



# Perceptions on Aboutness of Documentary Films:

## Comparing FAST Headings to User-Created IMDb "Plot Keywords"

Christopher S. Dieckman, M.L.I.S., Iowa State University, Ames Iowa

dieckman@iastate.edu ORCID ID: 0000-0002-3877-6387



### Introduction

This study assesses the extent to which aboutness, as indicated by Faceted Application of Subject Terminology (FAST) headings, in WorldCat records is covered by user-created "plot keywords" in the Internet Movie Database (IMDb).

### Abstract

Many popular websites allow users to assign descriptive metadata to individual resources. Previous studies have analyzed user-created metadata with several comparing it to metadata created by information professionals. This study analyzed Faceted Application of Subject Terminology (FAST) headings and IMDb "plot keywords" for the same documentary films in order to determine how much of the aboutness identified by the former was covered by the latter.

A randomized list of documentary films held at the Iowa State University Library, a large academic research library, was created. Records for titles without IMDb headings were excluded from analysis. The FAST headings were compared to each "plot keyword" used in the corresponding IMDb record to determine if they met one or more of the match types defined for this study: *exact match*, *close match*, *closely-related match*, *broad match*, and *narrow match*. In total, 604 FAST headings from 100 OCLC records were analyzed. This study found that relatively little of the aboutness expressed by FAST headings was represented by "plot keywords" in the corresponding IMDb records, with only 22.19% of FAST headings containing some type of match.

By advancing our knowledge of the similarities and differences between subject metadata created by professionals and non-professionals for documentary films, this work supports the development of future innovations to increase the discoverability of motion pictures and other resources.

### IMDb and "Plot Keywords"

Founded in 1990<sup>1</sup>, the Internet Movie Database (IMDb) contains information for 6,534,894 movies, television episodes, and other media titles as of June 2020<sup>2</sup>.

In addition to other descriptive metadata, IMDb uses user-supplied "plot keywords"—a word (or group of connected words) attached to a title (movie / TV series / TV episode) to describe any notable object, concept, style or action that takes place during a title<sup>3</sup>.

Figure 2 shows some of the "plot keywords" users added for the 2005 documentary film "Time of Fear"<sup>4</sup>.



Figure 2. Some of the descriptive metadata from the IMDb page for "Time of Fear"<sup>6</sup>

### Storyline

In World War II, more than 110,000 Japanese Americans were forced to leave their homes and relocate to military camps. This documentary tells the story of the 16,000 men, women and children who were sent to two camps in southeast Arkansas, one of the poorest and most racially segregated places in America. *Written by Anonymous*

Plot Summary | Add Synopsis

Plot Keywords: [japanese american](#) | [arkansas](#) | [f rated](#) | [xenophobia](#) | [world war two](#) [See All \(9\)](#) »

Genres: [Documentary](#)

### Study Background

This study was inspired by previous research on user-created metadata—particularly, studies comparing it to metadata created by information professionals (examples include Rolla (2009)<sup>7</sup>, Peter (2012)<sup>8</sup>, and White (2014)<sup>9</sup>. In addition, Naun and Elhard (2005) compared descriptive metadata in IMDb to that found in MARC records<sup>10</sup> including commentary on "plot keywords" in subject access<sup>11</sup>.

Documentary films held by Iowa State University were chosen because the author wanted to analyze titles likely to be owned by a large academic institution and the list was readily available. Faceted Application of Subject Terminology (FAST) was selected as the standard to use for comparison because its relatively simple syntax led the researcher to speculate that they would be closer to user-created metadata and because they are comprehensive in subject coverage. Finally, IMDb was chosen because it is a popular website, allows for user-created metadata, and contains records for a large number of motion pictures increasing the likelihood that titles held by the library would be found.

### Research Questions

The research question was as follows: Is the aboutness expressed by FAST headings in OCLC WorldCat also represented by user-created "plot keywords" in the IMDb for a chosen set of documentary films?

More broadly, it sought insight into whether existing library subject metadata can enhance records for information resources, particularly audio/visual materials, in the IMDb and similar Web resources for increased discovery.

### Sample Selection

#### Selection of documentary titles

Iowa State University Library's catalog was searched for all audio-visual materials with the LCGFT "documentary films". The list order was randomized and, starting from the beginning, each title was analyzed to confirm that it was correctly identified as a documentary film and that its respective title had an IMDb record. This continued until 100 records were selected for analysis.

#### Selection of FAST headings

The FAST headings were taken from their OCLC records adding to a total 604 individual headings—this did not include headings used as genre/form terms (i.e., the 655 field), which do not express aboutness and were therefore excluded from analysis. Each of the 100 OCLC records contained at least one FAST heading (least = 1, greatest = 20, average = 6.04, median = 5, mode = 4).

#### Selection of "plot keywords"

IMDb "plot keywords" were taken from all records in which they were present (69 of the 100 records). Collectively, the IMDb records contained a total of 1,354 "plot keywords" with considerable range in the number of keywords used (least = 0, greatest = 98, average = 13.54, median = 3).

### Analysis

After a cursory review of the FAST headings and "plot keywords," SKOS-inspired relationships were created: *exact match*, *close match*, *broad match*, *narrow match*, *closely-related match*, and *no match* (see *Table* for definitions).

All 604 FAST headings were individually compared to the "plot keywords" used in the corresponding IMDb record. If one or more of the "plot keywords" used for the IMDb record were related to the heading, the heading was analyzed, and the FAST heading was then coded in accordance with the appropriate relationship type. If multiple match types were found for a FAST heading, it was coded only with the first relationship type in the order in which they appear in the table. This process was continued until each FAST heading was analyzed.

After the initial analysis, minor changes were made to the definitions of some of the types to more precisely account for ambiguous cases.

### Relationship Type Definitions

Table. Relationship definitions. \*(Example is hypothetical. All other examples were found in study.)

Relationship Type	Definition	Example FAST Heading	Example IMDb "Plot Keyword"
Exact Match	FAST heading matches "plot keyword" exactly (IMDb "plot keywords" only use lowercase letters and cannot include diacritical marks; therefore, casing and diacritics were not considered)	Hydraulic fracturing	hydraulic fracturing
Close Match	FAST heading matches "plot keyword" exactly except in spelling or compound structure	Counterculture	counter culture
	FAST heading is plural form of "plot keyword" or vice versa	Hoax	hoaxes
	FAST heading is synonymous or near-synonymous to "plot keyword"	Ebola virus disease	ebola
	FAST heading differs from "plot keyword" in syntax and/or presence of a parenthetical qualifier, but there is no significant change in meaning	New York (State) -- New York	new york city
Broad Match	One or more of the defined variant (i.e., "used for") terms for FAST heading is an exact or close match to "plot keyword"	Public housing [FAST heading has variant "Government housing projects"]	housing projects
	FAST heading is an example of entity described by "plot keyword"	Glass, Philip	Composer
	FAST heading is a geographic region or jurisdiction located within place specified by "plot keyword"	Germany	Europe
Narrow Match	There is a single equivalent LCSH for the FAST heading, and it has a "broader term" that is either an exact or close match to "plot keyword"	Communicable diseases [equivalent LCSH has broader term "Diseases"]	Disease
	FAST heading has multiple equivalent LCSHS, each with respective authority records, and one or more have a term that is an exact or close match to "plot keyword"	Abortion--Religious aspects--Christianity	abortion; christianity [appearing as separate plot keywords]
Closely-Related Match	"Plot keyword" is an example of entity described by FAST heading	Public buildings*	Federal prisons*
	"Plot Keyword" is a geographic region or jurisdiction located within place specified by FAST heading	United States	virginia
No Match	There is a single LC equivalent for the FAST heading, and it has a "narrow term" that is either an exact or close match to "plot keyword"	Human ecology [equivalent LCSH has narrow term "Sustainability"]	sustainability
	There is a single LC equivalent for the FAST heading, and it has a "related term" that is either an exact or close match to the "plot keyword"	Race relations [equivalent LCSH has related term "Racism"]	racism
	FAST heading is a demonym for "plot keyword" or vice versa	China	chinese
No Keywords	FAST heading and "plot keyword" have a verb-noun relationship	Bus travel	bus trip
	FAST heading defines demographic group from geographic region or jurisdiction specified by "plot keyword" or vice versa	Palestinian Arabs	palestine
No Keywords	FAST heading describes an action performed by type of agent specified by "plot keyword" or vice versa	Art forgers	art forgery
	No "plot keywords" in the IMDb record meet match criteria for FAST heading	--	--
No Keywords	IMDb record does not contain any "plot keywords"	--	--

### Results

The study found that 69% of the IMDb records contained one or more "plot keywords" and 44% contained at least one "plot keyword" related to a FAST heading in the corresponding OCLC record.

Of the 604 FAST headings, 135 (22.35%) contained a match, 252 (41.72%) had *no match*, and 217 (35.93%) had *no keywords*. The percentage of matches by type were as follows: *exact match* (n=32, 23.70%), *close match* (n=37, 27.41%), *closely-related match* (n=10, 7.41%), *broad match* (n=45, 33.33%), and *narrow match* (n=11, 8.15%).

Figure 3. (Rounded to nearest percentage point for clarity.)

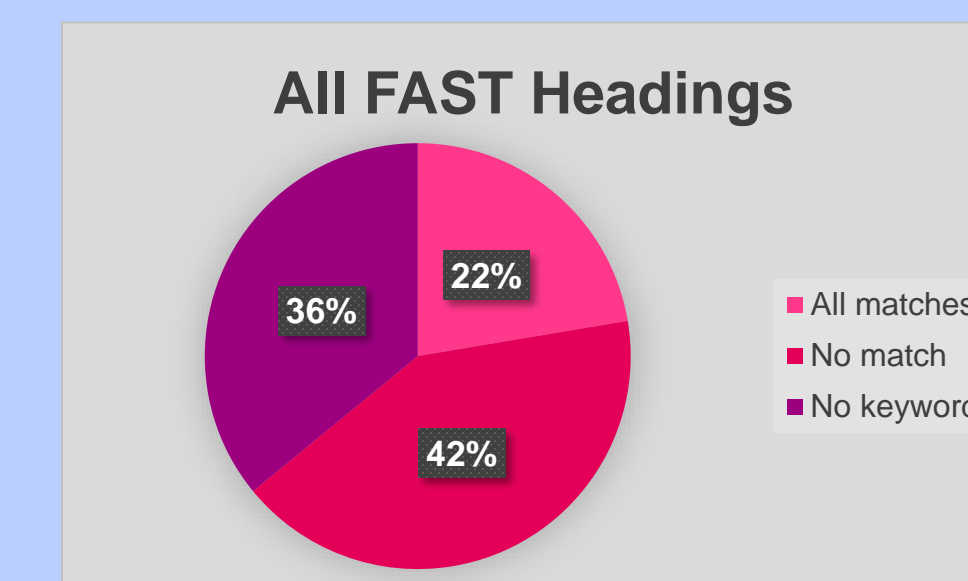
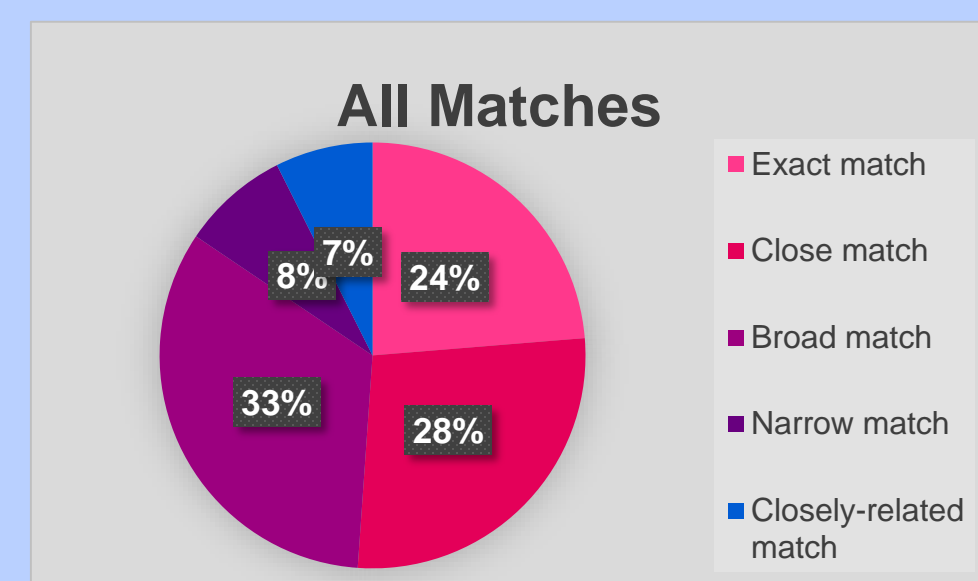


Figure 4. (Rounded to nearest percentage point for clarity.)



### Conclusions

While the majority of IMDb records contained at least one "plot keyword", the number of FAST headings with matches was relatively small. Notably, only 11.42% of headings had a *close* or *exact match*.

This evidences that most of the subject matter expressed in FAST headings is not represented by IMDb "plot keywords" for the same documentary films. It may also suggest that subject metadata created by librarians and information professionals is covering subject matter that is usually not represented by user-created metadata in popular websites; however, because of the narrow focus of this study, additional research is needed to test this hypothesis.

This leaves several other questions open for future research. Would similar results be found if FAST headings were compared to different resource types or if a different source of user-created metadata was used? What, if any, patterns exist in the aboutness of user-created metadata? Would expressing the same aboutness chosen by librarians and information professionals affect user discovery and access of resources on popular websites?

### Acknowledgement

The creator of this poster wishes to acknowledge Harriet Wintermute for her advice and recommendations on this project.

### References

1. *Encyclopædia Britannica*, s.v. "IMDb," by Robert Lewis, October 31, 2017, <https://www.britannica.com/topic/IMDb>
2. "Press Room," IMDb, accessed September 10, 2020, <https://www.imdb.com/pressroom/stats/>
3. "IMDb, Help," IMDb, accessed September 16, 2020, <https://help.imdb.com/article/contribution/titles/keywords/GXQ22G5Y72TH8MJ5#>
4. "Time of Fear (TV Movie 2005)," IMDb, accessed September 20, 2020, [https://www.imdb.com/title/tt0458502/?ref\\_=fn\\_al\\_tt\\_3](https://www.imdb.com/title/tt0458502/?ref_=fn_al_tt_3)
5. *Ibid.*
6. *Ibid.*
7. Peter J. Rolla, "User Tags Versus Subject Headings: Can User-Supplied Data Improve Subject Access to Library Collections?," *Library Resources and Technical Services* 53, no. 3 (2009).
8. Marija Peter, "Comparing User-Generated and Librarian-Generated Metadata on Digital Images," *OCLC Systems and Services* 28, no. 2 (2012).
9. Hollie C. White, "Descriptive Metadata for Scientific Data Repositories: A Comparison of Information Scientist and Scientist Organizing Behaviors," *Journal of Library Metadata* 14, no. 1: (2014).
10. Chew Chait Naun and K.C. Elhard, "Cataloguing, Lies, and Videotape: Comparing the IMDb and the Library Catalogue," *Cataloging & Classification Quarterly* 41, no. 1 (2005).
11. *Ibid.*, 36.