Update on the Unified Astronomy Thesaurus

Alberto Accomazzi

NASA Astrophysics Data System

Harvard-Smithsonian Center for Astrophysics

3 April 2014

AAHEP7

Rationale

- Merge existing divergent thesauri into an open, interoperable thesaurus covering all areas of A&A

- Work out legal issues related to copyrights, re-use

- Develop a stable maintenance process that maximizes UAT adoption and use

- Allow for input from community to be incorporated in a timely and transparent way

- Provide well-curated, periodic releases of the UAT

- Seek IAU blessing of Thesaurus

UAT Circa 2013

- The American Institute of Physics (AIP), the Institute of Physics (IOP), ADS, IVOA and AAS have been discussing a Unified Thesaurus since 2011

- IOP developed a thesaurus covering its literature, incorporated much of IVOAT in the process

- AIP and IOP merged and donated the astronomy parts of their thesauri to AAS which in turns released it under CC-BY

- The CfA library in consultation with ADS is currently managing the curation and dissemination of the thesaurus

UAT Lineage

- The current draft version of the UAT is essentially the astronomy portion of the IOP thesaurus enriched with terms from IVOAT

- What has been kept are *astronomy* concepts that appear in the astronomy journals published by IOP, i.e. *ApJ* and *AJ*

- Additionally, restructuring was done to group concepts, "balance" the thesaurus and facilitate indexing/retrieval

- This means that some sections of the thesaurus were not retained, others were moved or heavily modified

What we have now

- UAT:
  - ▸ Total number of concepts: 1907
  - ▸ Top-level concepts: 15
  - ▸ Total number of paths: 3552

- IVOAT:
  - ▸ Total number of concepts: 2889
  - ▸ Top-level concepts: 274
  - ▸ Total number of paths: 5202

The Website

- Main website: http://astrothesaurus.org

- Hierarchical browser of the thesaurus: http://astrothesaurus.org/thesaurus/hierarchical-browse/

- Alphabetical browser: http://astrothesaurus.org/alphabetical-browse/

- Dendogram: http://altbibl.io/astrothesaurus/uat/dendrogram/

- Download as Excel, SKOS/RDF: http://astrothesaurus.org/thesaurus/

Infrastructure Work

After long evaluation of all alternatives, ADS installing VocBench platform for UAT curation

- Open source platform for vocabulary management

- Supports various roles: user, editor, validator

- Allows multiple users to suggest changes to the UAT

- Captures detailed history, provenance of the term, and tracking of suggestions

- We expect this to be useful for managing other controlled vocabularies/ontologies by the community

Editorial Work

- Recruited a dozen volunteers from community to serve as editors of Thesaurus

- Actual editing has been delayed pending the creation of a web-based curation platform

- ADS performing concept extraction from astronomy corpus based on UAT, which will be useful in evaluation of thesaurus

- Still need to figure out how to deal with vocabulary mapping, linking to other ontologies

Going forward

- Guiding principle: the thesaurus should contain astronomical concepts found in the literature

- Depending on how we define "literature" things may be added or even taken away. Some options:

  ▸ The core journals: ApJ, AJ, A&A, MNRAS (~ 240K)

  ▸ The extended core set: add PASP, PhRvD, PASJ, PhRvL, Nature, Science (~ 330K)

  ▸ The refereed literature in ADS astronomy db (~ 1M)

  ▸ The entire ADS astronomy db (~ 2M)