

An Innovative Tour Recommendation System for Tourists in Japan

Quang Thai LE*, Davar PISHVA**

*Faculty of International Management, Ritsumeikan Asia Pacific University, Beppu, Japan

** Faculty of Asia Pacific Studies, Ritsumeikan Asia Pacific University, Beppu, Japan

quanle11@apu.ac.jp, dpishva@apu.ac.jp

Abstract— The paper demonstrates prototype of system that is capable of suggesting optimal touring plan which is composed of various points of interest (POI) and take travelers' preferences and context into account. By focusing on Tokyo metropolitan area as an example, it systematically collects and analyzes information on thousands of tourists attraction areas and geographical nodes of Japan Railway (JR) train stations together concurrent weather information, estimated travel time, associated expenses, and lists of multiple cultural events in order to demonstrate practicality as well as reliability of the system. A programmatic approach based on the heuristic greedy search is employed for transforming the obtained data into informative routes.

Keywords— *e-tourism, travel planning system, web scraping, modeling, and data mining.*

I. INTRODUCTION

A. Economic & Tourism Aspects of the Tokyo Olympic 2020

Olympic Games usually provide economic stimulus and since Japan won the bid to host the 2020 Summer Olympic Games in September 2014, various think-tank groups have announced their estimates for its economic impacts. The Olympic 2020 will therefore be imposing both challenges and opportunities to the current stream of globalization in the country [1]. According to Japan National Tourist Organization (JNTO), 12 million tourists visited Japan in the year 2014, but the number is expected to increase to 20 million by 2020 and as many as 8.5 million foreign tourists are project to visit the country just during the Olympic [2]. The next 5 years will therefore be a good opportunity for Japan to leverage its "world class" brand in tourism and hopefully become a tourism-oriented country in the near future [3]. Similarly, in April 2014, Japanese Ministry of Education decided to provide 37 selected universities in Japan with an annual subsidy of about \$3.6 million each for the next 10 years so as to support them to become the so-called top global universities [4]. This is just one of the many efforts used for attracting international professors, researchers, students and cooperating with other prestigious institutions around the world, welcoming about 300,000 talented foreign students to visit and study in Japan until 2020 [5]. Despite its academic nature, possible tourism dimension of such huge number of foreign consumers is undeniably an open issue for the tourism industry.

B. Characteristics of Tourism Activities in Japan:

Japan tourism has seasonal flavor and different marketing strategies are used during each of the four seasons. For example, Japanese Railway (JR) Company issues promotional free train tickets during a certain period in spring, summer or winter (e.g. Seishun 18 ticket) [6]. The country is also famous for its numerous annual cultural events. No wonder Japan was chosen as one of the World's top tourist destinations in 2011 by the CNN [7]. Presently, Japan is classified as the world's top 9-tourist destination. Its transportation infrastructure consists of modern buses and highly efficient train system equipped with rapid transit railway network that link thousands of stations located throughout the country [6]. Tokyo is by far the most popular destination in Japan, accounting for 57.4% of foreign visitors [8].

Even though touring Japan seems to connote economic affluence, recent depreciation of the Japanese Yen has made it possible for people from emerging countries to come and experience the Japanese life [9]. In order to cope with the increasing flow of foreign tourists, as well as with globalization project of the government, not only infrastructure metrics such as international airport capacity, wireless network or, foreign language signs availability, etc., but also intangible services such as multilingual promotion websites, tour-guide profession, etc. have to be further enhanced [9]. The task of selecting a tourism destination in Japan is very important [10], and it requires "comprehensive information" so as to be "comfortable" [11]. This is because; there is still a lack of communication means in order to effectively bridge the tourists with the domestic tourism assets. Hence, the exploration of a collective delivery of tourism related knowledge to the right target audience under a particular context is an important practical research topic.

C. A Brief Literature Review and Research Objectives

Previous approaches in implementing digital recommender system (RS) for the tourism industry mainly focused on providing more specialized products to customers. Gulcin and Buse (2011) introduced a generic intelligent system applying similarity algorithm to find and match customers' preferences to previous cases that shared common attributes the most [12].

The paper also summarized a list of factors that are influential to decision-making (e.g., travel budget, local knowledge, hobbies, sex, age, etc.), of which “origins of customers” plays an important role [12]. In another research, Chiang and Huang (2015) made use of the user interface to continuously receive feedback from the users, as well as facilitate them to adjust any unsatisfied recommended results [13]. In 2014, Damianos et al. did an intensive review on this topic, discovering that more and more researches take into account tourists’ contextual information in the recommendation algorithm such as “attracts already visited”, “user mobility pattern”, “transportation mode”, etc. [14]. The review also suggested consideration of weather conditions, more practical route and flexible usage of transportation as prospect future research directions [14].

This research proposes an empirical and practical approach to the described context above so as to support individual tourists, tourism policy makers, as well as tourism managers and provide more fruitful experiences to foreign tourists.

II. METHODOLOGY

A. Data Collection

1) *Data Source and Description*: Collection of appropriate data is crucial for any successful projects, and this is especially true in the field of tourism where relevant data is scattered across wide networks rather than being aggregated at a fixed location. Hence, this research employs an automated web-scraping technique that can crawl to all the necessary data sites which contains information; e.g., all the stations that have JR trains (4517 stations in total), 1021 tourism attractions and hundreds of cultural events located within Tokyo prefecture, etc. The result shown in this paper is based on the information that were obtained from three relevant websites; namely, Navitime (<http://www.navitime.co.jp>), TripAdvisor (<http://tripadvisor.com>), and Jalan (<http://www.jalan.net>).

The Navitime website contains a list of relevant information on all of the JR train stations in Japan (e.g., their names, addresses, geographical coordinates), which have already been categorized into 47 different prefectures and operation companies. The website has also an extensive list of tourism guidance offices and regional specialty shops located throughout Japan. The TripAdvisor is a worldwide famous entity for its rich database on tourism products. Its portal provides an extensive list of the most famous attractions in Japan, which are ranked by actual travelers, and are neatly classified into various categories such as “sight & landmarks”, “natural & parks”, or “museums”. The Jalan.net is a domestic tourism portal that provides information on regularly updated cultural events and tourism attractions related attributes such as average visiting time, restaurants’ price and rating, best time to visit, etc., that have been collected from travelers. Concurrent and projected weather condition is also retrieved through Yahoo Japan’s weather forecast API.

2) *Data Retrieving Using Web Scraping Technique*: In order to retrieve the above mentioned data, a program is

written in Ruby language which utilizes “Mechanize” and “Nokogiri” library to retrieve all of the mentioned information in a systematic and automatic manner. Moreover, “Parallel” library is also used to shorten the execution time by processing multiple URLs simultaneously. All of the retrieved data is subsequently stored in a local MySQL database, which acts as an intermediary data library in the recommendation system.

B. User’s Preferences and Grouping POIs

1) *Travellers’ Scenarios*: The ultimate goal of the research is to recommend the most enjoyable places and subject it to constraints specified by the users. In order to be as practical as possible, apart from elements such as “attraction popularity”, “origin location” (e.g., current accommodation hotel), “available time to visit”, “budget”, “weather condition”, “walking preferences”, etc., the paper suggests “selective category option” to picture two different empirical scenarios that can benefit the system’s users:

a) *The traveller who wants to visit a variety of places that fit into certain categories (e.g. Museums, natural parks, art galleries, cultural events, etc.).*

b) *The traveller who wants to visit a variety of places without any specific interest categories (e.g. recommended popular places).*

2) *Heuristic POIs Grouping Based on Nearest JR Train Station*: Walking preference is one of the interesting factors that the paper takes into consideration. Different people have different walking preferences when travelling which can be traced to differences in demographics and cultural preferences [15]. Since we want to take advantage of one-day free pass ticket to all of the JR stations in Japan to reduce huge amount of transportation cost, the most likely problem that have to be addressed is to search for POIs that are located within a preferred walking distance from a nearby JR station. To achieve this, we categorize users’ walking distance preferences into 3 levels of: “Not Really” (less than 500 meters), “Fairly” (less than 800 meters) and “Definitely” (up to 1 kilometer).

By doing this, we can match all of the POIs in the vicinity of different stations that are reachable within the relative distances. Instead of using available public API for finding the nearest station to/from a place which would require long processing time to go through thousands of places, this paper suggests the Euclidean measurement to heuristically facilitate the categorization. As shown in Figure 1, on a 2-dimensional geographical map, from each of the JR station, a circle, having the relative walking distance as its radius, can be drawn to indicate which POIs can be reached from the station. As a result, one can easily compare relative distances when travelling from places to places quite efficiently and obtain lists of nearby stations for each of the POIs for later access (e.g., Table 1). As can be clearly seen from the Table 1, all of the POIs collected within the Tokyo area are accessible within

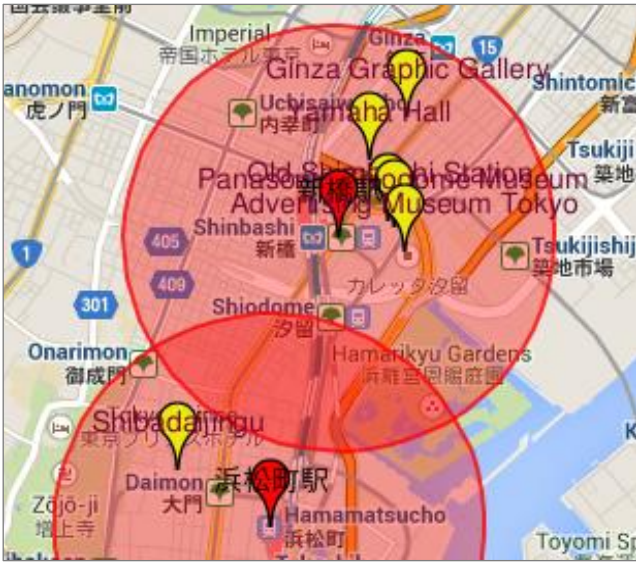


Fig. 1. Red circles with 800-meter radius are drawn to estimate the Euclidean distance from each of station (red markers) to nearby POIs (yellow markers).

1Km from the nearest JR station. Even though there are trade offs between walking preferences and POIs coverage-percentage, the number of uncovered POIs in the top 50 rankings looks quite reasonable.

TABLE I. CLUSTERING POIS BY NEARBY JR STATION

Distance (meter)	Average Walking Duration ^a	POIs Coverage Percentage	No of Uncovered POIs in Top 50
300	3.6 minutes	19%	26
500	6 minutes	33%	20
800	9.6 minutes	53%	14
1000	12 minutes	100%	0

^a Calculated based on average walking speed of 5kph

C. Tours Recommendation Using Greedy Algorithm

The process of this tour recommendation can be divided into the following three stages:

1) Category and Weather Based POIs Filtering

From a pool of all of available attractions, we filter based on the user's selected categories and the current weather information (if requested) to get a list of all of the POIs that meet the user's interest preference. Depending on the weather conditions (e.g., raining, snowing), outdoor POIs such as natural parks or bridges does not get included in the result list. When the category option is not selected, the algorithm outputs a list of the most popular places in the examining area.

2) POIs and Station Searching

By using greedy algorithm, we iteratively search for adjure stations surrounded by POIs that posses the best ranking positions. As illustrated in Figure 2, this searching process can be examined by modeling it on a directed graph. Each vertex of the graph describes location of a train station, which can

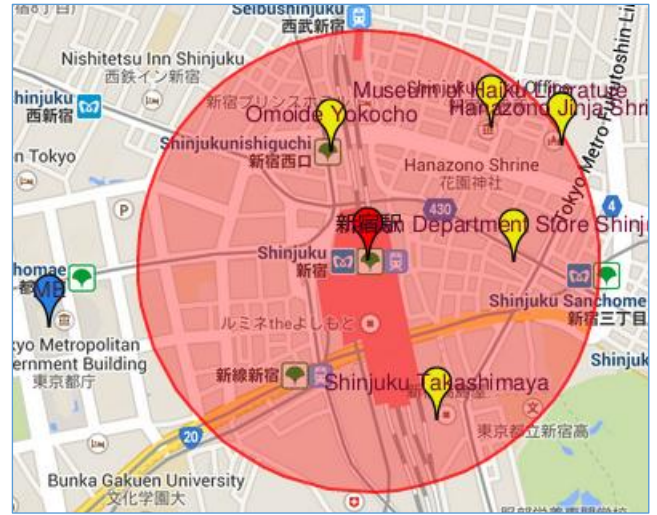


Fig. 2. Search result of a suggested trip of about 4 hours in cloudy weather condition, starting from the Sofitel Hotel Tokyo (blue marker), visit a nearby Shinjuku eki (red marker), and subsequently walks through each POIs (yellow marker) which include "Isetan department store", "Omoide Yokocho", "Hanazono Jinja", "Museum of Haiku", and "Shinjuku Takashimaya", elements of specified categories "Landmarks", "Specialty Museums", "Religious Sites", and "Shopping".

either be the one nearest to the user's current position, or to the desired destination, or the ones to pass through on the trip. Each of those nodes will have a weighted satisfaction score measured by summation of rankings of all of the filtered POIs that are grouped with that station. The lower the score, the more satisfying it is. Edges connecting vertexes represent the travel duration between them. The overall goal of the searching process is to accumulatively select the best POIs that are located as close to the current position as possible. Through the greedy algorithm, we try to find a good tour by searching locally from one vertex to another in order to choose the next best vertex to visit. The next best vertex (or station), must be as close to the current vertex (or station) as possible, and posses the most satisfied weighted score (e.g., Figure 2).

At each of the vertex, we subsequently apply greedy algorithm to select the best POIs located within a specified preferred walking distance, which both has high rankings, and satisfy the overall accumulated values of constraint variables such as the total available time, or budget. The search would stop once either the constraints are met or a balance number of

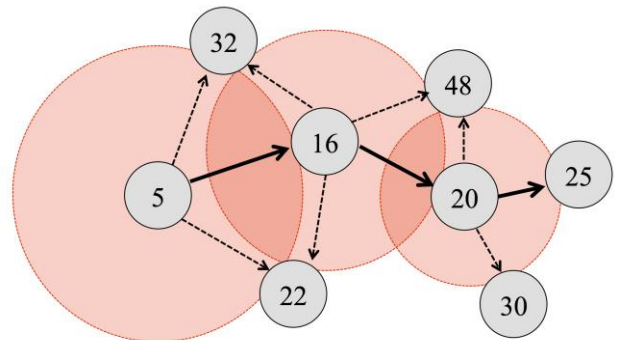


Fig. 3. Multiple circles with gradually increasing radius is drawn from each of vertex (small grey circles) to recognize its nearest next best station.

POIs from each of the desired categories is obtained. A sample result is illustrated in Figure 3.

3) Route Construction & Optimization

After obtaining a list of suggested POIs and its associated intermediary stations, an optimal sequence of traveling route that passes through all the places is investigated. Basically, the overall route will comprise of the traveling order through stations, and the walking order within a group of POIs located at a single station. However, after the searching process, there might be cases where POIs are more accessible from nearby POIs or a station belonging to another geographical cluster (e.g., Figure 4). In such cases, an optimization process is designed to iteratively examine each of the selected POIs and move them to a more appropriate central station, and eliminate any of the redundant stations from the constructed route.

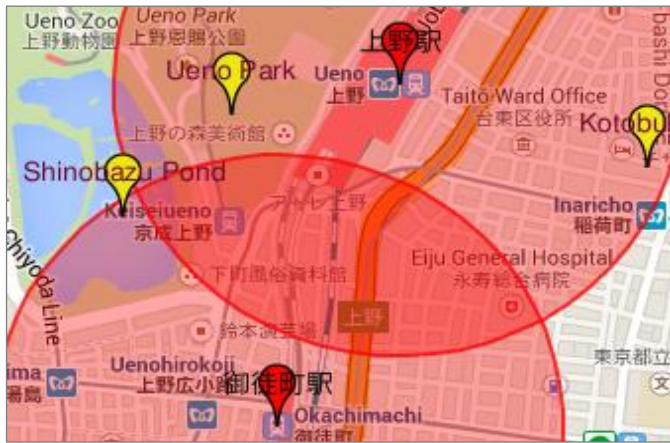


Fig. 4. A better solution can be achieved by matching the “Shinobazu Pond” POI (yellow markers) to the upper station to reduce the commuting time between the two stations (red markers).

The associated algorithm for each of the above mentioned 3 steps is shown in pseudo-code form in Figures 5, 6 and 7, respectively. Despite the fact that many genetic algorithms such as tabu search, etc., [16-17] facilitate the route optimization procedure, we have used Google Maps API because of its proven success and the practical nature of our research approach.

III. DISCUSSION

A. Comparison with Other Tourism Recommendation Systems

Even though a number of researchers have previously investigated tourism recommendation system, there is yet any commonplace regarding metadata (e.g. geographical coordinates), sources, and type of input data. This may be because of both the nature of employed algorithms as well as to whether a list of POIs or whole itinerary that are being achieved by the system. For example, some researchers have used collaborative filtering technique on previous tourism cases and collected data from a local tourism agency of a specific city to suggest a touristic planning for the user, while others have made use of available data sources such as public photos from Flickr [18], etc., to reach a list of favorable POIs. Hence, there is still a need of a widely acceptable generic

Result: Array filteredPOIs
foreach *POI* **in** *POIs* **do**
 if *POI* *match* *User.category* **then**
 if *POI* *match* *currentWeather* **then**
 ADD *POI* to filteredPOIs
 end
 end
end

Algorithm 1: POIs Filtering on Category and Weather

Fig. 5. Pseudo-code for step 1: Filtering POIs base on user’s preferences on choices of category and concurrent weather condition.

Result: Array finalPOIs, finalStations
nextStation = User.currentLocation;
while *checkTripSatisfied*(finalPOIs) == false **do**
 POIS = findPOIsNearStation(filteredPOIs, nextStation);
 POIS = sortByRankAscending(POIS);
 foreach *POI* **in** *POIS* **do**
 if *checkTripSatisfied*(finalPOIs) == false **then**
 if *POI.timeNeeded* < *User.timeConstraint* AND
 POI.budgetNeeded < *User.budgetConstraint* **then**
 User.timeConstraint -= *POI.timeNeeded*;
 User.budgetConstraint -= *POI.budgetNeeded*;
 POI.nearByStation = nextStation;
 ADD *POI* to finalPOIs;
 if nextStation not in finalStations **then**
 ADD nextStation to finalStations;
 end
 end
 end
 end
 else
 Exit For;
 end
end
nearStations = findNearStationsFromLocation(nextStation);
if nearStations is not null **then**
 nextStation = findStationHighestScore(nearStations);
end
else
 Exit While;
end
end

Algorithm 2: POIs and Station Searching

Fig. 6. Pseudo-code for step 2: Local search for POIs and Station with user’s constrains using Greedy algorithm.

Result: Array finalRoute
tripWayPoints = [];
foreach *POI* **in** finalPOIS **do**
 bestNearByStation = findBestNearByStation(*POI*, finalStations);
 if bestNearByStation not in tripWayPoints **then**
 ADD bestNearByStation to tripWayPoints;
 end
 ADD *POI* to tripWayPoints;
end
finalRoute = GoogleAPIDirection(origin: User.currentLocation,
destination: User.destinationLocation, waypoints: tripWayPoints);
Algorithm 3: Route Construction and Optimization

Fig. 7. Pseudo-code for step 3: Optimization of the list of stations, and construction of optimal travelling route through stations and POIs using Google API.

evaluation metrics, testing cases or approaches that could facilitate effectiveness and increase both satisfaction level and accuracy assessment of the recommendation system [14]. The following section briefly explains the uniqueness of our recommendation system by comparing its numerous features with other systems.

1) *Type of Recommendation System*: While many recommendation systems such as SigTur/E-Destination [19] mainly focuses on the act of POIs filtering, of which the final result is a list of POIs that match with user's preferences, our current research concentrates more on the touristic planning process, which embraces logistic consideration and geographical context of the user. In fact, the use of public transportation service into touristic recommendation system is yet claimed to be quite an open research topic [14].

2) *Scope of the Recommendation System*: Eventhough Tokyo's metropolitan area is chosen as a target vicinity in this paper, the availability and variety of data sources, as well as the examination of Japan's unique traits (e.g. complex train system, free train pass ticket) make the introduced methodologies applicable to other regions within Japan as well. In otherwords, other recommendation systems rarely take the localization aspect into consideration, implying that most of them are very generic (e.g. [18], [19]). As such, our approach can be considered as a pioneer tourism recommendation system for a specific national-level target (Japan), which undoubtedly gives no harm but more confidence in the feasibility of actual implementation.

3) *Input Data Space for POIS Filtering*: While many recommendation systems make use of restricted data sources, our system relies on data that are available to public. For example, algorithms such as collaborative filtering require an initial database or previous actual cases, which seems collectible solely through tours agencies. Geo-tagged sightseeing photos collected from various sharing platforms such as Flickr, Panoramio, etc. are also being captured to rank tourism attractions [18]. In our case, however, the system makes use of data available on the Internet and because two out of its three main data sources are very active Japanese online services related to the field of tourism, the input data could be considered more reliable, up-to-date and accessible by everyone.

IV. LIMITS AND FUTURE RESEARCH

The following limitations can provide space for improvement and direction for future research.

1) A systematic approach to update the current database on a regular basis is necessary for providing the most up-to-date information (e.g., new events) to the user.

2) The heuristic greedy algorithm finds the local optimal solution and the hope to meet the global optimum has proven its successful in many practical researches and applications [18, 20]. Despite that, there is a need for specific metrics to define the measurement of how good a tour recommendation is according to a specific user's profile.

3) The long-term goal of the research is to launch a system that not only able to draw a good tour plan in accordance with user's preferences, but also to recognize distinguished travelling interests of tourists from different countries while they visit and experience Japan. This can hopefully be achieved by applying machine learning and statistics to collected empirical data from surveys or preliminary mobile application and picture out the repeated pattern of a variety of cultural preferences. For example, Thai tourists tend to visit lots of temples while European travellers enjoy long-distance walk during the tour.

V. CONCLUSION

This paper shares the results of an on-going research in recommending tour planning to tourists in Japan under a practical perspective. It has collected a huge amount of supporting data, has built prototype of a system that uses greedy algorithm for POIs searching, and demonstrated its practicality. Despite its current limitations, potential benefits of such system that is generic to various tourism regions in Japan and tailorable to numerous users from different nations is shown to be high.

REFERENCES

- [1] JLL's Hotels & Hospitality Group, "Tokyo 2020 Olympics: expectations for the hotel industry, November 2014.
- [2] "Japan outlines plan to achieve 20m tourists by 2020 Olympics". (n.d.), retrieved May 3, 2015, from <http://www.travelweekly.co.uk/articles/2014/10/03/49577/japan-outlines-plan-to-achieve-20m-tourists-by-2020-olympics.html>
- [3] Japan Tourism Agency, "Efforts to promote a tourism-oriented nation", November, 2009.
- [4] "Not just international but 'Super Global Universities'", University World News. (n.d.), Retrieved May 3, 2015, from <http://www.universityworldnews.com/article.php?story=20141120233337379>.
- [5] "Promoting two-way student exchange". (n.d.), retrieved May 3, 2015, from <http://www.mext.go.jp/english/highered/1303572.htm>.
- [6] Japan National Tourism Organization, "Japan Rankings", retrieved May 3, 2015, from http://us.jnto.go.jp/survey/japan_rankings.php.
- [7] "Transportation - the official guide", Japan National Tourism Organization. (n.d.).
- [8] Han Jiho, Yotsumoto Yukio, "The trends regarding foreign tourists to Beppu, Oita prefecture in Japan", Journal of Ritsumeikan Social Sciences and Humanities, vol. 2, pp. 61-72.
- [9] "Attracting more tourists to Japan", The Japan Times, January, 2014. retrieved May 4, 2015.
- [10] João Romão, Bart Neuts, Peter Nijkamp, Asami Shikida, "Determinants of trip choice, satisfaction and loyalty in an eco-tourism destination: a modelling study on the Shiretoko Peninsula, Japan", Ecological Economics, vol. 107, November 2014, pp. 195-205.
- [11] Iponics Japan, "Marketing to Japan - appealing to the Japanese tourist". (n.d.). Retrieved May 3, 2015.
- [12] Gülçin Büyükoçkan, Buse Ergün, "Intelligent system applications in electronic tourism", Expert Systems with Applications, vol. 38(6), June 2011, pp. 6586-6598.
- [13] Hsiu-Sen Chiang, Tien-Chi Huang, "User-adapted travel planning system for personalized schedule recommendation", Information Fusion, vol. 21, January 2015, pp. 3-17.
- [14] Damianos Gavalas, Charalampos Konstantopoulos, Konstantinos Mastakas, Grammati Pantziou, "Mobile recommender systems in tourism", Journal of Network and Computer Applications, vol. 39, March 2014, pp. 319-333.
- [15] Wallace, Floyd, Associates Inc, "Massachusetts pedestrian transportation plan", 1998.

- [16] Kylie Bryant, I., & Arthur Benjamin. (n.d.). "Genetic Algorithms and the Traveling Salesman Problem".
- [17] Sylejmani, K., & Dika, A. "Solving touristic trip planning problem by using taboo search approach". *International Journal of Computer Science Issues (IJCSI)*, vol. 8(5). 2011.
- [18] Kai Jiang, Huagang Yin, Peng Wang, Nenghai Yu, Learning from contextual information of geo-tagged web photos to rank personalized tourism attractions, *Neurocomputing*, vol. 119, 7 November 2013, pp. 17-25, ISSN 0925-2312.
- [19] Antonio Moreno, Aida Valls, David Isern, Lucas Marin, Joan Borràs, SigTur/E-Destination: Ontology-based personalized recommendation of Tourism and Leisure Activities, *Engineering Applications of Artificial Intelligence*, vol. 26(1), January 2013, pp. 633-651.
- [20] Elias Koutsoupias, Christos H. Papadimitriou, "On the greedy algorithm for satisfiability", *Information Processing Letters*, vol. 43(1), 10 August 1992, pp. 53-55.