

РЕФЕРАТ

Магістерська дисертація присвячена проектуванню та розробці кластеризованої лінгвістичної бази знань з використанням *skipping n-grams*. Лінгвістична модель, що розробляється, є складовою частиною роботи програм з розпізнавання мовленнєвих сигналів і має за мету поліпшення розпізнавання мовлення.

Об'єктом досліджень є моделі розпізнавання мовлення, особливості розпізнавання української мови, лінгвістичні моделі у складі генеративної моделі, методи кластеризації, засоби розпізнавання мовлення, програмні продукти та системи побудови лінгвістичних баз.

Предмет дослідження: вдосконалена модель кластеризованої лінгвістичної бази знань зі застосуванням *skipping n-gram* для розпізнавання природномовного тексту українською мовою.

В роботі розглянуто та проаналізовано існуючі методи розпізнавання мовлення, визначені переваги ПММ та проаналізовано варіанти побудови лінгвістичних моделей. Проведено ряд експериментів, які показали доцільність використання *skipping n-grams* для побудови кластеризованої лінгвістичної бази знань.

Результатом дипломної роботи є сукупність модулів, що реалізує математичну модель лінгвістичної бази знань. Розроблене ПЗ впроваджено у використання в МННЦ інформаційних технологій та систем НАН України та МОН України. Основні положення та результати роботи доповідалися на конференціях «ССИИ'2013» та «IAI'2015».

Робота складається з вступу, 5 розділів та висновків і налічує 75 сторінок. Містить 10 ілюстративних матеріалів, 4 таблиць, 4 додатки та посиляється на 11 літературних джерел.

Ключові слова: приховані марківські моделі, розпізнавання мовленнєвого сигналу, розпізнавання природно-мовного тексту, кластеризація, лінгвістична база, *k-skip-n-gram*, *bigram*, *unigram*.