Autor Disambiguation with the Help of Wikidata

Adam Bannister

Outline

zbMATH
the first resource for mathematics

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

About

zbMATH

Zentralblatt MATH (zbMATH) is the world's most comprehensive and longest-running abstracting and reviewing service in pure and applied mathematics.

Me

I work on the project Scalable Author Disambiguation.

zbMATH
the first resource for mathematics

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

Outline

Introduction

Overview

- **Problem**
- **Aims**
- **External IDs**
- **Applications**

zbMATH

the first resource for mathematics

FIZ Karlsruhe

Leibniz Institute for Information Infrastructure

Outline

Problem

❑ Homonyms

- Müller, Michael
- Multiple Chinese characters with one transcription, eg. Wei

❑ Synonyms

- Chebyshev, Tchebychev, Tschebyscheff
- Name change due to marriage, fleeing from persecution

zbMATH
the first resource for mathematics

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

Outline

Aims

To date we can intellectually decide who wrote what but we need to scale.

With external IDs we gain metadata such as birth date, PhD date, etc.

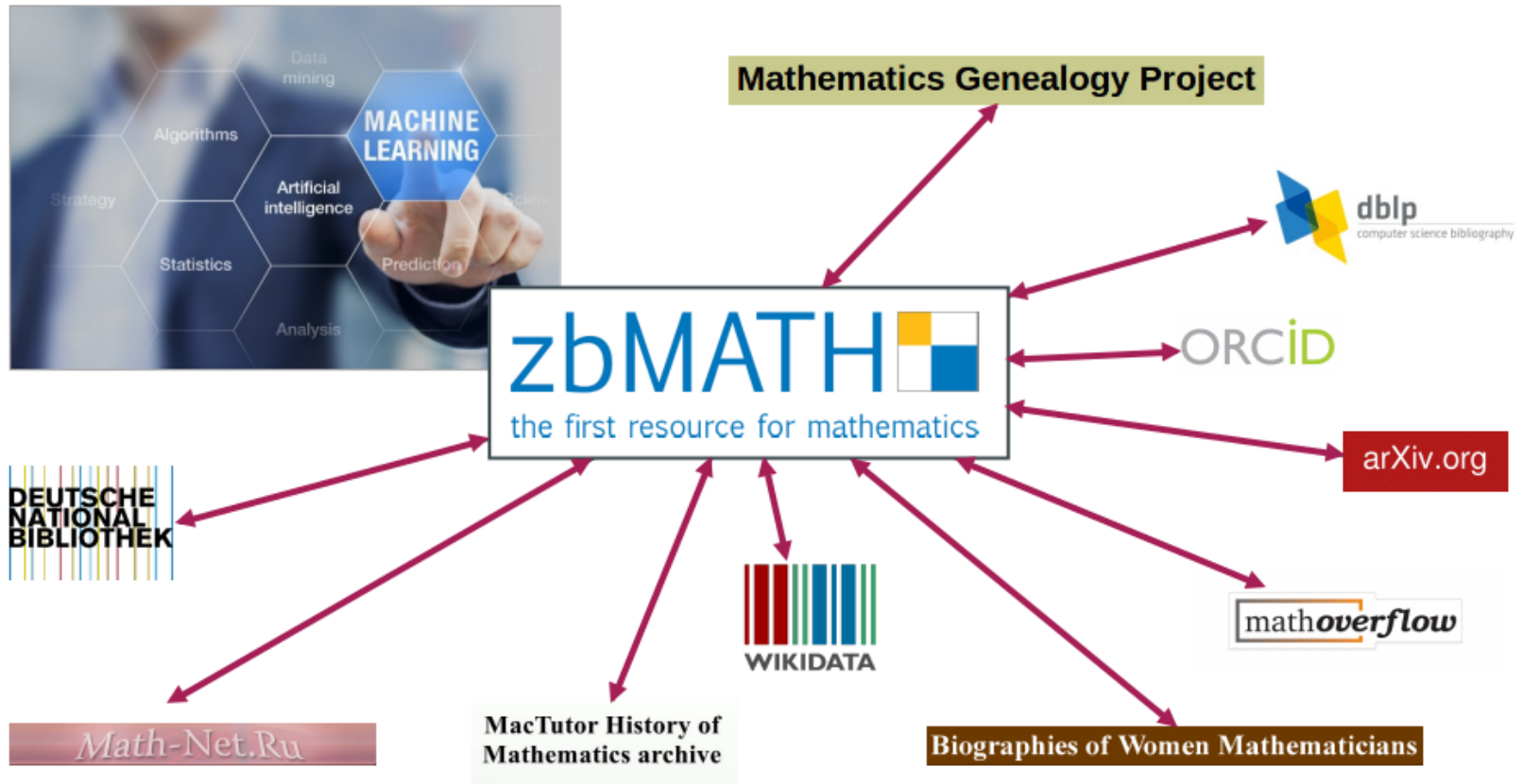Using this information we can exclude publications from an author's profile.

Autor Disambiguation with the Help of Wikidata
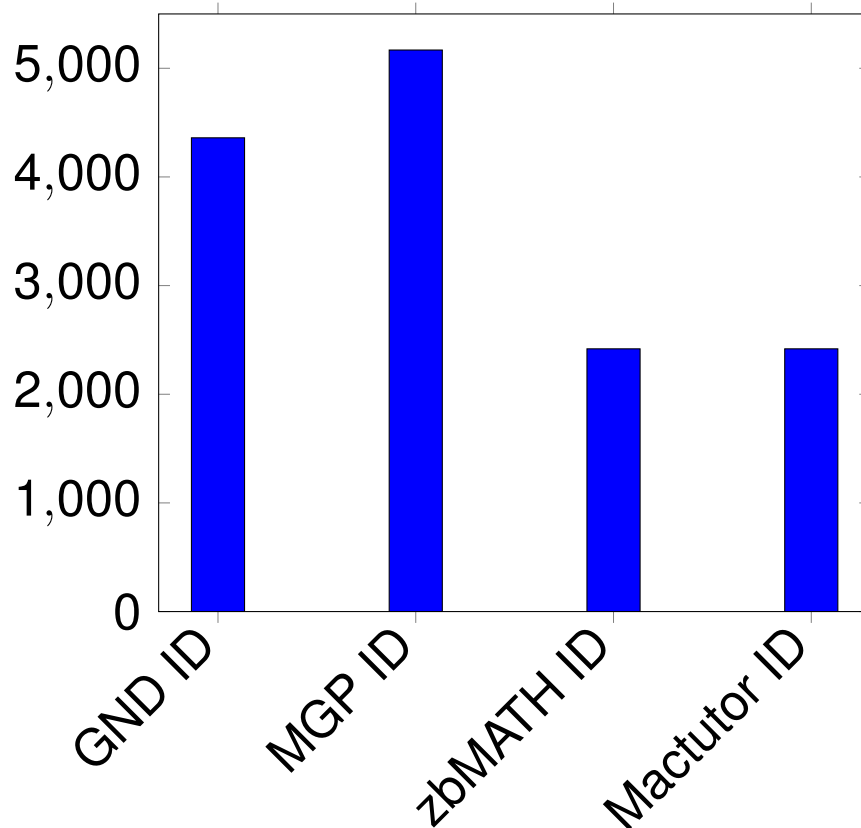
zbMATH
the first resource for mathematics

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

Outline

zbMATH
the first resource for mathematics

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

External IDs

Autor Disambiguation with the Help of Wikidata

External IDs

Taken from dataset of 8643 Wikidata persons (probably mathematicians)

Source: https://www.kaggle.com/joephilleo/mathematicians-on-wikipedia

Autor Disambiguation with the Help of Wikidata

zbMATH
the first resource for mathematics

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

Outline

Introduction

Overview

- **Problem**

- **Aims**

- **External IDs**

- **Applications**

Applications

Wikidata Timeline built in to zbMATH Author profile:

Paul Erdős.

Autor Disambiguation with the Help of Wikidata

Applications

Another use for wikidata is to secure other ids.

❑ We have links to:

- zbMATH
- MGP
- GND
- and many more.

Autor Disambiguation with the Help of Wikidata

zbMATH
the first resource for mathematics

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

Applications

If we have an MGP matched with a zbMATH ID we can find the wikidata ID.

1. We write simple sparql query that gives us all wikidata / MGP id pairings.

2. Go through our database and assign authors who have an MGP with the new wikidata id.

Positives:

❑ Extra event data

❑ Extra affiliation data

❑ Extra external data

zbMATH
the first resource for mathematics

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

Applications

One problem with this is when the source information is changed.

Domino Effect

❑ Wikidata ID
- ● MGP ID
  - ➜ PhD year
- ● Birth year

Autor Disambiguation with the Help of Wikidata

Applications

Fun idea:

Using Wikidata and zbMATH information to predict the next Fields' Medal winner.

- ❑ Younger than 40
- ❑ Coauthors
- ❑ Affiliations
- ❑ Fields of work

Outlook

What we have planned:

❑ Automatically save zbMATH IDs to Wikidata persons.

❑ Make better use of the semantic information (affiliations, timeline).

zbMATH
the first resource for mathematics

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

Thank you!

zbMATH
the first resource for mathematics

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure