International Journal of Engineering Research in Computer Science and Engineering (IJERCSE) Vol 2, Issue 10, October 2015

Integration of Hetergeneous Tourism Information Using Ontologies, Formal Concept Analysis and Bayesian Analysis

[1]Supriya Namdas, [2] Sapana Dolas [3] Shubham Sankale [4] Priya Surana
[2] Department of Computer Engineering ,
Pimpri Chinchwad College Of Engineering Nigdi pune 44,Pune University
[1] Supriya7723.namdas@gmail.com [2] sapanaldolas@gmail.com [3] shubhamxlr777@gmail.com [4] priyaasurana@yahoo.co.in

***Abstract*-- This paper presents an ontology-based approach, formal concept analysis (FCA) approach and a Bayesian Analysis to integrating heterogeneous tourism information for online tour planning. Ontology is playing a crucial role in knowledge management and semantic web. The tourism information ontology is becoming a core research field in the domain of information retrieval. The intent of this study is to scrutinize the potential role of formal concept analysis (FCA) for integration of heterogeneous tourism information. Two ontologies are developed, one for tourists and the other for tourism information providers. Both of these ontologies are mapped using FCA and Bayesian analysis to evaluate tourist preferences against the information published by the tourism information providers**

***Keywords*- formal concept analysis, ontology design, Bayesian analysis, analytical hierarchy process, tourism recommender system.**

I. INTRODUCTION

In recent years, the number of e-tourists has outstandingly increased in reaction to the expanding availability of on-line tourism information. Planning a tour on-line necessitates detailed data (information) regarding many aspects of tourist attractions, such as the activities provided, their open timings, fees for acceptance, and the course between them. Single tourism web site can hardly supply all required information, while different web sites may provide various and often contradictory information about the identical attraction. For a given section of information, such as admission fees, it may be put forward in various expressions, typically 'cost', 'admission charges', and 'entry charges', as well as the term 'admission price'. For a successful tour plan, the semantically heterogeneous information needs to be integrated and consistently represented, or else tour planning on-line is generally seen as a monotonous, unexciting and frustrating experience.

Recent progress in ontology research makes it feasible to integrate diverse information such as that accessible from on-line sources. Ontology consists of a set of concepts and the alliances between them. The definition of on-line information. Can be interpreted in a uniform way as stated by ontologism, and heterogeneous information can then be integrated.

There are two types of participants involved in tour planning, the tourists and the tourism information providers. Tourists are the people who travel to stay away from their usual habitat for some short period of time for leisure, adventure, for business purposes, to gather information, to study and for other purposes. This paper primarily focuses on those tourists who travel for leisure. The purpose of tourism information providers is to provide all the necessary information regarding the attractions of a place over the internet. The providers can be different government and non-profit organizations, intermediaries (e.g. travel agents), tourism service suppliers and tourists themselves. This paper concentrates mainly on the information provided by government and non-profit organizations as their information tends to be comprehensive, objective, and structured, thus ideal for ontology development. Although the information generated by tourists on social media has become significantly important for tour planning. it tends to be unreliable and unstructured. Tourism web sites, especially those which are provided by the government and non-profit organizations, still remain a better information system.

Tourists and the tourism information providers both have their own different approach in tour planning. Each perspective can be represented by ontology, and the two perspectives can be bridged through ontology mapping approaches for producing a tour plan that consists of both tourist's preference and the information published by the tourism information providers. For tourists, the tourism literature is mostly written from their point of view and presents the knowledge in formal and consistent forms. This offers an ideal basis to obtain ontology for tourists. For tourism information providers, there is little published literature that constitutes their perspective. On-line tourism information is generally available in heterogeneous forms.

Of the two ontologies, the development of the ontology for tourism information providers puts up a greater challenge to deal with.

II. TOURISM ONTOLOGY

Ontology consists of a set of concepts and the relationships between them. It is a formal representation of the common vocabulary of a knowledge area. These concepts and their relationships are under an agreement within an area Ontologism have been used in tourism research. One type of tourism information providers to uniformly represent their information. An example is the Thesaurus on Tourism and Leisure Activities defined by the World Tourism Organization (WTO). This standard can help unify tourism terminology to help effectively search for tourism information. However, there is a lack of central authority to enforce such a standard. Instead, various localized tourism ontologies are developed and these ontologies coexist over the Internet. These efforts primarily focus on the development of a single ontology for the tourism information providers. Presently, the challenge is to integrate these local ontologies. One proposed approach is to link a pair of local ontologies at a time, but this is a tedious and cumbersome task. The development of a central ontology can effectively integrate multiple local ontologies. This integrated ontology serves as a reference to understand the heterogeneous tourism information from various tourism information providers. Furthermore, an effective tour plan requires two ontologies to represent two different perspectives. Matching between the two ontologies is necessary to support the tour plan. Unfortunately, there has been less attention paid to mapping the ontologies of different perspectives, specifically in this case, between the ontology for tourists and the ontology for tourism information providers.

III. THE ONTOLOGY FOR TOURISTS

The ontology for tourists is concerned with a typical tourist's preferences. The concepts and their relationships for this ontology are derived from research in choice models, an important component of tourism literature. Choice models study factors that influence the selection of tourism services, such as what attractions a tourist may choose to visit [8]. For this research, a set of factors is extracted from the choice model literature using the 64 papers listed in Appendix A (in addition to the references). Information about these factors includes their name, definition, and properties. The extracted factors and 'sub-factors' in the original literature are used as concepts and sub-concepts, respectively, in the ontology for tourists. The concept and sub-concept relationships are specified based on the context of the literature. The properties associated with the factors specified in the original literature are treated as the properties of the concepts in the ontology. The derived ontology for tourists includes the following main concepts at the first level: preferred tour time, preferred tour budget, preferred tour transportation mode, and preferred activities of a tourist. Each concept may have multi-level sub-concepts and properties associated with both concepts and sub-concepts (Table 1). These concepts represent the factors that are consistently ranked the highest in terms of the frequency of their usage in the literature cited. This result is consistent with research findings that the selection of tourist attractions often depends on a small number of common factors, such as time, budget, and preferred activities. Personal profile information such as age, occupation, personality and interests is also a factor in choice models. This factor, however, is research is to develop a globally adoptable standard for mainly used in a situation where a tourist does not identify their preferred activities. In this research, personal profile information is not included in the ontology for tourists, as most tourists are able to identify their preferred activities when they plan a tour on-line.

IV. THE INTEGRATED ONTOLOGY FOR TOURISM INFORMATION PROVIDERS

The ontology for tourism information providers represents the perspective of various tourism information providers, and consists of descriptions of tourist attractions. Corresponding to the diverse and informal nature of on-line tourism infor- mation, a two-step approach is proposed for its development. The first step develops a large number of 'local ontologies,' one for each web site. In the second step, an integrated ontol- ogy is derived from all of these local ontologies to present a unified representation of on-line tourism information.

V. FORMAL CONCEPT ANALYSIS

Formal concept analysis (FCA) is a method of data analysis with growing popularity across various domains. FCA analyzes data which describe relationship between a Particular set of objects and a particular set of attributes. Such data commonly appear in many areas of human activities. FCA produces two kinds of output from the input data. The first is a concept lattice. A concept lattice is a collection of formal concepts in the data which are hierarchically ordered by a sub-concept- super-concept relation.

VI. BAYESIAN ANALYSIS

Bayesian analysis is a statistical paradigm that answers research questions about unknown parameters using probability statements. Bayesian inference uses the posterior

distribution to form various summaries for the model parameters, including point estimates such as posterior means, medians, percentiles, and interval estimates known as credible intervals. Moreover, all statistical tests about model parameters can be expressed as probability statements based on the estimated posterior distribution. Unique features of Bayesian analysis include an ability to incorporate prior information in the analysis, an intuitive interpretation of credible intervals as fixed ranges to which a parameter is known to belong with a pre-specified probability, and an ability to assign an actual probability to any hypothesis of inters.

VII. ARCHITECTURE DESIGN

A. *Ontology Development*

The ontology for tourism information providers represents the perspective of various tourism information providers, and consists of descriptions of tourist attractions. Corresponding to the diverse and informal nature of on-line tourism information, a two-step approach is proposed for its development. The first step develops a large number of 'local ontologies,' one for each web site. In the second step, an integrated ontology is derived from all of these local ontologies to present a unified representation of on-line tourism, information.

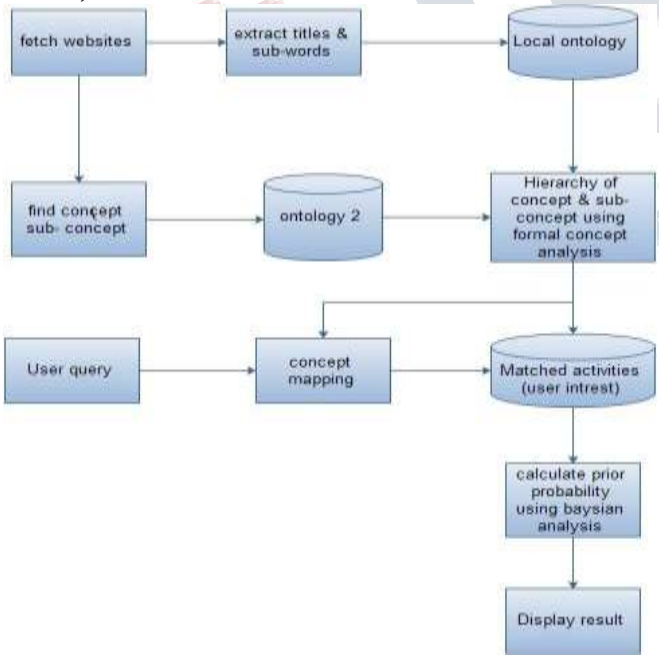


Fig: architecture Diagram

B. *Mapping between the two ontologies*

The ontology for tourists and the integrated ontology for tourism information providers are mapped in order to evaluate tourists' preferences against the information published by tourism information providers. The mapping between the two ontologies is performed at two levels. The first level maps the concepts between them, and the second level matches the properties for a given set of mapped concepts.

C. *Property Comparison*

For each set of mapped concepts, the second level of ontology mapping compares and matches their property values. The result of this process helps select appropriate attractions that satisfy the preferences of a tourist. Since out of the four categories of preferences (tour time, budget, transportation mode, and activities) preferred activities plays a primary role in the selection of attractions, property comparison between the pair of mapped concepts, 'preferred activities' and 'provided activities', is discussed first. Because the two concepts are both expressed in nominal values, a Bayesian analysis is used to compare their values. In a Bayesian analysis, a hypothesis of a concept, presented as the prior probability (P(H)), represents an initial belief. Additional evidence (P(ejH)) represents the likelihood of the hypothesis and is used to update the prior probability to the posterior probability (P(Hje)). The property comparison for other mapped concepts is through simple quantitative comparison.

D. *Shortest path algorithm*

In Dijkstras algorithm, maintain two sets, one set contains vertices included in shortest path tree, other set includes vertices not yet included in shortest path tree. At every step of the algorithm, we find a vertex which is in the other set (set of not yet included) and has minimum distance from source. Below are the detailed steps used in Dijkstras algorithm to find the shortest path from a single source vertex to all other vertices in the given graph.

- Create a set sptSet (shortest path tree set) that keeps track of vertices included in shortest path tree, i.e., whose minimum distance from source is calculated and finalized. Initially, this set is empty.
- Assign a distance value to all vertices in the input graph. Initialize all distance
- values as INFINITE. Assign distance value as 0 for the source vertex so that it is picked first.
- While sptSet doesnt include all vertices

(a) Pick a vertex u which is not there in sptSetand has minimum distance value.
(b) Include u to sptSet.
(c) Update distance value of all adjacent vertices of u. To update the distance values, iterate through all adjacent vertices. For every adjacent vertex v, if sum of distance value of u (from source) and weight of edge u-v, is less than the distance value of v, then update the distance value of v.

VIII. MATHEMATICAL MODEL

- Let, system S={I,OC,MC,B,BA,O,D}
  OC is ontology creation function
  OC= {LO,FCA,IO}
  LO is local ontology consist of user's attraction
  LO={ L1,L2,L3…… }
- FCA = {T,OF} is Formal Concept Analysis method applied on local ontology for creating integrated ontology
- T={G,M,I}

- T is triplet where
- G and M are two sets of elements called objects and attributes respectively, and I is a binary relationship between them.
- OF={r, n , c}
- OF is output from FCA which use for creating hierarchy concepts
- r is reference concepts
- n is new concepts
- c is local concepts
- I is input, I ={U,Q}
- U is users of Systems = {U1,U2,U3….)
- Q is queries of users = {q1,q2,q3…..} q1 contains four values ={T,B,A,M}
- T is preferred tour time
- B is preferred budget
- A is preferred activities
- M is preferred mode

- MC= {A,OF,OM}
- MC is concept mapping in which users preferred activities is map with provided activities.
- A is preferred activities
- OF is sub concepts hierarchy
- OM is user attraction places which map by preferred activity.
- BA is Bayesian analysis is used to compare preferred activities and provided activities values. In a Bayesian analysis, a hypothesis of a concept, presented as the prior probability (P(H)), represents an initial belief. Additional evidence (P(e | H)) represents the likelihood of the hypothesis and is used to update the prior probability to the posterior probability (P(H)) , as in
- Where *P*(*e*) is a normalization constant
- P={p1,p2,p3……}
- O is Output extracted from provided activities which have highest rank with tour planning. O={pa, pb , pc…..}
- Pa= {d1,d2,d3….} d=day
- Pb= {a1,a2,a3…..} a= city
- D={G,SP,So,En} G={V,E,ds}
- V={v1, v2, v3,…..} Where v1,v2,v3 are vertices
- E={e1,e2,e3….} Where e1,e2.. are edges
- Ds distance between vertices
- So is source point of tourism
- En is end point of destination
- SP is shortest path

IX. EXAMPLE

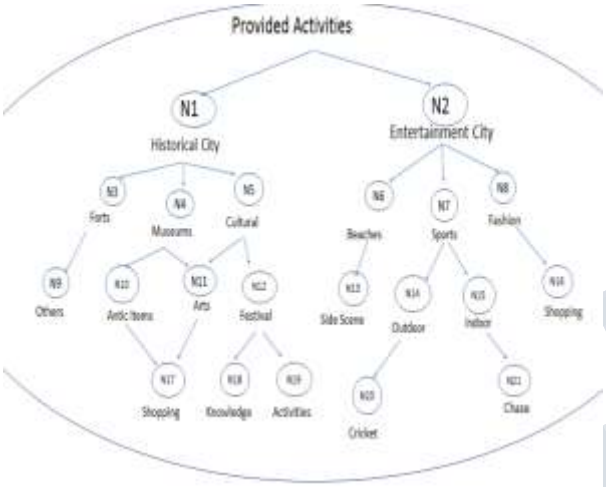
To illustrate the proposed ontology-based approach, an example for planning a tour in pune City is presented. The preferences of a hypothetical tourist are evaluated against the on-line information about attractions in pune City published by various tourism information providers. Information about attractions used in the illustration is from actual tourism web sites. The information about tourist behaviour is based on a survey that conducted tourist preference and choice of attraction in pune.It is supposed that the tourist, jully, plans to visit pune City on Friday, September 20, 2015. She has booked a hotel and plans to stay for two days and spend less than $100 on admission fees at attractions. jully's specific tour preferences are listed below.

- Preferred tour time
- Tour date: Friday, September 20, 2015
- Preferred length of tour: 2 days (approximately
- 8 hours a day
- Preferred budget: $100

Preferred transportation mode: walk, bus, or subway
Preferred activities: learning new things and enjoying nature Based on the proposed property comparison approach, jully's tour preferences are evaluated in three steps. The first step matches provided activities with jully's preferred activities using the Bayesian analysis. In the second step, the attractions in New York City that correspond to jully's preferred provided activities are retrieved. The third step ranks these retrieved attractions based on jully's preferences. In the first step, jully's two preferred activities, 'Learning new things' and 'Enjoying nature' are assumed equally important. Each is then given a probability of and the remaining two of the four preferred activities are given a probability of zero: 'Learning new things' 'Enjoying nature', 'Relaxing', and 'Venturing'. The probabilities of the four activities are used in the subsequent Bayesian process to estimate ram's choice of provided activities. For illustration purposes, of the provided activities in the six-level hierarchy only the activities at the top two levels are discussed here. Information about other tourists' behaviour is used to estimate the prior probability and the likelihood in the Bayesian analysis. Both the preferences and choices of

tourists are involved in the estimation. At the first level of provided activities hierarchy, there are two choices: 'Entertainment & outdoor' and 'Sightseeing'. The probabilities of choosing these two activities by other tourists, according to the survey output by Using Dijkstra's shortest path algorithm, the 2-day tour is arranged as following: '

Day 1: location 1, location 2, location 3…..
Day 2: location 1, location 2 …
Fig: hierarchical structure.



X. LITERATURE REVIEW

A Bayesian Network and Analytic Hierarchy Process Based Personalized Recommendations for Tourist Attractions over the Internet.	Y. Huang and L. Bian,	2009	proposed for recommendation of tourist attractions at a given destination and recommendation based on Location.	Not work on problem of *how to plan these attractions as a trip.*
Ontological Recommendation Multi-Agent for Tainan City Travel	S. Lee, Y.-C. Chang and M.-H. Wang,	2009	Proposed to recommend a trip plan for the traveller.	Not find optimized solution.
The Traveling Salesman Problem: A Computational Study	D. L. Applegate, R. E. Bixby, V. Chvatal and W. J. Cook,	2006	TSP finds the shortest path from the exact *n* given locations. In other words, all locations have to be involved.	Not optimal trip finding problem is to finds a sub-permutation of given *n* locations
Trip-Mine: An Efficient Trip Planning Approach with Travel Time Constraints	Eric Hsueh-Chan Lu and Chih-Yuan Lin Vincent S. Tseng	2012	Propose a novel data mining-based approach, namely *Trip-Mine*, to efficiently find the optimal trip which satisfies the user's travel time constraint based on the user's location. Achieves significantly high efficiency in optimal trip finding.	Not develop data mining-based approaches to automatically find the interesting attractions and estimate the stay time of attractions.
Personalized Location-Based Recommendation Services for Tour Planning in Mobile Tourism Applications,	-C. Yu and H.-P. Chang,	2011	proposed for trip recommending, including travel agent	The problem of automatically finding semantically annotated sequences
A Novel Approach to Mining Travel Sequences Using Collections of Geo-tagged Photos	S.Kisilevich, D. Keim and L. Rokach,	2010	Address the problem of automatically finding semantically annotated sequences based on geo-tagged photos.	It did not consider the travel constraints and user preferences.

XI. CONCLUSIONS

Ontology for tourists and in particular an integrated ontology for tourism information providers are developed, and the two ontologies are mapped to match the perspectives

between tourism service users and providers. In the context of ontology research, the approaches developed. The semantic web, the next generation of web, has been considered the potential direction for on-line tourism information systems, including tour planning. The integrated ontology for tourism information providers can help develop semantic web-based tourism information systems. This system focuses on tourism attractions it can be extended to other services that are also commonly involved in a tour planning, such as selecting accommodation and dining services. Similar to suggesting tourism attractions, a pair of ontologies, one for the service users and the other for the service providers are needed and then mapped between them to select appropriate hotels or restaurants.

REFERENCES

1. Yuxia Huang, Ling Bian "Using Ontologies and Formal Concept Analysis to Integrate Heterogeneous Tourism Information" School of Engineering and Computing Sciences, Texas AM UniversityCorpus Christi, Corpus Christi, TX 78412 USA,2015.

2. Y. Huang and L. Bian, A Bayesian network and analytic hierarchy process based personalized recommendations for tourist attractions over the Internet, Expert Syst. Appl., vol. 36, no. 1, pp. 933943, 2009

3. Y. Huang, toward semantic interoperability in travel planning on the semantic Web: Learning a reference ontology from online heterogeneous tourist attractions classification systems, Inf. Technol. Tourism, vol. 11, no. 1, pp. 5166, 2009.

4. J. Poelmans, D. I. Ignatov, S. O. Kuznetsov, and G. Dedene, Formal concept analysis in knowledge processing: A survey on applications, Expert Syst. Appl., vol. 40, no. 16, pp. 65386560, 2013.

5. L. Bian and S. Hu, Identifying components for interoperable process models using concept lattice and semantic reference system, Int. J. Geograph. Inf. Syst., vol. 21, no. 9, pp. 10091032, 2007.

6. U. Priss, Formal concept analysis in information science, Annu. Rev. Inf. Sci. Technol., vol. 40, no. 1, pp. 521543, 2006.

7. Z. Xiang and U. Gretzel, Role of social media in online travel information search, Tourism Manage., vol. 31, no. 2, pp. 179188, 2010.

8. C.-S. Lee, Y.-C. Chang and M.-H. Wang, Ontological Recommendation Multi- Agent for Tainan City Travel, Expert Systems with Applications, vol. 36, no. 3, pp. 6740-6753, Apr. 2009.

9. Eric Hsuch-ChanLu andCluh-Yuan Lin 'Trip-Mine:An Efficient trip planning approach with travel time constraint'

10. C.-C. Yu and H.-P. Chang, Personalized Location-Based Recommendation Services for Tour Planning in Mobile Tourism Applications, E-Commerce and Web Technologies, vol. 5692, pp. 3849, 2009.

11. D. L. Applegate, R. E. Bixby, V. Chvatal and W. J. Cook, The Traveling Salesman Problem: A Computational Study, Princeton University Press, 2006.

12. S. Kisilevich, D. Keim and L. Rokach, A Novel Approach to Mining Travel Sequences Using Collections of Geotagged Photos, in Proceedings of the 13th AGILE International Conference on Geographic Information Science, pp. 163-182, May 2010.