

UNIVERSITY COLLEGE LONDON

Department of Computer Science

British Library Big Data Experiment – Week 6 Progress Report

Date: 14th July 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

- Implementation of the feature that retrieves the frequency of specific words in a document.
- Implementation of the feature that allows users to select how many words are to be displayed in a list of most frequent words from a document.
- Implementation of the feature that retrieves the frequency of a bigram and a trigram.
- Implementation of the feature that retrieves the adjacent word for the most frequent words from a document.
- Connection to the Azure Storage Account through Visual Studio.
- Established a backend that allows connection to database.
- Provisioned a HDInsight cluster.
- Computed a test case within HDInsight framework.
- Meeting with James Baker to review requirements and project progress.
- Meeting with Simon Julier (UCL Lecturer & Wendy's supervisor) to discuss project.
- Attended the British Library Labs project event.

THIS WEEK KEY ACTIVITIES

- Developed primary components of the advanced search feature. The user can now enter a keyword about a title, publisher or publication place and retrieve relevant items (RQ01).
- Creation of Azure Tables (test data) that are used for testing the components which have been implemented.
- Designed and implemented a prototype home page.
- Designed and implemented an advanced search page.
- Preparation for the mid-project presentation.
- Skype meeting with James Baker to review project progress.

NEXT WEEK KEY ACTIVITIES

- Deliver mid-project presentation.
- Further develop the advanced search feature (author, contributors and publication date).
- Implement tracking of steps.
- Meeting with Melissa Terras to review project progress.

2. Deliverables

Description	RAG Status
Advanced feature set The advanced feature set has been implemented, it offers the functionality of returning items based on the fields of title, publisher or publication place. An important characteristic of this feature is prioritization where the most relevant results are shown first.	Green
Home page development This aspect of the user interface has been implemented as an initial prototype. In its simplest form it includes the logo of the British Library and a text box for keyword search. Further iterations of the home page will be developed.	Amber
Advanced search page development This aspect of the user interface has been implemented as an initial prototype. It consists of an input form to collect parameters for querying the data. Further iterations of the advanced search page will be developed.	Amber

3. Minutes

Meeting 1 – Team progress review

Date	Time	Location
Wednesday 18 th June 2014	11:00 – 11:30	Malet Place Engineering Building, Lab 1.21

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

Agenda Item – Points Discussed	Type Action, Decision or Info
1 – Progress overview	
1.01 NS reviewed last week's key activities.	Info
1.02 NS outlined the agenda for the meeting, a discussion of upcoming key activities for this week along with their respective actions: <ol style="list-style-type: none"> 1. Implementation of the advanced search feature. 2. Implementation of the home page and advanced search page. 3. Preparation for the mid-project presentation. 	Action
2 – Advanced search feature	
2.01 NS and SG explained the build plan for implementing the advanced feature set.	Info
2.02 The process for creating azure tables using test data was described by SG. This data shall be used for testing implemented components.	Info
2.03 The user enters a keyword (one or more words) on those specific fields, and then the system returns the items that are relevant with his field search term.	Info
2.04 The most relevant results must be shown first.	Info
3 – Home and advanced search page development	
3.01 WW and SA provided an explanation of the implementation of the home page and the advanced search page.	Info
4 – Preparation for presentation	
4.01 NS outlined that it will be necessary to plan and create content for both a slide-deck and video.	Info

New Action Items	Owner	Due Date
Advanced search feature.	NS, SG	11-07-2014
Implementation of the home page.	WW, SA	08-07-2014
Implementation of the advanced search page.	WW, SA	11-07-2014
Run a wordcount query within Hadoop.	SA	08-07-2014

Meeting 2 – Progress meeting with James Baker

Date	Time	Location
Friday 11 th July 2014	12:00 – 12:30	Skype Meeting

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

Agenda Item – Points Discussed	Type Action, Decision or Info
1 – Progress review	
1.01 The overall progress of the project was conveyed to JB.	Info
1.02 JB informed the team that the external comms piece was satisfactory and has been uploaded to the British Library digital scholarship blog.	Info
1.03 JB asked if the keyword search implementation considers multiple words. NS confirmed that the keyword search works with multiple words.	Info
1.04 WW described the home and advanced search page.	Info
1.05 JB requested that a link to the first prototype be forwarded for approval by British Library Press and Policy.	Action
2 – Future tasks	
2.01 Mid-project presentation.	Info

New Action Items	Owner	Due Date
Send link to the prototype	SA	11-07-2014