

## **ВИЗНАЧЕННЯ СМISЛУ ТЕКСТУ У ДИСКУРСІ АНАЛІЗУ ТЕКСТОВОЇ ІНФОРМАЦІЇ**

**Д'якова О.С., Бабкова Н.В., Угольнікова Н.С.**

*Національний технічний університет  
«Харківський політехнічний інститут»,  
м. Харків*

Аналіз текстової інформації, або Text Mining, поєднує в собі підходи різноманітних галузей знань, і є частиною напрямку "комп'ютерна лінгвістика". Сьогодні цей напрямок має всі підстави для розвитку. Постійне розширення доступу до ресурсів глобальної мережі, надає людям великі простори для обміну різними видами інформації. Неважко уявити, що переважна більшість цієї інформації представлена у вигляді тексту, й більша частина цього тексту, призначеного для сприйняття людиною, не має достатнього рівня структуризації для використання традиційних видів машинного аналізу. Сумарні обсяги інформації на природній мові значно перевищують аналітичні можливості людського сприйняття. Крім того, перегляд, вже й відібраної по заданій тематиці інформації, займає деякий, нерідко дуже тривалий, період часу, а відповідь на задане питання може бути прихована "між рядків". Однак найчастіше (особливо це стосується економічної сфери), саме своєчасність, актуальність та повнота інформації й складають її основну цінність.

Самі масштаби інформаційного обміну просто вимагають автоматизації детального аналізу текстової інформації. Але не можна ототожнювати процес мислення людини та роботи машини. Справа в тому, що початкові властивості думки, не мають за собою будь-якого знакового, символного уявлення. Тоді як, сучасні машини оперують тільки знаками, тому вони можуть лише імітувати деякі аспекти людської логіки, нерідко, цілком успішно. Наприклад, за допомогою технології нейронних мереж, машини, навчившись на наданій в приклад множині відповідностей, можуть скласти деякі критерії, що дозволяють надалі визначити подібні відповідності на розрізнених множинах. А семантична мережа здатна надати зв'язки між об'єктами та поняттями, що схоже зі створенням асоціативного рядку людським мисленням.

Цікавим і перспективним напрямком в аналізі текстів є процес виявлення дискурсу, або сенсу тексту. Незважаючи на узагальнену схожість з анотуванням, що стала вже «стандартною» процедурою для засобів Text Mining, виявлення дискурсу передбачає більш «глибоке осмислення». Склавши семантичну структуру тексту можна надати до машинного аналізу інформацію, що іноді називають «ієрархією смислів», яка могла б бути загублена при більш поверхневому аналізі. Такі структури відкривають широкий спектр можливостей по «видобутку» інформації. У дослідженнях, які спрямовані на вирішення саме цієї задачі, традиційні методи інтелектуального аналізу можуть послужити непоганою підмогою. Зокрема, орієнтуючись на структуру фрази, сучасні технології здатні визначати зв'язок формату «причина->слідство» між структурними елементами. За аналогією з цією процедурою, можливо виявити зв'язок подібного формату у тексті, а можливо і за його межами. Результати такого аналізу дозволять виявити причини, можливо неявні навіть для людини, що передують або впливають на описану подію.