

РОЗРОБКА МАСШТАБОВАНОЇ SAAS- ПЛАТФОРМИ ДЛЯ ВАЛІДАЦІЇ ЕЛЕКТРОННИХ АДРЕС В РЕАЛЬНОМУ ЧАСІ

Виконав: Ярошепта Б.П., студент 2-го року навчання
освітньо-наукової програми
«Комп'ютерні науки» спеціальності 122
Науковий керівник: Нагірна А. М.,
кандидат фіз.-мат. наук, доцент

МЕТА

Розробка платформи зумовлена потребою:

- у гнучкому налаштуванні логіки валідації, чого не забезпечують комерційні сервіси;
- у прозорості алгоритмів перевірки та контролі достовірності результатів;
- у застосуванні інтелектуального аналізу для виявлення фейкових адрес;
- у глибокій інтеграції в зовнішні системи через подієву модель;
- у масштабованій обробці запитів у режимі реального часу.

ВСТУП

Об'єкт дослідження: процес валідації електронних адрес у SaaS-середовищі.

Предмет дослідження: алгоритми та архітектура програмної платформи для перевірки адрес електронної пошти в реальному часі.

Завдання:

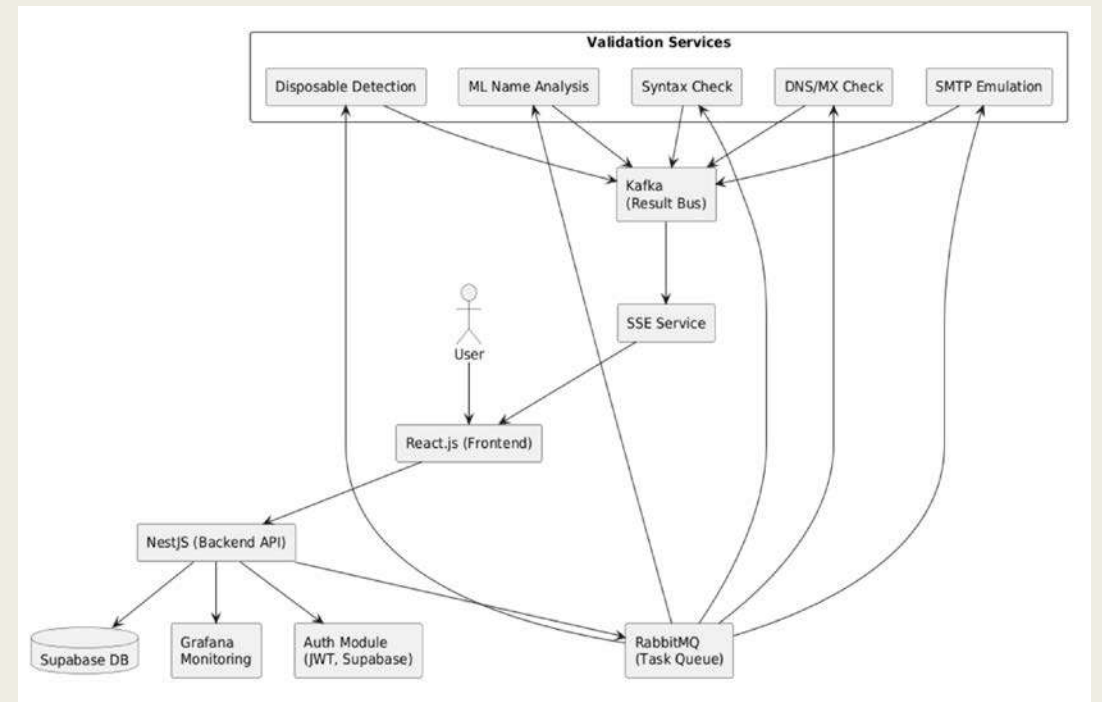
- проаналізувати існуючі сервіси та методи валідації;
- сформулювати технічні вимоги до платформи;
- розробити мікросервісну архітектуру;
- реалізувати повний цикл валідації електронних адрес;
- провести тестування та оцінити ефективність;

ІСНУЮЧІ РІШЕННЯ ТА ЇХ ОБМЕЖЕННЯ

- ZeroBounce - перевірка синтаксису, DNS, SMTP, одноразових адрес, але закритий код та відсутність кастомізації
- NeverBounce - швидка перевірка DNS і SMTP, API-інтеграція, спрощена логіка, проте немає аналізу імені користувача і обмежена аналітика
- Hunter - пошук і перевірка корпоративних адрес, не для масової перевірки, слабкий SMTP-аналіз

ПЕРЕВАГИ ЗАПРОПОНОВАНОГО РІШЕННЯ

- багаторівнева валідація: синтаксис, disposable-перевірка, DNS, SMTP, ML-аналіз;
- робота в реальному часі з асинхронною обробкою;
- масштабованість завдяки мікросервісній архітектурі;
- швидка інтеграція через API та подієву модель (SSE);
- підвищена точність за рахунок фільтрації та нейронної моделі.



ОПИС АЛГОРИТМУ

Синтаксичний аналіз: Перевірка відповідності структури адреси вимогам стандарту RFC 5322.

Визначення одноразових (disposable) адрес: Ідентифікація адрес, що належать до тимчасових поштових сервісів, на основі оновлюваного списку доменів.

Оцінка достовірності імені користувача: Застосування нейронної моделі для виявлення keysmash-подібних імен (автоматично згенерованих чи підозрілих).

Перевірка наявності домену: Аналіз DNS-записів (типів A та MX) для визначення, чи існує зазначений домен.

SMTP-емуляція: Встановлення з'єднання з поштовим сервером і спроба перевірки наявності скриньки без фактичної доставки листа.

Висновок про валідність адреси: Адреса визнається дійсною лише за умови успішного проходження усіх попередніх етапів перевірки.

ПРИКЛАД ЗАСТОСУВАННЯ

- Синтаксис: some-text - invalid
- Перевірка «одноразової» адреси: yaroshepta.b@tempmail.com - invalid
- Перевірка імені користувача: dfgdfghj@gmail.gom - invalid
- Перевірка домену: yaroshepta.b@notexisting.domain - invalid
- Перевірка SMTP сервера: yaroshepta.b@notmail.com - invalid
- yaroshepta.b@gmail.com - valid

mit 9967bcf pushed by yarosheptab	main	2 months ago 1m 1s
mit a883a99 pushed by yarosheptab	main	2 months ago 52s
q using to kafka	main	2 months ago 1m 5s
to sse	main	2 months ago 53s
o sse	main	2 months ago 48s
ifications sctream	main	2 months ago 1m 2s
ard	main	2 months ago 58s
mit 3d17291 pushed by yarosheptab	main	2 months ago 1m 0s

Email Validation
Enter an email address to validate it against our backend rules.

Email

Validate Email

	Valid	Date	Text Validation	Disposable Check	DNS Check
.com	✗	03/06/2025, 00:24:00	✓	✗	✗
	✓	03/06/2025, 00:23:55	✓	✓	✓
mail.com	✓	03/06/2025, 00:23:50	✓	✓	✓
	✗	03/06/2025, 00:23:40	✓	✓	✓
.com	✓	03/06/2025, 00:23:35	✓	✓	✓
	✓	03/06/2025, 00:23:30	✓	✓	✓
	✓	03/06/2025, 00:23:22	✓	✓	✓
blager@gmail.com	✓	03/06/2025, 00:23:17	✓	✓	✓
m	✓	03/06/2025, 00:23:08	✓	✓	✓
	✗	27/05/2025, 13:14:36	✓	✓	✓
m	✗	27/05/2025, 13:04:09	✓	✓	✓

Email
Enter your email

Password
Enter your password

Confirm Password
Confirm your password

Create account

Already have an account? Sign in

Login
Enter your credentials to access your account

Email
Enter your email

Password
Enter your password

Login

Don't have an account? Sign up

ПРОГРАМНА РЕАЛІЗАЦІЯ

НАУКОВА НОВИЗНА

Наукова новизна полягає в розробці багаторівневого алгоритму валідації електронних адрес, який поєднує синтаксичний аналіз за стандартом RFC 5322, перевірку DNS- і MX-записів, емуляцію SMTP-діалогу без доставки листа, а також машинне навчання для виявлення ймовірно автоматично згенерованих імен користувачів.

Запропонований підхід дозволяє зменшити кількість хибно визначених результатів, оптимізувати кількість SMTP-запитів і забезпечити точну, масштабовану валідацію в реальному часі.

ПЕРСПЕКТИВИ

- Додавання підтримки Webhook та OAuth2 для гнучкої інтеграції;
- розширення ML-моделі для виявлення спаму та фішингових адрес;
- впровадження звітів з аналітикою якості бази електронних адрес;
- розробка SDK для популярних мов (Node.js, Python, PHP);
- можливість white-label інтеграції для B2B-клієнтів.

ВИСНОВКИ

У межах роботи реалізовано масштабовану SaaS-платформу для валідації електронних адрес із багаторівневим алгоритмом перевірки, що включає синтаксичний аналіз, виявлення «тимчасових» адрес, перевірку DNS/MX, SMTP-емуляцію та машинне навчання для виявлення фейкових імен користувача. Платформа дозволяє знижувати ризики недоставки, покращує якість контактних баз та забезпечує надійність електронних розсилок у бізнес-процесах.



ДЯКУЮ ЗА УВАГУ!

