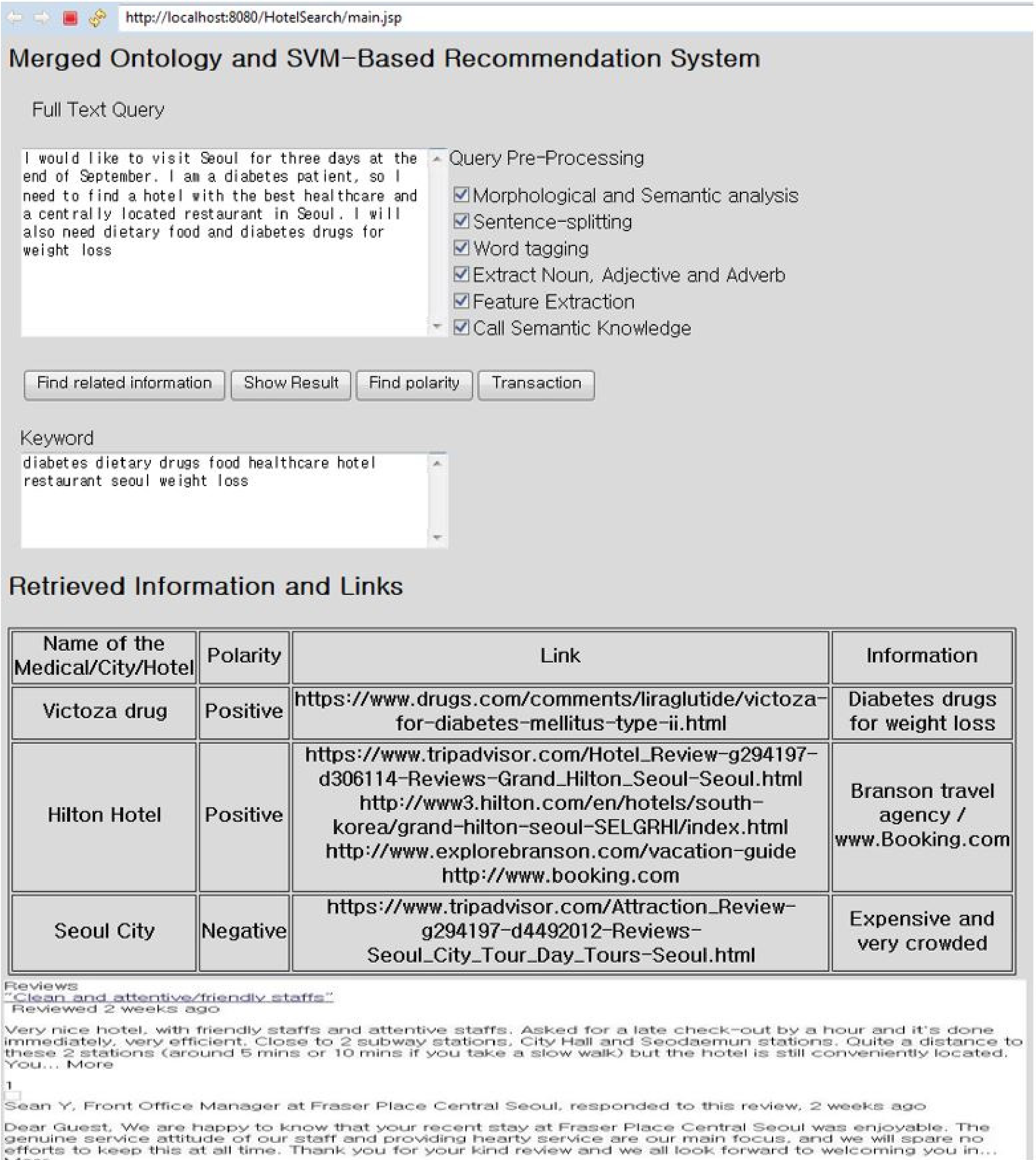


FIGURE 8. The proposed system for precise information extraction.

data are extracted by the input query of the disabled user. The information selection of the hotel and the hotel features relies on the input query and the retrieved information and reviews. Fig. 5 shows the output of this query, which illustrates that the polarity of the Hilton Hotel is positive, and the Soho travel agency can make the reservations.

Syntax of DL query 2: *HasBloodSugarSensorTestResult value BSVHigh and HasWeight value High and DiabetesDrugs.*

Explanation: In DL query 2, the ontology analyst wants to extract the drugs for a diabetes patient whose blood sugar is high and whose weight is heavy. *High* and *BSVHigh* are the linguistic terms of the fuzzy variable *Weight* and *Blood Sugar*, while *DiabetesDrugs* is a class name that contains drug information. The output of this query in Fig. 6 shows that *Victoza* is a drug for patients whose blood sugar is very high and who want to lose weight.

Syntax of DL query 3: *City and Environment and Polarity.*

TABLE 3. The proposed system's precision, recall, and accuracy using a regular ontology.

Ontology	Ontology Features	Total Extracted Features	Regular Ontology						
			TP	FN	TN	FP	Precision (%)	Recall (%)	Accuracy (%)
Medical Ontology	Concepts	562	171	110	180	101	63	61	62
	Properties	741	312	96	155	178	64	76	69
	Instances	355	102	73	111	69	60	58	59
Hotel Ontology	Concepts	321	120	52	79	70	63	70	66
	Properties	522	200	87	101	134	60	70	64
	Instances	218	70	35	64	49	59	67	63
City Ontology	Concepts	378	152	50	120	72	68	75	71
	Properties	394	160	35	109	90	64	82	72
	Instances	263	90	25	87	61	60	78	68

TABLE 4. The proposed system's precision, recall, and accuracy using the merged ontology with the SVM.

Ontology	Ontology Features	Total Extracte d Features	Merged Ontology with SVM						
			TP	FN	TN	FP	Precision (%)	Recall (%)	Accuracy (%)
Medical Ontology	Concepts	562	300	80	175	7	98	79	87
	Properties	741	400	120	213	8	98	77	86
	Instances	355	150	45	150	10	94	77	85
Hotel Ontology	Concepts	321	185	39	90	7	96	83	89
	Properties	522	310	90	112	10	97	78	86
	Instances	218	80	30	103	5	94	73	82
City Ontology	Concepts	378	160	55	150	13	92	74	82
	Properties	394	179	30	173	12	94	86	90
	Instances	263	111	40	100	12	90	74	81

Explanation: In the above query, the domain user needs to extract the information on the city and the city environment along with polarity. *Positive* and *clean* are the linguistic terms of the fuzzy variable for *city* and *environmen*t polarities, respectively. Fig. 7 shows the output of this query.

B. THE OVERALL SYSTEM EFFICIENCY MEASUREMENT

The proposed system received an input query from disabled user, pre-processed the query, and then assigned the query to the merged ontology-based web crawler that retrieves the information and the reviews related to the input query. The system retrieved 7639 sentences related to city, hotel, and diabetes drugs. The average length of the sentences was 50 words. For polarity computation, the extracted sentiment words numbered 9631. The downloaded text comprises unnecessary information. Therefore, this irrelevant information was filtered out, and the system made sure that every sentence is a complete clause. In filtering, the text document was transformed into suitable vectors of an SVM for classification. Sentences were classified by using the SVM classification function:

$$f(sentence) = 0.5^{*}diet + 0.1^{*}weight\ gain + 0.6^{*}blood\ sugar + 0.3^{*}liver\ disease.$$

The classification result is

$$f(sentence) = 0.5^{*}1 + 0.1^{*}1 + 0.6^{*}1 + 0.3^{*}1 = 0.9.$$

If f (sentences) > 0, it is positive sentence (related to diabetes drugs); otherwise, the sentence is negative. First, the text was manually categorized by domain experts to mark the text as relevant or irrelevant sentences. The text documents were then separated into training and testing sets. The RBF functions of the SVM were employed in the training phase. The retrieved sentences related to diabetes drugs, the city, and the hotel were used to train the classifier. In the training phase, the ontology classes (features) and retrieved sentences are designated as samples, and then the process splits the sentences for ontology features into two sectors. The first

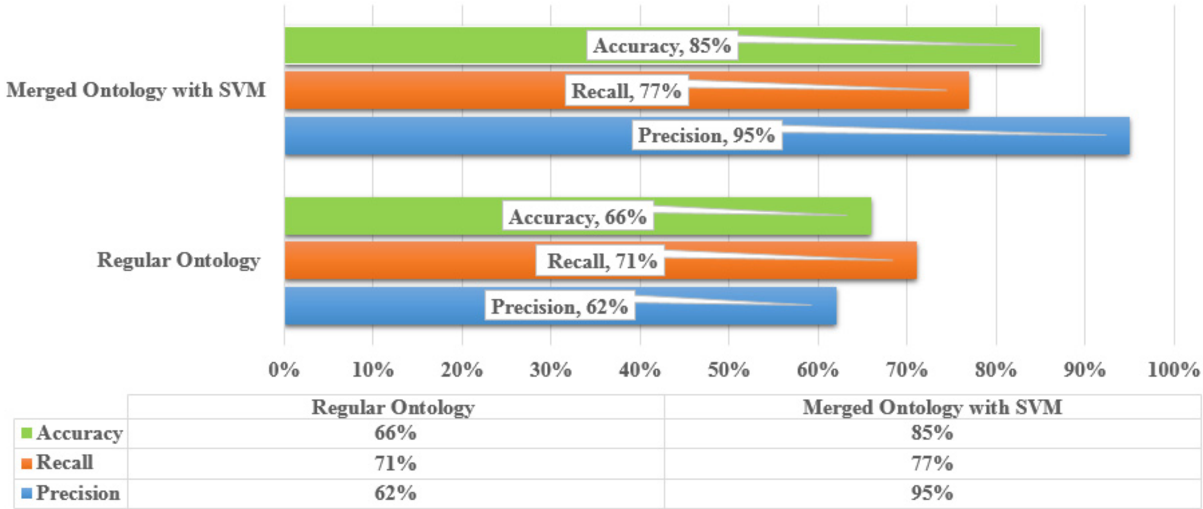


FIGURE 9. Graphical comparison of the regular ontology and the merged ontology with the SVM.

TABLE 5. Performance comparison of the proposed system with existing systems.

References	Algorithms	Precision	Recall	Accuracy
[1]	Type-2 fuzzy ontology	72%	61%	81%
[27]	Fuzzy ontology with SVM	74%	60 %	82%
Proposed system	Merged ontology with SVM	95%	77%	85%

sector is to train the information, and the second sector is to test the information. The system employed a collection of 4902 sentences as the training set and 2737 sentences as the testing set for the text categorization model. The relevant and irrelevant sentences for the items were employed to generate a 10-fold cross validation, with six-folds and four-folds for the training and testing data sets, respectively. The merged ontology then extracts precise information from the filtered data and finds the polarity of those features that are mentioned in the input query. Fig. 8 shows the full-text query reformulation module and item recommendations from the retrieved text using sentiment analysis.

Various experiments were performed to evaluate the efficiency of the proposed merged ontology and the recommendation system. Precision and recall are basic measures for recommendation efficiency. Precision (P) is the ratio of extracted elements that are relevant, and recall (R) is the ratio of relevant elements that are extracted. The concepts are defined in Table 2. Mathematically, the rate of precision, recall, and accuracy can be defined as shown below.

$$\text{Precision} = \frac{TP}{(\text{TP} + \text{FP})} \times 100\% \tag{11}$$

$$\text{Recall} = \frac{TP}{(\text{TP} + \text{FN})} \times 100\% \tag{12}$$

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(TP + FP + FN + TN)} \tag{13}$$

In the above equation, TP, TN, FP, and FN, represent true positive, true negative, false positive, and false negative, respectively, in the recommendation. Tables 3 and 4, respectively, show experimental results for the regular ontology and the merged ontology with the SVM. The average rates of precision, recall, and accuracy using the regular ontology were 62%, 71%, and 66%, respectively. In case of the proposed system, the statistics were 95%, 77%, and 85%. The comparative results of the experimental regular ontology and merged ontology with SVM in terms of precise information extraction and recommendation are shown in Fig. 9. Note that for regular ontology-based recommendations, the system achieves a 62% average precision, whereas the proposed merged ontology succeeds 95% of the time, which is a 33% improvement. The main reason is that the proposed system can efficiently handle blurred information because of the fuzzy logic, SWRL rules, and the SVM. It shows that our proposed system is more effective than a regular ontology-based system in terms of recommendation and information extraction.

As mentioned in the previous sections, the type-2 fuzzy ontology-based system studies only a limited amount of reviews and cannot filter out the irrelevant reviews, which affects the performance. The existing fuzzy ontology-based systems employ the fuzzy value to make the relationship between the feature and its polarity term. However, these systems are designed for a single domain and cannot handle the real time reviews of various domains. However, this paper proposes a merged ontology with SVM, which effectively handles real time reviews of different domains during recommendation and information extraction. The performance comparison of the proposed system with existing systems [1], [27] based on three performance metrics is shown in table 5. These results clearly show that the proposed system outperforms the other two systems in terms of *precision, recall, and accuracy.*

V. CONCLUSION

In this paper, we propose a merged ontology and SVM-based recommendation and information extraction system. The system automates the extraction of precise data from the Internet and suggests accurate items for disabled users. A number of reasonable issues are effectively considered; for example, oral query conversion into the proper format for a keyword-based search engine, morphological analysis of a full-text query, tokenization, POS tagging, information retrieval using a merged ontology, irrelevant data filtering employing an SVM, and sentiment analysis-based recommendation. The proposed system improves the performance in recommendations and information extraction. Moreover, the proposed system efficiently categorizes the retrieved data and intelligently computes the polarity of the desired items for recommendation. Our proposed system can effectively help disabled users to easily retrieve information from webpages, classifies the text, and then suggests correct items, since it has the capability to eliminate noise, extract the item feature information and sentiment words, and categorize the sentiment words into an extended degree of polarity terms (positive, neutral, and negative). In future research work, we will enhance the accuracy of the recommendation system. We will also concentrate on employing an artificial neural network and ontology-based sentiment analysis for drug recommendations.

REFERENCES

[1] F. Ali, E. K. Kim, and Y.-G. Kim, "Type-2 fuzzy ontology-based opinion mining and information extraction: A proposal to automate the hotel reservation system," *Appl. Intell.*, vol. 42, no. 3, pp. 481–500, Apr. 2015.

[2] P. Dahiwale, M. M. Raghuwanshi, and L. Malik, "PDD Crawler: A focused Web crawler using link and content analysis for relevance prediction," in *Proc. Int. Conf. SEAS*, Nov. 2014, pp. 1–9.

[3] M. Shiomi *et al.*, "Recommendation effects of a social robot for advertisement-use context in a shopping mall," *Int. J. Soc. Robot.*, vol. 5, no. 2, pp. 251–262, Apr. 2013.

[4] K. Kamei, K. Shinozawa, T. Ikeda, A. Utsumi, T. Miyashita, and N. Hagita, "Recommendation from robots in a real-world retail shop," in *Proc. Int. Conf. Multimodal Interfaces Workshop Mach. Learn. Multimodal Interact. (ICMI-MLMI)*, Nov. 2010, Art. no. 19.

[5] H. Baazaoui, M.-A. Aufaure, and R. Soussi, "Towards an on- line semantic information retrieval system based on fuzzy ontologies," *J. Digit. Inf. Manag.*, vol. 6, no. 5, pp. 375–385, 2008.

[6] Y.-H. Feng, N.-F. Huang, and C.-H. Chen, "An efficient caching mechanism for network-based URL filtering by multi-level counting bloom filters," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2011, pp. 1–6.

[7] N. Lee, E. Kim, and O. Kwon, "A real-time combination of text-mining methods for flexible movie recommendation in human robot interaction," in *Proc. Int. Conf. Platform Technol. Service (PlatCon)*, Feb. 2016, pp. 1–6.

[8] Y. Pan, H. Okada, T. Uchiyama, and K. Suzuki, "Listening to vs overhearing robots in a hotel public space," in *Proc. ACM/IEEE Int. Conf. Human-Robot Interact.*, Mar. 2013, pp. 205–206.

[9] G. Ermacora, S. Rosa, and A. Toma, "Fly4SmartCity: A cloud robotics service for smart city applications," *J. Ambient Intell. Smart Environ.*, vol. 8, no. 3, pp. 347–358, 2016.

[10] A. Prange, I. P. Sandrala, M. Weber, and D. Sonntag, "Robot companions and smartpens for improved social communication of dementia patients," in *Proc. 20th Int. Conf. Intell. User Interfaces Companion*, 2015, pp. 65–68.

[11] T. L. Nguyen, R. Boukezzoula, D. Coquin, E. Benoit, and S. Perrin, "Interaction between humans, NAO robot and multiple cameras for colored objects recognition using information fusion," in *Proc. 8th Int. Conf. Hum. Syst. Interact. (HSI)*, Jun. 2015, pp. 322–328.

[12] H. Bast, F. Bäurle, B. Buchhold, and E. Haussmann, "A case for semantic full-text search," in *Proc. 1st Jt. Int. Work. Entity-Oriented Semant. Search*, 2012, p. 4.

[13] E. Riloff and J. Wiebe, "Learning extraction patterns for subjective expressions," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2003, pp. 105–112.

[14] N. Kaji and M. Kitsuregawa, "Automatic Construction of Polarity-tagged Corpus from HTML Documents," in *Proc. COLING/ACL Main Conf. Poster Sess. (COLING-ACL)*, Jul. 2006, pp. 452–459.

[15] M. Chaves, L. Freitas, and R. Vieira, "Hontology: A multilingual ontology for the accommodation sector in the tourism industry," in *Proc. Int. Conf. Knowl. Eng. Ontol. Develop. (KEOD)*, 2012, pp. 149–154.

[16] P. Yan, Y. Zhao, and C. Sanxing, "Ontology-based information content security analysis," in *Proc. 5th Int. Conf. Fuzzy Syst. Knowl. Discovery (FSKD)*, vol. 5. Oct. 2008, pp. 479–483.

[17] L. Zhao and C. Li, "Ontology based opinion mining for movie," in *Proc. 3rd Int. Conf. Knowl. Sci., Eng. Manag. (KSEM)*, Nov. 2009, pp. 204–214.

[18] M. K. Dalal and M. A. Zaveri, "Linguistic hedges," *J. Appl. Comput. Intell. Soft Comput.*, vol. 2014, Jan. 2014, Art. no. 2.

[19] J. Polpinij and A. K. Ghose, "An ontology-based sentiment classification methodology for online consumer reviews," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell. (WI)*, Dec. 2008, pp. 518–524.

[20] B. O. Olusanya, K. J. Neumann, and J. E. Saunders, "The global burden of disabling hearing impairment: A call to action," *Bull. World Heal. Organ.*, vol. 92, no. 5, pp. 367–373, 2014.

[21] D. McColl, W.-Y. G. Louie, and G. Nejat, "Brian 2.1: A socially assistive robot for the elderly and cognitively impaired," *IEEE Robot. Autom. Mag.*, vol. 20, no. 1, pp. 74–83, Mar. 2013.

[22] H. Cunningham, R. J. Gaizauskas, and Y. Wilks, "A general architecture for text engineering (GATE): A new approach to language engineering R & D," Dept. Comput. Sci., Univ. Sheffield, Tech. Rep., 1995.

[23] N. O. Schiller and A. Caramazza, "Grammatical feature selection in noun phrase production: Evidence from German and Dutch," *J. Memory Lang.*, vol. 48, no. 1, pp. 169–194, Jan. 2003.

[24] H. Jeong, D. Shin, and J. Choi, "FEROM: Feature extraction and refinement for opinion mining," *ETRI J.*, vol. 33, no. 5, pp. 720–730, 2011.

[25] V. N. Vapnik, "An overview of statistical learning theory," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 988–999, Sep. 1999.

[26] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, 2013, Art. no. 27.

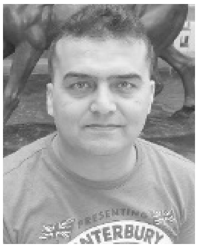
[27] F. Ali, K.-S. Kwak, and Y.-G. Kim, "Opinion mining based on fuzzy domain ontology and Support Vector Machine: A proposal to automate online review classification," *Appl. Soft Comput.*, vol. 47, pp. 235–250, Oct. 2016.

[28] F. Bobillo and U. Straccia, "Fuzzy ontology representation using OWL 2," *Int. J. Approx. Reason.*, vol. 52, no. 7, pp. 1073–1094, Oct. 2011.

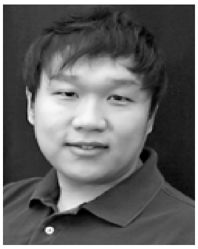
[29] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, Jun. 1965.

[30] F. Ali, E. K. Kim, and Y.-G. Kim, "Type-2 fuzzy ontology-based semantic knowledge for collision avoidance of autonomous underwater vehicles," *Inf. Sci.*, vol. 295, pp. 441–464, Feb. 2015.

[31] J. De Bruijn, M. Ehrig, C. Feier, F. Martíns-Recuerda, F. Scharffe, and M. Weiten, "Ontology mediation, merging and aligning," *Semant. Web Technol. Trends Res. Ontol. Syst.*, Jul. 2006, pp. 95–113.
[32] M. K. Dalal and M. A. Zaveri, "Semisupervised learning based opinion summarization and classification for online product reviews," *Appl. Comput. Intell. Soft Comput.*, vol. 2013, 2013, Art. no. 10.
[33] F. Ali, D. Kwak, P. Khan, S. M. R. Islam, K. H. Kim, and K. S. Kwak, "Fuzzy ontology-based sentiment analysis of transportation and city feature reviews for safe traveling," *Transp. Res. C, Emerg. Technol.*, vol. 77, pp. 33–48, Apr. 2017.



FARMAN ALI received the B.S. degree in computer science from the University of Peshawar, Pakistan, in 2011, and the M.S. degree in computer science from Gyeongsang National University, South Korea, in 2015. He is currently pursuing the Ph.D. degree at the Department of Information and Communication Engineering, Inha University, South Korea. His current research interests include opinion mining, ontology, fuzzy logic, artificial intelligence, and machine learning.



DAEHAN KWAK received the B.S. degree in information and computer engineering from Ajou University, South Korea, in 2005, and the M.S. degree from the Korea Advanced Institute of Science and Technology, South Korea, in 2008. He is currently pursuing the Ph.D. degree in computer science at Rutgers University, New Brunswick, NJ, USA. He was with the Telematics and USN Research Division, Electronics and Telecommunications Research Institute, during his graduate work. He had been with the UWB Wireless Research Center, Inha University, South Korea, as a Researcher. His current research interest includes mobile and pervasive computing, vehicular computing and networks, u-health networks/applications, and mobile applications.

PERVEZ KHAN received the B.S. and M.S. degrees in computer science from the University of Peshawar, Pakistan, in 2006 and 2003, respectively, and the Ph.D. degree in information and communication engineering from the Graduate School of IT and Telecommunication Engineering, Inha University, Incheon, South Korea, in 2015. His current research interests include wireless communications, fuzzy logic, semantic analysis, wireless sensor networks, wireless adhoc networks, wireless body area networks, performance evaluation, and MAC protocol design.

SHAKER HASSAN A. EI-SAPPAGH received the B.S. degree in computer science from the Information Systems Department, Cairo University, Egypt, in 1997, the M.S. degree from Cairo University in 2007, and the Ph.D. degree in computer science from the Information Systems Department, Mansura University, Egypt, in 2015. In 2009, he joined as a Lecturer with the Collage of Science, King Saud University. Since 2016, he has been with the Department of Information Systems, Minia University, as a Lecturer. His current research interests include medical informatics, ontology engineering, distributed and hybrid clinical decision support systems, semantic data modeling, distributed database systems, big data, semantic query languages, description logic, and fuzzy logic.

S. M. RIAZUL ISLAM (M'10) received the B.S. and M.S. degrees in applied physics and electronics from the University of Dhaka, Bangladesh, in 2003 and 2005, respectively, and the Ph.D. degree in information and communication engineering from Inha University, South Korea, in 2012. He was with the University of Dhaka, Bangladesh, as an Assistant Professor and a Lecturer with the Department of Electrical and Electronic Engineering from 2005 to 2014. In 2014, he was with the Samsung Research and Development Institute Bangladesh, as a Chief Engineer with the Department of Solution Lab for six months. From 2014 to 2017, he was a Research Professor with the UWB Wireless Communications Research Center, Inha University, South Korea. He has been an Assistant Professor with the Department of Computer Science and Engineering, Sejong University, South Korea, since 2017. His current research interests include wireless communications, signal processing for communications, and enabling technologies for 5G and beyond.

DAEYOUNG PARK received the B.S. and M.E. degrees in electrical engineering and the Ph.D. degree in electrical engineering and computer science from Seoul National University, Seoul, South Korea, in 1998, 2000, and 2004, respectively. He was with Samsung Electronics as a Senior Engineer from 2004 to 2007, contributing to the development of next-generation wireless systems based on MIMOâĂŞOFDM systems. He has held visiting positions with the University of Southern California, Los Angeles, CA, USA, and the University of California at San Diego, La Jolla, CA, USA. Since 2008, he has been with Inha University, Incheon, South Korea, where he is an Associate Professor. He is a coauthor of *Wireless Communications Resource Management* (Wiley, 2009). His current research interests include communication systems, wireless networks, multiuser information theory, and resource allocation.

KYUNG-SUP KWAK (M'81) received the Ph.D. degree from the University of California at San Diego, La Jolla, CA, USA, in 1988. From 1988 to 1989, he was with the Hughes Network Systems, San Diego, CA, USA. From 1989 to 1990, he was with the IBM Network Analysis Center, Research Triangle Park, NC, USA. Since 1990, he has been with the School of Information and Communication Engineering, Inha University, South Korea, as a Professor, where he was the Dean of the Graduate School of Information Technology and Telecommunications from 2001 to 2002. He has been the Director of the UWB Wireless Communications Research Center, South Korea, since 2003. In 2006, he served as the President of the Korean Institute of Communication Sciences and, in 2009, the President of the Korea Institute of Intelligent Transport Systems. He has authored over 200 peer-reviewed journal papers. His research interests include wireless communications, UWB systems, sensor networks, WBAN, and nano communications. In 2008, he was selected as an Inha Fellow Professor. He is currently an Inha Hanlim Fellow Professor. He was a recipient of the number of awards, including the Engineering College Achievement Award from Inha University, the LG Paper Award, the Motorola Paper Award, the Haedong Prize of research, and various government awards from the Ministry of ICT, the President, and the Prime Minister of Korea, for his excellent research performances. He served as the TPC/Track Chair/Organizing Chair for several IEEE related conferences.