### Analysis of the digitised newspapers held at the National Library of Wales (multiple projects) (2013-11-20) - Amanda Clare

Keywords: text processing, statistics, algorithms, protocols

The National Library recently digitised its collection of newspapers. A [hacathon](http://hacathonwiki.llgc.org.uk/w/index.php/Main_Page) in January 2013 had a go at making use of the data for various projects. Things I'd like:

* Biological species browser: in May we held a [Bioblitz](http://bioblitz.dcs.aber.ac.uk/) and counted the number of species that members of the public could find on campus. How many of these are also found in the historical newspapers? How could their stories be automatically summarised and presented in an accessible manner? Which species are not found?
* Scientists in the news: who were the scientists in these newspapers from 1850-1910? Can we build a database and website about them, based on the information that can be automatically extracted from these newspapers? What subjects did they practise and where did they work? Why were they in the news? What words were most frequently used to describe them? "distinguised"? "Welsh"? "chemist"? "summoned"? "clever"? "saved"? "fined"?

Problems include dealing with (suggesting) spelling variations and coping with OCR-misrecognition issues.

- purpose/use

- programming language

- what type of processing? searches or graphics of processed data and connections?

### Mini-analytics (2013-11-20) - Edel Sherratt

Keywords: big data, data analytics

This project entails extracting from the web publicly accessible data about members of the Department of Computer Science and building a set of profiles for those members.

The project will include a review of data analytics techniques, adaptation of those techniques for use in this mini-analytics project, reviews of two or three NoSQL database systems and systematic data analysis and design for the profiles. The project will require ethical approval and consent by the people who are profiled.

- more details;

- what does mini analytics means?

- what languages to use;

- how to go about ethical approval from the people?

- where would we get data from? anywhere or specific websites? what would that include?

- who would be the users.. the people who would use this data

### Picture chase (2013-11-20) - Edel Sherratt

Keywords: picture, game

This is an outdoor game based on a paper chase, but instead of following a trail of paper, the ‘hounds’ try to recreate a sequence of pictures taken by the ‘hare’ during the chase.

Your software will allow pictures to be uploaded to a website by the hare, together with the times at which the pictures are taken. The pictures will be displayed as they are uploaded; the hounds will follow trying to re-create the same sequence of pictures with as near as possible the same sequence of times.

If the hare reaches a pre-arranged destination first, or if none of the hounds re-creates the same sequence of pictures the hare wins. Otherwise the hound with a matching sequence of pictures and the smallest cumulative time difference (calculated by the software) wins.

- does the image taken by the hare needs to be compared using image processing to the one taken by the hounds?

- would the hounds know the location just because of the image posted online?

- what programming language?

**Alternative Approaches to Data Persistence (2013-11-20) - Edel Sherratt**

Keywords: persistent data, data modelling alternatives

This project entails comparing and contrasting different approaches to data persistence.

You will choose a reasonably simple application that requires some data to be stored and manipulated. This data should involve different formats – for example, sounds and images as well as text.

You will then implement the application round three radically different kinds of data storage – for example, a relational database, a native XML database, the Java Persistence API, and you will evaluate the three approaches used.

<http://www.ted.com/talks/jonathan_harris_collects_stories.html>

<http://www.ted.com/talks/jonathan_harris_tells_the_web_s_secret_stories.html>

### Ontology-based document clustering (2013-11-20) - Georgious Gkoutos

Keywords: text mining, semantic similarity, automated reasoning, computational biology

The aim is to develop a web-based software that can analyze large sets of documents (primarily from the biological domain) by combining text mining and measures of semantic similarity. Semantic similarity utilizes the structure of an ontology to derive a measure of relatedness between sets of concepts. The software will use Semantic Web technologies, in particular automated reasoning, natural language processing, apply similarity measures, analyse data and present the results to users.

**Develop an RDF store for phenotype data (2013-11-20) - Georgious Gkoutos**

Keywords: semantic web, bioinformatics, RDF, database, ontology

Information about phenotypes derived from animal model experiments is required for a large number of computational analyses. To facilitate these analyses, information must be integrated and accessible in a common format. The aim of the project is to use Semantic Web technologies, in particular RDF and SPARQL, and develop a store that integrates phenotype information of multiple species. In the project, an RDF store should be set up with a SPARQL query interface so that information about animal model phenotypes can be accessed an queried.

### NHS Wales mobile device applications (up to 4 projects) (2013-11-20) - Andy Starr

Keywords: iOS, Android, Medical Information

We have been talking to the applications group in NHS Wales. As part of its work, the group provide information resources via a web-based interface. This work would build applications on mobile platforms, with a proposed emphasis on iOS and Android platforms. There is scope for multiple applications.

There are two sets of data that the group would like to have available for offline access by doctors. These are:

* Formulary / antimicrobial guide app - The Formulary is a list of drugs recommended for use with a Local Health Board and sets out the order that should be prescribed and in which setting. The antimicrobial guide is similar and sets out which antimicrobials / antibiotics should be used for which clinical condition.

A similar app has been developed for NHS Lothian, available via the following link, The [screenshots](https://itunes.apple.com/ag/app/lothian-joint-formulary-antimicrobial/id680536192?mt=8) give a good indication of the sort of app we are seeking to develop.

Injectable Medicines Guide - The Injectable Medicines Guide contains a set of monographs of injectable medicines and the prescribing and administration information needed by doctors and nurses to give these drugs to patients safely. It is used by 85% of UK Hospitals. The monographs are all written to a standard template and the app would present this information for medical staff.

Neil Taylor has been in contact with the group and can provide some further information. Note: you need to create an app that can store and search the data. You do not need a background in the data.

This would be a great opportunity to undertake work for a major organisation. At the least, it would be an entry on your CV. Our original contact with the group was via the recruitment department. Whilst there are absolutely no guarantees, if you take on the work and impress them, there is a possibility that it could lead to further work.

- two databases- one local(SQLite) and the one provided

- confidentiality

-very good testing - people life depends on it

**Tourist attraction catalogue - mobile applications (2 projects) (2013-11-20) - Andy Starr**

Keywords: iOS, Android, location aware

The company proposing the project wants to enhance the visitor experience in Ceredigion by providing an event, activity and attraction calendar and location details via a smart phone app. It would like to create an app that:

* pinpoints the user's location to make content relevant
* provides details of events, activities and attractions within selectable radii of the user (say 1 mile, 5 miles, 10 miles and 25 miles)
* provides a calendar of events for a selected area (as above) for a selected date range (say within 24hrs, within a week, on a specific day, within a specified date range)

If it is possible, it would help if the app can utilise the [Google Map data](http://ceredigiontct.org/resources.html) that the company is already using and the calendar that it has incorporated in the [Cardigan Bay Events](http://www.cardiganbayevents.co.uk/news.html) web site

The company would like an iOS and an Android App, which can be done as two separate projects.