

Open Repositories 2014

Geo Tools for
Repository Augmentation

Overview

Who

- context
- (geo) standards

What

- Unlock (Places & Text)
- Tagger

Why

- rationale and motivation

Who?

EDINA, based at University of Edinburgh a Jisc designated center for digital expertise and online service delivery

Deliver thematic content and services to UK Higher and Further Education sector

e.g. Digimap, Repository Junction Broker, Opendepot, Sharegeo

A sidebar on (geo) standards

“The nice thing about standards is that you have so many to choose from.” A. Tanenbaum

ELF - European Location Framework (EU)

ISO/CEN/OGC

W3C

Library of Congress

‘Popular’ - georss

Formal (authoritative) vs informal

Vernacular geographies?

The barriers to interoperability of geospatial
PSI

- No consistent common reference data in all level of details at Pan-European level
- No uniting specifications on national data in a pan-European context
- Technology or policy lag at agencies prevents use of open architectures

WRT *Placenames* - *EuroGeonames*

BUT \$\$\$\$

What?

Unlock (Places & Text)

Middleware (API) services:

Unlock Places - an authority placename / feature gazetteer

Unlock Text - an advanced geoparser service

Tagger

A multimedia (mainly images) metadata enhancement and management service

Unlock Places

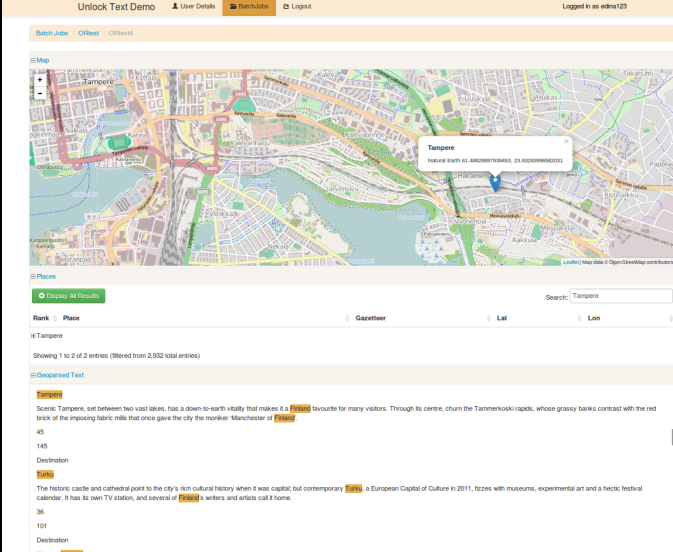
[unlock.edina.ac.uk]

- a place search service with API
- open data sources (OS, Geonames, Natural Earth)
- developer friendly format support (KML, (Geo)JSON)
 - developer tools - JS and Python packages
 - used (pending) by data.gov.uk to power search
 - provides "footprints" for features
 - each feature is a name,type,geometry set
- supports basic spatial relationship queries e.g. within a BBOX
- supports feature "crosswalking" e.g. postcodes for Edinburgh
 - just added +1M historical GB features via DEEP

Unlock Text

[unlock.edina.ac.uk]

- a natural language **GEOPARSER**
- text input -> marked up text as output
- speciality is place but can also do time, currency, names etc.
- builds on the "Edinburgh Geoparser" developed by local expert language processing group
- wraps above as a simple to use web service and demonstrator
- recent changes allow targeting of DEEP as the historical gazetteer target



The screenshot displays the 'Unlock Text Demo' web interface. At the top, there are navigation links for 'Batch Jobs', 'About', and 'Help', along with a 'Logout' button and the text 'Logged in as edina03'. Below the navigation is a map of Tampere, Finland, with a search box containing 'Tampere'. The map shows a city layout with roads, green spaces, and water bodies. Below the map, there is a search results section titled 'Showing 1 to 2 of 2 entries (filtered from 2,932 total entries)'. The results are displayed in a table with columns for 'Rank', 'Place', 'Gazetteer', 'Lat', and 'Lon'. The first result is for 'Tampere' with a rank of 45. The second result is for 'Tampere' with a rank of 145. The text below the table is partially obscured by redaction boxes, but it appears to be a description of the city of Tampere, mentioning its location between two vast lakes, its status as a favorite for many visitors, and its historical significance as a European Capital of Culture in 2011.

Tagger

[tagger.edina.ac.uk]

- a resource metadata miner
- exposes "inherent" properties of resource
 - e.g. exif tags GPSLatitude & GPSTLongitude, BitDepth etc
- an API over EXIFtool supporting read/write
- trivial use as Location Obfuscator
- simple demo exposes Open content - Geograph
- open sourced and available on github

The screenshot shows the 'Multimedia tagger' web application. It is divided into two main sections: '1 UPLOAD YOUR MEDIA' and '2 VIEW AND EDIT TAGS'.

1 UPLOAD YOUR MEDIA: This section features a 'BROWSE' button, a checkbox for 'I agree to the terms & conditions', and an 'UPLOAD' button. A preview of a landscape image is shown on the right.

2 VIEW AND EDIT TAGS: This section has tabs for 'ALL TAGS', 'EXIF TAGS', 'EXP TAGS', 'XMP TAGS', and 'CUSTOM TAGS'. A search bar is present. Below the tabs is a table of EXIF tags:

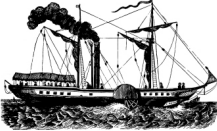
Name	Value
BitsPerSample	8
Color Components	3
ExifByteOrder	MM
ExifToolVersion	8.75
FileModifyDate	2012:04:12 19:14:56+01:00
FileName	1335507_998e99f7.jpg
FilePermissions	644
FileSize	131431
FileType	JPEG
GPSLatitude	55.1719473829778
GPSLatitudeRef	N
GPSLongitude	-2.51704202144889
GPSLongitudeRef	W

To the right of the table is a map of Scotland with a red location pin. A 'FIND' button is located above the map. Below the map are buttons for 'SAVE', 'EXPORT', 'DOWNLOAD', and 'ANONYMISE'. The bottom right corner says 'Powered by Unlock'.

Some examples: Trading Consequences

[<http://tradingconsequences.blogs.edina.ac.uk/>]

The resource provides multiple interfaces to a rich database of mentions of commodities and locations associated with commodities from the 18th century and up to the mid-20th century. One interface is a [Commodity Search](#) tool, which allows you, for example, to search for all documents that refer to [jute](#) and display a world map that clusters references by region until you zoom in for individually geolocated entries. You may also search by location to show, for example, all commodities associated in the documents with a location like [Hong Kong](#). In this interface, you can refine the search by collection or by decades. Another powerful way to interact with the material is through an [Interlinked Visualization](#) (use the Chrome browser) where, at a glance, you can view the distribution for mention of commodities across the decades. Another [Location Cloud Visualization](#) lets you view relative frequency of mentions of a commodity in various places over time.



Trading Consequences


Search Location

Search by: Commodity Location

Location edinburgh Search

Search results for 'edinburgh' (Page 1 of 1)

Location	Commodity Mentions
Edinburgh	33674
Edinburgh	94
Edinburgh District	80




Trading Consequences

Search Commodity Mention

Search by: Commodity Location

Locations related to the trade of 'Blubber'



Documents containing 'Blubber' (Page 1 of 795)

- 6005 Railway accidents, 1874. Return of accidents attended and unattended with loss of life or personal injury which have been reported to the Board of Trade by the several railway companies in England and Wales, Scotland, and Ireland, in pursuance of the Regulation of Railways Act (1871), 34 & 35 Vict. cap. 78, as having occurred on their lines during the year ending the 31st of December 1874. (Part I and Part II.)
- 5853 Reports of the inspecting officers of the Railway Department to the Board of Trade, upon certain accidents which have occurred on railways during the months of June, July, August, September, October, November, and December, 1873. (Part I and Part II.)
- 5793 Reports of the inspecting officers of the Railway Department to the Board of Trade, (including reports of public enquiries), upon certain accidents which have occurred on railways during the months of June, July, August, September, October, November, and December, 1872. (Part I and Part II.)
- 5094 Railway accidents, 1873. Part I. Return of accidents attended with loss of life or personal injury which have been reported to the Board of Trade by the several railway companies in England and Wales, Scotland, and Ireland, in pursuance of the Regulation of Railways Act (1871), 34 & 35 Vict. cap. 78, as having occurred on their lines during the year ending the 31st of December 1873. Part II. Return of accidents unattended with loss of life or personal injury which have been reported to the Board of Trade by the several railway companies in England and Wales, Scotland, and Ireland, in pursuance of the Regulation of Railways Act (1871), 34 & 35 Vict. cap. 78, as having occurred on their lines during the year ending the 31st of December 1873.
- 5158 Railway accidents.
- 4766 Royal Commission on Railways. Report of the commissioners.
- 4410 Royal Commission on Railway Accidents. Report of the commissioners.
- 3903 Railway accidents. Returns of accidents and casualties as reported to the Board of Trade by the several railway companies in the United Kingdom, during the year ending 31st December 1873, in pursuance of the Regulation of Railways Act (1871), 34 & 35 Vict. cap. 78, together with reports of the inspecting officers of the Railway Department to the Board of Trade upon certain accidents which have occurred on their lines.

Some examples: Palimpsest

The project, a collaboration between literary and informatics scholars, aims to develop an interactive, mobile friendly website, called Palimpsest, that provides access to vivid, evocative and dramatic excerpts of a wide range of **geolocated Edinburgh-based literary texts** either through the web or via smartphones and tablets.

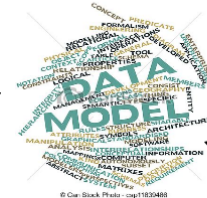


LTG Candidate and NLP processed input



User interaction (main website)

Parsing to relational data model (iterative)



DB population (iterative)



Why?

- as a repository QA control (place thesaurus)
- as a repository enrichment tool
- as a repository search tool
- as an easy way to "map" (visualise) content
- as a way to explore relationships between items
- because the tools are free to use
- because we are amenable to modify tools to meet bespoke workflows
- spatial content is ubiquitous
- naive use of spatial content can have unforeseen implications...

Questions?

**PLEASE BE CAREFUL WHEN
SHARING PHOTOS ON SOCIAL
MEDIA. THEY CAN LEAD POACHERS
TO OUR RHINO**

**TURN OFF GEOTAG FUNCTION
AND DO NOT DISCLOSE WHERE THE
PHOTO WAS TAKEN**

james.reid@ed.ac.uk

unlock.edina.ac.uk tagger.edina.ac.uk