

# An Exploratory Curriculum Study of Master Programs in Data Analytics

Tai-Hua Chen, San-Yih Hwang  
College of Management  
National Sun Yat-sen University  
Kaohsiung, Taiwan  
ctaihua.home@gmail.com, syhwang@mis.nsysu.edu.tw

## Abstract

According to a report by McKinsey & Company and IDC published in 2011, the shortage of Big Data or Data Analytic talent for the analytical technicians and data-literate managers in US would reach 190,000 and 1.5 million respectively, which is about 50 to 60 percent higher than its regular supply in 2018. Universities around the globe have found this trend and start to catch up. Within the last few years, many universities have launched Data Analytics master programs. Due to the cutting-edge nature of data analytics, no common course structure is available. This research presents a survey of curriculums of 79 Data Analytics master programs. Through the analysis, we provide an overview of current program competencies and curriculums.

**Keywords:** Big Data, Business Analytics, Data Analytics Programs, Curriculum design

## 1. Introduction

Information Technology has become one of the most significant infrastructures for many businesses. For example, they can use the ERP system to track the business processes whenever they want. In our daily life, people can reach their friends through Social Networking Services with a simple touch on cutting-edge mobile devices. Thus, a massive amount of complex data are stored in servers and distributed on the Internet. The size of data is too big to be handled by regular technologies and applications. As a result, the term, Big Data, was proposed to refer to the huge amount of data and the advanced techniques used to analyze them. These analytical techniques aim to reduce the operational cost, minimize their losses and increase the returns from investing on the analytical tasks.

It is commonly recognized that the growth of data is much faster than our abilities to analyze them, as witnesses by Tim Waddell (2014): “What we don’t need is more data, we need actionable data.” Nowadays, enterprises do not put their concerns on the size of their collected data. They want to get the forecast information for their business. They also want to know reasons for the forecast answers. Analysis

methods can help enterprises to reach that target.

How to equip enterprise with that kind of weapon? They need to find staffs with data analysis talent, called data scientists in recent press conferences. As a result, in recent years, quite a few master programs in data analytics around the globe are established. We collect and analyze the curriculum and competency information of 79 data analytics programs, and identify the trends revealed by these programs. The research result could be beneficial for those who intend to establish a new Data Analytics master program.

This paper is organized as follows. In Section 2, we give some background information about data analytics and its programs. In Section 3, we describe our research methodology. In Section 4, we present our survey results from different perspectives. Finally, Section 5 summaries our findings.

## 2. Background

In this section, we describe some fundamental knowledge of data analytics master programs.

### A. *Volume, Variety and Velocity*

Big Data are commonly characterized by three characteristics, namely volume, variety and velocity (3Vs). Volume refers to the enormous size of Data. Variety refers to the complex types of data. Velocity represents the frequency of Data produced, which means the flow of data is continuous without any interruption.

### B. *Without Analytics, Big Data is Just Noise*

In the old days, because the amount of data is less and only represents sampling data, people often use statistics techniques on these sample data, often resulting in misleading information. Fortunately, things have changed with modern technologies. The size of data is not an issue anymore. However, analytical problems arise with tremendous staved flying data. Because the stored enormous data are massive, noisy and un-organized, the raw data is not worth for anybody.

Data must be re-assembled to generate worthy

information for mankind. This data-transformation process is called data analysis. Enterprises can get the summary of customer opinions from analysis results. They can execute the best business strategies based on those outcomes. The perspective of analysis can be categorized into three types, which are descriptive analytics, predictive analytics and prescriptive analytics.

In November of 2011, Larry Feinsmith, the CIO of JPMorgan, announced that they will pay 10% more than other guys for Hadoop expert hiring. This is because JPMorgan found that Hadoop technology has the potential to gain lots of insights from the huge amount of data they collected in a way which was not possible before. This is the remarkable example for Big Data experts wanted.

According to the Big Data report of McKinsey & Company, Big Data expert workforce shortages will not be reduced until 2018. The United States alone would confront the Big Data talent shortage of 190,000 workers and 1.5 million managers. Big Data talents required knowledge.

Data Science is a generic description of Big Data talents required knowledge. It is the knowledge about discovering the objective results from the widespread and ubiquitous data. The set of knowledge can be divided into four areas, namely hacking skills, domain expertise, statistics, and society sciences (The Fourth Bubble in the Data Science Venn Diagram: Social Sciences, 2014). The term of hacking refers to required skills and knowledge for analyzing and interpreting the large amounts of data. The required skills and knowhow of data science are shown in Figure 1.

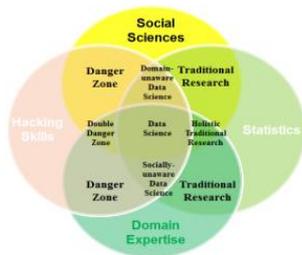


Figure 1. Required knowledge of Data Science

### C. Data Science required skills

The first skill is about finding required information in enormous data set. The second skill is the ability of defining a solid algorithmic for data analysis. Data scientists need to categorize and recompose the massy data into significant information for business decision usages. The third one are result generation and communication abilities. They have to clearly express the analysis result to all stakeholders. The conceptual diagram of skills for Data Scientists is shown in Figure 2.

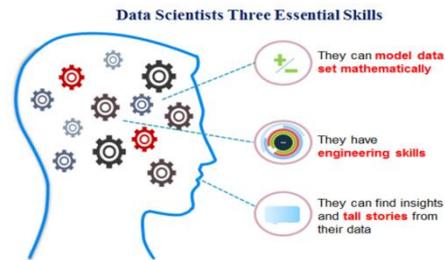


Figure 2. Three essential skills of data scientists

## 3. Methodology

To gain the insights about the newly established data analytics programs, we need a method. The method can simply be divided into three steps. First, we need to collect data about these programs. Second, the collected data need to be categorized into different types. We hire experts to perform coding for this step. Finally data analysis from different perspectives can be performed. The entire framework is shown in Figure 3.

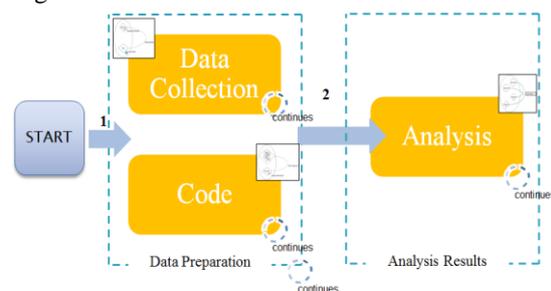


Figure 3. Data Analysis Framework

### A. Data Collection

Data collection is to obtain relevant data. As indicated earlier, the purpose of this study is to represent the statistic results of Master of Science programs on Data Analytics (MSDA), especially on their courses and competencies. Fortunately, a non-profit organization, KDnuggets™ with the mission of interconnecting data mining research community, has collected information about MSDA programs (<http://www.kdnuggets.com>). From KDnuggets, we can identify seventy nine MSDA programs' web pages. We retrieved the contents of these program web pages that describe the program competencies and courses information until July 2014. An example MSDA program web page is displayed in Figure 4. The courses and competencies samples are displayed in Table 1.

Table 1. Collected course and competency sample

<b>University name</b>		Southern Methodist University		
<b>Degree</b>		Master of Science in Business Analytics		
<b>Duration(Months)</b>		12		
<b>Department</b>		Business		
<b>Program start</b>		2014		
<b>Region</b>		USA		
Curriculum				
<b>Name</b>	<b>Field</b>	<b>Credit</b>	<b>Type</b>	<b>Description</b>
Managing Your Career	Business	1	C	This course empowers MSBA students with the knowledge and tools to effectively manage their own careers. Topics include, but not limited to: exploring career opportunities within the Business Analytics space, building and leveraging a professional network and developing a personal marketing plan.
<b>University Name</b>	<b>Program Name</b>	<b>Competency</b>		<b>Ref Web Page</b>
1	Central Connecticut State University (CCSU)	Master of Science in Data Mining	Be able to approach data mining as a process, by demonstrating competency in the use of CRISP-DM, the Cross-Industry Standard Process for Data Mining, including the business understanding phase, the data understanding phase, the exploration or analysis phase, the modeling phase, the evaluation phase, and the deployment phase.	<a href="http://web.ccsu.edu/datamining/learning_outcomes.html">http://web.ccsu.edu/datamining/learning_outcomes.html</a>
		Master of Science in Data Mining	Be proficient with leading data mining software, including WEKA, Clementine by SPSS, and the R language.	
		Master of Science in Data Mining	Understand and apply a wide range of clustering, estimation, prediction, and classification algorithms, including k-means clustering, BIRCH clustering, Kohonen clustering, classification and regression trees, the C4.5 algorithm, logistic Regression, k-nearest neighbor, multiple regression, and neural networks.	

Business, Information Technology, Data Science, and Capstone. Table 2 shows some example course instances, their course modules, and course types.

Table 2. Course categories sample

Course Name	Course Module	Course Type
Core CA674 Cloud Architectures	Information Technology	Architecture
CS 246 Advanced Computer Architecture	Information Technology	Architecture
Enterprise Systems and Architecture	Information Technology	Architecture
MIS 730 Integrating IT Architecture	Information Technology	Architecture
Service Oriented Architectures	Information Technology	Architecture

Table 3. Program competency categories sample and results

Technology	Description	
T1	Be proficient with data mining tools	
T2	Be able to master the data analysis process	
T3	Understand the capabilities of various data mining methods, such as clustering, prediction, association, relevance, and text mining	
T4	Have a broad understanding about relevant information technologies, including programming, databases, software engineering, networking, etc.	
No	Competency	Type
CP1	Be able to approach data mining as a process, by demonstrating competency in the use of CRISP-DM, the Cross-Industry Standard Process for Data Mining, including the business understanding phase, the exploration or analysis phase, the modeling phase, the evaluation phase, and the deployment phase.	T2
CP2	Be proficient with leading data mining software, including WEKA, Clementine by SPSS, and the R language.	T1
CP3	Understand and apply a wide range of clustering, estimation, prediction, and classification algorithms, including k-means clustering, BIRCH clustering, Kohonen clustering, classification and regression trees, the C4.5 algorithm, logistic Regression, k-nearest neighbor, multiple regression, and neural networks.	T3

In addition, each MSDA program may have some program capability descriptions. Again, these descriptions may differ, and we apply coding techniques to summarize them into a number of competences. We have employed one senior scholar in data analytics to code curriculum names and competencies. Expert has defined five competencies modules, which are Technology, Application, Theory, Management Issues, Soft Skills and Others. The program competences sample results are shown in Table 3. Competency Module details can be found in Table 4.

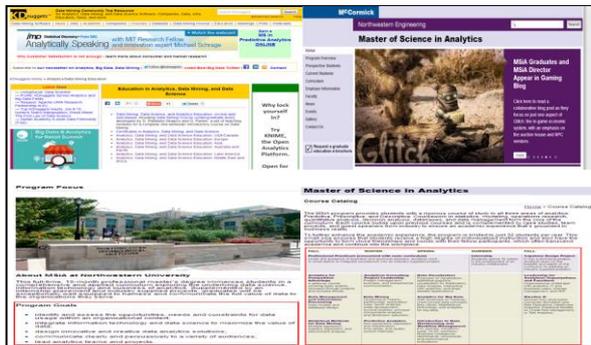


Figure 4. An Example MSDA web site

B. Coding

Coding is a process used by experts to quickly recognize some specific information. For example, someone could assign a specific key-word on any music for searching. This study applies coding techniques to categorize courses and program competences. As can be imaged, courses offered by different MSDA programs may have different course names even though they are highly similar. To prevent data explosion and to make sense of data, we have carefully examine the courses listed for each MSDA program and categorize courses into a small number of course types are indeed needed. In our collected MSDA programs, there are totally 1235 course instances, which are grouped into 84 course types, where courses in the same course type can be considered the same. We further group several course types into five course modules, namely Application,

Table 4. Competency Modules

Module Name	Description
Technology	This module contains all aspects of Information Technology which includes concepts about data mining, tools, programming, etc.
Application	This module is talking about introducing or expressing some specific domain knowledge.
Theory	This module mainly related to all concepts of computing and statistics.
Management Issues and Soft Skills	This module mainly related to all concepts of communication, ethics and management.
Others	All un-recognized competencies would be assigned in this module. Currently, 'Others' is something with certification and accreditation.

C. Data Analysis

Data analysis is a statistic technology which is used to investigate and measure the results of program basic information, courses coding and competencies coding. This research applies the SpagoBI for statistical data analysis (<http://www.spagoworld.org/xwiki/bin/view/SpagoBI/>).

Table 5. MSDA programs by locations

		Measures
Country		Count
All Countries		79
Asia		2
Australia		1
Belgium		1
Canada		2
EU		2
France		2
Hungary		1
Ireland		4
Latin America		1
Middle East and Africa		1
Spain		2
Sweden		1
UK		9
USA		50

## 4. Results

### A. Statistics about MSDA programs

Table 5 shows the MSDA programs by locations. As can be seen, about two thirds of the MSDA programs are launched by universities in USA. Only two MSDA programs are in Asia, and one in Australia. It is evident that western countries are much more enthusiastic in establishing MSDA programs.

### B. Statistics about MSDA program competencies

The MSDA programs may have different competencies that involving five different modules and Figure 5 display their distributions. It shows that technology and application are the top two competency modules emphasized by MSDA programs.

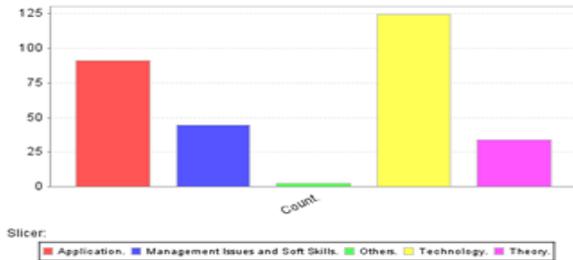


Figure 5. Distribution of modules in MSDA programs

#### 1) Distribution of Technology module

As technology competency module is mostly emphasized, we then look into the more detailed technology competencies. As can be seen from Figure 6, T1 and T3, which are “Be proficient with data mining tools,” and “Understand the capabilities of various data mining methods, such as clustering, prediction, association, relevance, and text mining,” respectively are mostly mentioned.

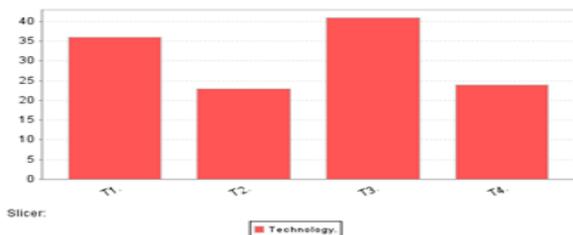


Figure 6. Distribution of competencies in the technology module

#### 2) Distribution of Application module

Figure 7 shows the distribution of competencies for application module. The top two MSDA application competencies are A2 and A3, which are “Have concentration on business data analytics” and “Be able to analyze large data sets in a variety of industries,” respectively.

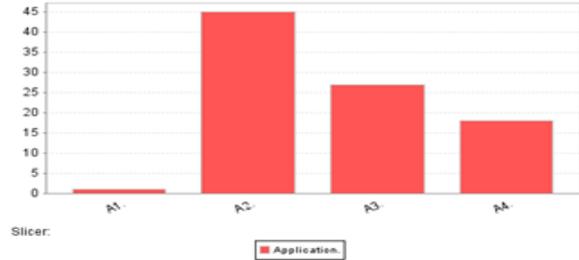


Figure 7. Distribution of competencies in the application module

#### 3) Distribution of Management Issues and Soft Skills module

Figure 8 shows the two top MSDA competencies in management issues and soft skills modules: M3 and M1, which are “Understand management issues such as leadership, marketing, human resources, and project management” and “Be able to communicate across disciplines and collaborate in a team” respectively.

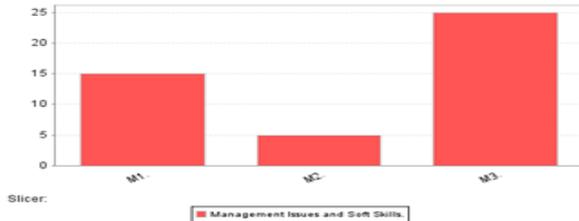


Figure 8. Distribution of competencies in the management issues and soft skills module

#### 4) Distribution of Theory module

Figure 9 shows the top MSDA competency in theory module is R2, which is “Be able to understand the foundation behind statistics, modeling, simulation, and optimization techniques”.

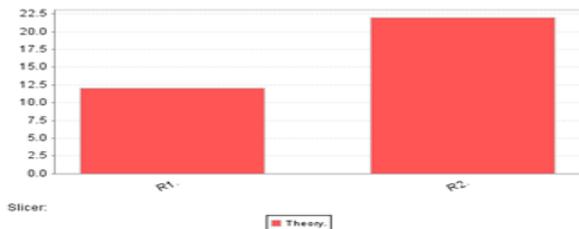


Figure 9. Distribution of competencies by the theory module

### C. Statistics about MSDA program curriculums

The MSDA program contains five different course modules, which are Business, Information Technology, Data Science, Capstone, and Application. The results displayed in Figure 10 indicate that Business and Information Technology are the two most popular modules in MSDA programs.

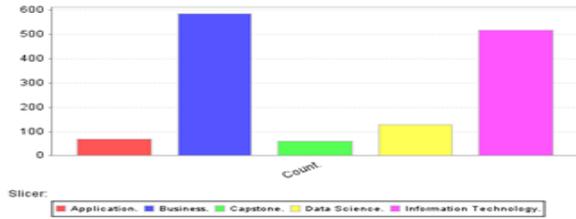


Figure 10. Distribution of MSDA course modules

#### 1) Distribution of Information Technology module

It can be indicated in Figure 11 that the two most popular course types in the information technology module are data mining and database management.

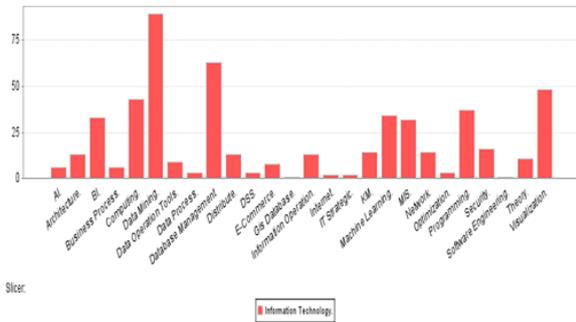


Figure 11. Distribution of MSDA course types in the information technology module

#### 2) Distribution of Application module

It can be represented in Figure 12 that top two MSDA course types in application module are bioinformatics and supply chain.

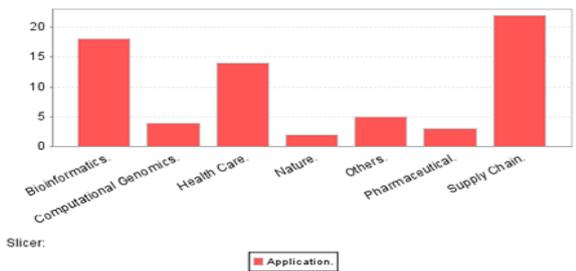


Figure 12. Distribution of MSDA course types in the application module

#### 3) Distribution of Business module

It can be presented in Figure 13 that top two MSDA course types in business modules are statistics and marketing.

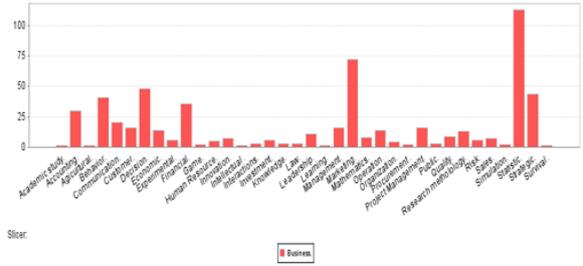


Figure 13. Distribution of MSDA course types in the business module

#### 4) Distribution of Data Science module

It can be seen in Figure 14 that the top two MSDA course types in data science module are big data and predictive analytics.

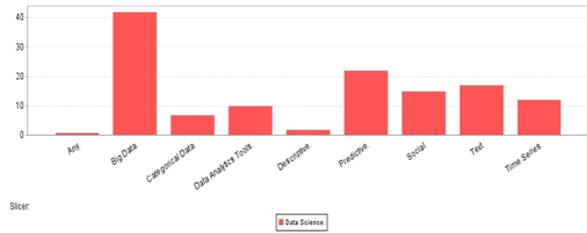


Figure 14. Distribution of MSDA course types in the data science module

#### 5) Distribution of Capstone module

It can be seen in Figure 15 that the top two MSDA course types in the capstone module are research project, workshop/seminar and intern.

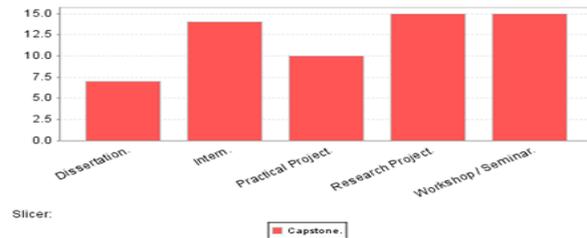


Figure 15. Distribution of MSDA course types in the capstone module

### D. Trend of MSDA programs

#### 1) Country distribution of MSDA degree

It can be seen in Figure 16 that most MSDA programs are offered by universities in western countries, where USA tops the list, followed by UK.

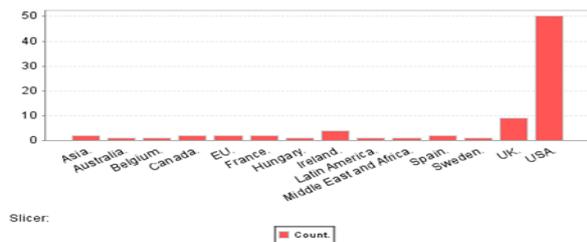


Figure 16. Distribution of MSDA programs by countries

## 2) Establishment Year of MSDA program

Figure 17 indicates that most MSDA programs were started after 2010.

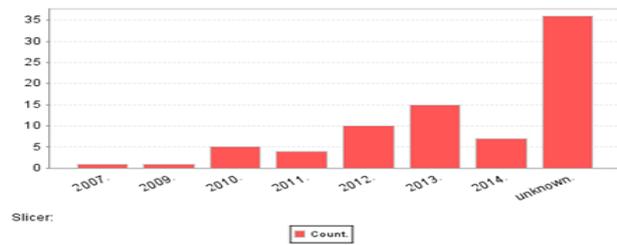


Figure 17. Distribution of MSDA programs by establishment years

## 3) Offering School of MSDA program

It indicates in Figure 18 that the two MSDA offering school types are business and information technology.

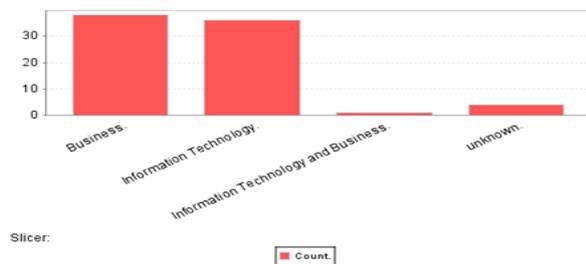


Figure 18. Distribution of MSDA programs by offering schools

## 5. Conclusions

According to the result section, we think the most impressive finding is that both Business and Technology are almost equal weight in MSDA program. This is because for a successful MSDA student, he or she needs to equip with both domain knowledge (especially business) and Information Technology. This is also witnessed by the fact that business school and information technology schools are most likely to offer MSDA programs. In addition, the number of MSDA programs keeps increasing in recently years.

## References

- [1] Viktor Mayer-Schönberger and Kenneth Cukier (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. New York, NY: Eamon Dolan & Houghton Mifflin Harcourt. pp.1-50, March 2013.
- [2] Choudaha, Rahul (2008). Competency-Based Curriculum for a master's program in service science. Management and Engineering (SSME): An online Delphi Study. Retrieved from Dissertations and Theses database. (UMI No. 304636413)
- [3] Structured Data, SQL and Unstructured Data (n.d.) Retrieved from [http://www.webopedia.com/TERM/S/structured\\_data.html](http://www.webopedia.com/TERM/S/structured_data.html)
- [4] Wikipedia. (n.d.). Semi-structured data. Retrieved from [http://en.wikipedia.org/wiki/Semi-structured\\_data](http://en.wikipedia.org/wiki/Semi-structured_data)
- [5] Wikipedia. (n.d.). Unstructured data. Retrieved from [http://en.wikipedia.org/wiki/Unstructured\\_data](http://en.wikipedia.org/wiki/Unstructured_data)
- [6] MbaLib. (n.d.). Unstructured Information. Retrieved from <http://wiki.mbalib.com/zh-tw/%E9%9D%9E%E7%BB%93%E6%9E%84%E5%8C%96%E4%BF%A1%E6%81%AF>
- [7] Fred Chiang. (2012). All about Hadoop and Big Data. Retrieved from <http://fredbigdata.blogspot.tw/2012/08/blog-post.html>
- [8] Hortonwork. (2012). 7 Key Drivers for the Big Data Market. Retrieved from <http://hortonworks.com/blog/7-key-drivers-for-the-big-data-market/>
- [9] Rosebt. (2013). Descriptive Diagnostic Predictive Prescriptive Analytics. Retrieved from <http://www.rosebt.com/blog/descriptive-diagnostic-predictive-prescriptive-analytics>
- [10] Andrew Tarantola. (2013). How Prescriptive Analytics Could Harness Big Data to See the Future. Retrieved from <http://datascienceasn.org/content/fourth-bubble-data-science-venn-diagram-social-sciences>
- [11] Michaelmalak. (2014). The Fourth Bubble in the Data Science Venn Diagram: Social Sciences. Retrieved from <http://www.boxuk.com/blog/importance-domain-knowledge/>
- [12] Carey Hiles. (2013). The importance of domain knowledge. Retrieved from <http://www.boxuk.com/blog/importance-domain-knowledge/>
- [13] Corey Vilhauer. (2011). DOMAIN KNOWLEDGE: WHAT YOU NEED – OR DON'T NEED – TO KNOW. Retrieved from <http://eatingelephant.com/2011/07/domain-knowledge/>
- [14] Dewey. (n.d.). Expertise and Domain Specific Knowledge. Retrieved from [http://www.intropsych.com/ch07\\_cognition/expertise\\_and\\_domain\\_specific\\_knowledge.html](http://www.intropsych.com/ch07_cognition/expertise_and_domain_specific_knowledge.html)
- [15] Project-Skills. (n.d.). Why Domain Knowledge Is Important In Project Management. Retrieved from <http://www.project-skills.com/why-domain-knowledge-is-important-in-project-management.html>
- [16] Thomas H. Davenport and D.J. Patil. (2012). Data Scientist: The Sexiest Job of the 21st Century. Retrieved from <http://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/ar/1>
- [17] Mark Rockwell. (2013). Gold in the data, but a shortage of miners. Retrieved from <http://fcw.com/articles/2013/09/17/big-data-analyst-shortage.aspx>
- [18] Stan Ahalt. (2013). The Big Data Talent Gap. Retrieved from <http://renci.org/wp-content/uploads/2013/04/The-Big-Data-Talent-Gap-White-Paper1.pdf>
- [19] J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
- [20] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in *Magnetism*, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
- [21] K. Elissa, "Title of paper if known," unpublished.
- [22] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.
- [23] M. Young, *The Technical Writer's Handbook*. Mill Valley, CA: University Science, 1989.
- [24] SpagoBI. (2014). SpagoBI, the only 100% open source, complete and flexible Business Intelligence suite. Available: <http://www.spagoworld.org/xwiki/bin/view/SpagoBI/>. Last accessed 20th June 2014.