

بازیابی پیشرفته اطلاعات (فاز سوم)



نیم سال 4042

مدرس: دکتر اسکندری

دانشکده‌ی مهندسی کامپیوتر – آزاد شیراز

زمان تحویل: 27 خرداد

طراحان: حمید نامجو، امیرحسین همتی و

علی مجاهد

- پروژه انفرادی و یا نهایتاً 2 نفره است.
- فقط در موارد ذکرشده مجاز به استفاده از کتابخانه‌های آماده هستید.
- هر گروه باید کد پروژه خود را در مهلت اعلام شده ارسال نماید.
- علاوه بر گزارش پروژه، در انتهای ترم یک ارائه و دفاع نیز از این پروژه گرفته خواهد شد. مهلت ارسال پاسخ تا ساعت 23:59 روز مشخص شده است.
- همکاری و هم فکری شما در انجام پروژه مانع ندارد اما پاسخ ارسالی هر گروه حتماً باید توسط خود او نوشته شده باشد.
- در صورت هم فکری و یا استفاده از هر منابع خارج درسی، نام هم فکران و آدرس منابع مورد استفاده برای حل سوال مورد نظر را ذکر کنید.

• هدف اصل این تمرین، درک مفاهیم بازیابی عصبی و RAG و تحلیل نتایج است؛ نه صرفاً

پیاده‌سازی مکانیکی کدها.

- امضای توابع را تغییر ندهید، در صورت نیاز توابع کمکی کوچک میتوانید تعریف کنید (به شرطی که ساختار نوتبوک حفظ شود)، پیش از تحویل حتماً چک کنید که کل نوتبوک با گزینه restart & run all یکبار بدون ارور و مشکلی ران بشود، پاسخ سوالات تشریحی را در سلول‌های مشخص شده بنویسید و از کپی‌برداری متنی خودداری کنید و در نهایت تنها خروجی مورد انتظار برای تحویل، نوتبوک تکمیل‌شده است.

1 مقدمه

در این تمرین، هدف آشنایی عملی شما با مفاهیم بازیابی عصبی (Neural Retrieval) و تولید تقویت شده با بازیابی (Retrieval-Augmented Generation یا RAG) است. در این چارچوب، یک سیستم بازیابی و پاسخ گویی مبتنی بر داده‌های فیلم‌های IMDB طراحی و پیاده‌سازی می‌شود که ترکیبی از بازیابی معنایی و تولید متن را به کار می‌گیرد.

2 نمای کلی بخش‌های تمرین

این تمرین در قالب ی نوتبوک ژوپیتتر ارائه شده است و تمام قسمت‌هایی که باید توسط دانشجو تکمیل شوند، با #TODO مشخص شده‌اند. نوت بوک شامل ۹ بخش اصلی به‌صورت زیر است:

۱. Setup و نصب کتابخانه‌ها

۲. بارگذاری و پیش پردازش داده‌ها

۳. تولید embedding‌های عصبی

۴. فرایند Chunking اسناد

۵. ساخت ایندکس برداری با FAISS

۶. Query Expansion و بازیابی

۷. پیاده‌سازی زنجیره‌ی RAG

۸. ارزیابی کیفیت بازیابی

۹. تحلیل و بحث نهایی

موفق و موید):