

# EVENT DENSITY MAP

Ceren YAŞAR  
Elif Şahin

17 January 2019

## 1 Abstract

”Event Density Map” project includes the development of a web application which presents the event information by collecting and saving the event information from the sites which are regularly publish the events. It also includes creating the statistical results by using event database and filtering the events according to the user’s choices.

One of the objectives of this study is to eliminate the need to use more than one website when investigating the activities will be performed. In addition to this purpose, the detailed filters provided in the application provide faster access to the desired. For this purpose, in addition to obtaining all information about the events from the sites, the irregular data given on the sites have been tried to be reached by using regex.

Another aim of this study is to ensure that detailed activity data collected can be used for various researches. These data were obtained by scanning sites using the BeautifulSoup and Selenium libraries. Figure 1 shows the data collected from published events. In addition, the database is visualized with graphs.

At the end of the project, the web application was implemented. An admin page has been added to allow the content of the web application to be edited by the admin. The pages that users can log in and the register were created. A web crawler system that can update the event database with appropriate intervals was developed. The event database was visualized by statistical graphs.

## 2 Introduction

”Activity Density Map” project investigate activities in a region with the help of event sharing sites after that it decomposes the activities into different clusters by time, category, subcategory, location, capacity etc. and show the numerical statistics of clusters on a website.

The purpose of creating an Activity Density Map is to search for events in a region and eliminate the need to use more than one site when reaching their location. Activities that are closest to the point set by a single site will be

visible. For this purpose, the data collected from the sites like Biletix, TFF etc. where the events are published will be presented to the user on the site [1, 3].

The need for this type of site is because the activities gathered from different sites can be put together in a single system to create a better user experience and to present the events in an interface focused on the activity locations. The largest part of this interface will be the map, and the activities will be searched via filters such as the date, activity category, radius, and the appropriate activities are shown by marking their location on this map. The data was taken from the event ticket sales sites by "web crawling", and this data was stored in a NoSQL database for processing this information.

A statistics page has been created by using the activities recorded with various features in the database. On this page, graphs are presented that showing the distribution of activities by provinces in the desired date range, the distribution of activities in the districts of the big cities, the distribution of activities by categories, etc.

In accordance with these purposes, the data were withdrawn from the event ticket sales sites, data saved in the database in an appropriate format, updated at regular intervals and a mobile / website that uses stored data is designed. In order to make inferences by utilizing the activity data, the above mentioned statistical page has been implemented.

### 3 Materials and Methods

This application is aimed at presenting data on multiple activity sites from a single site to the user with an interface that highlights location data. The system was temporarily worked on a single site during the implementation phase, and after the completion of the development, the TFF site was added to the application. In this respect, the BeautifulSoup and Selenium libraries were used to access the data on the sites [2]. With the help of these libraries, links were visited in the content of the site and the desired data was recorded in the database.

Some of the activity data we need are not clearly available on the sites. This data is the end time and event location capacity of the events. The end times of the events are irregular in the text where the activities are defined. Sitedeki bazı etkinlik yerlerinin kapasitesi hakkında bilgi yoktur. The capacity information was added to the database by accessing the Biletix event location information page and by using regular expressions in texts accessed by searching the event places on Wikipedia. Due to the fact that the information about the duration of the event was found with very different variations in the texts, the success rate would be low, it is intended to give an average value by type of activity. These data and activity name, date, start time, event venue, latitude-longitude position, category, subcategory, activity visual link, event description, the most appropriate ticket price are recorded in the database.

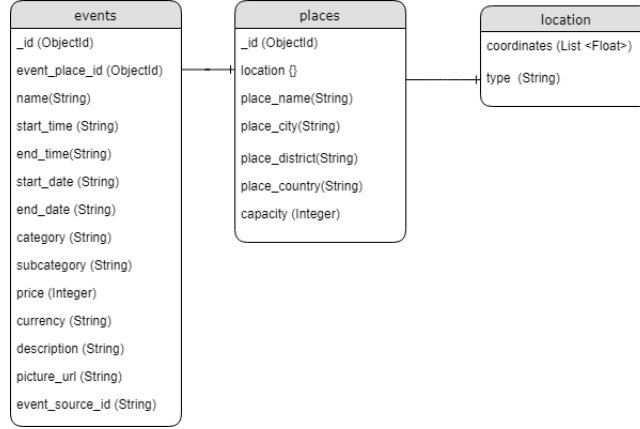


Figure 1: İlişkisel olmayan veri tabanı tabloları

In order to keep the data to be stored in the database up to date, data should be drawn from the sites at regular intervals. It will be costly to do this frequently for all events. Since the changes in the activities that will take place within 3-4 days in front of the current day will be more critical, these activities are planned to be updated several times during the day and activities other than those will be updated once a day.

When the site was viewed, it was seen that all activities on the site were not easily accessible. This is because the site can show only 300 events in a search. In order to solve this problem, the activities of each day were planned to be filtered by that day.

## 4 Conclusion

The topic of the project contains four related objectives. The first objective is to draw data from different event sites. The second objective is decomposing the activities into different clusters according to properties such as category, time, location etc. The third objective is statistical graphing of the data in the database. The fourth objective is to implement a web application that can be used by the end user, covering all of these objectives. For this purpose, a system that can scan multiple websites, transfer them to the database and update the data is implemented with Python programming language.

A web application that uses the database created by this system is implemented with HTML, CSS and JavaScript. In order to enable the users to filter the data in detail and to make the database comfortable to use in various researches, some data that are not structural on the sites were obtained from the texts with the help of regex.

In the project, the goals of creating the activity database and implementing the activity map web application have been reached. However, the efficiency

of the collected activity data on the traffic and the interpretation of this data could not be performed sufficiently.

In future studies, users can share their favorite activities and notify the user when similar activities arrive in the system. Traffic alerts can be sent by mail for very large events that are very close to the user's home and workplace.

The application can be designed as a mobile application, so it can instantly notify the user when he has an activity that he might love when he gives access to the location data.

## References

- [1] Biletix mobil app.
- [2] Python with selenium, beautiful soup.
- [3] Turkey football federation.