

РОЗРОБКА МОДЕЛІ ЛЮДИНО-МАШИННОГО ДІАЛОГУ НА ОСНОВІ ОБРАЗНОЇ ІНДЕКСАЦІЇ ПРИРОДНО-МОВНОГО КОНТЕНТУ

Анна Лісовенко – аспірант, Вінницький національний технічний університет (ВНТУ), Україна

Науковий керівник – **Олег Бісікало** – д.т.н., проф., декан ФКСА, ВНТУ, Україна

В даний час стає все більш очевидно, що найбільш перспективний шлях підвищення якості інформаційних процесів в людській діяльності, соціумі, ноосфері – спільне використання можливостей людини і програмно-технічних засобів, зокрема, інформаційних систем. Внаслідок цього стає актуальною задача вивчення основних аспектів людино-машинної взаємодії, а саме параметрів інтелектуальної діалогової системи. Метою розробки людино-машинної діалогової системи, що надаватиме відповідь у вигляді цитати з тексту на основі визначення сили зв'язків між словоформами за допомогою лінгвістичної технології DKPro.

Візьмемо за основу розгляду та порівняння систему DKPro. Метою розробки системи DKPro – перетворення неструктурованої інформації в структуровані за типами знання за допомогою технології семантичної обробки інформації. Система базується на Unstructured Information Management Architecture (UIMA) фірми IBM [1]. DKPro має гнучкий, масштабований і простий у використанні набір інструментів, що забезпечують технологію обробки тексту. Головними компонентами системи DKPro (рисунок 1) є наступні:

1. DKPro-UGD – даний інструмент призначений для гнучкої очистки текстових даних вилучених з різних інтернет-ресурсів, таких як блоги, чати, соціальні платформи і т.п.

2. MayoClinicNLP-Core[2] – інструментарій, який досліджує семантичну подібність на рівні речення.

3. DKPro Similarity [3] – це інструмент, що включає широку різноманітність мір текстової подібності в межах від простих, основаних на n-gram і на розповсюджених підпоследовностях, до високо розмірного вектору порівняння та структури, стилістики і фонетичної міри.

4. DKPro WSD – інструмент для вирішення неоднозначності смислу слова. Основною його перевагою є те що цей інструмент є модульним і має можливість розширюватись.

5. DKPro Keyphrases [4] – структура вилучення ключових фраз основана на архітектурі управління неструктурованою інформацією (UIMA).

6. DKPro TC – структура для проведення контрольованих навчальних експериментів на текстових даних.

7. DKPro Agreement – Java-бібліотека для вимірювання між експертної оцінки. Надійність – необхідна попередня умова для отримання високоякісних наборів даних і, таким чином, для підготовки правильного висновку з

дослідження аотації. Тому, даний інструмент є досить важливим і об'єднує загальний інтерфейс та модель даних.

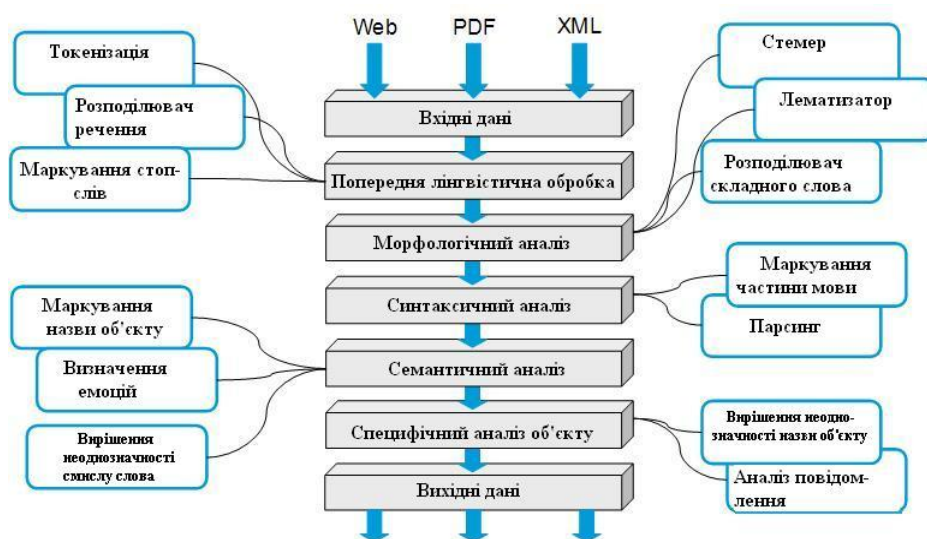


Рис. 1. – Ядро системи Darmstadt Knowledge Processing Repository (DKPro)

Існує шість типів діалогу [5] раніше визнаних аргументаційною літературою, такі як: діалог-запит (допит); діалог-переговори; діалог пошуку інформації; діалог обговорення; діалог полеміка; діалог відкриття.

Найбільш підходящим типом діалогу для розроблюваної нами

людино-машинної діалогової системи є діалог пошуку інформації. Даний тип діалогу – це діалог, який ґрунтується на запитанні та відповіді, або на пошуковій системі типу Google. Проте основною проблемою пошуку в інтернеті є «сирість» інформації, визначення її достовірності та якості.

Враховуючи усе наведене вище робимо висновок, що технологія DKPro має великий та зручний набір інструментів для обробки текстової інформації, проте вона не використовується для підтримки діалогу. Саме тому постає задача створення такої системи на основі DKPro, яка, використовуючи лінгвістичні інструменти виявлення зв'язків між словоформами, забезпечує участь користувача у діалозі з системою по певному введеному фаховому тексту та повертає у відповідь на запитання користувача цитату з попередньо опрацьованого системою цього фахового тексту.

Література

1. David Ferrucci. UIMA: An Architectural Approach to Unstructured Information Processing in the Corporate Research Environment /D.Ferrucci, A.Lally // Natural Language Engineering. – №10. – IX. 2004. – pp. 327-348.
2. S.Wu. MayoClinicNLP-CORE: Semantic representations for textual similarity / S.Wu, D. Zhu, B.Carterette, H.Liu // Second Joint Conference on Lexical and Computational Semantics. – №1. – VI.2013. – pp.148-154.
3. D.Bar. DKPro Similarity: An Open Source Framework for Text Similarity / D.Bar, T.Zesch, I.Gurevych // Annual Meeting of the Association for Computational Linguistics. – №51. – VIII.2013. – pp. 121-126.
4. N.Erbs. DKPro Keyphrases: Flexible and Reusable Keyphrase Extraction Experiments / N.Erbs, P.B.Santos, T.Zesch, I.Gyrevych // Annual Meeting of the Association for Computational Linguistics. – №52. – VII.2014. – pp. 31-36.

5. Douglas Walton. Types of dialogue and burdens of proof: Proceedings of the 2010 conference on Computational Models of Argument: Proceedings of SOMMA 2010 [Електронний ресурс] / Douglas Walton // – 2010. – pp. 13-24. – Режим доступу до статті: <http://www.dougwalton.ca/papers%20in%20pdf/10commaBoP.pdf>.