****

**Завдання Онлайн Раунду**

**Backend | DEV Challenge XVII**

**Зміст:**

1. [Завдання](#_cyrz6xdobf)
2. [Формат презентації результатів](#_ki0zjas8xc7e)
3. [Обмеження та критерії оцінювання](#_7qfi3fq5x58)
4. [Контакти](#_ybnj1ochs2bb)

## **Завдання**

Написати сервіс пошуку дубльованих статей. Стаття вважається дублікатом,   
якщо хоча б 95% її тексту співпадає в точній формі. Мова статей - англійська. Вибір технологій для виконання завдання не обмежується и лягає на плечі Учасника.

**Вхідні дані**: стаття визначається лише її вмістом, полем content. Створити сервіс, який зберігає статті в БД, та повертає список статей-дублікатів.

**Вимоги:**

* стаття зберігається в БД незалежно від того, чи вона є дублікатом, чи ні;
* алгоритм дублікації має враховувати змінений, видалений або добавлений текст;
* поріг в 95% має бути винесеним в конфігурацію, і може бути зміненим.

**Додаткові вимоги**:

* API виклик, який повертав би усі наявні групи дублікатів в БД;
* API виклик, який повертав би список статей уникаючи дублікуючі (повертається лише перша стаття із групи дублікатів);
* Нормалізація слів: одне і теж слово в різних формах рахуються однаковим.

**Очікуваний результат:** Серверний API з базовою функціональністю:

* **POST /articles,** який приймає  
  {"content": "..."}  
  та повертає  
  {"id": 5, "content": "...", "duplicate\_article\_ids": [1, 3]}
* **GET /articles/:id**, який повертає  
  {"id": 3, "content": "...", "duplicate\_article\_ids": [1, 5]}
* **GET /articles**, який повертає усі унікальні статті без дублікатів:  
  {"articles": [  
   {"id": 1, "content": "...", "duplicate\_article\_ids": [3, 5]},  
   {"id": 2, "content": "...", "duplicate\_article\_ids": []},  
   {"id": 4, "content": "...", "duplicate\_article\_ids": []}  
  ]}
* **GET /duplicate\_groups**, який повертає всі групи дублікатів:  
  {"duplicate\_groups": [  
   [1, 3, 5]  
  ]}

## **Формат презентації результатів**

1. Рішення вивантажувати в кабінеті Учасника на [сайті](https://devchallenge.it) в ОДНОМУ файлі-архіві   
   з назвою у форматі **Ім’я\_Прізвище.zip**.
2. Рішення має бути надано у вигляді серверної частини. Підніматись в контексті віртуального оточення Docker, але використання Vagrant також допускається.   
   Для старту вашого додатка має бути необхідним запустити єдину команду: **docker-compose up**, або **vagrant up.**
3. README файл, в якому обов’язково вказуйте:

* Кроки для старту сервiсу;
* Кроки для запуску тестiв;
* Опис API;
* Mетодологію, яку ви обрали, та пояснення, чому на ваш погляд вона найкраще підходить;
* Pечі, на які ви б хотіли звернути увагу або наступнi кроки для вдосконалення вашого сервiсу.

1. SCALEME файл, в якому описано, як ваша система маштабується, які для цього необхідні конфігурації.

Зверніть увагу, що назва архіву - єдине місце, де ви вказуєте свої персональні дані. Назви файлів всередині архіву не мають містити вашого ім’я чи прізвища.

Розмір архіву з рішенням не має перевищувати 10 МВ.

Організатори та Судді залишають за собою право дискваліфікувати роботу Учасника, якщо робота:

* Містить будь-яку вказівку на ім’я, прізвище, електронну пошту, компанію, адресу чи інші персональні дані учасника;
* Виконана у форматі, відмінному від того, що вказаний у завданні;
* Виконана за допомогою сторонніх осіб, а не Учасником особисто.

## **Обмеження та критерії оцінювання**

|  |  |  |
| --- | --- | --- |
| **Категорiя** | **Критерiї** | **Бали** |
| Lite | Працездатність POST /articles, GET /articles/:id | 100 |
| Lite | Дотримання кодстайлу, структурованість та простота коду | 30 |
| Lite | Вибір та реалізація алгоритму пошуку дублікатів | 36 |
| Bonus | Наявність юніт тестів | 24 |
| Bonus | Наявність інтеграційних тестів | 24 |
| Hard | GET /articles для повернення лише унікальних статей | 36 |
| Hard | GET /duplicate\_groups для повернення груп дублікатів | 38 |
| Hard | Слова нормалізуються, так що “go”, “goes” та “went” рахуються однаковими | 19 |

## **Контакти**

Питання та уточнення щодо змісту завдань:   
[канал Slack](https://devchallengehq.slack.com/) : #nomination-backend.   
Судді ігноруватимуть питання, які не стосуються завдань Чемпіонату.   
Організаційні питання: [team@wechallenge.it](mailto:team@wechallenge.it).