Learning Script Knowledge with Web Experiments

Scripts are fundamental pieces of commonsense knowledge that describe stereotypical event sequences of human activities (like "eating in a restaurant" or "visiting a doctor"). There have been a couple of attempts to learn script data from corpora, however, many scripts are shared implicit knowledge and usually not elaborated in detail. We present an approach for learning this kind of script knowledge by "crowdsourcing" the Internet and generalizing over the gathered data instances. We collect natural-language descriptions of script-specific event sequences from volunteers over the web. With this data, we feed an algorithm that computes a graph representation of the script’s temporal structure using a multiple sequence alignment algorithm.

The evaluation of our system shows quite promising results for this explorative study.