

UNIVERSITY COLLEGE LONDON

Department of Computer Science

British Library Big Data Experiment – Week 9 Progress Report

Date: 4th August 2014
Course Instructor: Dr. D. Mohamedally
Module: COMPGS99:
MSc SSE/CS Project 2014

Authors:

Nektaria Stavrou

Stelios Georgiou

Wendy Wong

Stefan P. Alborzpour

1. Overview of Activities

LAST WEEK KEY ACTIVITIES

- Developed advanced search feature using Azure Search API.
- Prepared and produced a project video.
- Updated user interface design.
- Implemented Boolean operators for advanced search (AND, OR).

THIS WEEK KEY ACTIVITIES

- Integration of advanced search and Boolean functionalities from Azure Search with the user interface.
- Implementation of hint text for the advanced search interface.
- Begin an investigation of the traceability feature.
- Implementation of download feature for OCR text.
- Implementation of multiple download feature (OCR text files) as a zip file.
- Meeting (Skype) with James Baker to review project progress.

NEXT WEEK KEY ACTIVITIES

- Integration of download feature into user interface (results page).
- Integration of statistical analysis.
- Development of traceability feature.
- Run performance tests using HDInsight.
- Improvement to the user interface.

2. Deliverables

Description	RAG Status
Advanced Search Feature Set The advanced search facility utilises the Azure Search Service API for title, publisher, publication place, date and OCR text. These fields support multiple keyword input from the user returning the results in the order of relevance.	Green
Download Feature This feature enables the user to select and download the OCR text of a document. It also supports the download of multiple files at once as a zip file.	Green
Contribution Report This is a document which highlights the contribution of the British Library Big Data Experiment.	Green
Traceability feature Aspects of the implementation of this feature are currently being investigated.	Amber

3. Deliverables

Meeting 1 – Project progress review with James Baker

Date	Time	Location
Thursday 31 st August 2014	11:30 – 12:00	Malet Place Engineering Building, Lab 1.21

Attendees	James Baker (JB), Nektaria Stavrou (NS), Stelios Georgiou (SG), Wendy Wong (WW), Stefan P. Alborzpour (SA)
Apologies	
Minutes	Stelios Georgiou

Agenda Item – Points Discussed	Type Action, Decision or Info
1 – Progress overview	
1.01 NS outlined the current status of the project along with upcoming tasks for the following weeks.	Info
1.02 JB highlighted that the requirements document and project plan should be revised and reprioritised. The deadline is fast approaching and JB stressed that the team should decide which requirements will be implemented.	Action
1.03 Considering the objective of creating a user friendly web portal, WW proposed some possible designs to JB for the Boolean operations on the advanced search web page.	Info
2 – Data	
2.01 SG and NS raised an issue regarding the Azure Search Service API. The service, as it is in the current version (experimental), does not support nested JSON data. SG explained that the team is currently working with a small set of data (flat format) however, for the final product the original data must be adapted to a flat structure. JB assured the team that will contact Ben'Osteen (Technical Lead of BL Labs) regarding this issue.	Info
3 – User group	
3.01 WW explained that she has scheduled user groups for the week commencing 11-08-2014. JB advised WW to start approaching the members of the user group as soon as possible, since he will be away from 11-08-2014.	Action

New Action Items	Owner	Due Date
Invite focus group participants to user group (via e-mail).	WW	05-08-2014
Evaluate and revise the requirements document.	NS,SG,SA,WW	08-08-2014