

Web Mashup Oriented Digital Information Extraction for Tourism Enhancement

Satyen Amonkar, Nandan Doddihai, Tanvi Gangan, Chintan Pasad, Mrs. Gresha Bhatia

Vivekanand Education Society's Institute of Technology, Mumbai, India

Abstract— Travelling and exploring new locations have become a very common and most attractive hobby for a lot of people. Ranging from a long holiday to some faraway places to a weekend getaway to a comparatively nearby unique spot, every type of touristic activity is highly cherished among the avid explorers. This leads to a demand for a resource that would cater the need of this rapidly growing community. Certain previous research suggests important further development to achieve this goal. It becomes necessary to integrate the relevant researches done previously so as to create a product with better usage. This paper discusses an approach of using Web Mashup technology to extract attributes and spatial data from multiple websites and use it for data analysis thereby enhancing the touristic experience. The main objective of this paper is to provide users with filtered information which is convenient to the users for selecting a tourist spot to visit. It also aims to improve degree of satisfaction of tourists and optimize touristic routes.

Index terms—web mashup, tourism, big data, ontology, web crawling, recommender system.

I. INTRODUCTION

The tourism industry has been flourishing worldwide. Many of the countries see the tourism sector as one of the most important source of revenue. Several developing economies are built on the further development in the tourism sector.[1] A country with natural beauty having an untarnished view of nature is bound to attract visitors. Under this circumstances,

having a good tourism based application proves to be a boon. Concepts of web crawling and web mashup[2] can be integrated well for enhancing the sector of tourism. A list of URL of tourism based websites can be compiled and crawled to extract information about a particular place from all aspects. Such an application would be useful to combine several useful attributes from several websites and then using appropriate mechanism, it would relate all of them and provide a relevant search result.[3]

A mashup is a web application that uses content and knowledge from multiple online sources to create a new software product that delivers a better result and output for the user. For example: a person can combine google maps with his collection of images and pictures to create a mashup that finds utility in some other application.[4] It implies usage of fast integration with application programming interfaces (API) leading to optimized results that were originally not found in the application. Client application and online search are mainly based on web mashup mechanisms.

Mashups are of different types which includes: Business mashup, Data mashup and consumer mashup. In content mashup, data is taken from different sources having different format and are organized in a single representation on the graphical user interface. In the data mashup, similar data are extracted from multiple sites to provide an elaborate result.[5] Another concept used in this paper is web crawler. It browses the World Wide Web for the purpose of web indexing which is also called web spider. Web crawling is also capable of copying data from the sites for further processing. An automatic

update of the content on the web page uses web crawling mechanism.[6]

This paper also provides a design to provide an option to the user for pin-pointing the relevant search result locations on the google map in order to plan their trip accordingly. The design thus provides answers to frequently asked questions by the users like:

- Is the attraction in close proximity to my current location?
- Are there any events going around that match my preferences?
- Is the location feasible and likable considering all the previous reviews and current preferences?

II. EXISTING SYSTEMS AND LACUNAE

Literature survey covered various fields mainly tourism, web mining, usage of web mining in tourism field. It led to better understanding on the state-of-the-art of the area and its fundamentals. Latest theories, methods and approaches were accessed leading to our research topic based on previous researches. One of the paper helped with in-depth knowledge in web mining and recommender system. Web data mining is used to extract useful pattern in the user search engine in order to understand the user preferences thus extracting user interests and potentially useful data. It discovers required information from different types of data such as images, audio, videos and other formats. It analyses server log, cookie log in order to analyze user's browsing pattern.[7]

Another useful information required for this project is the recommender system. Recommender systems help users find their information quickly by learning the users browsing pattern. Correct prediction of the user's preference plays a vital role here. Such systems help prevent web spamming by controlling the data to be displayed only up to the extent of user preferences. There are three types of recommender systems based on the requirement of an application: Non-personalized, User based, Product based. Most

prominent recommender systems are Collaborative Filtering, Associative Rule.[7]

The concept of Ontology helps in understanding a way to handle voluminous amount of information available on the internet. Manual integration of information available on the internet is difficult and so concepts of Ontology comes to use.[8]

Another important feature commonly present is mobility. Mobility is the basis of any touristic activity. Thus any development in tourism model must support this activity. A mobile application must maintain the capability of

- Knowing the state of activity
- Learning the state of time
- Knowing the state of place

Such capabilities are only held by newer smartphones these days as a result of the unavailability of some sensors on the previous generation phones. In order for the mobile system to be compatible with a touristic application it must have minimum tools such as an internet connection in order to sync the time, have GPS installed on the phone to track location, consist of in build sensors like gyroscope, accelerometer and compass.[9]

Literature survey also included newspaper and magazine articles providing information on web mining and its real-time application across the globe. Few of the previous work included mining one type of data and performing analysis over it [10] [11] [12] [13]. However the output of such an experiment led to a conditional usage due to restricted data. For example, work was performed only on tweets from Twitter or images and tags on Flickr. These applications used closed data set causing it to be of circumstantial usage.

There has been a constant need of an appropriate level of management and service so as to improve touristic experience. Any compromise in this leads to a negative impact on the image of tourism in that country. There should be a constant flow of competitive tourism product.[14] Very few of the previous researches have proven to be attractive due to lack of in-depth knowledge or analysis. Several technological integration has helped in streamlining and improvement resulting into rapid growth of cultural tourism sector.[15]

III. PROPOSED METHODOLOGY

Looking at the limitations we identify the need for providing quick, faster mechanism towards an efficient tourism industry through the use of web crawler, web mashup and recommender system. Thus, we present an approach for utilizing web mashup technology to extract different types of data from multiple tourism websites and filter this data as per the preferences of the user, and also analyze the extracted data to recommend the user ideal places to visit and favorable touristic routes as per his/her preferences; thereby providing a better touristic experience to the user.

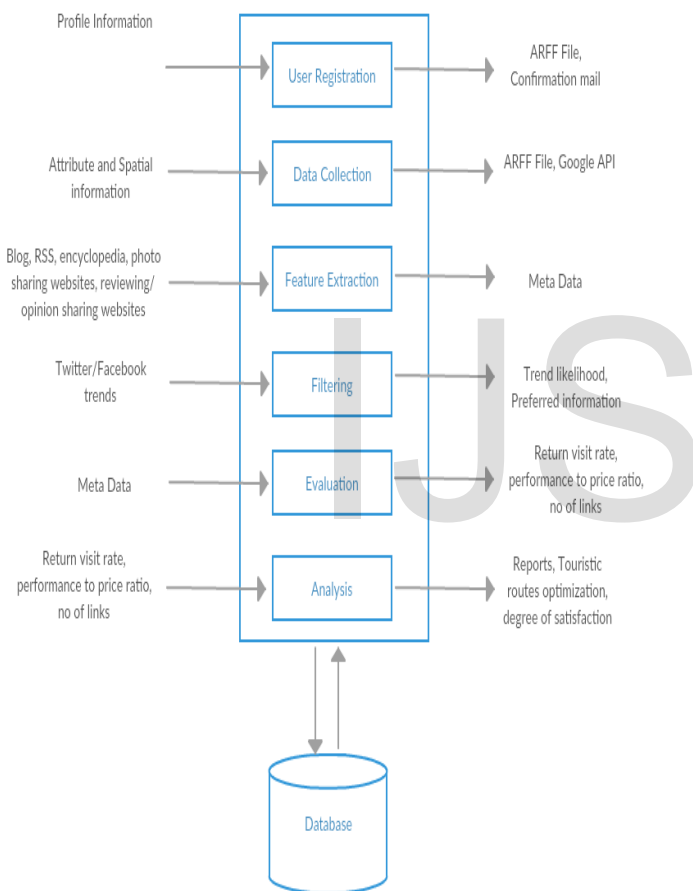


Fig.1 Block Diagram of the Proposed Design

For any user to use this service, he/she needs to register and create an account. The user would provide basic personal information such as name, username, password, email, contact number, gender, age, city, etc. The user may optionally also provide

personal preferences which would enable him/her get personalized recommendations about places to visit and others. These preferences include places of interest, cultures of interest, starting place, destination, tour theme, food, hotels, travel modes, visiting time, travel partners, payment mode, etc.[16] All this information provided by the user gets stored in the database as attributes and can be retrieved in the form of ARFF (Attribute-Relation File Format) files. The user may also book a trip for which he/she provides details like location of visit, time, duration, number of people budget, etc. On successful registration the user receives a confirmation mail.

The data is collected from various touristic websites using the tool like Kimono Labs. The data extracted includes list of categorized destinations like world, India, heritage, hill station, beach, riverside, forest, etc. Data specific to each destination such as activities that could be done when visiting there, entertainment options, shopping, night life, music concerts and/or sporting events, etc.[17] [18] is also collected. Other important data extracted includes all kinds of images and videos displayed on these websites. Data regarding user reviews is also collected like best routes, ratings, testimonials, quality of food, entertainment, etc.[19] Data about extra features provided by websites is also taken, like offers and deals, buying/selling of foreign exchange, insurance, etc.[20]

Feature extraction is done from various blogs, RSS feeds, encyclopedias, photo sharing websites, review/opinion sharing websites, etc. and stored in the database. These can be used in the form of metadata. Another important part of feature extraction is geo coding. Geographical information and co-ordinates for different places are collected and stored in the form of spatial information. This data is then utilized with Google Maps API in order to pin the recommended locations on the map. Attribute extraction is also done which includes acquisition of basic queries, selection of related words, query generation and retrieval.[21]

Data is also extracted from popular social networking sites like Facebook and Twitter about the

trending topics. Filtering is then done to get the trend likelihood and preferred information.

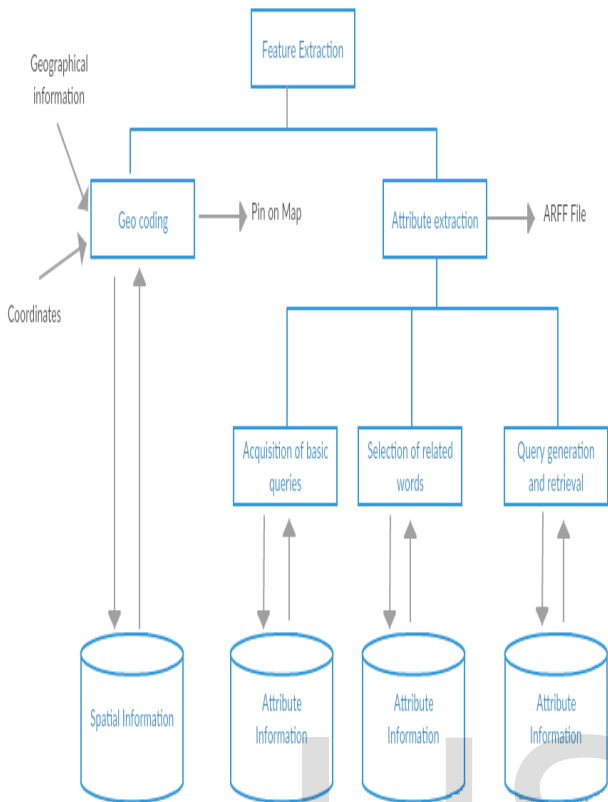


Fig.2 Modular Diagram for Feature Extraction

Evaluation is then done on all the collected attribute and spatial information along with metadata to calculate performance ratio for destinations, return visit rate, number of links, etc.

Finally the evaluated information is analysed. Analysis broadly includes two parts, link analysis and sentiment analysis. The link analysis part ascertains the range of touristic destination, confirms touristic attractions in the range, calculates travel attractiveness of each touristic attraction and confirms the travel attractiveness. The sentiment analysis part uses the opinions and feedback of users as well as their degree of satisfaction to determine the accuracy, recall and precision of the entire web mashup. A user-friendly interface is provided to the end user which is simple, easy to navigate and user interactive. The user is provided with the recommended destinations list with several images for each location. Reviews, ratings, quality of services, etc. are all provided to the user. Images and videos regarding touristic attractions of that location are displayed to the user by pinning them on the Google Map for that

location. All the latest events like music concerts, shopping festivals, cultural festivals, sporting events, etc. would be displayed to the user for that location. The user can at any time add or modify his/her preferences and the updated information would be used for recommendations.

IV. DATA FLOW DIAGRAMS

The data flow diagram is a diagrammatic representation of the flow of control of the system. It helps to visualize the system in a better way. It leads to a modular decomposition of the system that helps to check for system error module wise. How the system reacts for a set of input? What data does the module consist of? Where does the flow of control move for a particular set of input? These questions are easily answered with the help of data flow diagrams of the system. Data flow diagrams are prepared at the initial stage of the project to create an overview of the system. It supports appropriate processing of data to ensure the collected data is suitable to fulfil the system goals. Data flow diagrams are of multiple levels. The Level 0 is the most abstract level. It consists of only a single process. Breaking this process into different tasks is explained in further levels.

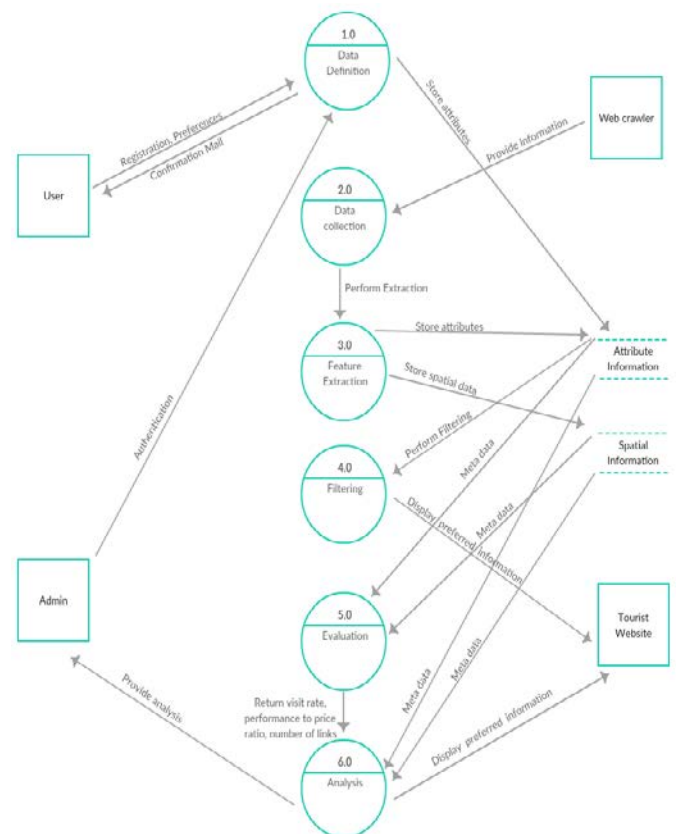


Fig.3 Level 1 Data Flow Diagram of the System Design

V. CONCLUSION

This paper is an enhancement over the currently persisting systems working in the tourism sector and which are based on web mining and mashups. It provides a comprehensive study of a proposed model which is a mashup itself of all the specific and important concepts already in use in tourism field. It picks up few lacunae by performing literature survey of previous researches in relevant field of study. Based on these lacunae and then considering all necessary attributes, a proposed model was developed. This model aims at fulfilling the user requirement at an individual level. Every person is different and so is his likings and preferences. Considering the requirement of such a dynamic set of users, this models plans to serve the purpose of an enhanced service. This service would initially get the personal details from the user providing an input to the data collection module. This set of extensive data is stored with our local database at the backend. Parallel implementation is performed about the feature extraction mechanism. Here data about multiple location is obtained. This data has several important attributes which helps mainly for segregating the data and providing an output to the recommender system. This recommender system must ensure that the data is uses from feature extraction module but be relevant to the needs of the user. Otherwise it may lead to web spamming. Thus, a quick and handy tool would provide a solution to the problem of guiding a user who is a new to a city. This tool not only provides an information about the touristic spots based on user's interest but also mentions a piece of considerable future work by identifying the integration of the web mining results with Google Maps so as to mark the geographical location of the touristic spot.

The system thus supports not only travelers and explorers but also assist in providing information to a naïve user visiting a particular location for once.

VI. REFERENCES

- [1] www.tpdco.org/dynaweb.dti?dynasection=tourismenhancement...main
- [2] [https://en.wikipedia.org/wiki/Mashup_\(web_application_hybrid\)](https://en.wikipedia.org/wiki/Mashup_(web_application_hybrid))
- [3] <https://www.webcrawler.com/>
- [4] Edoardo Ardizzone et al, "Extracting Touristic Information from Online Image Collections", 2012, 8th International Conference on Signal Image Technology and Internet Based Systems, Palermo, Italy.
- [5] Gabor Imre, Gergely Mezei, "Beyond Mashups – Graph Transformations on Web Data", 2014, IEEE 18th International Conference on Intelligent Engineering Systems, Budapest, Hungary.
- [6] Jen-Hsiang Chen, "Hybrid Recommendation System for Tourism", 2013, IEEE 10th International Conference on e-Business Engineering.
- [7] Kaifan Ji, Xuesong Zhao, "Tourism E-Commerce Recommender System Based on Web Data Mining"
- [8] S.Yasodha, S.S. Dhenakaran, "An Ontology-Based Framework for Semantic Web Content Mining", 2014, International Conference on Computer Communication and Informatics, Coimbatore, India
- [9] Ridi Ferdiana, Bimo Sunarfri Hantono, "Mobile Tourism Services Model: A Contextual Tourism Experience Using Mobile Services", 2014, 6th International Conference on Information Technology and Electrical Engineering (ICITEE), Yogyakarta, Indonesia
- [10] Kazutaka Shimada et al, "Analyzing Tourism Information on Twitter for a Local City", 2011, First ACIS International Symposium on Software and Network Engineering, Fukuoka, Japan.
- [11] Lili Song et al, "Evaluation of the Attractiveness of Tourism Destinations Based on Link Analysis", 2010, 18th International Conference on Geoinformatics, Beijing, China.
- [12] Shanshan Gao et al, "The Application and Comparison of Web Services for Sentiment Analysis in Tourism", 2015, 12th International Conference on Service Systems and Service Management, Guangzhou.
- [13] Kazutaka Shimada et al, "On-site likelihood identification of tweets for tourism information analysis", 2012, IIAI International Conference on Advanced Applied Informatics.
- [14] Jiang Yan-xia, MA Yi-liang, "Comparative Analysis of Tourism Satisfaction on Five Cities in Northeast China", 2012, International Conference on Management Science & Engineering, Dallas, USA

- [15] LIU Xiuhua, "Design and Implementation of Folk Tourism E-Commerce System Based on Web2.0 and Mashup", 2011, International Conference on E - Business and E -Government, Shanghai, China.
- [16] Xuwei Zheng et al, "Tourism Service Composition Based on Multi-objective Optimization", 2015, IEEE Twelfth International Symposium on Autonomous Decentralized Systems, China.
- [17] Xiangxiang Xie, Baohong Li, "Research on the Pattern of Tourism E-commerce", 2012, IEEE 26th International Parallel and Distributed Processing Symposium Workshops & PhD Forum.
- [18] Chin-Chih Chang, Kuo-Hua Chu, "A Recommender System Combining Social Networks for Tourist Attractions", 2013, Fifth International Conference on Computational Intelligence, Communication Systems and Networks.
- [19] Marta Rey-Lopez et al, "moreTourism : Mobile Recommendations for Tourism", 2011, IEEE International Conference on Consumer Electronics, Spain.
- [20] Haiying Zhao et al, "A Mobile Service Platform for Xinjiang Tourism", 2015, Seventh International Conference on Measuring Technology and Mechatronics Automation, China
- [21] Atae Rezaei Aghdam et al, "Identifying Places of Interest for Tourists using Knowledge Discovery Techniques ", 2014, University Teknologi Malaysia Johor Bahru, Malaysia