

Házi feladat AI asszisztens fejlesztése és tesztelése

Ebben a feladatban egy AI asszisztens webapp-ot hoztunk létre, amely rendelkezik dokumentum feltöltési funkcióval, dokumentumokat elemez, darabol, vektorizál és tárol egy vektor adatbázisban. Rendelkezik még funkcióra szabott system prompt-al és lehetővé teszi a teljeskörű tesztelést.

Lépések:

1. Projekt setup instrukciók
2. API dokumentáció
3. Dokumentumok feldolgozása és a feldolgozott dokumentumok tárolása egy vektor adatbázisban. System prompt elkészítése.
4. Tesztelés és ennek kapcsán pár vizsgálat a válaszok pontosságának javítására.
5. Fejlesztési lehetőségek

Részletek:

1. SETUP INSTRUKCIÓK

1.1 Előfeltételek

A projekt futtatásához az alábbi szoftverek szükségesek:

- **Node.js:** v18.x vagy újabb
- **pnpm:** v8.x vagy újabb (telepítés: `npm install -g pnpm`)
- **Python:** v3.8 vagy újabb
- **Docker:** v20.x vagy újabb (Docker Desktop Windows-on)
- **Git:** verziókezeléshez

1.2 Telepítési lépések

1.2.1 Projekt klónozása

```
git clone <repository-url>
cd ai-sdk-rag-starter
```

1.2.2 Node.js függőségek telepítése

```
pnpm install
```

1.2.3 Python függőségek telepítése

```
pip install requests
```

1.2.4 Környezeti változók beállítása

Létrehoztam egy .env fájlt a projekt gyökerkönyvtárában:

```
OPENAI_API_KEY=sk-proj-your-api-key-here
DATABASE_URL=postgresql://postgres:postgres@localhost:5432/ai-sdk-rag-starter
```

1.2.5 Docker konténer indítása (PostgreSQL + pgvector)

`docker-compose up -d` - Ez elindítja a PostgreSQL adatbázist pgvector kiterjesztéssel.

Ellenőrzés:

```
docker ps
```

1.2.6 Adatbázis séma inicializálás

```
pnpm db:push
```

Ez létrehozza a szükséges táblákat (resources, embeddings) az adatbázisban.

Ellenőrzés az adatbázisban:

```
docker exec -it ai-sdk-rag-starter-postgres-1 psql -U postgres -d ai-sdk-rag-starter
```

1.2.7 Next.js development szerver indítása

```
pnpm dev
```

A szerver elindul a `http://localhost:3000` címen.

1.3 Receptek feltöltése

1.3.1 Egyszeri feltöltés (teszteléshez)

A recept fájlokat (.md) bemásolása a recipes/ mappába

```
python upload_folder.py recipes/
```

1.4 Adatok ellenőrzése

```
docker exec -it ai-sdk-rag-starter-postgres-1 psql -U postgres -d ai-sdk-rag-starter
SELECT COUNT(*) FROM embeddings;
SELECT COUNT(*) FROM resources;
```

1.5 Leállítás

Next.js szerver leállítása: Ctrl+C
Docker konténer leállítása: docker-compose down

2. API dokumentáció:

A projekt két fő API endpoint-tal rendelkezik:

2.1 /api/upload

Dokumentum chunk-ok feltöltése az adatbázisba. Ezt a Python script (upload_folder.py) hívja meg automatikusan.

Python Script Használata

Funkció: Mappa tömeges feltöltése - .txt és .md fájlokat dolgoz fel

Futtatás:

```
bash
python upload_folder.py recipes/
```

Chunking stratégia:

- **YAML front matter** kinyerése metaadatoknak (title, prep_time, stb.)
- **Markdown headingek** szerinti darabolás (## szekciók)
- Minden heading = 1 chunk
- Fallback: mondatonkénti darabolás (.split('.')) ha nincs markdown struktúra

Hibakezelés:

- Részletes exception handling minden lépésben
- File olvasási hibák (FileNotFoundError, PermissionError, UnicodeDecodeError)
- Hálózati hibák (ConnectionError, Timeout)
- Folyamatos logging minden fájlról

Request példa (a Python script által generált):

```
json
{
  "chunks": [
    "Recipe metadata: title: Aloo Matar, prep_time: 45 min, servings: 4",
    "Ingredients\n* 2 medium tomatoes...",
    "Steps\n1. In a blender, puree tomatoes...",
    "Notes\n* Try adding paneer..."
  ],
  "filename": "aloo-matar.md"
}
```

Konzol kimenet:

```
Received 6 chunks from aloo-matar.md. Processing...
Successfully inserted embeddings into the database.
```

2.2 /api/chat

Chat üzenetek küldése és RAG-alapú válaszok fogadása. A válasz streaming formátumban érkezik.

Tool használat: Az endpoint automatikusan meghívja a getInformation tool-t, amely:

1. Embedding generálást végez a felhasználói kérdésből
2. Similarity search-öt futtat az adatbázisban (threshold: 0.80)
3. Visszaadja a top 4 releváns chunk-ot similarity score-okkal

Konzol kimenet példa:

```
Search results: [
  { content: '# Pad Thai', similarity: 0.8716 },
```

```
{ content: 'ingredients...', similarity: 0.8664 }
]
Total results found: 2
```

Ha nincs találat:

Total results found: 0
No results found - checking total embeddings in database...
Total embeddings in database: 484

3. Dokumentumok feldolgozása:

- 3.1 A feladathoz én Jeff Thomson receptjeit választottam. Az előző házi kapcsán már letöltöttem a receptgyűjtemény markdown file-jait, így azokat most csak fel kellett dolgozni.
- 3.2 Először a 85 file-ból kiválasztottam 6 receptfile-t és azzal teszteltem az elkészült kódrészleteket.
- 3.3 Megírtam a feldolgozást és feltöltést végző kódrészletet a kapott anyagok alapján.
- 3.4 Elindítottam a Next.js szerveret.
- 3.5 Lefuttattam a feltöltő script-et (upload_folder.py)
- 3.6 Az upload script sikeresen fel tudta dolgozni a minta file-okat és a feltöltés is sikerült.

```
Talált fájlok: 6
Feltöltés kezdése:
[1/6] aloo-matar.md Sikeres: aloo-matar.md (6 chunk)
[2/6] austro-hungarian-cream-of-horseradish-soup.md Sikeres:
austro-hungarian-cream-of-horseradish-soup.md (4 chunk)
[3/6] baba-ganoush.md Sikeres: baba-ganoush.md (5 chunk)
[4/6] bagels.md Sikeres: bagels.md (6 chunk)
[5/6] barbeque-sauce.md Sikeres: barbeque-sauce.md (5 chunk)
[6/6] beef-tacos.md Sikeres: beef-tacos.md (5 chunk)
--- Összegzés ---
Sikeres: 6
Sikertelen: 0
Összes: 6
```

3.7 Ellenőrzés az adatbázisban:

- 3.7.1 `docker exec -it ai-sdk-rag-starter-postgres-1 psql -U postgres -d ai-sdk-rag-starter`
- 3.7.2 `select count(*) from embeddings;`
count: 31

3.8 Teszt futtatása konzolról:

- 3.8.1 Egy böngészőben megnyitottam a <http://localhost:3000-at>
- 3.8.2 Amikor elindult az AI asszisztens, a prompt-ba beírtam a keresést:
„What can I make with potatoes?”

```
user
What can I make with potatoes?

assistant
Here are a couple of delicious recipes you can make with potatoes:

### 1. Creamy Potato Soup
**Ingredients:**
- 1 lb potatoes (high-starch, like russets), chopped into 1/2" chunks
- 2 medium onions, coarsely chopped
- 1.5 oz butter
- 2 cups vegetable stock
- 1 cup chicken stock (or more vegetable stock)
- 1 cup heavy whipping cream
- 1 large egg yolk
- 2 tbsp dry white wine
- 1.5 oz freshly grated horseradish
- Black pepper and salt to taste

**Garnish:**
- Fresh dill, chopped
- Good quality olive oil

**Steps:**
1. In a large saucepan, heat the butter over low-to-medium heat until bubbling.
2. Add potatoes and onions, increase heat to medium, stirring occasionally until onions are translucent.
3. Add stocks, cover, and bring to a boil.
4. Reduce heat and simmer uncovered until potatoes are tender (about 20 mins).
5. Puree using an immersion blender.
6. Stir in whipping cream, heat until just below boiling, then remove from heat.
7. Whisk egg yolk with white wine and a few ladles of the hot soup, then return to the pot, whisking until incorporate.
8. Stir in horseradish; season with salt and pepper.
9. Top with fresh dill and a drizzle of olive oil.

### 2. Potato and Pea Curry
**Ingredients:**
- 1 lb potatoes, chopped
- 2 cups peas
- Fresh tomatoes, garlic, and ginger (for puree)
- Cinnamon, cloves, and cardamom (whole spices)
- Cumin seeds
- Onion and whole chilies
- Spices (cumin powder, turmeric, etc.)
- Fenugreek leaves and garam masala
- Cream (optional)
- Fresh cilantro for garnish

**Steps:**
1. In a blender, puree tomatoes, garlic, and ginger.
```

Az AI két receptet talált burgonyából:

Creamy Potato Soup

Potato and Pea Curry

3.8.3 Egy másik teszt:

„What vegetarian dishes do you know?”

```
user
What vegetarian dishes do you know?

assistant
Here are some popular vegetarian dishes:

1. **Baba Ganoush**: A creamy dip made from roasted eggplant, tahini, garlic, and lemon juice. It's typically served with pita bread.

2. **Aloo Matar**: A comforting Indian dish made with potatoes (aloo) and green peas (matar), cooked in spices and often served with rice or flatbreads.

3. **Vegetable Stir-Fry**: A quick dish made by sautéing a mix of vegetables—like bell peppers, broccoli, and carrots—with soy sauce and spices.

4. **Chickpea Salad**: A refreshing salad of chickpeas mixed with cucumbers, tomatoes, onions, and a lemon-olive oil dressing.

5. **Caprese Salad**: A simple Italian salad of fresh tomatoes, mozzarella cheese, basil, olive oil, and balsamic vinegar.

6. **Stuffed Bell Peppers**: Peppers filled with a mixture of rice, beans, vegetables, and spices, then baked until tender.

7. **Vegetarian Chili**: A hearty dish made with various beans, vegetables, and spices, often enjoyed with cornbread.

8. **Lentil Soup**: A nutritious soup made from lentils, vegetables, and spices, perfect for a warming meal.
```

Az AI itt 8 találatot adott ki. Azonban összesen csak 6 recept lett feltöltve.

Az AI hallucinációt csinált - kitalált recepteket, amik nincsenek az adatbázisban.

Az adatbázisban csak ezek vannak:

1. Baba Ganoush (megtalálta)
2. Aloo Matar (megtalálta)
3. Bagels
4. Barbeque Sauce
5. Beef Tacos
6. Horseradish Soup

A többi az AI kitalálta ezeket:

- Vegetable Stir-Fry

- Chickpea Salad
- Caprese Salad
- Stuffed Bell Peppers
- Vegetarian Chili
- Lentil Soup

Megvizsgáltam miért történhetett ez. Több okot is feltételezhetünk:

Egyik ok lehet az, hogy a similarity score túl alacsony, ezért az AI a saját tudásából válaszolt a tool eredmények helyett.

Score értékek:

Baba Ganoush: **0.7881**

Beef Tacos: **0.7773**

Aloo Matar: **0.7759**

A másik ok lehet az, hogy a system prompt nem elég részletes vagy pontos. Igazítani kell rajta, hogy ha nincs találat, ne találjon ki semmit, hanem írja meg, hogy nem talált receptet.

Harmadik oka lehet, hogy a feltett kérdés túl általános, a receptekben nincs külön jelölve, hogy vegetáriánus az étel vagy sem.

Megoldási lehetőségek:

Score emelése 0,8-ra.

Receptekben jelölni, hogy vegetáriánus étel (ez most nem opció).

System prompt-ot szigorítani:

You MUST ONLY use information from the getInformation tool to answer questions.

3.8.4 Javítás utáni tesztek eredménye:

A válasz most már csak az adatbázisban lévő recepteket tartalmazza:

1. Aloo Matar (Potato and Pea Curry)
2. Baba Ganoush
3. Creamy Potato Soup

Nincs hallucináció - nem találta fel a Chickpea Salad-ot, Lentil Soup-ot, stb.

3.8.5 A következő lépésben a System prompt-on szigorítottam és feltűnt, hogy a válaszadás hangnemére nincs utasítás benne. Javítottam és pótoltam.

Újabb teszt lekérdezést futtattam. Eredménye:

[Score: 0.7931] info - About 45 minutes...

[Score: 0.7881] # Beef Tacos ← NEM vegetáriánus!

[Score: 0.7877] # Baba Ganoush [Score: 0.7750] info - About 15 minutes...

user

What vegetarian dishes do you know?

assistant

It looks like I didn't find specific vegetarian dishes in my recipe database. Would you like to try searching for a different type of dish or maybe a specific ingredient you want to use?

▼ ► Show how I found this (1 tool call)

Called `getInformation`

```
Input: {  
  "question": "What vegetarian dishes are available?"  
}
```

```
Output: [Score: 0.7931] info  
* About 45 minutes  
* 4 servings as part of a meal with other dishes
```

```
[Score: 0.7881] # Beef Tacos
```

```
[Score: 0.7877] # Baba Ganoush
```

```
[Score: 0.7750] info  
* About 15 minutes  
* 2 servings
```

Látható, hogy nem talált megfelelő receptet és ezt ki is írta udvarias formában.

3.9 Feltöltöttem az összes receptet.

Sikeres: 85

Sikertelen: 0

Összes: 85

3.9.1 Ellenőrzés az adatbázisban:

3.9.1 `docker exec -it ai-sdk-rag-starter-postgres-1 psql -U postgres -d ai-sdk-rag-starter`

3.9.2 `SELECT COUNT(*) FROM embeddings;`

`SELECT COUNT(*) FROM resources;`

93 db resource sor (benne maradt az előző feltöltés is)

484 db embeddings

3.10 SYSTEM PROMPT TESTRESZABÁSA

3.10.1 Szerepkör és személyiség:

3.10.1.1 "Knowledgeable and friendly cooking assistant specializing in recipe guidance"

3.10.1.2 Warm, encouraging, and enthusiastic about cooking

3.10.1.3 Casual, conversational language

3.10.1.4 Offer helpful tips and alternatives when appropriate

3.10.1.5 Show excitement about sharing recipes

3.10.2 Adatforrás korlátozás:

3.10.2.1 "Your knowledge comes exclusively from a carefully curated recipe database"

3.10.2.2 "This database contains detailed recipes with ingredients, steps, timing, and serving information"

3.10.2.3 "Each recipe has been tested and verified"

3.10.3 Válaszadási irányelvek:

3.10.3.1 ALWAYS call `getInformation` tool before answering questions about recipes or cooking

3.10.3.2 ONLY respond using information from tool results - NEVER use general cooking knowledge

3.10.3.3 If no relevant recipes are found (empty or low-quality results), say: "I don't have information about that in my recipe database. Would you like to try a different search?"

3.10.3.4 Present recipes in a clear, easy-to-follow format

3.10.3.5 Include cooking times, servings, and any important notes from the recipes

3.10.4 Tiltások:

3.10.4.1 Never invent or suggest recipes not in your database

3.10.4.2 Don't provide cooking advice outside of the recipe context

3.10.4.3 Don't make substitutions unless mentioned in the recipe notes

system: `You are a knowledgeable and friendly cooking assistant specializing in recipe guidance. Your primary goal is to help users discover and prepare delicious dishes from your curated recipe collection.

ROLE & PERSONALITY:

- Be warm, encouraging, and enthusiastic about cooking
- Use casual, conversational language
- Offer helpful tips and alternatives when appropriate
- Show excitement about sharing recipes

DATA SOURCE:

- Your knowledge comes exclusively from a carefully curated recipe database
- This database contains detailed recipes with ingredients, steps, timing, and serving information
- Each recipe has been tested and verified

RESPONSE GUIDELINES:

- ALWAYS call getInformation tool before answering questions about recipes or cooking
- ONLY respond using information from tool results - NEVER use general cooking knowledge
- If no relevant recipes are found (empty or low-quality results), say: "I don't have information about that in my recipe database. Would you like to try a different search?"
- Present recipes in a clear, easy-to-follow format
- Include cooking times, servings, and any important notes from the recipes

WHAT TO AVOID:

- Never invent or suggest recipes not in your database
- Don't provide cooking advice outside of the recipe context
- Don't make substitutions unless mentioned in the recipe notes`,

4 Tesztesetek a teszteléshez:

4.1 Specifikus alapanyag keresése:

What can I make with chickpeas?

Score-ok: 0.88, 0.86, 0.85, 0.84 - mind nagyon magasak (>0.75)

Talált 3 receptet:

1. Hummus
2. Chickpea Salad (Vegetarian)
3. Channa Masala (Chickpea Curry)

Mind a három valóban csicseriborsót tartalmaz. A válasz részletes, barátságos, és csak az adatbázisból válaszolt.

4.2 Ételtípus keresése:

Show me some soup recipes

Score-ok: 0.84, 0.84, 0.83, 0.83 - határ körül (közel a 0.75-höz)

A találatok közül nem mind származott az adatbázisból. Megint HALLUCINÁCIÓ történt.

Ami tényleg az adatbázisból van:

1. Simple Rice Soup (Biryani Shorba?)

Amit kitalált:

2. "Tomato and Basil Soup" - ez nincs az adatbázisban.

Az AI látta, hogy a "soup" keresésre csak 1 jó találat van (rice soup)

A "Homemade Pasta" (0.8393) találat nem is leves.

Ezért kiegészítette a saját tudásával egy általános paradicsomleves recepttel.

A Horseradish Soup-ot, ami fel van töltve az adatbázisba, nem találta meg.

Valószínűleg a Horseradish Soup csak 0.74 körüli score-ral jött vissza, ami alatt van a threshold-nak.

A kérdéssel a másik probléma, hogy a "soup" szó ritkán szerepel a receptekben explicit módon.

4.3 Konyhai stílus keresése:

Do you have any Thai recipes?

Talált 2 Thai receptet:

1. Pad Thai
2. Thai Fried Rice (Khao Phat Muu)

Mindkettő tényleg az adatbázisban van, részletes hozzávalókkal (ingredients) és lépésekkel. Az összes találat >0.86

4.4 Nehezebb, általános kérdés:

What desserts can I make?

Score-ok: 0.84-0.81 - ezek közel vannak a határhoz, de átmertek.

Amit talált:

3. Chocolate Chip Cookies (test_recipe.txt-ből)

Amit kitalált: A teljes recept részletei hallucináltak.

"Chocolate Chip Cookies are a classic dessert" - csak egy mondat

"Drop spoonfuls onto a baking sheet" - csak egy lépés

"They require flour, sugar, butter, eggs, and chocolate chips"

Az AI ebből kitalált egy teljes receptet pontos mennyiségekkel:

2 1/4 cup flour, 350°F, stb., ami nincs az adatbázisban.

4.5 Olyan alapanyag, ami nincs a receptekben:

What can I make with quinoa?

Válasz: No relevant information found in the knowledge base.

Nem talált megfelelő választ, udvariasan közölte a tényt.

Total results found: 0

No results found - checking total embeddings in database...

Total embeddings in database: 484

Output: No relevant information found in the knowledge base.

Az AI helyesen visszautasította a kérdést, mert:

Quinoa: minden találat <0.80 volt

0 chunk került vissza

Válasz: "I don't have information about that..."

Nincs hallucináció! Nem találta ki a quinoa recepteket.

4.6 Tanulság:

A jelenlegi 0.80 threshold olyan esetekben jól működik, amikor egyértelműen nincs találat. De 0.80-0.85 közötti tartományban (desszert teszt) még mindig hajlamos kiegészíteni.

4.7 Kritikus felfedezés: Score tartományok viselkedése (Boundary values)

Megfigyelés: A RAG rendszer viselkedése erősen függ a similarity score értékétől.

4.7.1 Egyértelmű hiány (score <0.75):

Az AI helyesen visszautasítja a kérdést

Válasz: "I don't have information about that in my recipe database"

Nincs hallucináció

Példa - Quinoa teszt:

Query: "What can I make with quinoa?"

Results: 0 találat (minden <0.80)

Válasz: Tiszta visszautasítás, nincs hallucináció

4.7.2 Határérték környékén (0.75-0.85):

Az AI hajlamos kiegészíteni a hiányos információkat

Saját tudásból ad hozzá konkrét értékeket

Részlegesen hallucinál.

Példa - Dessert teszt (0.80 threshold):

Query: "What desserts can I make?"

Találat score-ok: 0.83, 0.80, 0.80, 0.80

Eredeti chunk: "They require flour, sugar, butter, eggs..."

Válasz: Általános ingrediensek, de pontos mennyiségek nélkül

Viselkedés: Határeset, enyhe kiegészítés

Példa - Soup teszt (0.75 threshold):

Query: "Show me some soup recipes"

Találat score-ok: 0.84, 0.84, 0.83

AI válasz: 1 valódi recept + 1 kitalált "Tomato Basil Soup" X

Viselkedés: Hallucináció történt

4.7.3 Magas score (>0.85):

Az AI megbízhatóan válaszol

Szigorúan az adatbázis tartalmát használja

Nincs hallucináció

Példa - Thai receptek teszt:

Query: "Do you have any Thai recipes?"

Találat score-ok: 0.87, 0.86, 0.86, 0.86

AI válasz: 2 pontos recept (Pad Thai, Thai Fried Rice)

Viselkedés: Tökéletes, nincs hallucináció

Példa - Chickpeas teszt:

Query: "What can I make with chickpeas?"

Találat score-ok: 0.88, 0.86, 0.85, 0.84

AI válasz: 3 pontos recept (Hummus, Chickpea Salad, Channa Masala)

Viselkedés: Tökéletes, nincs hallucináció

5. Fejlesztési lehetőségek:

5.1 Feltöltési folyamat optimalizálása:

- 5.1.1 **Batch embedding generation:** Több chunk egyszerre az OpenAI API-nak
- 5.1.2 **Aszinkron feldolgozás:** Python asyncio használata párhuzamos feltöltéshez
- 5.1.3 **Eredmény:** a kb. 30 recept feltöltése 10-15 percről 2-3 percre csökkenthető

5.2 Metadata strukturált kezelése:

- 5.2.1 Új metadata JSON mező az embeddings táblában
- 5.2.2 Strukturált tárolás: {"title": "Pad Thai", "prep_time": 30, "cuisine": "Thai", "vegetarian": false}
- 5.2.3 **Előny:** Szűrési lehetőség (pl. "Thai recipes under 30 minutes")

5.3 Monitoring és logging:

- 5.3.1 Score eloszlás követése (hány query esik 0.75-0.85 tartományba)
- 5.3.2 Hallucináció detektálás
- 5.3.3 User feedback gyűjtés (thumbs up/down) rossz válaszoknál