Librarian

Проект по предметот Мобилни информациски системи

*Тодор Јовановски (213133) ; Ведран Десковски (213159)*

Librarian претставува мобилна апликација во која корисникот може да поставува книги (или било каков друг документ) во PDF формат и тој може да поставува прашања во врска со таа книга и апликацијата да одговори на тоа прашање. Апликацијата е наменета за корисници кои наидуваат на нејасни делови во некоја книга или кога текстот е тежок за разбирање, овозможувајќи им да добијат појаснување во секое време.

Технологии користени во нашиот back-end:

* FastAPI, LangChain, HuggingFace (Python)
* Pinecone (векторски бази на податоци)

**Објаснување на нашиот back-end:**

Откога корисникот ќе постави PDF фајл во апликацијата, се извршуваат следните чекори:

1. Целиот PDF фајл се конвертира во еден голем String формат.
2. String форматот се дели на парчиња (chunks), каде што едно парче се состои од 1024 карактери.
3. Од HuggingFace се зема модел за ембедирање и секој од овие парчиња (strings) се ембедира на 768 димензии. На овој начин се креираат вектори.
4. Откога ќе се креираат векторите, тие се внесуваат во векторска база на податоци (Pinecone).
5. Корисникот веднаш е навигиран до соодветниот chat room.

При поставување на некое прашање, се извршуваат следните чекори:

1. Се зема соодветен LLM модел од HuggingFace. Во оваа апликација го одбравме Mistral. Исто така се дефинира и prompt темплејт, кој служи за полесно моделот да го создаде одговорот.
2. Во зависност од тоа во кој чет се наоѓа корисниот, до API-то се испраќа идентификатор на книгата, кој служи да направиме пребарување на сличност (similarity search). Стандардната метрика за ова пребарување е cosine similarity (косинусова сличност). Поставеното прашање се ембидира во иста димензија во која е ембедирана книгата и преку cosine similarity се враќаат 4 вектори кои што се најслични со прашањето.
3. Овие 4 вектори служат како контекст на LLM моделот. Моделот го анализира контекстот и соодветно праќа одговор на поставеното прашање.

При бришење на чет, до API-то се испраќа идентификаторот на книгата и според истиот се бришат сите вектори од базата на податоци кои одговараат на тој идентификатор.

Технологии користени за креирање на мобилна апликација:

* .NET MAUI (C#)
* SQLite за перзистирање на податоци

**Објаснување на текот на мобилната апликација:**

При отворање на апликацијата, корисникот може веднаш да почнува со внесување на книга. Исто така има копче кое го навигира корисникот до листа која ги прикажува сите четови кои ги има креирано.

За креирање на нов запис:

1. Корисникот внесува име на документот и го поставува во PDF формат.
2. Пред да кликне на копчето “Start new chat”, корисникот може да го поништи овој документ со кликање на копчето X.
3. Откога корисникот ќе започне нов чет, истиот се перзистира во SQLite базата и се навигира кон нов екран кој е одговорен само за еден документ.

За поставување на прашање:

1. Корисникот го пишува своето прашање преку тастатура, клика “send” и апликацијата му враќа одговор на тоа прашање.

* Корисникот исто така има шанса за “speech to text”, каде се вклучува микрофонот на телефонот, тој зборува и се креира прашањето.
* Доколку корисникот се предомисли за прашањето, тој може да го ресетира. Или пак да се врати на тастатура.

1. Прашањето на корисникот и одговорот на апликацијата се зачувуваат во SQLite база.

Во екранот за прегледување на сите четови, корисникот може да избрише чет при “swipe”. Со клик се навигира кон соодветниот чет, каде се покажува последната интеракција меѓу корисникот и апликацијата. Ако нема интеракција се покажува празен чет со соодветна порака.