Sharing and mapping out scholarly funding opportunities

Report Name	Outline Project Specification
Author (User Id)	Emanuil Evgeniev Tolev (eet9)
Supervisor	Reyer Zwiggelaar (rrz)
Module	CS39440
Date	October 22, 2012
Revision	1.0
Status	Release

1. Project description

The aim of this project is to produce a web application which will perform two major functions:

- ✓ collect data about scholarly funding opportunities. These opportunities can be any event/grant/policy/calls for bids/invitations to tender where the intended end result is **money** allowing somebody in the scholarly (Humanities + Sciences) community to do some research work. (Intended) recipients can be at any level – from Undergraduates to Professors.
- ✓ present this data to users of the application. Data will be anonymised (submissions will not be tied to a particular user in publicly accessible datasets). "Users" means both human users viewing an HTML5 representation of funding opportunities via a web browser and machines (programs) using a ReSTful API to query the data.

A few further notes on the two major pieces of functionality:

I. Data collection will happen via crowdsourcing. Scholars, research development officers and other interested parties (incl. organisations) will be able to submit information about funding opportunities they're aware of or wish to advertise.

   Automatic understanding (using artificial intelligence / text mining techniques) of policy/grant/tender documents *may* be added to enhance the data and usefulness of the application, but *is not* included in the initial scope of the project.

II. Presenting the data will happen in several ways:

   a. the familiar "catalogue" style – a searchable, browseable list of funding opportunities

   b. different graphical visualisations of the data. This may include laying a summary of the data on a map (of the Earth) so that the user can see (for example) which subfields of Computer Science they can get money for if they were working in a research institution in Wales.

   c. the machine-readable ReSTful API mentioned above – basically allowing other software to interact with this project and reuse the knowledge/value stored in this project's databanks

The project *may* also include functionality which would allow users to save their searches, create e-mail alerts when new opportunities matching their criteria come up, and other personal account/work management features. This has been described as quite useful for both academics and research development officers by one of the latter and will definitely be contemplated – at this time only the basic functionality of collecting and representing data is included in the scope of the project, however.

2. Work to be tackled

I. Technical spike work: the project will initially involve lots of rapid, investigative work to determine the best tools for the task (back-end, front-end).

   a. Back-end: investigate elasticsearch, an indexing server which allows for sophisticated search queries against a body of text and other data. This will be used in place of SQL relational databases due to its powerful features. it has a ReSTful API which outputs JSON and there are tools available which would allow for the creation of a high-quality searchable/browseable catalogue as mentioned in 1.I.a. above. The fact that it produces JSON also enables rapid prototyping,

evaluation and integration with Javascript visualisation libraries (examples in the Initial Bibliography section below).

b. Front-end: project will likely be written in Python due to available functionality, fast and elegant web & testing frameworks which all enable more rapid development. However, the actual "face" of the web application will have to be designed as a series of HTML5/CSS3/Javascript web pages.

   i. To that end, a few HTML/CSS and Javascript libraries like Bootstrap and jQuery will need to be investigated.

   ii. Javascript libraries which do on-the-fly visualisations (feature described in 1.II.b.) will need to be investigated.

   iii. Toolkits for rapidly building a search/browse catalogue interface need to be looked at (mentioned in 2.a. above)

II. It would also be prudent to read around scientific work published on topics such as "mapping science" or "modelling research activity" since this tool fits almost perfectly within those scholarly areas.

III. Another pocket of time should be set aside for communicating with potential users – academics, research development officers, governmental institutions and NGO-s which sponsor science or broadcast funding information.

3. Project deliverables

I. Work piece 2.I. should produce a working web application by the 10th December 2012, which will: collect data via an HTML form, put it into (presumably) an elasticsearch index, and represent it. Representation will be via a basic catalogue (fully working search and pagination of results) and one graphical visualisation. The results may not be perfect, but should be stable.

II. Work piece 2.II. should produce the summaries of at least 10 relevant publications, written down by the 14th December 2012 to be integrated into the dissertation's literature review section.

III. Work piece 2.III. should produce a small set of *prioritised* story cards – A6 pieces of cardboard with features that academics, research dev. officers etc. would find useful. This doesn't have a deadline since a prioritised set of cards will be produced by me anyway, and new ones will be added as discussions with interested parties progress.

IV. After the first "draft" version of the software is produced as per 3.I., the next steps would involve several further "draft" versions which should become more useful to interested parties with each new version. These will be produced every two weeks, starting from 1st February 2013.

4. Initial bibliography

[1] A. Rip. Chapter 9, "Handbook of quantitative studies of Science and Technology", Elsevier Science Publishers B.V. (North-Holland), 1988

[2] Snake Oil? Country and Regional Analysis. (Visualisation example). http://www.informationisbeautiful.net/visualizations/snake-oil-supplements/ . D. Mccandless, P. Doughty-White, A. Wdowski, A. Perkins. October 2010. Accessed October 2012.

[3] Interactive Grants Map. (Visualisation example). http://science.energy.gov/universities/interactive-grants-map/ . US Department of Energy, Office of Science. February 2012. Accessed October 2012.

[4] Where Does My Money Go? (Visualisation example). http://wheredoesmymoneygo.org/bubbletree-map.html . Open Knowledge Foundation. Accessed October 2012.

[5] Our Vision. (On the topic of reusable knowledge, inspired the ReST API ideas described above and the idea of the project itself). http://okfn.org/about/#our-vision . Open Knowledge Foundation. Accessed October 2012.

[6] Facetview. (Toolkit for rapidly building a search interface/catalogue on top of Elasticsearch). http://okfnlabs.org/facetview/ . Open Knowledge Foundation. Accessed October 2012.

[7] Get The Data (sharing datasets under open licences, I plan to see if anybody has filtered down some scientific funding data by asking there). http://getthedata.org/ . Accessed October 2012.

[8] Grano - open source social network analysis web platform (analyses relationships between actors in a certain subject, has visualisations – could be used for modelling funding data). https://github.com/pudo/grano . Accessed October 2012.

[9] Elasticsearch home page. (Back-end engine of choice for this project, at least for the time being). http://www.elasticsearch.org/. October 2012. Accessed October 2012.