Promotion MM for Banner

# Overview of the project

The millions of biomedical publications that exist are a valuable, but difficult to handle resource. Identifying those documents that are most relevant to a particular disease or health condition is currently a costly, human intensive activity. The goal of this MM is to develop new algorithms to aid in the automated Named Entity Recognition (NER) of biomedical publications.

# Overview of the problem

The National Institute of Health (NIH) has built a system on how crowd labeling can be used to annotate abstracts from Pubmed so disease characteristics can be identified. This open-source, supervised learning system called BANNER achieves a good level of prediction power.

After training on about 500 abstracts manually annotated by *experts*, BANNER currently accomplishes this task with precision and recall around 0.8. While the result is successful, the training capabilities of the current algorithm are restricted to a very small (expert) dataset, which is limited by expensive expert's time. There is an idea that this limitation can be overcome if we teach BANNER how to further improve its accuracy by training on abstracts annotated by *non-experts* (Mechanical Turkers), potentially available in much larger quantities.

The goal of this contest is to improve BANNER accuracy by teaching it on MTurk-annotated abstracts.

# Why members should be interested in?

It is an experimental contest. If you participate, you will be given the opportunity to compete in new and fun types of MM competition.

This an exclusive event for-rated-only members (MM or Algo) and participation is limited to 300 registered members!

You will be competing in *small* virtual rooms. Room prizes will be awarded to the 1st and 2nd of each room, in addition to several grand prizes for the best competitors overall.