**Якименко Костянтин Миколайович. Системотехніка та технологія лексикографічних систем семантичного типу : дис... канд. техн. наук: 05.13.06 / НАН України; Національна бібліотека України ім. В.І.Вернадського. - К., 2006.**

|  |  |
| --- | --- |
| |  | | --- | | Якименко К.М. Системотехніка та технологія лексикографічних систем семантичного типу. – Рукопис.  Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 "Автоматизовані системи управління та прогресивні інформаційні технології". Національна бібліотека України імені В.І. Вернадського. – Київ, 2006.  У дисертації розвинуто теорію лексикографічних систем семантичного типу. На базі теорії розроблено та сформовано лексикографічну базу даних тлумачного Словника української мови (ЛБД СУМ), створено технологічний комплекс укладання нової версії тлумачного словника, розроблено ЛБД та клієнтську програму Українського національного лінгвістичного корпусу для поповнення словника текстовими ілюстраціями. Отриманий комплекс адаптовано до створення та редагування ЛБД російського тлумачного словника. Розроблена технологія дозволяє значно підвищити продуктивність лексикографічної праці.  Автором також розроблено технологічний інструментарій для формування семантичних ЛБД, таких як фразеологічна, синонімічна та етимологічна. Створено засоби для роботи з цими базами даних (їх перегляду та редагування), а також для інтеграції їх із тлумачною ЛБД, зокрема, з використанням розроблених технологій переформовано фразеологічну підсистему СУМа.  На основі ЛБД СУМа створено лексикографічні бази даних для проведення лінгвістичних досліджень окремо за частинами мови. З використанням ЛБД "Дієслово" одержано нову класифікацію українських дієслів та інші дані про структуру їх граматичної та лексичної семантики. | |
| |  | | --- | | У ході дисертаційного дослідження розв’язано ряд актуальних науково-технічних проблем української лінгвістичної технології та отримано низку практично цінних результатів.  Так, у дисертації побудовано лексикографічну модель даних для тлумачних словників, яка при застосуванні принципу рекурсивної редукції Л-системи дозволила вичленити у їх структурі основні структуротвірні елементи. У такий спосіб теорію лексикографічних систем розвинено для систем семантичного типу, що надає можливість для побудови універсального словника, в якому зафіксована максимальна лексична система.  З використанням розвиненої теорії розроблено концептуальну модель Л-системи тлумачного Словника української мови, на основі якої створено структуру ЛБД цього словника. Розроблено технологію, яка дозволила в автоматизованому режимі сформувати цю ЛБД на основі тексту 11-томного СУМа. Зауважимо, що етап виокремлення структурних елементів у розпізнаному та відредагованому тексті словника та рознесення їх до відповідних полів ЛБД виконувався цілком автоматично засобами розробленого програмного забезпечення.  Для укладання нового тлумачного Словника створено технологічний інструментальний комплекс, який дає можливість редагувати будь-які структурні елементи у ЛБД СУМа. Також створено ЛБД Українського національного лінгвістичного корпусу та програму роботи з нею, яка значно прискорила процес поповнення СУМа текстовими ілюстраціями. Таким чином, розроблений у ході дисертаційного дослідження технологічний комплекс дозволив створити новий 20-томний тлумачний Словник української мови всього за 4 роки. Крім того, на основі розвиненої технології ЛБД СУМа розширено, щоб забезпечити підтримку електронних тлумачних словників будь-якої мови, якщо їх структура є подібною до структури СУМа. Зокрема, розроблену технологію адаптовано для укладання російського тлумачного словника й створено відповідний інструментальний комплекс.  У дисертації також розроблено методи та засоби формування інших ЛБД семантичного типу на основі текстів, отриманих з паперових словників. Сформовано фразеологічну ЛБД на основі Словника фразеологізмів української мови, синонімічну ЛБД на основі Словника синонімів української мови та етимологічну ЛБД на основі Етимологічного словника української мови. Для ЛБД СФУМа розроблено технологію автоматизованої інтеграції з ЛБД СУМа, використання якої привело до впорядкування підсистеми стійких сполучень слів СУМа та поповнення СУМа значною кількістю фразеологізмів. Засоби інтеграції синонімічної та тлумачної ЛБД дали можливість побудувати систему тезаурусного типу для української мови. Для синонімічної та етимологічної ЛБД також розроблено відповідні програмні комплекси їх редагування.  ЛБД СУМа було використано для створення ЛБД “Дієслово". Ці ЛБД разом з розробленими програмними засобами зарекомендували себе як потужні інструменти для проведення лінгвістичних досліджень. З використанням ЛБД “Дієслово" одержано нові дані про структуру граматичної та лексичної семантики українського дієслова, зокрема нова фундаментальна класифікація українських дієслів, а також побудовано нові типи словників.  Результати дисертаційної роботи дозволяють стверджувати, що створено концептуальну та технологічну основу систем автоматичного семантичного маркування природномовних текстів, що відкриває перспективи для розробки інтелектуальних семантичних механізмів пошуку, розуміння та діалогу на рівні неструктурованої (текстової) інформації. | |