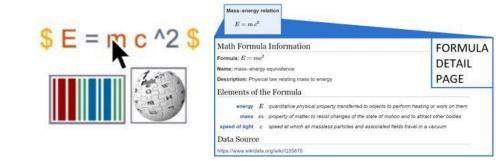
## MR01 Wikipedia Formula Annotation

### Background

Mathematical Question Answering systems, such as <a href="https://mathqa.wmflabs.org">https://mathqa.wmflabs.org</a> require labeled formulae. However, very little data is available so far.



#### Goal

• Develop a formula annotation recommender system for Wikipedia.

### **Tasks**

- Build an API for the AnnoMathTeX (<a href="https://annomathtex.wmflabs.org">https://annomathtex.wmflabs.org</a>)
   annotation recommendations.
- Build a seeding service to transfer the annotated formula concept knowledge to Wikidata.
- Build an integrated system and deploy it to Wikipedia. If successful, you can proudly claim to be a Wikipedia developer.

Philipp Scharpf scharpf@gipplab.org



Moritz Schubotz Moritz.Schubotz@ fiz-karlsruhe.de



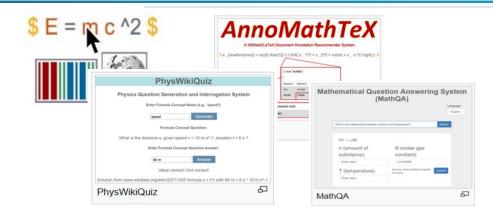


# MR02 A Large Dataset for Mathematical Entity Linking

## Background

Mathematical Information Retrieval systems (MathIR) and Machine Learning applications require labeled formulae.

However, very little data is available so far.



### Goal

Develop a large annotated dataset of annotated (linked) mathematical formulae.

### Tasks

- Research design principles for large annotated datasets (annotation guidelines).
- Annotate formulae in a corpus (e.g., Wikipedia articles or arXiv documents) using the AnnoMathTeX formula and identifier annotation recommender system.
- Test selected MathIR systems (e.g., MathQA or PhysWikiQuiz) on your dataset.
- Document insights from the annotation and test process to improve the systems.





Moritz Schubotz Moritz.Schubotz@ fiz-karlsruhe.de



