|  |  |
| --- | --- |
| Project Status Report |  |

# Project Name: Science & Technology Search Engine

February 8, 2017 Team Members: Liyi Li, Yunyan Wu, Jie He, Jinjin Xiao, Yi Luo

|  |  |
| --- | --- |
| The project, during the week of 1/31/2017 – 2/8/2017, is due to the following: | 1, Start to explore Docker,AWS,Spark  2, Compare pros and cons of Elasticsearch and Solr |
| Issues: | 1, Report output of domain list, urls and plaintext files of urls  2, Discuss data explosion problem  3, Plan cloud and spark solutions |
| Milestones accomplished the week of 1/31/2017 - 2/8/2017: | 1, Online meeting with UN officer- Mr. Jorge  2, Made Progress reporting  3, Discussed big data problem  4, Learned AWS services and decided to build Search Engine in Elasticsearch  5, Collected domain list for governments, UN, and online repositories  6, Based on domain list, extracted list of URLs, and plaintext from URLs |
| Milestones planned this week, but not achieved with variance: | 1, Data storage for plaintext of target URLs  2, Explore all potential cloud solutions for data explosion problem |
| Milestones planned for next week: | 1, Learn and try Docker tutorial  2, Learn and connect to AWS  3, Learn and Elasticsearch tutorial |
| Areas/questions for discussion: | what is the pipeline between these cloud platforms?  what is the estimated budget for cloud solutions?  what is the expected output and size for prototype SE based on the cloud services? |
| Last week’s issues forwarded to this week: |  |