

УДК 004.89:004.822

Використання автоматично побудованих синтаксичних шаблонів для виявлення та обробки семантичних зв'язків у тексті

Левашко О.Л., старший викладач кафедри програмування та захисту інформації,
Levoshkoel@ukr.net

Кіровоградський національний технічний університет, м. Кіровоград

Дана робота присвячена дослідженню й розробці методів автоматичного виявлення та обробки семантичних відношень в тексті. Під обробкою семантичного зв'язку мається на увазі визначення в тексті пар імен сутностей, що перебувають в даному відношенні (зв'язку). Приклади семантичних відношень - «працювати_на» (ім'я людини і його місце роботи), «створена_в» (назва організації та дата її створення). Завдання обробки семантичних відношень належить до галузі інтелектуального аналізу текстів, вкрай актуальної та затребуваної в даний час. Алгоритми обробки семантичних відношень можуть застосовуватися для покращення методів пошуку ключових слів у текстах, для автоматичного створення баз даних, а також для розширення інтерфейсу пошукових систем.

Для вирішення завдання обробки семантичних відношень з текстів був використаний метод, заснований на автоматичному створенні синтаксичних шаблонів, що є релевантними для шуканих відношень. Метод побудовано на алгоритмі активності, котрий розповсюджується на семантичній мережі. Запропонований метод складається з двох етапів:

1. На першому етапі автоматично будується набір синтаксичних шаблонів. Кожен шаблон моделюється піддеревом дерева синтаксичного розбору речень. Для побудови шаблонів використовується текстовий файл і набір пар імен сутностей, які перебувають у початковому відношенні. Також на даному етапі обчислюються оцінки, що характеризують якість шаблонів.

2. На другому етапі побудовані раніше синтаксичні шаблони застосовуються до речень цільового тексту. При цьому формуються пари імен сутностей, що претендують на наявність шуканого семантичного відношення між ними. Рішення про включення кожної пари імен до результуючої множини приймається на підставі оцінок застосованого шаблону, за допомогою якого була вибрана дана пара імен, семантичної близькості слів речення, з якого була вибрана пара імен, до набору ключових для даного відношення слів, точності відповідності речення до шаблону.

Висновок. Запропонований метод обробки семантичних відношень в тексті показав досить високу точність (70%) на тестових даних для відношення «працювати_на».

Список літератури

1. [Електронний ресурс]. – Режим доступу: <http://www.dictum.ru>. Матеріали з автоматичної обробки текстів.
2. [Електронний ресурс]. – Режим доступу: <http://www.aot.ru>. Матеріали з автоматичної обробки текстів.
3. Lei Shi. An Algorithm for Open Text Semantic Parsing / Lei Shi, Rada Mihalcea // Proceedings of the ROMAND 2004 Workshop on «Robust Methods in Analysis of Natural Language Data». – 2004.