

РОЗРОБЛЕННЯ МОВНОЇ МОДЕЛІ ДЛЯ ЗАДАЧ ОБРОБКИ ПРИРОДНОМОВНИХ ТЕКСТІВ

Автоматизований аналіз текстів наразі широко використовується для пошуку інформації, побудови анотацій та перекладів. Однак для побудови баз знань та знання-орієнтованих систем необхідно виявлення не тільки ключових елементів тексту, а й смислових зв'язків між ними. Це завдання вирішується за допомогою створення мовної моделі із процедурами семантичного аналізу.

Метою даної роботи є опис структури запропонованої системи семантичного аналізу тексту, розробка якої дозволить поліпшити створення баз знань та імплементувати на основі неї підсистему «штучної совісті» як елементу розроблюваної моделі штучного інтелекту [1].

Вхідними даними до системи семантичного аналізу є текст у символічному кодуванні. Першим етапом обробки є граматичний аналіз фрагменту тексту, який відносить кожне слово до певної частини мови за ознаками словотворення та базою граматик відповідної мови. У випадках, коли однозначно визначити частину мови неможливо, слово тимчасово позначається відношенням до декількох можливих частин мови. В залежності від частини мови, до властивостей слова додаються інші граматичні ознаки (рід, відмінок і т.д.).

На етапі синтаксичного аналізу з'ясовується роль слів у реченні. Ця операція виконується на основі пунктуації, порядку слів та їх класифікації за частинами мови. На даному етапі також завершується розв'язання неоднозначностей минулого етапу [2].

Етап трансляції у терміни бази знань передбачає виявлення у висловлюванні на природній мові семантичних одиниць, після чого результати синтаксичного аналізу формуються у вигляді гіпотез висловлювання.

Оновлення бази знань представляє собою порівняння нових семантичних одиниць та гіпотез висловлювання із наявними в базі. У разі відсутності до бази будуть додані нові елементи. Окремим випадком є режим оновлення існуючих правил, який може застосовуватися виключено при обробці висловлювань із джерел з високим ступенем довіри.

На ефективність подальшої роботи із базою знань окрім інформативності вхідних даних суттєво буде впливати структура її організації. У роботі пропонується використання реляційної бази даних для зберігання наступних сутностей: словника термінів на природній мові, трансляції у внутрішні сутності та опис залежностей між ними у вигляді типізованих атрибутів концептів.

Висновки: у роботі запропоновано структуру системи семантичного аналізу текстів, описано основні її складові та обґрунтовано взаємозв'язок між ними; дана робота може отримати розвиток у вигляді програмної реалізації компонентів системи та проведення досліджень щодо ефективності її застосування для створення баз знань.

Список використаних джерел

1. Shevchenko, A.I., Klymenko, M.S. Developing a Model of Artificial Conscience / CSIT 2020 - Proceedings, 2020, 1, pp. 51–54. DOI: 10.1109/CSIT49958.2020.9321962
2. Клименко М.С. Розроблення знання орієнтованої підсистеми підтримки прийняття рішень інтелектуальної інформаційної системи / Штучний інтелект. – 2020. - 25(1). – с. 51-56 DOI: <https://doi.org/10.15407/jai2020.01.051>