Relation Extraction for Tourism Domain

P.Shanthi Bala[1], Dr.G.Aghila and Jayanthi

Department of Computer Science, Pondicherry University, Puducherry

[1]shanthibala.cs@gmail.com

Abstract

Information Retrieval is crucial in today's Internet World. System should have complete knowledge to provide the information to the user based on their query. This paper explores the use of relations in information retrieval system for Tourism domain. A relation provides syntactic and semantic information between pair of concepts. Relation extraction captures the relationship between multiple concepts. In this paper, concepts and binary relations have been identified for Tourism domain. It extracts the relation between the concepts. Our approach explores the new relationship even if it is not available. The extracted information gives meaningful and complete information. It also appends additional information into knowledge base.

Keywords-- **Information Retrieval, Reasoning, Tourism**

I. INTRODUCTION

Many researches focus towards Information Retrieval for effective retrieval of information. To achieve, the information should be represented properly. Knowledge representation technique is used to represent the complete knowledge of a particular domain. It explores complete knowledge of the domain by the suitable knowledge representation techniques. It could be useful to extract meaningful information. It could be achieved by representing concept; relationship exists between the concepts, its attribute and instances. Semantic relation identification between the concepts automatically is a challenging task. But it is very useful for many applications. It can be applied to information retrieval to retrieve relation between concepts. It plays a critical role in information extraction. This paper identifies semantic relations for tourism domain and provides information based on their query. Information retrieval involves identifying a relation which exists between various concepts in Tourism domain.

Reasoning is the ability to make inferences from the available knowledge automatically. Understanding the meaning of multiple concepts should allow a learner to predict the kind of relations exists between these concepts. But system should understand it and should make inferences. It is necessary in modern world. Reasoning must have ability to reason under certain conditions like compensate for lack of knowledge, resolve ambiguities and contradictions and update the knowledge base over time.

Relations provide more semantics between multiple concepts. Without relation, it will not provide complete information.

Relation is used to provide more semantic between pair of concepts. The semantic relation is not limited to is-a, part-of relation. In addition to these basic relations, each domain has many semantic relations.

Basic relations are very complicated to achieve high accuracy in identifying semantic relations from various concepts. It varies depends on the context, semantics and algorithm used to infer the knowledge, representation of the knowledge, etc. Thus relations provide more information about concept and how they are related to each other's. This approach involves concept identification and extracting relationships between given concepts.

This paper is organized as follows: Section 2 describes how the semantic relation can be extracted and used. Section 3 explains knowledge explored in tourism domain. Section 4 provides implementation details. Finally, section 5 concludes the paper.

II. RELATED WORKS

Sung-Hyon Myaeng and Christopher Khoo described the method to identify semantic relations automatically and accurately in the text are increasingly important in information retrieval and information extraction systems [1].

Sridevi and Nagaveni explained information retrieval systems rely on textual keywords to index and retrieve documents. Keyword-based retrieval may return inaccurate and incomplete results when divergent keywords are used to describe the same conception in the documents and in the queries [10].

Josh King and Venu Satuluri explained lexical semantic relations to encode knowledge about the world (common-sense or otherwise) and hence they are essential for many natural language applications such as question-answering, natural language understanding and information retrieval [3].

Yan Li and JianWen Zhoujun Li describes about relation using an Improved Relation-Based Information Retrieval Technique. The most significant effectiveness of relation is to improve the performance of the information retrieval systems, because currently none of the information retrieval system can search implicit or unknown information based on user's query [2].

The extraction of relations is a recent research focus in the area of information extraction. Relation can be extracted in supervised or unsupervised manner. But only few types of relations can be extracted. One of the main problem is used to extract the biological relations is indexing. Usually in data mining, the documents are retrieved by the number of occurrence the given terms in the document [10].

Next section describes tourism domain knowledge which is used to extraction the relation.

III. CONCEPTS IN KNOWLEDGE BASE

Now a day, most of the people engages their vacation in some visiting places. In modern world, the individual can gather information about places of visit, transport facility and stay, etc. Still an individual faces many problems for visiting unknown places. Now, many Tourism corporations fulfill their wishes. This paper explores the tourism domain and it helps to the end user to know about any information about the domain.

Tourism plays major role in entertaining all sorts of people from young age to elder, poor person to rich one's. It includes information such as accommodation, transportation, season to visit, guide services, places of visit in the destination. Due to the heterogeneity of the tourism package, a tourist agency guide or a person who is planning the tour itself must have access to different sources of information. This paper explores all the essential information.

The following Knowledge is identified in Tourism domain:

- General information like places of visit, distance from source to destination and duration to reach and mode of transportation, etc.
- Mandatory information like schedule for the travel based on mode of transportation, Accommodation, Fare, etc.

- Timing schedule information such as arrival time, Departure time, etc
- Food services in the hotel nearby visiting places based on the location of the tourist.
- Further procedures like preparation for tour based on the places of visit, room booking based on the person in the family, etc
- Other affordable services based on the tourists.
- Extra services based on the tourist expense and distance from beach, shopping, etc

In this paper, the following concepts are identified.

- *Customer*
  Customer is the person who is availing Tourism facility.
- *Customer Details*
  Details include name, age, native, address, family details and contact details, etc
- *Tourism*
  It includes packages, places of visit, etc
- *Transport*
  It includes road, air and water transport.
- *Duration*
  Duration regarding the travel during the vacation is mentioned.
- *Date &Time*
  Date and time of the travel is noted.
- *Location*
  Location is a place where the customer desires to visit during his travel.
- *Accommodation*
  It includes places and hotel, location, time and amount details etc.
- *Air*
  It can be provided to the tourist based on their budget.
- *Water*
  It includes ship or boat travel.
- *Road*
  Usually majority of the places can be visited only with the use of road transport.
- *Business*
  It can be available on airlines. It is intermediate airline service between first class and economy class. Now many airlines offer this service as the high level of service.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)
International Conference on Information Systems and Computing (ICISC-2013), INDIA.

- *Domestic:*

  This kind of services provided within the country. The person belong to the country can easily move around with the uses of this service.

- *Car*

  It provide luxurious mean of transportation. It is cheaper and took less time to reach particular destination. With the use of this facility, a tourist can reach any place (particularly hill area).

- *A/C*

  The A/C cars are provides luxury to the customers in their travel. Apart from A/C, all the other basic facilities are provided at nominal price rate.

- *Non A/C*

  Non A/C cars are cheaper than A/C cars and contain all other requirements at nominal rate affordable to the user.

- *Bus*

  Buses play a major role in tourism. Buses used to visit the local attractions or scenery.

- *Sleeper*

  A sleeper bus provides more affordability with some increase in cost.

- *Semi Sleeper*

  A Semi sleeper provides some more affordability as compared to ordinary buses.

- *Cruise*

  Cruises are concentrated towards short sea journeys of about a week. It provides floating resort services to enjoy luxury and entertainment while moving towards multiple destinations.

- *Classic Cruise*

  It is associated with relaxing, friendly and informal style of cruising and a memorable "value for money" cruising experience.

- *Family Cruise*

  It offers facilities to take of the family including the kids. Some family cruise is friendlier than others.

- *Luxury Cruise*

  It provides separate ships may be smaller in size and the staff to guests. It offer luxury services and the expectations of the user have been exceeded.

- *Hill Station*

  Huge number of tourists visits hill stations. It is popular world-wide for its ethereal attractiveness, healthy climate, breathtaking landscapes and attractive charisma. Many tourists relax their vacation in the lap of the nature.

- *Temples*

  India is a traditional country. It has many historical temples and many foreign tourists visit temples for its unique architecture. Beaches, Hill stations, Wildlife safari, historical monuments, pilgrim centers are some of the major attraction of India Tourism.

- *Desert*

  Rock outcrops, mountains with strange shapes, and deep gorges are the most popular things in the desert.

- *Guide*

  The tourist guide's plays a major role in the tourism. They provide more historical information about particular place.

- *Details*

  The more details provided in the tour guides in terms of entertainment. It helps to keep all the tourist to be happy, comfortable journey and satisfaction.

Figure 1 shows some of the concepts and relation between those concepts in the tourism domain. Next section provides implementation details of relation extraction.

IV. IMPLEMENTATION

Representation of the knowledge is very important in the reasoning task. Many languages have been used to represent the knowledge. Some of them are RDF, OWL, etc. We have used triple format to represent the knowledge of Tourism domain. It has many advantages. It is very useful for reasoning. Our knowledge contains 25 concepts and 15 relationships between the concepts. The knowledge is represented in the following format:

<Concept1, Concept2, Relation>

Some of the concepts and relations are represented in Table 1.

In this work, we have analyzed all the new relations. It proves that the relation retrieved is correct. Figure 2 shows the architecture representation on relation extraction for Tourism domain.
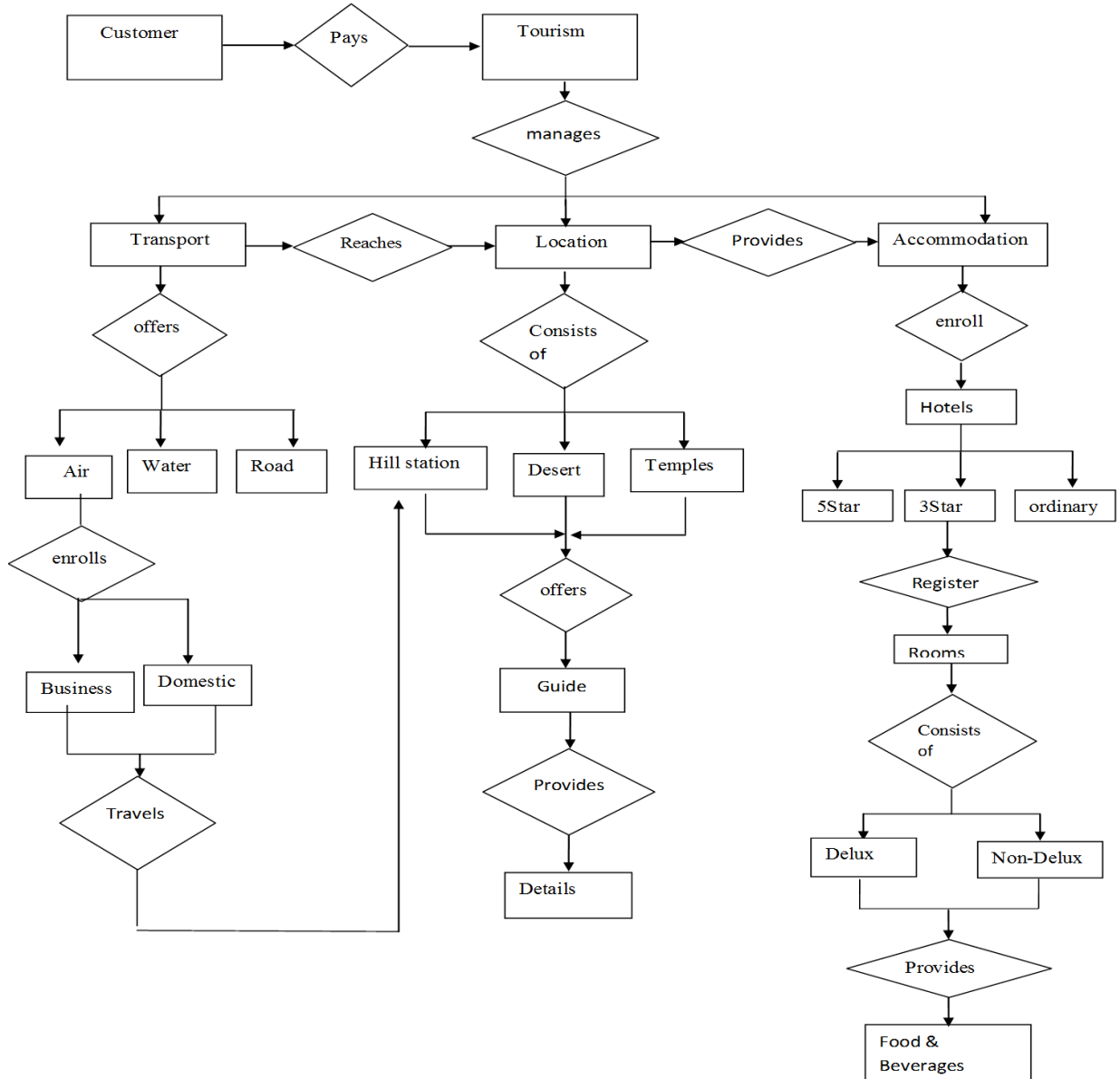
Fig 1. Concepts and relations in Tourism

International Conference on Information Systems and Computing (ICISC-2013), INDIA.

Concept1	Concept2	Relation
Customer	Tourism	Pay
Tourism	Location	Manages
Tourism	Transport	Manages
Tourism	Accommodation	Manages
Transport	Location	Reaches
Location	Accommodation	Provides
Transport	Air	Offers
Transport	Water	Offers
Transport	Road	Offers
Air	Business	Enrolls
Air	Domestic	Enrolls
Business	Hill station	Travels
Business	Desert	Travels
Business	Temples	Travels
Domestic	Hill station	Travels
Domestic	Desert	Travels
Domestic	Temples	Travels

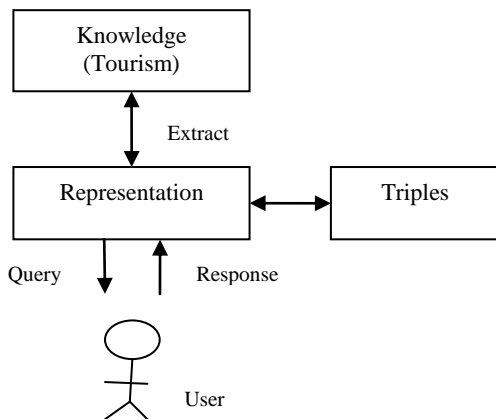**Table 1. Concepts and Relations**

Fig.2 Architecture of the proposed system

It has been implemented in Java (JDK 1.6). The user interface has been completely designed using Java Swings. This is to enable the users to use the system independent of the platform. The user can access the system in any operating environment. The modules in the system are implemented with the classes. The implementation of the system strictly follows the coding standard. It extracts the knowledge and provides the information to the user according to their wishes. It also extracts the information which does not reside in knowledge also. Figure 3 shows the relation retrieved for the identified concepts.

Figure 4 shows the assertion on new knowledge into the knowledge base.
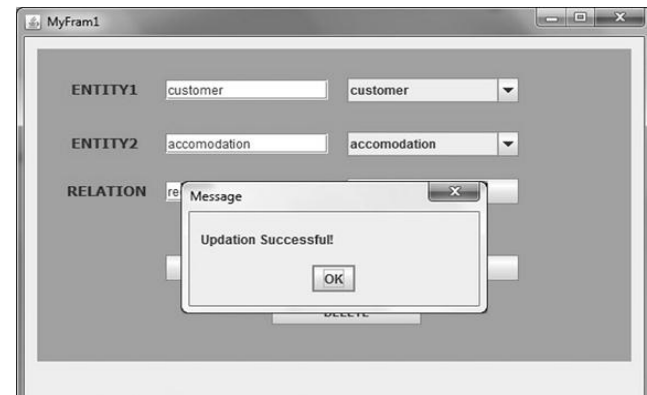

Fig 3. Relation Extraction

Fig 4. Relation Updation

V. CONCLUSION

This paper discussed about the representation of knowledge of Tourism domain. Relation is extracted for all possible combination of concepts that were identified for the tourism domain. New relations are identified and it has been asserted into the database. It has been implemented using Java. The results obtained on a real test triples show the effectiveness of our approach. In future, the proposed system can be modified so that the relation can be extracted from any web documents.

REFERENCES

[1] Christopher Khoo and Sung-Hyon Myaeng, "Identifying Semantic Relations in Text for Information Retrieval and Information Extraction", Chapter, Semantics of relationships: An Interdisciplinary Perspective, pp.161-180, 2002

[2] Yan Li and Jian Wen Zhoujun Li ," An improved Relation-Based Information Retrieval Technique for Bioinformatics", International Conference on Information and Automation,pp.1536-1541,2008

[3] Josh King and Venu Satuluri "Extracting semantic relationship". 2010.

[4] Shubin Zhao and Ralph Grishman," Extracting Relations with Integrated Information Using Kernel Methods", 43d ACL Meeting, pp.419-426, 2005.

[5] Punuru, J., Chen, J., "Extraction of Non-hierarchical Relation from Domain Texts", IEEE Symposium on Computational Intelligence and Data Mining, 2007

[6] Gaizauskas, R., &Wilks, Y. (1998). Information extraction beyond document retrieval. Journal of Documentation, 54(1), 70-105.

[7] Rau, L.F., Jacobs, P.S., &Zernik, U. (1989). Information extraction and text summarization using linguistic knowledge acquisition. Information Processing & Management, 25(4),419-428.

[8] Farradane, J.E.L. (1967). Concept organization for information retrieval. Information Storage and Retrieval, 3(4), 297-314.

[9] MassimilianoCiaramita, Aldo Gangemi, Esther Ratsch, JasmimSaric, Isabel Rojas. 2005. Unsupervised learning of semantic relations between concepts of a molecular biology ontology. 19[th]IJCAI, pp. 659-664

[10] Sridevi U.K. and Nagaveni N, "Ontology Based Semantic Measures in Document Similarity Ranking", IEEE International Conference on Advances in Recent Technologies in Communication and Computing, 2009.