An Ontology Framework For Context-Based Multilingual Document Retrieval

Priyadarshini M.R
Department of PG Studies
S.A. Engineering College
Chennai, India
priyaram1985@gmail.com

Abstract— This paper aims to present a multilingual ontology model for retrieving documents. This technique is based on context based multilingual translation. Though English is a common language in the world, at times there are situations where there is no chance in rural and remote places. For each given language, multilingual ontology is used to map to a given set through indexed processing and querying. The users may be software engineers, students or working professionals who want to understand the given document in their native language rather than the language in which the document is given. It helps in understanding and conceiving the concepts of document in their native language. Thus, it increases the efficiency of the users who are unaware of the language where they set in by increasing the level of understanding.

Keywords- Multilingual ontology, Indexed processing and querying, context-based multilingual document retrieval

I. INTRODUCTION

The importance of software development at multisite has been drastically increasing, since its advantages have been realized by major corporations throughout the world. Most of the corporations have moved their software development to various countries where employees can be paid low wages and the same sort of work can be retrieved from the employers. In this regard, employers sometimes become inefficient due to lack of proficiency in Multilanguage. This produces the financial gain that has driven people to have their business to multisite development and it is facilitated by the internet.

This globalization of software development has led to linguistic problems. In a software development team, team members who perform the various tasks and activities assigned by their respective team leaders who in turn control the tasks and activities. It is not necessary that managers and higher officials of the software development need to be present at the same place. It is quite natural that all the members who are all involved in the project would possess the same native language or all of them are well-versed in a single language from the top hierarchy to bottom hierarchy of the organization of the project.

Software engineering training and practices are also different between various cities and countries. It can be difficult to explain the document in each and every language to everyone present in the team.

Sometimes, when issues are raised, they are ignored because they are not understood and it becomes difficult to get clarified because of diversity of native languages in the team. Thus, there is a lack of understanding which may lead to incorrect model of the software development.

Despite this, Software development has common understanding body of knowledge and is an easily learnt subject that includes technologies that are recently developed and new methodologies that are easily adopted. However, different team members would possess different native language and will be well versed in that language. There are situations that a document given to a team member in one particular language, where the member does not know the language or knows partially, issues will be raised in understanding the theory.

Consequently, these problems that arose and the underlying issues are important to consider because understanding is the real challenge that is faced day-to-day. This is very tedious and it is not always possible to resolve with the help of other team members. Thus, to eradicate this kind of issue, once the document which contains information about the problem definition or the requirement document with the method of multilingual ontology the document can be translated into their own respective native language without changing the concept of the original information contained in the document.

Thus, this helps the better understanding of the concept by the individual team members by clearing up the ambiguities in understanding the concept.

II. MULTILINGUAL ONTOLOGY AN OVERVIEW

A. Requirement of Ontology for Multilanguage

Currently, most of the organizations working in multilingual environment demand ontologies supporting different natural languages. Consequently, the inclusion of the multilingual information retrieval is not an option but a must. In general, ontology is the study of reality. More specifically, ontology is an expression of a particular model of reality, including a specification of concepts, relationships among concepts, and constraints that exist in the model.

It has enough information in the model to be able to drive the process of extracting concepts and relationships from source document. Ontology in computer science and information science formally represents knowledge as a set of concepts within a domain, and the relationships between pairs of concepts.

B. Procedural Overview of Multilingual Ontology

Interoperability plays the major role in multilingual ontology. Here, the matching methods are important because it requires automatic searching and pattern matching of words of similar pattern or dissimilar pattern. In this case, the script of the language plays a major role for efficient comparison and retrieval. The frequency of word usage is also concerned here where the retrieval rate is higher than the infrequent or new word usage. This word frequency usage is found by highest number of repetitions of the words in the source language and its requirement in the target language.

III. RELATED WORK

A. Exploiting the Document Retrieval

Wiktionary is freely available. This contains vast data with a large database of words that has incorporated the translations to many languages. Document retrieval has become a very tedious task for search engines on the Web, because the growth speed of the number of web site and. In the other hand, the searching methodologies used currently are not suitable to better exploit the vast quantity. This element makes the document retrieval by the current search engines very difficult and does not meet the users' needs.

B. Need For Multilingual Translation

This concept will be more useful for the engineers who have their educational background in their native language. Suppose, consider a person from any of the rural parts of the country will be well versed in coding, but lacks with the knowledge of understanding the concept given in the document which may not be in his/her native language. So he/she has to go in search for another team member for helping to understand the concept.

Instead, he/she can use the document translator and can translate the document into his own native language, and conceive the contents of the document. Even though the performance of translating the document may not be high, but it will some the engineer to capture the actual content in the document, which is really a great boon when he/she is working at multisite/multilingual software development environment.

Even now, there are countries that are language patriotic and give importance to their national language. The engineers who may be well verse in coding but may lack in conceiving the actual content of the given document other than his national/native language, because the learning process restricted to one language. In such scenarios, this translation for any document which contains any information related to project will be very useful for the software engineer at his/her early stages of the career development.

IV. PROPOSED MODEL FOR DOCUMENT RETRIEVAL

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

A. Structural Document

The structural overview of the proposed concept is a framework and the proposed framework is called software document knowledge management (SDKM). This framework with multilingual ontology helps to retrieve the documents in their language which the user wants.

This framework has three phases (a) Document processing (b) Translating process (c) Retrieval of required document. The translation process is more responsible for obtaining the most probable translation for each document given. The given document is analyzed and most frequently used words and their translations are retrieved from the knowledge database and the term which occurs for the first time and their translations are retrieved from the web-dictionary.

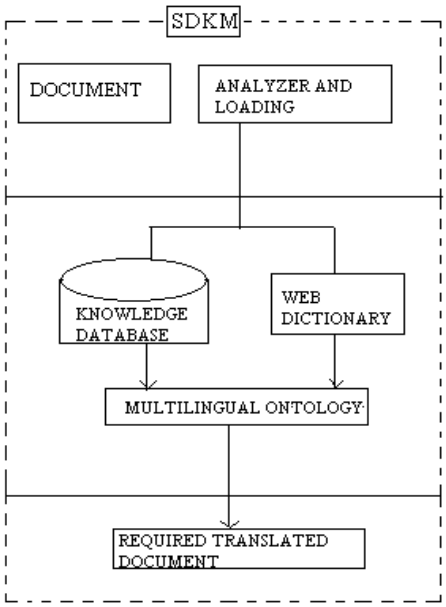

Figure 1. Software Document Knowledge Management

This (figure 1) framework relies on a relatedness measure based on the translations. This is based on comparing the text associated to each possible translation and their text. The analyzer captures all the linguistic information associated with concepts. This framework is designed for the representation of multilingual information ontologies. In the proposed framework, the link we establish between the ontology and the associated translations is characterized by simple references between concepts.

The document is analyzed and loaded to retrieve the required content from the knowledge database and web-dictionary. This retrieval process is a parallel process and is obtained as follows, (i) The document is searched for more frequently used terms that is stored in the database and the translated term is easily retrieved and the term which occurs for the first time it searches in web-dictionary which is a collection of terms of all languages then the translated term is retrieved.

B. Functional Overview

This functional overview briefly describes the interaction of the components shown.

The main activities are

(i) The user chooses the document to be translated.
(ii) The document is analyzed and loaded in to the system.
(iii) From the loaded system, using the multilingual ontology looks for the translations parallel in both knowledge database and web-dictionary.
(iv) The system is updated if there is any new linguistic information.
(v) Finally, the linguistic information associated with the document shown to the user.

Thus, these are the functions related to the Software document knowledge management.

C. Evaluation For Multilingual Ontology

The software engineering ontology evaluation is an important task that is needed in many situations. For example, during the process of building of ontology, ontology evaluation is important to guarantee that what is built meets the application requirement. The way data is placed within ontology is also a very important measure of ontology quality because it can indicate the effectiveness of the ontology design and the amount of real-world knowledge represented by the ontology.

V. CONCLUSION AND FUTURE ENHANCEMENTS

In this paper, the need for document translation in the native language is analyzed. Then the structure of the framework is designed and its functions are specified. The document translation that has been discussed is with respect to the concept oriented translation which in turn helps the user to conceive the actual concept.

However, the efficiency of the performance of the proposed concept would be best if the source language and the target language have similarities. The limit of the proposed system is that efficiency of the performance degrades, if both the source language and the target language completely vary in their script.

REFERENCES

[1] Salam, F., "New Semantic Indexing and Search System Based on Ontology", Fourth International Conference on Emerging Intelligent Data and Web Technologies (EIDWT), 2013, pp. 313–318, Sep 2013.

[2] M. C. Suarez-Figueroa, (coordinator). NeOn Development Process and Ontology Life Cycle. NeOn Project Deliverable 5.3.1 ., 2007.

[3] Cross, V.V., "Intelligent query construction for multilingual document exploitation" 19th International Conference of the North American Fuzzy Information Processing Society, pp. 232 - 236, 2000.

[4] Almeida, J. and Simes, A. (2006). T2O: Recycling Thesauri into a Multilingual Ontology. Fifth international conference on Language Resources and Evaluation, LREC 2006, Genova, Italy, May.

[5] Enriching Ontology with Multilingual Information Mauricio Espinoza and Eduardo Mena., 2006.

[6] Yoshinaga, K, "Multi-lingual intelligent information retriever with automated ontology generator", Third International Conference Knowledge-Based Intelligent Information Engineering Systems, pp-62-65,1999.

[7] Zagorulko, Y , Zagorulko, G ., "Ontology-based program shell for building and editing multilingual thesauri of subject domains" , IEEE 12th International Conference on Intelligent Software Methodologies, Tools and Techniques (SoMeT), pp. 96-106, 2013 .

[8] Achour, H., Zouari, M., "Multilingual learning objects indexing and retrieving based on ontologies", World Congress on Computer and Information Technology (WCCIT), pp-1-6, 2013.