

نحوه راه اندازی پروژه

جستجوی هوشمند ابیات حافظ

یک برنامه وب مبتنی بر Streamlit که امکان جستجوی معنایی در کل دیوان حافظ را با استفاده از پیشرفته‌ترین امبدینگ‌های جملات فارسی و جستجوی برداری توسط Qdrant فراهم می‌کند.

مرور پروژه

این پروژه یک رابط جستجوی هوشمند برای غزلیات حافظ ارائه می‌دهد. متن کامل **دیوان حافظ** را دریافت می‌کند، آن را به ابیات (بیت) تقسیم می‌کند، امبدینگ‌ها را با استفاده از مدل `heydariAI/persian-embeddings` محاسبه کرده و داده‌ها را در پایگاه داده برداری Qdrant ذخیره می‌کند تا امکان جستجوی سریع و تقریبی نزدیک‌ترین همسایه فراهم شود. کاربران می‌توانند هر عبارت فارسی را وارد کرده و مرتبط‌ترین ابیات از نظر معنایی را بازیابی کنند.

ویژگی‌های کلیدی ✨

- **جستجوی معنایی:** ابیات را بر اساس معنا پیدا می‌کند، نه فقط تطابق کلمه به کلمه
- **پشتیبانی از زبان فارسی:** چیدمان راست به چپ (RTL) کامل و فونت فارسی (وزیرمتن)
- **امتیازات ارتباط:** امتیاز فاصله را برای هر نتیجه نمایش می‌دهد
- **داکرایز شده:** برای استقرار یکپارچه به راحتی کانتینری شده است
- **رابط کاربری تعاملی:** ساخته شده با Streamlit برای توسعه و اشتراک گذاری سریع

ساختار مخزن

— Dockerfile	# تعریف تصویر داکر
— docker-compose.yml	# تنظیمات سرویس‌ها
— LICENSE	# مجوز MIT
— README.md	# این مستندات
— requirements.txt	# نیازمندی‌های پایتون
— hafez.txt	# متن خام دیوان (منبع)
— run_embedding.py	# اسکریپت تجزیه و امبدینگ ابیات
— embed_query_tester.py	# اسکریپت تست امبدینگ و کوئری (اختیاری)
— query.py	# رابط برنامه‌نویسی جستجو
— streamlit_app.py	# برنامه وب Streamlit
— docker-compose.yml	# برای ذخیره برداری Qdrant شامل سرویس

پیش‌نیازها

- پایتون 3.12+
- داکر و Docker Compose (برای راه‌اندازی کانتینری)
- دسترسی به اینترنت برای دانلود مدل heydariAI/persian-embeddings

نصب داکر (Docker)

ویندوز:

1. دانلود [Docker Desktop](#)
2. اجرای فایل نصبی و دنبال کردن دستورالعمل‌ها
3. پس از نصب، سیستم را ریستارت کنید

مک‌اواس:

1. دانلود [Docker Desktop for Mac](#)
2. باز کردن فایل dmg و کشیدن Docker به پوشه Applications
3. اجرای Docker و تأیید مجوزهای لازم

لینوکس (Ubuntu/Debian):

```
sudo apt update
sudo apt install docker.io
sudo systemctl start docker
sudo systemctl enable docker
sudo usermod -aG docker $USER # نیاز به لاگ‌اوت/لاگ‌این مجدد
```

نصب و راه‌اندازی

1. کلون کردن مخزن:

```
git clone https://github.com/your-username/bachelor-project.git
cd bachelor-project
```

2. ایجاد محیط مجازی پایتون (اختیاری اما توصیه شده):

```
python3 -m venv .venv
source .venv/bin/activate
```

3. نصب نیازمندی‌ها:

```
pip install --upgrade pip
pip install -r requirements.txt
```

4. تولید امبدینگ‌ها:

```
python run_embedding.py
```

این اسکریپت فایل `hafez.txt` را خوانده، غزلها و ابیات را تجزیه می‌کند، امبدینگ‌ها را محاسبه کرده و به مجموعه Qdrant آپلود می‌کند.

اجرای برنامه به صورت محلی

```
streamlit run streamlit_app.py --server.port=7777 --server.address=0.0.0.0
```

در مرورگر خود به آدرس `http://localhost:7777` بروید تا از رابط جستجو استفاده کنید.

استقرار کانتینری شده

1. ساخت تصویر داکر:

```
docker build -t hafez-search .
```

2. اجرا با Docker Compose:

```
docker-compose up --build
```

این دستور پورت `7777` را روی هاست باید کرده و دایرکتوری جاری را mount می‌کند تا امکان ویرایش زنده فراهم شود.

استفاده از Makefile

یا به جای همه این پیچیدگی‌ها می‌توانید از Makefile موجود استفاده کنید:

```
make build && make up
```

نمونه‌های استفاده

- **جستجوی ساده:** کلمات کلیدی مانند `عشق` یا عبارات کامل مانند `آزادگی` را وارد کنید
- **پرسش مفهومی:** `سفر در زمان` برای یافتن ابیات مرتبط با سفر و زمان

دسترسی برنامه‌نویسی

از `query.py` برای انجام جستجو در پایتون استفاده کنید:

```
from query import search
from sentence_transformers import SentenceTransformer
```

```
model = SentenceTransformer('heydariAI/persian-embeddings')
vec = model.encode(['عشق'])[0]
results = search(vec)
print(results)
```

سفارشی‌سازی

- **مدل:** مدل heydariAI/persian-embeddings را با هر مدل سازگار SentenceTransformer جایگزین کنید
- **منبع داده:** فایل hafez.txt را با هر متن فارسی دیگری جایگزین کنید
- **رابط کاربری:** فایل streamlit_app.py و CSS را برای شخصی‌سازی برند تغییر دهید

توسعه‌یافته برای پروژه کارشناسی توسط علیرضا خرمی