

Георги Илиев

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

СЪОТНАСЯНЕ НА ПАРАЛЕЛНИ КОРПУСИ С ПРИЛАГАНЕ НА МОДЕЛИ ЗА СТАТИСТИЧЕСКИ МАШИНЕН ПРЕВОД

(Резюме)

Статията разглежда класическите модели за статистически превод на IBM, които намират широко приложение в съотнасянето на паралелни текстове по думи в съвременето. Представя се процедурата за машинно самообучение с алгоритъм от тип EM за оценка на параметрите на IBM модел 1 с примери от експериментален българско-шведски паралелен корпус. Прави се преглед на параметрите на модели IBM-M2 до IBM-M5 и на процедурите по съставяне на извадките, от които се научават тези параметри. Математическите модели са дефинирани асиметрично, ето защо обсъждането завършва с представяне на някои от предложените в литературата методи за обединяване на пословните съотнасяния, научени в двете посоки на превод, във фразово съотнасяне.

Ключови думи: статистически машинен превод, съотнасяне на паралелни корпуси, преводен модел, машинно самообучение, EM, IBM

✉ Георги Илиев
georgi@dcl.bas.bg

Публикувано: 18 април 2013