International Journal of Fuzzy Logic and Intelligent Systems, vol. 7, no. 4, December 2007, pp. 301-306

Time Variant Event Ontology for Temporal People Information

Yong-Jin Han, Se-Young Park, Seong-Bae Park, Young-Hwa Lee and Kweon-Yang Kim*

Department of Computer Engineering, Kyungpook National University
1370 Sankyuk-Dong, Buk-Gu, Daegu, 702-701, Korea
Tel: +82-53-957-4846, E-mail: {yjhan,sypark,sbpark,yhlee}@knu.ac.kr
School of Computer Engineering, Kyungil University*
33 Buho-ri, Hayang-up, Gyeongsan-si, Gyeongsang buk-do, 712-701, Korea
Tel: +82-53-850-7287, E-mail: kykim@kiu.ac.kr

Abstract

The people information is distributed in various forms such as database, web page, text, and so on, where the world wide web is one of the main sources of publicly-available people information. It has a characteristic that the information on people is intrinsically temporal. Therefore, the reconstruction of the information is needed for an individual or a company to use it efficiently. In order to maintain or manage the temporal people information, it must distinguish the variable information from invariable information of people. In this paper, we propose a method that constructs an ontology based on events to manage the variable people information efficiently. In addition, we present a system which reconstructs people information that satisfies the users' demand with the ontology.

Key Words : Temporal people information, variable information, Event Ontology

1. Introduction

The people information is itself the historical information of many individuals. In general, it is distributed in various forms such as database, web page, text, and so on. Particularly, the web is an enormous source of publicly-available information on people. However, if someone wants to know the information of a person, he should retrieve all relevant kinds of information from various sources and manually reconstruct them. Therefore, it requires much time and efforts in order to get the consistent person information.

Existing information retrieval systems in general use a vector model of which potential is from the frequencies of matched keywords between documents and a query. For example, when a query is given "From 1960 to 2000, Microsoft Bill Gates' personal history?", then a system defines the keywords as "1960", "2000", "Microsoft", "Bill", "Gates", "personal", "history", and retrieves documents including them using the frequencies. The first document of Google's search results for this query is an introduction of a book related to "Bill Gates", but no information about who "Bill Gates" actually is. Most of the highly ranked documents contain the name "Bill Gates", but their content is just about personal computers.

Many documents hold recent information about "Bill Gates", and they have similar content one another. However, in general, the interests of users about a person are not only about recent information but also about past information. If a document which is about "Microsoft Bill Gates" exists in the form well-organized by someone else, it would be helpful to get personal information about "Bill Gates". However, this is not the general case. In general, the users have to search the information by themselves by using an information retrieval system and have to reconstruct it according to their interests. What is worse, sometimes some keywords are irrelevant with the existing search engines. For instance, among the keywords above, "1960" and "2000" make no sense in searching information about "Bill Gates".

In this paper, we define an event as variable information of people according to time or place, and represent it as an ontology. A search system using this ontology outputs the well-organized reconstructed information on a specific individual in interest, and then the reconstructed information is suitable for users' demand. The ontology for people information is called "Event-Ontology" in this paper.

Manuscript received Oct. 29, 2007; revised Dec. 17, 2007.
*Corresponding author
This work was supported in part by MIC & IITA through IT Leading R&D Support Project.

2. Representing People Information Using Event-Ontology

When representing people information with an ontology, the permanence and the practical use must be considered[1, 2]. It is very important in the variable situation how to define concepts and relations to keep the permanence of an ontology. Therefore, a method is required which constructs an ontology schema for temporal people information.

The information on people can be divided into two types. One is variable and the other is constant. For instance, the birthday and hometown of a person are constant information, but the job and position are temporal. The constant information can be used to discriminate a homonym. The persons whose name is "Bill Gates" can be discriminated according to their birthday or home town. This kind of information can be derived from databases, blogs or homepages.

The *datatype property* of OWL can define the intrinsic properties of persons that are not varying for time or external factors, while the *objecttype property* can define the intrinsic relations between two concepts.

The temporal information for a person is a typical example of variable information. For instance, the graduate information, career, position, awards and all events related to a person are all time-varying. The natural language sentences containing such information are used to extract the temporal information. For instance, consider the following sentence.

Microsoft Bill Gates is_educated_at Harvard.

This sentence is false in this moment. In order to let an ontology have permanence, we represent concepts and relations that do not vary according to time and any external factors.

Actually, one may think that "Bill Gates" who is a chairman of Microsoft is the same person irrespective of time. Even though "Bill Gates" is one person, each event in which he was involved is different. First, Bill Gates was a Harvard student. Then he is the chairman of Microsoft. One solution for this phenomenon is to define time concept, which results in the events represented with concepts and relations. For this reason, we do not represent only the current information of a person, but also represent the time-varying events. Figure 1 shows the schema of Event-Ontology related with People, Place, Time, and Organization, and some examples of instances that represent a temporal event with these relations.

Event-Ontology is defined based on fundamental factors which describe an event. That is, subject, object, predicate and time is represented as properties of an event. A factor which is a subject of an event becomes the property value of *subject*, an action of a subject or a content of an event is the property value of *hasPredicate*. *object* has place or an object of an event. *hasTime* has time information .

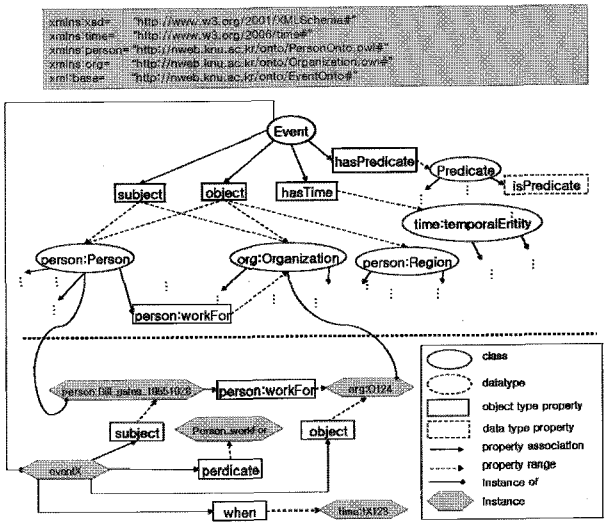


Fig. 1. Event-Ontology and an instance

A subject can be *Person* or *Organization*. In Figure 1, a value of *subject* is *person: Bill_gates_19551028* which is for Microsoft Bill Gates. The property *object* has *Person, Organization,* or *Product* as a value. In this example, *object* has *org:O124* which is the Harvard.

The property *hasPredicate* contains *Predicate* that represents the content of an event. In Figure 1, the original form of syntactic representation like "work for" is a value of *predicate_workFor*, and *hasPredicate* has this instance as a value. In general, one single meaning can be represented in various syntactic forms. The *Predicate* is needed to be represented a concept given in these various forms. A *datatype property, isPredicate* has an original form of the list for these syntactic representations. For example, an instance, *Person_workFor* can be restricted to "work", "belong to", "employ", and "serve". The following sentences have the same meaning but different syntactic forms.

1. Bill Gates works for Microsoft
2. Bill Gates belongs to Microsoft
3. Bill Gates is employed by Microsoft
4. Bill Gates serves the Microsoft

In these example sentences, a predicate information is represented as *Person_workFor*. One can recognize them by preparing syntactic expression in a text and the value of *isPredicate* of the *predicate* instances.

Time factor is defined with existing time ontology, OWL-Time[3]. It can express facts about topological relations among instants and intervals, together with information about durations, and about datetime information. We mainly use it to represent the intervals, it needs a starting point and an ending point. Each

is defined as *datatype property* of *hasBeginning,* and *hasEnd.* If only *hasBeginning* has a value, its instance is the starting point and the event is still progressing. *hasBeginning*, and *hasEnd* take *date* type defined in xml, as a value range. All of these are expressed in the form of "year-month-date".

An event can be recognized by *subject, hasPredicate, object,* and *hasTime*. The same events by a subject are recognized differently by object or time. In addition, the events that occur at the same time can be different events according to the subject of the events.

These properties that describe an event let a system reconstruct event information according to users' demand. Through the query "From 1960 to 2000, Microsoft Bill Gates' personal history", a user wants Microsoft Bill Gates' personal history between 1960 and 2000. To get appropriate results of this query, a retrieval system just searches *Event* instances according to the time axis in the Event-Ontology.

There have been a lot of information extraction researches from structured or unstructured documents. STALKER is a wrapper induction system that performs hierarchical information extraction[4]. The PAPIER system uses relational learning to construct unbounded pattern-match rules for information extraction given semistructured text[5]. BBN IE system extracts information based on statistical method from texts[6]. These researches are applicable to automatic generation of instances from structured or unstructured text for an ontology. Celjuska proposes a method which generates candidate instances from unstructured text[7]. The instances are semi-automatically generated for the user's final decision. Bernardo et al. propose a solution based on CRF for ontology population[8]. KnowItall system extracts instance by using patterns learned from corpus. Yet, generating instances from raw text is under study.

3. A System for Managing People Information

The people information exists in various forms such as database, web page, text, and human knowledge. An ontology is automatically or manually constructed from these information. The various kinds of information are saved as instances or data value according to concepts and relations defined in the ontology.

Figure 2 is the configuration of a system that searches people information based on Event-Ontology by using Protégé API.

This system searches the information within the Event-Ontology through 4 steps.

1. A query is inputted and the concepts of the keywords are identified.
2. Inference an instance which is matched with each concept of the keywords.
3. Inference the event information which satisfies the user's demand.
4. Reconstruct results.

A semantic query processing is performed based on Event-Ontology. Through searching instances which is matched with concepts of keywords, the system can recognize user's demand.

For example, in step 1, a user queries "From 1960 to 2000, Microsoft Bill Gates' personal history." The system searches concepts for "From 1960 to 2000", "Microsoft", "Bill Gates", and "personal history" based on concepts and relations defined in the Event-Ontology. "From 1960 to 2000"can be identified as a time condition by using a pattern. "Microsoft" and "Bill Gates" are identified as *Organization* and *Person* respectively by using instance information. The system searches instances of predicates for "personal history" by using previously constructed mapping table.

For processing a query, this system infers twice based on the Event-Ontology. The first is done in step 2. The system searches the candidates which are *Person* instances, and then, in step 3, infers the event information for user's demand. The persons whose name is "Bill Gates" are restricted by "Microsoft" as *Organization*. The system should search persons named "*Bill Gates*" and who belongs to "*Microsoft*". It searches the event instances of which property *subject* has a *Person* instance and the value of its *hasName* is "*Bill Gates*". At the same time, it checks whether the Property *hasPredicate* of the event instances has a *Predicate* instance, and whether the property *object* has an *Organization* instance and the value of its *hasName* is "*Microsoft*". The property *subject* of finally searched instances has a *Person* instance which the user searches.

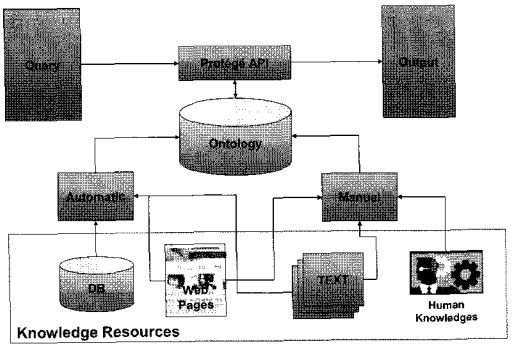


Fig. 2. System configuration

In step 3, the system searches the event information which satisfies user's demand. It searches *Event* instances of which *subject* has the instances searched in the step 2 and *hasPredicate* has an instance related to "personal history". Finally, it searches Event instances between 1960 and 2000. The event information is sorted according to time.

There can be several persons whose name is "Bill Gates" in

"Microsoft". In this case, the result is the event information about more than one person. This system outputs the event information about a person with his property information. Thus, the users can distinguish them.

This system reconstructs the information in an ontology. The instances and data value in an ontology are obtained from various resources such as DB, web page, text, and so on. When using the existing information retrieval system, the users can hardly collect necessary information from the distributed information sources. However, an ontology manages the information in an integrated form, and offers the information efficiently. The event instances which meet the query can be reconstructed in a new form according to users' interests.

4. Evaluation and Discussion

4.1 Correct Semantic Search

Actually, the information retrieval system like Google searches mainly recent articles in high rank and the contents of articles from different web sites are similar one another. Users have to search necessary information from various web pages and reconstruct the results by themselves. However, the proposed system solves this problem. In addition, it is possible to infer new information from existing information by using Event-Ontology. Of course it is unfair to compare the performance of a semantic search system based on the Event-Ontology and a system that searches the web pages, because the semantic search system has a process that constructs an ontology automatically or manually. Nevertheless, the significance of the proposed system is that it can offer the precise results for meeting the users' demand.

4.2 Representing Variable People Information

The Event-Ontology that represents people information with time information maintains and manages not only the recent information but also the past. For the same person, if a property is given with a different value, then it is distinguished by other properties such as time or place. Thus the width of usable people information is broadened. The reliability of information can be guaranteed, because the extracted people information can be used permanently.

4.3 Reconstruction of Results

For several years, the search portals yet have just a list form as results. Only to highly-ranked documents a thumbnail is added or the keyword is shown in bold. Recently, according to the format of searched documents, the results are grouped into some categories such as web page, blog, image, knowledge, cafe, movie, and so on. However, the basic form cannot overcome the list.

The suggested system can output various forms such as generation of text, graph, table, galaxy, cone, tree, and so on (see Figure 3). The semantic contents of the result can be arranged by time or importance. These various forms of result can help users understanding the search results.

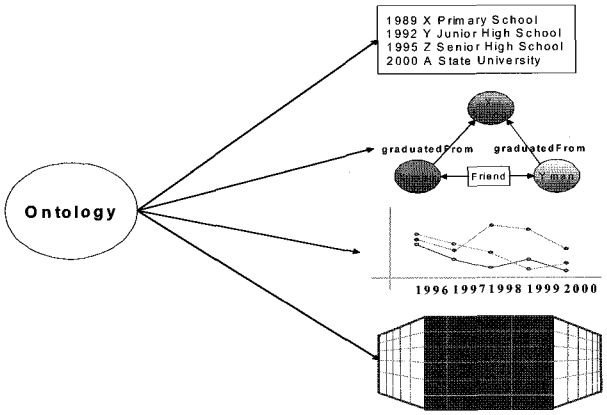


Fig. 3. Various output

4.4 Privacy for Using People Information

Companies can find people's interest using each person's historical information and offer personal service. Already some companies have a marketing strategy based on people information. Each service gets accepted through joining agreements for using personal information, so people offer their information by themselves. This has a limit that it is impossible to get personal information for the people who do not agree. Because the proposed Event-Ontology in this paper uses only the public information, it is free of privacy infringement and thus it is legally possible to get amount of people information online.

5. Related Work

Several ontologies have been constructed for holding people information. The Semantic Web Portal Project managed by Deri proposes a semantic web meta-portal for providing information on people[9]. It describes the necessity of validity to update recent personal information, but ignores the past information for persons which is also important to be a portal on people.

John et al. construct ontologies for information of LiveJournal by using FOAF[10]. However, what they actually do is just to construct a kind of FOAF which is independent of time even though a person is defined in various ways. This is because FOAF cannot represent the temporal information. JRC constructs an ontology to represent the relation between people related to terrorism and presents a retrieval system that can process semantics of query such as organization, event, person

and place[11]. However, they leave a future work to represent temporal information.

For SHOE-Personal Ontology, there is a research about ontology versioning to solve a problem that the concepts in an ontology can change according to time[12]. However, they miss that those which are rapidly changing are not ontologies but the information which is usually represented as instances in the ontology. For example, while a person's job, position and salaries change over time, but some concepts such as job, position and salary hardly change. There have been researches to represent a temporal data in the field of database. Paolo et al. propose a model to represent temporal medical data and a query language[13]. Faiz et al. introduce the tXSchema model and a notation to annotate XML schemas to support temporal information[14]. Since these researches are working with a database, they record temporal information over time and show the information itself, but cannot make a reasoning result.

6. Conclusions

It is an important issue for realizing semantic web how to construct an ontology. The Event-Ontology, an event-based ontology proposed in this paper can represent the variable information. In addition, the existing ontology on the web such as people or organization ontology can be applicable with the Event-Ontology. Therefore, the Event-Ontology can be useful for various applications.

References

[1] Riichiro Mizoguchi, "Tutorial on ontological engineering - Part 1: Introduction to Ontological Engineering", *New Generation Computing*, Vol. 21, No. 4, Ohm-Sha&Springer, pp. 365-384, 2003.

[2] Riichiro Mizoguchi, "Tutorial on ontological engineering - Part 3: Advanced course of ontological engineering", *New Generation Computing*, Vol. 22, No. 2. Ohm-Sha&Springer, pp. 198-220, 2004.

[3] Jerry R. Hobbs and Feng Pan, "An Ontology of Time for the Semantic Web", *ACM Transactions on Asian Language Processing (TALIP): Special issue on Temporal Information Processing*, Vol. 3, No. 1, pp. 66-85, 2004.

[4] Ion Muslea, Steve Minton, and Craig Knoblock, "STALKER: Learning extraction rules for semistructured Webbased information sources", *In Proceeding of the Association for the Advancement of Artificial Intelligence*, pp. 74-81, 1998.

[5] Mary Elaine Califf and Raymond J. Mooney, "Relational Learning of Pattern-Match Rules for Information Extraction", *In Proceedings of the Sixteenth National Conference on Artificial Intelligence*, pp. 328-334, 1999.

[6] Scott Miller, Heidi Fox, Lance A. Ramshaw, and Ralph M. Weischedel, "A novel use of statistical parsing to extract information from text", *In Proceeding of the 1st Association for Computational Linguistics North American*, pp. 226-233, 2000.

[7] David Celjuska and Maria Vargas-Vera, "Ontosophie, A Semi-Automatic System for Ontology Population from Text", *kmi-04-19, Tech Report*, Knowledge Media KMi Institute, 2004.

[8] Bernardo Magnini, Emanuele Pianta, Octavian Popescu, and Manuela Speranza, "Ontology Population from Textual Mentions: Task Definition and Benchmark", *In Proceeding of the 2nd Workshop on Ontology Learning and Population at COLING/ACL2006*, pp. 26-32, 2006.

[9] Anna V. Zhdanova, *People's Community Semantic Web Portal and Metaportal*, Semantic Web Portal Project, DERI SWPP, deliverable 14, 2004.

[10] John C. Paolillo, Sarah Mercure, and Elijah Wright, "The social semantics of LiveJournal FOAF: Structure and change from 2004 to 2005", *In Proceedings of the ISWC 2005 Workshop on Semantic Network Analysis*, pp. 69-80, 2005.

[11] Pinar Oezden Wennerberg, *Ontology Based Knowledge Discovery in Social Networks*, Final Report, JRC Joint Research Center, 2005.

[12] Jeff Hein and James Hendler, "Dynamic ontologies on the web", *In Proceedings of the 7th National Conference on Articial Intelligence*, pp. 443-449, 2000.

[13] Paolo Terenziani, Richard T, Snodgrass, Alessio Bottrighi, Mauro Torchio, and Gianpaolo Molino, "Extending Temporal Databases to Deal with Telic/Atelic Medical Data", *In Proceedings of the American Invitational Mathematics Examination*, pp. 58-66, 2005.

[14] Faiz Currim, Sabah Currim, Curtis Dyreson, and Richard T. Snodgrass, "A Tale of Two Schemas: Creating a Temporal XML Schema from a Snapshot Schema with tXSchema", *In Proceeding of the Extending Database Technology*, pp. 348-365, 2004.

Yong-Jin Han

Yong-Jin Han received his B.S. degree in the Department of Computer Engineering from Kyungpook National University, Korea, in 2006. He is currently in a Master's Program in the Kyungpook National University, Korea. His research interests are in semantic web, social network analysis, information retrieval and natural language processing.

E-mail : yjhan@sejong.knu.ac.kr

Seong-Bae Park

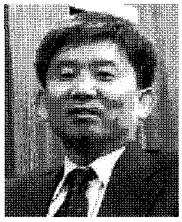
Seong-Bae Park received his B.S. degree in the Department of Computer Science from Korea Institute of Science and Technology in 1994, and M.S. and Ph.D. degree in the Department of Computer Engineering from Seoul National University in 1996 and 2002 respectively. He has been a faculty member of the Department of Computer Science at Kyungpook National University since 2004, in which he is an assistant professor. His research interests are machine learning, natural language processing, information retrieval, and semantic web.

Phone : +82-53-950-7574
Fax : +82-53-957-4846
E-mail : seongbae@knu.ac.kr

Young-Hwa Lee

Young-Hwa Lee received her B.S. , M.S. and Ph.D. degree in the Department of Computer Engineering from Kyungpook National University, Korea, in 1991, 1993 and 1999 respectively. Since 2001, she has been a faculty member of the Kyungpook National University. Her research interests are in Natural Language Process, Ontology.

E-mail : allyh@knu.ac.kr



Se Young Park

Se Young Park received his B.S. degree in the Department of Electronic Engineering from Kyungpook National University, Korea in 1980, and M.S. degree in the Department of Computer Engineering from KAIST, Korea in 1982, and Ph.D. degree in the Department of Computer Engineering from University of Paris VII, France, in 1989. From 1982 to 2000, he had been a principle researcher in ETRI, Korea. And from 2003 to 2005, he had been an expert adviser in IITA. Since 2005, he has been a faculty member of the Department of Computer Engineering at Kyungpook National University, where he is currently a Professor. His research interests are in information retrieval, semantic web, digital contents, and social network analysis.

Phone : +82-53-950-5550
Fax : +82-53-950-2122
E-mail : seyong@knu.ac.kr
http://nweb.knu.ac.kr

Kweon-Yang Kim

Kweon-Yang Kim received his B.S., M.S. and Ph. D. degrees in the School of Computer Engineering from Kyungpook National University, Korea, in 1983, 1990 and 1998 respectively. Since 1991, he has been a faculty member of the School of Computer Engineering at Kyungil University. Before he join the Kyungil University, he was employed by ETRI as a member of technical staff. He had been visiting academy of University of Central Florida from 1999 to 2000. His research interests are semantic web and Korean engineering.

Phone : +82-53-850-7287
Fax : +82-53-850-7609
E-mail : kykim@kiu.ac.kr
http://ce.kyungil.ac.kr/~kykim