

ПОБУДОВА МОДЕЛІ ЗНАНЬ ПРО ПРЕДМЕТНУ ОБЛАСТЬ НА ОСНОВІ КОРПУСА РІЗНОМОВНИХ ТЕКСТІВ ПЕВНОЇ ТЕМАТИЧНОЇ СПРЯМОВАНОСТІ

Балабін Віктор Володимирович
Замаруєва Ірина Вікторівна

*Військовий інститут Київського національного університету
імені Тараса Шевченка
Київ, Україна*

irina@zamaruev.com

В доповіді висвітлюються особливості побудови формальної моделі знань про предметну область на основі різномовних текстів заданої тематичної спрямованості та її використання для усунення синтаксичної і лексико-граматичної омонімії в багатомовних системах машинного перекладу. Під формальною моделлю знань про фрагменти навколишнього світу (предметну область) будемо розуміти предикатну структуру, що включає поняття, відношення та характеристики понять і відношень. Процес побудови моделі знань відбувається у декілька етапів. На першому етапі фахівцем (експертом) укладаються базові поняття із заданої предметної області. Системні відношення між базовими поняттями формуються в результаті аналізу відповідних текстів однієї мови (в нашому випадку – української) і представляються у вигляді тезауруса. Призначення даного тезауруса – преставити так звані «вертикальні» відношення між базовими поняттями, що існують в предметній області, і які не залежать від їх контекстного вживання. Визначаються наступні відношення, які зафіксовані в тезаурусі: *частина-ціле, рід-вид, синонімії, антонімії*. На другому етапі за усіма дискрипторами, що увійшли до тезауруса, формується корпус різномовних текстів відповідної заданої тематики. На цьому етапі корегується тезаурус, побудований на попередньому етапі. Це пов'язане з тим, що класифікація знань про предметну область в різних мовах, як правило, не співпадають. На третьому етапі за корпусом різномовних текстів формуються так звані «горизонтальні» відношення, тобто ті відношення, які проявляються безпосередньо в контексті. Результати третього етапу представляються у вигляді так званої матриці істинності, кожний рядок якої є предикатна формула.

Запропонована модель представлення знань про предметну область дозволяє усувати синтаксичну омонімію при автоматичному перекладі, а також у разі появи у тексті нового поняття, відсутнього у перекладному словнику, автоматично добрати контекстний синонім до нового слова.