VisTA: Visual Terminology Alignment Tool for Factual Knowledge Aggregation

A. Axaridou, K. Konsolaki, M. Theodoridou, A. Kozlov, P. Haase & M. Doerr
Contents of presentation

• Overview
• A short discussion about Alignment
• Motivation
• Related work
• The Alignment problem in VisTA
• Key features
• Alignment rules
• Alignment result
• VisTA GUI & features
• Configuration
• Conclusions

This work is the result of collaboration between the British Museum, metaphacts and FORTH in the context of the ResearchSpace project.
Overview

• **Exact** (not approximate) alignment for RDF/SKOS terminologies => enables factual knowledge aggregation

• A simple and friendly **web-based** user interface for the alignment between two terminologies

• **Visualizes** the terminology hierarchies, enables the **interactive** alignment process, and presents the alignment result

• Component of the **Synergy Reference Model** (an initiative of the CIDOC CRM Special Interest Group for manipulation of data provisioning and aggregation processes)
What is Terminology Alignment

Terminology alignment = Thesauri mapping

Terminology A  “Translation”  Terminology B

Communication

Individuals/Communities

knowledge

knowledge
About Alignment 1/2

• The broader area of Alignment comprises: schema matching, schema mapping, ontology alignment

• Alignment is: The process of creating correspondences (direct associations) between terms/concepts of different vocabularies. E.g. “alignment” may also be known as “matching”

• Proposed methods: applying automatic or semi-automatic procedures based on configurable algorithms and workflows. Many approaches usually deal with literal matching of the terms along with a structural matching

• Focusing on: large datasets/schemas/ontologies/vocabularies

• Challenge: the development of approaches trying to achieve results of as high as possible precision and recall
About Alignment  2/2

Takes place (usually)
• between two terminologies

Results in:
• the estimation of the similarity among the terms
  
  *e.g.*
  
  \[ Sim (member, remember) = 87\% \]
  \[ Sim (member, membership) = 75\% \]

Even if more accurate results...
The question is:
• How a percentage value can factually contribute to a semantic knowledge aggregator?
  (What does a percentage value semantically mean?)
Motivation

- The approximate results of estimated similarity among terms, give no semantics about the similar terms.

- To make the Alignment result useful to a knowledge aggregator specific semantic relations have to be defined at least between the most “similar” terms.

- Last phase of an approximate alignment: the human convention (manual process). The only method preserving precision with high recall.

- Lack of GUI for supporting visual interactive alignment. Existing GUIs operate usually in 2 separated steps: a configuration phase followed by a rendering phase of the result.
## Related work

<table>
<thead>
<tr>
<th>Method</th>
<th>Supportive GUIs</th>
<th>Input size</th>
<th>Output accuracy</th>
<th>Exploitability of result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automatic</td>
<td>Yes, for configs</td>
<td>Large</td>
<td>Approximate, e.g. estimation of similarity between terms</td>
<td>?</td>
</tr>
<tr>
<td>Semi-Automatic</td>
<td>Yes, for configs&lt;br&gt;Sometimes yes, for the manual phase but without attribution of the semantics</td>
<td>Medium/Small</td>
<td>Exact, i.e. define the exact relation between terms</td>
<td>Yes</td>
</tr>
<tr>
<td>Manual</td>
<td>VisTA. Other?</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The alignment problem in VisTA 1/4

- Alignment of two terminologies as an *asymmetric* process, aiming to the *subordination* of a source to a target terminology.

- The process results in $n \times m$ correspondences among the terms of the two terminologies.
The alignment problem in VisTA 2/4

• We consider three *principles*:
  • Terminologies and the alignment result are *acyclic* graphs preserving the taxonomy *subsumption* (*is-a*)
  
  • The *broader* and *narrow* relations are *symmetrically inversed*

• *Target-driven* reconciliation of the source terminology: the source conforms to the target but not the reverse
The alignment problem in VisTA 3/4

• Goal:
  • Integration of different terminologies under one target terminology used as a core terminology in a knowledge aggregator. The core terminology gets extended with new specialized terms
  • Empowering searching capabilities in a semantic network, as the users of different terminologies are enabled to make queries using the common target vocabulary together with their own familiar vocabularies, to find more resources in their results.
• Example of factual aggregation
• Searching for resources related to term **disciplines**, we get all the instances related to its specialization
VisTA key features

• The adaptation of the structure of the source terminology is supported by the tool: removal of the incompliant subterms that break the subsumption of the target (see an example next)

• During the process multiple-inheritance of terms, i.e. a term may have multiple parents, and subhierarchy overlaps may occur. These situations are allowed unless they break subsumption

• The extension of the target terminology is based on broader match and the exact match relations

• Alignment rules (...more)

• Native RDF/SKOS Alignment result (...more)
Example of incompliance of the source terminology structure

Source

Art & Architecture Thesaurus (AAT)

aat: Activities Facet encompasses areas of endeavor, physical and mental actions, etc.

aat: natural phenomena

Target

Backbone Thesaurus (BBT)

bbt: activities classifies intentional actions

bbt: activities
Alignment rules

• **Check for explicit (direct) alignment relation.** The source term must not be already aligned to the same target term, otherwise the user is informed and the process is **canceled**.

• **Check for the existence of the source term in the target tree.** When the source term to be aligned is already an *original* term of the target terminology then alignment of that term causes the change of the target terminology which is not allowed. In this case the user is informed and the process is **canceled**.

• **Check for implicit (indirect) alignment relation.** The source term may already be indirectly aligned to the target term. In this case the user gets a warning whereas the alignment is **allowed**.

• **Check for aligned descendants of the source term.** The source term may contain already aligned sub-terms. In this case the user gets a warning whereas the alignment is **allowed**.
Alignment result

For each pair of aligned terminologies

• Alignment graph
• Native RDF/SKOS
• Contains:
  • the correspondences
  • the children hierarchy of the terms coming from the source terminology
• Searching space:
  • Alignment graph + Target terminology graph
VisTA GUI: Import/Export page
VisTA GUI: Edit page
VisTA Features 1/6

- Visualization of the Terminologies
  - Tree structures as indented lists
  - Expandable/collapsible nodes
- Interactive alignment
- Drag’n’drop
- Run time visualization of the alignment result on the target terminology tree
• **Visualization** of a Term

  • **States of a Term:**
  
  ➤ - Directly aligned  e.g. *aat:mental activities, bbt:human interaction*
  ➤ - Indirectly aligned  e.g. *aat:functional concepts (general)*
  ➤ - Non-aligned  e.g. *bbt: intentional destructions*
Visualization of Correspondences

- Highlight of the related terms at both hierarchies
  e.g. `aat:mental activities` and `bbt:human interactions`
VisTA Features 5/6

Visualization of matching relations

- exact-match
  - represented with “=“
  - the source term shares its children with the target term

- broader-match

**exact match to bbt:functions**

**broader match to bbt:human interactions**
VisTA Features 6/6

- **Searching**
  - By term-id
  - Free-text
- **Visualization** of the search result
  - The search result is highlighted in the full hierarchy
  - The parent hierarchy of the searched term is highlighted too
Configuration of the tool

- **metaphacts** platform provides an **extensible template mechanism** based on modern technologies (HTML5, Web Components, Handlebars)

- Easy configuration
  - SPARQL **query** templates
    - Displaying hierarchies
    - Searching hierarchies
  - RDF **properties** (types, hierarchy, alignment)
    - *e.g.* `rdf:type`, `skos:broader`, `gvp:broader`, `skos:exactMatch`, `skos:broaderMatch`
  - RDF **classes** of concepts
    - *e.g.* `skos:Concept`, `gvp:Concept`, `gvp:Subject`, `gvp:Facet`, `gvp:Hierarchy`, `gvp:GuideTerm`
Conclusions

• Terminologies and thesauri, contrary to other ontologies can be handled in more precise ways regarding alignment in order to produce accurate and exploitable results.

• VisTA, provides an interactive solution for the exact terminology alignment problem, required in the context of data provisioning and aggregation processes.

• We propose VisTA not as a competitive but as a complementary supportive solution for the work pending the manual phase of an alignment procedure.

• This is a work in progress:
  • More features
  • Integrate new alignment constraints
  • Evaluation
End of presentation

Thanks for your attention!

• Questions or comments?