Czech University of Life Science Prague

Faculty of Economics and Management

Department of Information Technologies

DIPLOMA THESIS ABSTRACT

Analysis and design of web application based on open data

Author: Damir Pirija

Supervisor: Ing. Miloš Ulman, Ph.D.

1 SUMMARY

As trends in web application development change almost every year, it is important to follow them, and follow the best practice of web application design and development in order to produce a product which will be widely used and which will produce results and conclusions as are intended. Every information system, including web applications, should be created by following systems development life cycle. It guides us through the process of creating an application or a system from the moment the idea appears in our mind, until, the last, testing and maintenance period. Furthermore, the term "Open data" will be discussed. This type of data is generated by different companies and government institutions and is open for free use. This type of data has proven to be very useful due to the different attributes and situations from which it is collected. Then, this data can be used to produce meaningful calculations, predictions and conclusions which can benefit the community. During the reading of this work, reader will find out more about the systems development life cycle, technologies used to create the actual web application, and the results, benefits and drawbacks of development of application which makes use of open data. Furthermore, in the practical part, a web application will be created using HTML5, CSS3, JavaScript and PHP, based on dataset from United Kingdom's police, which includes details about various crimes, such as type, date, location and other attributes. This data will be inserted into a SQL database, which will feed the data to the web application. Then, the web application will present this data using a Google Maps API and other meaningful sections, which will provide assistance and insight to the local community and tourists as well.

Keywords: Open Data, PHP, Data set, Crime, Database, SQL, Web application, HTML, CSS, Analysis, Design, Open Data application

2 INTRODUCTION

In order to describe analysis and design process of a web application which is based on open data, it is necessary to shortly describe what open data actually is. The idea behind open data is that certain data which can serve a greater purpose should be available to everyone to use. Main goals of the movement which is supporting the use of open data are that this data should be without any restrictions and should not be bounded by copyrights and patents. Open data should be open source, with open content and open access to everyone. The concept itself is not new, but when put in the use with World Wide Web (WWW) it gains different dimension of its meaning. Open data is often made of non-textual material, such as mathematical formulas, statistical data, various parameters, medical data, and other. It is often provided by government institutions.

Systems development life cycle (SDLC) is a cycle that guides us through different stages of systems development, such as software specification, design, implementation, validation and evolution. During the reading of this diploma thesis, reader will find details on processes that are happening inside these stages, depending on project requirements. Furthermore, a web application will be developed which will make use of open data in order to display the information in an understandable manner. Different sections of the web application will display the data by different means. These include mapping each row of the dataset to a particular location on a map, providing the dataset in tabular view, and computing certain statistics based on open data attributes and values.

Finally, possible benefits of a web application which uses open data will be described, and different ways of usage of the web application will be explained in order to provide maximum benefits to the end-user.

3 AIMS AND METHODOLOGIES

The diploma thesis investigates the advantages and disadvantages design and development of web application based on open data. Main goal of this diploma thesis is to develop a web application which will make use of open data in order to provide meaningful insight to final consumers based on the dataset provided. Partial goals are:

- To explain the concept and purpose of open data;
- To provide overview of web application development life cycle which includes analysis, design and implementation stages;
- To develop a web application which will make use of open data which will be imported to the database;
- To present valuable information and draw conclusions from the web application which uses open data dataset.

The several methodology parts are defined to accomplish the thesis. First part will be focused on the theory behind open data, its use, and different models of open data. Furthermore, technologies and methods which will be used for the development process will be discussed. Development life cycle will be described and explained. Second part consists of actual design and development of a web application which uses a dataset related to crime. Information about crimes will be loaded into the database. Then, the application will present that data in an understandable way to be interpreted by the user, from which he or she can draw conclusions. Based on the literature review and results of practical part, conclusions will be formulated.

4 RESULTS AND DISCUSSION

As soon as Crime Spotter, the web application has started gaining a final form of its structure and functionality, it was already possible to see the output of this kind of web application. In this case, raw open data, which was supplied by the police institution of United Kingdom, was processed in order to be highly interpretable by the end user.

The main page contains all the main functionalities of Crime Spotter. An interactive map, which was placed in the first section, was fed raw data that was imported into a MySQL database, processed, and stylized in order to improve user experience while using Crime Spotter. The map was created with the use of Google Maps API, and it contains crime that was mapped according to its location. End user was provided with the ability to check criminal activity that surrounds him and his loved ones. Based on these insights, user can

choose what to do with this information, and how to use it to gain advantage in decision making process. These decisions come in various forms, such as choosing where to rent or buy a real estate, or through which street to pass on his way back home. Families with children can make a decision on whether a certain neighbourhood is safe for their children to grow up in, or go to school in.

After the map section was successfully developed, a new section was introduced, a section containing charts. As volume of criminal activity changes from month to month, a way to follow these changes was necessary to be implemented. By using Crime Spotter, a person can see how criminal activity has increased or decreased in their place of living. For the purpose of testing the application, a small dataset containing criminal activity data in the city of London has been used, but the database can be expanded with data from all parts of the world. Users can see whether they should increase their security measures regarding any type of property, based on the insights gained from provided charts. Third section that was created displays a table, with all information on criminal activity from the database. If the user is looking for information on specific crime, he can use a built in filter to search through the database of crimes for a particular activity of interest.

Figure 1 - Final view of the main section
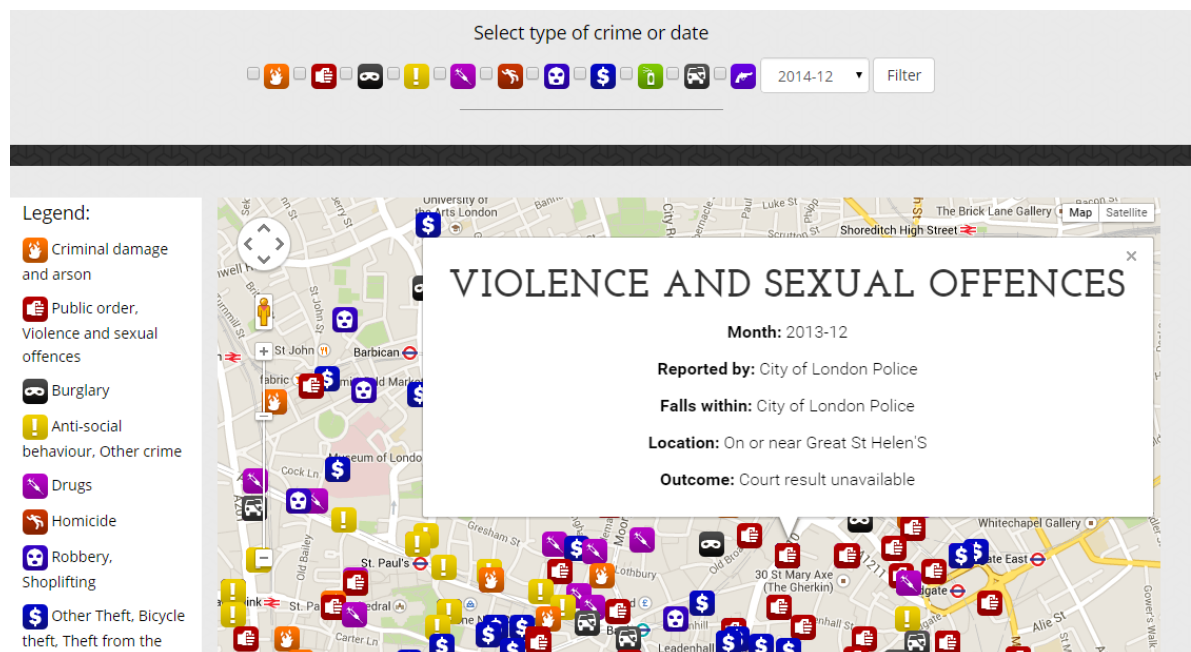
In order to compare features of Crime Spotter to those of similar existing web applications,

two crime mapping platforms have been selected, **CrimeReports^{TM}** which is located at https://www.crimereports.com/ and **CzechCrime** located at http://www.czechcrime.org/. It is important to note that these two mapping platforms are developed and maintained by teams of developers, and therefore have a higher set of functionality than **CrimeSpotter.** The following table introduces an overview of main features of each web application.

Feature	CrimeReports^{TM}	CzechCrime	CrimeSpotter
Data	Criminal activity for USA, Canada, UK	Criminal activity for Czech Republic	Criminal Activity for London, UK (test dataset)
Filter	Based on type of crime, exact date, time period, particular address, distance from current location, area, department	Based on type of crime, time period, address, crime rate	Based on type of crime, time period
Markers	Supported, show exact location of a particular crime	Not supported	Supported, show exact location of a particular crime
Areas	Not supported	Supported	Not supported
Statistics	Not supported	Supported; Number of crimes per area, number of crimes per type, graphical representation of crime rates, comparison between areas, charts per area	Supported; Number of crime per type, pie chart representing crime count per type in a certain period, bar chart providing crime counts per type in different periods
Other	Possibility of user registration, crime tip submission	No possibility of user registration, possibility of providing feedback.	Possibility of user registration, crime tip submission

Table 1 - Comparison of crime mapping web applications

The table above shows main differences between the three crime mapping web applications being compared. If we focus on **CrimeReports^{TM}**, we can see that the main, and only section is the map section, which has markers placed on exact locations where the crimes have happened. It is possible to filter the data by selecting crime type of interest, a certain date, or even a particular police agency. However, what is missing is an in-depth view of criminal activity. Each marker provides basic information on a particular crime, but there is no possibility of comparing overall criminal activity for a certain area or time. **CzechCrime** provides a very detailed user interface, along with the map of criminal activity. A user can find many comparisons, charts, statistics about crime overall, but also for a particular crime type. The map of Czech Republic is divided in smaller sections, and user can get information on number of crimes for a particular section. However, if a user wants to find a certain criminal activity in one of the streets, there is no possibility of that. The reason for this can be a constraint in data usage, in order to protect privacy. Overall, **CzechCrime** is a very good example of crime mapping web application with an in-depth overview of statistical information and trends in criminal activity in Czech Republic. **CrimeSpotter** is using a small dataset only for testing purposes. It provides basic statistical information in form of pie and bar charts. In order to expand a set of features, CrimeSpotter could implement some of the features of **CzechCrime**, similar to area filtering support. This can be useful if datasets from more countries would be included in the database, as displaying markers for every city in every country increases loading time and requires more processing power. So, dividing the information into larger areas would decrease the time and power required for display of that information, which would lead to better user experience and usability of web application. However, it is important to note that a crime mapping web application, such as CrimeSpotter, is fully dependable on the open data it uses. If the data that is provided is inaccurate, or contains certain constraints, usage of such data will also be under constraint, which means that the end-user will not reap full potential benefit of using it. Furthermore, some countries still do not support publishing of data regarding criminal activity inside the country. This poses a barrier in actual development and usage of the application, because not all citizens from different countries will be able to properly use it. In order to overcome these barriers, it is important to further promote the usage and publishing of open data, as well as educating the society about the potential benefits which could be gained.

5 CONCLUSION

In order to conclude the writing of this diploma thesis, it is important to review what the expected objectives were, and in what measure were they fulfilled. The main aim of this work was to analyse, design, and develop a web application which uses open data. This was done with the development of Crime Spotter, which uses raw open data and displays it in several easy to understand views. Before the actual development has started, concepts of open data, its benefits and barriers were described in detail. Furthermore, prior to the development of web application, the software and web development life cycles were presented, that would serve as a developers' guide to forming an approach to the development process.

As a part of development, the open data that was to be used was analysed and described, in order to understand the dataset before usage. Furthermore, models were created that show the structure and intended functionality of the web application. In order to proceed with the development, several wireframes were designed which have shown the actual design idea of Crime Spotter. Based on completed analysis and design, Crime Spotter was developed using client-side and server-side technologies. The web application provides a map of criminal activity for a particular location, which can be interpreted and used in many situations.

Furthermore, charts were included which show trends in the numbers of criminal activity. Lastly, a table populated with detailed information on criminal activity was provided to be inspected, used and searched based on a particular crime in mind. As a result of production of this sort of web application, it was shown how end users can draw meaningful conclusions from the usage, and how the same application can be applied to provide benefits for small communities, as well as whole countries and continents.

6 BIBLIOGRAPHY

Chun, S. A., Shulman, S., Sandoval, R. & Hovy, E., 2010. Government 2.0: Making. *Information Polity,* Issue 15, pp. 1-9.

Frain, B., 2012. *Responsive web design with HTML5 and CSS3.* vi ed. Birmingham: Packt Publishing.

Gurin, J., n.d. *Open data now: the secret to hot startups, smart investing, savvy marketing, and fast innovation.* XVI ed. s.l.:s.n.

Heeks, R., 2006. *Implementing and managing eGovernment: an international text.* IX ed. s.l.:Thousand Oaks.

Holdener, A. T. I., 2008. *AJAX - The Definitive Guide.* 2nd ed. s.l.:O'Reilly Media, Incorporated.

Janssen, M., Charalabidis, Y. & Zuiderwijk, A., 2012. *Benefits, Adoption Barriers and Myths of Open Data and Open Government.* s.l.:Taylor & Francis.

Lee, T.-B., n.d. *Five Star Open Data.* [Online]
Available at: http://5stardata.info/
[Accessed 12 February 2015].

Lockhart, J., 2015. *PHP The Right Way.* [Online]
Available at: http://www.phptherightway.com/
[Accessed 05 March 2015].

McFarland, D. S., 2013. *CSS3: The Missing Manual.* Third ed. Sebastopol, CA 95472: O'Reilly Media Inc..

Meeker, M., 2010. *Internet Trends 2010 by Morgan Stanley's Mary Meeker.* [Online]
Available at: http://bbh-labs.com/internet-trends-2010-by-morgan-stanleys-mary-meeker/
[Accessed 25 February 2015].

Moore, R. et al., 2011. *Understanding HTML5: Today's realities and its future market potential.* [Online]
Available at: http://www.slideshare.net/iLoopMobile/understanding-the-state-of-html-5-and-its-potential
[Accessed 02 March 2015].

Network, M. D., 2015. *Mozilla Developer Network: Cascading Style Sheets.* [Online]
Available at: https://developer.mozilla.org/en-US/docs/Web/CSS/Syntax
[Accessed 25 February 2015].