



Organization: OCLC
Primary mentor: Ray Karpman
Secondary mentor(s): Devon Smith, Jenny Toves

Project Title	Mapping DDC to FAST
Description	The desired outcome will be a suggested Dewey number for each FAST topical. The strength of the mapping will be affected by the occurrence counts and by whether there is a strong “winner” or several equally likely options. The report at the end should indicate how many FAST topicals can be mapped with what degree of strength – how many have a clear value, how many have several possible values, how many are too infrequently used to make a strong statement about the mapping.
Problems/ Research Questions	A book can be cataloged in a variety of ways and can have a variety of subjects applied to it from differing vocabularies. Mapping vocabularies allows us to gain a more comprehensive understanding of a works subject(s) and facilitates many downstream processes.
Techniques	Clustering, fuzzy matching.
Tools/ Languages used	Python, or any language the fellow can justify for the specific needs of the project
Data	Description: The provided data will be Dewey numbers and FAST topicals (tag 650) that co-occur in Worldcat bibliographic data and will include an occurrence count for the pair of values. Data Type: CSV Data Size: Millions of rows

LEADING 2023

Outcome	A file or database matching FAST to Dewey with a strength metric.
Milestone Timeline	Month 1: Introduction to material, literature review, getting systems and data access. Months 2-4: Analysis, iterative development and experiment Months 5-6: Synthesis and presentation
References	https://fast.oclc.org https://www.oclc.org/en/dewey/resources.html