

АНАЛІЗ ДІАЛОГУ ІСQ НА БАЗІ ОБ'ЄКТНО-ОРІЄНТОВАНОГО ПІДХОДУ

Костенко П. М.

Науковий керівник – проф., д.т.н. Дубовой В. М.

Робота присвячена створенню програмного комплексу для аналізу взаємодії між комп'ютерною та людською (природною) мовами. Комплекс складається із двох частин: генерація природної мови та сприймання природної мови.

Система генерації природної мови відповідає за перетворення інформації із комп'ютерної бази даних у людську мову.

Система сприймання природної мови відповідає за перетворення людської мови у більш формальний вигляд, такий як представлення у вигляді дерева граматичного розбору або структур логіки першого порядку, якими легше маніпулювати за допомогою комп'ютера.

У роботі вирішуються такі проблеми як створення морфологічної моделі (структури) слів для коректного аналізу змісту речення, а також для генерації граматично коректного речення; синтаксична невизначеність, тобто наявність декількох дерев розбору для одного речення, вибір найбільш вдалого з яких вимагає наявності семантичної або контекстної інформації; смислова невизначеність слова, тобто вибір такого смислу, який підходить найбільш вдало у заданому контексті.

Більшість даних проблем вирішується за допомогою методів статистичної обробки природної мови, а саме: стохастичних, імовірнісних та статистичних методів, особливо для вирішення труднощів пов'язаних із великими за довжиною реченнями, які є найбільш невизначеними, що призводить до появи тисяч або мільйонів можливих аналізів. Методи вирішення проблем невизначеності включають у себе використання корпусної лінгвістики та ланцюгів Маркова. Технологія статистичної обробки походить від машинного навчання та інтелектуального аналізу даних.

Обробка природної мови є однією з основних галузей розробки штучного інтелекту, а тому може знайти використання у багатьох областях, таких як до цього часу потребували живого спілкування із людьми, наприклад, перший рівень служби підтримки компанії або інтерактивна база знань, яка здатна відповідати на нечітко поставлені питання.