

A Domain Case-based Knowledge Model for Supporting of High-level Information Retrieval using Reducing Ambiguity in Language

Sukyoungh Kim, Hyunsuck Seo, Kihong Anh*

*Computer Engineering, Hanbat National University, San16-1 Duckmyoung-dong, Daejeon, Korea

kimsk, khahn@{hanbat.ac.kr}, skinahitoda@gmail.com

Abstract— This study is about a domain case-based knowledge base model that can improve the level of human value judgement and decision for information retrieve as reducing a ambiguity in language. The big data problem occurred which is decreased the user confidence in search result by a dramatic increase of data or information, and The increasing use of mobile device, is emphasized to increase the simplicity that reduce the human involvement for searching and deciding of result. Therefore, in this study, for increasing the information retrieve performance in area of needing the high-level decision. Indeed, this study has the objective as search for improved performance that required high-level decision-making in the field. For showing high-level concepts or knowledge, in a specific domain area, we used TBox and RBox search and inference mechanism for solving the ambiguity in language. In order to improve search performance which represented concepts, on the processing of knowledge base design and quality through the use of qualified, and is to validate the configuration. Drugs that require expertise in the field of knowledge base Search performance analysis of the knowledge base, Keyword-based search with the ambiguity of the language problem can be solved.

Keywords— ontology model, ambiguity, information retrieval

I. INTRODUCTION

A. Background of Study

According to the Web 2.0 era, Internet spread is being rapidly. In this situation, the rapid growth of unstructured data and their users and companies are struggling to find the desired information. The task of the refining data by user, these operations are generated many problems. Moreover, the amount of data was that degrade the accuracy and recall. Renowned in the field of IT research firm Gartner suggest, 10 strategic technologies in 2012 at Table 1. Big Data problems experienced by IT, Next-generation analytical techniques by Business Experience developed as expected.

Big Data problems were disposed of more data than before to user and company. And the world Internet access is expected to grow more. As we can see table 2, there are

technologies to solve Big Data problem, that is Hadoop, GFS (Google File System), NoSQL, etc. However, these techniques have problem that is for large amounts of data storage and retrieval, and indexing. Not how efficiently a large amount of data users can access. In these circumstances, this study aims was that compliance is granted to retrieve the data. High-level supporting, and resolving the ambiguity of the language on Big Data.

TABLE 1. GARTNER'S 2012 10 STRATEGIC TECHNOLOGY

Human Experience	1	Media tablets and beyond
	2	Mobile-centric applications and interface
	3	Contextual and social user experience
Business Experience	4	internet of things
	5	App stores and marketplaces
	6	Next generation Analytics
IT Dept, Experience	7	Big Data
	8	in-memory computing
	9	Extreme low-energy servers
	10	Cloud Computing

TABLE 2. TECHNOLOGIES RELATED WITH BIG DATA

Technology	Description
Hadoop	Hadoop is distributed file system that consist of MapReduce programming model
GFS	Hadoop distributed file system, such as ideas, Large-scale cluster-based file systems, service platforms
NoSQL	If a relational database to store large amounts of data on its own scale-out approach is difficult to expand. These problems are resolved, and hundreds of millions - billions more than storing data reliably without the need to join and have access to disability in situations that could help you in the sense that generic skills are usually used.

B. Object of Study

For solving the Big Data Problem, we suppose that the user to easily access the data easily in order to induce.

① We are choice the text mining technology for the non-refined and informal collected data.

② To analysis the collected data distributed processing. Finally,

③ The data stored in the user to provide more relevant results ontology technology.

Prior to the Big Data problems of this study, we focused the how can reduce the amount of data, and solve the ambiguity in language. The main point is the accuracy of the information searching according to the domain that user want search it. We design a domain case-based knowledge model for supporting of high-level information retrieval for reducing ambiguity in web site or text. First of all, in order to design NLP ontology, we collected web pages related with specific domain and than applied the text-mining technology analysis important vocabulary, and capture the concept and classify the cluster it. And we design an ontology model belong to the specific area, especially need to expert experience.

It also provides for the general user searches. For improving accessibility with the data on professional. Expertise is needed during the study Data mining techniques for reading the data between the html tag, Crawling algorithm for the collection of unstructured data and Ontology technical for knowledge base If need a relationship between the data model. The result is derived during the research outcomes Ontology-based reasoning model drug knowledge base, crawler for unstructured data is stored in the database and specialized drug knowledge base model. Finally, these studies are essential to the user. More data is exploding, and Search for information via the Internet has become essential to everyday.

II. RELATED STUDY

A. Big Data process framework

1) Hadoop-based traffic analysis, system architecture for Multi-user

This study focused. Due to the increase of Internet traffic analysis and management. Mobile device penetration rates are increasing, music, and video services has increased. As a result, increases the amount of traffic that must be analyzed. As such, more and more rapidly increasing traffic handled at a time to do a lot of data to a computer, which charged distributed processing techniques, such as Hadoop, it costs more to process data efficiently and helps ensure that. The results of applying the Hadoop technology for increased data processing without having to run the server extensions are unnecessary. Furthermore the software to purchase, rather than if user want to connect to the server to take advantage of the data derived, better traffic analysis to build a system model that can be presented.

2) Hadoop-based distributed filtering service model for Handling large amounts spam mailr

MapReduce in Hadoop distributed parallel systems that support a programming model based on the most widely used mail server sends mail and uses MapReduce. MapReduce

distributed processing using these ontologies distributed through the data analysis should be performed. Thus, in the presence of large amounts of data faster than it can be treated as distributed to service will present the possibility of creating.

MapReduce, as shown in Figure 2, the structure part of the Map and Reduce parts are separated. Through a combination of both to support operations for distributed parallel systems. MapReduce data {Key, Value} by a pair of support to handle this process. If possible, is utilized it for reasoning.

C. Semantic Social Network for text-mining

Research in social network, a semantic social network research which recently uses the ontology or the semantic web technology as follows; [2] introduce a three-layered model which involves the network between people(social network), the network between the ontologies they use(ontology network) and a network between concepts occurring in these ontologies. and explain how relationship in one network can be extracted from relationships in another one based on analysis techniques relying on this network specificity [3] describe a Semantic Web application that detects Conflict of Interest(COI) relationships among potential reviewers and authors of scientific papers, and discovers various 'semantic associations' between reviews and authors in a populated ontology to determine a degree of Conflict of Interest.

There is the analysis of the social network with the research to consider in order to re-form the social network. This understands the architectural feature of the social network and provides the analysis about the density of the relation or the concentricity. Generally, we can grasp the feature of a network through the analysis of this because of having the small world and scale free characteristic [4].

III. PROPOSED APPROACH

All paragraphs must be indented. All paragraphs must be justified, i.e. both left-justified and right-justified.

A. Ontology Schema Model Design

as described in the above, actively has to connect the information of many. The active connection has to facilitate an expansion and an embedding between the mutually different domain ontologies[5]. The many research that it is like an ontology mapping and an ontology matching is progress to provide this automatically with the ontology mapping and By the way, the offer of the automatized expansion and embedding, formerly, a model has to be designed so that a service can be provided through the explicit ontology model of the manually base. The criteria that has consider to design the explicit ontology model is as follows:

- (1) Representation scope of the designed ontology model
- (2) Explicit inference and implicit inference scope provided through the ontology
- (3) Integrability with the old data infrastructure like the web or the pervasive computing.
- (4) Ontology representation language and a performance of reasoning engine.

- (5) Accessibility of the ontology information and usage frequency and possibilities of growth and transformation

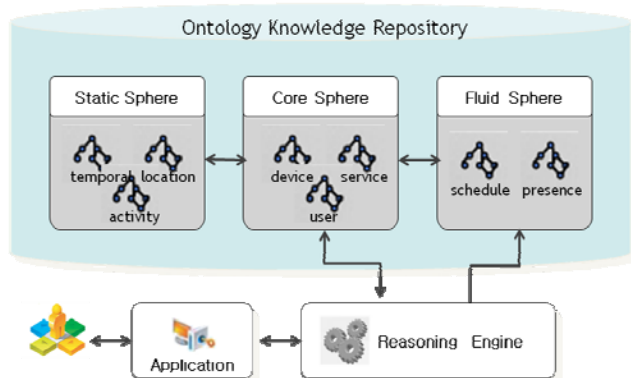


Figure 1. Structure and Layout Ontology Repository

The design criteria in which we present can use as the guide-line in which it can be satisfied the feature of the constructed ontology. Fig. 1 shows that each design criteria can determine the several ontology feature. That is, the representation scope of the designed ontology model is used for the ontology structure, the service coverage, that it is provided by using the ontology, and the ontology expandability, and etc. as the guide-line information.

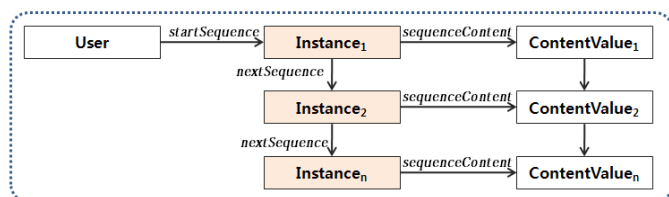


Figure 4. Pattern for NLP process in ontology model

Fig. 3 is the ontology model framework in which it comprises according to the result of being analyzed by the condition of the upper part. The ontology repository is separated with the three spheres which generalized the condition that it described at above and the role of the constructed ontology and data fluidity and communization and etc.

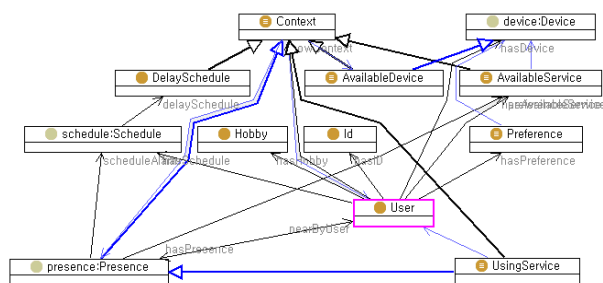


Figure.5 Class diagram around ontology

B. Knowledge base model Through Verify the reliability of information

Shown in Figure 3, prior to the overall design model ① ontology and, ② Database focuses on the design model should be. Among them, as an important model ontology focuses on the design. Returns the data to the user the most Knowledge base, focused on the drug. "A Domain Case-based Knowledge Base Model for Supporting of High-level Information Retrieval using Reducing Ambiguity in Language" As you can see from the research subject for the user would like to support high-level information retrieval. important role in giving since.

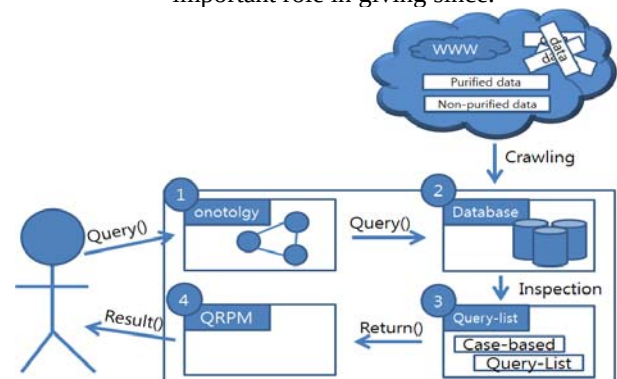


Figure 6. Overall processing suggested model

TABLE 3. similary vocabulary related with drugs

Drug Efficacy	Synonyms(Sentence)
Cold	Goppul, Cough, Sneeze (Goppul is korea dialect)
Headache	Head

C. Database Modeling through Proven Knowledge base

There is work to be followed making the inference engine. ① data collect ② data modeling based on data collected.

This study database is based on data from the drug. According to research the background of expertise in how the public could be accessed without difficulty, is the reason for verifying. Also Symptoms and effects modeling applied to the inference engine to verify the relationship to fit (Table 3) were selected from the reason that.

Figure 7 modeling of the database to search for drugs that show the variety of keywords. But ordinary users to retrieve data accurately is difficult. This problem is that the user to access data, accessible from the start down a search. These problems can be overcome. According to the data in the appropriate ontology modelling.

IV. TEST

User can be the following situation. ① I know my disease name ② I know My symptoms but I don't know my disease name ③ I don't know my disease name and symptoms but I can describe my states. When the three conditions classified as Information about their medicines when you want to know the

```

graph TD
    Q[쿼리] --> S[시작]
    Q --> E[종료]
    S --> SS[시작_시작]
    S --> SE[시작_종료]
    S --> EE[시작_종료]
    E --> ES[종료_시작]
    E --> EE[종료_종료]
    SS --> SS_S[시작]
    SS --> SS_E[종료]
    SE --> SE_S[시작]
    SE --> SE_E[종료]
    EE --> EE_S[시작]
    EE --> EE_E[종료]
    ES --> ES_S[시작]
    ES --> ES_E[종료]
    EE --> EE_S[시작]
    EE --> EE_E[종료]
    
```

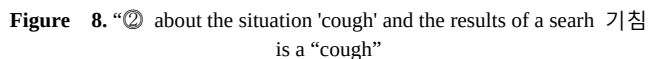
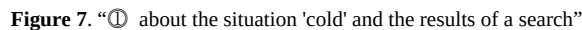
Figure 1 illustrates the decomposition of a query into subqueries. The root node is '쿼리' (Query), which branches into '시작' (Start) and '종료' (End). '시작' branches into '시작_시작' (Start_Start), '시작_종료' (Start_End), and '시작_종료' (Start_End). '종료' branches into '종료_시작' (End_Start) and '종료_종료' (End_End). Each of these further branches into '시작' (Start) and '종료' (End). The diagram shows the decomposition of a query into subqueries, with each subquery having its own set of attributes and relationships. The relationships are labeled 'Rel_01' through 'Rel_05'.

TABLE 4. Assumed Situation Searching Table

Situation	Information needed
①	Keyword about name of a disease
②	Keyword about symptom
③	Explanation about symptom

V. RESULT AND FUTURE WORK

knowledge, the frequency of using, accessibility, and the degree of a change to be similar to the preexistence MVC pattern is proposed. Secondly, we are described which is the situation of the various entity related to a user and are presented the situation-oriented ontology representation pattern for the adaptable decision according to this. Thirdly, we propose for the smooth induction of the inference rule according to the dynamic service progress, and the visible knowledge processing and knowledge connection method base on a flow chart.



This study was carried out with 2011 Korean Government (the Ministry of Education, Science and Technology) budget and the Basic Research Project Support of Korea Research Foundation (2011-0005893).

- [1] Dean Allemang, "Semantic Web for Working Ontologist", Schtec Media, pp3-10,
- [2] SPICE Ontology Definition of User Profiles, Knowledge Information and Services pp.3-5, 2006.
- [3] MobiLife, <http://www.ist-mobilife.org>, 2008.
- [4] Chantal Taconet, Zakia Kazi-Aoui, "Context-awareness and Model Driven Engineering", E-Commerce application scenario. ICDIM 2008.
- [5] Bright, M.W., Hurson, A.R., Pakzad, S.H. "Automated resolution of semantic terogeneity in multidatabases", ACM Transaction on Database Systems 19, pp.212-253, 1994.
- [6] Guarino, N., Masolo, C., Verete, G. "Ontoseek: Content-based access to the web", IEEE Intelligent Systems 3, pp.70-80, 1