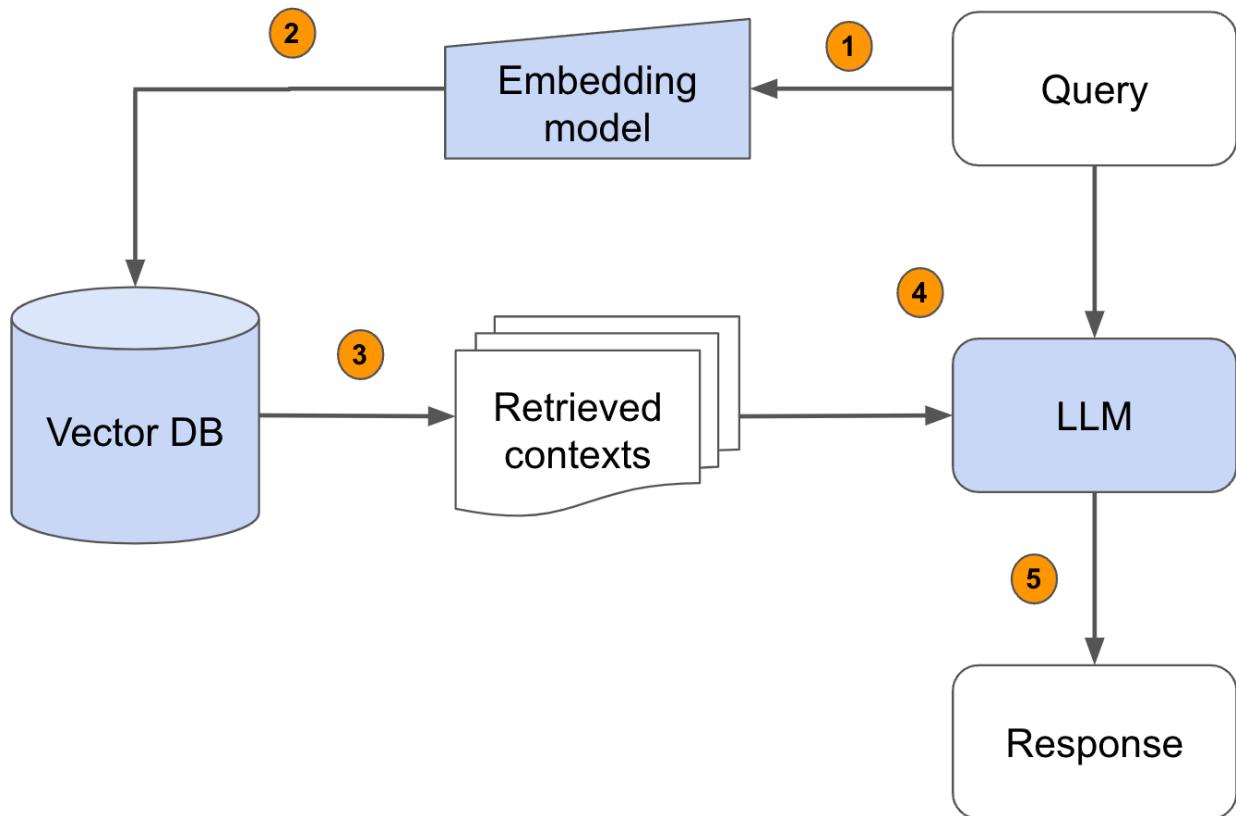


پیاده‌سازی سامانه RAG جهت مشاوره انتخاب واحد هوشمند



نام دانشجو: کیانا نصیری

استاد مربوطه: جناب آقای دکتر تاجبخش

درس: بازیابی اطلاعات

۱۴۰۴ پاییز

فهرست

- ۱. چکیده
- ۲. زیرساختها و ابزارهای فنی
 - ۲.۱. مدل Embedding ○
 - ۲.۲. Vector Database ○
 - ۲.۳. Knowledge Base ○
- ۴. فرآیند بازیابی و Semantic Search
- ۵. دو روش بازیابی و Prompt Engineering
 - ۱.۵. استخراج و تزریق پویا (RAG-Driven Prompting) ○
 - ۲.۵. استفاده از سوالات پیشفرض برای سناریوی انتخاب واحد ○
- ۶. نتایج و خروجی نهایی
 - ۱.۶ خروجی روش اول ○
 - ۲.۶ خروجی روش دوم ○
- ۷. منابع

۱. چکیده

این پروژه با هدف طراحی یک دستیار هوشمند انتخاب واحد برای دانشجویان مهندسی کامپیوتر پیاده‌سازی شده است. سیستم از معماری **RAG: Retrieval-Augmented Generation** بهره می‌برد تا با ترکیب قوانین آموزشی، چارت درسی و کارنامه اختصاصی دانشجو، پیشنهادات دقیقی برای ترم آینده ارائه دهد.

۲. زیرساخت‌ها و ابزارهای فنی

در این پیاده‌سازی از تکنولوژی‌های زیر استفاده شده است:

۲.۱. Embedding

برای تبدیل متون به بردار، از سرویس Google و مدل 001 models/embedding استفاده شد.

۲.۲. Vector Database

برای ذخیره‌سازی و بازیابی اطلاعات، از دیتابیس [Pinecone](#) استفاده شده است. یک Index به نام courseinfo ایجاد شد.

ساختار ذخیره‌سازی: داده‌ها همراه با متادیتا (شامل chunk_text (متن اصلی داده‌ی اینزرتی) و doc_name (اسم موضوع داده مثلاً "پیش‌نیاز" یا "درس‌های پاس شده")) در دیتابیس اینزرت شدند تا در زمان بازیابی، متن اصلی نیز قابل دسترسی باشد.

Knowledge Base .۳

اطلاعات اینزرت شده به ایندکس شامل چهار بخش حیاتی است که منطق تصمیم‌گیری سیستم را تشکیل می‌دهند: (متن همه داده‌های اینزرتی، در لینک نوت بوک ضمیمه شده در بخش ۷. منابع وجود دارد).

- **قوانین آموزشی (hamniaz و pishniaz):** متونی که تعاریف دقیق «پیش‌نیاز» (مانند لزوم پاس کردن ریاضی ۱ برای ریاضی گستته) و «هم‌نیاز» (مانند اخذ آزمایشگاه‌ها) را شرح می‌دهند.
- **چارت تحصیلی (list_of_all_courses):** لیست کامل دروسی که دانشجو باید در طول ۸ ترم بگذراند.
- تفکیک شده بر اساس ترم‌بندی استاندارد. این اطلاعات از طریق Gemini و از روی پی‌دی‌اف چارت درسی کامپیوتر ۹۷ به بعد، به شکل متن استراخ شدند.
- **اطلاعات دروس (courses):** دسته‌بندی دروس به عمومی، پایه، اصلی و تخصصی و قوانین مربوط به هر دسته. از پی‌دی‌اف راهنمایی درسی امیرکبیر استخراج شده که درس‌های دانشگاه ارومیه برای ورودی ۱۴۰۱ هم، بر این اساس است.

- تاریخچه تحصیلی دانشجو (**history**): کارنامه واقعی دانشجو (کیانا نصیری) شامل ریز نمرات از نیمسال ۱۴۰۱ تا ۱۴۰۴. این داده شامل نمرات قبولی و مردودی (مانند فیزیک ۲ در ترم ۱۴۰۱۲) است که برای تشخیص دروس پاس نشده حیاتی است. این داده هم از پی دی اف کارنامه با استفاده از Gemini به شکل متن استخراج شدند.

:Pinecone از پنل دیتابیس پردازی تصویری

۴. فرآیند بازیابی و Semantic Search

تابع `retrieve_context` هسته اصلی جستجوی معنایی در این سامانه است که تعامل با دیتابیس Pinecone را مدیریت می‌کند. مکانیزم عملکرد سیستم هنگام دریافت یک کویری به شرح زیر است:

- **Vectorization:** ابتدا متن سوال کاربر توسط مدل Embedding پردازش شده و به یک فضای برداری با ابعاد ۳۰۷۲ بردگی می‌شود. مدلی با ابعاد بالا انتخاب شده تا بشود خروجی با جزئیات تر خوبی داشت.
 - **محاسبه شباهت:** بردار تولید شده با تمام بردارهای ذخیره شده در دیتابیس (Knowledge Base) مقایسه می‌شود. برای سنجش میزان قربت معنایی، از معیار Cosine Similarity استفاده شده است. این معیار زاویه بین بردار سوال و بردارهای دیتابیس را اندازه‌گیری می‌کند؛ هرچه زاویه کمتر باشد، شباهت معنایی بیشتر است. مطابق بحث کلاسی یعنی هم جهت ترند.
 - **استخراج نتایج (Top-K Retrieval):** در نهایت، سیستم رکوردهایی که متنشان بالاترین امتیاز شباهت کسینوسی را به کویری داشته باشند، به عنوان مرتبه‌ترین بخش‌ها استخراج کرده و پرمیگرداند.

نمونه‌های بازیابی موفق در سیستم:

در پاسخ به کوئری "چه درس هایی پاس کرده؟"، سیستم با تشخیص شباهت معنایی، دقیقاً متن مربوط به (تاریخچه نمرات دانشجو) را بازیابی کرد.

در پاسخ به "لیست درس هایی که باید پاس شود چیست؟"، بردار سوال با بخش `list_of_all_courses` تطبیق داده شد و چارت کامل درسی بازیابی گردید.

در پاسخ به "هم نیاز چیست؟"، قوانین آموزشی مربوط به همنیازها از بخش قوانین استخراج گردید.

```
passed_courses = retrieve_context(['جده درس هایی پاس کرده؟'])
print("\n")
print(passed_courses)
]
✓ 3.0s
Retained : 3

=====
نیمسال اول (۱۴۰۱):
زبان خارجی ، ۲۰ ۳ واحد -
زبان فارسی ، ۱۸ ۲ واحد -
اندیشه اسلامی ۱ (هدایاء و معاد) ، ۱۶ ۲ واحد -
ریاضی عمومی ۱ ، ۱۷ ۲ واحد -
مبانی کامپیوتر و برنامه نویسی ، ۱۷/۵۰ ۲ واحد -
کارگاه مبانی کامپیوتر و برنامه نویسی ، ۲۰ ۱ واحد -
فیزیک ۱ (حرارت و مکانیک) ، ۱۲/۷۵ ۳ واحد -
=====

نیمسال دوم (۱۴۰۱۲):
عرفان علی اسلام ، ۱۸ ۲ واحد -
ریاضی عمومی ۱۸/۷۵ ۲ ، ۱۸/۷۵ ۲ واحد -
ریاضیات گسسته ، ۱۷ ۳ واحد -
برنامه نویسی پیشرفته ، ۲۰ ۲ واحد -
کارگاه برنامه نویسی پیشرفته ، ۲۰ ۱ واحد -
داده های منطقی ، ۱۹/۵۰ ۲ واحد -
کارگاه برق ، ۱۹/۵۰ ۱ واحد -
فیزیک ۲ (الکتریسیته و مغناطیس) ، ۰ ۲ واحد -
=====

نیمسال اول (۱۴۰۲):
segments ⌂ ۲ ⌄ ۴ ⌁ Connect
```

۵. دو روش بازیابی و

برای تولید پاسخ نهایی، از مدل زبانی `gemini-3-flash-preview` استفاده شد. دو روش برای حل این مسئله پیشنهاد میشود:

۱.۵. استخراج و تزریق پویا (RAG-Driven Prompting)

در این روش، ابتدا کوئری کاربر به بردار تبدیل شده و با استفاده از شباهت کسینوسی، مرتبطترین دادهها از پایگاه داده Pinecone بازیابی میشوند.

- **تزریق داده:** سیستم ۳ ریکورد Top-3 RAG Results را که دارای بالاترین امتیاز شباهت هستند استخراج کرده و به عنوان Context به پرامپت نهایی اضافه میکند.

- هدف: این کار باعث می‌شود مدل به جای پاسخ‌های کلی، بر اساس داده‌های واقعی (مانند نمرات دقیق دانشجو در کارنامه و سرفصل دروس) پاسخ دهد.

```

prompt2 = """You are an academic assistant for computer science students.
First, carefully analyze the following information:
1. Review the student's transcript to identify all courses they have already passed (grades of 10 or higher).
2. Compare these with the complete list of required courses (Core, Basic, and Elective).
3. Identify the required core courses the student still needs to complete.

Based on your analysis, recommend courses for the next semester:
- Priority 1: Required core courses for which the student has met all prerequisites.
- Priority 2: Elective courses, if core courses are unavailable or to reach the 15-unit elective requirement.
- Ensure the total course load is between 12-20 units.
- Strictly follow the prerequisite and co-requisite rules as per the provided chart.
- IMPORTANT: The next semester is an even semester (2, 4, 6, 8). Only recommend courses that are offered in even semesters. Courses from odd semesters
- Lab(ازمایشگاه) courses are available in the all semesters.
- Only one course from the group of "محیط زیست", "دانش مندس", "زبان ترکی" must be chosen during the 4 years of bachelor's degree.
Student's passed courses and grades:
ENSURE YOU DONT RECOMMEND COURSES THAT ARE ALREADY PASSED:

USER QUERY:
{user_query}

What You Know about the user background and courses:
{RAG_context}

Based on the user's query and the information provided above:
1. Analyze the relevant academic information
2. Provide a clear, accurate, and helpful response
3. If the query relates to course planning, ensure you follow prerequisite and corequisite rules
4. If the query is about specific courses, provide detailed information about them
5. Answer in the same language as the user's query (Farsi or English)
6. Keep your response concise but comprehensive

Please respond directly to the user's query using only the information provided above.
"""
formatted_prompt = prompt2.format(
    RAG_context=RAG_context,
    user_query=user_query
)

```

۲.۵. استفاده از سوالات پیش‌فرض برای سناریوی انتخاب واحد

اگر سیستم تشخیص دهد که سوال کاربر مربوط به «انتخاب واحد» است، (این بخش سیستم پیاده نشده است، چونکه با فرض اینکه تست بر اساس انتخاب واحد خواهد بود.) از یک راهبرد تکمیلی استفاده می‌کند:

- **بازیابی ۵ سوال کلیدی:** سیستم ۵ سوال استاندارد و آماده را که برای تحلیل وضعیت تحصیلی ضروری هستند (مانند وضعیت پیش‌نیازها، دروس باقی‌مانده از چارت، و تفکیک دروس عمومی و تخصصی) به صورت خودکار بازیابی می‌کند.
- **تغذیه به LLM:** این ۵ سوال به همراه پاسخ‌های استخراج شده از پایگاه دانش در پرامپت قرار می‌گیرند تا مدل بتواند تحلیل چندجانبه‌ای از وضعیت تحصیلی دانشجو ارائه دهد.

```

prompt="""You are a student assistant in course planning and academic advising.

First, carefully analyze the following information:
1. Review the student's transcript to identify all courses they have already passed (grades of 10 or higher).
2. Compare these with the complete list of required courses (Core, Basic, and Elective).
3. Identify the required core courses the student still needs to complete.

Based on your analysis, recommend courses for the next semester:
- Priority 1: Required core courses for which the student has met all prerequisites.
- Priority 2: Elective courses, if core courses are unavailable or to reach the 15-unit elective requirement.
- Ensure the total course load is between 12-20 units.
- Strictly follow the prerequisite and co-requisite rules as per the provided chart.
- IMPORTANT: The next semester is an even semester (2, 4, 6, 8). Only recommend courses that are offered in even semesters. Courses from odd semesters
- Lab( ) courses are available in the all semesters.
- Only one course from the group of "Defense Thesis", "Zibian Turki", or "Majlis Ziyest" must be chosen during the 4 years of bachelor's degree.

Student's passed courses and grades:
ENSURE YOU DONT RECOMMEND COURSES THAT ARE ALREADY PASSED:
{passes_courses_info}
-----
Complete list of courses required for graduation (Core and Elective categories):
{all_courses}
-----
Information about prerequisites:
{pishniaz_info}
-----
Information about corequisites:
{hamniaz_info}
-----

YOUR RESPONSE MUST BE IN Farsi AND INCLUDE:
1. A list of recommended courses for the next semester (names only).
2. Total units for the recommendation.
"""

```

این ۵ سوال به همراه پاسخهای استخراج شده از پایگاه دانش در پرامپت قرار می‌گیرند تا مدل بتواند تحلیل چندجانبه‌ای از وضعیت تحصیلی دانشجو ارائه دهد.

۶. نتایج و خروجی نهایی

۱.۶ خروجی روش اول:

با سلام، من به عنوان دستیار آموزشی شما، وضعیت تحصیلیتان را بررسی کردم. با توجه به اینکه ترم آینده یک *ترم زوج* است و بر اساس دروسی که تاکنون گذرانده‌اید (نمرات بالای ۱۰) و پیشنهادهای تعیین شده، برنامه پیشنهادی برای ترم بعد به شرح ذیل است:

تحليل وضعیت فعلی
شما اکثر دروس اصلی و پایه را با موفقیت گذرانده‌اید. موارد باقیمانده شامل دروس اختصاصی ترم ۸، چند دروس پیشنهادی در آن نداشتید. همچنین با گذراندن «دفاع مقدس»، نیازی به اخذ دروس «محیط زیست» یا «زبان ترکی» ندارید.

دوروس پیشنهادی برای ترم آینده (ترم زوج):

***اولویت ۱:** دروس اصلی و پایه باقیمانده (مطابق جارت ترم‌های ۴ و ۸)
پژوهه نرم افزار (۳ واحد): با توجه به گزارش مهندسی نرم افزار ۲، مجاز به اخذ درس مستبدله
میانی رایانش ابری (۲ واحد): پیشنهادهای این درس (سیستم عامل و شبکه) را گذرانده‌اید.
۳۰ *** انتقال داده (۳ واحد): باز دروس اصلی ترم

مدارهای الکتریکی و الکترونیکی (۲ واحد): است که در گزارش مهندسی مشاهده شده است.

آزمایشگاه مدارهای الکتریکی و الکترونیکی (۱ واحد): همچنان درس بالا لازم است.

انتدیشه اسلامی (۲ واحد): با توجه به نمره ۱۴۰/۲۱ در ترم ۸، باید مجدد در این ترم زوج اخذ شود.

آزمایشگاه فیزیک عمومی (۱ واحد): از دروس پایه که باقیمانده است (آزمایشگاه‌ها در تمام ترم‌ها از نظر معرفتی متشوند).

***اولویت ۲:** دروس اختیاری (برای رسیدن به متفق ۱۵ واحد اختیاری)
یک درس اختیاری تخصصی (۲ واحد): شما تاکنون ۸ واحد اختیاری (یادگیری ماشین، یادگیری عقیق و دفاع مقدس) گذرانده‌اید.
۸۰ *** برای نزدیک شدن به سقف ۱۵ واحد اختیاری، پیشنهاد می‌شود یک درس ۳ واحدی دیگر اخذ کنید.

خلاصه واحدها برای ترم آینده
* دروس اختصاصی و پایه: ۱۰ واحد***
* دروس عمومی (انتدیشه ۲): ۲ واحد***
* دروس اختیاری: ۳ واحد***
* مجموع واحد پیشنهادی: ۱۹*** واحد*

دروس مهم
درس **اصول طراحی کامپیوتر** چون مزبور به ترم فرد (۷) است، در این ترم ارائه نمی‌شود و باید در ترم بعد از آن اخذ کنید.
تمامی پیشنهادهای دروس بالا (مانند ریاضی ۲ و فیزیک ۲ برای مدار، و مهندسی نرم افزار برای پژوهه) توسط شما با موفقیت گذرانده شده است.

آیا تمایل دارید لیست دروس اختیاری ارائه شده در دانشکده را هم بررسی کنیم؟

۲.۶ خروجی روش دوم:

ترم آینده یک **ترم زوج** است و بر اساس دروسی که تاکنون گذرانده اید (نمرات بالای ۱۰) و پیشنهادهای تعیین شده، برنامه پیشنهادی برای ترم بعد به شرح زیر است ...

تحلیل وضعیت فعلی:

کارنامه شما نبودند و درس اندیشه ۲ است که نمره قبولی در آن نداشتید. همچنین با گذراندن «دفاع مقدمی»، نیازی به اخذ دروس «محیط زیست» با «زبان ترکی» ندارید.

دروس پیشنهادی برای ترم آینده (ترم زوج):

اولویت ۱: دروس اصلی و پایه باقیمانده (طبقه باریت ترم‌های ۲ و ۸)

بروزه نرم افزار (۲ واحد): با توجه به گذراندن مهندسی نرم افزار، مجاز به اخذ این درس هستید**.

مبانی رایانش ابری (۲ واحد): پیشنهادهای این درس (میستم عامل و شبکه) را گذرانده اید**.

انتقال داده (۲ واحد): از دروس اصلی ترم ۳۰ ***.

مدارهای الکتریکی و الکترونیکی (۲ واحد): این درس در چارت ترم ۴ (زوج) است که در کارنامه شما مشاهده نشد**.

آزمایشگاه مدارهای الکتریکی و الکترونیکی (۱ واحد): همینیاز درس با**.

اندیشه اسلامی ۲ (۲ واحد): با توجه به نمره ۰ در ترم ۱۴۰۲۱، باید مجدداً در این ترم زوج اخذ شود**.

آزمایشگاه فیزیک عمومی ۲ (۱ واحد): از دروس پایه که باقی مانده است (آزمایشگاه‌ها در تمام ترم‌ها ارانه می‌شوند)**.

اختیاری (یادگیری ماشین، یادگیری عمیق و دفاع مقدمی) گذرانده اید. برای نزدیک شدن به سقف ۱۵ واحد اختیاری، پیشنهاد می‌شود یک درس ۳ واحدی دیگر اخذ کنید**.

اولویت ۲: دروس اختیاری (برای رسیدن به سقف ۱۵ واحد اختیاری) گذرانده اید. برای نزدیک شدن به سقف ۱۵ واحد اختیاری، پیشنهاد می‌شود یک درس ۳ واحدی دیگر اخذ کنید.

خلاصه واحدها برای ترم آینده**

* دروس اختصاصی و پایه: ۱۴ واحد**.

* دروس عمومی (اندیشه ۲): ۲ واحد**.

* دروس اختیاری: ۲ واحد**.

* سجموع واحد پیشنهادی: ۱۹*** واحد**.

**نکات مهم

درس **اصول طراحی کامپیالر** چون مربوط به ترم فرد (۷) است، در این ترم ارانه نمی‌شود و باید در ترم بعد از آن اخذ کنید.

تمامی پیشنهادهای دروس بالا (مانند ریاضی ۲ و فیزیک ۲ برای هدار، و مهندسی نرم افزار برای پروره) توسط شما با موقتیت گذرانده شده است.

آیا تمایل دارید لیست دروس اختیاری ارانه شده در دانشکده را هم بررسی کنیم؟

در هر دو روش خروجی مورد انتظار بسیار مناسب بوده و درس‌های ترم بنده دانشجوی مورد بررسی را به درستی پیشنهاد می‌دهد.

۷. منابع

- چارت درسی مصوب دانشگاه صنعتی امیرکبیر
- کارنامه تحصیلی دانشجو (نیمسال ۱۴۰۱۱ الی ۱۴۰۴۱)
- Google Gemini v3 Flash Preview
- Google Generative AI Embedding
- Pinecone Serverless
- [Source notebook](#)