

## **Постановка завдання**

В рамках даної роботи буде досліджено алгоритм ітеративної побудови термінології у вигляді RDF-графу (триплетів суб'єкт-предикат-об'єкт згідно стандарту) за колекцією наукових текстів спільної тематики. Схема побудови зв'язків між термінами буде враховувати статистичні властивості термінів, а також лексикографічні евристики пов'язані зі структурою україномовних документів. Принципова схема методу складається з двох кроків: початкове виділення з тексту важливих термінів за статистичним методом, і наступне видобування з тексту пов'язаних термінів за рахунок застосування лексикографічних шаблонів. В рамках дослідження передбачається розробити структурну схему алгоритму, з подальшою формалізацією алгоритму, а також програмне забезпечення для моделювання роботи алгоритму. Кінцевою метою розробки є програма, що дозволить отримати RDF граф термінів колекції документів, наданої у форматі pdf. Результати даної роботи мають вирішити задачу складання термінології по колекції документів і можуть слугувати основою для створення інтелектуальних систем пошуку документів у семантичному вебі.