

Laure Thompson
U. of Massachusetts Amherst

Building Large-Scale Collections of Genre Fiction



Curated list of works: author-title pairs



Extensive fan-built database of speculative fiction with rich work-level metadata:

- genre, subgenre
- award nominations & wins
- inclusion in notable book lists

18,809 works by 3,718 authors
published from 1900–2010

Book Type	# Works
Novel	13,585
Collection, Anthology, Omnibus	3,567
Novella, Novelette	1,657

Genre	% Works
Science Fiction	53.7%
Fantasy	33.7%
Horror	8.3%

Searching the HathiTrust catalog



Total: 3,241 works & 5,160 volumes

Automated Search

Produced 4,920 work-volume pairs

- 4,382 good matches
- 538 mismatches

Non-Exhaustive Manual Search

Identified 3,155 work-volume pairs

- 778 missed by our automated search

Main causes of mismatches:

- Distinct works with similar titles
- Overlapping content for compilations & contained works

Main causes of misses:

- Series-level records
- Title & author name variation
- Alternate titles & author names

What did we find?



Mostly novels but also compilations

- 2,231 novels, 367 collections, 130 anthologies

Authors with the most work-level matches are fairly prolific

- Top 3: Robert Silverberg, Robert Heinlein, & Isaac Asimov

Out of copyright works are over represented at the volume level

Some prolific authors have little representation

- James Axler, Tanith Lee, C.J. Cherryh, & Mercedes Lackey

...also true for award-winning ones

- Only 2 works by Connie Willis

Other surprise absences:

- Octavia E. Butler's *Parable of the Sower* and its sequel
- Works by S.P. Somtow

Beyond the catalog: content-based methods

We use HathiTrust Extracted Features to produce volume-level signatures which allow us to identify duplicate works and ignore bad catalog-level matches.

Sim.	HTID	Title / Author	Year
1	nyp.33433112045251	The chessmen of Mars / by Edgar Rice Burroughs ... ; illustrated by J. Allen St. John	1922
0.988	osu.32435017883182	The chessmen of Mars / by Edgar Rice Burroughs ... ; illustrated by J. Allen St. John.	1922
0.802	osu.32435017174004	Thuvia, maid of Mars / by Edgar Rice Burroughs, illustrated by J. Allen St. John.	1920

Sim.	HTID	Title / Author	Year
1	mdp.39015013315810	The Hugo winners, edited by Isaac Asimov.	9999
0.963	pst.000012384754	The Science fiction hall of fame.	9999
0.962	mdp.39015000656127	Dangerous visions; 33 original stories. Illus. by Leo and Diane Dillon.	1967