Data Pipeline

ORCID ←Cross-match by title→ astrophysics data system
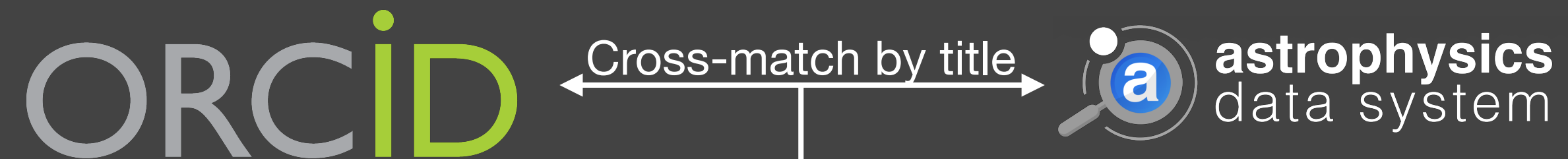
554,962 matches for researcher records

e.g., [Author Name, Affiliation, Title, Abstract]

Data Representation

e.g., [Vicente Amado Olivo, Michigan State University, ...]

Author Names and Affiliation
Character-level Representation

Title and Abstract
Semantic Representation

A = Chars2vec(Li, Zhao..) + SPECTER(FeAmGen.jl: A Julia program for Feynman Amplitude Generation)

$$\text{Similarity}(\mathbf{A}, \mathbf{B}) = Cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|}$$

B = Chars2vec(Li, Zhao…) + SPECTER(Qualitative analysis and modulation instability for the extended …)

Baseline Model

Author Record #1

Chars2vec(Li, Zhao..) + SPECTER(FeA

Feed Forward NN

Embedding #1

Shared Weights

Distance Metric → Similarity

Author Record #2

Chars2vec(Li, Zhao…) + SPECTER(Quali

Feed Forward NN

Embedding #2