Open Repositories 2014

Geo Tools for Repository Augmentation

James S Reid
EDINA
University of Edinburgh
Overview

Who
- context
- (geo) standards

What
- Unlock (Places & Text)
- Tagger

Why
- rationale and motivation
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
“The nice thing about standards is that you have so many to choose from.” A. Tanenbaum

ISO/CEN/OGC
W3C
Library of Congress
‘Popular’ - georss
Formal (authoritative) vs informal
Vernacular geographies?

ELF - European Location Framework (EU)

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
- 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
- a resource metadata miner
- exposes "inherent" properties of resource
  e.g. exif tags GPSLatitude & GPSLongitude, 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 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.

[http://tradingconsequences.blogs.edina.ac.uk/]
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.
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 unforseen implications...
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