# Integrating the UAT into the ADS

#### Alberto Accomazzi

aaccomazzi@cfa.harvard.edu

NASA Astrophysics Data System

Unified Astronomy Thesaurus Webinar | 8 December 2020







# Improving Discovery in ADS with UAT

- Better Understand what the user is looking for
- Properly identify concepts discussed in papers
- Disambiguate meaning of words in corpus
- Leverage curated knowledge graph
- Provide better insights to users

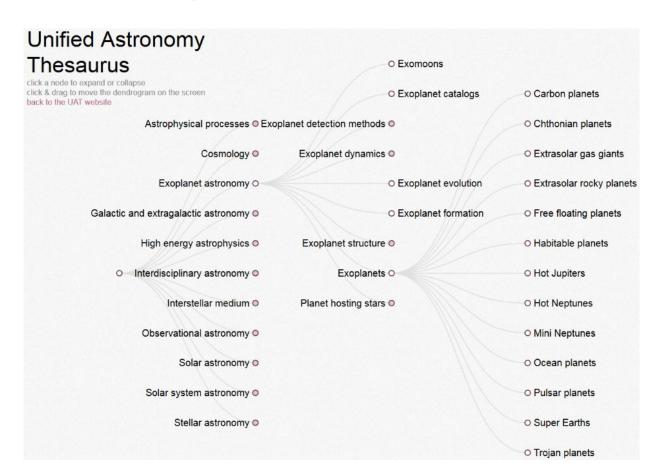
#### **Normalization**

```
gamma-ray bursts
Gamma-ray bursts (629)
Gamma ray burst [Gamma-ray bursts (629)]
GRB [Gamma-ray bursts (629)]
Gamma ray bursts [Gamma-ray bursts (629)]
Gamma-ray burst [Gamma-ray bursts (629)]
Î<sup>3</sup>-ray bursts [Gamma-ray bursts (629)]
Cosmic gamma ray burst [Gamma-ray bursts (629)]
Cosmic gamma-ray burst [Gamma-ray bursts (629)]
```

### **Disambiguation**

```
accretion
Galaxy accretion disks (562)
Galactic accretion disks [Galaxy accretion disks (562)]
Galaxy accretion discs [Galaxy accretion disks (562)]
Stellar accretion (1578)
Stellar accretion disks (1579)
Stellar accretion discs [Stellar accretion disks (1579)]
Accretion (14)
Particle accretion [Accretion (14)]
Galaxy accretion (575)
Bondi accretion (174)
```

# **Hierarchy**



# What about subject heading keywords?

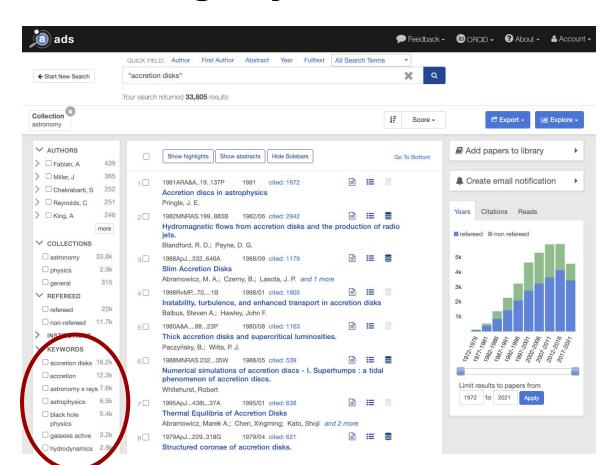
Incomplete: don't cover all subject areas

**Subjective**: assigned, not validated

**Ambiguous**: meaning not always grounded

**Sparse**: most ADS records don't have any

Not a good filter!



### What about discovery?

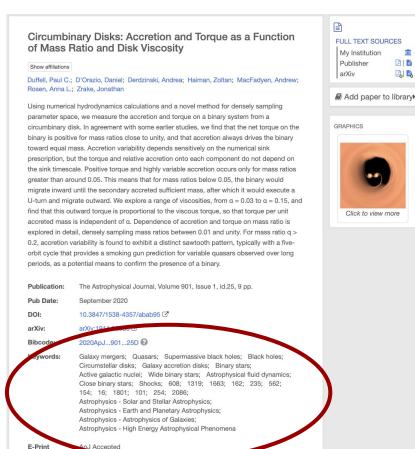
Currently ADS collects and uses all keywords available for each paper

Since multiple keyword systems are in use, cross-linking papers not effective

Ideally, with wide UAT adoption and labeling:

- 1. All papers have UAT terms assigned
- 2. Concepts used for filtering, browsing
- 3. Papers cross-linked via UAT terms:

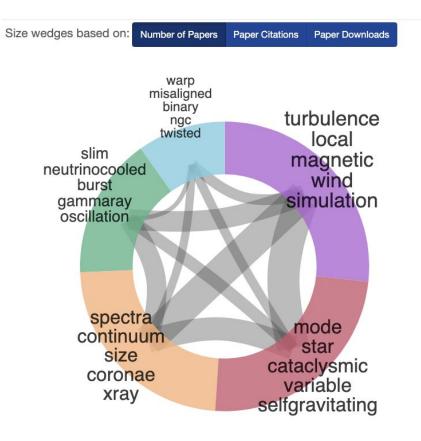
Galactic and extragalactic astronomy > Galaxy Physics > Galaxy Mergers (608)



Comments:

A | B

#### From words to concepts: paper networks



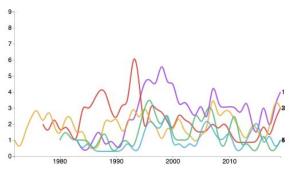
Summary

Detail

#### Paper Network for Query

The segments of the visualization to the left represent groups of papers from your result set which cite similar papers.

Group Activity Over Time (measured in papers published)



Click on a group to learn more about the papers within the group, as well as the papers cited by those papers.

Learn more about the paper network.

### From words to concepts: word clouds

```
region thin surface galactic discuss property present disc form model find include parameter disk rate large viscous inner hole effect object study

accretion time structure angular angular radius produce discuss form model find include pressure numerical study accretion time structure nucleus
                                     observation momentum value
```

#### Plan

- Promote the use of the UAT as a the standard system for assigning concepts to Literature, Data, Software
- Start enabling cross-linking of papers with UAT concepts in ADS in 2021, increasing their exposure to the users
- Use Machine Learning techniques to automatically assign UAT concepts to prior literature
- Use UAT concepts rather than extracted keywords in ADS visualizations, analytics
- Extend system to use concepts drawn from other thesauri for content in other disciplines as appropriate