**Валенда Наталя Анатоліївна. Методи та моделі функціонально-семантичної обробки текстів природної мови у системах штучного інтелекту : Дис... канд. наук: 05.13.23 - 2007.**

|  |  |
| --- | --- |
| |  | | --- | | Валенда Н. А. Методи та моделі функціонально-семантичної обробки текстів природної мови у системах штучного інтелекту. – Рукопис.  Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.23 – системи та засоби штучного інтелекту - Харківський національний університет радіоелектроніки, Харків, 2006.  Дисертація присвячена розробці методів і моделей аналізу текстів природної мови на основі функціонального подання семантики. У дисертації розроблено модель формалізації семантики природно-мовних конструкцій у вигляді суперпозицій семантичних функцій, що дозволяє поліпшити якість систем автоматичної обробки текстів під час аналізу багатозначних мовних одиниць. Розроблено метод перетворення мовних конструкцій у суперпозиції семантичних функцій. Розглянуто метод уніфікації, що дозволяє встановлювати значеннєву еквівалентність мовних конструкцій, які подано семантичними функціями. Набув подальшого розвитку метод аналізу на основі атрибутних транслюючих граматик, завдяки введенню семантичних функцій у вигляді символів дії. Розглянуто застосування розроблених методів і моделей у системах штучного інтелекту, для поліпшення якості аналізу текстів. | |
| |  | | --- | | У дисертаційній роботі наведено результати, які у відповідності з поставленою метою є вирішенням актуальної задачі створення методів обробки текстів природної мови у системах штучного інтелекту на основі використання формалізованої семантичної інформації. Отримані результати мають важливе наукове і практичне значення для створення ефективних систем обробки текстової інформації: аналізаторів природної мови, інтелектуальних пошукових систем. У процесі виконання наукових досліджень отримано такі результати:  1. Проаналізовано сучасні підходи до аналізу текстової інформації, розглянуто найбільш використовувані рішення у цій галузі, обґрунтовано доцільність удосконалення методів обробки тексту за рахунок використання семантичної інформації, що дає можливість підвищити якість аналізу текстів природної мови.  2. Розроблено модель формалізації семантики природно-мовних конструкцій у вигляді суперпозицій семантичних функцій, яка дозволяє будувати методи та алгоритми перетворення мовних конструкцій у формальне подання з урахуванням багатозначності мовних одиниць, що дає можливість отримувати більш якісний аналіз конструкцій природної мови.  3. Розроблено метод аналізу мовних конструкцій на основі заповнення вільних позицій семантичних функцій. Це дозволяє побудувати транслятор, який здійснює автоматичне перетворення мовних конструкцій на суперпозиції семантичних функцій. Проведено порівняння з семантичним аналізом на основі відмінкових фреймів та встановлено, що завдяки зіставленню семантичної інформації усім словам досягається аналіз, який не залежить від чіткого порядку слів, що дає можливість застосовувати його до мов з вільним порядком слів у мовних конструкціях.  4. Розроблено метод виведення на основі логіки предикатів, що відрізняється від існуючих введенням уніфікації для суперпозицій семантичних функцій, яка дозволяє існування розбіжностей в уніфікованих формулах. Метод дозволяє проводити зіставлення формул, які відрізняються за структурою, але передають близькі значення, та може використовуватися для визначення еквівалентності мовних конструкцій, при встановленні значень багатозначних слів, у пошукових системах під час порівняння результатів пошуку з анотацією.  5. Розроблено модель вибору значення лексичних одиниць на основі заповнення вільних позицій та уніфікації суперпозицій семантичних функцій, яка дозволяє покращити якість роботи транслятору природної мови.  6. Набув подальшого розвитку метод аналізу природної мови на основі атрибутних транслюючих граматик, за рахунок використання семантичних функцій у символах дії, що дозволяє визначати значеннєву сумісність елементів допущеної конструкції. Перевірка семантики словосполучень на рівні синтаксичного аналізу веде до спрощення подальшого аналізу. За рахунок звуження області дії синтаксичного аналізу з речення до словосполучення спрощується структура транслятора.  7. Удосконалено метод семантичного аналізу природної мови, який відрізняється від аналізу на основі фрейму предиката за рахунок заповнення вільних позицій фрагментами, що вже пройшли семантичне узгодження, та використанням виводу на семантичних функціях під час вибору значення багатозначних слів. Це дає можливість підвищити якість обробки багатозначних слів та отримувати єдине формальне подання.  8. Розроблено спосіб організації словника на основі навантаженого дерева, який дозволяє зберігати інформацію у вигляді композиції векторів переходів та списків переходів. Це дає можливість зберігати словник частково або повністю в оперативній пам’яті чи на диску. Розроблено алгоритм пошуку помилок у слові, що дає можливість обробки некоректно введених слів.  9. Отримані методи аналізу текстів природної мови впроваджені в інформаційно-пошуковій системі підприємства "Торгівельна компанія "ЕлектроМир". Застосування результатів дисертаційної роботи дозволяє поліпшити якість пошуку інформації природною мовою, за рахунок порівняння суперпозицій семантичних функцій для пошукового запиту та анотацій документів (акт упровадження від 04.06.2005). Теоретичні результати дисертації використано у навчальному процесі на кафедрі Програмного забезпечення ЕОМ (акт упровадження від 17.11.2005). | |