

Diving into information lakes with the PoolParty API

(and coming up with treasure!)

Ian Piper
tellurasemantics.com

10 September 2019

What's in store

- ◆ A word about APIs
- ◆ Improving content classification
- ◆ Classification without integration
- ◆ Taxonomy building through online card sorting

A word about APIs

Application Programming Interface

A communication channel between two systems

Using the PoolParty API in a nutshell

- ◆ Build an http request with parameterised data
- ◆ Send the request to PoolParty
- ◆ Retrieve a package of results
- ◆ Unpack and present in the application UI

A quick example


Getting useful data out of PoolParty

Let's get some data out of this

▼ CMS legal classification taxonomy

- ▶ Content types (9)
- ▶ Glossary of legal terms and definitions (73)
- ▼ **Regions (7)**
 - ▶ Cross border (3)
 - ▼ England (8)
 - East of England (0)
 - London (0)
 - ▶ Midlands (2)
 - North East (0)
 - North West (0)
 - ▼ South East (2)
 - Lower South East (0)
 - Upper South East (0)
 - South West (0)
 - Yorkshire and the Humber (0)



Regions

 <http://tellura.poolparty.biz/CMSClassification/1bd8d749-020d-4a48-8461-14eae220c874>

Metadata	Triples	Visualization	History
----------	---------	---------------	---------

Top Concepts

- [Cross border](#)
- [England](#)
- [National](#)
- [Northern Ireland](#)
- [Rest of world](#)
- [Scotland](#)
- [Wales](#)

The subtree method

GET /PoolParty/api/thesaurus/{projectid}subtree?root=<http://tellura.poolparty.biz/CMSClassification/{conceptid}>

{projectid} : the project identifier

subtree : the name of the method

root : one of the parameters to use with this method

conceptid : the point from which you want to see a tree of child concepts

The subtree method response

The method retrieves a JSON package showing the subtree structure starting from the specified root concept

```
"concept": {
  "uri": "http://tellura.poolparty.biz/CMSClassification/a196ffaa-2049-4d9e-bdbb-397bbf3b4d8d",
  "prefLabel": "England",
  "narrowers": [ {
    "concept": {
      "uri": "http://tellura.poolparty.biz/CMSClassification/57d7fd70-342e-4b60-9398-b7e4affcf611",
      "prefLabel": "South East",
      "narrowers": [ {
        "concept": {
          "uri": "http://tellura.poolparty.biz/CMSClassification/5c81feed-4f59-46c8-9532-71b4a9a14276",
          "prefLabel": "Lower South East",
          "narrowers": [ ],
        },
        "concept": {
          "uri": "http://tellura.poolparty.biz/CMSClassification/20cb33e7-4cbd-4f38-8f62-c574c5dd303c",
          "prefLabel": "Upper South East",
          "narrowers": [ ],
        }
      ]
    }
  }
],
}
```


How it looks in an application

CMS legal taxonomy

- + Subjects (88 items)
- + Glossary of legal terms and definitions (73 items)
- + Content types (9 items)
- Regions (7 items)
 - Wales
 - Scotland
 - National
 - Rest of world
 - Northern Ireland
 - + Cross border (3 items)
 - England (8 items)
 - South West
 - + Midlands (2 items)
 - Yorkshire and the Humber
 - North East
 - North West
 - London
 - South East (2 items)
 - Upper South East
 - Lower South East
 - East of England

Subtree down
from
England

Some real-life examples

Part I

Improving content classification

Better user experience and
future-proofed information storage

Conventional CMS tagging

- ◆ Content item in a CMS is linked (“tagged”) with a taxonomy term
- ◆ Taxonomy terms are little more than **keyword** lists in the CMS database
 - ◆ That is, buried away from external systems
- ◆ The **tagging fact** is stored inside the CMS
 - ◆ So that information is inaccessible too
- ◆ When you change CMS, you **lose** the tagging information

How could we improve on this?

- ◆ Use a taxonomy stored in a **taxonomy management system**
 - ◆ Rich **concepts** rather than just **keywords**
- ◆ Store the tagging fact **outside** the CMS
 - ◆ Tagging is independent of the CMS, so it's **future-proofed**
- ◆ Store the **tagging fact** in a graph database
- ◆ The outcome is a **content graph**
 - ◆ A network of linked content objects and taxonomy concepts

The Content Graph Explorer

- ◆ A tool to improve content tagging
 - ◆ Better user experience for authors and editors
 - ◆ Future-proofing tagging information
- ◆ A bridge between a CMS, PoolParty and a graph database
- ◆ An essential step in building a content graph
- ◆ Uses the PoolParty Extractor API

Using the Extractor API

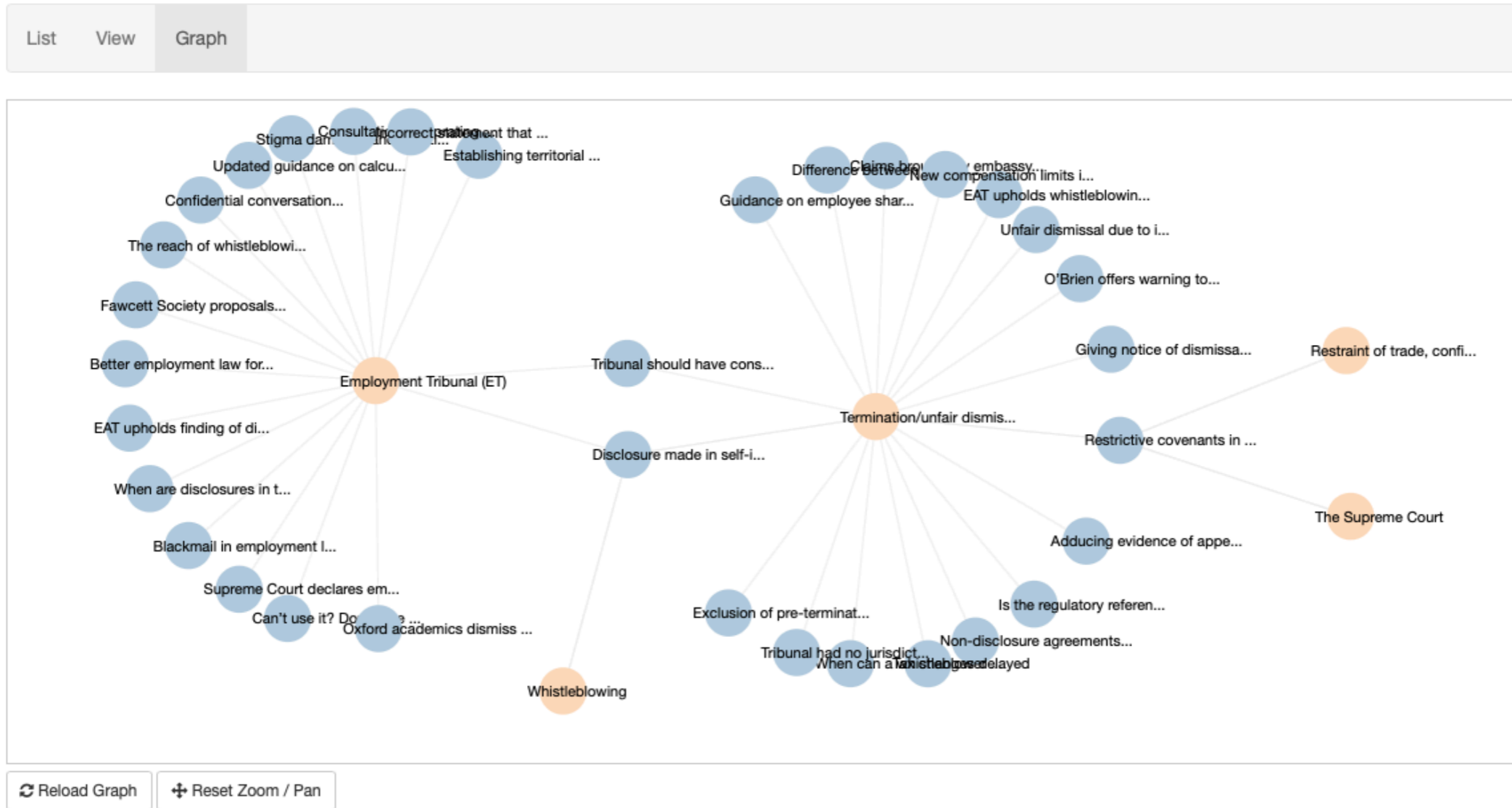
- ◆ `/extractor/api/extract`
- ◆ Create a request containing
 - ◆ A block of text or a file
 - ◆ The taxonomy project ID
- ◆ Returns a package of concepts from the taxonomy that match the content

Content Graph Explorer in more detail

- ◆ Loads content items out of Drupal in real time
- ◆ Sends content to PoolParty to extract matching concepts
- ◆ Allows editor to choose the relevant tags
- ◆ Stores the tagging fact in GraphDB
- ◆ Provides a visual map of the content graph

Exploring a content graph

Restrictive covenants in the Supreme Court



Content Graph Explorer demonstration

Part 2

Classification without integration

...because sometimes we can't integrate systems

We can't always integrate systems

- ◆ The IT department doesn't like new systems coming in
- ◆ CMSes are often proprietary and tagging mechanisms are opaque
- ◆ There may not even be an API to hook into

So...

How can we classify CMS content with taxonomy concepts but without integrating the CMS and the TMS?

Classify externally... using TurboTagger

TurboTagger

- ◆ Chrome browser extension
- ◆ Does **not** require invasive integration
- ◆ Stores tagging facts in a graph database or in PoolParty

TurboTagger demonstration

Part 3

Making taxonomy building easier

Using online card-sorting for
collaborative, real-time taxonomy building

Card sorting for taxonomy building

- ◆ **Stage 1** - Gather colleagues together with a set of cards
 - ◆ Create or select cards that make sense for the knowledge domain
- ◆ **Stage 2** - Group and order the cards into logical structures
 - ◆ Taxonomist captures structure in notes or photographs
- ◆ **Stage 3** - Taxonomist adds concepts to taxonomy management system
- ◆ Effective, but manually intensive and everyone has to be in the room

What if we could make this easier?

Wouldn't it be nice if we could directly build a taxonomy using a collaborative online tool?

The Cardsort tool

- ◆ Interactive web application
 - ◆ Use it anywhere, on many different devices
- ◆ **Stage 1** - Each colleague can suggest new concepts
- ◆ **Stage 2** - Real time addition of concepts directly into the PoolParty hierarchy
- ◆ **Stage 3** - There is no stage 3
- ◆ Uses the PoolParty suggestions API and other methods
- ◆ No need to have everyone in the room
- ◆ No tedious transcription of concepts from cards to TMS

Cardsort demonstration

In closing

- ◆ The PoolParty API offers opportunities to build innovative tools for creating and exploring content/knowledge graphs
- ◆ It provides the basis for building better, future-proofed content classification
- ◆ Novel ideas can turn into practical products



Thanks to:

Bronwen Reid

Joe Pritchard

Sharon Hemming

Clair Honeywill

Semantic Web Company

Ontotext

Employment Lawyers Association

ian.piper@tellura.co.uk

<https://tellurasemantics.com>