Semantic Matching Using Ontology in Multilingual Environment

Makrant I., Mansi M.

Abstract: The tremendous increase in usage of data over the past few years and the ease of availability of things across the globe any time lead to the Advancement of multilingual database. Storage, retrieval and archiving of data for multilingual system has been a challenge. Our research project “Semantic Matching using Ontology in Multilingual Environment” is an extension to look into addressing multilingual data. The report focuses on providing design and implementation of multilingual system. These comprise of two main components being (i) Cross Lingual Information Retrieval, and (ii) Indian Language to Indian Language Machine Translation. Consider the context of large scale natural language processing applications in the areas of Cross Lingual IR and Machine Translation, wherein such a model for multilingual dictionary is established. When contrasted to traditional single lingual or bilingual dictionary, the model uses the core concept of Synonym Groupings (synsets) that is used as a way to connect different languages in a crisp and efficient manner.

Keywords: Multilingual database, Storage, Retrieval, Hierarchy, multilingual dictionary, Translation, Algorithm

I. INTRODUCTION

The primary goal of the system is to help the people to get familiar with expressions in everyday conversations and typical dialogues in the target languages (Marathi, Hindi) that can take place in public places like a shopping mall, a railway station, a bus-stop, a restaurant, a post office, a bank etc in a systematic way. To help them better understand the spirit of the city an idea of the most used dialogues in Marathi can be helpful.

Multilingual Database is a database in three languages – Hindi, Marathi and English. It is a search and retrieval system for digital Documents. With the increasing integration of global economy and proliferation of languages other than English into information systems, capability to store and manage data in multiple languages simultaneously is of vital importance. A hierarchical tree structure called ‘Ontology’ is used. Applications that are catered to domain-specific knowledge being to use Ontology extensively. This is used for semantic matching.

The existing system has following features:
   a. Ontology is in English.
   b. Synsets are available in different languages.

Computational Framework for the Multilingual Dictionary: For effective implementation of our idea of synset based multilingual dictionary, we carefully designed the dictionary development process, which is, in fact, expected to involve a number of human lexicographers. Figure below shows the complete semi-automatic data flow in the dictionary development process.

II. PROBLEM STATEMENT WITH SCOPE & OBJECTIVE

The customer opens the web-page. He will select the transaction he wants to perform. There are two transactions to be performed. In case of INTER-LANGUAGE TRANSLATION, the user will select a source language and a target language. Then he will enter a word from the database to be translated. In case of SEARCHING OPERATION, he can either input a certain price range and get the products in that range, or can search a certain brand in the hierarchy. The important factors considered are: whether methods used are fast and convenient? and is the database complete for the shop?

III. PROPOSED SYSTEM

Our proposed system will include:
   a) Ontology in 2 other languages: Hindi and Marathi.
   b) Semantic matching in the three languages for words.

An ontology is a shared conceptualization of knowledge in a particular domain. There is need to separate the knowledge about the target domain from the rest of the application code. Ontology helps to facilitate the above. The key benefits of this approach are: simplification of the application code, possible sharing of knowledge among multiple applications, and the flexibility of evolving the knowledge without requiring changes to the application.
This paper addresses the problem of supporting ontology-based semantic matching in RDBMS.

**Word-Alignment in the Proposed Model:** The problem of correct lexical choice on the target language side was tackled by proposing a novel approach of word-alignment across the synsets of languages. Word alignment refers to the mapping of each member of the synset with the most appropriate member of the synset of another language. Consider the example wherein the word under selection is लड़का 'boy' in Hindi. In this case of 'a young male person' needs to be lexically transferred to Marathi, there are four choices available in the synset, as illustrated in figure 2.

![Figure 2 Illustration of aligned synset members for the concepts: a youthful male person](image)

Considering Hindi as the pivot, we proposed that each of the four words in Marathi synset be linked to the appropriate Hindi word in the direction Marathi → Hindi and each of the two words in English synset has to be linked with appropriate Hindi word in the direction English → Hindi. As a result, the first and the third member of the Marathi synset (i.e. मुलगा and पोर) are mapped to two different Hindi words (मुलगा → लड़का, पोर → बच्च) in the Hindi synset. The second and the fourth member in Marathi synset are linked to one word (पोरगा → छोकरा and पोरगे → छोकरा) in the Hindi synset. Three words in Hindi synset i.e. HW4, HW5, HW7) are left without being linked, as seen in the figure 2. In a situation, when a Marathi word is aligned with a single Hindi word (eg मुलगा → लड़का) for a particular concept in the direction of Marathi to Hindi, from our past experience we assume that the lexical transfer in the reverse direction also holds good, yielding लड़का → मुलगा.

**IV. PROJECT DESIGN**

1) System design: The flow diagram is illustrated in the figure 3. NetBeans IDE 6.7.1 is used for the Front End. Glassfish server is used to connect the JSP and the database in MS Access.

![System architecture](image)

2) **Algorithms:**
   a) Specific algorithm:
   1- Choose source language
   2- Find hierarchy of key word in source language
   3- Start from the ROOT node.
   4- Search among the children.
   5- If the keyword selected by the user is found, display its children in the next dropdown for the next selection.
   6- If translation is selected as the operation, assign it as the curr_node_level
   7- Now we search hierarchy level in target language
      A- If ( curr_node_level = target_level1)
      Target language Translate
      B- Else If ( curr_node_level = target_level2 : leaf node)
      Target language Translate
      C- Else If ( curr_node_level = synset of source_level2 : leaf node)
      Target language Translate
      Else
      D- Word not found in database

   b) General algorithm:
   i) To find a match between Hindi & English synsets:
      for each synset identity english_synset_id in English Ontology do
      if (english_synset_id == hindi_synset_id) do
      for each relation r pointed by hindi_synset_id do
      if (relation type of r is semantic) do
      clamp the synset identity linked by relation r in to english_synset_id
      else
      clamp the synset identity linked by relation r in hindi_synset_id to english_synset_id AND manually insert the corresponding lexical element
      end if
      end if
      end else
      clamp the synset identity linked by relation r in hindi_synset_id to english_synset_id AND manually insert the corresponding lexical element
      end else
      end for
      end if
      end for
   Eg.: Here we have hindi input to English output
The example is as follows:

<table>
<thead>
<tr>
<th>Hindi Word</th>
<th>English Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>दुकान</td>
<td>shop</td>
</tr>
</tbody>
</table>

Take the word दुकान.
Consider the id number of the word.
Now search for the id number in the English set.
If match is found, display the output.
Else manually insert the word in the table with an id number.

ii) To find a match between Hindi & Marathi synsets:
for each synset hindi identity marathi_synset_id in Marathi Ontology do
if (marathi_synset_id == hindi_synset_id)
do
for each relation r pointed by
hindi_synset_id do
if (relation type of r is semantic) do
clamp the synset identity linked by
relation r in to marathi_synset_id
end if
else
clamp the synset identity linked by
relation r in hindi_synset_id to
marathi_synset_id AND manually insert
the corresponding lexical element
end else
end for
end if
end for

Eg: Here we have hindi input to English output
Marathi word | English Word
दुकान        | boy

Take the word दुकान.
Consider the id number of the word.
Now search for the id number in the English set.
If match is found, display the output.
Else manually insert the word in the table with an id number.

3) Uml Diagrams:

Fig.4 Use case diagram

Fig.5 Class diagram
5) Data Diagrams:

6) Interface Design (User Interface):

The following are the screenshots of the user interface:

![User Interface Screenshot](image1)

This Project help locate the price of a Product.

Choose the language:
- English
- हिंदी
- मराठी

Choose Dept: Software

Choose Category: Mobile

Choose Brand: Nokia

Available products (with price) are:

<table>
<thead>
<tr>
<th>Product</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>H_N45</td>
<td>200.0</td>
</tr>
<tr>
<td>H_E63</td>
<td>300.0</td>
</tr>
</tbody>
</table>

![User Interface Screenshot](image2)

Back to Home Page
This Project helps locate the products in the specified Price range.

Choose the language:
- English  Ṣiṣẹ́  Ṣiṣẹ́

Enter Price Range:
- From: 100  To: 100

Available products (for the specified price range):
<table>
<thead>
<tr>
<th>Price</th>
<th>Product Brand</th>
<th>Category</th>
<th>Dept</th>
</tr>
</thead>
<tbody>
<tr>
<td>200.000 N45</td>
<td>NOKIA Mobile</td>
<td>Electronic</td>
<td></td>
</tr>
<tr>
<td>300.000 N45</td>
<td>NOKIA Mobile</td>
<td>Electronic</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 11 Front end for Price Range operation

V. PROJECT TESTING

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Test Name</th>
<th>Test Objective</th>
<th>Test Configuration</th>
<th>Procedure</th>
<th>Action</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>VIEW TRANSACTIONS:</td>
<td>View Transactions:</td>
<td>Installation of application</td>
<td>Launch the application.</td>
<td>1.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>The user should be able to see</td>
<td></td>
<td>2. Select VIEW TRANSACTION option</td>
<td>2.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>all the previous transactions</td>
<td></td>
<td>from the screen</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>he has performed from the current</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>account. He can also select a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>particular transaction from the</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>list to view the complete</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>details.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>CANCEL TRANSACTION:</td>
<td>Cancel Transaction:</td>
<td>1. Installation of application</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>The screen allows a user to</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>place a request for the</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>cancellation of a particular</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>transaction and only if the</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>customer selects to cancel, a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Expected Results
- If there are any transactions performed then a list of all will be displayed else an error message stating “No Transactions” will be displayed.

Actual Result
- AS EXPECTED

PASS/FAIL
- PASS

Sr. No. 2

Test Name
- CANCEL TRANSACTION: This screen allows a user to place a request for the cancellation of a particular transaction and only if the customer selects to cancel, a
request the transaction is cancelled.

<table>
<thead>
<tr>
<th>Test Objective</th>
<th>Client Testing(functional)</th>
</tr>
</thead>
</table>
| Test Configuration | 1. Installation of application  
2. Initiation of application  
3. Display of the first screen |
| Procedure | Launch the application. |
| Action | Response |
| 1. Launch the application | 1. Confirmation displayed |
| 2. Select CANCEL TRANSACTION option from the screen | 2. Cancel transaction if both the participating entities have placed a cancellation request |
| Expected Results | The cancellation-accepted request message will be displayed |
| Actual Result | AS EXPECTED |
| PASS/FAIL | PASS |

VI. REFERENCES


