

Мій абстрактний тип даних MyMultiset призначений для зберігання запитань та відповідей. Це динамічний масив, елементами якого є об'єкти власного класу Question. Цей клас має три атрибути:

- 1) запитання англійською мовою(над ним будуть проводитись усі операції);
- 2) відповідь на запитання українською мовою(представлення значення ключа);
- 3) вектор запитання, який характеризує його(представлення ключа словника).

1)Запитання англійською мовою отримується шляхом перекладу даного запитання українською мовою(передається як аргумент) за допомогою Google Translate API.

2) відповідь на запитання передається як аргумент після попереднього її пошуку у каналі Slack

3) вектор початково є None, формується в результаті створення TF-IDF або LSI(LSA) моделі речення(запитання)

При ініціалізації створюється пустий масив. Елементи до нього додаються за допомогою методу add_key, який аргументами приймає запитання та відповідь у формі стрічки. Інші операції: вилучення елемента, отримання списку елементів(у формі генератора), доступ до атрибутів класу Question, отримання списку запитань у формі стрічок.

За допомогою даного ADT створюється корпус запитань: створюється список запитань(тип:рядок), з кожного запитання вилучаються так звані stop-слова(прийменники, артиклі), кожному унікальному слову в цьому списку присвоюється цифра від 0 до n (n - к-сть унікальних слів), для кожного речення створюється список списків, де кожен внутрішній список містить порядковий номер унікального слова і к-сть його повторів у запитанні, тобто створюється вектор запитання. Створюється [TF-IDF](#) або [LSI](#) модель кожного запитання: визначається вага кожного слова в запитанні. На основі сформованої моделі встановлюються коефіцієнти подібності кожного запитання із заданим. В результаті знаходиться найбільш схоже запитання до заданого.