

Никола Обрешков

Институт за български език „Проф. Л. Андрейчин“
при Българската академия на науките
София, България

КЛАСИФИКАЦИЯ НА ВЪПРОСИТЕЛНИ ИЗРЕЧЕНИЯ ЗА БЪЛГАРСКИ ЕЗИК

(Резюме)

Тази статия разглежда въпроса за класификация на въпроси за български език. Два експеримента са проведени върху корпус от около 1400 въпроса, тагирани ръчно, според таксономията разработена от Ли и Рот. Изпробвани са три популярни обучаващи се алгоритма (machine learning algorithms): най-близки съседни k (k -nearest neighbors – означен като k NN), алгоритъм, базиран на теоремата на Бейз (naïve Bayes – NB) и дърво на решенията (decision tree – DT). Системата WEKA (Хол и др. 2009) се използва като платформа за провеждане на експериментите, а резултатите са измерени по метода на десетократната кръстосана проверка (10-fold cross-validation).

✉ Никола Обрешков
vocabo@abv.bg

Публикувано: 31 март 2010