Text to RDF - and thence to the Web of Data

Open Knowledge Scotland, 13th May 2010

Language Technology Group, School of Informatics
Two points to take away

- txt2rdf
- grounding
We’re used to communicating in natural language
We’re used to communicating in natural language
We’re used to communicating in natural language
We’re used to communicating in natural language
We’re used to communicating in natural language

...but it’s still difficult to search well.
txt2rdf pipeline

fancy NLP processing + rdfisation
Example - RCAHMS text
Evidence of a quartz knapping site was found within the confines of the [stone circle][SITETYPE], and in conjunction with several structures within the inner ring, strongly suggests a domestic site.

Besides the quartz implements and corresponding waste, several other artifacts of local origin occurred including a split pebble axe of greenstone with [Shetland][PLACE] [Early Bronze Age][PERIOD] affinities. [B Beveridge][PERSNAME], [1972][DATE].

Field survey and [excavation][EVENT], as a response to continual wind and marine erosion, was carried out at the [Sands of [Breckon][ADDRESS]][SITENAME] between [1982][DATE] and 1983. [HP50NW 11.00][ADDRESS] was recorded as a [stone settings][SITETYPE] surrounded by occupational debris (Site 22). [Excavation][EVENT] revealed midden deposits of an early [Iron Age][PERIOD] date and a surface scatter of artefacts of mixed dates. The [stone settings][SITETYPE] were tentatively interpreted as the basal stones of [long cists][SITETYPE].

[Historic Scotland][ORG] Archive Project (SW) [2002][DATE]
Evidence of a quartz knapping site was found within the confines of the [stone circle][SITETYPE], and in conjunction with several structures within the inner ring, strongly suggests a domestic site. Besides the quartz implements and corresponding waste, several other artifacts of local origin occurred including a split pebble axe of greenstone with [Shetland][PLACE] [Early Bronze Age][PERIOD] affinities. [B Beveridge][PERSNAME], [1972][DATE]. Field survey and [excavation][EVENT], as a response to continual wind and marine erosion, was carried out at the [Sands of [Breckon][ADDRESS]][SITENAME] between [1982][DATE] and [1986][DATE], yielding further deposits of an early [Iron Age][PERIOD] [stone][settings][SITETYPE]. [Historic Scotland]
Disambiguation and grounding
Disambiguation and grounding
Disambiguation and grounding
Disambiguation and grounding

ground classification terms, placenames, people, organisations...
against
domain ontologies, geonames, dbpedia, etc
Grounding turns data into Linked Data
Grounding turns data into Linked Data
Grounding turns data into Linked Data
Grounding turns data into Linked Data
Grounding turns data into Linked Data
Grounding turns data into Linked Data
Grounding turns data into Linked Data
Grounding turns data into Linked Data
Grounding turns data into Linked Data
Grounding turns data into Linked Data.
Grounding turns data into Linked Data

grounding local URIs against “authority” nodes is the next big challenge!
Thank you!