

УДК 007.5

МЕТОДИ ОТРИМАННЯ ЗНАТЬ З ТЕКСТІВ ДЛЯ ПОДАЛЬШОГО ВИКОРИСТАННЯ В ЗАДАЧАХ ЛІНГВІСТИЧНОЇ ОБРОБКИ

Бісікало Олег Володимирович, Кравчук Ірина Анатоліївна

Вінницький національний технічний університет, Україна

Анотація

Оглядовий матеріал містить розгляд методів отримання знань з природномовних текстів. Для кожного методу розглянуто можливості його використання в лінгвістичних системах.

Material includes consideration of methods of obtaining knowledge from natural language texts. Possibility of use each method in linguistic systems is considered.

Вступ

Задача отримання корисної інформації з природномовного тексту являє собою ідентифікацію в ньому певних елементів. Таку задачу можна розв'язувати, використовуючи:

- Синтаксично-семантичний підхід;
- Моделі, побудовані на основі запитів до бази даних;
- Онтології;
- Асоціативно-статистичний підхід до аналізу текстів.

Синтаксично-семантичний підхід

Синтаксично-семантичний підхід до отримання знань із тексту природною мовою опирається на лінгвістичну модель, відповідно до якої основу семантичної структури висловлювання являє так званий пропозиційний компонент плану змісту. Основним елементом у структурі пропозиції речення є предикати, що позначають відносини між об'єктами – обов'язковими учасниками ситуації, які називають аргументами або актантами предиката [1]. Значимість елементів речення, як прийнято вважати, характеризує комунікативний ранг, що визначається їхньою віднесеністю до теми або реми й співвіднесенням з членами речення. Ця інформація, поряд з статистикою вживання й іншими факторами, дозволяє виділити ключові елементи тексту для порівняння документів при пошукуві й класифікації [1].

Моделі, побудовані на основі запитів до бази даних

Для полегшення збору та зберігання інформації для опису елементів словника, що використовується для лінгвістичного аналізу текстової інформації, з усіма характеристиками використовують бази даних [2]. На даний момент в комп'ютерній лінгвістиці не існує спеціалізованих лексикографічних баз даних, проте такі бази даних, як ACCESS, FOX-Base, PARADOX, D-Base можуть бути використані для ведення словникових картотек, а також для їхнього редагування.

Онтології

В рамках комп'ютерної лінгвістики онтологія розуміється як система абстрактних понять, що існують тільки у свідомості людини, яка може бути виражена природною мовою. Лінгвістичний підхід до створення і дослідження онтологій заснований на вивченні природної мови (зокрема, семантики) і побудові онтологій на великих текстових масивах, так званих корпусах [3]. Використання онтологій дозволяє представити природномовний текст у вигляді, придатному для автоматичної обробки. Додатково онтології можуть використовуватися в якості посередника між користувачем та інформаційною системою, що дозволяє формалізувати використовувані терміни між усіма користувачами проекту [3].

Асоціативно-статистичний підхід до аналізу текстів

Багато років розробники штучного інтелекту намагалися навчити комп'ютер логічному мисленню, заснованому на маніпулюванні формалізованими знаннями й правилами їхнього перетворення. Такий тип мислення характерний для обробки інформації лівою півкулею мозку. Проблема полягає в неможливості самонавчання без участі людини [4]. Асоціативно-статистичний підхід використовується в програмних продуктах RCO і UCO for Oracle та в RCO і UCO Semantic Network (<http://www.rco.ru>, <http://uco.ua>), що використовуються для пошуку для Oracle та для Microsoft.

Список використаних джерел:

1. Ермаков А. Е. Неполный синтаксический анализ текста в информационно-поисковых системах / А. Е. Ермаков // Компьютерная лингвистика и интеллектуальные технологии: труды Международного семинара Диалог'2002. – Москва, Наука. – 2002. – Т. 2. – С. 180 – 185.
2. Юрков Н. Интеллектуальные компьютерные обучающие системы / Н. Юрков. – Пенза: Изд-во ПГУ, 2010. – 304 с.
3. Никоненко А. Обзор баз знаний онтологического типа / А. Никоненко // Искусственный интеллект. – 2009. – № 4. – С. 208 – 219.
4. Букин А. Г. Проект: поиск знаний в интернет / А. Г. Букин, Ю. А. Попов // Научная сессия МИФИ-2005: сб. научных трудов МИФИ. – Москва, 2005. Т. 11. – С. 72 – 75.