The NASA Tournament Lab at Harvard University and the Scripps Research Institute are excited to announce our next Marathon Match for TopCoder members. Reserved exclusively for rated members (MM or Algo) this nine day contest will challenge our community to solve an important information extraction problem in the biomedical domain. The contest will offer $41,000 in prizes and exclusive limited edition NASA Tournament Lab t-shirts.

Registration for the Marathon Match is between March 2, 2015 - March 5, 2015 and the contest launches on March 8, 2015. Below we provide you with a sneak peek of some of the details.

# Challenge Overview

Everyone is well aware of the explosion in data, information, and knowledge within the life sciences literature. Every year hundreds of thousands of new scientific publications are added to the millions of existing articles (e.g. the PubMed database) making the entire corpus of the medical literature a very valuable resource. However, identifying specific documents that are most relevant to a particular disease or health condition is currently a costly, error-prone, human intensive activity.

The goal of this Marathon Match is to develop new algorithms to aid in the automated Named Entity Recognition (NER) of biomedical publications. To accomplish these tasks effectively, algorithms are needed that can learn to accurately merge data collected from multiple annotators of varying quality and integrate this data into predictive models.

# Problem Overview

The United States National Institute of Health (NIH) has built a system that uses expert labeling to annotate abstracts from Pubmed so disease characteristics can be more easily 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 current results are an important advance, the training capabilities of the current algorithm are restricted to a very small (expert) dataset, and is further constrained by relying on experts to generate the label.

The Scripps Research Institute is investigating if 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).

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

# Why Should You Be Interested?

Like other challenges sponsored by the NASA Tournament Lab, this contest is quite challenging, highly practical and has the potential to dramatically improve the state of the art in information extraction from textual data. The top solutions will be implemented and exposed to a broad community of information science researchers.

The contest format is experimental and if you participate you will be given the opportunity to compete in new and fun types of MM competition. This is 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.

# Prizes & T-shirts

* $41,000 in prizes
* 24 1st place prizes of $1,000 each
* 24 2nd place prizes of $200 each
* 3 grand prizes of $4,000 each
* In addition to the prizes listed above, registered competitors will be awarded a special, limited-edition t-shirt upon completion of a brief post-event survey.

# Schedule

* 03/02 - Registration opens
* 03/05 - Registration closes
* 03/08 - Marathon Match begins and problem statement revealed
* 03/16 - Marathon Match closes

Don't forget to stay up to date on BANNER as well as all other topcoder happenings by subscribing to our blog.

We look forward to your participation.