

СИНТАКСИЧНИЙ АНАЛІЗАТОР ДЛЯ СКАНЕРА QR-КОДІВ

Соха В.М.¹⁾, Порплиця Н.П.²⁾

Тернопільський національний економічний університет

^{1)магістрант, ^{2)к.т.н., старший викладач}}

I. Постановка проблеми та мета роботи

Проблема ефективного планування та обліку власного бюджету є актуальною для будь-кого. Існує багато додатків для обліку доходів та витрат користувача, за допомогою яких користувач може додавати свої щоденні витрати та надходження, переглядати статистику, формувати звіти. Але цей процес займає певний період часу і його можна ще більше автоматизувати. Останнім часом у європейських країнах доволі розповсюдженою стає практика, коли за допомогою самого лише QR-коду товарного чеку записують інформацію про покупки в магазинах у електронний гаманець. Тому актуальною є розробка програмного модуля, який би коректно зчитував QR-код товарних чеків та автоматично формував витрати по відповідних категоріях мобільного гаманця для обліку доходів та витрат.

II. Синтаксичний аналізатор для сканера QR-кодів

Загальний вигляд товарних чеків із QR-кодами подано на рисунку 1. Після сканування QR-коду у додатку буде показано сторінку з інформацією про зроблену покупку, кожен товар буде відноситися до певної категорії, після чого можна відредагувати або зберегти його.

November 11, 2011	12:33 PM
Transaction # 678456	Register: 3
Men's Large Thermal Shirt	\$10.99
Men's Size 34 Hiking Shorts, Brown	\$29.95
Women's Ankle Running Socks	\$6.99
Water Bottle	\$10.95
Sub Total	\$58.88
6% Sales Tax	\$2.94
Total	\$61.82
CASH	\$70.00
Change	\$8.18


Scan this QR Code for special savings good NOW through December 24th	
	



Рисунок 1 – Приклади товарних чеків із QR-кодами

У магазинах інформація про покупки може бути записана у різному вигляді, наприклад молоко: «мол. Молокія 1л» або «молоко Молокія, 1л», капуста: «кап. Пек.» або «капуста Пекінська».

Тому даний програмний модуль повинен аналізувати такі слова і розпізнавати який продукт до якої категорії належить.

З цією метою використовується синтаксичний розбір (парсинг), у лінгвістиці та інформатиці – процес порівняння лінійної послідовності лексем (слів) природної або формальної мови з її формальною граматиною. Результатом є дерево розбору (синтаксичне дерево). Інакше кажучи, парсинг – це процес аналізу або розбору тексту на компоненти з використанням спеціального програмного забезпечення. Парсер – це програма, яка аналізує текстові документи, зберігає аналіз даних у своїй базі даних, а потім видає їх при пошуку актуальних і поточних даних. Аналізатор може виявити велику кількість корисної інформації та обробляти її, залежно від завдань. Синтаксичний

аналіз є важливою складовою опрацювання тексту і спрямований на розпізнавання, виділення та групування даних.

Будь-який аналізатор складається з трьох частин, які відповідають за три окремі процеси розбору:

- отримання тексту у його первісному вигляді. Отримання тексту часто означає завантаження текстового документа, з якого необхідно отримати певні дані.

- вилучення та перетворення даних. Необхідні дані, які були отримані на першому етапі на цій фазі «втягаються». Для вилучення найчастіше використовуються регулярні вирази. Крім того, на цьому етапі здійснюється перетворення витягнутих даних у певний формат, якщо це необхідно;

- генерування результатів. Це заключний етап аналізу. Він досягається шляхом виводу або запису даних, отриманих на другому етапі, у бажаному форматі. Часто записують безпосередньо в базі даних. Завдання синтаксичного аналізатора, по суті, полягає у визначенні, чи є і в який спосіб вхідні дані можуть бути отримані з початкового символу граматики.[1]

Завдання синтаксичного аналізатора, по суті, полягає у визначенні, чи є і в який спосіб вхідні дані можуть бути отримані з початкового символу граматики.

Методи синтаксичного розбору:

- 1) нисхідний;
- 2) висхідний;
- 3) комбінований.

Нисхідний синтаксичний аналіз можна розглядати як спробу знайти найлівіший диференційований елемент вхідного потоку за рахунок пошуку дерев розбору, використовуючи розширення даних формальних правил граматики проходом згори донизу.[2] Лексеми «поглинаються» зліва направо. Виключний вибір використовується для узгодження неоднозначностей, використовуючи розширення всіх альтернативних правобічних граматичних правил [4]. Такий аналіз можна легко здійснити вручну, наприклад методом рекурсивного спуску. Висхідний синтаксичний аналіз може починати розбір з входу і пробувати переписати його на початковий символ. Інтуїтивно зрозуміло, що аналізатор намагається знайти основні елементи, потім елементи, які містять основні, тощо. LR-аналізатор є прикладом висхідного аналізатора. Послідовність розбору визначає, в який спосіб буде побудовано дерево розбору на кожному кроці підстановки. Ці підстановки можна здійснювати зліва направо, справа наліво або довільно[3]. Слід зазначити, що використання упорядкованого розбору прискорює його виконання за рахунок зменшення кількості правил, що перебираються.

Алгоритм роботи синтаксичного розбору організований за принципом, зображеним у вигляді діаграми на рис.2.



Рисунок 2 – Діаграма потоків даних процесу парсингу

Висновки

Аналіз сучасного ІТ ринку в Україні показав, що зараз відсутні спеціалізовані модулі для синтаксичного аналізу товарних чеків із магазинів. Зважаючи на той факт, що використання QR-кодів у побуті, бізнесі, сфері послуг та туризмі стає все більш популярним та необхідним, актуальною є задача розробки програмного модуля для синтаксичного аналізу інформації QR-коду. У праці запропоновано підхід для синтаксичного аналізу даних з чеків за допомогою алгоритмів висхідного та нисхідного аналізів.

Список використаних джерел

1. Ахо А. Теория синтаксического анализа, перевода и компиляции. Т. 1. Синтаксический анализ / А. Ахо, Дж. Ульман. – М. : Мир, 1978. – 650 с.
2. Хантер Р. Основные концепции компиляторов / Робин Хантер ; пер. с англ. – М. : Вильямс, 2002. – 256 с.
3. Севбо И. П. Графическое представление синтаксических структур и стилистическая диагностика / И. П. Севбо. – Київ : Наук. думка, 1981. – 192 с.
4. Дарчук Н. П. Автоматичний синтаксичний аналіз текстів корпусу української мови / Н. П. Дарчук // Укр. мовознавство. – 2013. – Вип. 43. – С. 11–19.