UDC 00.004

DEVELOPMENT OF INTELLECTUAL SYSTEM FOR GENERATING PERSONALIZED ROUTES FOR SAINT-PETERSBURG

Mikhail Galperin (ITMO University, faculty of informational communication technologies)

Academic adviser – docent of ICT faculty, doctor of technical sciences Maksim Khlopotov (ITMO University)

The task of developing a system for generating personalized walking routes for Saint-Petersburg is being reviewed in this report. The system contains of the server application that provides collected and structured data to all client applications. The major focus of development process is placed on two tasks: collecting appropriate data, filtering it and converting to the format needed and on routes generation algorithm which takes user interests into a consideration.

Data preparation task. Routes generation system requires a huge amount of trustworthy data with a special structure. Relevant data that already exists does not respect all structural requirements or sometimes does not contain information that is needed. At the same time the sources that were not parsed ever exist. By reviewing the example of collecting persons related to Saint-Petersburg the entire process of data preparation is being explained.

The suggested solution includes a set of technical instruments to work with Wikipedia providing opportunities to retrieve, handle and convert intro an appropriate structre from the specific categories organized on the Wikipedia side. The term of “scoring system” is being introduced to ranging all the information located in the category that matches conditions of this task. Every data category is being tagged by an expert group, then integer value is being calculated for each tag which is later being used in a final score formula for each entity. Also an expert group sets up minimal values for every set of tags that allows us to add a step of removing least relevant or low in content records.

84k records were collected while solving the data preparation task. Passing it through the abovementioned scoring system led to 3.2k records removal. Processed and converted records that left are ready to be imported into a database and can be used for the recommendation system later on.

A set of automatic tests was made for every tool implemented during the work process that lets stay sure about accuracy of the system in general and each of its component. This is very much useful in terms of adapting to any kind of change on the data source origin whether they can be content or structure related.

Personalized routes generation algorithm. Routes generation is being performed using the data collected in the previous task. The dataset is about famous persons and famous places of Saint-Petersburg, their relations a major part of which is automatically matched and a set of user interests like their location, preferred route duration and length. Taking the distance between route stops, availability of every stop to either be added or removed from the route and the level they match the user interest into a consideration a completely new route can be generated or early prepared route can be restructured at the moment of being requested from the client application. To decide whether a stop can be removed or added to the route a lot of factors regarding of how much favorable the walk will be for the user are being considered.

The final system is offered for testing purposes to the focus group of potential users and to the expert group. Later on it is also ready for being deployed to the local and international tourism markets.

Mikhail Galperin (author) Signature

Maksim Khlopotov (academic adviser) Signature