Data Visualization for Multimedia Super Corridor Malaysia using Dot Distribution Maps

Mohd Suffian Sulaiman¹, Amira Yasmin Raimey¹ and Azri Azmi²
¹Universiti Teknologi MARA, Malaysia {suffian@tmsk.uitm.edu.my, amira.yasmin0510@gmail.com}
²Universiti Teknologi Malaysia, Malaysia {azriazmi@utm.my}

ABSTRACT
Multimedia Super Corridor (MSC) is a Malaysian government designated zone developed as a catalyst for the advancement of the information and communication technology (ICT). Currently, the locations of each MSC Malaysia status companies are represented in a list. Thus, it is time-consuming and inconvenient for users to go through a list of 3213 of MSC Malaysia companies. This study proposed to develop a webpage application to visualize the locations of MSC Malaysia companies in Klang Valley using dot distribution maps technique. Google Maps Javascript API is used to display the map in the webpage application. The data of the companies are collected and reorganized with additional information. The webpage application also utilizes markers as symbols to display the exact location of the MSC Malaysia companies. Besides, the tooltips feature is also included. Pan and zoom control features are provided by Google Maps Javascript API. From the result of system testing, it shows that the webpage application functions, work properly and the locations of MSC Malaysia companies are presented in a more understandable way in which ease for users to search for the exact location. By applying dot distribution maps technique, the users can easily view and identify the locations of MSC Malaysia companies in Klang Valley.

Keywords: Data Visualization, JavaScript, Multimedia Super Corridor, Dot Distribution Maps, Google Maps, API

1 INTRODUCTION
A structured and unstructured of large and complex data are referred as big data in which the traditional processing techniques are unable to operate on these data (Husin & Abdullah, 2019; Taylor-Sakyi, 2016). The characteristics of big data are 5V’s which consist of volume, velocity, variety, veracity and value (Hilda, Srimathi, & Bonthu, 2016). According to International Data Corporation (IDC), the collection of data is growing at an exponential rate (Olshannikova, Ometov, Koucheryavy, & Olsson, 2015). A massive volume of data is complex to be analyzed, therefore the data visualization techniques will be used to solve this issue. Data visualization is vital in visualizing the big data efficiently (Razak, Fauzi, Gining, & Jamaluddin, 2019; Sulaiman, Mohamad, Hanafi, & Wahab, 2018). The visualized data using the data visualization techniques are more understandable as they are presented in graphical or pictorial form (Sadiku, Shadare, Musa, & Akujobi, 2014).

In this study, MSC status companies are chosen as the dataset domain. MSC is a Malaysian government designated zone developed as an accelerator for the progression of the information and communication technology (ICT). MSC Malaysia companies are initiative projects by the fourth Malaysian Prime Minister, Tun Dr. Mahathir Mohamad. MSC Malaysia companies are classified into four different clusters of ICT focus area which are Creative Content Technologies (CCT), Global Business Services (GBS), Institutes of Higher Learning (IHLs) and Information Technology (InfoTech). The advancement of ICT is important to the society to transform Malaysia into a knowledge-rich economy by the year 2020 to achieve the Vision 2020 (Injau, 2011). MSC Malaysia is committed to nurture and guide the talent groups from primary and secondary students, undergraduates, fresh graduates and working professionals, digital maker, and premier IHL. This study is proposed to focus on fresh graduates from local universities. It is a job opportunity for the fresh graduates because MSC Malaysia is a great platform for the fresh graduates to pursue their career in digital technology to enhance their knowledge in ICT area.

Based on our study, we found that the user faces difficulties to identify the location and search the MSC Malaysia company on the current web site. It is due that the user can only search for the name of the MSC Malaysia company. Alternatively, the user can search for each company by scrolling up and down through the list. However, it is very time-consuming for the user to identify the location of the company because the user can only view the address once clicked on the company's name to view more details information of the company. Currently, the MSC Malaysia company has 2,954 active companies. The data stored for each company includes the company name, core activities, address and contact and these data are still presented as a text list. This presentation of data is inefficient to be analyzed.

Data visualization can be categorizing to five types which are temporal, multidimensional, hierarchical,
network and geospatial (Ward, Grinstein, & Keim, 2015) (Wilke, 2019). In this study, the geospatial or spatial data visualizations type is chosen because it is suitable category in order to display the real-life physical locations of the MSC Malaysia companies in Klang valley.

In geospatial category, there are several techniques that can be used to visualize the data such as cartogram, choropleth map, proportional symbol map, dot distribution map and dasymetric map. Cartogram is map that represents a mixture of statistical and geographical data in the geographical sections are adjusted in which the statistics of the population and income are presented (Nusrat, Alam, & Kobourov, 2018). Cartogram can be categorized into four different types which are contiguous cartograms, non-contiguous cartograms, dorling cartograms and rectangular cartograms. Choropleth map is used to visualize the location data such as countries, states or regions which varies by the variable attributes and the locations data are presented using circle, square, bar or a pictographic symbol. The advantage of using proportional symbol map technique is the attribute values of the measured location data and conceptual location data are represented together (Kunigami, Rezende, & Souza, 2014). Dot distribution map also known as density map uses dots in visualizing the location data. It is a technique used in representing the point-based data in which each point displayed on the map represents data with a known geo-location or a value aggregation (Polisciuc, Alves, & Machado, 2015). Dasymetric map is used in assigning population data to display the spatial data and its auxiliary data. The outcomes of dasymetric mapping can be validated by evaluating the relationship between the density of population and the auxiliary data using the statistical linear regression (Liu, Peng, Wu, Jiao, & Yu, 2018).

Different techniques of geospatial can be applied on the particular scenario and case study. Therefore, the dot distribution map is chosen since it is suitable with the objective of this study to visualize the MSC Malaysia companies in Klang valley.

II RESEARCH METHOD

This section will focus on the process of developing the data visualization for MSC Malaysia status companies. The waterfall model (Sulaiman & Ghani, 2011) was chosen as the software process model to represent the comprehensive study process as depicted in Figure 1. This model is used because it works well for smaller projects where the requirements are well understood.

In requirements definition phase, the system’s services, constraints and goal is established by consultation with target user and literature review from selected website, books and articles from journal and conference proceeding. The data of active MSC Malaysia companies were collected from Malaysia’s Open Data Portal (Jamil, n.d.). The data extracted from Malaysia’s Open Data Portal is in .xlsx format as shown in Table 1. The extracted data consists of 3213 active MSC Malaysia companies. It consists of 370 CCT cluster, 605 GBS cluster, 2119 InfoTech cluster and 119 IHLs and Incubators cluster.

Table 1. Sample Of Extracted Data From Malaysia’s Open Data Portal

Figure 2 shows the use case diagram (Sulaiman, 2003) representing the functional requirement of data visualization for MSC Malaysia companies (Sulaiman, Rosli, & Ghazali, 2009). The target user can be an internship student, fresh graduates or any qualified local or international people.
In system and software design phase, an overall system architecture model is proposed as depicted in Figure 3. The architecture partitions the requirements to either hardware or software systems (Gani & Sulaiman, 2005; Sulaiman, Tamizi, Shamsudin, & Azmi, 2020). All the data are stored inside the web server. The user need to use any browser in order to access the data visualization for MSC Malaysia companies. In implementation and unit testing phase, it is realized by using related software and hardware such as notepad++ as programming editor for Hypertext Markup Language (HTML), Cascading Style Sheets (CSS) and JavaScript, Google Chrome as web browser, XAMPP as a web server and laptop with Windows 10 operating system.

The array list is utilized to store the data companies which include the company’s name, cluster, contact, website, address, latitude, longitude and main activities. All the data will be sorted according to its cluster and city. Figure 6 shows a sample code excerpt using initCyberjaya() function.

```javascript
function initCyberjaya(){
    var locations8 = [['Xhai Studios Sdn Bhd', 2.922712, 101.661302, "Creative Content Technologies", "+6018-2862168 (T)", "info@xhaistudios.com", "xhaistudios.com", "i-Tech Tower, Jalan Impact, Cyber 6, 63000 Cyberjaya, Selangor", "Provision of pre-production, production and post production services for: 1) Xhai Mobile Games Development, Provision of implementation, maintenance and technical services related to the above services 2) Research, development and commercialization of the following solution: Gamer's Diary (GD), Provision of implementation, maintenance and technical services related to the above solution"],];
}
```

The marker is used as a symbol to indicate the exact location of MSC Malaysia companies. The coordinates for the marker to be placed will be retrieved from an array list named locations. Once the coordinates have been retrieved, the marker for each company will be generated. Figure 7 shows the excerpt code of the marker.

```javascript
for(var i=0; i<locations.length; i++){
    var loc = locations[i];
    var marker = new google.maps.Marker({
        position: {lat: loc[1], lng: loc[2]},
        map: map
    });
}
```

The discrete piece of information of the company which includes company name, address, contact, core activities and URL website is generated using tooltips. The tooltips will display when the user hover their mouse to the particular marker. Figure 8 shows the tooltips coding excerpt. As shown in Figure 8, the name of company is stored at loc[0], address stored at loc[5], contact that consists of telephone number is stored at loc[3] while email is stored at loc[4] in the array list. Core activities is stored at loc[6] and website is stored at loc[7] in the array list.
The drop-down menu features on the dot distribution map enable the users to display different datasets on the map. Figure 9 shows the drop-down menu for cluster which consists of CCL, GBS, InfoTech and IHLs and Incubators. This drop-down menu enables user to display the MSC Malaysia companies according to selected cluster. For example, if a user clicked on CCT, the application will only display the MSC Malaysia companies that belong to that cluster. This similar approach will be applied for cities in Klang Valley such as Petaling Jaya, Shah Alam, Cyberjaya, Subang Jaya, Bangsar and Kuala Lumpur. All the functions, variables and reserved words are tested in unit testing phase to confirm that the code is doing things right (Azmi & Ibrahim, 2011a).

In integration and system testing phase, the functional testing is done to ensure all the functional requirements as stated in the requirements definition phase were verified (Azmi & Ibrahim, 2011b). The last phase, operation and maintenance is ignored since it is beyond of our study scope due to the main objective is to develop the data visualization using dot distribution map (Sommerville, 2015).

### III RESULTS AND DISCUSSION

The data visualization for selected 80 MSC Malaysia companies are presented in the graphical user interface as illustrated in Figure 10. By a single click on map button, the web page will display all the selected companies from all clusters. It will ease the target user to identify the location and distribution of the companies.

#### A. Map based on cluster

Figure 11 illustrates the location of MSC Malaysia companies based on CCT cluster. The user can select an appropriate cluster by using the drop-down menu. The user can select another cluster such as GBS, InfoTech and IHLs & Incubator as well.

#### B. Map based on city

Figure 12 shows the drop-down menu for selected companies in particular city in Klang Valley such as Petaling Jaya, Shah Alam, Cyberjaya, Subang Jaya, Bangsar and Kuala Lumpur.

#### C. Tooltips

Figure 13 shows the tooltips that will display the information such as company name, cluster name, contact, address, core activities and website URL when the user mouseover to the particular marker.
The list of MSC Malaysia companies previously were shown in MSC Malaysia directory on the Malaysia Digital Economy Corporation Sdn Bhd (MDeC)’s official website (Malaysia Digital Economy Corporation, n.d.). However, the current official website no longer displays the list of active MSC companies. It only shows the number of active companies which are 2954 as of May 2019. The data collected from the Malaysia’s Open Data Portal listed an active MSC companies as of year 2017 consists of 3212 companies. Thus, it is difficult to determine whether the MSC companies in 2017 are still active in 2019. In addition, the current data do not provide the complete company profile such as contact number, email, address and coordinate of the company’s location. Therefore, the data need to be updated when the new current active MSC Malaysia companies are obtained.

![Figure 13. Tooltips for Each Companies](image)

### IV CONCLUSION

This paper presents the development of data visualization for MSC Malaysia status companies using dot distribution maps technique. The target users for this study is for fresh graduates that are looking for job or any qualified local or international people that would like to pursue a career whether for full time, part time or an internship with MSC Malaysia companies located especially in Klang Valley. This study covers all clusters of the MSC companies which are Creative Content Technologies (CCT), Global Business Services (GBS), Information Technology (InfoTech) and Institutes of Higher Learning (IHLs) and Incubators. This study can be further improved by adding searching feature that will enable the users to search the nearest location of MSC Malaysia companies based on city or state.

### ACKNOWLEDGMENT

The authors would like to thank Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM) for sponsoring this paper.

### REFERENCES


