

# Типи семантичних зв'язків у медичних текстах

Антоніна Заяць<sup>1</sup>, Андрій Романюк<sup>2</sup>

Кафедра систем автоматизованого проектування, Національний університет  
“Львівська політехніка”, УКРАЇНА, м. Львів, вул. С. Бандери, 12,  
E-mail: 1. tonya-zayats@yandex.com, 2. anrom@polynet.lviv.ua

*Abstract – This article is dedicated to semantic relations types in medical textdata in context of the relation extraction process and further development of the relation extraction system for the Ukrainian language.*

Ключові слова – екстракція семантичних зв'язків, опрацювання природної мови, семантичні сутності, схема розмітки.

## I. Вступ

Сучасний науковий світ приділяє значну увагу сфері автоматичного опрацювання текстів. Серед основних завдань цього напрямку виділяють автоматичний морфологічний, синтаксичний і семантичний аналізи, серед складніших – створення систем машинного перекладу, інформаційного пошуку, реферування, редагування текстів тощо [1].

Часто для реалізації цих завдань необхідне розуміння семантичних зв'язків між сутностями. На сьогодні відомо чимало методів реалізації цього процесу. Найчастіше використовують вручну створені патерни, кероване, напівкероване та некероване автоматичне навчання [3]. Кожен з цих підходів має свої переваги та недоліки, а вибір залежить від поставлених дослідницьких цілей та можливостей їх реалізації. Проте, початковим етапом, який не залежить від подальшого алгоритму створення системи розпізнавання, є побудова анотованого корпусу текстових даних.

## II. Основна частина

Ця стаття присвячена вибору семантичних зв'язків при створенні анотованого корпусу текстів для подальшої розробки системи екстракції зв'язків у медичній галузі. Вибір предметної області продиктований нагальною потребою у таких засобах для спрощення роботи медичних систем при структуруванні великих об'ємів даних.

Загалом процес створення анотованого корпусу медичних текстів можна поділити на такі етапи:

1. Збір та систематизація множини текстових документів.
2. Вибір програмних засобів для реалізації процесу анотування.
3. Розробка схеми анотацій.
4. Анотування зібраних матеріалів програмними засобами.

Складність вибору типів іменованих сутностей та зв'язків для подальшого маркування полягає в особливості обраної предметної області та відсутність аналогів для української мови. Значна частина закордонних досліджень у цій сфері спирається на Unified Medical Language System (UMLS), що містить свою ієрархію сутностей та зв'язків. [2, 3]. На основі низки наукових статей та експериментів, цю систему було взято за основу реалізації схеми розмітки майбутньої системи видобування зв'язків для української мови.

Оригінальна схема складається з 54 типів та підтипів семантичних зв'язків. Серед п'яти основних:

- physically related to;
- spatially related to;
- temporally related to;
- functionally related to;
- conceptually related to.

## Висновок

Вибір типів семантичних зв'язків є початковим етапом процесу побудови анотованого корпусу текстових даних для подальшої розробки системи розпізнавання зв'язків у медичних текстах. Для реалізації цього завдання було використано схему уніфікованої системи медичної мови (Unified Medical Language System).

## Література

1. Katrenko S. A Closer Look at Learning Relations from Text. PhD thesis, University of Amsterdam. – 2009, – 222 p.
2. UMLS – current semantic types. Режим доступу: [http://www.nlm.nih.gov/research/umls/META3\\_current\\_semantic\\_types.html](http://www.nlm.nih.gov/research/umls/META3_current_semantic_types.html).
3. UMLS – current relations. Режим доступу: [http://www.nlm.nih.gov/research/umls/META3\\_current\\_relations.html](http://www.nlm.nih.gov/research/umls/META3_current_relations.html).