**Meeting Minute 2**

|  |  |
| --- | --- |
| **Date:** | 2nd February 2015 |
| **Time:** | 10.00am – 11.00am |
| **Venue:** | SIS Meeting Room 4-3 |
| **Attendees:** | Prof Kam, Le Hoang Trinh, Zheng Tianwei |
| **Absentees:** | Nil |
| **Things discussed:** | 1. Team updates on progress    1. Team has crawled data from OneSource Database. There are 2 sets of data: company & executives.    2. For data cleaning, team has removed the duplicates found in dataset using Excel    3. Propose to use OneSource Classification Definition instead of SSIC Definition    4. Team made some effort to visualize a small sample of data (1000 rows), using NodeXL to see the connections among the variables 🡪 the real datasets are too big to run on NodeXL, team will try Gephi to visualize later. 2. Prof’s advices on team progress    1. The data is not completed. There is some fields missing that need to be added on, i.e. parent companies, date founded, etc. Advise team to get more related data from OneSource    2. Approve team to use the classification definition from OneSource    3. Data Preparation should not be underestimated. All steps performed in data preparation should be documented, i.e. describe what you have done with the data, to ensure data is ready & appropriate to be analysed.    4. Keep the raw data and the analyzing data separate. (i.e. in the analysis data only keep the relevant columns) 3. Data Preparation 4. Check if the data is consistent 5. Listing all the duplicates and choosing what to delete. Suggestion: check the names (if names being duplicated & not consistent in spelling or missing parts of the executive names) 6. Document: what are the changes, how many firms are duplicated, what are the criteria to decide which one to remove. By the end, how many rows of data do you get back? 7. Discuss during next meeting if there is any problem. 8. Overview of Firm Distribution 9. Distribution of company: by industry, by private/public company, listed/non-listed. E.g. general account of how many persons sits in 10 companies, etc. 10. Distribution of the board of executive (executive title) 11. Tools & Software   Use **JMP** instead of SAS. |

**Tasks**

|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **Task** | **Assigned Person** | **Due Date** |
| 1 | Update and refinement of proposal | All | 04.02.2015 |
| 2 | Re-extract data from OneSource | All (to be divided among team) | 05.02.2015 |
| 3 | Download & Install JMP | All | 06.02.2015 |
| 4 | Data Preparation – Clean Duplicates & Refine missing Data | All (to be divided among team) | 07.02.2015 |
| 5 | Overview Summary on Company Distribution | Trinh | 08.02.2015 |
| 6 | Overview Summary on Executive Distribution | Tianwei | 08.02.2015 |
| 7 | Update Wiki Page | Tianwei | 08.02.2015 |

The meeting was adjourned at 11.00am. These minutes will be circulated and adopted if there are no amendments reported in the next three days.

Prepared by,

Le Hoang Trinh

Vetted and edited by,

Zheng Tianwei