

Светлозара Лесева, Ивелина Стоянова
Институт за български език „Проф. Л. Андрейчин“
Българска академия на науките
София, България

КЪМ АВТОМАТИЧНОТО РАЗРЕШАВАНЕ НА РЕФЕРЕНЦИИ В БЪЛГАРСКИЯ ЕЗИК

(Резюме)

Статията представя първите стъпки към разработването на система за автоматично разрешаване на анафората за български език. В настоящото изследване се разглеждат следните типове анафора: (а) местоименна анафора и (б) неизразен подлог. Системата се основава на правила, които отразяват редица особености на българския език като свободния словоред, елипсата при подлога, безличните изречения и съгласуването. Правилата са два вида: филтриращи, които налагат ограничения и така се отстраняват невалидните кандидати, и ранкиращи, които служат за сортиране на валидните кандидати и подбор на най-надеждния от тях. За различните анафори се използват различни комбинации от така формулираните правила, свързани с обсега за откриване на референции, както и със синтактичните и морфосинтактичните особености. Методът е приложен върху семантично аотирания корпус на български език БулСемКор и резултатите са съпоставени с базовия метод, който използва само съгласуване и приписва референция на най-близкия възможен кандидат. Резултатите имат предварителен характер.

Ключови думи: анафора, разрешаване на анафората, автоматични методи, български език

✉ Светлозара Лесева
zarka@dcl.bas.bg

✉ Ивелина Стоянова
iva@dcl.bas.bg