

Розробка універсального парсера для обробки фінансових звітів на основі Ві-платформи QlikView

Вінницький національний технічний університет

Анотація

У даній роботі проведено загальний аналіз структур існуючих фінансових звітів від різних ретейлерів у сфері продажу одягу, взуття і аксесуарів. Проведено порівняння загальних і відмінних рис між звітами різних компаній, наведено аналіз спільних рис, на базі яких можна розробити універсальний парсер для подальшого застосування отриманих даних із загальною структурою. Створено алгоритм роботи спроектованого парсера, наведено його переваги та недоліки.

Ключові слова: парсер, фінансовий звіт, QlikView.

Abstract

In this work an overall analysis of the structure of existing financial statements from various retailers in selling clothes, shoes and accessories has been performed. A comparison of common and distinguishing features between the reports of various companies, the analysis of common, on which a universal parser for further usage of the data with the overall structure has been done. A parser algorithm has been also designed with all its advantages and disadvantages.

Keywords: parser financial report, QlikView.

При отриманні великої кількості фінансових звітів різних типів, форматів, з різною логікою обробки та різними датами звітності дуже важливою є швидка та чітка обробка даних без втрати чи видозміни потрібної інформації та фільтрацією зайвих даних, так як перше у подальшому призведе до неможливості використання отриманих даних в цілому, а друге у багато разів уповільнить швидкість їх обробки.

Так як звіти різних ретейлерів не підлягають жодній стандартизації, подальша робота з ними або їх порівняння є вкрай складним завданням у першопочатковому вигляді. Таким чином, для того щоб мати можливість далі працювати з отриманою інформацією необхідна певна універсальна структура, формат, що дозволить звести усі ці джерела даних у один загальний вигляд для подальшої ручної або автоматизованої обробки. Це дозволить будувати статистичні залежності, швидше знаходити необхідні записи та порівнювати велику кількість звітів між собою.

Для вирішення даної мети необхідна розробка та застосування універсального парсера, що є метою даної роботи. Це дозволить звести усі дані у однакову структуру для подальшого застосування, у першу чергу – автоматичного, а також для візуального представлення даних користувачу.

У кожному окремому типі звітів присутня як спільна так і унікальна інформація. Якщо остання не несе жодної користі для статистики через неможливість співставлень, то перша є необхідною та важливою. Так, будь-який звіт включає дату продажів, стилі та опис товару, що був проданий або присутній на складі, а також кількість одиниць продукції, її ціна, колір, та ін. Саме ці показники будуть застосовані у подальшій роботі, тому саме ці показники мають бути отримані з першопочаткових джерел, структуризовані (відповідність стилів товарів між різними джерелами, однакова валюта при співставленні даних, її конвертація при необхідності) та виведені (у файл загального формату для подальшої обробки або користувачу на екран).

Так як звіти різних джерел занадто різні по структурі, то для кожного з джерел алгоритм обробки інформації буде відрізнятися до останнього кроку – виведення даних з парсера (цей крок має бути однаковим для усіх джерел). Тобто, розроблений парсер буде складатися з наступних блоків:

- отримання загальної не деталізованої інформації без обробки звітів-таблиць – назва компанії-

ретейлера, дата звіту (NRF-тиждень у форматі YYYY/MM/DD, де YYYY – рік, MM – місяць, DD - день), назва аркуша.

- знаходження інформації, необхідної для обробки звіту – кількість рядків заголовку, що буде ігноруватись (header), номер відділу (department), функції пошуку назв колонок (у випадку, коли для одного ретейлера вони можуть називатись по-різному)

- обробка даних звіта (визначення ретейлера, визначення структури звіту, назв чи порядок колонок, що будуть оброблені, та рядків що будуть проігноровані (наприклад, сумарні значення, що й так будуть підраховані програмою), зчитування даних.

- загальна структуризація отриманих значень та їх виведення як єдиного цілого у файл або на екран.

Для розробки парсера буде застосовуватись BI-платформа QlikView, яка включає широкий спектр функцій для роботи з даними що стосуються фінансів та роботи с текстовими/числовими значеннями та є унікальним інструментом візуалізації даних. [2]

Універсальність досягається кількістю різних джерел і структур, що можуть бути оброблені одним загальним парсером. Його перевагами є також швидкодія, кількість маневрів щодо калькуляції даних, можливостей щодо представлення інформації. Недоліком є залежність від початкових даних – при отриманні некоректних звітів (структура, змінені назви стовпців або змінені типи даних) можливі помилки щодо вихідних результатів, тому у таких випадках необхідне ручне втручання та корекція алгоритмів обробки. Таким чином, стабільність роботи залежить від дотримання першопочаткових звітів схожої структури, наявності усіх обов'язкових полів тощо.

Розроблений парсер матиме широке застосування у задачах бізнес-аналітики, з можливістю його доповнення для поширення спектру використання (додання нових типів структур, нових типів даних тощо).

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Анализ статистических зависимостей – [Електронний ресурс] – Режим доступу : [URL] http://www.distanz.ru/feed/lectures/analiz-statisticheskikh-vzaimosvyazey_3572
2. QlikView® Guided Analytics – [Електронний ресурс] – Режим доступу : [URL] <http://www.qlik.com/us/products/qlikview>

Науковий керівник: Компанець Микола Миколайович – доцент кафедри автоматизації та інформаційно-виміральної техніки, Вінницький національний технічний університет, м. Вінниця
Глуценко Андрій Андрійович – студент групи ІАКІТ16-м факультет комп'ютерних систем та автоматизації, Вінницький національний технічний університет, м. Вінниця, andrewglu94@gmail.com