Recent Progress on the Unified Astronomy Thesaurus

Alberto Accomazzi

IVOA Interop, Trieste, Italy

23 October 2016

History of Vocabularies in Astronomy

- 1973: The Apj starts publishing “Subject Headings” to help characterize published papers; list gets updated every few years
- 1992: The IAU Thesaurus is created by R&R Shobbrook (AAO); last update and paper go back to 1995
- 2008: Rick Hessman and IVOA members resurrect the thesaurus, update it, and name it the IVOA Thesaurus
- 2010: Various publishers in physics express an interest to collaborate on an interoperable thesaurus, focus on Astronomy because of IVOAT
- 2012: Access Innovations merges concepts from IOP, AIP, IVOAT and releases an updated thesaurus

The Unified Astronomy Thesaurus

- 2012: Copyright for the draft thesaurus is assigned to the AAS
- 2013: Work on the UAT begins at the CfA library
  - Review of concepts, merge of SPIE terms
- 2013: A first beta version is released and named Unified Astronomy Thesaurus
  - 1,920 concepts, 15 top level categories, a depth of 15, and 224 ‘related terms’
- 2015: Workshop on Shared Curation at the CfA brings librarians, editors together
  - Consensus is reached that thesaurus needs major review / restructure
  - Librarians and data scientists are identified to carry out the review
- 2016: New Release of UAT (v. 1.1)
  - 1,836 concepts, 11 top level categories, a depth of 10, and 319 ‘related term’
  - Top-level structure mirrors IAU divisions

Use Cases

- Journals: publishers are interested in identifying topics discussed in papers so they can provide recommendations, notifications to authors, readers
- ADS: having a list of concepts associated with papers from controlled dictionary allows filtering of results, consistent searching and recommendations
- Archives (STScI): being able to classify observing proposals with concepts provides better management of reviews and classification of observations
- Archives: associating concepts to data products enhances discovery, supports analytics, data management efforts
- Community at large: if every archive / system were to use a common vocabulary for describing a resource, discovery and linking would benefit from it
- What is your use case? Tell us!

Governance and Stewardship

- The UAT is owned by the AAS for the benefit of the worldwide astronomy community
- The thesaurus is made available under a CC-BY-SA license and can be downloaded from its website or github
- An advisory board has been setup to coordinate future efforts and guide its development; your participation welcome!
- Guiding principle for content is that it should be able to completely describe the core literature in astronomy (think ApJ, A&A, MNRAS)
- More work is needed to review parts of the thesaurus:
  - Review current thesaurus and offer feedback
  - Areas where topics are missing (Techniques and Methods, Astrobiology)
  - Ensure that community use cases are covered

Curation Platform

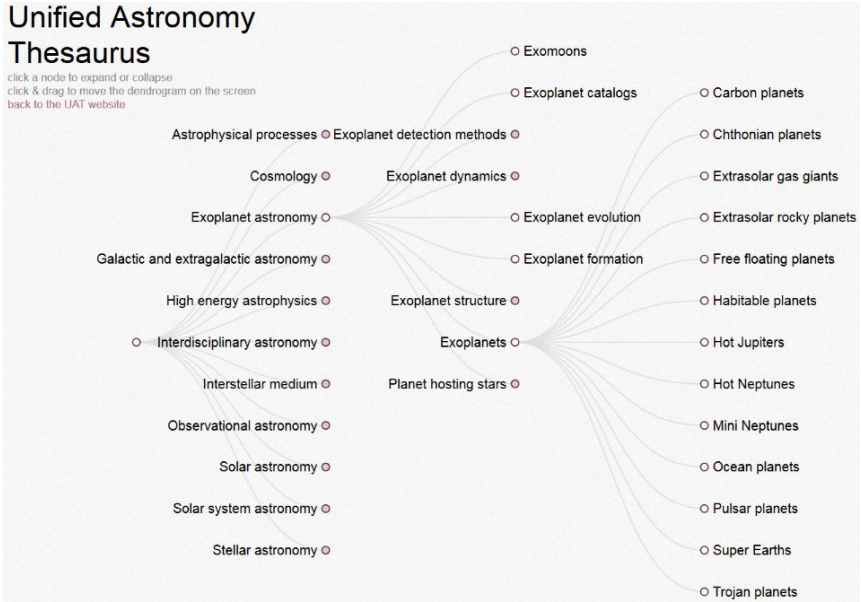
- 2013: Early editorial work relied on Access Innovation's curation platform (Data Harmony)
- 2014: The ADS takes over hosting of the thesaurus
  - Curation is done on the Vocbench platform (ontology management system)
  - Outreach / tools hosted on the project's website: http://astrothesaurus.org
  - SKOS versions of the thesaurus released on Github
- 2016: Curation is moved to the PoolParty platform hosted by the Australian National Data Service (ANDS): http://www.ands.org.au
  - Better support for provenance and history tracking
  - Linked data APIs built into the system
- Is this what people want / need?

Releases and Versioning

- Current plan is to provide regular releases of the thesaurus at least once a year, but community feedback will determine eventual schedule
- Current proposal is to name releases using criteria similar to the Semantic Versioning pattern (see http://semver.org): MAJOR.MINOR.PATCH
  - MAJOR change -> modification in hierarchy, deletion of concepts
  - MINOR change -> addition of terms, links, deprecation of concepts
  - PATCH change -> addition of alternates, preferred terms, documentation
- Could this be a template for the versioning of other vocabularies in IVOA?

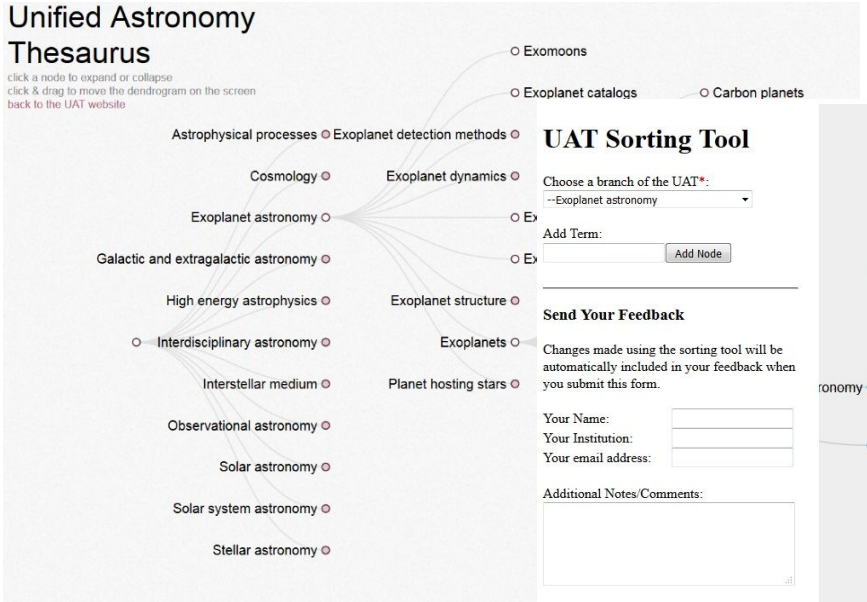
astrothesaurus.org/tools

astrothesaurus.org/tools

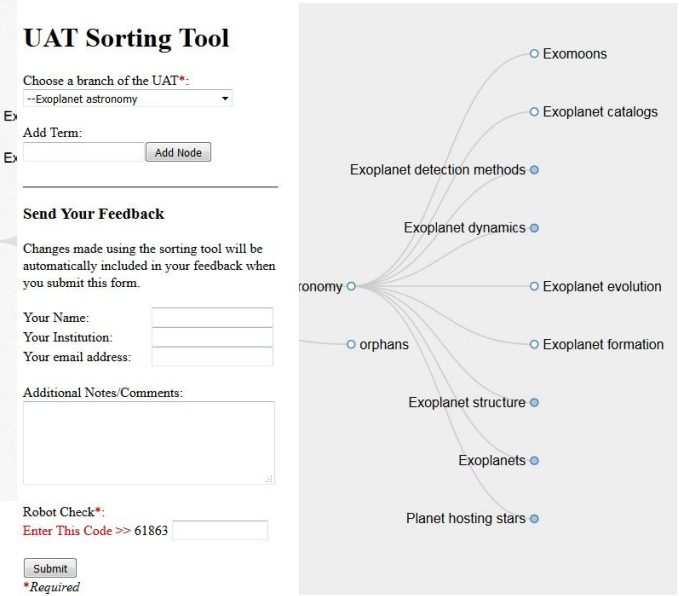


Visual Browser

astrothesaurus.org/tools

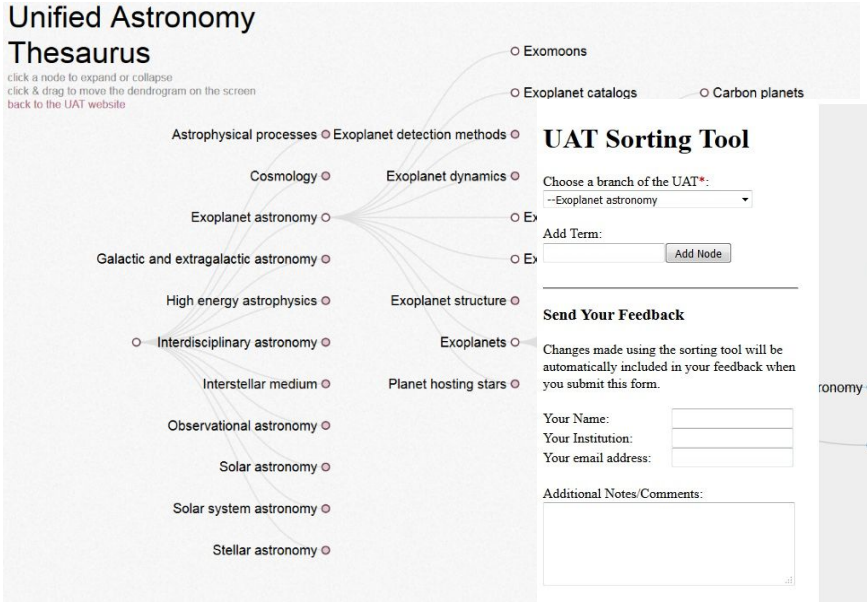


Visual Browser

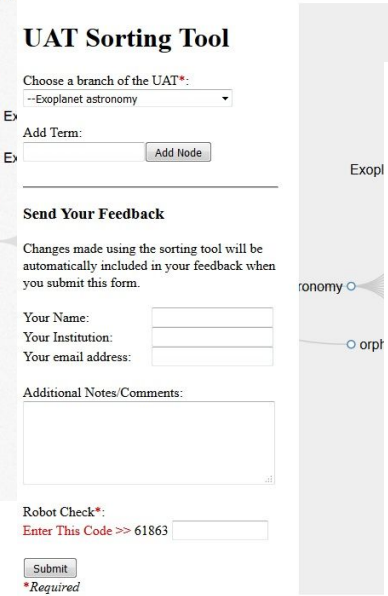


Sorting Tool

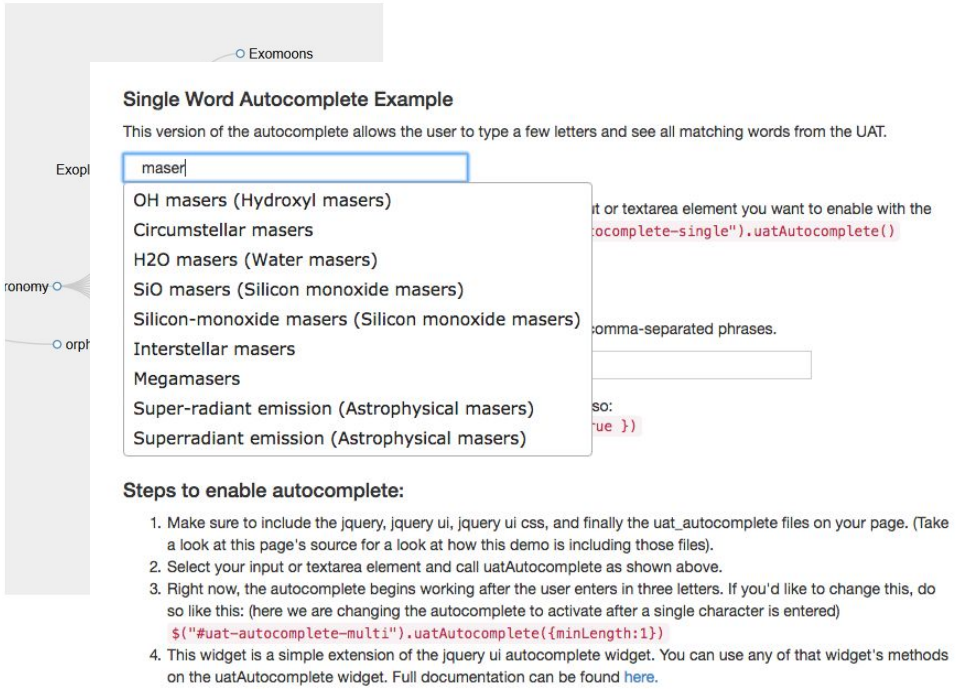
astrothesaurus.org/tools



Visual Browser



Sorting Tool



Autocomplete widgets

How and why you should get involved

- A substantial amount of work has gone into updating a legacy thesaurus and make it relevant for current Astronomy research
- ADS will use the UAT to classify documents in our system, and AAS journals will do the same at publication time
- Using these concepts to describe data products will make your data more discoverable, provide additional ways to link to it
- We are now trying to validate the current thesaurus and could user your help, especially if you have additional use cases
- Does the UAT work for you? Great!
- Does it fall short? Help us make it better!
- Plan to use something else instead? Please talk to us first and tell us why.

Acknowledgments

- Chris Erdmann (NCSU) and Alberto Accomazzi (ADS) brought the UAT to the CfA, with help from Norman Gray (U. Galsgow)
- Katie Frey (CfA library) has assumed the role of UAT “gardner” since 2013 and is primarily responsible for its curation and state
- Chris Biemesderfer and Julie Steffen (AAS) gave the UAT a sanctioned home
- The following people were instrumental in curating different branches of the thesaurus: Sarah Weissman, Josh Peek, Jill Lagerstrom (STScI), Dianne Dietrich (Cornell), Jane Holmquist, Kayleigh Bohemier (Yale), Barbara Kern (U. Chicago)
- Dozens of astronomers at the institutions listed above offered their expertise in the development and validation of the thesaurus