

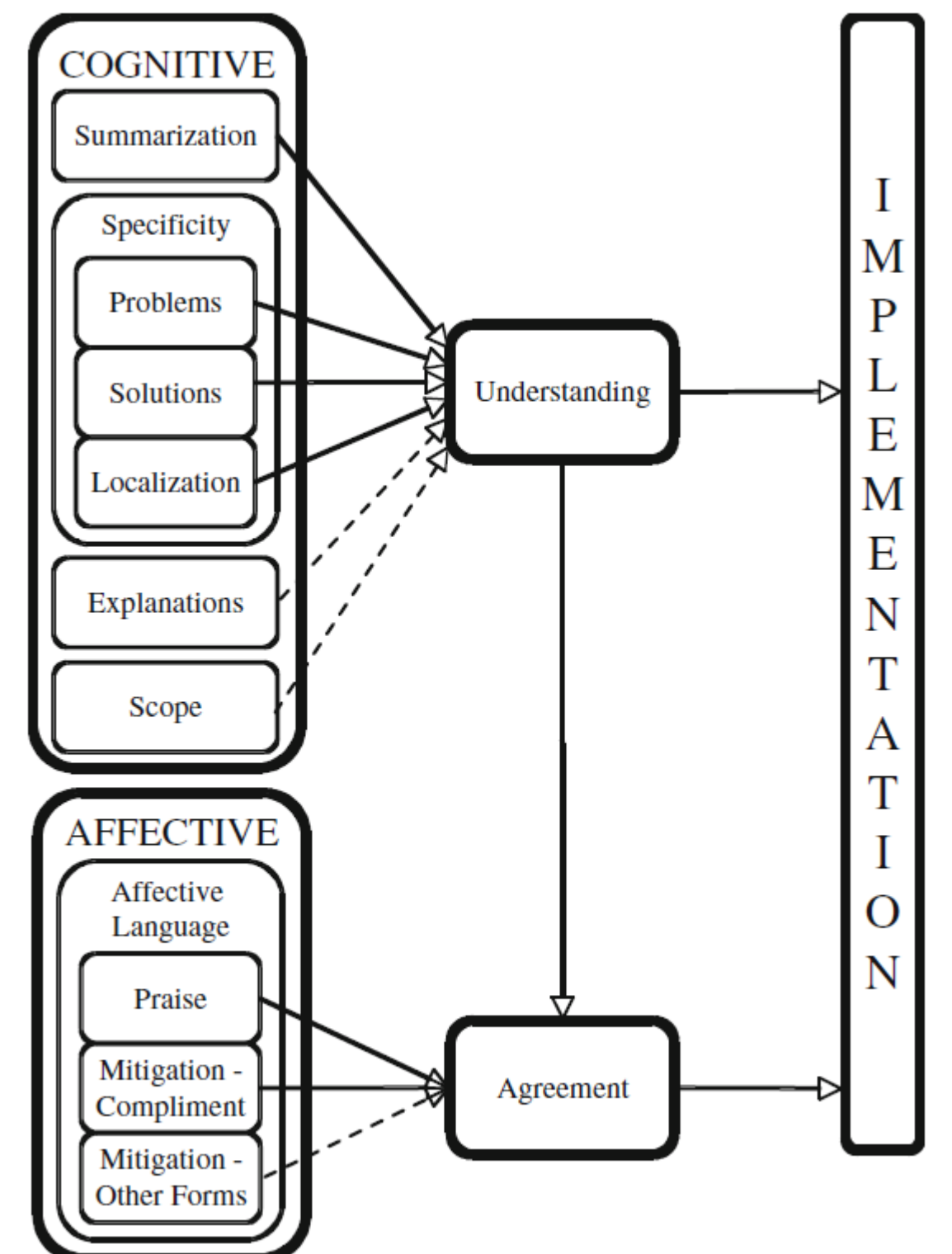
## Educational Data Mining

- Peer assessment analysis

With more and more people switching on to online platform seeking knowledge, MOOC becomes a popular way of learning. Due to the amount of participants, there's hardly anyway that a TA or a team of TAs would be able to grade everyone's work on time. Most MOOC platforms adapted a peer review system to let participants grade each other's work, but there's hardly any quality control on the content of reviews.

## Problem Definition

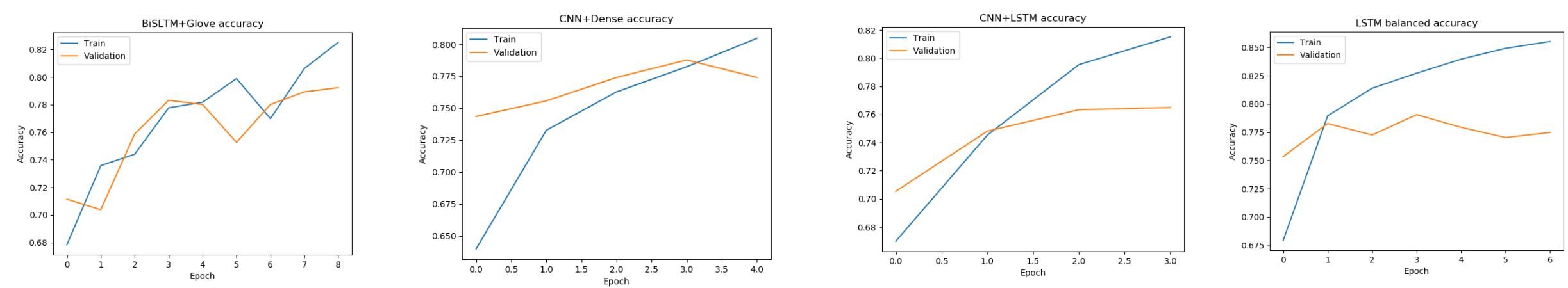
The lab has been working on machine learning towards detecting numerals language features in peer reviews. We focus on things that would make a review valid, such as whether it mentioned a problem of the submitted work; did it point out where the problem is; was there any solutions given, etc.. These are the things that truly helps reviewer to improve & re-implement their work.



[1] Nelson and Schunn, 2009

## Methodology

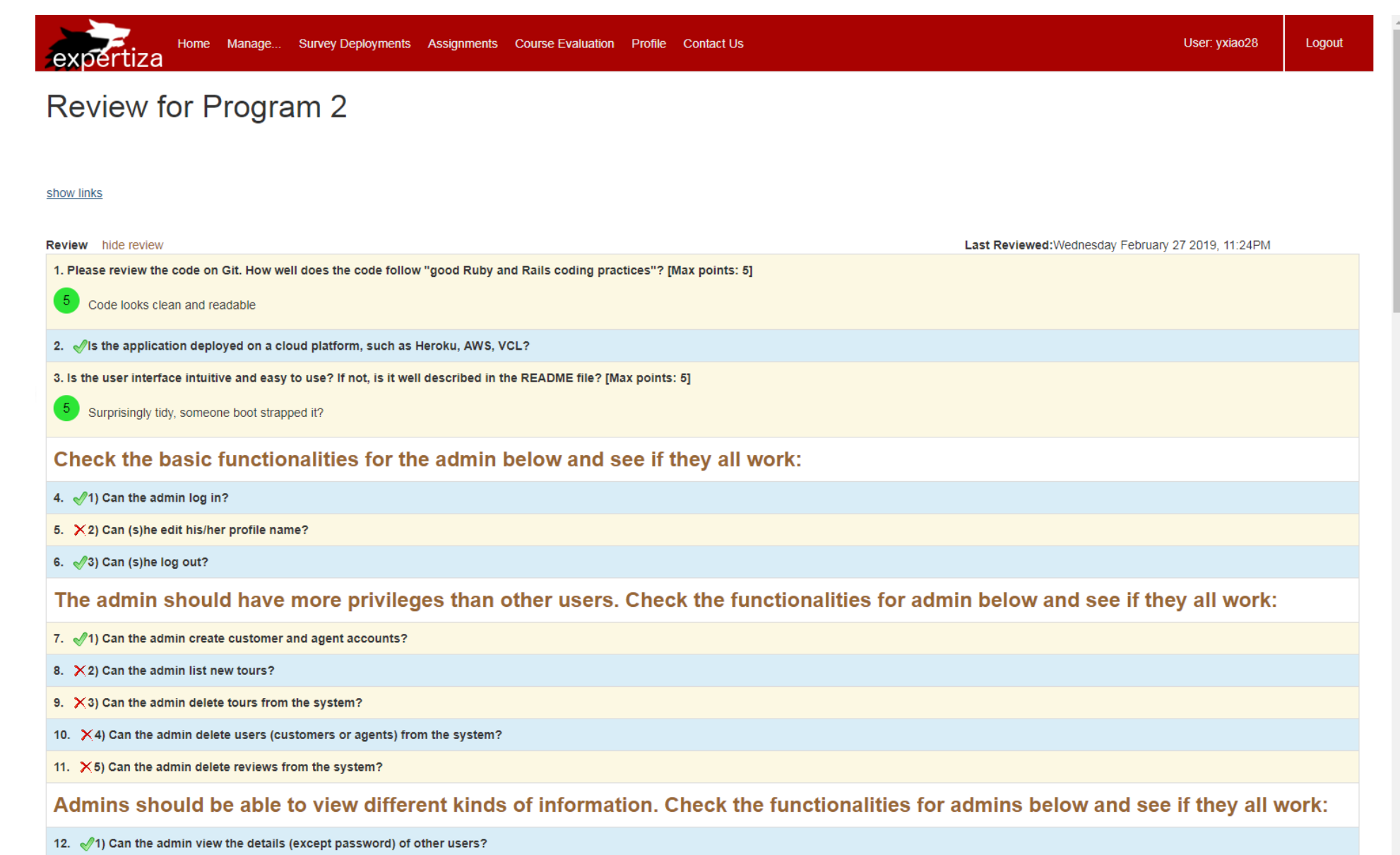
The process of detecting language features from text involves several different technical component, we apply a variety of **Machine Learning** models, **Data Mining** techniques as well as **Natural language Processing** skills towards this problem. In the past, we have applied Neural Network, specifically LSTM, CNN, Bi-LSTM, as well as some traditional Machine Learning methods such as Bayes, Decision tree & Random Forest, Regression Classifiers and Rule-based Classifiers.



## Expertiza

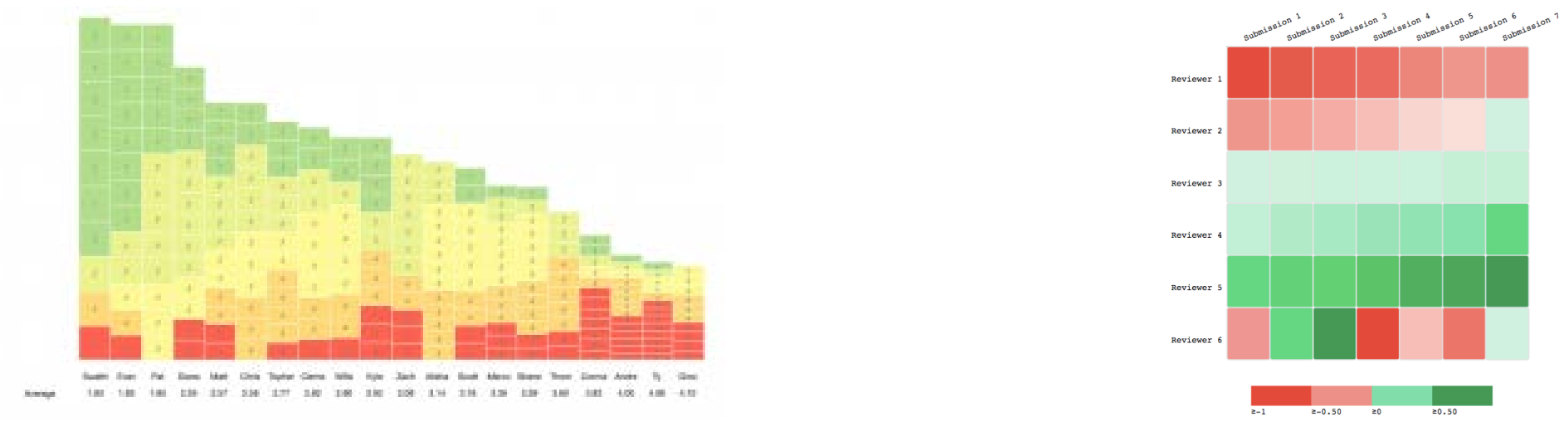
### The Active Learning Platform

The Expertiza platform is a divide-and-conquer approach to producing reusable learning objects through active-learning exercises. Students select from a list of tasks to be performed, with several students selecting each task. They prepare their work and submit it to an electronic peer-review system. The work is reviewed by other students, who offer comments to help the submitters improve their work. The best submissions for each task are selected for use in later courses. These learning objects can be improved iteratively; next year's class can refine and improve the previous year's contributions. Expertiza gets students working together to improve others' learning experiences. It helps them learn; by making them think through the lecture material and apply it to a real-world situation, just as they might do on the job. Because the interactions are asynchronous over the web, distance education students can participate on the same basis as on-campus students.



## Benefit from usage

Through using Expertiza, students could gain valuable insights from their peers, instructors and TAs could save time organizing the class. Visualization helps everyone to have a bird eye view if the status and progress of class thus being able to make improvements over time.



Besides the above perks, we were able to use **data** collected to do **Machine Learning** studies. Continuous improvements on the platform involves optimizing performance, advancing visualization, putting our research out comes into the system to help students writing more helpful reviews, etc.. All of these involves putting up some flavor of **web services** or **ruby on rails** developments.

## Peer-logic Data Warehouse

With all these demands on mining data, securing enough becomes critical. The peer logic data warehouse is created to serve this purpose.

### Data Source

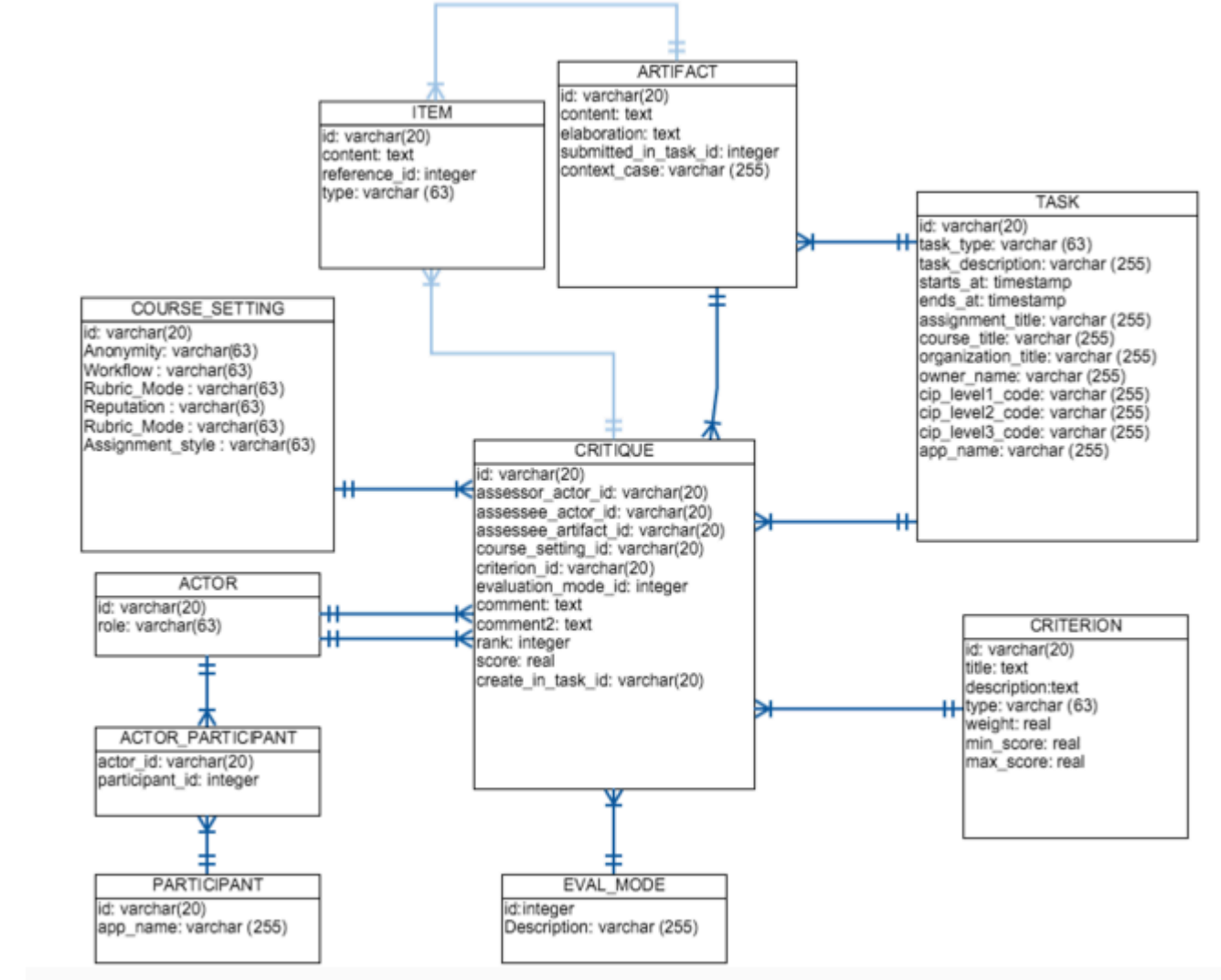
Out there, there are institutions with already well-established learning systems similar to Expertiza, such as the SWoRD system at the University of Pittsburg, CritViz at Arizona State University. If we could collaborate with all the universities having such system, we would end up with more than enough data for our research. As a matter of fact, a lot of published study only have a dataset with around 1000~ reviews in it.



In the past, we have collaborated with Arizona State University and received data from the CritViz system. With some aggregation and **ETL in SQL**, we have successfully constructed a warehouse with student reviews and progress in it. As we expect more **data** to come, the **warehouse** would be **expanded** and **organized**.

## The PRML language

PRML was used to define a data warehouse model that can be used to share data from different peer-review systems. We designed the schema based on dimensional modeling approach. Dimensional modeling requires the data that contains measurements, metrics, or facts of the business process to be stored as Fact tables.



The Fact tables also contains foreign keys to the dimension tables that can be used to group the facts into multidimensional arrays of data, known as OLAP cube or hypercube. Dimensional modeling encourages data warehouse schema to follow a star topology, in which fact tables are placed in the center.