PART - IV

RESULTS AND DISCUSSION
A. PHONETIC MATCHING

In this thesis, we had proposed two approaches for phonetic matching. First, is based on different writing styles of the string considered as various phonetic structures for the string. Second, is based on phonetic rules for the Hindi and Marathi languages. We compared both approaches with existing proposed approaches. The result shows that the proposed method is an efficient for retrieval of multilingual data, when search is given in local languages than using WordNet, which may need more processing power and memory than our approach. We found our approaches are simple, easy to implement for any domain and gives accurate results.

The sample screen-shots of the result for the application are as shown in figure 5.5 and figure 5.6. Figure 5.5 is taking local language keyword as input in Hindi and figure 5.6 is the outcome for the same query. We can select any local language outcome, depending on user’s need. The result shows that the efficiency does not depend on number of phonetic primitives, but depends on minimum number of keys to be mapped to enter a local string [5].

In this approach, each string is being parsed and translated into English for comparison. If matched is found from database, the result are being shown as IR result.

In the second approach, it is supported by the phonetic rules in order to match the strings for Hindi and Marathi languages. We explored and compared our approach with some of the phonetic matching approaches like soundex, Q-gram and so on for Indian languages such as Hindi and Marathi. We have found out that these approaches are lagging in accommodating all the characters from an alphabet for both Hindi and Marathi languages. This approach also supported for IR after phonetic matching with phonetic form of a string stored in database. This approach is without using IPA code, thus reducing the ambiguity for matching. Advantages of our approach are that it gives the user and developer a simple, easy and efficient way for phonetic matching.

As we are using IME approach for inputting the text, the efficiency of an algorithm has increases exponentially since it reduces the efforts and time to enter. More memory is utilized in phonetic approach. So, the performance will not be an issue with respect to memory.
consumption. For better performance, we can restrict to use this algorithm to only text-based processing applications. The future scope for the research can be extended for multilingual multimedia matching using different approach.

**B. SEMANTIC MATCHING**

Ontology plays an important role to represent knowledge, which can be used for semantic matching. There are many ontology building methodologies suggested for various domains. With our proposed methodology, we can effectively build the domain ontology with all possible user scenarios using simple and complex keywords. We had developed the concepts and their relationships dynamically. Also, this methodology is faster than the earlier one, since we are using top-down approach to build the ontology.

Figure A shows the main user interface screen, where either customer or user may use the system in order to select sub-domain. Figure B shows the choice of language for the customer or user for sub-domain.

Figure C and figure D show the questions designed for the sub-domain in order to select the items required for the sub-domain. It has advantages, first, the customer gets acquainted about the items required for the sub-domain, second, the customer gets choice to select from the probable answers provided or if he may use his knowledge to answer the question. With those answers, the system has prepared the ontology terms to build ontology.

Finally, we built the ontology for the selected sub-domain as per figure E. In this figure, the main concept has been shown as recipe and sub-concepts are ‘chhole’ and ‘puri’. For each sub-concept, concept levels have been added as per answers from the Q-A module and finally its properties as values have been shown as a leaf node.

Once the data is represented in ontology form for a sub-domain, it is being used for inference system. The system has been extended to compute any inference from this ontology. Figure F and figure G shows the user interface for inference queries for Hindi and Marathi users respectively. We formulate inference queries for Hindi and Marathi users for the sub-domains, which uses ontology as a backend database to acquire the knowledge as a result for the query.
Figure A: Main User Interface for Selection of Sub-domain

Figure B: Main User Interface for Selection of Language
Figure C: Question-Answer (Q-A) Interface1 to build Ontology

Figure D: Question-Answer (Q-A) Interface2 to build Ontology
Information Retrieval in Multilingual Environment: Phonetic and Semantic Along with Ontology Approaches for Hindi and Marathi

Figure E: Ontology for Sub-domain
Finally, we evaluated the entire system by using performance parameters like precision, recall, F-measure and accuracy which gives the overall accuracy of the system. So, our phonetic and semantic approach deals with solving the problems arising when the user doesn’t know the exact keywords or meaning but he can acquire the information from the system.