

Leveraging the Power of Social Media and Data Visualisation

A thesis submitted to the
University of Dublin, Trinity College
for the degree of
Masters of Science

Lin Yue
Computer Science (Network and Distributed System)
School of computer science and statistics
Trinity College, Dublin,
yuel@tcd.ie
Supervised by Dr. Melika Sah & John McAuley

Declaration

I, the undersigned, declare that this work has not been previously submitted as an exercise for a degree at this or any other University, and that, unless otherwise stated, it is entirely my own work.

Lin Yue
August 2013

Permission to Lend or Copy

I, the undersigned, agree to deposit this dissertation in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Lin Yue
August 2013

Acknowledgement

This project lasts 10 months which is the longest one in my previous experience. First of all I really appreciate my two supervisors Dr. Melike Sah and John McAuley, without their help, I cannot achieve the goals of this project. Furthermore I have to thank my mentors in Ava Trade. They give me a lot of free time to finish my dissertation. My friend Jinwu Li who is the writer of the other relevant thesis also gives me lots of help during the whole process. In the end, I would like to thank myself. You made it!

Abstract

Leveraging power of social media and data visualization is a project to implement a novel visual search and interaction interface for LinkedIn Ireland using semantic web and data visualization technologies.

This project is based on the data which extracted from the user generated content of LinkedIn and produce three scenarios for different user statuses. They are government staffs, HRs, job seekers and college perspective students. This project aims to provide useful insights about professions for employment rate exploration, career path selection, recruiting investigation. To achieve these objectives, this project is able to support complex queries to retrieve, analyze the data and then offer graphical user interaction interfaces, which can support interpretation and exploration of large, complex user generated data.

Table of Contents

1.	Introduction	1
1.1.	Research Questions	2
1.2.	Research Objectives.....	2
1.3.	Contribution.....	2
1.4.	Thesis Outline.....	2
2.	State of the art	4
2.1.	Data visualization development	4
2.2.	Implication for this project	5
2.3.	Relevant technical knowledge:	5
2.3.1.	Classification of data types	5
2.3.2.	Classification of data visualization technique	7
2.3.3.	Major Challenges for Visualization	12
2.4.	A categorization of visualization approaches	13
2.4.1.	Discussion.....	16
2.5.	Implementation Frameworks	20
2.6.	Chapter Summary	20
3.	Design.....	22
3.1.	Methodology.....	22
3.1.1.	Scenario Developments	22
3.2.	Data and Schema	29
3.3.	Data visualization and data interaction	31
3.3.1.	Data visualization and interaction design Sketches.....	31
3.3.2.	Data visualization and interaction design-Version1:	33
3.3.3.	Data visualization and interaction design-Version 2	34
3.3.4.	Data visualization and interaction design-Version3	35
3.4.	Chapter Summary	35
4.	Implementation	37
4.1.	Interface implementation	37
4.2.	Complex query implementation	37

4.3.	Visualization Implementation:.....	39
4.4.	User interaction implementation:	40
4.5.	Queries test and performance test.....	40
4.6.	Chapter summary.....	41
5.	Evaluation	42
5.1.	Comparison with LinkedIn	42
5.1.1.	Main functions comparison LinkedIn VS My project	42
5.1.2.	User interaction comparison LinkedIn VS My project	43
5.2.	User study	46
5.2.1.	Hypotheses.....	46
5.2.2.	Evaluation setup.....	46
5.2.3.	Results.....	47
5.3.	Chapter summary.....	56
6.	Conclusion.....	57
6.1.	Future Work.....	57
6.1.1.	Automatically generate relevant analysis respect to requests.....	57
6.1.2.	Personalize design for individual users.....	58
6.1.3.	Refine the data structure and add time line to each scenario in order to check the trend of each area involved.	58
6.2.	Closing Remarks	58
7.	Reference:.....	59
8.	Appendix	63
8.1.	Appendix 1:	63
8.2.	Appendix 2:	65
8.3.	Appendix 3:	67
8.4.	Appendix 4:	69
8.5.	Appendix 5:	69
8.6.	Appendix 6:	70
8.7.	Appendix 7:	71

Table of Figures

Figure 1:Nominal data 1	6
Figure 2:Nominal data 2	6
Figure 3:Ordinal data 1	6
Figure 4:Ordinal data 2	6
Figure 5: Interval data	7
Figure 6:Ratio data	7
Figure 7: 2D Scatter Plot	8
Figure 8: 3D Scatter Plot	8
Figure 9: 2D Line Graph	9
Figure 10: 3D Line Graph	9
Figure 11: Pie Chart	9
Figure 12: Bar Chart	10
Figure 15: Word Cloud	12
Figure 16 : Methodology.....	22
Figure 17: Knowledge Model for LinkedIn public profile and company profile.....	30
Figure 18: Government scenario sketches.....	31
Figure 19: HR scenario sketches	32
Figure 20: Job seeker and college perspective students scenario sketches	32
Figure 21: Layout version1	33
Figure 22: Layout version2	34
Figure 23: Layout version3	35
Figure 24: Design architecture	36
Figure 25: A basic query example.	38
Figure 26: A mid-complex query example	38
Figure 27: A complex query example.....	39
Figure 28: A super complex query example.....	39
Figure 29::LinkedIn visualization page	44
Figure 30: My project visualization page	44
Figure 31: LinkedIn preference data searching page	45
Figure 32: My project visualization page	45
Figure 33: User satisfaction questionnaires government score	49
Figure 34: User satisfaction questionnaires HR score.....	51
Figure 35: User satisfaction questionnaires job seeker and college perspective students score	52
Figure 36: SUS usability comparison	54
Figure 37: SUS Scores	55
Figure 38: SUS Score standard	55
Figure 39: Sample code of regular expression	69
Figure 40: Sample code of Ajax.....	69
Figure 41: Sample code of case classification	70
Figure 42: Sample code of data visualization.....	71

Table of Tables

Table 1: Illustrate the different data types.....	6
Table 2: Illustrate the different data visualization techniques.....	8
Table 3: Illustrate a categorization of visualization approaches	14
Table 4: Implementation framework comparison	20
Table 5: Scenario questions gathering	23
Table 6: Government Scenario data requirements and visualization	24
Table 7: HR Scenario data requirements and visualization.....	26
Table 8: Job seeker and college perspective student Scenario.....	28
Table 9: Main functions comparison between LinkedIn and my project.....	42
Table 11: User interface comparison between LinkedIn and my project	44
Table 12: Task-based question answer sheet summary	47
Table 13: User satisfaction questionnaires government score	48
Table 14: User satisfaction questionnaires HR score	50
Table 15: User satisfaction questionnaires	52
Table 16: SUS questionnaire score for three scenarios	54
Table 17: SUS level	56

1.Introduction

Leveraging power of social media and data visualization is a project which utilizes the large user generated content from LinkedIn by using technologies of semantic web and data visualization to produce a new power to public.

This project is inspired from the Career Galaxy project [1]. This paper illustrates the situation that it is hard for one user to randomly retrieve and analysis all of profiles with similar background to get more idea about future by using the resource of LinkedIn. In addition, LinkedIn does not have income information and visualized information which is important for career planning. The thesis finds a good way to address the issues by using ontologies to build the standard relationship between different entities and extract data from LinkedIn API, then by using different data visualization to show the trend of professions situation in graphical ways is.

However, this project only focuses on one user type (students) and lacks complex queries which can only show simple graphics. Thus, the objective of my project is to provide three different user interfaces which allow complex queries to explore the facts for three specific scenarios:

- 1. Government - government personnel who are interested in understanding industry developments and employment statistics in Ireland.**
- 2. Human resource - HR personnel who are looking for to recruit new personnel**
- 3. Job seekers and college perspective students - who are seeking employment.**

There is no doubt that the big data era is here. With the development of web-based applications, the growing numbers of users would like to share information on social media. LinkedIn has a great number of users and being used by the people all around the world. [2] The large volumes of user generated data have been accumulated capturing by prosperous applications on LinkedIn.com. [3]

LinkedIn and other social network's stable user groups and bright prospect make people interested in one area, which is finding out how to utility the ability to make timely decision based on available data which is very helpful to business success and career path. The plan is all based on the large user generated data and at the same time, almost every field of study eventually has the same challenge of big data problem. [4]

As a consequence, it is not difficult to imagine that users cannot find out the result in such big volume data in very short time, let alone navigate the collection and perform convincing

analysis. However, data visualization is a useful tool that can enable the analysis of increasingly large datasets.

1.1. Research Questions

How can a visual search and interaction interface be implemented for LinkedIn Ireland considering following the following user profiles:

- Government personal
- HR personal
- Job seeker and college perspective students personal

1.2. Research Objectives

- Review the state of the art in information and data visualisation and identify the affordances and constrains of each approach
- Design a visual search and interaction interface that can be used to address each of the scenarios identified in the introduction
- Evaluate this interface using task-based questions answer sheet, user satisfaction questionnaire and SUS (system usability scale) for each of the identified scenarios.

1.3. Contribution

A novel visual search and interaction interface for LinkedIn Ireland using semantic web and data visualization technologies that can support complex queries and different visualization of statistical information for three different scenarios, which hasn't been done before.

1.4. Thesis Outline

Chapter 2 reviews the state of the art in data visualization.

Chapter 3 outlines the design process. We adopted an iterative participative design process in which several early sketches were refined.

Chapter 4 presents the system requirement and technology chooses for implementation.

Chapter 5 describes the evaluation, presents our results and our analysis.

Chapter6 describes the achievements, summarizes the contribution of the thesis and discusses future work.

2.State of the art

This chapter reviews the state of the art in data visualization. More specifically, we look at data visualization from the following perspectives. The data visualization development gives me a clear way to know the brief history of this area and also shows the strong motivation of my project. In order to implement a user satisfied interface, I have to make sure which is the best match between different types of data and a plenty of visualization technologies. Thus a lot of work went to the related work part. More than 20 papers I read illustrate a good way to implement my project and lead me to find out the most suitable technologies for my project. At the end, I find out which are the biggest challenges of data visualization. They are very important notifications during the whole design and implementation process. Section 2.1 will discuss data visualization development. Section 2.2 will talk about the implication for this project. Section 2.3 provides few tables to compare the relevant technical knowledge. Section 2.4 is the related work part about the categorization of visualization approaches. Section 2.5 tells us the challenge of data visualization. In the end, section 2.6 summarizes the whole chapter.

2.1. Data visualization development

“Data visualizations are initially design by a human, but are then drawn algorithmically with graphing, charting, or diagramming software”. [5] However, large-scale data is too complex to understand. Therefore, interpretations of large-scale data seem very challenging [6]. One of the main goals of data visualization is to improve the user understanding of data by showing integrated information as a graphical representation, which can help and support a better comprehension, memory and decision making. This kind of data idea had been suggested hundreds of years ago, in 1860s [7], the statistical development of data visualization enabled two dimensional implementations. The visualizations show the data as graphs. This traditional concept of data visualization kept going until the computer revolution. In recent years, interactive data visualization provides a good way to response to new data quickly, which means through new algorithms and techniques, increasingly sophisticated visualizations are being investigated. The best solution of hiding the complexity of information to users is to map the relationship between users, concept or words before using visualization tools. [8]

Today, the data visualization tools play a good role of a bridge to make a connection between the large volume data and human mind. The purpose of data visualization tool is to get the information from the normal database and show them into an explicit and concise way to the users. There are numerous kinds of graphics to represent the statistical data in different ways, for example, the curve graph based on time will show the development trend of the matters, the pie chart based will illustrate the ratio of individual thing in a whole event. [9]

2.2. Implication for this project

However, if the visualized data only come from database without respect to any particular requirements, the goal of data visualization to provide a better way of helping users make decisions cannot be reached. The interactions between users and data are the most significant area. In this part, the extracted data which are useful shown in terms of the specific project in the database will be queried with database query statements which based on the requirements of users. The interactive events will be triggered by different kinds of users and different kinds of requirements. By using the knowledge of semantic web, and through the understanding of information architecture, the interactions should be more real to express the linkages between facts. [10]

The overall achievements indicate that data visualization is useful for data analysis and comparison. [11] When this concept combines the advantage of semantic web, the potential for improving the efficacy of decision making in terms of user generated content is increasingly impressive.

2.3. Relevant technical knowledge:

In this section, we will discuss the classifications of data types and visualization techniques in order to find out the appropriate techniques to each type of data. From the discussion, the optimal match between data types and visualization techniques, it is relatively easy to suggest the best solution for each type of data. Finally, the readable and persuasion of the graphics which we will provide in the project will be massively enhanced.

2.3.1. Classification of data types

Generally speaking, there are two basic types of visualization: Information Visualization and Data Visualization. Information Visualization represents the abstract data to improve the cognition of users. [12] The information comes from text and graphics. Thus, information visualization can be widely used in a range of different areas., Data visualization has a similar purpose to Information visualization, however, data visualization is used to represent data that is abstracted into some schematic form. Data Visualization abstract useful data from all above information and finally represents the data in graphic way.

According to diverse professional fields, there are different ways of classifying data. With respect to data visualization, the most appropriate criteria typically refer to the definition in statistics. This type of classification came from Stanley Smith Stevens, [9] who proposed that Nominal data, Ordinal data, Interval data and Ratio data are four basic types of data as shown in Table 1.

Table 1: Illustrate the different data types

Data Type	Definition	Examples														
1. Nominal data	Can be counted, but it cannot be measured. No level between nominal data. They are logically equal. [13]	Name, Professions, Gender <i>(Figure1&Figure2)</i>														
<p>Figure 1: Nominal data 1 [14]</p>	<p>Nominal Data</p> <table border="1"> <tr> <td>Point</td> <td>airport ✕</td> <td>town ●</td> <td>mine ⚒</td> <td>capital ★</td> </tr> <tr> <td>Line</td> <td>river —</td> <td>road —</td> <td>boundary - - -</td> <td>pipeline</td> </tr> <tr> <td>Area</td> <td>orchard 🌿</td> <td>desert 🏜</td> <td>forest 🌲</td> <td>water 💧</td> </tr> </table> <p>Figure 2: Nominal data2 [15]</p>	Point	airport ✕	town ●	mine ⚒	capital ★	Line	river —	road —	boundary - - -	pipeline	Area	orchard 🌿	desert 🏜	forest 🌲	water 💧
Point	airport ✕	town ●	mine ⚒	capital ★												
Line	river —	road —	boundary - - -	pipeline												
Area	orchard 🌿	desert 🏜	forest 🌲	water 💧												
2. Ordinal data	The data have natural order and it can be count and order but no measure. [16]	Strong dislike, dislike, neutral, like, strong like <i>(Figure3&Figure4)</i>														
<p>Figure 3: Ordinal data 1 [17]</p>	<table border="1"> <tr> <td>Like very much</td> <td>1</td> </tr> <tr> <td>Like</td> <td>2</td> </tr> <tr> <td>Like slightly</td> <td>3</td> </tr> <tr> <td>Neither like nor dislike</td> <td>4</td> </tr> <tr> <td>Dislike slightly</td> <td>5</td> </tr> <tr> <td>Dislike</td> <td>6</td> </tr> <tr> <td>Dislike very much</td> <td>7</td> </tr> </table> <p>Figure 4: Ordinal data 2 [18]</p>	Like very much	1	Like	2	Like slightly	3	Neither like nor dislike	4	Dislike slightly	5	Dislike	6	Dislike very much	7	
Like very much	1															
Like	2															
Like slightly	3															
Neither like nor dislike	4															
Dislike slightly	5															
Dislike	6															
Dislike very much	7															
3. Interval data	There is a distance (space or time) between two neighbor units. The zero point is arbitrary. [19]	Changes of height from 1989-2013 <i>(Figure5)</i>														

Total Medicare Private Health Plan Enrollment, 2000-2010

Enrollment in millions:

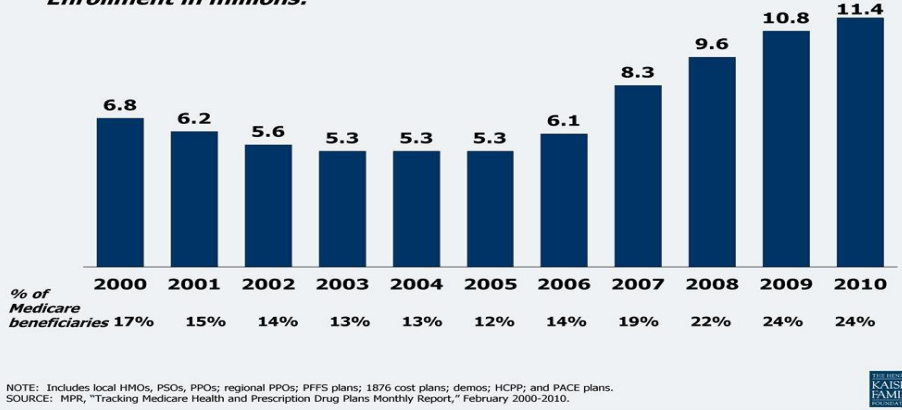


Figure 5: Interval data [19]

4. Ratio data

There is a measure word to describe the data which can be used to measure the data according to a unit data of the same type. [20]

Length, Energy, Electric charge
(Figure6)

US women weight vs. height

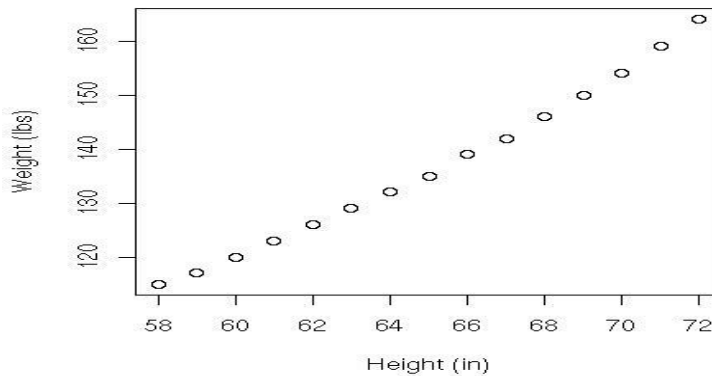


Figure 6: Ratio data [21]

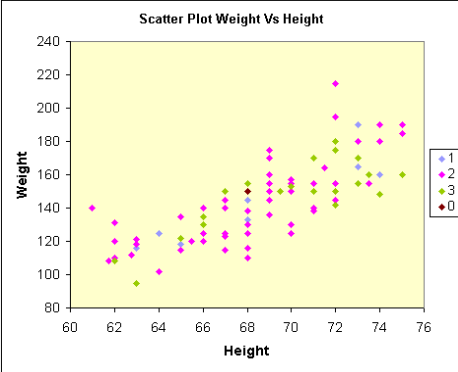
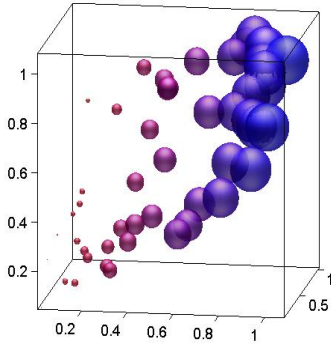
2.3.2. Classification of data visualization technique

With the progress of data visualization techniques, the data can be shown not only on the two-dimension graphics, but also extend to three-dimension graphics. It is enough to use two-dimension graphics to explain the content of data. It is a simplest way to give the user a straightforward understanding about the complex data. What is more, two-dimension graphics are loading faster than three-dimension ones. Thus two-dimension graphics are widely used in data visualization. Three-dimension graphics are generally used in Medical Science and Geography Science visualization. In additional, with respect to illustrate the

variation of data, the animation will be used into data visualization [22].

However, no matter the how the format of visualization technologies change, the basic frameworks of representing data are fixed. These three ways all aims to enhance the user understanding of comprehension. Ignoring the phenomenal difference, according to this, there are Scatter Plot, Line Graph, Pie Chart, Bar Chart, Tree Diagram, Map and Word Cloud, 7 distinct and basic data visualization techniques responsible illustrate majority types of techniques ass explained in Table2

Table 2: Illustrate the different data visualization techniques

Data visualization techniques	Advantages	Disadvantages
<p>1. Scatter Plot and 3D Scatter Plot</p>	<p>It can show the plots more efficient if there are not too many individual points need showing on the graph and it can show the relationship among individual points detailed.[23]</p>	<p>It cannot be used if thereis no common feature among points which means the points cannot be shown on one graph. <i>(Figure7&Figure8)</i></p>
<div style="text-align: center;">  <p>Figure 7: 2D Scatter Plot [24]</p> </div>	<div style="text-align: center;">  <p>Figure 8: 3D Scatter Plot [25]</p> </div>	
<p>2. Line Graph and 3D Line Graph</p>	<p>It can illustrate time series data effectively with one event or multi events.[25]</p>	<p>It cannot be used to describe the individual point. <i>(Figure9&Figure10)</i></p>

Time Needed to Locate Suitable Library Materials

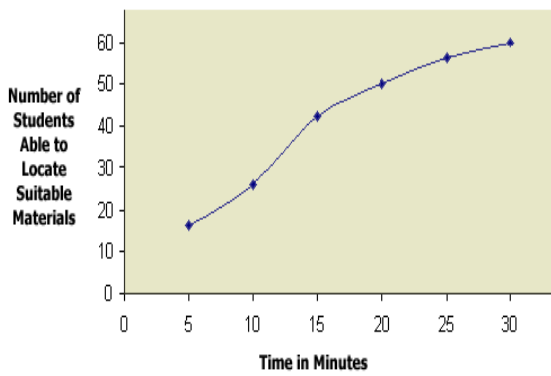


Figure 9: 2D Line Graph [27]

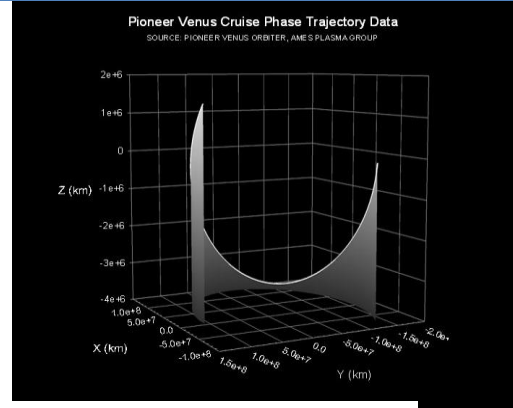


Figure 10: 3D Line Graph [28]

3. Pie Chart

It can be used to show the classification data and the area of each part can easily tell reads the truth of each category.[29]

It cannot be used to describe the interaction between two objects. It is not useful for ordinal data.

(Figure11)

Pie Chart of Countries

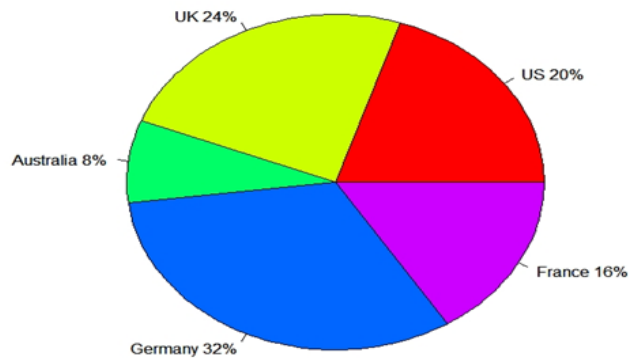


Figure 11: Pie Chart [30]

4. Bar Chart

It can be used to show the classification data and the length of the bar can easily tell reads the truth of each category. [31]

It cannot be used to describe the interaction between two objects.

(Figure12)

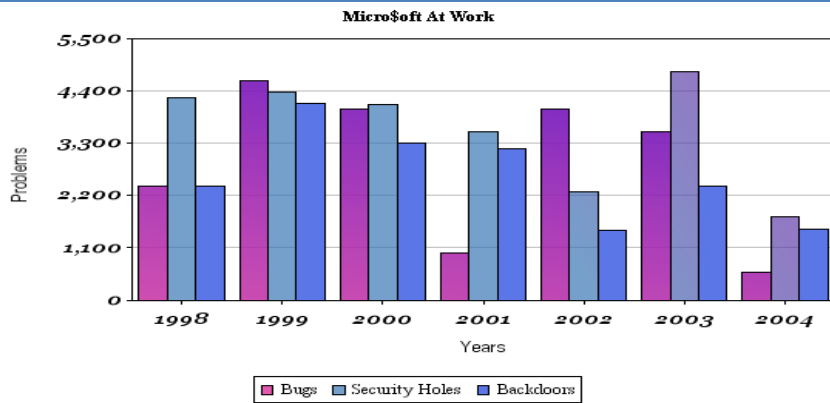


Figure 12: Bar Chart [32]

5. Tree Diagram

It can be used to illustrate the relationship between two or more nodes such as inheritance, aggregation, and causal connection.[33]

It can only be used to describe a fixed relationship between nodes and the variation of relationship cannot be shown by Tree Diagram.

(Figure13)

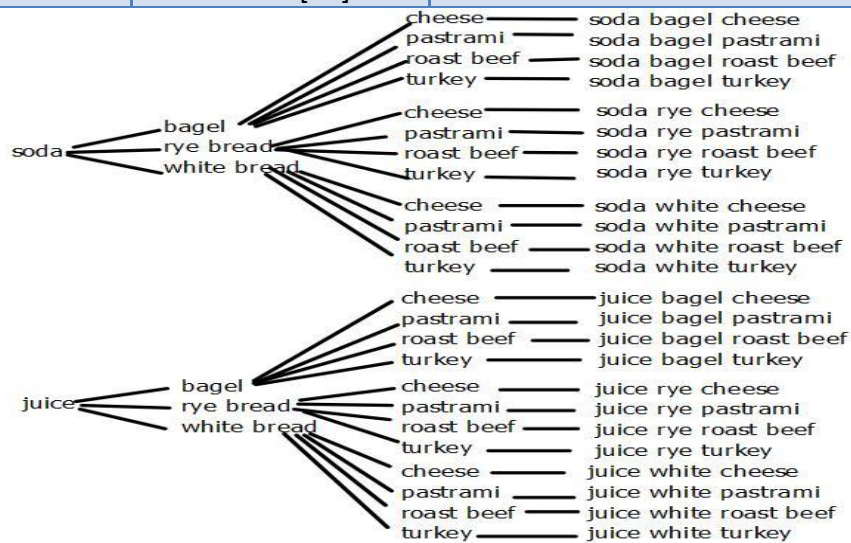


Figure 13: Tree Diagram [34]

6. Map

The geographical areas can be communicated with using different colours and the map can combine the information and the location easily. It is a

It can only be used to represent data related to geography.

(Figure14)

comprehensive way to represent more than one kind of data to users.[35]

Cause 7 Age-adjusted Death Rates, 1988–1992

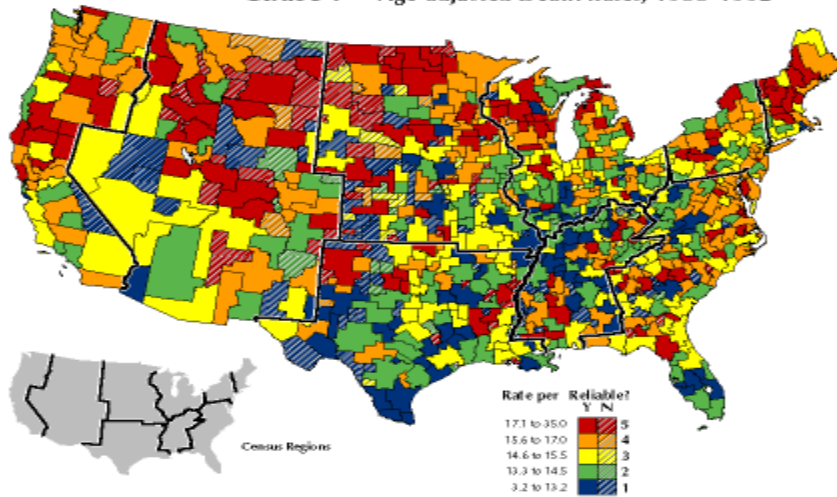


Figure 14: Map [36]

7. Word Cloud

It can be used to show the words frequency of the appearance in the source data. The size of the words can illustrate the frequency is high or low and also from the word cloud generator the users can choose the match of colours they like.[37]

It is just reflect the words frequency of appearance. It cannot be used to show the interactions among data and the truth of statistical data we mentioned above. It is questionable how useful this representation is for data analytics.

(Figure15)

constrained by the limits of representation and flexibility of users' cognition. The evaluation of interactivity only is based on the feedback from users, thus it is difficult to improve the interactivity of data visualization system can only after it is released to users. [41]

1.1.1.3 Scalability

Because of the increasingly large datasets, and the difficulty of identifying useful information in these datasets, scalability is a challenge for data visualization. [42]

1.1.1.4 Evaluation

The above issues indicate that data visualization implementation is a long and complex process. The challenges have shown the complexity to produce a high-performance data visualization product. Apparently, we should pay more attention to the challenge of evaluation. Only in this way can we achieve an accurate result, develop a deeper understanding of how visualization can be used to address specific problems and make more useful suggestions to improve the products and services.

In addition, according to the thesis, to evaluate a visualization tool which has no baseline system to compare should be the most difficult part to complete.

Evaluating data visualization involves:

- Selecting the evaluation questions appropriately;
- Selecting the correct participants who have knowledge of the domain
- Measuring for Usability (using standardized tools) and Usefulness (measuring the time it takes to complete a task).

For instance, a whole product are consisted by different parts, in other words, each part should be evaluated individually. Furthermore, the testers should be recruited based on different parts of product. Otherwise the result will not be convincing and leads a sequence of problems in practice. [43]

On the whole, an excellent evaluation will be very helpful to refine the data visualization product and provide a direction for enhancing the user experience.

2.4. A categorization of visualization approaches

In this section, I will discuss more than 20 papers which can represent the updated development status of data visualization. From comparing the complex types of visualization techniques and data types to confirm the result I got from the previous section and also, to mark the comments for each specific data type in order to make it more convenient to be visualized, such as the small limitations and requirements for expressing data I have not been aware of from the previous work. (The discussion results are showing in advantage and disadvantage parts after the table.)

What is more, the tools also give me a good direction to search the best visualization tools for the design and implementation of our system. Table 3 summaries all of the papers that I have reviewed for this thesis.

Table 3: Illustrate a categorization of visualization approaches

Instance number	Paper	Author	Type of visualization techniques	Data type	Tools
1	Automated multi-modality image registration based on information theory[44]	Collignon, A Maes, F Delaere, D Vandermeulen, D Suetens, P Marchal, G Leuven, K U Gasthuisberg, U Z	Scatter Plot and 3D Scatter Plot	Ratio data	Not mention
2	Rolling the dice: multidimensional visual exploration using scatterplot matrix navigation. [45]	Elmqvist, Niklas Dragicevic, Pierre Fekete, Jean-Daniel	Scatter Plot and 3D Scatter Plot	Ordinal data(dice face probability)	Not mention
3	Visualization Methods for Personal Photo Collections : Browsing and Searching in the PhotoFinder[46]	Kang, Hyunmo Shneiderman, Ben	Scatter Plot and 3D Scatter Plot	Nominal data	Not mention
4	AMS Mathematics Subject Classification (2000): 05C50 Key Words: graph theory, graph spectra, line graph, signless Laplacian 1. [47]	Laplacians, Signless Graphs, Line	Line Graph and 3D Line Graph	Ordinal data(question derivation process) and ratio data	Not mention
5	Applications of Graph Visualization[44]	North, Stephen C Hill, Murray	Line Graph and 3D Line Graph	Ordinal data(sequence of event)	Dotty and layout tools: Dotmakes

6	Many Eyes : A Site for Visualization at Internet Scale[48]	Viégas, Fernanda B Wattenberg, Martin Ham, Frank Van Kriss, Jesse Mckeon, Matt	Line Graph and 3D Line Graph	Ordinal data(Axis labels:t Value:N+)	“Many Eyes”
7	Many Eyes : A Site for Visualization at Internet Scale[49]	Viégas, Fernanda B Wattenberg, Martin Ham, Frank Van Kriss, Jesse Mckeon, Matt	Pie Chart	Ordinal data(Axis labels:t Value:N+)	“Many Eyes”
8	Visualization of Power System Data[50]	Overbye, Thomas J Weber, Jamie D	Pie Chart	Nominal data	EMS
9	Protovis: a graphical toolkit for visualization.[51]	Bostock, Michael Heer, Jeffrey	Pie Chart	Nominal data	Protovis
10	Protovis: a graphical toolkit for visualization.[52]	Bostock, Michael Heer, Jeffrey	Bar chart	Nominal data	Protovis
11	Visualization of Test Information to Assist Fault Localization Georgia Institute of Technology[53]	Jones, James A Harrold, Mary Jean Stasko, John	Bar Chart	Nominal data and Ratio data(probability)	Eichel et al
12	P. J. D. andD.[54]	Momke, J	Bar Chart	Nominal data and Interval data	Not mention
13	Thread Arcs: an email thread visualization[55]	Kerr, B.	Tree Diagram	Interval data and nominal data	Not mention
14	EulerView: a non-hierarchical visualization component[56]	Chiara, Rosario De Fish, Andrew	Tree Diagram	Nominal data(inheritance)	Euler View

15	A space-optimized tree visualization[52]	Nguyen, Quang Vinh	Tree Diagram[53]	Nominal data	Not mention
16	Visualization of Power System Data[57]	Overbye, Thomas J Weber, Jamie D	Map	Nominal data	EMS
17	Visualization and Data Mining of Pareto Solutions Using Self-Organizing Map[58]	Obayashi, Shigeru Sasaki, Daisuke	Map	Nominal data and Ratio data	SOM
18	P. J. D. andD.[59]	Momke, J	Map	Nominal data and Interval data	Not mention
19	Tag-Cloud Drawing Algorithms for Cloud Visualization[60]	Kaser, Owen John, Saint	Word Cloud	Nominal data and interval data	EDA
20	Tag Clouds: Data Analysis Tool or Social Signaller?[61]	Hearst, Marti a. Rosner, Daniela	Word Cloud	Nominal data and interval data	Web2.0 relevant technology
21	SparkClouds: visualizing trends in tag clouds.[62]	Lee, Bongshin Riche, Nathalie Henry Karlson, Amy K Carpendale, Sheelash	Word Cloud	Nominal data and interval data	SparkC loud

2.4.1. Discussion

2.4.1.1 Scatter Plot and 3D Scatter Plot with Ratio data

The advantage of this case is the scatter plot illustrates the RMI situation and other kinds of medical data. This kind of data is separated in different ways. In the scatter plot, the individual data can be described obviously and the relative locations also express the important information at the same time. The disadvantage of this case is sometimes, the data have a distinct trend which means the ratio data can be illustrated by line graph. We do not need to illustrate each point in detail, the special one is the more significant to point it out. [44]

2.4.1.2 Scatter Plot and 3D Scatter Plot with Ordinal data

The advantage of this case is the scatter plot, especially the 3D scatter Plot shows the dice face probability of the ordinal data. It is easily to understand the percentages situation of each probability. There is no need to catch the individual figure. The general state of data is what the users want in this case. The disadvantage of this case is the general state of the data distribution can get easily, but the exact numbers can not been seen clearly on the graph. [45]

2.4.1.3 Scatter Plot and 3D Scatter Plot with Nominal data

The merit of this case is there is a big difference between nominal data and other kinds of data, no relationship between nominal data. The authors just put the individual kind of data into one table. At this time, the scatter plot and 3D scatter plot will be used to describe them individually. The challenge for scatter plot is to classify the data from one table into different graph. When the developer use this kind of graph to visualize the nominal data, it is hard to put data into one unit which means the data are able to be express on one graph. [46]

2.4.1.4 Line Graph and 3D Line Graph with Ordinal data, Ratio data

Line graph is easy to show the changes trend of the data. In this paper, line graph can describe the individual data which are generated by the formula. It is a good way to translate the formula into graph. The best choice to illustrate ordinal data is using line graph, because the abscissa axis of the line graph can describe the degree of each kind of data. However, in terms of the ratio data, the abscissa axis cannot express the appropriate relationship between different data unit. [47]

2.4.1.5 Line Graph and 3D Line Graph with Ordinal data

In this paper, the author uses line graph to describe the development process of the object. It is a little bit like the tree diagram. However, the only difference between line graph and tree diagram is the relationship between two individual points is very general. In other words, only if exist there a linkage between two points, the line graph could be used to describe that clearly. On the other hand, when the data is time-related or degree related, line graph is not the advantage of line graph is obvious. However, if the feature is not such distinct, the line graph can be easily instead by bar chart. [48]

2.4.1.6 Line Graph and 3D Line Graph with Ordinal data

The advantage of line graph in this paper is also the disadvantage. Because the author puts bar chart, line graph and stack graph into one unit and the matched data schema is totally same, which means, in this paper the impacts of bar chart and line graph are the same. [49]

2.4.1.7 Pie chart with Ordinal data

From the pie chart, the percentage of each unit of data can be seen obviously. It is a better

way to compare the share of each data unit. In other words, the pie chart is designed to show relative totals, thus the individual data can be aggregated in a session. That is the biggest data between pie chart and scatter plot. In addition, the relative relationship of each data unit cannot be expressed clearly. If the data is time-related or degree-related data, there is no axis in the graph, which means the attributes of each data unit cannot be described by the pie chart. It just illustrates the relative share of each data unit. [49]

2.4.1.8 Pie chart with Nominal data

In this paper, the author puts more than one pie charts into a map. The author uses the map to describe the location information and use the pie chart to express the relative share information in each specific location. It is very clear to utilize pie charts to show the percentage of each data unit involved in an entire event. It is a little bit complex when users find out the key figures in the map because the main part of this paper used 3D pie chart to show the data. This is the main reason which makes the graphs are disorderly and unsystematic. [50]

2.4.1.9 Pie chart with Nominal data

The author uses pie chart to show the nominate data which there is no relationship between any data unit. However, the percentage of each data unit is the most significant point to illustrate in the graph. According to the above research, pie chart is the most appropriate techniques to show the nominate data. There is a significant situation that if the data cannot be aggregated which means the data should be expressed individually, in this case, the pie chart cannot be used to describe the nominate data. [51]

2.4.1.10 Bar chart with Nominal data

Compared the feature of bar chart and pie chart, the nominate data can be shown in both of the charts. However the bar charts are better in describing the time-related data and degree-related data. Other advantages are as same as pie charts. As same as the situation in the disadvantage of pie chart in reference [47], the individual data cannot be shown by bar charts. [51]

2.4.1.11 Bar chart with Nominate data, Ratio data

In this paper, the ratio data are based on the relationship about version updated process. The horizontal axis and vertical axis are described the linkage and percentage of each data unit. The comparison and the share of data unit can be illustrated easily. [52]

2.4.1.12 Bar chart with Nominate data, Interval data

The author use bar chart to illustrate the time-dependent data which means the interval data. Form this way the bar chart can display the changes in data over time. Different colors in the bar chart can describe the increases of each time step. The solution also leads a new way to display the ordinal data in the bar charts. Through different colors to display the various degree of the objects meet the requirements of illustrate the relationship between

different data unit at the same time. Thus this disadvantage can be fixed. [53]

2.4.1.13 Tree diagram with Interval data, Nominal data

The tree diagram shows the generation updating process of “thread arcs” in the paper. Because of the features of tree diagram, the rank correlation of each data unit can be shown in the graph. The evolutionary process is better to be translated by tree diagram. On the other way, if the relationship between two data units are not quite clear, the lines which show the linkage will be overlapped. That is the point the developer need to concern especially in showing the effects of a larger thread. [54]

2.4.1.14 Tree diagram with Nominal data

This paper adopts the concept of tree diagram and applies it into the file system classification. By using color and different types of arrows show the relative relationship between the data units. EulerView is designed show the hierarchal data. In other words, the limitations of EulerView are obvious. EulerView can only show the key words from the change process, the specific figures are hard to be described in the tree diagram. [55]

2.4.1.15 Tree diagram with Nominal data

There are four different views of tree diagram: the classical hierarchical view, the h-tree layout, the radial view and the balloon view. These four views can easily display the siblings of sub-trees, the classical hierarchical relationship, the circular relationship and the hyperbolic relationship. The complexity of the data model is not a question to a tree diagram. If there is only one tree diagram to show the whole process or the architecture of the data, the space of tree diagram should be too large. The solution to this drawback is using animation to show the relationships step by step or section by section.[56]

2.4.1.16 Map with Nominal data, Ratio data

The author uses the map with 3 dimension techniques to solve the problems which the bar chart are not able to show the data in the case of objectives are changing in an irregular change. In the paper the map also express how the pressure value distributed on different part of an irregular object. This feature is irreplaceable. [57][58]

2.4.1.17 Word Cloud with Nominal data, Interval data

Word cloud in these three papers is used to describe the keyword that is associated with the users’ behaviors and information. That is very interesting and the interaction between the users and system can be enhanced obviously. By using word cloud, the users’ behaviors can be shown individually and no matter the nominal data or the interval data which are based on changes during the time period will be expressed in a novel way. [59][60][61]

2.5. Implementation Frameworks

During the whole process of relate work study, I found that my project is very similar to the paper who is using Many eyes to visualize data. They call it casual visual users. Many Eyes: very successful and provided simple visualizations that could be shared in communities. In additional, the visualization application tool kits: Protovis, D3 and Canvasjs are widely used in lots of papers. However they have plenty of different. Table 4 compares the different implementation frameworks that I considered for this thesis.

Table 4: Implementation framework comparison

Visualization Application tool kit	Similarity (for all tools)	Dissimilarity	Feature
Protovis	1. Visualize data in the browser	Good at expressing straightforward and static scenes.	Static scenes
D3	2. The library is easy to use and the source codes are easily understood.	Pay more attention to the scene changes	Interaction and Animation
Canvasjs	3. All three tools can visualize trends, graphs and text and also tag clouds.	Expert in illustrate the big volume data and make the figure readable in the graph	Static scenes and simple animation

According to my project, the animation and dynamic visualization is not such common, thus I will choose the tool kit has the same advantages like Canvasjs, which is more suitable for my project.

2.6. Chapter Summary

Generally speaking, before the research, there were three issues to be addressed.

- Which matches between the types of data and the types of visualization techniques are more suitable for user to read and understand?
- What are the affordances and constraints to each case study?
- How to avoid the recognized difficulties during the course of implementation?

In this chapter, we have briefly analyzed the development of data and information visualization approach and techniques. From our analysis, we present the following findings:

- The common and different points between information visualization and data visualization: data visualization is a part of information visualization and the data is extracted from information.
- The classification of data type with respect to data visualization: they are Nominal data, Ordinal data, Interval data and Ratio data are four basic types of data and features of each type.
- The classification of data visualization technique: they are Scatter Plot, Line Graph, Pie Chart, Bar Chart, Tree Diagram, Map and Word Cloud and evaluating each technique.
- The usability, interactivity, scalability and evaluation are the main challenges during the course of data visualization implementation.
- Protoviz, D3 and Canvasjs are the most popular visualization tools used in current status.
- The discussion of related work shows visualization tools' advantage which can be taken and the disadvantage of should be avoided during the design and implementation phases.

Compared different types of data and matched with appropriate visualization techniques, this research on the current state of data visualization gives me a clear way to implement the project. Referencing each table I mentioned above, there has been established an explicit criterion to visualize the extracted data from LinkedIn. In other words, the GUIs will be designed as a standard and the project should be presented in a scientific way to users.

We believe that after this research, the power of social media and data visualization will be enhanced during design and implementation. From this project, the LinkedIn's impacts to the society are not only helping users to look for jobs, but also offering a way to different types of users to get out of the normal sight, and then gain useful advices and assistances to realize a better profession understanding to the relevant decision making.

3.Design

In this chapter, we will introduce the design process which starts as the explanation of methodology. The rest part of design should follow the established methodology. There is no doubt that this part was refined several times in order to make the best preparation for implementation. Thus this part is significant and it will impact the implementation and evaluation in the end.

In section 3.1 we describe the methodology, in section 3.2 we outline data and schema and in sect 3.3 we explain the data visualization and data interaction. In chapter summary part, an architecture figure will illustrate the whole process of my design.

3.1. Methodology

My research approach is based on case study with three types of users – HR, individuals and public services. The system will support three separate user cases on the same dataset. The design methodology that I have selected is outlined in Figure 16.



Figure 16 : Methodology

3.1.1. Scenario Developments

There are three different user types. The relevant data according to different users' requirements should be extracted from LinkedIn which means the project needs to know the requirements of users first and then by analyzing the extracted data to implement data visualization.

3.1.1.1 Requirements survey

This is the first step to establish a well data set and it is the key factor to decide the user

satisfactions. Because of the importance of this part, I referenced some latest papers about techniques for requirements gathering. For example the paper from IBM shows the whole process and the essential portions during study. [62] The instructions and suggestions make the result more and more accurate and definitely improve the user satisfaction in the end. In this session, I will introduce the process of requirements gathering and show the result of requirements analysis.

I designed three questionnaires (appendix 1&2&3) based on different scenarios. The aim of the questionnaire is to assess the feasibility of the project and lead a better way to achieve a better product. Through analyzing the result of questionnaire, I got very useful information I would use in the future.

The participants I choose from two groups, master student and HR officers. These two groups of user can meet the status requirements of HR scenario and Job seeker scenario. However, the government case, I assume and indicate the situation in front of the questionnaires in order to make the participants fully understand the meaning of government case.

For instance, “Alex is working for Irish government under the statistical department. He is usually confused by the complex data which show the trend of the employment rate and employment situation. He is looking for a better way to get the result in this area more clearly and conveniently.”

3.1.1.2 Requirements analysis

From the questionnaires I find that there are few questions are very essential and popular. Almost every participant has consideration in that area which will be provided in the implementation. After the survey, I summarize some examples of the questions in Table 5 and this makes a good start to complete the scenario development.

Table 5: Scenario questions gathering

Scenario	Example of questions:
Government	<ol style="list-style-type: none"> 1. What is the industrial distribution in Ireland? 2. What is the number of employee distribution on different industrial types? 3. What is the general situation for education background of employees among different cities?
HR	<ol style="list-style-type: none"> 1. Which college has the most popular graduates on different majors in Ireland? 2. What are the most popular skills that company required in different areas? 3. Which city has the relevant strong education background architecture?
Job seeker and college	<ol style="list-style-type: none"> 1. What are the most popular or essential skills the other employees have on different positions?

perspective student	<ol style="list-style-type: none"> 2. Which job has more positions in Ireland? 3. Which major or degree is more popular in Ireland? 4. What are the available jobs related to different degrees?
----------------------------	---

3.1.1.3 Data requirements and visualization

According to the requirement analysis, I generated three question tables for each scenario. They are Table 6 for government case, Table 7 for HR scenario and Table 8 for job seeker and college perspective student. Each table includes the involved data in each case, and also combined the knowledge I get from the previous chapter. Finally I make a decision about which visualization technologies will be used for each questions.

Table 6: Government Scenario data requirements and visualization

Factors (Key Words)	Questions can be solved	Involved Data	Task	Visualization
All cities(Whole Ireland)/Specific City/Cities + Time	<ol style="list-style-type: none"> 1. For Industrial What is the industrial distribution of the whole Ireland/Specific City/Cities? 2. For Education Background What is the education background distribution of the whole Ireland/Specific City/Cities? 3. For Company Size What is the company size's distribution of the whole Ireland/Specific City/Cities? 	<ol style="list-style-type: none"> 1. Industry data with location and time 2. Personal education background with location and time 3. Company size with location and time 	Locate the different cities and compare the industries/education background/company size of each city get an extract value of one year or an average value during a period of time.	<ol style="list-style-type: none"> 1. Single: pie chart 2. Multiple: bar chart
All cities(Whole Ireland)/Specific	<ol style="list-style-type: none"> 1. What is the employee's education 	<ol style="list-style-type: none"> 1. Industry data with 	Locate the different cities and compare the	<ol style="list-style-type: none"> 1. Single: pie chart

City/Cities + Industrial(One or two or three) + Education Background(ALL) + Time	background distribution of the specific industry/industries based on one city/ comparison between multiple cities?	location and time 2. Personal education background with location and time	industries and show the situations in terms of personal education background of each city get an extract value of one year or an average value during a period of time.	2. Multiple: bar chart
All cities(Whole Ireland)/Specific City/Cities + Education Background(One or two or three) + Industrial(ALL) + Time	1. What is the industries' distribution of the specific employee's education background based on one city/ comparison between multiple cities?	1. Industry data with location and time 2. Company size with location and time	Locate the different cities and compare the education level and show the situations in terms of individual industry background of each city get an extract value of one year or an average value during a period of time.	1. Single: pie chart 2. Multiple: bar chart
All cities(Whole Ireland)/Specific City/Cities + Industrial(One or two or three) + Company size(ALL) + Time	1. What is the company sizes' distribution of the specific industry/industries based on one city/ comparison between multiple cities?	1. Industry data with location and time 2. Company size with location and time	Locate the different cities and compare the industries and show the situations in terms of company size of each city get an extract value of one year or an average value during a period of time.	1. Single: pie chart 2. Multiple: bar chart
All cities(Whole Ireland)/Specific City/Cities + Company size(One or two or three) + Industrial(ALL)	1. What is the industrial distribution of the specific company size/sizes based on one city/ comparison between multiple cities?	1. Industry data with location and time 2. Company size with location and time	Locate the different cities and compare the company size and show the situations in terms of individual industry background of each city get an extract value of one year or an average value	1. Single: pie chart 2. Multiple: bar chart

+ Time			during a period of time.	
All cities(Whole Ireland)/Specific City/Cities + Education level (One or two or three) + Company size(ALL) + Time	1. What is the company sizes' distribution of the specific education level/ levels based on one city/ comparison between multiple cities?	1. Personal education background with location and time 2. Company size with location and time	Locate the different cities and compare the education level and show the situations in terms of company size of each city get an extract value of one year or an average value during a period of time.	1. Single: pie chart 2. Multiple: bar chart
All cities(Whole Ireland)/Specific City/Cities + Company size(One or two or three) + Education level (ALL) + Time	1. What is the education level's distribution of the specific company size/sizes based on one city/ comparison between multiple cities?	1. Personal education background with location and time 2. Company size with location and time	Locate the different cities and compare the company size and show the situations in terms of education level of each city get an extract value of one year or an average value during a period of time.	1. Single: pie chart 2. Multiple: bar chart

Table 7: HR Scenario data requirements and visualization

Factors (Key Words)	Questions can be solved	Involved Data	Task	Visualization
All colleges/Specific college/ colleges + Time	1. For Degree What is the degree distribution of the all colleges/Specific college/ colleges 2. For Skill What is the skills distribution of the all colleges/Specific college/ colleges 3. For Work experience	1. Degree with location and time 2. Skill with location and time 3. Personal work experience with location and time	Locate the different colleges and compare the degrees/skills/work experience of each college get an extract value of one year or an average value during a period of time.	1. Single: pie chart 2. Multiple: bar chart

	What is the work experience distribution of the all colleges/Specific college/ colleges			
All colleges/Specific college/ colleges + Degree(One or two or three) + Skill(ALL) + Time	1. What is the employee's skill distribution of the specific degree/degrees based on one college/ comparison between multiple colleges?	1. Degree with location and time 2. Skill with location and time	Locate the different colleges and compare the degrees and show the situation in terms of skills in each college get an extract value of one year or an average value during a period of time.	1. Single: pie chart 2. Multiple: bar chart
All colleges/Specific college/ colleges + Skill(One or two or three) + Degree(ALL) + Time	1. What is the degree of the specific employee's skill/skills based on one college/ comparison between multiple colleges?	1. Degree with location and time 2. Skill with location and time	Locate the different colleges and compare the skills and show the situation in terms of degrees in each college get an extract value of one year or an average value during a period of time.	1. Single: pie chart 2. Multiple: bar chart
All colleges/Specific college/ colleges + Skill(One or two or three) + Work experience(ALL) + Time	1. What is the work experience distribution of the specific employee's skill/skills based on one college/ comparison between multiple colleges?	1. Degree with location and time 2. Personal work experience with location and time	Locate the different colleges and compare the skills and show the situation in terms of work experience in each college get an extract value of one year or an average value during a period of time.	1. Single: pie chart 2. Multiple: bar chart
All colleges/Specific college/ colleges	1. What is the skill's distribution of the specific employee's	1. Personal work experience with	Locate the different colleges and compare the work experience	1. Single: pie chart 2.

+ Work experience(One or two or three) + Skill(ALL) + Time	working experience based on one college/ comparison between multiple colleges?	2. Skill with location and time	and show the situation in terms of skills in each college get an extract value of one year or an average value during a period of time.	Multiple: bar chart
All colleges/Specifi c college/ colleges + Work experience(One or two or three) + Degree(AL L) + Time	1. What is the degree's distribution of the specific employee's working experience based on one college/ comparison between multiple colleges?	1. Personal education backgrou nd with location and time 2. Company size with location and time	Locate the different colleges and compare the work experience and show the situation in terms of degree in each college get an extract value of one year or an average value during a period of time.	1. Single: pie chart 2. Multiple: bar chart
All colleges/Specifi c college/ colleges + Degree(On e or two or three) + Work experience(ALL) + Time	1. What is the working experience's distribution of the specific degree/degrees based on one college/ comparison between multiple colleges?	1. Personal education backgrou nd with location and time 2. Company size with location and time	Locate the different colleges and compare the degrees and show the situation in terms of work experience in each college get an extract value of one year or an average value during a period of time.	

Table 8: Job seeker and college perspective student Scenario

data requirements and visualization

Factors (Key Words)	Questions can be solved	Involved Data	Visualiz ation
Specific(on e or two or	1. What do I have based on skills?	1. Job title with location and time	Bar chart

three) job title + Time		2. Personal working experience 3. Offered jobs from all of the companies	
Specific(one or two or three) job title + Skills + All cities	1. What do I want based on positions?	1. Personal working experience 2. Offered jobs from all of the companies 3. Personal education background	Pie chart
Specific(one or two or three) Skills + Degrees + All cities	1. What do I need to fight for in future 3-4 years in college?	1. Offered jobs from all of the companies 2. Personal education background 3. Locations	Bar chart
Specific(one or two or three) Skills + Positions + All cities	1. Which position I will get after I grasp the skill during college?	1. Offered jobs from all of the companies 2. Personal education background 3. Locations	Bar chart

3.2. Data and Schema

This whole project is divided into two parts. The other project “Developing knowledge models of social media over world wide web” is responsible for data model design. First of all, the data model is designed based on the “involved data” one of the columns in data mapping tables which I show above. In addition, Jinwu Li who is the writer of the other project using Python and its libraries as the main programming language to complete the data model. The third step is using Lucene text search engine to get full-text indexing and searching for data normalisation from 13000 profiles from LinkedIn Ireland. At last, the extracted data from LinkedIn are stored into” 4 store” in RDF format [64]. This is a remote server. Through this approach, my project can access the dataset and get result according to different queries using semantic web query language SPARQL. [65]

In order to complete the queries in a convenient way and show the relationship between two different objectives more straightforward, Jinwu generated a graphic which can show all the information of the data model as illustrated in Figure 17.

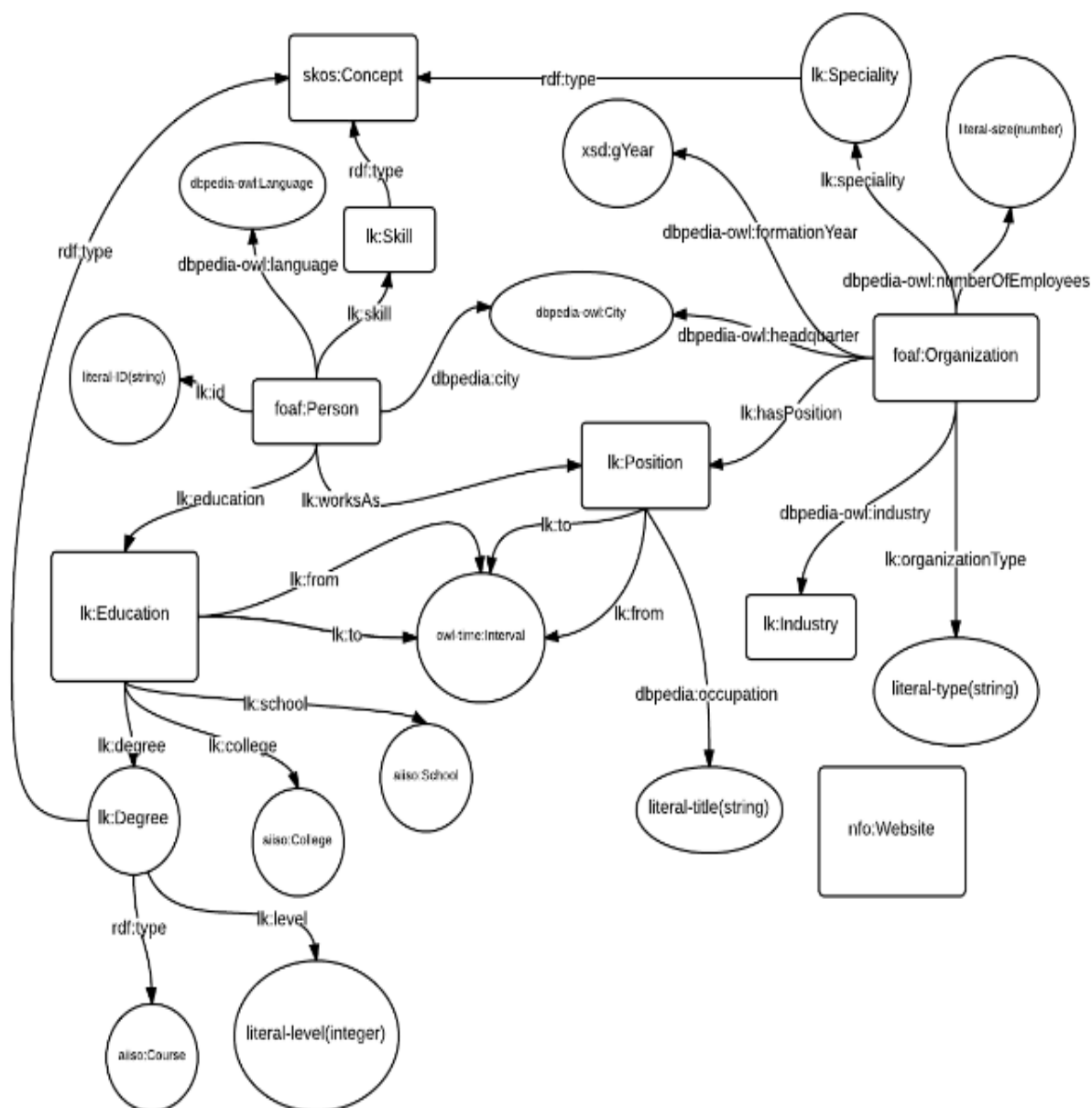


Figure 17: Knowledge Model for LinkedIn public profile and company profile

This figure contains information about public profiles such as person, skills, their occupation, education background, organizations and locations. We use the relationships (e.g. object properties) between these concepts to search across the graph and provide the visualizations.

Jinwu Li did extensive analysis on the accuracy a completeness of the extracted data from the LinkedIn Ireland. It seems that the extracted data is high quality. [66]

3.3. Data visualization and data interaction

In order to improve the user experience I combine requirements of the involved data and task proposed before, data visualization and data interaction design worth a long process to refine it again and again. At first I had several discussions with my two supervisors. With consideration of their comments, we draw different sketches as reference. After the sketches phase, I designed the layout of the interface in three different versions. This resulted in three interactions on the original design. Here we show the iterations of the initial design including sketches for each scenario and three versions graphics.

3.3.1. Data visualization and interaction design Sketches

We discussed the tasks for each scenario according to the data requirements and visualization as we discussed in Table 6, Table 7 and Table 8, we decided the main function and objectives of the whole interface for each scenario. Figure 18, Figure 19 and Figure 20 these three sketches show some sample data visualization and user interaction design. From The design process we interactively changed the design of interactions based on our ongoing discussions.

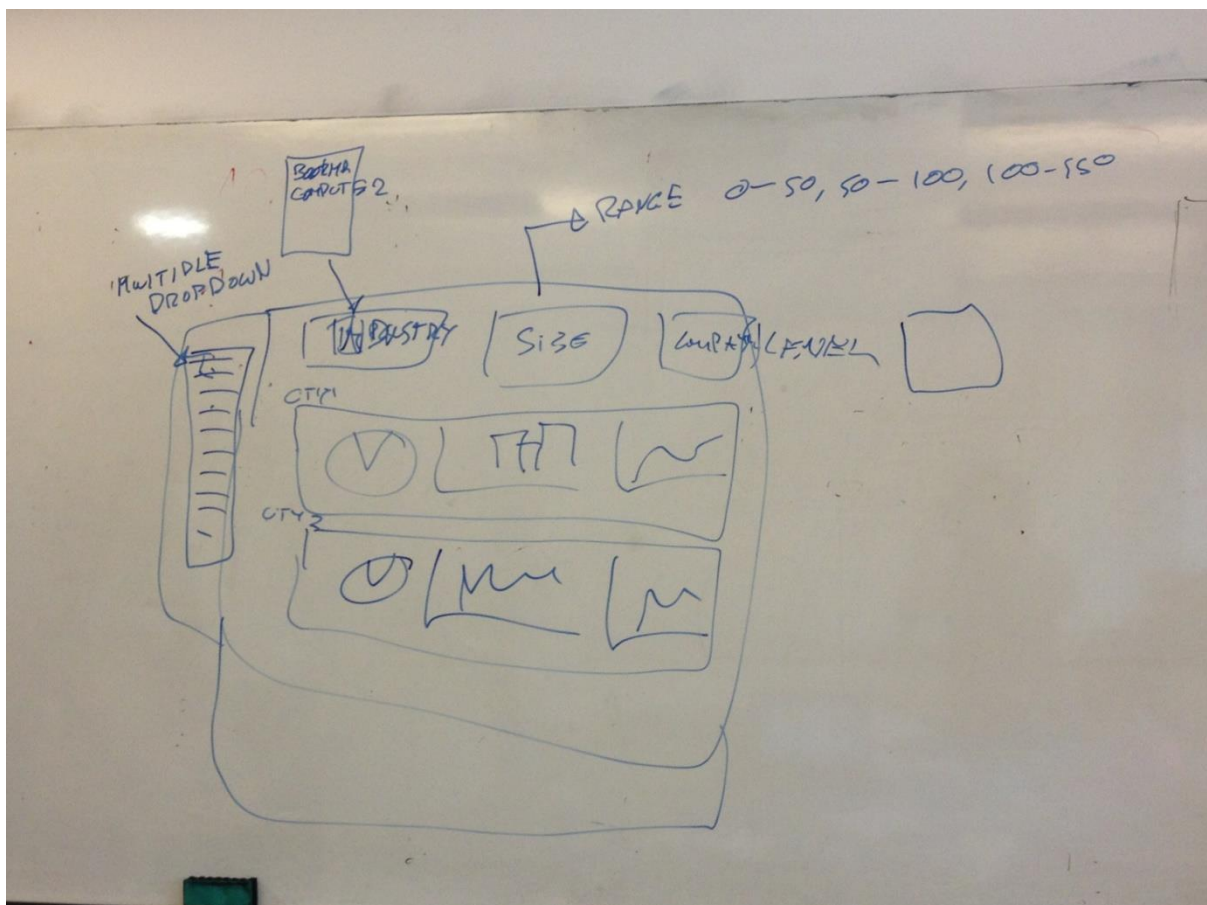


Figure 18: Government scenario sketches

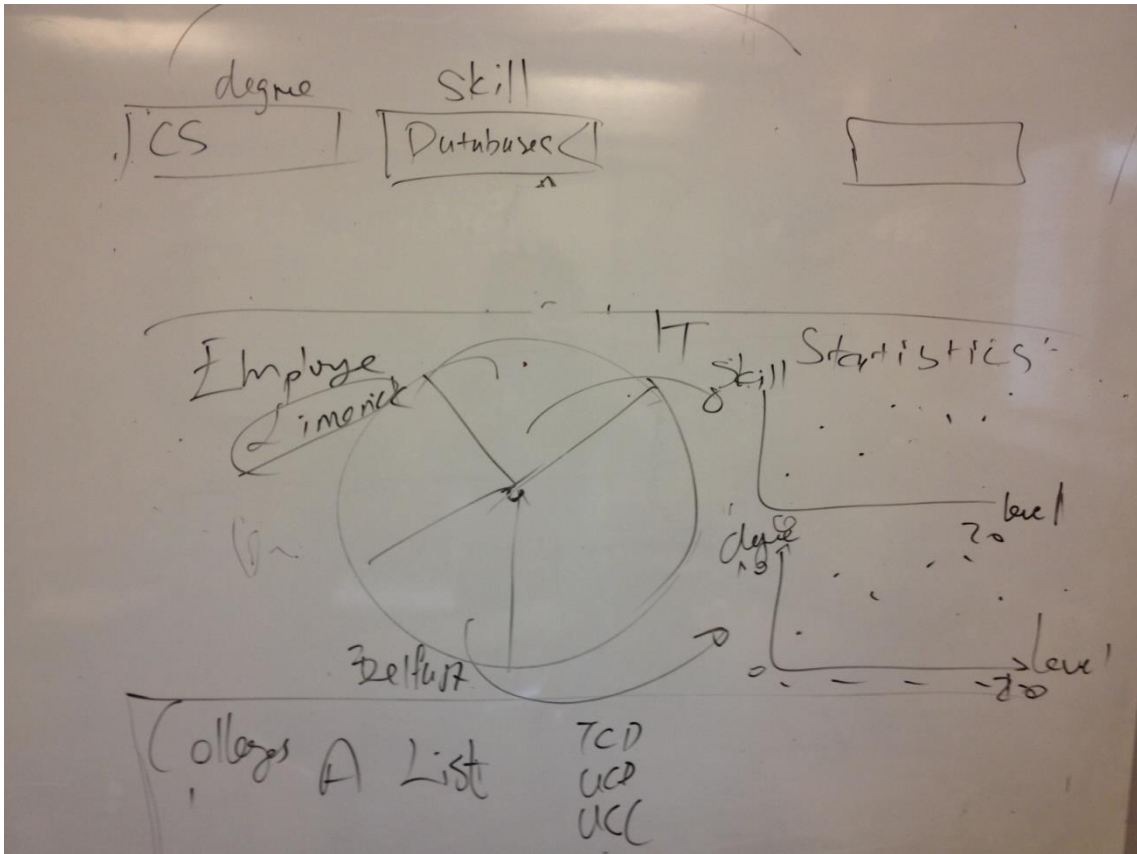


Figure 19: HR scenario sketches

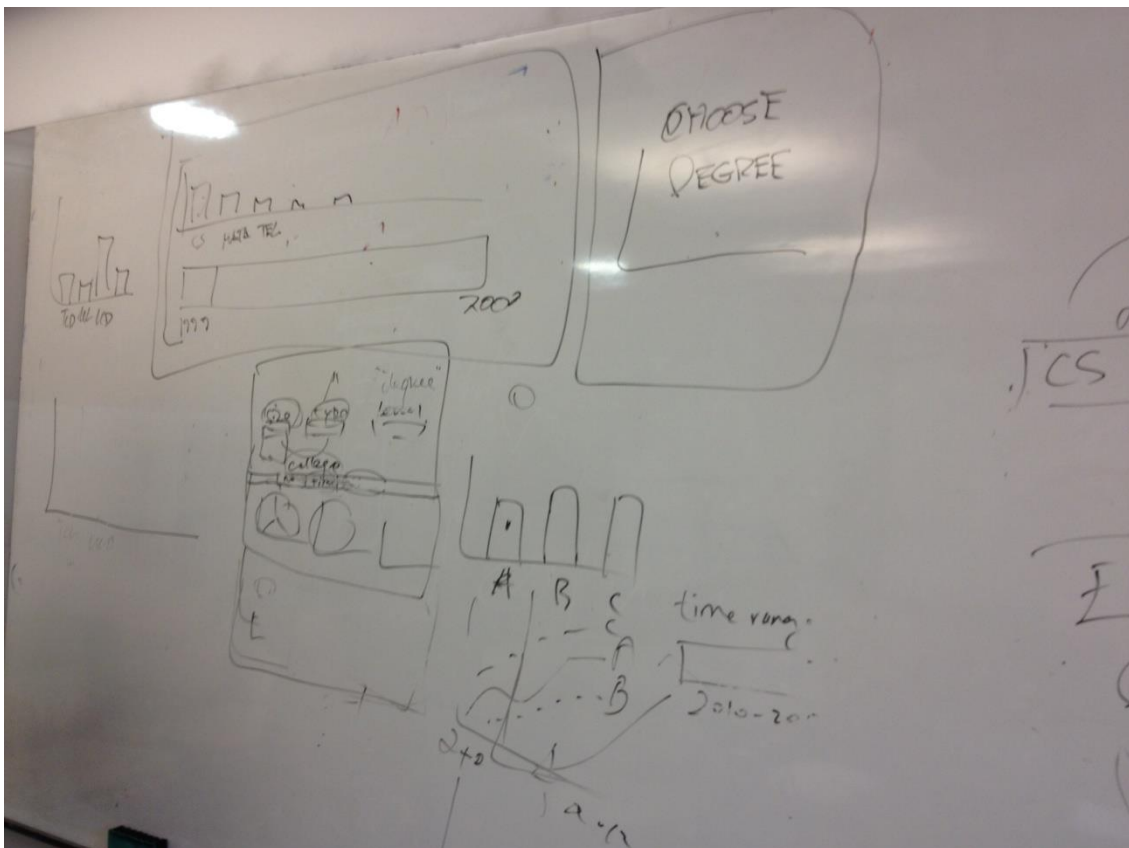


Figure 20: Job seeker and college perspective students scenario sketches

3.3.2. Data visualization and interaction design-Version1:

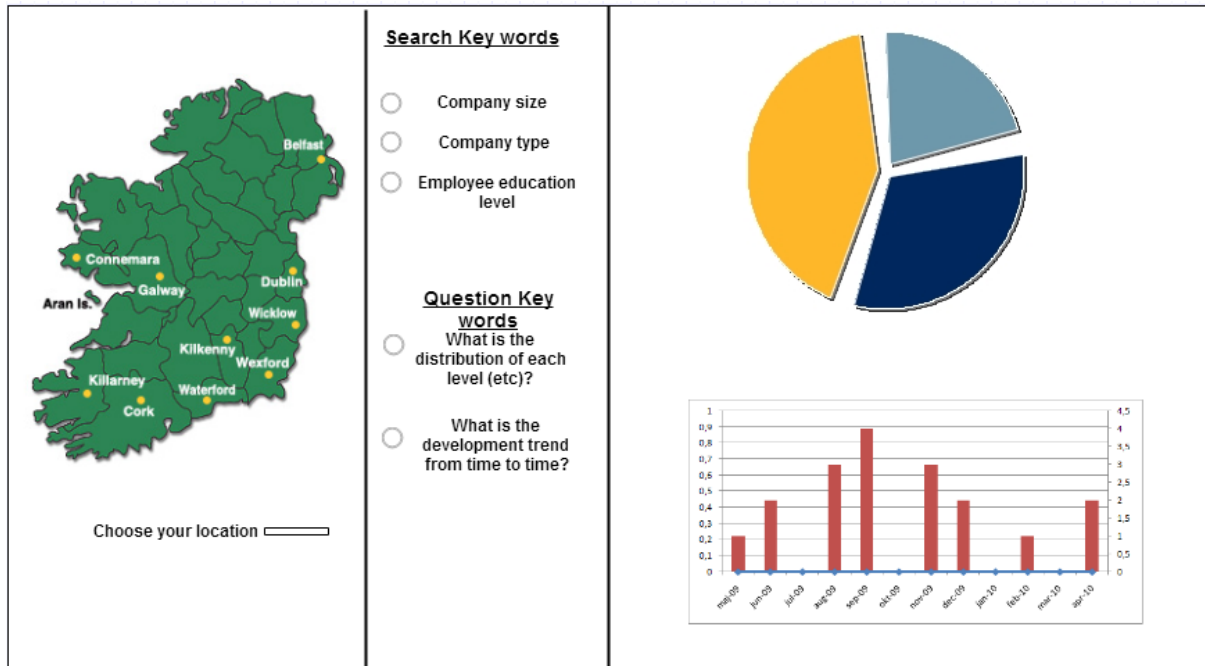


Figure 21: Layout version1

As we illustrated in Figure 21, this version emphasizes the locations in each scenario, but the query interface cannot be expanded when lots of comparisons requested. For instance, if I want to compare the employment rate between two cities, the graphics area cannot be shown the chart at one time. The whole interfaces need updating two times. Users are not able to check the comparison. Thus, we changed the design as shown in the next section.

3.3.3. Data visualization and interaction design-Version 2

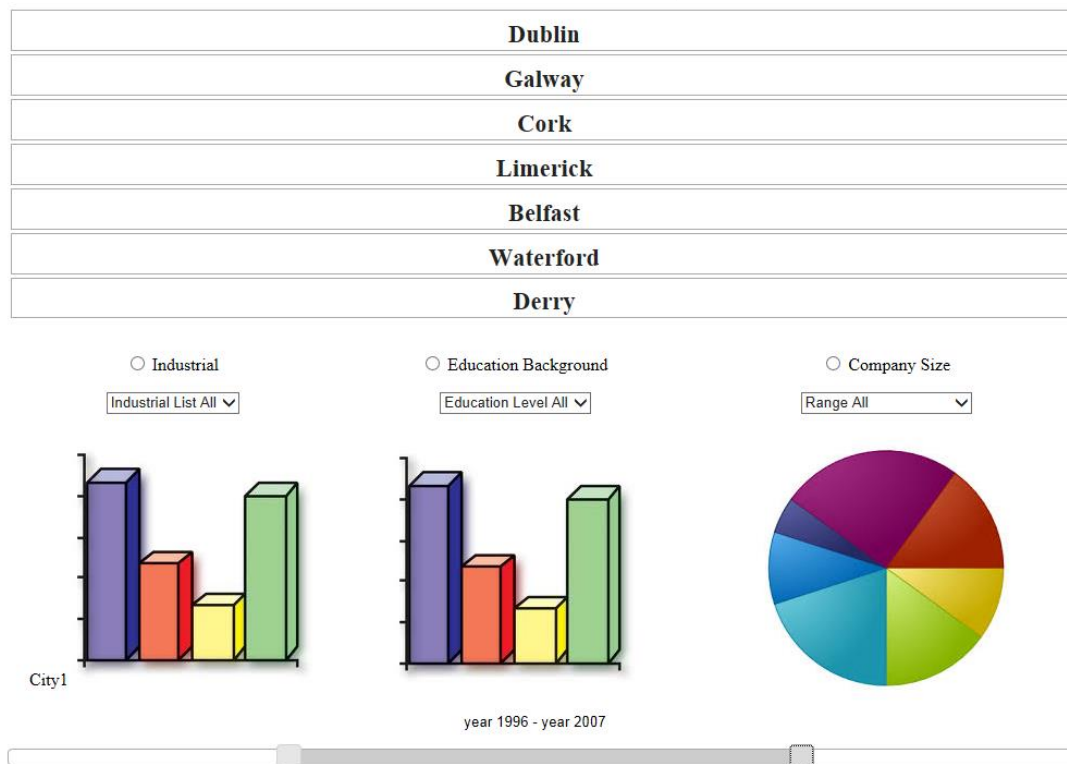


Figure 22: Layout version2

According to the drawback analysis of the first version, I decided remove the maps in order to offer a better way to compare the same situation between different cities as shown in Figure 22. Although, this version solve the problem of expanded the whole page when user request to compare, the graphic should be the main part of the whole page not the locations tabs. This is the issue for this version. Thus I designed the following version in the next section.

3.3.4. Data visualization and interaction design-Version3

The interface is titled "Data Visualization" and features a yellow header bar with a "City:" label and six radio button options: Dublin, Galway, Cork, Wexford, Waterford, and Limerick. An "Execute" button is located on the right side of this bar. Below the header, there are two main sections. The first section, labeled "Job Seeker", is divided into two panels: "You know what you have" and "You know what you want". Each panel contains two input boxes for "Skill:" and "Position:", with a pair of navigation buttons (> and <) between them. The second section, labeled "College Perspective Student", is titled "What do I need to fight for?" and contains three input boxes for "Skill:", "Degree:", and "Position:", with a pair of navigation buttons (> and <) between the first two. A green footer bar at the bottom contains the text "Copyright © 2012-2013 Lin Yue".

Figure 23: Layout version3

As we can see in Figure 23, this is the final version of the data visualization and interaction design. This version can emphasize the locations in each scenario. This query interface put the graphic in first status. The Figure 23 can be clearly shown the companions such as cities, skills, and education backgrounds and also attract users' attention.

3.4. Chapter Summary

In this chapter, I make a conclusion for all parts above. This architecture graph illustrates the whole essential parts of design and the perspective requirements of technique layer. For example, the data are stored in a remote server, in order to access the data and bring the result to the interface we need to generate the queries automatically. In addition, the queries are all based on the requests users made and the questions users asked. The general objectives for each scenario are included in Figure 24. This summarizes my design process.

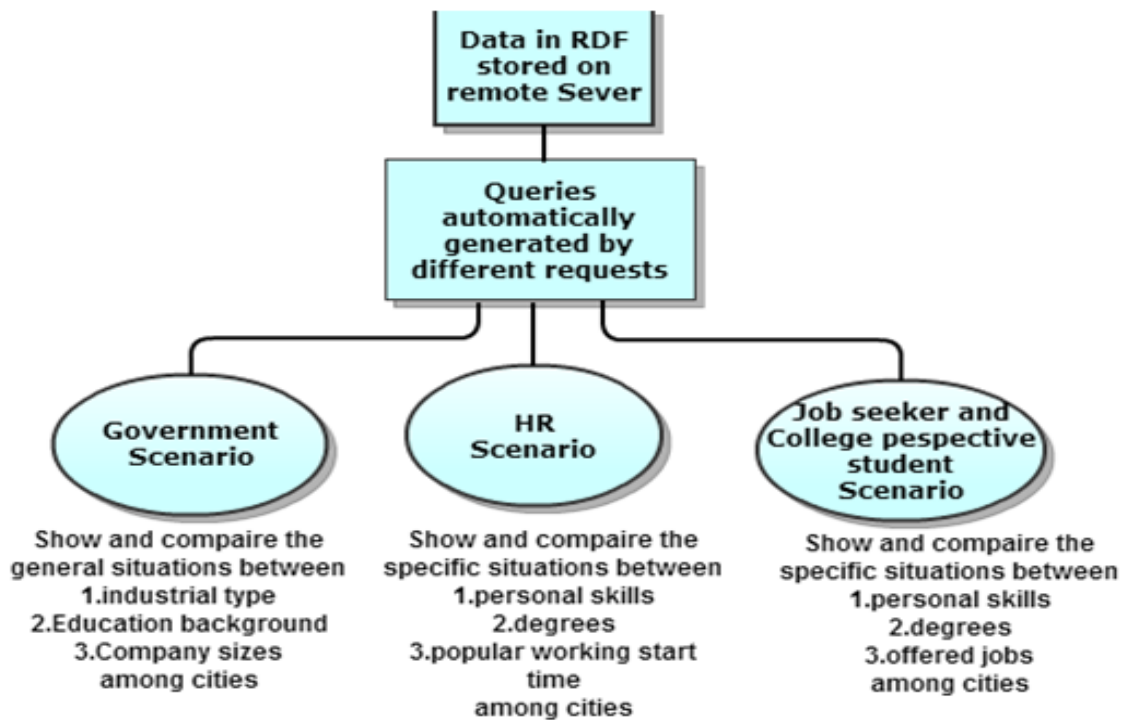


Figure 24: Design architecture

This architecture is a good way to summary what I have done in design chapter and also shows a very explicit direction and the procedures for implementation.

4. Implementation

This chapter discusses the technologies which I am using during the implementation. There are lots of issues in front of the implementation. The following text illustrates the reasons for choosing the appropriate technologies and the how can I solve the key technical issues. There are four main parts in this chapter. They are interface implementation, complex queries implementation, visualization implementation and user interaction implementation. In interface implementation part, I use HTML5 and CSS3 to improve the user impressions. In complex query implementation part, I use SPARQL to complete the query clauses. For the visualization implementation part, Canvasjs and D3 are very helpful. JavaScript and JQuery take the responsibility for user interaction implementation. In the end, I tested my queries and the performance of my project. This is not the part of implementation but supporting a good implementation.

The chapter is organized as follows. In section 4.1 I discuss the interface implementation, in section 4.2 I give some samples on complex queries, in section 4.3 I explain the codes of visualization implementation, in section 4.4 I illustrate the process of user interaction implementation and in section 4.5 I talk about the queries test and performance test during the implementation. The chapter summary shows the improvement after the implementation

4.1. Interface implementation

The main idea of my project is to achieve a query interface for LinkedIn.com which should be a web based application. There are three reasons I choose HTML5 and CSS3 for interface design:

- One new feature of HTML5 is the <canvas> element supporting 2D drawing which is very important to my project.[63]
- HTML5 makes code simple, by using HTML and its API, we can achieve lots of new functions just by browser instead of complex code in JavaScript which improve the efficiency of codes.
- CSS3 enhances the quality of rendering images on interface. Drawing the correct and vivid graphs is the most important thing I need to consider. CSS3 addresses this issue.

4.2. Complex query implementation

In this section, we need to produce the queries which based on dynamic interface user interaction. Thus we use SPARQL to query the data which store in the remote server in RDF format.

RDF (Resource Description Framework) is a W3C standard data representation model that allows machine-process able data communication between computers.

SPARQL (SPARQL Protocol and RDF Query Language) is a standard query language of the Semantic Web, which allows search and retrieval on an RDF graph. [64][65]

As I mentioned in the previous design part, the other project who takes the responsibility to extracted data from LinkedIn. That project should generate the RDF graph to express the relationship between the involved data which we figure out from the questionnaire analysis.

The benefits of SPARQL are: SPARQL can be used for open data which is not just for keyword matching, it can query properties/ relationships between concepts which other query language cannot achieve. A lot of work went into designing these queries based on user interactions for data requirements. Some SPARQL queries examples are shown in Figure 25, Figure 26, Figure 27 and Figure 28.

For example in Figure 25, a simple query of generating the result of all industry types in Ireland is using “select ? industry”. “select ? industry” means the result should only show the data belong to industry.

```
select ?industry where {  
  ?org a foaf:Organization;  
  dbpedia-owl:industry ?industry.  
  ?p a foaf:Person.  
}
```

Figure 25: A basic query example: generate the result of all industry types in Ireland.

Figure 26 is a mid- complex query which generates the result about the distribution of industry types in Ireland. In this sample, “count(?industry) as ?oCount “means count the data you choose. “Gourp by “can group the selected type of data

```
select ?industry (count(?industry) as ?oCount)where {  
  ?org a foaf:Organization;  
  dbpedia-owl:industry ?industry.  
  ?p a foaf:Person.  
  } Group by ?industry  
// If the user wants to check the distribution rank of industry type, the query will be  
  
select ?industry (count(?industry) as ?oCount)where {  
  ?org a foaf:Organization;  
  dbpedia-owl:industry ?industry.  
  ?p a foaf:Person.  
  } Group by ?industry Order by ?industry
```

Figure 26: A mid-complex query generate the result about the distribution of industry types in Ireland.

Figure 27 is a complex query which can generate the result about the distribution of industry types and compare the distribution between more than one cities which the users choose at the front end. In this query “FILTER “means the specific items are selected as limitations, the result will only show the data related to the item you choose.

```
select ?industry (count(?industry) as ?oCount)where {
  ?org a foaf:Organization;
  dbpedia-owl:industry ?industry.
  dbpedia-owl:city lk:** , lk: **.
  ?p a foaf:Person.
} Group by ?industry
```

Or

```
select ?industry (count(?industry) as ?oCount)where {
  ?org a foaf:Organization;
  dbpedia-owl:industry ?industry.
  dbpedia-owl:city lk:** , lk: **.
  ?p a foaf:Person.
FILTER (?dbpedia-owl = lk:** || ?dbpedia-owl = lk:**)
} Group by ?industry
```

Figure 27: A complex query: generate the result about the distribution of industry types and compare the distribution between more than one cities which the users choose at the front end.

Figure 28 is a super complex query that users choose the more than one items and at the same time add different limitations to each one. This query shows the rank and comparison according to a complex situation

```
select ?skill ?occupation_name(count(?occupation_name) as ?oCount) where Protovis {
  ?p a foaf:Person;
  lk:skill ?skill;
  dbpedia-owl:city lk:** lk:**;
  lk:worksAs ?position.
  ?position dbpedia:occupation ?occupation_name.
  FILTER (?skill = lk:** || ?skill = lk:**)
} Group by ?occupation_name order by desc(?oCount)
```

Figure 28: A super complex query: when you choose one or two skills, generate the result about the relevant distribution of positions and compare the distribution between more than one cities which the users choose at the front end.

4.3. Visualization Implementation:

As I mentioned in the state of art chapter, D3 and Protovis are the most popular visualization tool used by the up to date visualization applications.

According to my project, the animation and dynamic visualization is not such common thus I choose the tool kit has the same advantages like Protovis, but more suitable for my project.

Canvasjs is the tool I choose to implement visualization. There are three reasons for me to pick this API:

- CanvasJS excels in drawing static graphic and the performance for each type of charts is excellent- resulting in lightweight, beautiful and responsive dashboards.[67]
- CanvasJs improves the user interaction and it is clear to show big volume data. According to my project, I need to show the result among more than 10000 LinkedIn profiles. If the range of data is not limited heavily, the result should be huge.
- The most import aspect in data visualization is to show data in a very clear way, and make the graphics have a strong readability. This API adds the label to each point of data, thus when your click no matter which data point you want, the label will help you to screen the information you want.

4.4. User interaction implementation:

Because the visualization API is based on JavaScript and JQuery has a plenty of API and framework to implement an excellent user interaction, I choose JavaScript and JQuery to complete my project in this session.

Please see the appendix 4,5,6,7 for illustrating the sample of codes.

I give four sample codes which can provide a whole process of implementing user interaction. The codes are including JavaScript, JQuery and Ajax technologies. The four samples will be:

- Regular expression for retrieving the data from Json
- Set up a post with Json result format
- Case classification(according to the requests users make in front end)
- Visualize the data which I got from Json

These four parts are the essential codes. Please see the detail in appendix 5 6 7 8.

4.5. Queries test and performance test

According to the complexity of the queries, I cannot make sure every visualized result is correct or not. But this is the most important standard to evaluate my project is valuable or not. Because the volume of the extracted data is huge, we cannot calculate the correct

answers manually, the only way we can do is to test my queries one by one which means I wrote two similar queries and changed some limitations which I thought would impact the result and then compared the result to find out the difference is reasonable or not. After several times test, I got the result and combined the real world information to check the result. During the process, I also got some errors and then I changed it to correct. But sometimes I found the result is far away from my common sense, I double checked the queries I did not find any logical error which means the queries is correct. If this situation happened, I chatted with some professionals such as HR officers in a real company and staffs working in career office and talked about the result I got. I was trying to find out the reasonable explanation for the result. In addition, the data are extracted from LinkedIn not the official data from government, as a consequence, the data all generated by individual users. You would better think about and allow the situation which the information in a small area is a little bit exaggerating than reality.

In terms of performance test, I found the data loading time is longer than I planned. But this issue did not appear at the first stage of implementation. Because at first, there were only 6000 profiles in the remote server, the result can return in 2 seconds. After the server updated, the loading time became long. Thus the implementation almost finished, I could not change the codes from beginning, and I added some functions into the front page to notice the user the loading is processing. This is the way I improve the performance after the performance test.

4.6. Chapter summary

In this chapter, I illustrate what the four main parts of implementation. During the implementation, I recorded the issues I met, compared the technologies I choose, integrated the tasks I listed in the previous chapter. The most significant part is to test my product. From queries test and performance test, I changed the source codes to improve the quality of my implementation and get ready for the evaluation.

5.Evaluation

In this Chapter I compare the search interface with LinkedIn and at the same time I use three different ways to evaluate my project. Since there is no similar application which is designed for LinkedIn and the involved areas are different, it is inappropriate to compare a bespoke visual interface with the existing LinkedIn interface. That is the biggest challenge of the evaluation.

In section 5.1 I will make a comparison between LinkedIn and my project. In sections 5.2 I will illustrate the user study I conduct for user satisfaction evaluation. Section 5.3 discusses the SUS process and analysis the result for each scenario.

5.1. Comparison with LinkedIn

The comparison will focus on the main functions offered and the user interaction provided in both search interfaces. From these two angles, we can easily find out the similarities and dissimilarities between two interfaces in order to evaluate the value of my project in this area.

5.1.1. Main functions comparison LinkedIn VS My project

As shown in Table 9, purpose of LinkedIn and my interface is completely different. LinkedIn focuses on social network analysis and limited statistical analysis, whereas in my project we focus on analysis of facts from a wider perspective. For example job situation, employment rates, skill distribution Ireland.

Table 9: Main functions comparison between LinkedIn and my project

Category	LinkedIn	My project
➤ Similarity	<ol style="list-style-type: none"> 1. Raw job seeker’s good guide 2. Human resource officers’ and head hunters’ essential reference 3. Statistical data demonstration and related analysis 4. Personalize designed profiles and functions for job seekers and human resource officers. 	
➤ Dissimilarity	<ol style="list-style-type: none"> 1. Editing personal profile 2. Making a perfect Interpersonal relationship net 3. Automatically establish personal contact list 4. Automatically establish the relevant job vacancy list 5. Grouping users’ 	<ol style="list-style-type: none"> 1. Supporting three different user status: government staffs, HR and job seekers(college perspective students) 2. Mastering the overall distributions of industrial type, education background, relevant personal skills, offered jobs with location for each user by using statistical data and showing in graphical way

	<p>connection list</p> <p>6. Checking the connections' profile updating</p> <p>7. Applying jobs</p> <p>8. Releasing latest recruitment information</p> <p>9. Advertising enterprises' newest events</p>	<p>3. For government staffs:</p> <ul style="list-style-type: none"> ➤ being aware of the whole industrial development structure, education background distribution and the recruiting time phases ➤ supporting comparison between different factors among different cities/ whole Ireland <p>4. For HRs:</p> <ul style="list-style-type: none"> ➤ finding out the most suitable group of job seekers and setting the recruiting requirements for certain vacancies ➤ supporting comparison between different factors among different cities/ whole Ireland <p>5. For Job seekers and college perspective students:</p> <ul style="list-style-type: none"> ➤ finding out the most suitable group of vacancies and setting the aims for future career plan ➤ supporting comparison between different factors among different cities/ whole Ireland
--	---	--

5.1.2. User interaction comparison LinkedIn VS My project

In these two projects, they both offered statistical data demonstration pages and preference data searching pages as we shown in Table 11. Although the services are similar, the main idea of each project is different. For instance, as you can see Figure 29: LinkedIn visualization page, LinkedIn.com has already implemented some visualized data and the statistical data can be shown for a small group, but the data LinkedIn shows are all for individual group or users, it cannot show the general situation of a city or a whole country. The figure just shows the situation of Trinity College, Dublin. In addition, the data it visualized are not based on the requirements users have. Comparing to my project there are three benefits. At first, my project can visualize the data all according to the questions which generated by users in the front end, which means the visualized data in my project are more valuable. In addition, we use professional visualization API to draw the graphs, thus the visualized data we provided are more straightforward, readable and comparable. The last but the most important is we provide a novel visual search of interaction interface for LinkedIn Ireland using semantic web technologies this is the service LinkedIn.com did not provide.

Because the functionality of LinkedIn is varies greatly from my project. It was not

appropriate to compare both systems in a formal evaluation. Thus, I evaluate each system I individually triangulate the results.

Table 10: User interface comparison between LinkedIn and my project

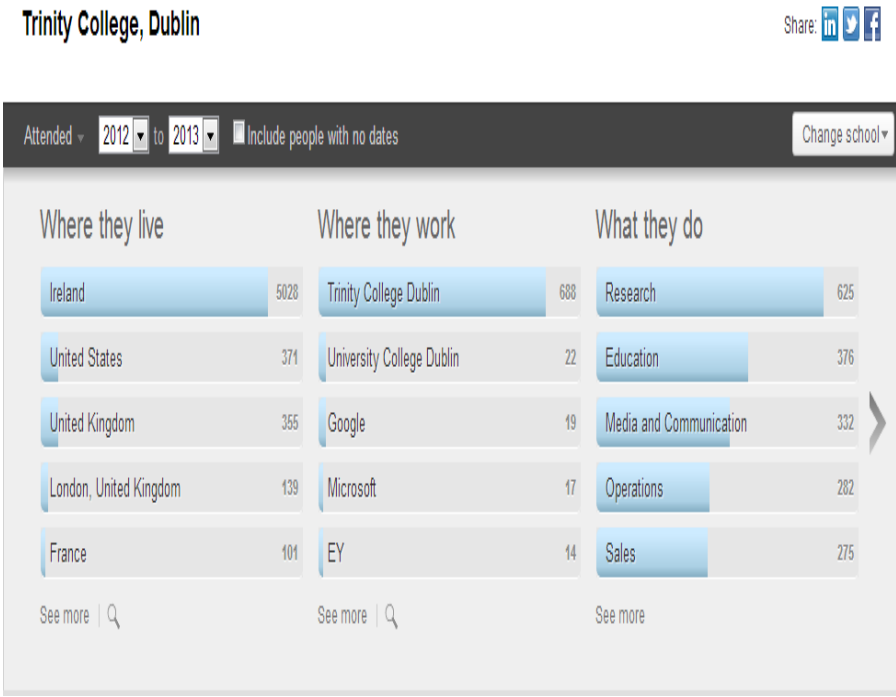

Category	LinkedIn	My project
<p>➤ Similarity</p>	<p>1. Statistical data demonstration page</p> 	

Figure 29: LinkedIn visualization page

Figure 30: My project visualization page

2. Preference data searching page

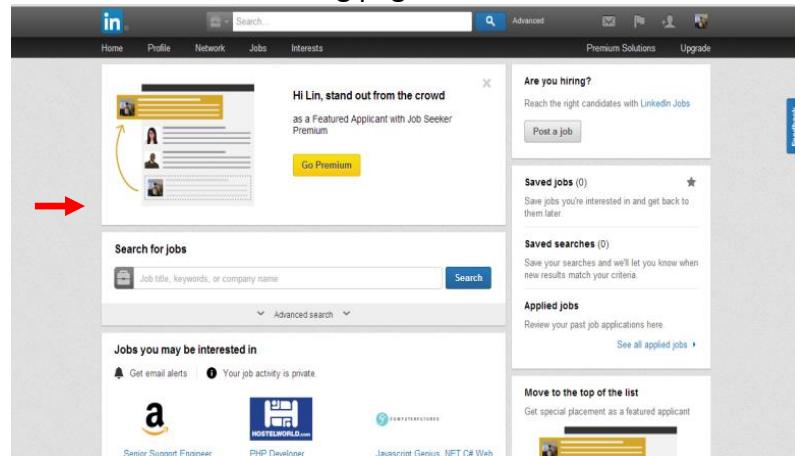


Figure 31: LinkedIn preference data searching page

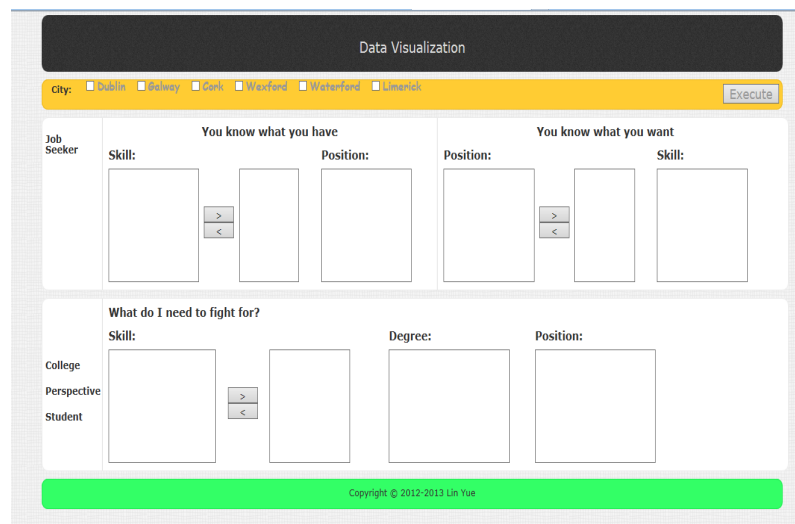


Figure 32: My project visualization page

➤ Dissimilarity

Although there are different functions, the interfaces are similar. However, all information LinkedIn give are specific data or individual results.

All functions are implemented by using graphics which are more straightforward and easy to understand.

5.2. User study

As I mentioned in State of the art, the evaluation is a big challenge in data visualization. We cannot just use one experiment result to judge the project is successful or not. Furthermore, there are three scenarios in my project. In order to make the difference between each case, I conducted:

1. Various task-based answer sheets to test the efficiency and effect my project bring to users for each scenario;
2. Distinct satisfaction questionnaires to evaluate how satisfied after users test my each scenario;
3. System Usability Scale questionnaires to assess usability designed for each scenario.

5.2.1. Hypotheses

The user evaluations have the following hypothesis:

- H1: My project can show the visualized data based on the requirement for three different user statuses and the performance is efficient.
- H2: My project provides high scores of user satisfaction that satisfies users.
- H3: My project is a well-established system which can meet the requirements of users in a high level.

5.2.2. Evaluation setup

There were 12 students from M.Sc. in computer science Trinity College Dublin and the other 6 participants are from Ava Trade Ltd in IFSC, Dublin1. These 12 students were responsible for Government, Job seekers and college perspective students' scenarios. 6 HR officers from Ava Trade Ltd were arranged in HR scenario.

This project has three different scenarios. Each scenarios has own responsibilities and tasks which I mentioned in design part. Thus the evaluation for each case should be individual. According to the results the project provide are all in graphics and the main idea of the project is to make the data more readable into graphics. We need to calculate the efficiency of the project which means I need to design task-based questions to record the answer and time at the same time I checked the consistency of the data the project visualized. To evaluate the user satisfaction is a good standard to weight the value of the project. I conducted the user satisfaction evaluation after the users finished the task answer sheet.

Finally, SUS (System Usability Scale) was used which the standard to evaluate the similar systems. The result will be convinced.

We used a task-based user evaluation to assess the efficiency of my project. The questions on the sheets are all based on the tasks we design in the previous chapter. We set the goals of my project for each scenario and the questions are all around the goals we want to achieve. The questions are long and short, some of them you need to get familiar with my project first then you will get the correct answer and some of them are very straightforward. The design of the questions is very fair and valuable to evaluate my project. For the user satisfaction questionnaire, we reference other paper that show a good way to evaluate an online system, from the services, efficiency, user interfaces and system four angles to evaluate the project. [68] My usability questions are designed based on the paper I referenced and also combine the specific situations- three different user status. Thus the result of the usability questionnaires is valuable. For the SUS (System Usability Scale) test, the questions are already set up.

5.2.3. Results

First the participants were asked to complete the task-based question answer sheet. The questions are shown in Table 12. I measured the completion time for each question and also check the consistency of the answers from each one. Next, they were asked to complete two surveys: user satisfaction questionnaire and SUS.

5.2.3.1 Task-based question answer sheet

Task completion time is presented in Table 12. As we can see from the table, Users can easily find out the answer they wonder, all questions can be completed in 5 minutes. The longest one the second question in HR case. However, due to the design of the question, the answer includes lots of comparison. Even though, the answer time is still a relatively short time.

Table 11: Task-based question answer sheet summary

Questions	Average completion time
<i>Government:</i> Q1: How is the employee education background different between Cork and Galway?	1'55''
Q2: What is smallest and largest company sizes in each different city?	3'09''
Q3: What is the most general education background regards to the industrial type of accountants and broadcast media in Dublin city?	2'03''

HR:	1'41''
Q1: What city has the highest number of PhDs?	4'28''
Q2: What most and least popular skills in each city?	1'41''
Q3: What's the most popular start date across all cities?	3'26''
Q4: For M.Sc./PHD/BA degree, what are the top skills in each city?	
Job seekers and College perspective students:	3'25''
Q1: If you had Machine-learning skill, what is the available job positions distribution in Ireland?	2'46''
Q2: What is most popular skill you may have if you want to be a Developer/Marketing manager?	1'18''
Q3: What is skills requirement difference between Junior IT developer and Senior IT developer?	

5.2.3.2 User satisfaction questionnaires

In this section, I describe the results for each scenario

- **Government scenario**

First, I will describe a sample scenario. Then I list the result for user satisfaction questionnaires and present the analysis.

Example user persona:

Kevin is working for Irish government under the statistical department. He is usually confused by the complex data which show the trend of the employment rate and employment situation. He is looking for a better way to get the result in this area more clearly and conveniently.

Survey result:

Table 12: User satisfaction questionnaires government score

Questions	Average scores
1. The visualized data which come from this system can be trusted.	4.6
2. The provided information was helpful.	4.0
3.The visualized data made it easier to understand the complex and huge volume data	4.6

4. The layout is clear and the content is understandable.	4.4
5. The system provided me with various functionalities to search the statistical data for recruiting in Ireland.	4.2
6. It was easy to compare the job seekers 'personal degree according to specific industrial type.	4.0
7. It was easy to compare the industrial types according to different cities.	3.6
8. It was easy to check the company sizes according to different cities.	3.6
9. I found the presented graphics was useful and understandable.	4.2
10. I found the interactions with the system were easy to comprehend.	4.2

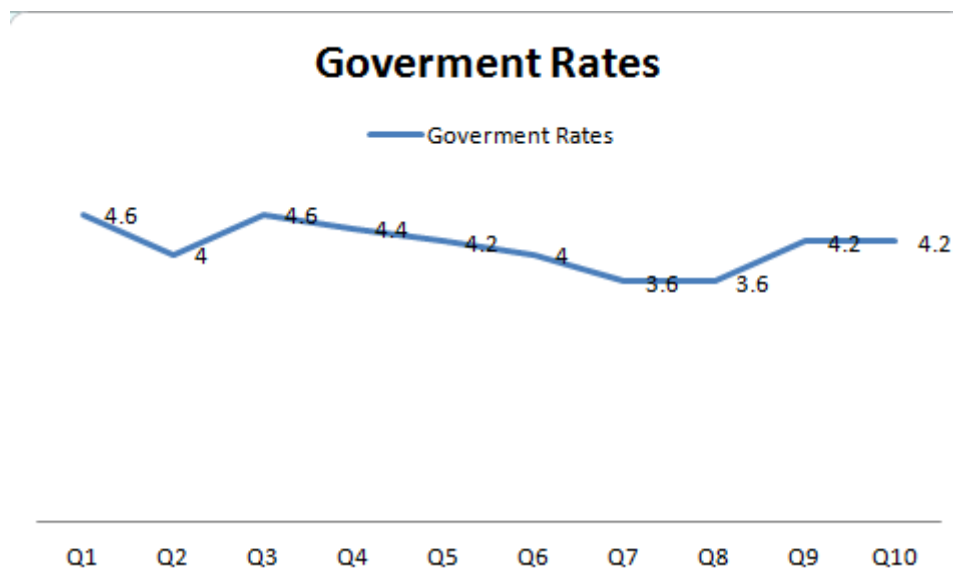


Figure 33: User satisfaction questionnaires government score

Analysis:

As you can see from the Figure 33, the scores are all above 3.6. The full marks for each question are 5, which mean the overall satisfaction is in a high level in government scenario. The highest score in this case is question 1 and question 3. The lowest score in this case is question 7 and question 8. The questions are shown in Table 13.

Question 1 and question 3 showed that that the visualized data are useful, this visual search interface can make the data more readable. It is helpful when government staffs is searching the general situation information which is related to their work.

Question 7 and question 8 were lower than other questions. According to the feedback I gathered from the participants, majority part of them found that the “execute” button should not be designed besides all cities toolbar. This will lead a misunderstanding that after the users choose the specific cities and specific industrial types, the graphic should be generated automatically. It is better to get rid of the “execute” button. Actually, I think about this issue before, since the loading data takes more than 5 seconds after you change

your choosed item, the execute button gives user a impression of you need wait for a second. This is the biggest problem in this scenario. However, all the other two scenarios are desginded by the same button on the page, this similar problem also happened on other cases.

➤ **HR scenario**

First, I will describe a sample scenario. Then I list the result for user satisfaction questionnaires and present the analysis.

Example user persona:

Alex is working for a trade company under the human resource department. He is usually confused on looking for the most suitable job seekers and recruiting new staffs. He is looking for a better way to make a good preparation to know the general employment situation in terms of colleges before commencing recruiting.

Survey Results:

Table 13: User satisfaction questionnaires HR score

Questions	Average scores
1. The visualized data which come from this system can be trusted.	3.7
2. The provided information was helpful.	3.83
3.The visualized data made it easier to understand the complex and huge volume data	3.7
4. The layout is clear and the content is understandable.	3.5
5. The system provided me with various functionalities to search the statistical data for recruiting in Ireland.	3.5
6. It was convenient to find the most and least popular skills in each city.	4.2
7. It is easy to find the highest number of PhDs and compare among Galway, Limerick and Wexford.	3.5
8. For Bs/BA/PhDs degrees, the graphics are clear to illustrate the top skills in each city.	3.83
9. It was handy to get the most popular start date in each city.	3.16
10. I found the presented graphics was useful and understandable.	3.83
11. I found the interactions with the system were easy to comprehend.	3.83

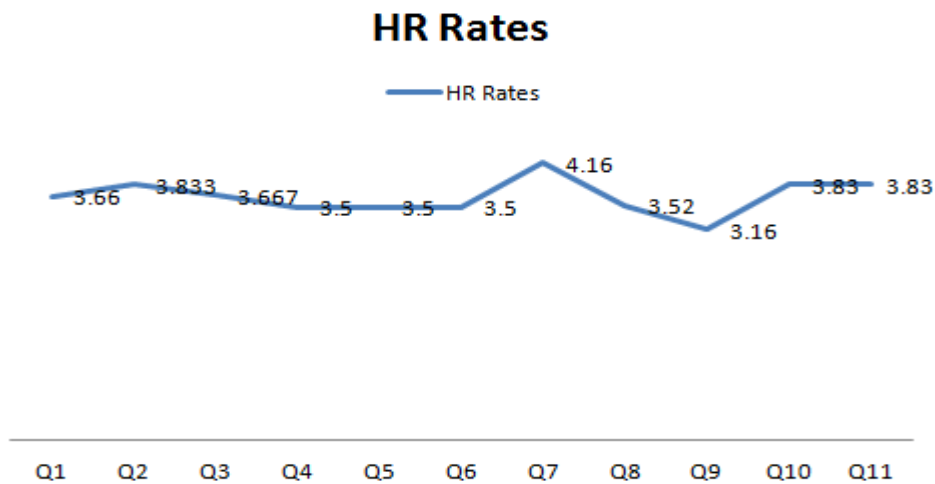


Figure 34: User satisfaction questionnaires HR score

Analysis:

As the information shown in Table 14, the scores in this scenario were lower than the other cases. From the feedback, I summarize two reasons. The first one is that the participants have been working of HR for companies for a long time, as their perspective this scenario can be used not only just for HR but more suitable for the company's founder and headhunters. At the first stage of design, we added the college information into this scenario. It will be helpful if the HR can use the information as reference in order to find out which college will be the better to get more suitable graduates. For example, there are 70% of employees who are working for computer science are from TCD in Ireland, which means if you want to recruit new staffs will work for computer science, the best choice is going to TCD to have a presentation to introduce your company. However, the data model had a problem of generation the college relevant data. We have to use city information to replace the college information. These changes directly made the scores of the whole scenario are lower than expected. Thus the average scores lower than other cases.

However in Figure 34, question 6 illustrates that although the average scores were lower than the other two cases, the overall information of this scenario are valuable. Most of the HR are concern about the skills distribution, so during the test, they all focused on the skill data visualization. From this result, we can easily find that the involved data in this scenario is useful and helpful to HR when they are recruiting new staffs.

➤ Job seeker and college perspective students scenario

First, I will describe a sample scenario. Then I list the result for user satisfaction questionnaires and present the analysis.

Example user persona:

Ruth is a job seeker who graduated from Trinity College, Dublin with the M.Sc. in computer science degree. She has a LinkedIn account but she cannot know the competition situation of the whole computer science industry from LinkIn.com She wants to make more preparation before the interviews.

Daire is a college perspective student who finished the Leaving Cert exam. He is looking for the new college and he does not know which course is suitable for him and which degrees are easier to get a job in Ireland.

Survey Result:

Table 14: User satisfaction questionnaires job seeker and college perspectives students score

Questions	Average scores
1 The visualized data which come from this system can be trusted.	4.0
2 The information is helpful when the job seekers approach to look for a more suitable job.	4.5
3 The visualized data is easier to understand the complex and huge volume data	4.0
4 The layout is clear and the content is understandable.	4.16
5. The system is very functionality and interface is comprehensive.	4.33
6. The online system will be used frequently when uses are confused to choose a more suitable job for them.	4.16
7 The system is very interacting.	3.67

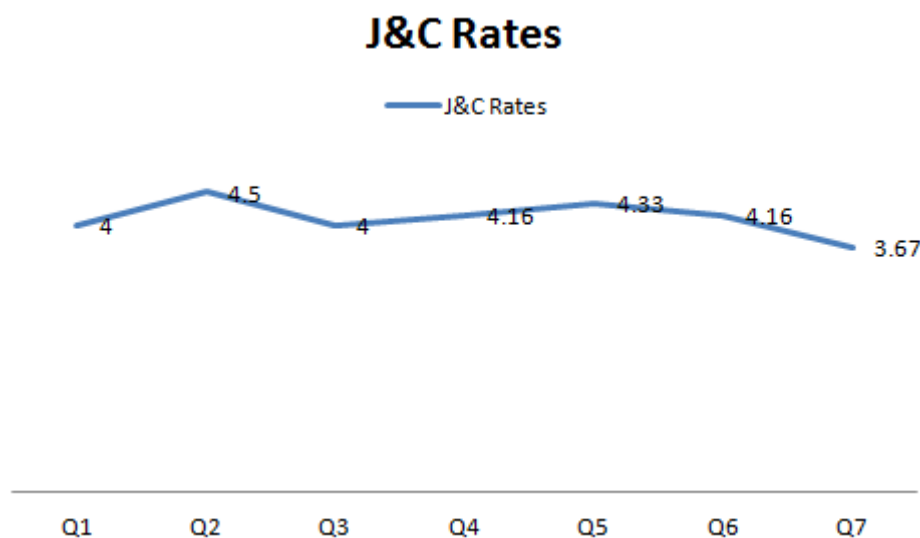


Figure 35: User satisfaction questionnaires job seeker and college perspective students score

Analysis:

The performance was found high by the participants. From the scores as we showed in Table 15, we can see the average level is around 4. There is no big movement between different questions. The reason is LinkedIn are designed for the job seekers, the relevant data are more convinced and more reasonable. In addition, I am also a job seeker and I also had the same experience with "Daire". The layout design and the visualized graphic design are more suitable than other two cases. This is the special point of my project and from the score we can prove that this case is more successful in user satisfaction evaluation.

Although the overall scores were very high, the question 7 was still not very satisfied level. When the information can provide the enough useful information to user, the requirements for user are moved up to a new layer. They want more interaction between the graphic and themselves which means the visual search interface should be more intelligent to meet the users various preference.

5.2.3.3 SUS Usability

The System Usability Scale (SUS) is a widely used and well-established standard to test the usability of the whole system [69].

Figure 36 and Table 16 illustrate that the performance of government scenario was better than the other two. The reason for this ranking is that government scenario has the best impression in every part of the whole project. Actually, the performances of three scenarios were slightly different. However, government case was excellent in question 4 and question 8. These two questions show that government scenario is handy to use and the user interactions are designed appropriately. However there is a big issue about the data consistency in government case. The industry result is far away from the expectation. I tested the same query in other case, the answer was fine. Before the remote server upgraded, the result is all in the normal range. This is a main reason leading the lower score in question 5. Other cases all get high average scores. Thus we can figure out the question 5 is a good judgment for queries.

HR scenario won the first place at question 1 which means this system will be used by HR in future. The design of this case is successful and the product is valuable to users. At the same time, many questions showed that the HR scenario is a little bit complex to use. This is the reason why in the previous user satisfaction test, the scores of HR case was lower than the others. We used SUS which can find out the reason to illustrate what is the merit of each case, and also find out the reason why this case has the drawbacks.

Job seeker and college perspective student case is the special point of the project. Participants were confident to use it. That is the reason why the participants would like to use this system in future and they found it useful for looking for jobs. This is the most significant objective we wanted to achieve. On the other hand, the system was not as easy to use as government scenario. The user interaction part is also the biggest challenge in front of us.

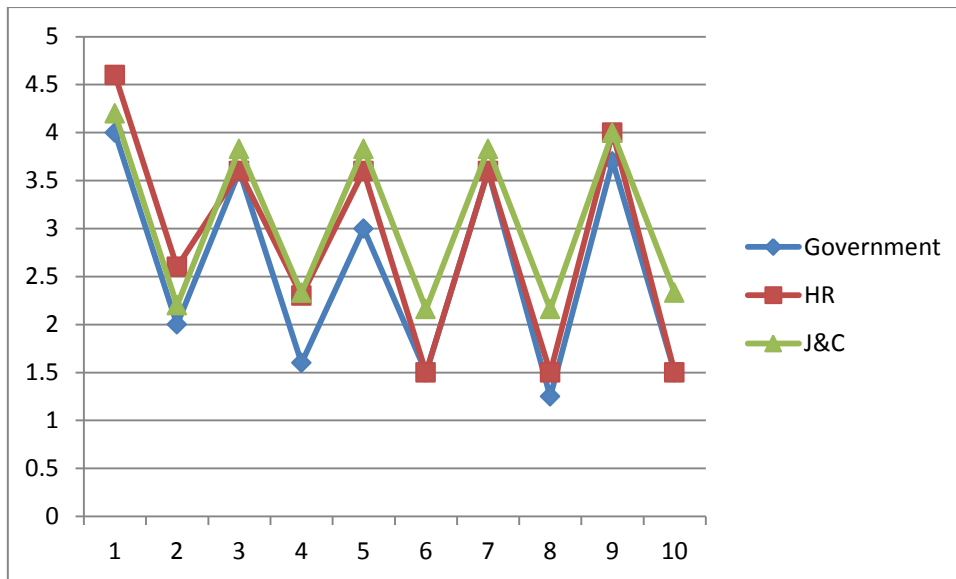


Figure 36: SUS usability comparison

Table 15: SUS questionnaire score for three scenarios

Question	Government	HR	Job seeker and Perspective Student
1. I think that I would like to use this system frequently.	4.0	4.6	4.2
2. I found the system unnecessarily complex.	2.0	2.6	2.2
3. I thought the system was easy to use.	3.6	3.6	3.83
4. I think that I would need the support of a technical person to be able to use this system.	1.6	2.3	2.33
5. I found the various functions in this system were well integrated.	3.0	3.6	3.83
6. I thought there was too much inconsistency in this system.	1.5	1.5	2.16
7. I would imagine that most people would learn to use this system very quickly.	3.6	3.6	3.83
8. I found the system very cumbersome to use.	1.25	1.5	2.16
9. I felt very confident using the system.	3.7	4.0	4.0
10. I needed to learn a lot of things before I could get going with this system.	1.5	1.5	2.33

Then we calculated SUS scores according to the paper illustrates. [69] This is the procedure to score SUS:

1. SUS score range from 0 to 100.
2. For odd number questions, the contribution is the scale position minus 1.
3. For even number questions, the contribution is 5 minus the scale position

4. Add all the contribution of each question and multiply 2.5 to get the final scores [70].

We got the result for three scenarios. The results which are shown in Figure 34 match the analysis for Figure 36 and Table 16.

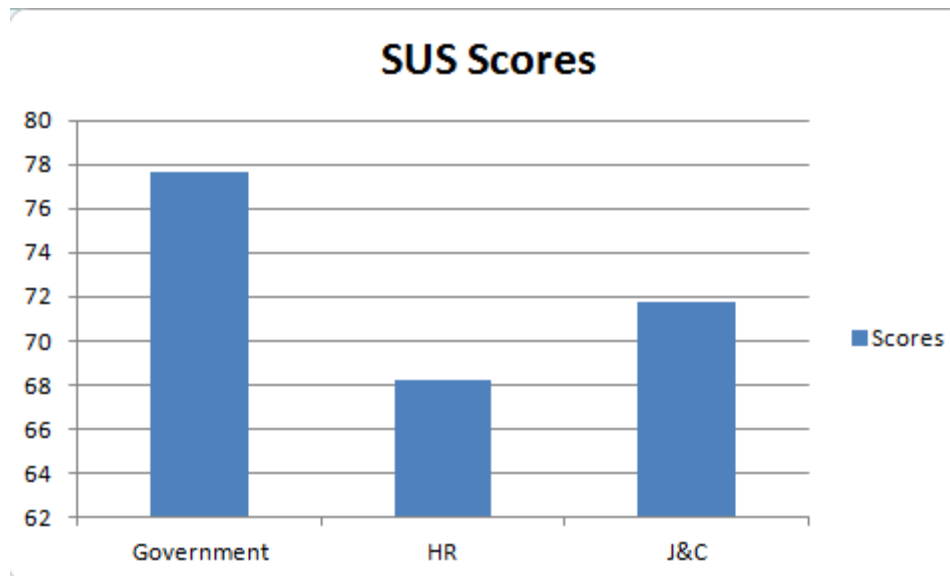


Figure 37: SUS Scores

Figure 37 showed the comparison result. The government scenario was higher than the other one. The government scenario achieved 77, job seeker and college perspective student scenario got 72. HR case achieved 69.

Because SUS already evaluated thousands of system, there is a standard to make a obvious level to decide the system is successful or not. The situation is, if the system awards the final score more than 60 it can be decided as fair enough system. If the system awards more than 70, it really means the system is above average level. [71]

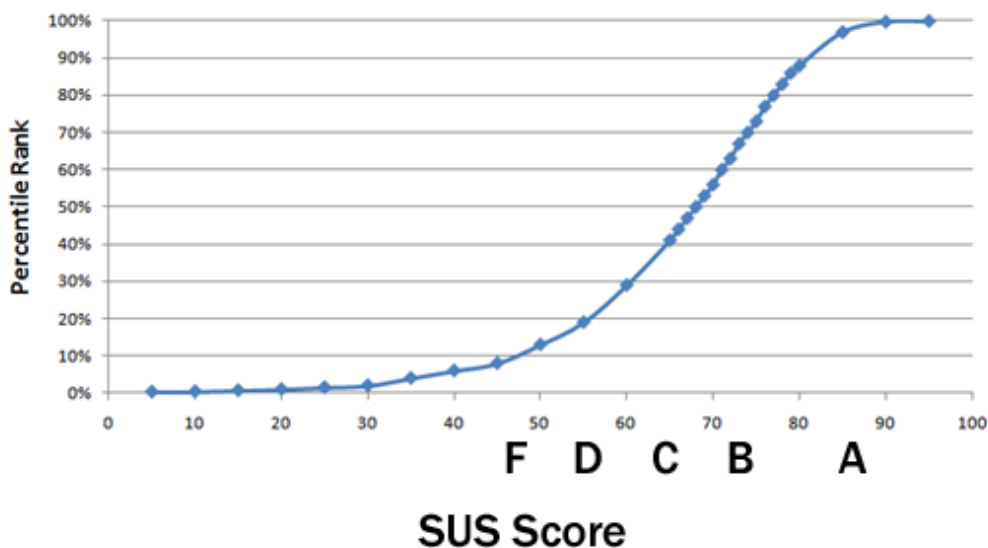


Figure 38: SUS Score standard

Table 16: SUS level

Scenario	Score
Government Scenario:	B
HR Scenario:	C
Job seeker and college perspective student Scenario:	B

The standard makes a classification for each level. As you can see from the graph, the final results of all scenarios are above pass line C. Two of them are higher than average level which means my project is successful in the end.

5.3. Chapter summary

This chapter records the whole process of evaluation for my project. Three different angels support a comprehensive, trustworthy and impartial result. As I mentioned in state of the art part, evaluation is the biggest challenge for data visualization. In addition, this project does not have any baseline system to compare which becomes an extra difficult point to evaluate my project.

- To evaluate project's efficiency, we use task-based answer sheet, we record the completion time for each question.
- To evaluate the project's user satisfaction, we use user satisfaction questionnaire which the questionnaires are design according to the objectives for each scenario and also reference the similar user study. From reading the research paper, I found a better way to evaluate the user satisfaction with appropriate questions.
- To evaluate the whole system performance, we use SUS, a well-established scale to get the final scores and comparing the scores with the standard which generated by thousands of evaluations.

In the three evolutions and the comparison with LinkedIn search interface, my project all gets satisfactory result. However, there are some issues need addressing in future. For instance user interaction needs to be improvement. A good way is to generate relevant analysis respect to different user requests. Queries result need to be more accuracy and the college data should be modified and add it into college scenario. At the same time to add time line into each case will be convenient to show the trend. Personalize design will be a new part to implement in order to keep the user group. Thus, the evaluation shows some a good direction to refine my project and make it more effective.

6. Conclusion

In this thesis I implemented three different scenarios:

- 1. Government scenario is convenient to explore the general situation such as employment rate, industrial distribution.**
- 2. HR scenario can be very helpful for recruiting investigation.**
- 3. Job seeker and college perspective student scenario will be a good suggestion to the career path selection.**

These three scenarios are supported by a plenty time of design and established by semantic web and visualization technical tools. From this project, the data we extracted from LinkedIn tell the truth, show the facts and lead the correct understanding to Government staffs, HRs, job seekers and college perspective students.

The evaluation which is taken from thee different angles to assess the user satisfactions, the product efficiency and system achievement. The result of each angle shows the product is successful and the product achieves the goals we set before implementation. Although there are some problems that need addressed in future, the overall scores can prove that the project is well established.

In conclusion, I implemented a novel visual search and interaction interface for LinkedIn Ireland using semantic web technologies, which has not been done before to the best of our knowledge.

6.1. Future Work

According to the evaluation result, there are three issues need addressed in future.

6.1.1. Automatically generate relevant analysis respect to requests.

One of the biggest challenges of data visualization is how to improve the user experience. Users would like to use the visualized data to make decision quicker and accurately. There is a situation that user cannot fully understand the meaning of the graphics. Another instance is user cannot grasp the main points when they are comparing one factor but in different circumstance. If we can add some analysis besides the graphs automatically and show some case study after the visualized data, it will be very helpful to improve the user experience in very short time.

6.1.2. Personalize design for individual users

User career path selection, HR recruiting staffs investigation and Government employment rate exploration are long term process. If the system can record of each query for individual users and mark the development for individual users, it will attract more and more attention on my product.

6.1.3. Refine the data structure and add time line to each scenario in order to check the trend of each area involved.

Time line is a very important factor which impacts the result of visualized data. With a timeline, you can see the development and trend of one area you interested. That should be the best way to be aware of the movement and make more accurate decisions.

6.2. Closing Remarks

Now data visualization of social media is becoming a new trend. This project can provide a new data visualization interface for LinkedIn.com which is novel. This allows complex query of Irish industry and employments just extracted from LinkedIn user generated content. Before this project, data visualization is mainly used in scientific research such as medicine study and biotech pharmaceuticals studies. This project will be a good trigger to lots of ordinary people not only the professional to experience the convenience the data visualization bring to us.

7. Reference:

- [1] Li, Y., Shi, Y., Fan, X., & Bhavsar, M. (n.d.). CareerGalaxy – A planner for future.
- [2] Dar, H., & Shah, A. (2013). Analysis of SNs popularity from different perspectives among users. 2013 5th International Conference on Information and Communication Technology for the Muslim World (ICT4M), 1–4. doi:10.1109/ICT4M.2013.6518875
- [3] Auradkar, A., Botev, C., Das, S., De Maagd, D., Feinberg, A., Ganti, P., Gao, L., et al. (2012). Data Infrastructure at LinkedIn. 2012 IEEE 28th International Conference on Data Engineering, 1370–1381. doi:10.1109/ICDE.2012.147
- [4] Keim, D., Qu, H., & Environments, H. (2013). Big-Data Visualization, 2013–2014
- [5] Nagel, H. R. (n.d.). Scientific Visualization versus Information Visualization, 2–5.
- [6] Pellegrino, D., Robinson, A., Stryker, M., Weaver, C., Mitra, P., Turton, I., & MacEachren, A. (2008). Grand challenge award: Data integration visualization and collaboration in the VAST 2008 Challenge. 2008 IEEE Symposium on Visual Analytics Science and Technology, 197–198. doi:10.1109/VAST.2008.4677384
- [7] Keim, D. a. (2002). Information visualization and visual data mining. IEEE Transactions on Visualization and Computer Graphics, 8(1), 1–8. doi:10.1109/2945.981847
- [8] Dang, N., Yang, F., Xiao, B., & Zhu, Y. (2009). WebScope: A new tool for fusion data analysis and visualization. 2009 16th IEEE-NPSS Real Time Conference, 141–143. doi:10.1109/RTC.2009.5322112
- [9] Ramloll, R., Washington, D. C., Trepagnier, C., & Sebrechts, M. (n.d.). Gaze Data Visualization Tools : Opportunities and Challenges, 1–8.
- [10] Van Ham, F., & Van Wijk, J. J. (2004). Interactive Visualization of Small World Graphs. IEEE Symposium on Information Visualization, 199–206. doi:10.1109/INFVIS.2004.43
- [11] Van Dijck, J. (2013). “You have one identity”: performing the self on Facebook and LinkedIn. Media, Culture & Society, 35(2), 199–215. doi:10.1177/0163443712468605
- [12] Cardoso, J. S., Sousa, R., & Domingues, I. (2012). O Guan, T., & Feng, B. (2004). Fuzzy clustering of incomplete nominal and numerical data. Fifth World Congress on Intelligent Control and Automation (IEEE Cat. No.04EX788), 3(1), 2331–2334. doi:10.1109/WCICA.2004.1342007
- [13] Shih, J. (2011). Rating and Financial Ratio Data.
- [14] M. Plonsky, 2011, Psych. Statistics Graphing.htm, <http://www4.uwsp.edu/psych/stat/4/graphing.htm> Accessed by 01/03/2013
- [15] rdinal Data Classification Using Kernel Discriminant Analysis: A Comparison of Three Approaches. 2012 11th International Conference on Machine Learning and Applications, 473–477. doi:10.1109/ICMLA.2012.86
- [16] Houghton Mifflin Company, Grade 5 Data and Graphs Overview.htm, http://www.eduplace.com/math/mw/background/5/06a/te_5_06a_overview.html Accessed by 01/03/2013
- [17] Peng, W., & Li, T. (2006). Interval Data Clustering with Applications. 2006 18th IEEE International Conference on Tools with Artificial Int Natural Resources Canada, 2001, Carto Corner - Cartographic Symbology.htm, http://www.geography.siu.edu/people/oyana/Teaching/381_550/symbol.html Accessed by 01/03/2013

- [18] Total Medicare Private Health Plan Enrollment, 2000-2010 - Kaiser Slides.htm, <http://facts.kff.org/chart.aspx?cb=58&sctn=165&ch=1757> Accessed by 01/03/2013
- [19] elligence (ICTAI'06), 355–362. doi:10.1109/ICTAI.2006.71
- [20] Wikimedia project,2012,File Data plot women weight vs height.jpg - Wikimedia Commons.htm
http://commons.wikimedia.org/wiki/File:Data_plot_women_weight_vs_height.jpg Accessed by 01/03/2013
- [21] Cowperthwaite, D. J., & Fracchia, F. D. (1997). Distortion Viewing from 2D to 3D, (August), 42–51.
- [22] Kang, H., & Shneideman, B. (2000). Visualization Methods for Personal Photo Collections : Browsing and Searching in the PhotoFinder, 00(c), 1539–1542.
- [23] Nist Sematech, 1.3.3.26. Scatter Plot.htm, <http://www.itl.nist.gov/div898/handbook/eda/section3/scatterp.htm> Accessed by 05/03/2013
- [24] Scatter Plots and Trend Lines.htm, https://gcps.desire2learn.com/d2l/lor/viewer/viewFile.d2lfile/6605/5579/ScatterPlotLessonPart1_print.html Accessed by 05/03/2013
- [25] Trémeau, a, & Colantoni, P. (2000). Regions adjacency graph applied to color image segmentation. IEEE transactions on image processing : a publication of the IEEE Signal Processing Society, 9(4), 735–44. doi:10.1109/83.841950
- [26] Systat SoftwareInc. 2013, Dynamic and Sophisticated Graphs Made Easy with SigmaPlot!.htm, <http://www.sigmaplot.co.uk/products/sigmaplot/graph.php> Accessed by 05/03/2013
- [27] Liu, J. Y., Xie, M., Goh, T. N., & Ranjan, P. (2004). Time-Between-Events Charts for On-line Process Monitoring, 1061–1065.
- [28] Scatter Plots and Trend Lines.htm, https://gcps.desire2learn.com/d2l/lor/viewer/viewFile.d2lfile/6605/5579/ScatterPlotLessonPart1_print.html Accessed by 05/03/2013
- [29] Franklin, K. M., & Roberts, J. C. (2003). Pie chart sonification. Proceedings on Seventh International Conference on Information Visualization, 2003. IV 2003., 4–9. doi:10.1109/IV.2003.1217949
- [30] Robert I. Kabacoff, 2012, Quick-R Pie Charts.htm, <http://www.statmethods.net/graphs/pie.html> Accessed by 05/03/2013
- [31] Transweb Global Inc, 2007, Bar Charts Help, Get Tutors for Bar Charts Homework - Transtutors.com.htm, <http://www.transtutors.com/homework-help/statistics/describing-data-visually/types/bar-charts/> Accessed by 05/03/2013
- [32] Plaisant, C., Grosjean, J., & Bederson, B. B. (2002). SpaceTree: supporting exploration in large node link tree, design evolution and empirical evaluation. IEEE Symposium on Information Visualization, 2002. INFOVIS 2002., 2002, 57–64. doi:10.1109/INFVIS.2002.1173148
- [33] Mrs. Grieser's Algebra Wiki WikiGrieser Probability Tree Diagrams.htm, <http://mrsgalgebra.pbworks.com/w/page/28245155/Probability%3A%20%20Tree%20Diagrams> Accessed by 05/03/2013
- [34] Borland, D., & Taylor, M. R. (2007). Rainbow color map (still) considered harmful. IEEE computer graphics and applications, 27(2), 14–7. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/17388198>

- [35] Pennsylvania State University, 1998, NCHS Research Visualizing Health Statistics.htm
<http://www.geovista.psu.edu/grants/nchs/reliability.htm> Accessed by 05/03/2013
- [36] Nagel, W. E., Arnold, A., & Weber, M. (n.d.). VAMPIR : Visualization and Analysis of MPI Resources 2 The Message Passing Interface (MPI), 1–12.
- [37] Depends, I. A. (2009). Data Security in the World of Cloud Computing, 61–64.
- [38] Fearlessflyer.com, 2013, Reportr.net » What a word cloud says about this blog.htm
<http://www.reportr.net/2008/08/18/what-a-word-cloud-says-about-this-blog/> Accessed by 05/03/2013
- [39] Nagel, H. R. (n.d.). Scientific Visualization versus Information Visualization, 2–5.
- [40] Nagel, W. E., Arnold, A., & Weber, M. (n.d.). VAMPIR : Visualization and Analysis of MPI Resources 2 The Message Passing Interface (MPI), 1–12.
- [41] Fernando, J., Jr, R., & Traina, A. J. M. (n.d.). Enhancing Data Visualization Techniques.
- [42] Bruls, M., Huizing, K., & Wijk, J. J. Van. (n.d.). Squarified Treemaps
- [43] Mazza, R. (2004). Using Information Visualisation to Facilitate Instructors in Web-based Distance Learning A dissertation presented by, (June).
- [44] Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., Marchal, G., Leuven, K. U., et al. (n.d.). AUTOMATED MULTI-MODALITY IMAGE REGISTRATION BASED ON INFORMATION THEORY.
- [45] Elmqvist, N., Dragicevic, P., & Fekete, J.-D. (2008). Rolling the dice: multidimensional visual exploration using scatterplot matrix navigation. *IEEE transactions on visualization and computer graphics*, 14(6), 1141–8. doi:10.1109/TVCG.2008.153
- [46] Kang, H., & Shneiderman, B. (2000). Visualization Methods for Personal Photo Collections : Browsing and Searching in the PhotoFinder, 00(c), 1539–1542.
- [47] Laplacians, S., & Graphs, L. (2005). AMS Mathematics Subject Classification (2000): 05C50 Key Words: graph theory, graph spectra, line graph , signless Laplacian 1.
- [48] North, S. C., & Hill, M. (n.d.). Applications of Graph Visualization.
- [49] Viégas, F. B., Wattenberg, M., Ham, F. Van, Kriss, J., & Mckeon, M. (2007). Many Eyes : A Site for Visualization at Internet Scale, 13(6), 1121–1128.
- [50] Overbye, T. J., & Weber, J. D. (2000). Visualization of Power System Data, 00(c), 1–7.
- [51] Bostock, M., & Heer, J. (2009). Protovis: a graphical toolkit for visualization. *IEEE transactions on visualization and computer graphics*, 15(6), 1121–8. doi:10.1109/TVCG.2009.174
- [52] Jones, J. A., Harrold, M. J., & Stasko, J. (n.d.). Visualization of Test Information to Assist Fault Localization Georgia Institute of Technology, 467–477.
- [53] Momke, J. (2003). P. J. D. and D.
- [54] Kerr, B. (2003). Thread Arcs: an email thread visualization. *IEEE Symposium on Information Visualization 2003 (IEEE Cat. No.03TH8714)*, 2003, 211–218. doi:10.1109/INFVIS.2003.1249028
- [55] Chiara, R. De, & Fish, A. (2007). EulerView: a non-hierarchical visualization component. *IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC 2007)*, 145–152. doi:10.1109/VLHCC.2007.42
- [56] Nguyen, Q. V. (2002). A space-optimized tree visualization. *IEEE Symposium on Information Visualization, 2002. INFOVIS 2002.*, 2002, 85–92. doi:10.1109/INFVIS.2002.1173152

- [57] Chiara, R. De, & Fish, A. (2007). EulerView: a non-hierarchical visualization component. IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC 2007), 145–152. doi:10.1109/VLHCC.2007.42
- [58] Obayashi, S., & Sasaki, D. (2003). Visualization and Data Mining of Pareto Solutions Using Self-Organizing Map, 796–809.
- [59] Kaser, O., & John, S. (2007). Tag-Cloud Drawing : Algorithms for Cloud Visualization.
- [60] Hearst, M. a., & Rosner, D. (2008). Tag Clouds: Data Analysis Tool or Social Signaller? Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008), 160–160. doi:10.1109/HICSS.2008.422
- [61] Lee, B., Riche, N. H., Karlson, A. K., & Carpendale, S. (2010). SparkClouds: visualizing trends in tag clouds. IEEE transactions on visualization and computer graphics, 16(6), 1182–9. doi:10.1109/TVCG.2010.194
- [62] Persson, K. (n.d.). Techniques for Requirements Gathering and Definition Requirements Lifecycle Management.
- [63] May, D. R. (2009). Html 5, (May).
- [64] Hutchison, D., & Mitchell, J. C. (1973). The Semantic Web – ISWC 2006.
- [65] Hutchison, D., & Mitchell, J. C. (1973). The Semantic Web - ISWC 2009.
- [66] Jinwuli Developing knowledge models of social media over world wide web 2013 Master Thesis
- [67] Beautiful HTML5 & JS Charts CanvasJS.htm, <http://canvasjs.com/> Accessed by 25/08/2013
- [68] Lam, H., Bertini, E., Isenberg, P., Plaisant, C., Carpendale, S., & Lam, H. (2011). Seven Guiding Scenarios for Information Visualization Evaluation Seven Guiding Scenarios for Information Visualization Evaluation.
- [69] Finstad, K. (2006). The System Usability Scale and Non-Native, 1(4), 185–188.
- [70] Brooke, J. (n.d.). SUS - A quick and dirty usability scale.
- [71] Lin, H. K., Hsieh, M., Wang, C., Sie, Z., & Chang, S. (2011). ESTABLISHMENT AND USABILITY EVALUATION OF AN INTERACTIVE AR, 10(4), 181–187.

8. Appendix

8.1. Appendix 1:

1.1 User requirements gathering-Government scenario

1. Do you usually use the internet to get work-relevant information?
 Yes
 No

2. Do you have a LinkedIn account?
 Yes
 No

3. Do you believe data extracted from LinkedIn could potentially be helpful?
 Yes
 No

4. Have you already organized an official website for job seekers?
 Yes
 No
 Perspective

5. Select the from the following key words the area most relevant to your work
 Locations
 Industrial classify
 Company size
 Education structure
 Offered job titles
Please specify

6. Select the key words that most accurately reflect your involvement with the organization of Job seekers' and employee's data.
 Education background
 Personal Skills
 Certificates
 Work experience
 Gender
 Age
 Nationality

Please specify

7. What is the instrument you are using to manage the work relevant data?

Online internal data management system

Internal localization designed database

Documents

Manual operation

Please specify

8. What is the instrument you using to analysis the work relevant data?

Online internal data management system

All depend on statistical department

Search separately on the internet and gathering the result together

Please specify

9. How do you deal with the huge volume of data?

Please specify

10. Have you find out a better way to make the complex data clear>

Yes

No

Please specify

8.2. Appendix 2:

User requirements gathering- HR scenario

1. Do you usually use internet to get work-relevant information?
 - Yes
 - No

2. Do you have a LinkedIn account?
 - Yes
 - No

3. Do you believe the data which extracted from LinkedIn which will be helpful?
 - Yes
 - No

4. What do you usually do when you recruit new staffs and notice the job seeker?
 - Prepare the campus speech
 - Advertise the job vacation on the internet
 - Search on LinkedInPlease specify

5. Select the following key words you are offering on the internet when you are recruiting new staffs in terms of the your company
 - Locations
 - Industrial classify
 - Company size
 - Education structure
 - Offered job titlesPlease specify

6. Select the following key words you are concerning when you are recruiting new staffs in terms of job seekers' personal information
 - Education background
 - Personal Skills
 - Certificates
 - Work experience
 - Gender
 - Age
 - NationalityPlease specify

7. What is/are the way/ways you are using to make preparation for recruiting new staffs?
 - LinkedIn or other SNS
 - Contacting the Career departments of Colleges
 - No preparation just advertise the recruited information

Please specify

8. What do you usually want to do when you gathering the huge volume data from the internet?
- Leave it to the specific staff to analysis
 - Find a better way to make the complex data clear
 - Manually classify the data and analysis it later

Please specify

9. Have you ever heard data visualization?

- Yes
- No

10. If you can search all the information you want in one website and it will help you to deal with the questions you concern, do you want to try it?

- Yes
- No

8.3. Appendix 3:

User requirements gathering- Job seeker and College perspective student scenario

1. Do you usually use internet to get work-relevant information?
 Yes
 No

2. Do you have a LinkedIn account?
 Yes
 No

3. Do you believe the data which extracted from LinkedIn which will be helpful?
 Yes
 No

4. What do you usually do when you are looking for a new job?
 To know the general employment situation
 To know the general requirements for the jobs you are looking for
 To search on the job seeking internet
 To search on the enterprise's website
 Search on LinkedIn
Please specify

5. Select the following key words you are concerning about when you are looking for a new job in terms of the company
 Locations
 Industrial classify
 Company size
 Education structure
 Offered job titles
Please specify

6. Select the following key words you could offer of your personal information when you are looking for a new job
 Education background
 Personal Skills
 Certificates
 Work experience
 Gender
 Age
 Nationality
Please specify

7. What do you usually do after you searched the relevant job information?
 Manually make some comparisons among different jobs
 Analysis them individually rely on the information from the internet

Use online system to manage them

Please specify

8. Have you ever heard data visualization?

Yes

No

9. If you can search all the information you want in one website and it will help you to deal with the questions you concern, do you want to try it?

Yes

No

10. What is/are the most important question/questions you want to know by the new website which mentioned in the previous one?

Please specify

8.4. Appendix 4:

Regular expression for retrieving the data from Json

Figure 39: Sample code of regular expression

```
function getName(urlText){
var regex = new RegExp("[a-zA-Z][a-zA-Z0-9\\_]*$");
var m = regex.exec(urlText)
```

8.5. Appendix 5:

Set up a post with Json result format

Figure 40: Sample code of Ajax

```
xmlhttp.open('POST', endpoint, true); // GET can have caching probs, so POST
xmlhttp.setRequestHeader('Content-type', 'application/x-www-form-urlencoded');
xmlhttp.setRequestHeader("Accept", "application/sparql-results+json");

// Set up callback to get the response asynchronously.
xmlhttp.onreadystatechange = function () {
    if (xmlhttp.readyState == 4) {
        if (xmlhttp.status == 200) {
            // Do something with the results
            if (isDebug)
                //alert(xmlhttp.responseText);
                callback(xmlhttp.responseText, "from");
        } else {
            // Some kind of error occurred.
            alert("Sparql query error: " + xmlhttp.status + " "
                + xmlhttp.responseText);
            $('#from1Loading').hide();
        }
    }
};
```

8.6. Appendix 6:

Case classification (according to the requests users make in front end)

Figure 41: Sample code of case classification

```
if (numberOfCityChosen >= 1 && numberOfDegreeChosen == 0 && numberOfSkillChosen == 0 && numberOfWorkExpChosen == 0) {
    //alert(1000);
    case_num = 0;

    } else if (numberOfCityChosen >= 1 && numberOfDegreeChosen >= 1 && numberOfSkillChosen == 0 && numberOfWorkExpChosen == 0) {
    case_num = 1;
    //alert(1100);
    } else if (numberOfCityChosen >= 1 && numberOfDegreeChosen == 0 && numberOfSkillChosen >= 1 && numberOfWorkExpChosen == 0) {
    case_num = 2;
    alert(1010);
    } else if (numberOfCityChosen >= 1 && numberOfDegreeChosen == 0 && numberOfSkillChosen == 0 && numberOfWorkExpChosen >= 1) {
    case_num = 3;
    alert(1001);
    } else if (numberOfCityChosen >= 1 && numberOfDegreeChosen >= 1 && numberOfSkillChosen >= 1 && numberOfWorkExpChosen == 0) {
    case_num = 4;
    alert(1110);
    } else if (numberOfCityChosen >= 1 && numberOfDegreeChosen >= 1 && numberOfSkillChosen == 0 && numberOfWorkExpChosen >= 1) {
    case_num = 5;
    alert(1101);
    } else if (numberOfCityChosen >= 1 && numberOfDegreeChosen == 0 && numberOfSkillChosen >= 1 && numberOfWorkExpChosen >= 1) {
    case_num = 6;
    alert(1011);
    } else if(numberOfCityChosen == 0 && numberOfDegreeChosen >= 1){
    case_num = 7;
    }else {
    case_num = -1;
    }
```

8.7. Appendix 7:

Visualize the data which I got from Json

Figure 42: Sample code of data visualization

```
function callback0001_2(str, row) {

    console.log(str);
    var mydata = JSON.parse(str);
    var bindings = mydata.results.bindings;

    var dps2 = [];

    for (i = 1; i < bindings.length; i++) {

        var obj = {
            label : getSkillName(bindings[i].degree.value),
            y : parseInt(bindings[i].oCount.value)
        };
        dps2.push(obj);

    }
    var cycChart = new CanvasJS.Chart(("canvas" + (3 * row + 1)), {
        title : {
            text : "The degree distribution."
        },
        data : [{
            type : "column",
            dataPoints : dps2
        }
    ]
    });

    cycChart.render();

    xmlhttp.send(querypart);
```