**Intelligent itinerary creator and optimizer**

**Research Report**

**Galan Radu**

# Abstract

Making travelling easier has been a hot topic in the last few decades due to the continuous increase of the number of people that are investing money in touristic activities. This has also strongly supported the interests of public and private organizations to also make investments in the infrastructure (by conserving historical sites, constructing new parks/location for entertainment and leisure activities, etc.). Relying on human curiosity (and even thirst of knowledge and new experience) the number of touristic attractions has exploded worldwide and it is becoming harder and harder for tourists to organize their travel plans/itineraries and routes. In consequence, entire hours are wasted with this process for every trip, and even afterwards people might not be satisfied (due to the famous concept of paralysis choice). We will go through some research studies that attempt to resolve this complicated problem of creating an itinerary (and optionally a route) using a very big database of touristic location.

# Introduction

This is a well-known problem that we are trying to find a solution for, and consequently it has many names. One of the most popular names in the literature “ The orienteering problem” which was first stated in 1987 by Bruce Golden, but it had another nomenclature even before that. Most scientist know the problem as “The traveling salesman problem” (first named so by Karl Menger – the famous mathematician). Although both these names cannot quite comprehend the actual complexity of the problem we will describe, they are still a pretty good point to start. We will attempt to compare a few works that consider more parameters related to the travelling experience in order to obtain the route and/or itinerary.

A few of the most important parameters that we will be searching are the schedule of certain objectives, the recommended visiting time of each touristic objective, preferences of the user regarding categories of the touristic objectives, multiple day trip, variability in the categories, alternative means of transportation: walking, personal car, bus, etc.

Motivations can be found simply for this problem. Urban areas will increase in size due to population rising, which means that more museum, more parks, more theaters (more point of interest for tourists) will be constructed or even renovated (we are referring here of course to historical sites, museums, and building). The simple idea that so many people are travelling more every year is a problem already by itself. More tourists mean longer waiting lines, they also mean that restaurants and hotels will be more likely to be agglomerated or unavailable at all, whilst the network of transport might fail altogether the biggest cities. This reduces the value of the time spent in these touristic objectives drastically due to longer waiting time, longer travelling time from one point to another, and even impossibility of dining or checking in. The latter motivation is actually just beginning to be researched since until recently only solution only seek local optimal (these approaches are called Selfish Routing approach)

We are talking about an industry sector that is worth 1.2 trillion $, but even so, most trips are organized by tourist themselves. This is both inconvenient for the tourist, and inopportune for the visited city because they cannot be sure the traveler will choose the best attractions in order to make a correct opinion about the locations. and even so many of the tourist have to organize everything themselves. Not only this is an inconvenient, but it is also a disadvantage for cities themselves. The process occupies a lot of time and is also complicated to be done by one who has no experience or few information regarding the local activities and attractions. Even if a few local tour guides would get organized, the problem is way beyond means of non-technological measures and they would surely get overwhelmed. Also, not only it is complicated, but nearly impossible, to choose the best public transportation alternative, considering the time needed for visit every attraction, calculating for each individual location’s schedule, and then continue to properly experience and enjoy the touristic objectives relaxed and open. This is not likely to happen and the image of the city, locations, and even country will be affected.

Another very interesting fact is the uniqueness of each traveler. While most of them might be willing to visit the more popular location, it is very probable that they will likely desire their own unique experience and will most probably get bored after being in a crowd too long.

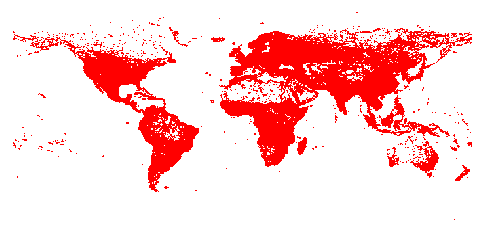
Freeing the user from all the tedious and repetitive decisions that he has to take during the planning of a touristic travel is the main concern of this study. Most important and hard of those tasks are the selecting of an ordered list (relative to one’s interest) of touristic objectives , designing an optimal route and itinerary which should contain as many of the point of interest as possible. Those are the information that we are also interested to find out.

During the following chapters the wider view on the problem will be explained along with the actual problem, and then some solutions will be presented and analyzed. We will end with some personal notes and a conclusion.

# Big picture

As we said before this problem has multiple names, but also has many variations. The TSP was studied and extended to multiple variations like the Travelling Salesman Problem with Priority Prizes (TSPPP) (Admilson Alcântara da Silvaa 2018) or Selective Travelling Salesman Problem with Profits (STSPwP) (Koszelew 2014), and then it turned into something more complex that is named today “The orienteering problem” (Bruce Golden 1987). The name actually comes from a sport in which each player seeks to reach multiple points on a map and for each of them he earns points. Whoever gets the most points, obviously, wins. This was a game that usually took place in hard to access outdoor areas (for example forests), now the strategies they empirically discovered playing this are strongly bound to our problem(and the general problem of mass navigation). The problem obviously developed many variations like Team Orienteering Problem (TOP), Team Orienteering Problem with Time Windows (TOPTW) (Koszelew 2014).

A rather more complex version of this approach that might be more accurate to our situation is the Tourist Tour Generation Problem (TTGP) very thoroughly explained by (Koszelew 2014).

Something very important which is worth mentioning is the The world TSP (University 2020) The university of Waterloo constructed open data library that consists of 1.9 million cities all over the world. A lower bound has already been found as a solution( 7’512’218’268 by CPLEX linear-programming solver). At the moment it also stores data on any other algorithms that attempts solve the problem (the best one has an error of 0.0474%)

# Problem description

Simply put we will start with the following information:

-a list of possible touristic objectives (POS) in multiple location and with multiple details about each of them regarding their schedule, activity type, coordinates, etc;

-a complete profile of the user and his preferences,

And we want to construct an intelligent algorithm that can deduce the following information:

-the ordered list for touristic objectives (sorted by the user’s preferences)

-an itinerary with the touristic objectives for the selected period taking into consideration as many parameters as possible.

This problem’s complexity is quite high in general, and it is also categorized as NP hard (Garey M 1979).

In her synthetization (Koszelew 2014) there are presented the majority of the important features an algorithms of this type could have:

• Selecting and Routing(SaR): Basic features that generated a route as short as possible and that respects some of the basic constraints.

• Possibilities of Returns (POR): The route can add a location twice (with 0 points given the second time). This is useful for algorithms that rely on graph theory methods.

• Obligatory POIs (OPs): Introduces the notion of “must see” attraction that must be found in the final route.

• OpeningHours (OH): The route will only include an attraction in the route if by the approximated time the tourist visits, it will still be open.

• ScenicRoutes (SR): Attempts creating routes that will pass near objectives like monuments, architectural buildings.

• PublicTransportation (PT): Considers multiple types of transportation aside from personal car ro walking, like: train, plane, bus, etc;

• BudgetLimitations (BL): The objectives on the route will not exceed a certain budget limit.

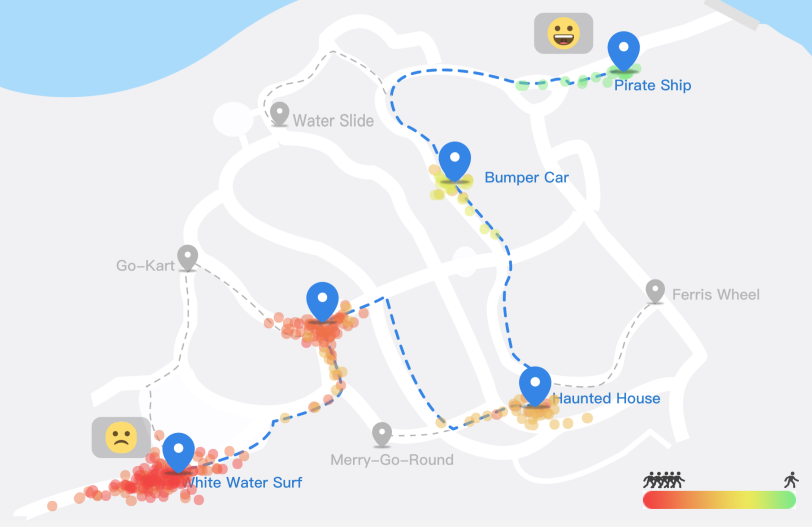
• DynamicRecalculation (DR): allows the recalculation of the route during the visit in order to cover unforeseen events.

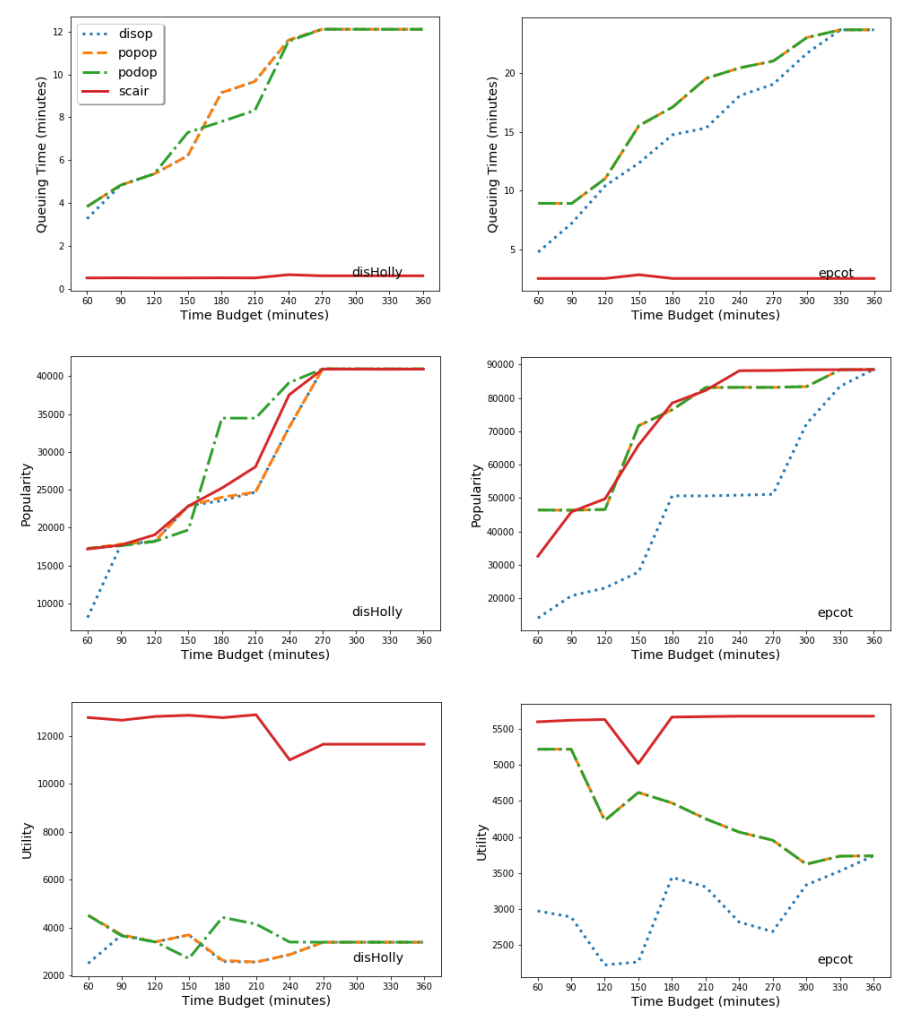
• Max − nT ype (MnT): Constraint that only allows a limited number of a certain category of touristic objectives.

• ObligatoryTypes (OT): Constraint that only allows a route to be valid if it contains a minimum from a certain category of touristic objectives.

# Strategic and Crowd-Aware Itinerary Recommendation (Junhua Liu 2020)

The problem is formulated here as “route recommendation as Markov chains”. It takes a very interesting approach on the problem using a big Flickr database of images that also contains geo-tagged data and queueing data about each individual location.

The main attribute of this research is the fact that it tries to extend the perspective for this very problem. Since most of the algorithms that are created or studied for the orienteering problem/travelling salesman problem are based on a Selfish Routing approach (as has been called in the community), the perspective is quite limited regarding the actual scalability of the algorithm. Although there are multiple decent approaches that manage to find a solution for a relatively big database and with multiple constraints, all of them only consider the case of a unique user (this is where the Selfish comes from). You cannot have reached the global minimum and a stable scalable solution unless you consider the fact that multiple users could be using the same route at once and the initial obtained route will no longer be viable, because the times will increase directly proportional with the level of crowdedness.

They introduce a crowd-aware itinerary recommendation algorithm (called SCAIR) that should be able to optimize the solution for multiple travelers. It is compared with other simple rule-based and greedy-based approaches on the forementioned dataset from Flickr (Epcot Theme Park and Disney Hollywood Studio locations).

It is compared with 3 algorithms:

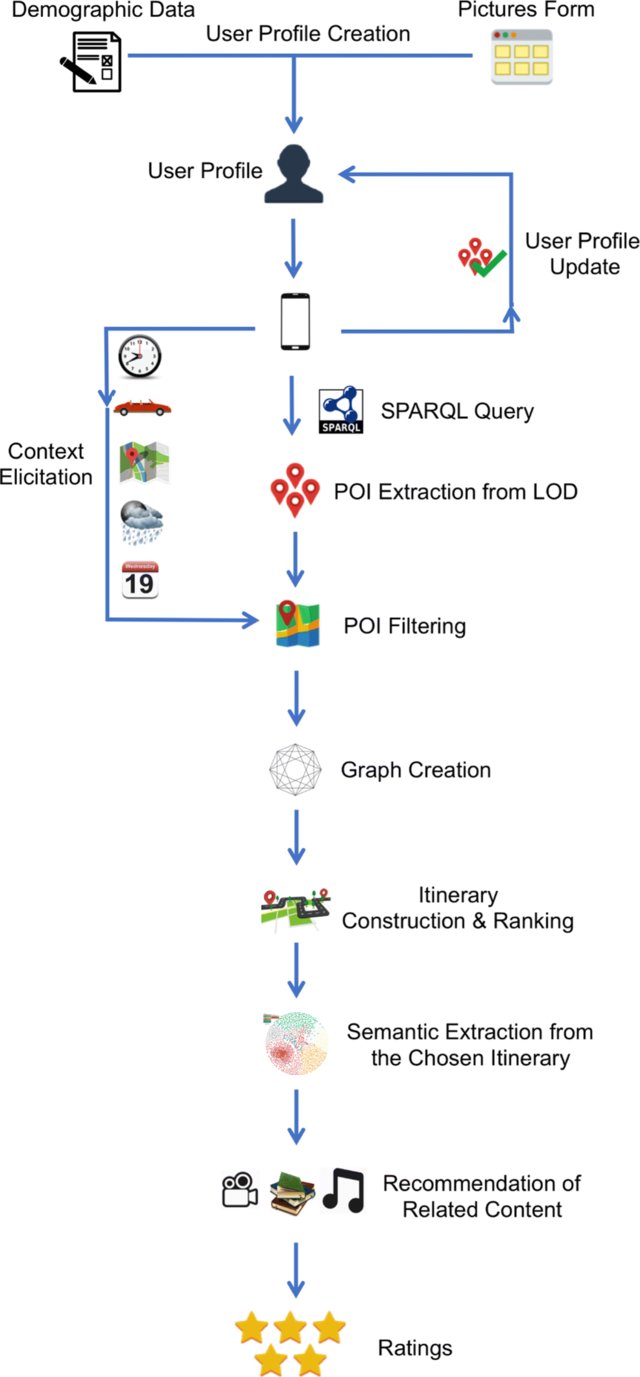
-Distance Optimization (step by step, finds closest objective) - disop

-Popularity Optimization (step by step, finds most popular objective) - popop

-Popularity over Distance optimization (greedy step by step) - podop

As can be seen it runs very good comparing to the other implementations (but let’s not forget that they are not state-of-the-art).

# Exploiting semantics for context-aware itinerary recommendation (Alessandro Fogli 2019)

A very interesting, complex, and comprehensive study has been recently written (about a year ago) and has created an entire system that is able to create accounts for users (containing preferences regarding different categories of touristic objectives.

The article takes advantage of linked open data (LOD) in order to achieve ‘contect-aware recommendation of personalized itineraries. Moreover, it also recommends for each point of interest (POI) on the generated route a piece of media (text, video, image) which is related to the user’s preferences.

The system has 6 main components:

-acquiring information about the user and its preferences through the user profile creation

-understanding the current context

-selecting the right POI from the LOD with proper queries

-generating k itineraries that respect the constraints

-use multiple factors of recommendation to sort the itineraries

-recommend textual and multimedia contents related to the itinerary

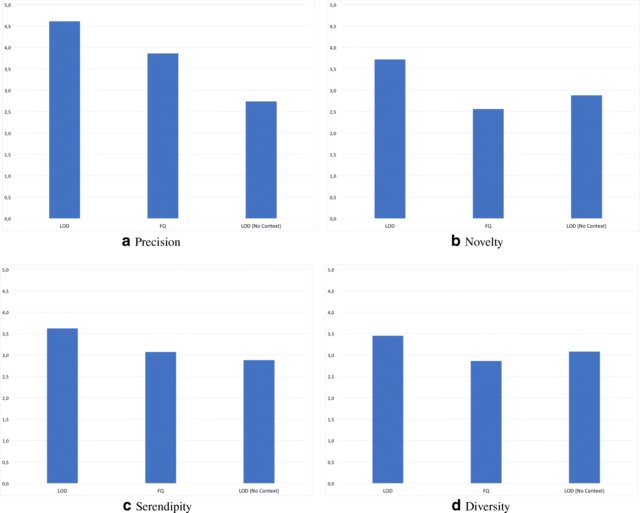
Some of the most interesting factors for recommendation are:

-POI popularity based on reviews

-POI diversity considering the number of categories

-the distance and traveling time of the itinerary

-the user preferences, the physical and social context.

The application of the algorithm has been tested fairly well with 50 real users and using normalized discounted cumulative gain, precision and other metrics as can be seen in the diagram.

# Solving tourist trip planning problem via a Simulated Annealing Algorithm (A. M. Kadri Sylejmani 2014)

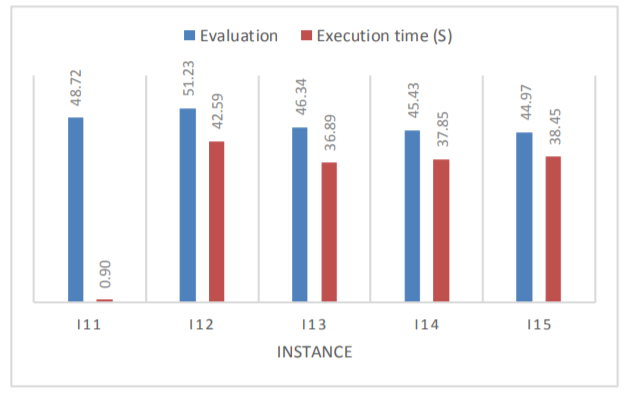
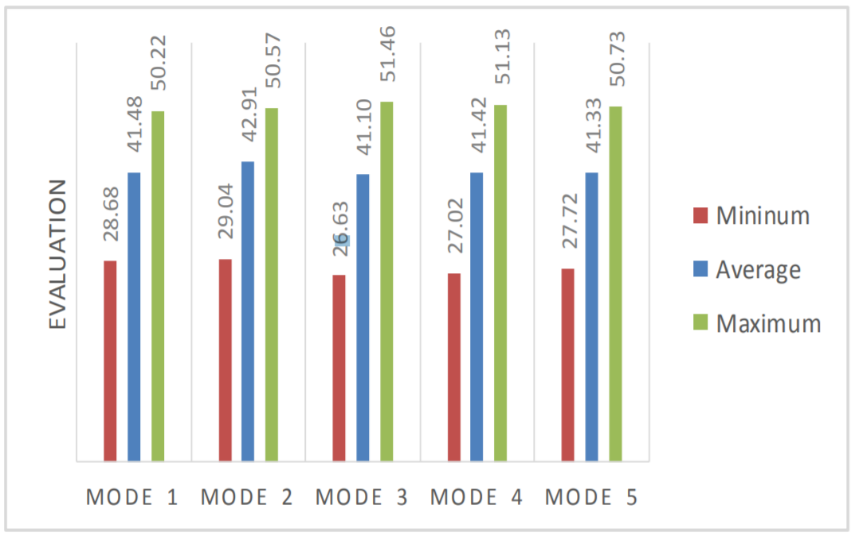
The following paper presents another algorithm that will create an itinerary which takes into consideration trip of multiple days, a limited budget to spend, the schedule of each location, breaks for eating/relaxation. Apart from the usual component that attempt maximizing the score of ‘satisfaction’ of the user, it also takes into consideration minimizing the traveling time along the route.

The trip begins and ends in exactly the same point (the hotel supposedly) and it is deviated from the TOPTW (Team orienteering problem with time windows). The solution uses simulated annealing (SA) as a metaheuristic from thermodynamics. The algorithm transforms into something rather difficult due to the number of dimensions and attributes. Using a self-defined fitness function (which considers two factors of satisfaction: regarding the quality of the objectives and the travelling time) they manage to evaluate multiple models creates.

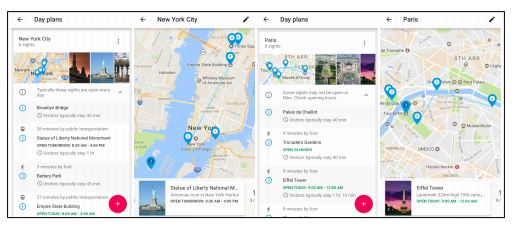
There are 5 modes created that all use a form of simulated annealing:

* Mode 1: randomly generates the POIs
* Mode 2: choose the POI ordered by the POIs popularity
* Mode 3: choose the POI ordered by the POIs popularity (but reversed)
* Mode 4: choose the POI ordered by the proximity to any other POIs
* Mode 5: choose the POI that have the shortest visiting time

As can be seen in the diagrams below, the algorithm behaves rather well on the given dataset (limited to only 15 entries), but it is moving rather slow (30-40 seconds).



# Other mentions

A rather important study comes right from a very experienced team from Google that researches travel itineraries generation. (Zachary Friggstad 2018) They have available huge amount of data, and they make good use of it to improve the algorithms of itinerary and route (which is usually more complicated due to lack of public available data and services) generation. An important point they are trying to accomplish is to balance out the categories and contest of interesting objectives in each day.

A few other papers that are rather interesting are:

* Tabu search for priority prizes - (Admilson Alcântara da Silvaa 2018) where they use tabu-search to improve based on previous steps. It was tested against CPLEX and received great results.
* Operating Research - (Ander Garciaa 2009): it calculated using a rather simple algorithm (Iterated Local Search), but which just enough to solve the Multi Path Team Orienteering Problem with Time Windows (MPTOPTW) and even consider multiple means of transportation and varying travelling hours.
* Group tour recommendation (Lim 2016): where they are using points of interests scrapped from the internet and geo-tagged photos to create travel sequences and routes based on tourist interests (using categories) and popularity of each point of interest.

# Conclusion

After seeing multiple methods we can say that this is a very active field of research and many intelligent and variated algorithms have already been tested on this problem, but there is still much to be improved and many other features of the problem are yet to be solved at the same time. Let us hope that the interest in this problem will not stop and it will still be studied and upgraded.

# Bibliography

Admilson Alcântara da Silvaa, Reinaldo Morabito, Vitória Pureza. 2018. "Optimization approaches to support the planning and analysis of." *Expert systems with applications* 321-330.

Alessandro Fogli, Giuseppe Sansonetti. 2019. "Exploiting semantics for context-aware itinerary recommendation." *Personal and Ubiquitous Computing Springer.*

Ander Garciaa, Maria Teresa Linaza, Olatz Arbelaitz, Pieter Vansteenwegen. 2009. "Intelligent Routing System for a Personalised Electronic Tourist Guide." *Information and Communication Technologies in Tourism* 185-197.

Bruce Golden, L. Levy, Rakesh V. Vohra. 1987. "The orienteering problem." *Naval Research Logistics 34(3)* 307-318.

Emmanuel L. Seetam, Ledisi G. Kabari. 2019. "Intelligent Agent Based Itinerary System for Route Optimizer and Tourist Guide." *International Journals of Advanced Research in Computer Science and Software Engineering* 90-97.

Garey M, Johnson D. 1979. *COmputers and intractability.* San Francisco: Freeman.

Junhua Liu, Kristin L. Wood, Kwan Hui Lim. 2020. "Strategic and Crowd-Aware Itinerary Recommendation." *Proceedings of the 2020 European Conference on Machine Learning and Knowledge Discovery in Databases (ECML-PKDD’20.*

Kadri Sylejmani, Atdhe Muhaxhiri, Agni Dika, Lule Ahmedi. 2014. *Solving tourist trip planning problem via a.* Kosova: MIPRO 2014/CIS.

Kadri Sylejmani, Jurgen Dorn, Nysret Musliu. 2017. "Planning the trip itinerary for tourist groups." *Inf Technol Tourism* 275-314.

Koszelew, Jolanta. 2014. "Optimization Problems and Methods Applicable in Intelligent Tourist Travel Planners." *Trends in Contemporary Computer Science, Podlasie* 127-134.

Lim, Kwan Hui. 2016. "Recommending and Planning Trip Itineraries for Individual Travellers and Groups of Tourists." *Association for the Advancement of Artificial.*

Zachary Friggstad, Sreenivas Gollapudi, Kostas Kollias, Tamas Sarlos, Chaitanya Swamy, Andrew Tomkins. 2018. "Orienteering Algorithms for Generating Travel Itineraries." *WSDM’18, February 5-9* 180-188.