Završni zadatak - Josip Šare

Ak. god. 2023./24.

Aplikacija za inženjering upita razgovornih agenata

Inženjering upita (engl. prompt engineering) je vještina oblikovanja precizno formuliranih upita (engl. prompts) koji učinkovito usmjeravaju razgovorne agente (engl. chatbots) u generiranju traženih odgovora. Ova tehnička vještina omogućava bolju i precizniju komunikaciju čovjeka i stroja, a danas postaje sve zanimljivija jer razgovorni agenti dobivaju nova i važna područja primjene. U ovom radu potrebno je upoznati se s osnovnim principima inženjeringa upita, uključujući kako pravilno formulirati upite koji potiču razgovorne agente na generiranje korisnih, preciznih i sigurnih odgovora. Cilj završnog rada je izraditi aplikaciju za oblikovanje upita definiranih formata i strukture temeljem principa inženjeringa upita. Pohraniti upite u relacijsku bazu podataka i preoblikovati ih. Odabrati referentne razgovorne agente i provesti eksperimentalno vrednovanje upita te statističku obradu rezultata. Radu priložiti izvorni i izvršni kod razvijenog sustava uz potrebna dodatna objašnjenja i dokumentaciju. Također, radu priložiti izrađene skupove podataka te citirati korištenu literaturu.

Application for chatbot prompt engineering

Prompt engineering is the ability to design precisely formulated queries (prompts) that effectively direct conversational agents (chatbots) to provide the desired responses. This technical skill enables better and more precise human-machine communication and is becoming increasingly interesting today as conversational agents open up new and important areas of application. In this thesis, it is necessary to familiarize oneself with the basic principles of prompt engineering. This includes how to formulate queries correctly so that conversational agents can generate useful, precise and secure responses. The goal of this thesis is to create an application for designing queries of defined formats and structures based on the principles of prompt engineering. Store and redesign queries in a relational database. Select reference conversational agents and perform an experimental evaluation of the queries and statistical analysis of the results. Include the source and executable code of the developed system with the necessary additional explanations and documentation. Include the generated data sets and cite the literature used.