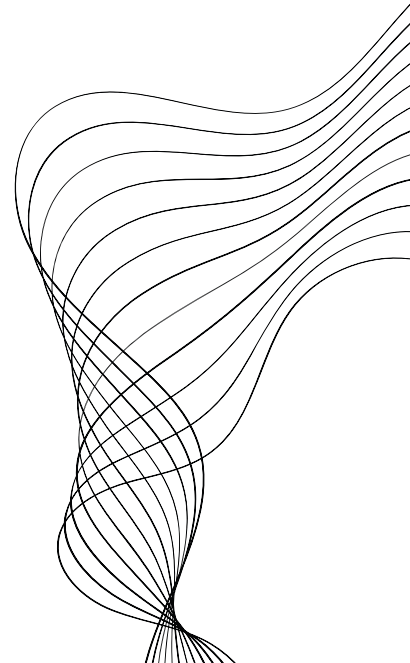


# ՀԱՅԱՍՏԱՆԻ ԱԶԳԱՅԻՆ ՊՈԼԻՏԵԽՆԻԿԱԿԱՆ ՀԱՄԱԼՍԱՐԱՆ

Էլեկտրոնային փաստաթղթերի (PDF)  
բովանդակության ճանաչման և ամփոփման  
ավտոմատացված համակարգի մշակում  
ԱԲ-ի կիրառությամբ

Խումբ՝ SS161  
Ուսանողուհի՝ Արևիկ Խաչատրյան  
Ղեկավար՝ Ռուբեն Մանուկյան



# Նախագծի նպատակը

Ստեղծել AI համակարգ, որը՝

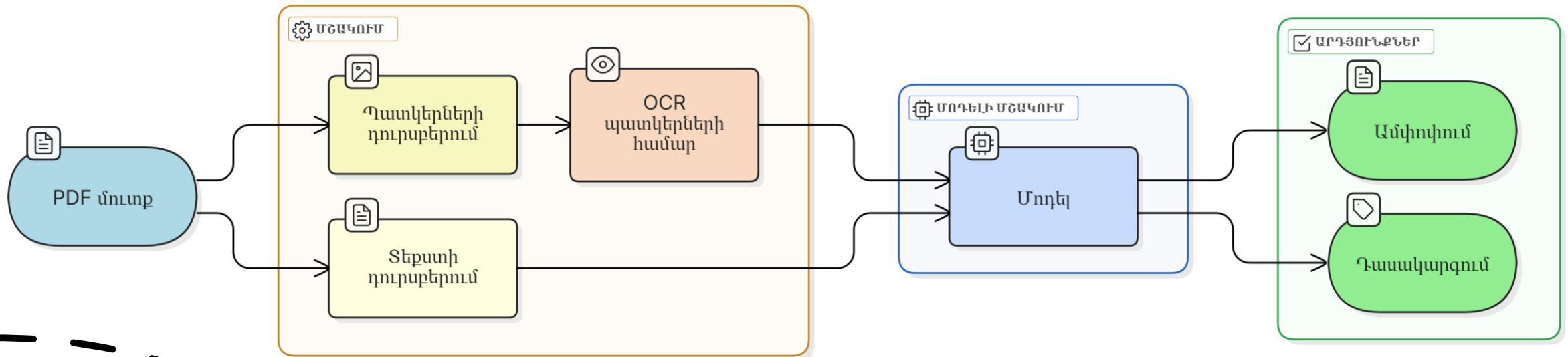
- Մշակում է PDF փաստաթղթերը
- Առանձնացնում է տեքստային և վիզուալ բովանդակությունը
- Անհրաժեշտության դեպքում կիրառում է OCR տեխնոլոգիա
- Ամփոփում է տեղեկատվությունը՝ պահպանելով բովանդակության իմաստը
- Դասակարգում է ամփոփումները ըստ թեմատիկ ուղղությունների

# Խ ն դ ի ր ն ե ր ի լ ու ծ ու մ ն ե ր

---

- PDF փաստաթղթերի կառուցվածքային վերլուծություն և մշակման գրադարանների ինտեգրում
- OCR տեխնոլոգիայի ներդրում պատկերային տեքստի ճանաչման համար
- Տեքստի ամփոփման մոտեցումների ուսումնասիրություն
- Համապատասխան LLM մոդելների ընտրություն և հարմարեցում
- Web ինտերֆեյսի նախագծում և իրականացում
- API ինտերֆեյսի մշակում երրորդ կողմի ինտեգրման համար
- Համակարգի արդյունավետության և ճշգրտության բարելավում

# Ընդհանուր կառուցվածքը



# PDF-ից տեքստի դուրսբերման մոտեցումներ

## Տեքստային շերտի ուղղակի դուրսբերում

- Արագ և ճշգրիտ մեթոդ  
թվային PDF-ների համար
- Չի գործում սկանավորված  
կամ բարդ կառուցվածքով  
փաստաթղթերի դեպքում

## OCR տեխնոլոգիաներ

- Սքանավորված PDF-ների  
տեքստի ճանաչում
- Հիմնվում է խոր  
ուսուցման մոդելների վրա  
(CNN, RNN)
- Կախված է պատկերի  
որակից և պահանջում է  
ռեսուրսներ

## Հիփրիդային մոտեցում

- Համադրում է ուղիղ  
դուրսբերումն ու OCR-ը
- Ունիվերսալ լուծում  
տարբեր տեսակի  
փաստաթղթերի համար
- Պահանջում է  
համակարգային  
բարդություն և համադրում

# Տեքստի ամփոփման մոտեցումներ

Տեքստի ամփոփումը նպատակ ունի ստեղծել կարճ տարբերակ՝ պահպանելով բովանդակային առանցքը: Առանձնացվում են երեք հիմնական մոտեցում՝

## Էքստրակտիվ ամփոփում

Ընտրում է բնագրի ամենակարևոր նախադասությունները՝ առանց վերաձևակերպման: Օգտագործվում են TF-IDF, TextRank, LSA, ինչպես նաև մեքենայական և խորը ուսուցման մոդելներ (BERT, CNN, LSTM):

- *Առավելություններ*՝ փաստացի ճշգրտություն, պարզ իրականացում, քիչ ռեսուրսատար:
- *Թերություններ*՝ կրկնվող կամ անկապ ամփոփումներ, ճկունության խնդիր:

## Աբստրակտիվ ամփոփում

Ստեղծում է նոր, վերաձևակերպված տեքստ՝ պահպանելով իմաստը: Օգտագործվում են Sequence-to-Sequence մոդելներ, Transformer-ներ (T5, BART, GPT):

- *Առավելություններ*՝ բնական, հակիրճ, բովանդակային ամփոփումներ:
- *Թերություններ*՝ հնարավոր սխալներ, մեծ հաշվարկային բարդություն:

## Հիբրիդային մոտեցում

Համադրում է էքստրակտիվ և աբստրակտիվ մեթոդների ուժեղ կողմերը՝ երկփուլ, ուղղորդված կամ բազմամակարդակ ռազմավարություններով:

- *Առավելություններ*՝ հավասարակշռություն ճշգրտության ու բնականության միջև:
- *Մարտահրավերներ*՝ բարդ իրականացում, ինտեգրման դժվարություններ:

# Gemma 3՝ ժամանակակից տեքստի ամփոփման մոդել

Gemma 3-ը Google DeepMind-ի բաց կոդով լեզվային մոդել է (2024), որը հիմնված է Transformer ապակոդավորիչ ճարտարապետության վրա՝ նման GPT-ին: Նախապես ուսուցանված է հրահանգների կատարման վրա:

## Առանձնահատկություններ

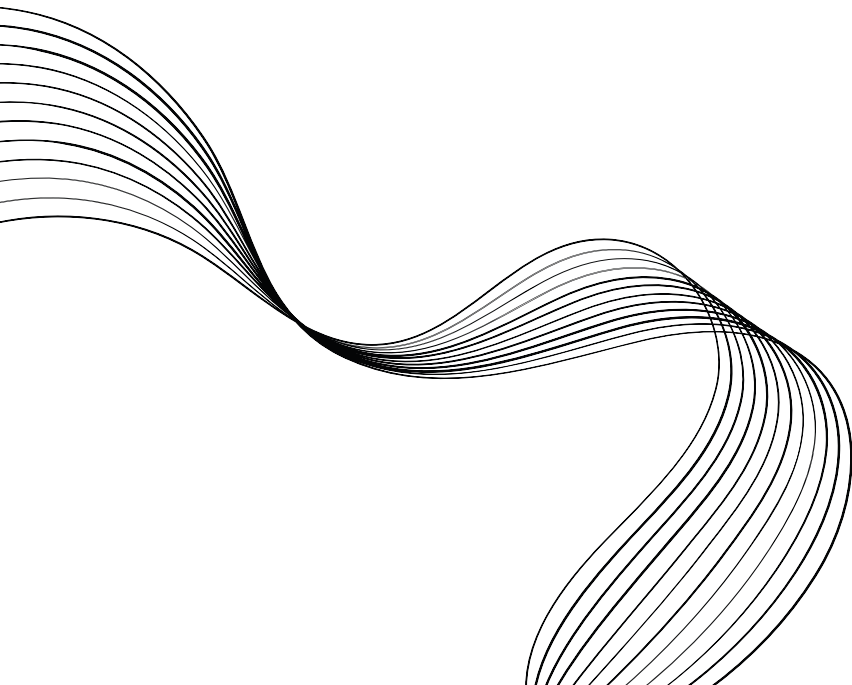
- Բարելավված ուշադրության մեխանիզմ՝ երկար տեքստերի արդյունավետ մշակման համար
- Բաց կոդով, հասանելի տարբեր չափերով
- Ճկուն հարմարեցում կոնկրետ խնդիրների և ոլորտների համար
- Աշխատում է բազմալեզու տեքստերի հետ

## Ամփոփման հնարավորություններ

- Արստրակտիվ ամփոփմամբ բովանդակության վերաձևակերպում
- Zero-shot / Few-shot / Fine-tuned կիրառումներ
- Հարմար է նույնիսկ սահմանափակ ռեսուրսներով միջավայրերում

Gemma 3-ը օպտիմալ ընտրություն է՝ ամփոփման բարձր որակ, բաց կոդ և արդյունավետություն համատեղելով:

# Օգտագործողի ինտերֆեյսի ընդհանուր տեսքը



AMENRU

Ամփոփման Գործիք

Այս գործիքը ամփոփում է PDF ֆայլերը հայերեն, անգլերեն և ռուսերեն լեզուներով: Ստեղծված ամփոփագիրը կլինի նույն լեզվով, ինչ վերբեռնված փաստաթուղթը:

Վերբեռնել PDF

Քաշեք և գցեք ձեր PDF-ը այստեղ կամ [Ընտրել](#)

☐ Միացնել Դասակարգումը

Ամփոփել

Ամփոփման Արդյունք՝

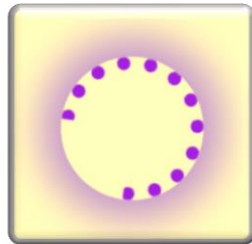
Պատճենել

Ներբեռնել

Դասակարգման Արդյունք՝



# Օգտագործողի փորձը բարելավող ինտերֆեյսային լուծումներ (UX տարրեր)



Վերբեռնել PDF

Քաշեր և գցեր ձեր PDF-ը այստեղ կամ [Ընտրել](#)

AMM.pdf x

☐ Միացնել Դասակարգումը

Ամփոփել



Ամփոփումը պատճենվել է  
սեղմատախտակին!

Մուտքագրեք Ֆայլի  
Անունը

ամփոփագիր.txt

Պահպանել

# Հայերեն ինտերֆեյսով և հայերեն բովանդակությամբ օրինակ

PDF ֆայլից  
հատված

Օրինակի  
դիտարկում

ԲԱԶՄԱՆԴԱՄՆԵՐԻ ԱՐԺԵՔՆԵՐԻ ՀԱՇՎՈՒՄԸ: ՍՏԱՆԴԱՐՏ  
ԳՆԱՀԱՏՄԱՆ ԱԼԳՈՐԻԹՄ և ՀՈՐՆԵՐԻ ՄԵԹՈԴ:

1. Ստանդարտ գնահատման ալգորիթմ (բազմանդամների հաշվարկ):

- Ունենաք  $P(x)$  ձևի բազմանդամ՝  $P(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n$ , որտեղ  $a_i$  բազմանդամի գործակիցներն են:

-  $P(x)$  բազմանդամի արժեքը  $x$ -ի որոշակի արժեքի համար հաշվարկելու համար պարզապես  $x$ -ը փոխարինում ենք բազմանդամի յուրաքանչյուր անդամի մեջ և կատարում թվաբանական գործողություններ՝ արդյունքը ստանալու համար: Օրինակ՝  $P(x)$  տրված  $x$ -ի համար հաշվելու համար կատարում ենք հետևյալը.

$$P(x) = a_0 + a_1 \cdot x + a_2 \cdot x^2 + \dots + a_n \cdot x^n$$

2. Horner մեթոդը (Horner scheme):

- Այս մեթոդը թույլ է տալիս ավելի արդյունավետ հաշվարկել բազմանդամի արժեքը, հատկապես բազմանդամների մեծ աստիճանի դեպքում:

- Ունենք  $P(x)$  բազմանդամը նույն ձևով.  $P(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n$ :

- Հորների մեթոդը սկսվում է բազմանդամը ավելի բարձրից ցածր աստիճանների հաշվարկով, օգտագործելով կրկնության բանաձևը.

## Ամփոփման Գործիք

Այս գործիքը ամփոփում է PDF ֆայլերը հայերեն, անգլերեն և ռուսերեն լեզուներով: Ստեղծված ամփոփագիրը կլինի նույն լեզվով, ինչ վերբեռնված փաստաթուղթը:

Վերբեռնել PDF

Քաշեք և գցեք ձեր PDF-ը այստեղ կամ [Հնարել](#)

Alg\_miji.pdf

☒ Միացնել Դասակարգումը

Ամփոփել

Ամփոփման Արդյունք

Ալգորիթմը հստակ սահմանված հաշվարկային ընթացակարգ է, որը մուտքային տվյալներից ստանում է որոշակի արդյունք: Ալգորիթմի բարդությունը ցույց է տալիս, թե ինչպես փոխվում է աշխատանքի ժամանակը մուտքային տվյալների չափի հետ: Ամփոփատվող վերլուծությունը օգտագործում է Big-O, Omega և Theta նշանները՝ ալգորիթմի բարդությունը գնահատելու համար:

Այս տեքստում քննարկվում են հետևյալ ալգորիթմները.

\* \*\*Երկուական որոնում:\*\* Տեսակավորված զանգվածում տարր գտնելու արդյունավետ ալգորիթմ (Ժամանակային բարդություն՝  $O(\log n)$ ).

\* \*\*Հերմիտյան տեսակավորում:\*\* Պարզ տեսակավորման ալգորիթմ, որը յուրաքանչյուր տարրը տեղադրում է իր տեղում

Պատճենել

Ներբեռնել

Դասակարգման Արդյունք

ակադեմիական

# Անգլերեն ինտերֆեյսով և ռուսերեն բովանդակությամբ օրինակ

## PDF ֆայլից հատված

НА ЗАМЕТКУ

Когда меня спрашивают, в чем секрет нашего успеха, я всегда даю совет: «Не занимайте слишком много тестировщиков».

Как Google справляется с таким маленьким штатом тестировщиков? Если бы мне нужно было ответить просто, я бы сказал, что в Google вся ответственность за качество лежит на плечах тех, кто пишет код. Качество никогда не бывает проблемой «какого-то тестировщика». Каждый, кто пишет код в Google, — уже немного тестировщик, а качество — это проблема всего коллектива (рис. 1.1). Говорить о соотношении численности «разработчики/тестировщики» в Google — то же самое, что рассуждать о чистоте воздуха на поверхности Солнца. Эту тему бессмысленно поднимать. Каждый инженер является и тестировщиком. Если в его должности есть слово «тестирование», то он помогает другим специалистам проводить качественное тестирование.



Рис. 1.1. Программисты Google предпочитают качество широте функциональности

## Օրինակի դիտարկում

AM EN RU

### Summarization Tool

This tool summarizes PDF files in Armenian, English, and Russian. The generated summary will be in the same language as the uploaded document.

Upload PDF:

Drag and drop your PDF here or [Browse](#)

G\_28\_36.pdf

☒ Enable Classification

Summarize

Summary Result:

## Глава 1. Первое знакомство с организацией тестирования в Google - Краткое содержание

В этой главе автор рассказывает о часто задаваемом вопросе: как Google тестирует программное обеспечение. Он объясняет, что подход Google к тестированию постоянно эволюционирует и что он долгое время откладывал написание книги на эту тему, пока не почувствовал, что накопил достаточный опыт и понимание.

Тестирование в Google является частью централизованной системы "направление продуктивности разработки", охватывающей инструменты для разработки, тестирования и выпуска программных продуктов. Google успешно справляется с масштабом и сложностью благодаря решениям в области производительности и масштабирования, позволяющим выпускать продукты с высокой скоростью.

Copy

Download

Classification Result:

Technical

# Ռուսերեն ինտերֆեյսով և անգլերեն բովանդակությամբ օրինակ

Սկանավորված PDF  
ֆայլից հատված

## The paradigm shift in advertising

In an era defined as the 'post-television age' (Cappo, 2003), the way advertising works is changing (Auletta, 2005). Traditional advertising is said to be losing ground to viral marketing, gaming, on-demand viewing, long-form content and other 'new marketing' (Jaffe, 2005). The idea itself of 'advertising's death' is present in several major titles. Ries and Ries (2002) argue that the future for brands is in public relations and not in advertising. Zyman (2002) warns about the end of advertising as we know it. Both books claim that advertising agencies are selling clients short. In these books, when death is mentioned, the authors are describing more a transformation than an end. But they want to underline what they see as unprecedented levels of change. Life after traditional advertising is the topic of another set of books. New marketing techniques, including Internet advertising, video gaming and product placement are the focus of works by Jaffe (2005) and Galician (2004). The synergies achieved between entertainment and advertising through product placement are also examined by Donaton (2004).

The strength of online advertising and its cost-effectiveness is discussed in a variety of works that explain how search engines such as Google are changing current advertising strategies, adding the Internet to the media mix. Marketing without the operational use of agencies has also become a frequent topic in both the academic and managerial debate. Nyren (2005) suggests that there is no longer any need to work with an advertising agency.

Օրինակի դիտարկում

## Инструмент Суммирования

Этот инструмент суммирует PDF файлы на армянском, английском и русском языках. Сгенерированный текст будет на том же языке, что и загруженный документ.

Загрузить PDF:

Перетащите сюда ваш PDF или [Выберите](#)

The Advertising Handbook.pdf x

☒ Включить Классификацию

Суммировать

Результат Суммирования:

## Summarization of "The Advertising Handbook" - Chapter: "Advertising fragmentation: the beginning of a new paradigm?"

This chapter by Francisco J. Pérez-Latre analyzes the significant shifts occurring within the advertising and media industries, arguing that a fundamental paradigm shift is underway. While traditional media still hold a substantial share of advertising expenditure, the rise of digital media and changing consumer behavior are forcing a re-evaluation of established practices.

\*\*Key points:\*\*

\*\*\*Fragmentation & Transformation\*\*\* The media landscape is becoming increasingly fragmented, driven by technological

Копировать

Скачать

Результат Классификации:

Рекламный



# API-ի միջոցով կանչի օրինակ

```
1 import requests
2
3 response = requests.post('http://127.0.0.1:5000/api/summarize',
4                           files={'file': open('C:/Users/user/Desktop/diploma codes/AMM.pdf', 'rb')})
5 print(response.json())
6
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

Python + - [ ] [X] ... ^ X


```
PS C:\Users\user> & C:/Users/user/AppData/Local/Programs/Python/Python311/python.exe "c:/Users/user/Desktop/diploma codes/api.py"
```

```
{'classification': '', 'filename': 'AMM.pdf', 'status': 'success', 'summary': 'Այս տեքստը ամփոփված ներկայացնում է արհեստական բանականության (ԱԲ) տարբեր ապակետներ, ներառյալ փորձագիտական համակարգերը, բազմազան համակարգերը, գենետիկ ալգորիթմները և ոչ հստակ համակարգերը:\n\n**Փորձագիտական համակարգերը** Նախատեսված են փորձագետներին փոխարինելու և որոշումներ կայացնելու համար, օգտագործելով գիտելիքների բազա, որտեղ պահվում են մասնագետների փորձը: Դրանք կիրառվում են ախտորոշման, կանխագուշակման, կառավարման և այլ բնագավառներում:\n\n**Բազմազան համակարգերը (ԲԱՀ)** Բաղկացած են ինքնավար ագենտներից, որոնք փոխգործակցում են միմյանց հետ: Դրանք օգտակար են բաշխված խնդիրների լուծման համար և կիրառվում են տրանսպորտում, ռոբոտատեխնիկայում և այլ ոլորտներում:\n\n**Գենետիկ ալգորիթմները** Էվոլյուցիոն մեթոդներ են, որոնք օգտագործում են ընտրության, խաչասերում և մուտացիա՝ լուծումներ գտնելու համար: Դրանք կիրառվում են օպտիմիզացիայի, մեքենայական ուսուցման և այլ խնդիրների համար:\n\n**Ոչ հստակ համակարգերը** Իմված են ոչ հստակ բազմությունների և ոչ հստակ տրամաբանության վրա, թույլ տալով աշխատել անորոշության պայմաններում: Դրանք օգտագործվում են կառավարման համակարգերում, փորձագիտական համակարգերում և այլ ոլորտներում:\n\nՏեքստը նաև ընդգծում է յուրաքանչյուր համակարգի առավելությունները, թերությունները և նախագծման փուլերը:}'
```

```
PS C:\Users\user>
```

# Եզրակացություն

- Հաջողությամբ մշակվել և իրականացվել է արհեստական բանականության վրա հիմնված համակարգ՝ PDF փաստաթղթերի բովանդակության ճանաչման և ամփոփման համար
- Ստեղծված համակարգը գերազանց մշակում է ինչպես տեքստային, այնպես էլ սկանավորված փաստաթղթեր՝ կիրառելով OCR տեխնոլոգիա
- Համակարգում հաջողությամբ ներդրվել է բազմալեզու աջակցություն (հայերեն, ռուսերեն, անգլերեն)՝ զգալիորեն ընդլայնելով հնարավորությունները
- Մշակվել են օգտագործողի համար հարմարավետ web ինտերֆեյս և ճկուն API՝ ապահովելով համակարգի հեշտ ինտեգրումը



Հնորհակալություն

