

Text Analytics in ITS 2.0: Annotation of Named Entities

Annotate named entities using Natural Language Processing to inform localization services about the intended meaning of text

Problem: Dealing with personal names, product names, or geographic names, chemical compounds, protein names

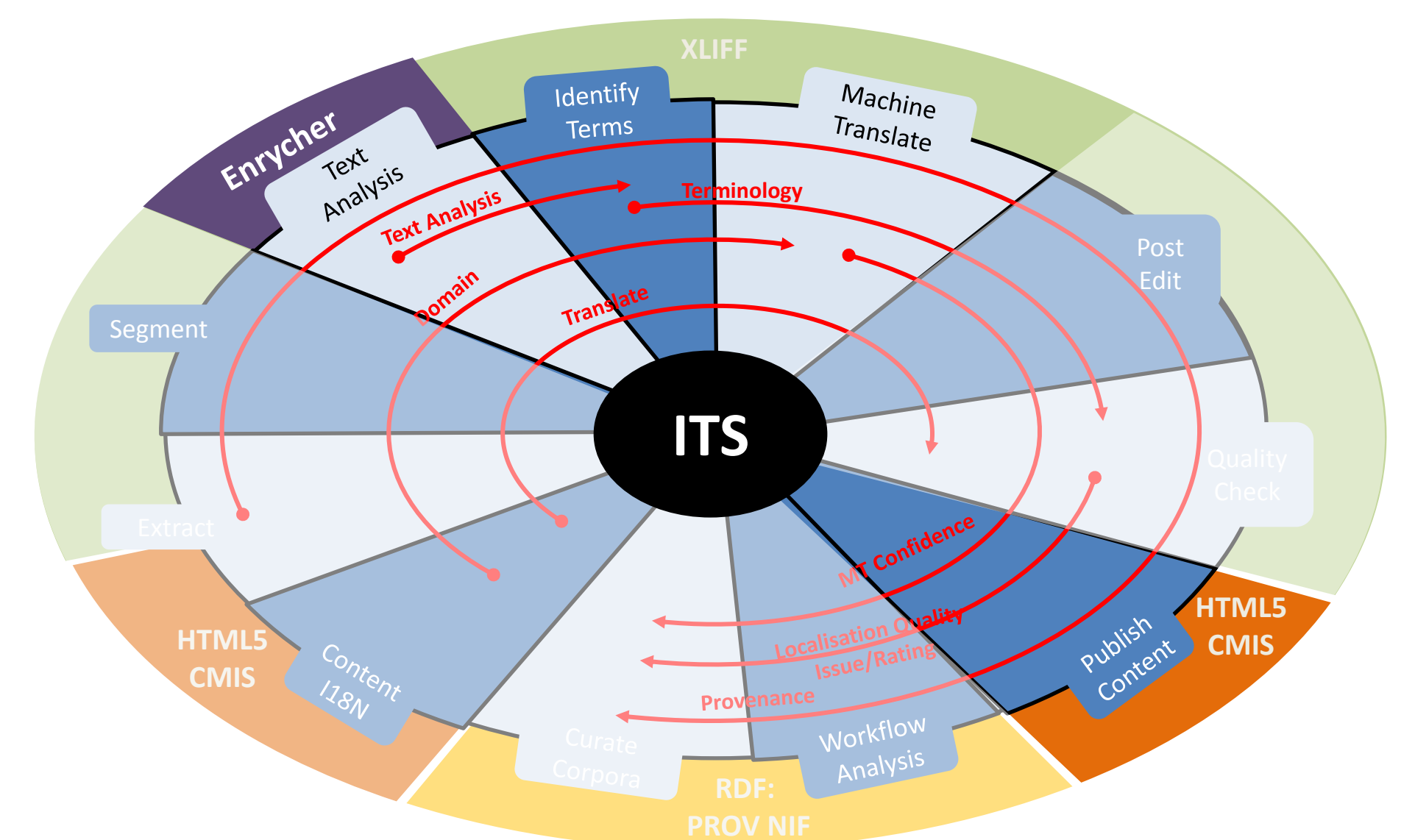
Solution: When we localize the content without sufficient context, we can use automatic annotations to provide context for ambiguous content:

- the identity (for example London - England, and London - Ontario can be distinguished)
- the category (e.g. geographic name/place, person, organization)

enrycher

THIS IS WHAT YOU'RE WRITING ABOUT:

```
<!DOCTYPE html>
<div
  its-annotators-ref="text-analysis|http://enrycher.ijs.si/mlw/toolinfo.xml#enrycher">
  Hello, <span
    its-ta-ident-ref="http://dbpedia.org/resource/Rome"
    its-ta-class-ref="http://schema.org/Place">Rome</span>!
</div>
```



ITS 2.0 Data Category: **Text Analysis**

Use cases

- ITS 2.0 markup provides the key information about which entities are mentioned, so they can be correctly processed within translation:
 - Use different handling mechanisms: specific translation, officially mandated translations, transliteration, or even keep the original.
 - see poster *Simple Machine Translation, Translation Package Creation, Quality Check*
- Besides providing important information for specific translation scenarios, it is also usable for text-data integration.
 - see poster *CMS implementation of ITS2.0. CMS to TMS System: Drupal to GBC Server / PLINT*

Demo: <http://enrycher.ijs.si/mlw/>

Source: <http://github.com/tadejs/enrycher-its20>