

IRTM - RAG

Răzvan-George Costea

January 2025

1 Introduction

In this project we had to build RAG system on top of a base retrieval one. The scope of it is to observe what system works the best for retrieving data from a given query. The comparison is between the basic retrieval system, the Large Language Model alone and the RAG, which is a combination of both. The language given is romanian and the language model is OpenLLM-Ro/RoLlama2-7b-Base and sentence-transformers/paraphrase-multilingual-mpnet-base-v2 for encoding whole documents.

2 Description of each system

2.1 Information Retrieval System

Those are basic systems that retrieve documents based on a given query and some dictionary. In this project we have two retrieval systems, one based on inverted dictionaries and one based on KDTrees with sentence-transformers/paraphrase-multilingual-mpnet-base-v2 as embedding model.

2.2 Large Language Model

The main Large Language Model used for both question answering and document retrieval is OpenLLM-Ro/RoLlama2-7b-Base. It is used alone for both tasks.

3 Retrieval Augmented Generation

The model from the previous subsection is also used in pair with an IR system for both document retrieval and question answering.

4 Method

4.1 Data

Data used are pdf files in romanian collected from wikipedia from different subjects. The titles are: "Batalia din dealul Spirii", "Indicele calitatii vietii", "Oceanul Atlantic", "Reolutia Americana", "Romania". For computational reasons, the texts are summarized for RAG experiments.

4.2 Retrieval

For retrieval we use inverted dictionaries, KDTrees, and Large Language Models.

Inverted dictionaries are used to search in what documents are the words that are present in the query.

KDTree is a data structure used for organizing points in a k-dimensional space. It is used Nearest Neighbor Search and Efficient Indexing.

The LLM is used as an embedding model that encode the query and the documents. The cosine similarity is used to find the document with the highest match with the query.

4.3 Question Answering

For QA we have two methods, LLM alone and LLM with context from retrieval systems.

For the LLM alone, we just pass the question to it and the model responds based on its knowledge. The problem with it is that the model might hallucinate.

In theory, a RAG system is more precise in answering a question, because it uses the given context as additional knowledge. A drawback for this system is the accuracy of the retrieval system that provides context.

5 Comparison

5.1 Retrieval

In this section we compare all retrieval systems, the inverted dictionary and KDTree.

The conclusion from the Table 1 is that KDTree retrieval is superior to inverted dictionary, therefore we continue with it as the IR system. The table below shows the retrieval tasks performed with the LLM alone and with LLM combined with KDTree.

In the Table 2 it is shown the retrieval of documents using LLM model alone, LLM paired with KDTree IR system and KDTree alone.

5.2 Question Answering

In this subsection we will compare the LLM alone and the LLM with KDTree on the questions answering task. The comparison table is listed bellow

Method	Query	Response
LLM	Când a avut loc Bătălia din Dealul Spirii?	În 1990, în urma unui referendum, Republica Moldova a declarat independența față de URSS.
LLM+KDTree	Când a avut loc Bătălia din Dealul Spirii?	13 septembrie 1848
LLM	Describe semnificația istorică a Bătăliei din Dealul Spirii în contextul Revoluției de la 1848?	În cazul în care unul dintre părinți este plecat în străinătate, iar celălalt părinte este în țară, atunci părintele care este plecat în străinătate va trebui să își ia concediu de odihnă pentru a putea să își petreacă timpul cu copilul.
LLM+KDTree	Describe semnificația istorică a Bătăliei din Dealul Spirii în contextul Revoluției de la 1848?	1848 a fost un an de revoluție în Europa, în care au avut loc schimbări politice majore în mai multe țări. În România, revoluția a fost declanșată de mișcarea lui Tudor Vladimirescu, care a încercat să obțină independența țării față de Imperiul Otoman. În 1848, revoluția a
LLM	Ce rol au avut evenimentele de la Dealul Spirii în dezvoltarea naționalismului românesc?	În cazul în care unul dintre părinți este cetățean străin, iar celălalt cetățean român, copilul va dobândi cetățenia română la naștere, dacă celălalt părinte este cetățean român și locuiește în România.
LLM+KDTree	Ce rol au avut evenimentele de la Dealul Spirii în dezvoltarea naționalismului românesc?	1. Evenimentele de la Dealul Spirii au avut un rol important în dezvoltarea naționalismului românesc, deoarece au arătat că românii sunt capabili să se opună cu succes unei armate mult mai numeroase și mai bine înarmată.

LLM	Care a fost rolul militarilor în această confruntare?	În primul rând, trebuie să știi că nu există un singur tip de păr. Există părul creț, ondulat, drept, ondulat, creț, drept, ondulat, creț, drept, ondulat, creț, drept, ondulat, creț, drept, on
LLM+KDTree	Care a fost rolul militarilor în această confruntare?	Și-au dat viața pentru țară.
LLM	Ce factori determină indicele calității vieții?	1.1.1.1
LLM+KDTree	Ce factori determină indicele calității vieții?	1. Bunăstarea materială măsurată în PIB pe cap de locuitor (în dolari americani , la PPPS constant în 2006) 2. Speranța de viață la naștere 3. Calitatea vieții de familie bazată în principal pe ratele de divorț 4. Starea libertăților politice 5. Siguranța locului de muncă
LLM	Cum influențează indicele calității vieții politica globală?	În cazul în care unul dintre părinți este cetățean străin, iar celălalt cetățean român, copilul va dobândi cetățenia română la naștere, dacă celălalt părinte nu a renunțat la cetățenia română.
LLM+KDTree	Cum influențează indicele calității vieții politica globală?	Țările cu un indice ridicat al calității vieții sunt mai puțin susceptibile de a fi implicate în conflicte armate, deoarece oamenii sunt mai puțin predispuși să se revolte împotriva guvernului. Țările cu un indice ridicat al calității vieții sunt mai puțin susceptibile de a fi implicate în conflicte armate,

LLM	Explică modul în care este calculat indicele calității vieții.	În cazul în care se constată că nu sunt îndeplinite condițiile de eligibilitate, solicitantul va fi informat cu privire la motivele care au stat la baza neacordării ajutorului de stat.
LLM+KDTree	Explică modul în care este calculat indicele calității vieții.	Indicele calității vieții este calculat prin combinarea rezultatelor unor sondaje de satisfacție a vieții cu previziunile privind PIB-ul pe cap de locuitor pentru a determina un scor al unei națiuni.
LLM	Cum este România clasificată în funcție de calitatea vieții?	În cazul în care nu se poate face dovada că persoana care a încasat venitul a fost angajată în baza unui contract individual de muncă, venitul va fi impozitat la sursă, prin reținere la sursă, de către plătitorul venitului.
LLM+KDTree	Cum este România clasificată în funcție de calitatea vieții?	Țara noastră se află pe locul 49 în lume în ceea ce privește calitatea vieții, conform clasamentului realizat de Banca Mondială.
LLM	Care este importanța acestui indice în politică și economie?	În primul rând, trebuie să știi că nu există un singur tip de păr. Există părul creț, ondulat, drept, ondulat, creț, drept, ondulat, creț, drept, ondulat, creț, drept, ondulat, creț, drept, on
LLM+KDTree	Care este importanța acestui indice în politică și economie?	Și care este importanța acestui indice în politică și economie?

LLM	Ce țări sunt învecinate cu Oceanul Atlantic?	Țările învecinate cu Oceanul Atlantic sunt: Europa, Africa, America de Nord și America de Sud.
LLM+KDTree	Ce țări sunt învecinate cu Oceanul Atlantic?	Țările care sunt învecinate cu Oceanul Atlantic sunt: America de Nord, America de Sud, Europa, Africa, Antarctica.
LLM	Ce descoperiri istorice au fost făcute în Oceanul Atlantic?	În 1999, într-un sat din nordul Chinei, a fost descoperit un schelet uman vechi de 11.500 de ani.
LLM+KDTree	Ce descoperiri istorice au fost făcute în Oceanul Atlantic?	1. Insula Madeira 2. Insula Azore 3. Insula Capului Verde 4. Insula Feroe 5. Insula Sfânta Elena 6. Insula Tristan da Cunha 7. Insula Ascension 8. Insula São Tomé 9. Insula Anobón 10. Insula Annobón 11. Insula Tristan da Cunha
LLM	De ce este Oceanul Atlantic considerat esențial pentru comerțul global?	În cazul în care unul dintre părinți este plecat în străinătate, copilul poate fi încredințat unei rude sau unei persoane de încredere, care locuiește în aceeași localitate cu părintele plecat.
LLM+KDTree	De ce este Oceanul Atlantic considerat esențial pentru comerțul global?	Oceanul Atlantic este cel mai mare ocean de pe Pământ
LLM	Explică influența curenților oceanici din Atlantic asupra climei globale?	În cazul în care se constată că nu sunt îndeplinite condițiile de eligibilitate, cererea de finanțare nu va fi evaluată și nu va fi transmisă către Autoritatea Contractantă.

LLM+KDTree	Explică influența curenților oceanici din Atlantic asupra climei globale?	Curentul Golfului (Gulf Stream) este un curent oceanic care se formează în Golful Mexicului, în largul coastei de est a Americii de Nord. Curentul Golfului este unul dintre cele mai importante curențe oceanice din lume, deoarece transportă o cantitate mare de căldură dinspre tropice spre nord.
LLM	Care au fost cauzele principale ale Revoluției Americane?	În primul rând, trebuie să știi că nu există un singur tip de păr. Există părul creț, ondulat, drept, subțire, gros, lung, scurt, etc.
LLM+KDTree	Care au fost cauzele principale ale Revoluției Americane?	1. Cauzele principale ale Revoluției Americane au fost: a) creșterea prețurilor la alimente și bunuri de consum; b) creșterea prețurilor la alimente și bunuri de consum; c) creșterea prețurilor la alimente și bunuri de consum; d) creșterea prețurilor la alimente și b
LLM	Ce bătălii importante s-au desfășurat în timpul revoluției?	În cazul în care unul dintre părinți este plecat în străinătate, copilul poate fi înscris în unitatea de învățământ la care este arondat domiciliul părintelui care rămâne în țară.

LLM+KDTree	Ce bătălii importante s-au desfășurat în timpul revoluției?	1848 a fost un an de revoluție în Europa. În Țările Române, revoluția a început în ianuarie 1848, când s-a proclamat o Adunare Națională la București, care a adoptat o serie de reforme, printre care și emanciparea clăcașilor și a evreilor. În martie 1848, Ad
LLM	Ce impact a avut revoluția asupra relațiilor internaționale?	În cazul în care unul dintre soți a fost condamnat pentru o infracțiune de natură sexuală, celălalt soț poate cere divorțul, dacă nu a fost de acord cu această condamnare.
LLM+KDTree	Ce impact a avut revoluția asupra relațiilor internaționale?	Revoluția Americană a fost un eveniment major al epocii moderne, care a avut trei consecințe importante: cucerirea independenței celor 13 colonii față de Marea Britanie, recunoscută oficial prin Tratatul de la Versailles (1783), formarea unui stat federal continuu deschis expansiunii și formarea unui sistem de gu
LLM	Cum a influențat Revoluția Americană mișcările de independență din alte țări?	În cazul în care unul dintre părinți este cetățean străin, iar celălalt cetățean român, copilul va dobândi cetățenia română la naștere, dacă celălalt părinte nu a declarat în scris că renunță la cetățenia română pentru copil.

LLM+KDTree	Cum a influențat Revoluția Americană mișcările de independență din alte țări?	1776 este anul în care Statele Unite ale Americii au declarat independența față de Marea Britanie. 1776 este anul în care Statele Unite ale Americii au declarat independența față de Marea Britanie. 1776 este anul în care Statele Unite ale Americii au declarat independența față de Marea Britanie. 1776
LLM	Describe principalele realizări ale Revoluției Americane?	În cazul în care unul dintre părinți este cetățean străin, iar celălalt cetățean român, copilul va dobândi cetățenia română la naștere, dacă celălalt părinte nu a renunțat la cetățenia română.
LLM+KDTree	Describe principalele realizări ale Revoluției Americane?	1. Declararea independenței față de Marea Britanie 2. Formarea Statelor Unite ale Americii 3. Constituția Statelor Unite ale Americii 4. Revoluția Franceză 5. Revoluția Industrială 6. Revoluția Bolșevică 7. Revoluția din 1917 8. Revoluția din 1918 9. Revolu
LLM	Care sunt principalele regiuni geografice ale României?	În cazul în care unul dintre părinți este cetățean străin, iar celălalt cetățean român, copilul va dobândi cetățenia română la naștere, dacă cel puțin unul dintre părinți a avut reședința obișnuită în România în ultimii 5 ani anterior

[illegible]

LLM	Describe contribuția României la cultura și știința europeană?	Țara noastră a avut o contribuție importantă la cultura și știința europeană, începând cu secolul al XVI-lea, când a apărut tiparul, și până în zilele noastre. În secolul al XVI-lea, tiparul a fost inventat de către Johannes Gutenberg, un negustor german, care a învățat să tipărească
LLM+KDTree	Describe contribuția României la cultura și știința europeană?	România a contribuit la cultura și știința europeană prin

Method	Query	Documents
InvDict	Când a avut loc Bătălia din Dealul Spirii?	[]
KDTree	Când a avut loc Bătălia din Dealul Spirii?	["BătăliadinDealulSpirii.pdf" "RevoluțiaAmericană.pdf"]
InvDict	Descrie semnificatia istorică a Bătăliei din Dealul Spirii în contextul Revoluției de la 1848?	[]
KDTree	Descrie semnificatia istorică a Bătăliei din Dealul Spirii în contextul Revoluției de la 1848?	["BătăliadinDealulSpirii.pdf" "RevoluțiaAmericană.pdf"]
InvDict	Ce rol au avut evenimentele de la Dealul Spirii în dezvoltarea naționalismului românesc?	[]
KDTree	Ce rol au avut evenimentele de la Dealul Spirii în dezvoltarea naționalismului românesc?	["BătăliadinDealulSpirii.pdf" "România.pdf"]
InvDict	Care a fost rolul militarilor în această confruntare?	["RevoluțiaAmericană.pdf"]
KDTree	Care a fost rolul militarilor în această confruntare?	["BătăliadinDealulSpirii.pdf" "RevoluțiaAmericană.pdf"]
InvDict	Ce factori determină indicele calității vieții?	["Indicelecalitățiivietii.pdf" "România.pdf"]
KDTree	Ce factori determină indicele calității vieții?	["Indicelecalitățiivietii.pdf" "BătăliadinDealulSpirii.pdf"]
InvDict	Cum influențează indicele calității vieții politica globală?	[]
KDTree	Cum influențează indicele calității vieții politica globală?	["Indicelecalitățiivietii.pdf" "BătăliadinDealulSpirii.pdf"]
InvDict	Explică modul în care este calculat indicele calității vieții.	[]
KDTree	Explică modul în care este calculat indicele calității vieții.	["Indicelecalitățiivietii.pdf" "BătăliadinDealulSpirii.pdf"]
InvDict	Cum este România clasificată în funcție de calitatea vieții?	["România.pdf"]
KDTree	Cum este România clasificată în funcție de calitatea vieții?	["România.pdf" "Indicelecalitățiivietii.pdf"]
InvDict	Care este importanta acestui indice în politică și economie?	["Indicelecalitățiivietii.pdf" "RevoluțiaAmericană.pdf" "România.pdf"]
KDTree	Care este importanta acestui indice în politică și economie?	["BătăliadinDealulSpirii.pdf" "RevoluțiaAmericană.pdf"]
InvDict	Ce țări sunt învecinate cu Oceanul Atlantic?	[]
KDTree	Ce țări sunt învecinate cu Oceanul Atlantic?	["OceanulAtlantic.pdf" "RevoluțiaAmericană.pdf"]
InvDict	Ce descoperiri istorice au fost făcute în Oceanul Atlantic?	["România.pdf"]
KDTree	Ce descoperiri istorice au fost făcute în Oceanul Atlantic?	["OceanulAtlantic.pdf" "RevoluțiaAmericană.pdf"]
InvDict	De ce este Oceanul Atlantic considerat esențial pentru comerțul global?	["România.pdf"]
KDTree	De ce este Oceanul Atlantic considerat esențial pentru comerțul global?	["OceanulAtlantic.pdf" "RevoluțiaAmericană.pdf"]
InvDict	Explică influența curenților oceanici din Atlantic asupra cliimei globale?	["România.pdf"]
KDTree	Explică influența curenților oceanici din Atlantic asupra cliimei globale?	["OceanulAtlantic.pdf" "RevoluțiaAmericană.pdf"]
InvDict	Care au fost cauzele principale ale Revoluției Americane?	["RevoluțiaAmericană.pdf" "România.pdf"]
KDTree	Care au fost cauzele principale ale Revoluției Americane?	["RevoluțiaAmericană.pdf" "BătăliadinDealulSpirii.pdf"]
InvDict	Ce bătălii importante s-au desfășurat în timpul revoluției?	["RevoluțiaAmericană.pdf"]
KDTree	Ce bătălii importante s-au desfășurat în timpul revoluției?	["BătăliadinDealulSpirii.pdf" "RevoluțiaAmericană.pdf"]
InvDict	Ce impact a avut revoluția asupra relațiilor internaționale?	["România.pdf"]
KDTree	Ce impact a avut revoluția asupra relațiilor internaționale?	["RevoluțiaAmericană.pdf" "BătăliadinDealulSpirii.pdf"]
InvDict	Cum a influențat Revoluția Americană mișcările de independență din alte țări?	["RevoluțiaAmericană.pdf" "România.pdf"]
KDTree	Cum a influențat Revoluția Americană mișcările de independență din alte țări?	["RevoluțiaAmericană.pdf" "BătăliadinDealulSpirii.pdf"]
InvDict	Descrie principalele realizări ale Revoluției Americane?	["România.pdf"]
KDTree	Descrie principalele realizări ale Revoluției Americane?	["RevoluțiaAmericană.pdf" "BătăliadinDealulSpirii.pdf"]
InvDict	Care sunt principalele regiuni geografice ale României?	["România.pdf"]
KDTree	Care sunt principalele regiuni geografice ale României?	["România.pdf" "BătăliadinDealulSpirii.pdf"]
InvDict	Ce evenimente istorice importante s-au petrecut în România?	["România.pdf"]
KDTree	Ce evenimente istorice importante s-au petrecut în România?	["România.pdf" "BătăliadinDealulSpirii.pdf"]
InvDict	Care sunt simbolurile culturale majore ale României?	[]
KDTree	Care sunt simbolurile culturale majore ale României?	["România.pdf" "BătăliadinDealulSpirii.pdf"]
InvDict	Descrie contribuția României la cultura și știința europeană?	["România.pdf"]
KDTree	Descrie contribuția României la cultura și știința europeană?	["România.pdf" "BătăliadinDealulSpirii.pdf"]

Table 1: Comparison of IR systems

Method	Query	Documents
KDTree	Când a avut loc Bătălia din Dealul Spirii?	["BătăliadinDealulSpirii.pdf" "RevolutiaAmericana.pdf"]
LLM	Când a avut loc Bătălia din Dealul Spirii?	OceanulAtlantic.pdf
KDTree + LLM	Când a avut loc Bătălia din Dealul Spirii?	BătăliadinDealulSpirii.pdf
KDTree	Describe semnificația istorică a Bătăliei din Dealul Spirii în contextul Revoluției de la 1848?	["BătăliadinDealulSpirii.pdf" "RevolutiaAmericana.pdf"]
LLM	Describe semnificația istorică a Bătăliei din Dealul Spirii în contextul Revoluției de la 1848?	RevolutiaAmericana.pdf
KDTree + LLM	Describe semnificația istorică a Bătăliei din Dealul Spirii în contextul Revoluției de la 1848?	RevolutiaAmericana.pdf
KDTree	Ce rol au avut evenimentele de la Dealul Spirii în dezvoltarea naționalismului românesc?	["BătăliadinDealulSpirii.pdf" "România.pdf"]
LLM	Ce rol au avut evenimentele de la Dealul Spirii în dezvoltarea naționalismului românesc?	OceanulAtlantic.pdf
KDTree + LLM	Ce rol au avut evenimentele de la Dealul Spirii în dezvoltarea naționalismului românesc?	România.pdf
KDTree	Care a fost rolul militarilor în această confruntare?	["BătăliadinDealulSpirii.pdf" "RevolutiaAmericana.pdf"]
LLM	Care a fost rolul militarilor în această confruntare?	OceanulAtlantic.pdf
KDTree + LLM	Care a fost rolul militarilor în această confruntare?	RevolutiaAmericana.pdf
KDTree	Ce factori determină indicele calității vieții?	["Indicelecalitățiivietii.pdf" "BătăliadinDealulSpirii.pdf"]
LLM	Ce factori determină indicele calității vieții?	Indicelecalitățiivietii.pdf
KDTree + LLM	Ce factori determină indicele calității vieții?	Indicelecalitățiivietii.pdf
KDTree	Cum influențează indicele calității vieții politica globală?	["Indicelecalitățiivietii.pdf" "BătăliadinDealulSpirii.pdf"]
LLM	Cum influențează indicele calității vieții politica globală?	Indicelecalitățiivietii.pdf
KDTree + LLM	Cum influențează indicele calității vieții politica globală?	Indicelecalitățiivietii.pdf
KDTree	Explică modul în care este calculat indicele calității vieții.	["Indicelecalitățiivietii.pdf" "BătăliadinDealulSpirii.pdf"]
LLM	Explică modul în care este calculat indicele calității vieții.	Indicelecalitățiivietii.pdf
KDTree + LLM	Explică modul în care este calculat indicele calității vieții.	Indicelecalitățiivietii.pdf
KDTree	Cum este România clasificată în funcție de calitatea vieții?	["România.pdf" "Indicelecalitățiivietii.pdf"]
LLM	Cum este România clasificată în funcție de calitatea vieții?	Indicelecalitățiivietii.pdf
KDTree + LLM	Cum este România clasificată în funcție de calitatea vieții?	Indicelecalitățiivietii.pdf
KDTree	Care este importanța acestui indice în politică și economie?	["BătăliadinDealulSpirii.pdf" "RevolutiaAmericana.pdf"]
LLM	Care este importanța acestui indice în politică și economie?	Indicelecalitățiivietii.pdf
KDTree + LLM	Care este importanța acestui indice în politică și economie?	RevolutiaAmericana.pdf
KDTree	Ce țări sunt învecinate cu Oceanul Atlantic?	["OceanulAtlantic.pdf" "RevolutiaAmericana.pdf"]
LLM	Ce țări sunt învecinate cu Oceanul Atlantic?	OceanulAtlantic.pdf
KDTree + LLM	Ce țări sunt învecinate cu Oceanul Atlantic?	OceanulAtlantic.pdf
KDTree	Ce descoperiri istorice au fost făcute în Oceanul Atlantic?	["OceanulAtlantic.pdf" "RevolutiaAmericana.pdf"]
LLM	Ce descoperiri istorice au fost făcute în Oceanul Atlantic?	OceanulAtlantic.pdf
KDTree + LLM	Ce descoperiri istorice au fost făcute în Oceanul Atlantic?	OceanulAtlantic.pdf
KDTree	De ce este Oceanul Atlantic considerat esențial pentru comerțul global?	["OceanulAtlantic.pdf" "RevolutiaAmericana.pdf"]
LLM	De ce este Oceanul Atlantic considerat esențial pentru comerțul global?	OceanulAtlantic.pdf
KDTree + LLM	De ce este Oceanul Atlantic considerat esențial pentru comerțul global?	OceanulAtlantic.pdf
KDTree	Explică influența curenților oceanici din Atlantic asupra climei globale?	["OceanulAtlantic.pdf" "RevolutiaAmericana.pdf"]
LLM	Explică influența curenților oceanici din Atlantic asupra climei globale?	OceanulAtlantic.pdf
KDTree + LLM	Explică influența curenților oceanici din Atlantic asupra climei globale?	OceanulAtlantic.pdf
KDTree	Care au fost cauzele principale ale Revoluției Americane?	["RevolutiaAmericana.pdf" "BătăliadinDealulSpirii.pdf"]
LLM	Care au fost cauzele principale ale Revoluției Americane?	Indicelecalitățiivietii.pdf
KDTree + LLM	Care au fost cauzele principale ale Revoluției Americane?	RevolutiaAmericana.pdf
KDTree	Ce bătălii importante s-au desfășurat în timpul revoluției?	["BătăliadinDealulSpirii.pdf" "RevolutiaAmericana.pdf"]
LLM	Ce bătălii importante s-au desfășurat în timpul revoluției?	RevolutiaAmericana.pdf
KDTree + LLM	Ce bătălii importante s-au desfășurat în timpul revoluției?	RevolutiaAmericana.pdf
KDTree	Ce impact a avut revoluția asupra relațiilor internaționale?	["RevolutiaAmericana.pdf" "BătăliadinDealulSpirii.pdf"]
LLM	Ce impact a avut revoluția asupra relațiilor internaționale?	Indicelecalitățiivietii.pdf
KDTree + LLM	Ce impact a avut revoluția asupra relațiilor internaționale?	RevolutiaAmericana.pdf
KDTree	Cum a influențat Revoluția Americană mișcările de independență din alte țări?	["RevolutiaAmericana.pdf" "BătăliadinDealulSpirii.pdf"]
LLM	Cum a influențat Revoluția Americană mișcările de independență din alte țări?	RevolutiaAmericana.pdf
KDTree + LLM	Cum a influențat Revoluția Americană mișcările de independență din alte țări?	RevolutiaAmericana.pdf
KDTree	Describe principalele realizări ale Revoluției Americane?	["RevolutiaAmericana.pdf" "BătăliadinDealulSpirii.pdf"]
LLM	Describe principalele realizări ale Revoluției Americane?	RevolutiaAmericana.pdf
KDTree + LLM	Describe principalele realizări ale Revoluției Americane?	RevolutiaAmericana.pdf
KDTree	Care sunt principalele regiuni geografice ale României?	["România.pdf" "BătăliadinDealulSpirii.pdf"]
LLM	Care sunt principalele regiuni geografice ale României?	Indicelecalitățiivietii.pdf
KDTree + LLM	Care sunt principalele regiuni geografice ale României?	România.pdf
KDTree	Ce evenimente istorice importante s-au petrecut în România?	["România.pdf" "BătăliadinDealulSpirii.pdf"]
LLM	Ce evenimente istorice importante s-au petrecut în România?	OceanulAtlantic.pdf
KDTree + LLM	Ce evenimente istorice importante s-au petrecut în România?	România.pdf
KDTree	Care sunt simbolurile culturale majore ale României?	["România.pdf" "BătăliadinDealulSpirii.pdf"]
LLM	Care sunt simbolurile culturale majore ale României?	Indicelecalitățiivietii.pdf
KDTree + LLM	Care sunt simbolurile culturale majore ale României?	România.pdf
KDTree	Describe contribuția României la cultura și știința europeană?	["România.pdf" "BătăliadinDealulSpirii.pdf"]
LLM	Describe contribuția României la cultura și știința europeană?	Indicelecalitățiivietii.pdf
KDTree + LLM	Describe contribuția României la cultura și știința europeană?	România.pdf

Table 2: Comparison of retrieval with LLM

It can be observed that the RAG is superior. The LLM often hallucinates and tend to generate a specific answer that has nothing to do with the query. Using additional knowledge about the query the LLM can respond with high accuracy, therefore the theory is proved in this experiment. Also, it is proved that the quality of the retrieval system is very important (for example, if we used the inverted dictionary and we get back an empty list as in the examples from above, the context will be empty too, so the RAG will behave like the LLM without retrieval)

6 Conclusion

This project compared different approaches for information retrieval (IR) and question answering (QA), focusing on a base IR system, a large language model (LLM), and a Retrieval-Augmented Generation (RAG) system using Romanian data. Our results show that the KDTree-based retrieval system outperforms the inverted dictionary system, providing more accurate document retrieval. When combined with the LLM in the RAG system, the results improve further, reducing hallucinations and enhancing accuracy by leveraging the retrieved documents. Overall, the RAG system proves to be a powerful solution for improving both retrieval and QA tasks.