

## ПІДСИСТЕМА АВТОМАТИЧНОГО ГЕНЕРУВАННЯ ОПИСІВ ГОТЕЛІВ ДЛЯ СИСТЕМИ ПОШУКУ ГОТЕЛІВ "HOTELSBROKER"

Крепич С.Я.<sup>1)</sup>, Літвинчук М.В.<sup>2)</sup>

Тернопільський національний економічний університет

<sup>1)</sup> к.т.н., ст. викладач; <sup>2)</sup> магістрант

### I. Постановка проблеми

Hotelsbroker – це проект який дозволяє порівняти ціни на готелі серед багатьох постачальників. Кожен готель представлений сторінкою-профілем [1].

Для готелів є багато структурованих даних, зібраних за час існування проекту. Зокрема, таблиці, наповнені їх характеристиками. Але у багатьох не вистачає якісних описів, які очікує побачити турист, коли переглядає сторінку готелю. Існує потреба написання ще близько 170 тисяч описів. Є декілька підходів для рішення такої проблеми. Перший – це долучити редактора і написати ці тексти. Другий – це використання шаблонних текстів. Єдиного рішення, яке можна було б використати – немає. У кожному проекті, де виникає така задача, її вирішують самостійно і налаштовують генератор під свій проект. Перевага надається автоматичній генерації тому, що характеристики готелю змінюються з часом і тексти треба переписувати. Основна ідея – це швидка адаптація опису до нових вимог і стану його характеристик. Це той випадок, коли робота програміста коштує менше від роботи редактора.

### II. Мета роботи

Метою роботи є розробка рішення для генерації описів готелів.

### III. Особливості програмної реалізації

Для реалізації програмної системи було використано мову програмування Python [2].

Для того щоб генератор вмів підбирати правильні форми слова підключається додатковий модуль rymorphy2 [3,4]. Це модуль текстового аналізу. Йому на вхід подається слово і відмінок, який хочемо отримати на виході.

В програмі є конфігурації. Конфігурації розміщені в уaml файлах. Для того, щоб працювати з ними використовується модуль PyYAML [5].

Дані по готелях зберігаються в базі даних PostgreSQL. Для того, щоб програма могла працювати з цими даними підключається модуль Psycopg2 [6].

Вирізняють два види характеристик готелю. Зокрема, ідентифікуючі: назва, тип готелю, кількість зірок, адреса і додаткові. До додаткових належать послуги які надаються готелем, комплектація номерів, елементи інфраструктури, визначні місця, які розташовані неподалік, відстані до цих місць, а також до аеропортів чи залізничних вокзалів. Також є години роботи рецепції, інформація про харчування, трансфери і мови, на яких розмовляють. Для підвищення унікальності опису текст розбивається на блоки. Текст може складатись з довільної кількості блоків. Блоки можуть бути в різному порядку. Схематично опис готелю представлений на рисунку 1.

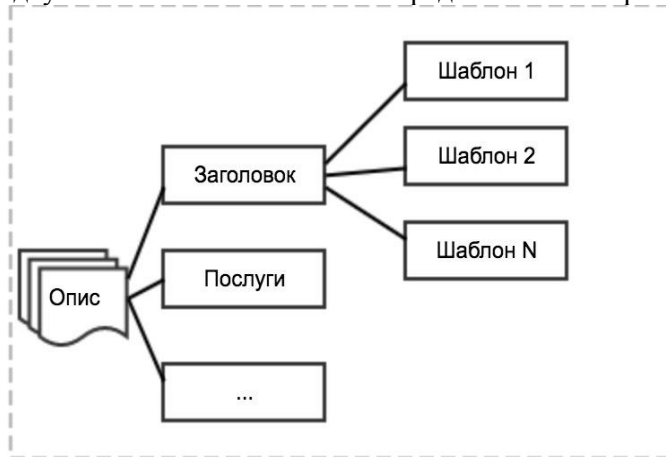


Рисунок 1 - Структура опису

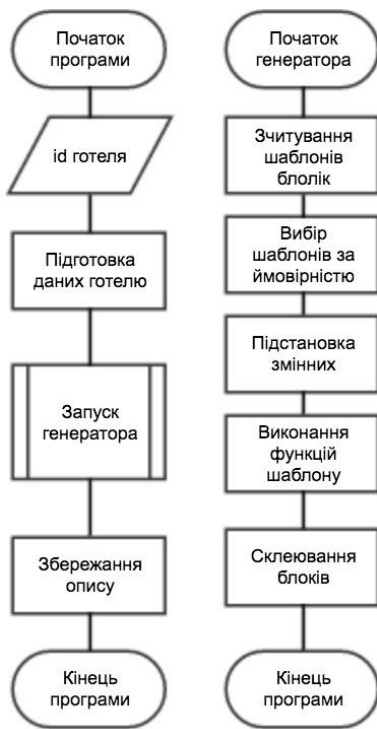


Рисунок 2 - Блок схема алгоритму генерації опису готелю

блоком заголовку опису. В звичайному режимі роботи генератора результат генерації зберігатиметься в базі даних.

Кожен блок відповідає за набір характеристик. В результаті блоки можна перемішувати. Блок може мати свої шаблони. Це свого роду декомпозиція, де кожен невеликий шаблон відповідає за відображення певної ознаки, характеристики чи їх групи.

Характеристик по групах може бути багато. Тому на кожен блок виділяється від 3 трьох до 5 таких характеристик, а на опис від 4 до 6 блоків. Управління шаблонами, редагування, проходить через клієнтський веб-інтерфейс проекту, на якому будуть використовуватись описи. Безпосередньо програма буде запускатись як через веб для генерації одного опису, так і через командний інтерфейс для великої кількості готелів.

На вхід генератор отримує 2 параметри. Перший – це масив шаблонів, які можна використовувати, а другий – це підготовлені характеристики готелю.

Шаблони підтримують вкладені функції, які можна застосувати до даних. Наприклад, зробити першу букву великою чи підібрати називний відмінок слова. Блок-схема з повною послідовністю інструкцій представлена на рисунку 2.

Шаблони блоку мають чисельну ймовірність появи в тексті. Це для того, щоб можна було управляти їх ймовірністю появи в тексті.

При генерації опису через веб-інтерфейс можна вибирати які блоки використати, і по якому шаблону згенерувати блок. Можна редагувати шаблони. При масовій генерації задаються ймовірності появи блоків і ймовірності використання певних шаблонів блоку.

На рисунку 3 представлено екранну форму із згенерованим результатом виводу генерованого блоку заголовку

```
example — -bash — 80x24
MacBook-Pro-NICO:example nico$ python generate.py
[Четырехзвездочной бутик-отель с номерами в классическом стиле с бесплатным WiFi ]
расположен в центре города Львова, в 600 метрах от Глинянских ворот, Бернардинск
ого костела и дворца Потоцких.
MacBook-Pro-NICO:example nico$
```

Рисунок 3 - Результат виводу генерованого блоку заголовку

### Висновок

У роботі представлено рішення для генерації описів готелю. Цю підсистему можна легко адаптувати під інші проекти. Генератор має свій інтерфейс прийому даних і від проекту, який його використовує, залежить лише підготовка вхідних даних.

### Список використаних джерел

1. Сайт проекту Hotelsbroker. [Електронний ресурс]. - Режим доступу: <http://hotelsbroker.com>
2. Офіційна сторінка розробників Python. [Електронний ресурс]. - Режим доступу: <https://www.python.org>
3. Korobov M.: Morphological Analyzer and Generator for Russian and Ukrainian Languages // Analysis of Images, Social Networks and Texts, pp 320-332 (2015).
4. Документація модуля морфологічного аналізу слів rymorphy2. [Електронний ресурс]. - Режим доступу: <https://rymorphy2.readthedocs.io>
5. Офіційна сторінка модуля PyYAML. [Електронний ресурс]. - Режим доступу: <http://pyyaml.org>
6. Документація модуля Pyscorp2. [Електронний ресурс]. - Режим доступу: <http://initd.org/psycorp/docs/>