

# ТРАНСЛЯЦІЯ ТЕКСТІВ ПРИРОДНОЮ МОВОЮ У ЗАПИТИ ДО БАЗ ДАНИХ

Вінницький національний технічний університет

## **Анотація**

*Обґрунтовано доцільність розробки систем трансляції текстів природною мовою в запити до баз даних. Досліджено методи та способи виконання трансляції фраз природною мовою у запити до баз даних.*

**Ключові слова:** *реляційна база даних, інформаційна система, лінгвістичний аналізатор, морфологічний аналіз, синтаксичний аналіз.*

## **Abstract**

*The expediency of developing systems for translating natural texts into database queries is substantiated. The methods of implementation of translation of phrases are investigational by a human language in queries to the bases given.*

**Keywords:** *relational database, information system, linguistic analyzer, morphological analysis, syntactic analysis.*

У процесі поширення і використання баз даних бізнес-менеджери та співробітники потребували доступу до інформації для ведення своїх компаній. Департаменти обробки даних компаній намагалися задовольнити цю потребу, починаючи з початку 50-х років. В даний час ведення обліку більшості організацій здійснюється комп'ютером, а відмова від усіх видів даних, що часто описують транзакції з докладною інформацією, перебуває на центральних комп'ютерах цих організацій. Теоретично всі ці дані доступні для перегляду співробітниками таких компаній. Проте в практиці користувачі такої інформації зіткнулися з серйозними перешкодами для отримання інформації, яка їм потрібна.

Часткою відповіддю на запит користувачів на отримання даних з бази даних є те, що дані не зберігаються таким чином, що дозволяло б використовувати їх для задоволення потреби користувача. Крім того, для складності існуючих систем бази даних потрібен кваліфікований спеціаліст, щоб з'ясувати, як дані, запитувані користувачем, можуть бути отримані з бази даних. Цей спеціаліст повинен інтерпретувати запит користувача або "запит", визначити саме те, що шукає користувач, і з'ясувати, як отримати цю інформацію з бази даних. Тоді, коли дані будуть завантажені, їх потрібно відформатувати і сформувати у звіт, який користувач може використовувати та зрозуміти.

В останні роки тип бази даних, відомий під назвою «реляційна база даних», широко використовувався у бізнес-спільноті. Модель "сутність-зв'язок" часто використовується при картографуванні системи реального світу в систему керування реляційною базою даних. Модель відносин і сутності характеризує всі елементи системи як об'єкт (наприклад, особу, місце або речі) або відносини між сутностями. Обидві конструкції представлені однаковою структурою, яка називається таблицею.

Таблиця - це сукупність даних, організованих у рядки та стовпці, і являє собою одиницю реляційної бази даних. Наприклад, в системі введення замовлення об'єкти будуть виключати частини та замовлення. Така інформація може бути представлена у двох різних таблицях. Зв'язок, яку частини вимагає замовлення, може бути представлена третьою таблицею.

Таким чином, при застосуванні моделі "сутність-сутність" об'єкти системи ідентифікуються і таблиці структуровані для представлення сутностей. Потім визначаються відносини між сутностями, а поточні таблиці розширені (або створюються нові таблиці) для представлення цих відносин. Нарешті, атрибути кожного суб'єкта ідентифікуються, а таблиці розширюються, щоб включати такі

атрибути. Спеціалісти в галузі техніки добре знайомі із застосуванням моделі "Сутність-зв'язок" до реляційних систем керування базами даних.

Використання складних речень природної мови в запиті користувача не передбачається природно-мовним інтерфейсом з базами даних. Попри це для розпізнавання фраз необхідно будувати лінгвістичний аналізатор як перетворювач, що складається з трьох рівнів пофразного представлення тексту – морфологічного, синтаксичного і семантичного. Кожен ківень передбачає відповідну складову моделі зі своїм масивом правил і побудовою формальної схеми запиту – морфологічної, синтаксичної та семантичної структурами. Також важливим є те, що для вирішення поставленої задачі не потрібно проводити повний семантичний аналіз, це потрібно для класифікації лексем, що виконується для проблемного аналізу[3].

Для правильної роботи механізму обробки запитів слід дотримуватися наступних етапів:

- введення користувачем інтелектуального запиту природною мовою з можливістю задання параметрів;
- виконання лексичного аналізу;
- виконання синтаксичного аналізу;
- виконання семантичного аналізу;
- виконання трансляції тексту запиту мовою запитів SQL;
- виконання інформаційною системою трансльованого програмного запиту;
- виведення користувачеві результатів роботи програми або повідомлення про помилку з конкретними причинами виникнення.

В теперішній час однією з провідних компаній з аналізу даних є компанія Tableau Software Inc.. Розробники цієї фірми додали у своє програмне забезпечення нову функцію під назвою Ask Data, яка дає змогу користувачам виконувати запити своїх даних природним голосом. Це дає змогу ставити прості і складні питання і отримувати на них негайну відповідь у формі інтерактивної візуалізації. Суть цієї функції полягає в точному розумінні мети запиту і здатності пов'язувати запити з даними, навіть якщо концепції пошуку явно не відтворені у заголовку бази даних.

**Висновки.** Інтерфейс природної мови заощадить велику кількість часу та грошей для компаній, що використовують реляційні бази даних, і дозволить користувачам, які мають обмежений досвід або практичний досвід, використовувати складну систему бази даних, просто вводячи (наприклад, за допомогою клавіатури) природну мову (наприклад, англійська) питання.

#### СПИСОК ЛІТЕРАТУРНИХ ДЖЕРЕЛ

1. Обробка запитів природною мовою на основі семантичних мереж і шаблонів [Текст] / Н.Ю. Барішнікова // Вісник АДТУ. – 2016. – № 4. – с. 36–45.
2. Обробка текстів і повідомлень природною мовою в інформаційних системах [Текст] / Я.В. Поточняк // Інформаційні технології. – 2016. – № 5. – с. 5–6.
3. Методи і алгоритми трансляції природно-мовних запитів до бази даних в SQL-запити: Монографія. – Улан-Уде: Вид-во ВСГТУ, 2004. – 148 с.

*Майданевич Ірина Олександрівна – студентка групи КІВ-166, факультет комп'ютерних систем і автоматизації, Вінницький національний технічний університет, Вінниця, e-mail: [maidanevych.i.99@gmail.com](mailto:maidanevych.i.99@gmail.com)*

*Науковий керівник: Гришук Тетяна Вікторівна – доцент кафедри КСУ, Вінницький національний технічний університет, Вінниця.*

*Maidanevych Iryna O. - Faculty of Computer Systems and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: [maidanevych.i.99@gmail.com](mailto:maidanevych.i.99@gmail.com)*

*Supervisor: Gryshuk T.V. - Associate Professor, Department of the KSU, Vinnytsia National Technical University, Vinnytsia.*