

Retrieval-Augmented Generation (RAG): *TravelBuddy in Europe*

Ludwig-Maximilians-Universität, München

Centrum für Informations- und Sprachverarbeitung

Masterseminar: Suchmaschinen und RAG

Dozent: PD Dr. Stefan Langer

Referentin: Evanthia Tsigkana

München, 03. Juli 2024

Übersicht

1. Einführung und Motivation
2. Aufbau und Implementierung des RAG-Systems: *TravelBuddy in Europe*
3. Evaluierung des Retrievers
4. Evaluierung des gesamten RAG-Systems
5. Zusammenfassung

Einführung und Motivation

Warum „TravelBuddy in Europe“?

- Die Reisetendenz steigt deutlich nach der Pandemie.

„Die Reisebranche befindet sich derzeit in einem starken Aufwärtstrend“

Quelle: Tagesschau.de (Link:<https://www.tagesschau.de/wirtschaft/verbraucher/reisetrends-2023-itb-101.html>, letzter Abruf am 01.7.2024)

„Die Buchungen steigen massiv an.“

Quelle: <https://www.zeit.de/entdecken/reisen/2022-02/tourismus-corona-pandemie-urlaub-reiseverhalten-zukunftstrend>, letzter Abruf am 01.07.2024)

- Nur in Europa, aber verwendbar für die ganze Welt

Aufbau und Implementierung des RAG-Systems

- Dokumente in deutscher und englischer Sprache (insgesamt 31) ausgewählt von folgenden Webseiten über Reisen allgemein: Wikivoyage, Lonely Planet, TripAdvisor, Travelocity, Expedia
- Format der Dokumente: html
- Pre-processing: Textextraktion mit „*BeautifulSoup*“ (Mit der in Python geschriebenen Software lassen sich XML- und HTML-Dokumente parsen.)

Aufbau und Implementierung des RAG-Systems

- Zunächst Elastic Search installiert, wurde aber durch LlamaIndex ersetzt.
- Die Kapazität des Rechners war nicht ausreichend, deshalb weiter mit LlamaIndex online.
- Online-Nutzung gegen Bezahlung (\$5) – keine lokale Nutzung
- `os.environ["OPENAI_API_KEY"] = "sk-proj-DwikG6Qsk0rlnchK8czNT3BlbkFJoAgxLhHtFmyLAvg8j2nJ"`
#dieses key ist bezahlt worden

Aufbau und Implementierung des RAG-Systems: Fragen – Antworten

Auswahl von allgemeinen Fragen – Antworten festgesetzt, in englischer und/oder deutscher Sprache:

1. "Where to travel with children?": "The Vulcania theme park in Auvergne is surprisingly educational and entertaining. The Science Museum in London is worth a visit."
2. "Wohin mit Kindern reisen?": "Der Vulcania Themenpark in der Auvergne ist überraschend informativ und unterhaltend. Das Science Museum in London ist einen Besuch wert.",
3. "What is springtime in Paris like?": "Springtime in Paris is a lovely season with Cafe terraces opening and parks and chestnut trees blossoming."
4. "What is the capital of Lithuania?": "Vilnius"
5. "Welche Fremdsprachen werden in Tschechien gesprochen?": "In Tschechien werden Slowakisch, Russisch, Englisch und Deutsch als Fremdsprachen gesprochen."
6. "What are the best beaches?": "Great beaches can be found in the South of France close to Nice and also in Ireland."
7. "What to visit in Hungary?": "The capital Budapest is the political, economic and cultural centre with cosmopolitan flair. Other places worth visiting include Debrecen, Győr, Miskolc, Esztergom and Sopron. There are also events such as the apricot feast and the Gulyas feast."

Aufbau und Implementierung des RAG-Systems: LlamaIndex

- LlamaIndex <https://www.llamaindex.ai/>: Installation
- Erstellen „*VectorStoreIndex*“ → vektorisierter Index der Rohdaten
- Default-Einstellungen verwendet
- Index ist lokal gespeichert und besteht aus mehreren Dateien, daher es braucht viel Zeit zum Öffnen

Aufbau und Implementierung des RAG-Systems: LlamaIndex

- Gespeicherte Dokumente (frühere 31 htmls) in 250 Chunks (Textabschnitte) aufgeteilt
- Diese Chunks, kombiniert mit Metadaten zu Nodes, sind Grundlage für die spätere Suchfunktionen und eine „ideale“ Liste ist erstellt für den späteren Vergleich bei der Berechnung von NDCG
- Chunks weisen einen Überlapp auf → ein folgender Chunk beginnt vor dem Ende des vorherigen Chunks → dieselbe Textstelle in 2 Chunks vorhanden

Aufbau und Implementierung des RAG-Systems: LlamaIndex

- Für jede Frage wird eine Liste von Nodes angelegt → ideale Position und Gewichtung der Nodes für eine bestimmte Frage angelegt
- Nodes mit bedeutender Information für die Frage erhalten Gewicht größer Null.
- Vermeiden die Erstellung von 250 Nodes für jede Frage

Aufbau und Implementierung des RAG-Systems: LlamaRetriever BM25

- LlamaRetriever BM25 : Eingeführt in den 1970ern von Robertson, Jones et al., City University, London, seither verwendet und weiter entwickelt
- Nimmt die Suche an und macht eine Rangliste der Chunks
- Ränge werden mit der BM25 (BM = BestMatch) Funktion berechnet

Aufbau und Implementierung des RAG-Systems: LlamaRetriever BM25

- IDF-gewichteter Bag-of-Words Score auf die Häufigkeit von Suchbegriffen, keine Berücksichtigung der Position eines Begriffes

- Benutzt die Implementierung `from`

```
llama_index.retrievers.bm25 import
```

```
BM25Retriever
```

Evaluierung des Retrievers: NDCG

NDCG

1. "Where to travel with children?": {"**result_answer**": "Head to the surprisingly educative Vulcania theme park in Auvergne if you're traveling with children.", "**reference_answer**": "The Vulcania theme park in Auvergne is surprisingly educational and entertaining. The Science Museum in London is worth a visit."} "**ndcg**": **0.726696042754045**
2. "Wohin mit Kindern reisen?": {"**result_answer**": "Consider traveling to destinations that offer family-friendly activities and attractions, such as amusement parks, nature reserves, interactive museums, or beach resorts. Look for places that cater to children's interests and provide a mix of fun and educational experiences to make the trip enjoyable for the whole family.", "**reference_answer**": "Der Vulcania Themenpark in der Auvergne ist \u00fcberraschend informativ und unterhaltend. Das Science Museum in London ist einen Besuch wert."} "**ndcg**": **0.046683762752523**

Evaluierung des Retrievers:NDCG

3. "What is springtime in Paris like?": {"**result_answer**": "Springtime in Paris is characterized by blooming chestnut trees, blossoming flowers in parks and gardens, and the emergence of cafe terraces. The weather is described as lovely, making it an ideal time to visit the city.", "**reference_answer**": "Springtime in Paris is a lovely season with Cafe terraces opening and parks and chestnut trees blossoming.", "**ndcg**": **0.9629689360508414**}
4. "What is the capital of Lithuania?": {"**result_answer**": "Vilnius", "**reference_answer**": "Vilnius", "**ndcg**": **1.0**}
5. "Welche Fremdsprachen werden in Tschechien gesprochen?": {"**result_answer**": "Slowakisch, Russisch, Englisch und Deutsch werden in Tschechien gesprochen.", "**reference_answer**": "In Tschechien werden Slowakisch, Russisch, Englisch und Deutsch als Fremdsprachen gesprochen.", "**ndcg**": **1.0**}

Evaluierung des Retrievers: NDCG

6. "What are the best beaches?": {"**result_answer**": "Europe's most stunning beaches to explore in 2024.", "**reference_answer**": "Great beaches can be found in the South of France close to Nice and also in Ireland.", "**ndcg**": **0.7038981622489484**}

7. "What to visit in Hungary?": {"**result_answer**": "You can visit various culinary events and festivals in Hungary, such as the Aprikosenfest in Borsod-Aba\u00faj-Zempl\u00e9n, the Lecs\u00f3fesztiv\u00e1l in Balatonkenese, the Gulaschfest in Szolnok, and the Paprikafest in Kalocsa. Additionally, you can explore the different national holidays and cultural celebrations that take place throughout the year in Hungary.", "**reference_answer**": "The capital Budapest is the political, economic and cultural centre with cosmopolitan flair. Other places worth visiting include Debrecen, Gy\u00f6r, Miskolc, Esztergom and Sopron. There are also events such as the apricot feast and the Gulyas feast.", "**ndcg**": **0.0**}

Evaluierung des Retrievers: NDCG

Übersicht der Evaluation mit der Metrik NDCG

1	"ndcg"	0.726696042754045
2	"ndcg"	0.046683762752523
3	"ndcg"	0.9629689360508414
4	"ndcg"	1.0
5	"ndcg"	1.0
6	"ndcg"	0.7038981622489484
7	"ndcg"	0.0

Die Ergebnisse variieren stark je nach Frage. Es geht von perfektes Ergebnis bis 0.0

Evaluierung des RAG-Systems: Rouge1, Rouge4, RougeL

Rouge: Rouge1, Rouge4, RougeL

1. "Where to travel with children?": {"result_answer": "Head to the surprisingly educative Vulcania theme park in Auvergne if you're traveling with children.", "reference_answer": "The Vulcania theme park in Auvergne is surprisingly educational and entertaining. The Science Museum in London is worth a visit." "rouge": {"rouge1": [0.4, 0.5, 0.4444444444444445], "rouge4": [0.11764705882352941, 0.15384615384615385, 0.13333333333333333], "rougeL": [0.3, 0.375, 0.333333333333333326]}}
2. "Wohin mit Kindern reisen?": {"result_answer": "Consider traveling to destinations that offer family-friendly activities and attractions, such as amusement parks, nature reserves, interactive museums, or beach resorts. Look for places that cater to children's interests and provide a mix of fun and educational experiences to make the trip enjoyable for the whole family.", "reference_answer": "Der Vulcania Themenpark in der Auvergne ist \u00fcberraschend informativ und unterhaltend. Das Science Museum in London ist einen Besuch wert.", "rouge": {"rouge1": [0.05, 0.02040816326530612, 0.02898550724637681], "rouge4": [0.0, 0.0, 0.0], "rougeL": [0.05, 0.02040816326530612, 0.02898550724637681]}}

Evaluierung des RAG-Systems: Rouge1, Rouge4, RougeL

3. "What is springtime in Paris like?": {"result_answer": "Springtime in Paris is characterized by blooming chestnut trees, blossoming flowers in parks and gardens, and the emergence of cafe terraces. The weather is described as lovely, making it an ideal time to visit the city.", "reference_answer": "Springtime in Paris is a lovely season with Cafe terraces opening and parks and chestnut trees blossoming.", "rouge": {"rouge1": [0.7647058823529411, 0.3611111111111111, 0.490566037735849], "rouge4": [0.07142857142857142, 0.030303030303030304, 0.0425531914893617], "rougeL": [0.4117647058823529, 0.19444444444444445, 0.2641509433962264]}}
4. What is the capital of Lithuania?": {"result_answer": "Vilnius", "reference_answer": "Vilnius", "rouge": {"rouge1": [1.0, 1.0, 1.0], "rouge4": [0.0, 0.0, 0.0], "rougeL": [1.0, 1.0, 1.0]}}
5. Welche Fremdsprachen werden in Tschechien gesprochen?": {"result_answer": "Slowakisch, Russisch, Englisch und Deutsch werden in Tschechien gesprochen.", "reference_answer": "In Tschechien werden Slowakisch, Russisch, Englisch und Deutsch als Fremdsprachen gesprochen.", "rouge": {"rouge1": [0.8181818181818182, 1.0, 0.9], "rouge4": [0.25, 0.3333333333333333, 0.28571428571428575], "rougeL": [0.5454545454545454, 0.6666666666666666, 0.6]}}

Evaluierung des RAG-Systems: Rouge1, Rouge4, RougeL

6. "What are the best beaches?": {"result_answer": "Europe's most stunning beaches to explore in 2024.", "reference_answer": "Great beaches can be found in the South of France close to Nice and also in Ireland.", "rouge": {"rouge1": [0.17647058823529413, 0.3333333333333333, 0.23076923076923078], "rouge4": [0.0, 0.0, 0.0], "rougeL": [0.17647058823529413, 0.3333333333333333, 0.23076923076923078]}}

7. "What to visit in Hungary?": {"result_answer": "You can visit various culinary events and festivals in Hungary, such as the Aprikosenfest in Borsod-Aba\u00faj-Zempl\u00e9n, the Lecs\u00f3fesztiv\u00e1l in Balatonkenese, the Gulaschfest in Szolnok, and the Paprikafest in Kalocsa. Additionally, you can explore the different national holidays and cultural celebrations that take place throughout the year in Hungary.", "reference_answer": "The capital Budapest is the political, economic and cultural centre with cosmopolitan flair. Other places worth visiting include Debrecen, Gy\u00f6r, Miskolc, Esztergom and Sopron. There are also events such as the apricot feast and the Gulyas feast.", "rouge": {"rouge1": [0.34210526315789475, 0.24074074074074073, 0.2826086956521739], "rouge4": [0.0, 0.0, 0.0], "rougeL": [0.18421052631578946, 0.12962962962962962, 0.15217391304347827]}}

Evaluierung des RAG-Systems: Rouge1, Rouge4, RougeL

Übersicht der Evaluierung mit Rouge

1. "rouge": {"rouge1": [0.4, 0.5, 0.44444444444444445], "rouge4": [0.11764705882352941, 0.15384615384615385, 0.13333333333333333], "rougeL": [0.3, 0.375, 0.333333333333333326]}
2. "rouge": {"rouge1": [0.05, 0.02040816326530612, 0.02898550724637681], "rouge4": [0.0, 0.0, 0.0], "rougeL": [0.05, 0.02040816326530612, 0.02898550724637681]}
3. "rouge": {"rouge1": [0.7647058823529411, 0.3611111111111111, 0.490566037735849], "rouge4": [0.07142857142857142, 0.030303030303030304, 0.0425531914893617], "rougeL": [0.4117647058823529, 0.19444444444444445, 0.2641509433962264]}
4. {"rouge1": [1.0, 1.0, 1.0], "rouge4": [0.0, 0.0, 0.0], "rougeL": [1.0, 1.0, 1.0]}
5. "rouge": {"rouge1": [0.8181818181818182, 1.0, 0.9], "rouge4": [0.25, 0.3333333333333333, 0.28571428571428575], "rougeL": [0.5454545454545454, 0.6666666666666666, 0.6]}
6. "rouge": {"rouge1": [0.17647058823529413, 0.3333333333333333, 0.23076923076923078], "rouge4": [0.0, 0.0, 0.0], "rougeL": [0.17647058823529413, 0.3333333333333333, 0.23076923076923078]}
7. "rouge": {"rouge1": [0.34210526315789475, 0.24074074074074073, 0.2826086956521739], "rouge4": [0.0, 0.0, 0.0], "rougeL": [0.18421052631578946, 0.12962962962962962, 0.15217391304347827]}

Evaluierung des RAG-Systems: Bleu

Bleu

1. "Where to travel with children?": {"**result_answer**": "Head to the surprisingly educative Vulcania theme park in Auvergne if you're traveling with children.", "**reference_answer**": "The Vulcania theme park in Auvergne is surprisingly educational and entertaining. The Science Museum in London is worth a visit.", "**bleu**": 0.18449721823522283}
2. "Wohin mit Kindern reisen?": {"**result_answer**": "Consider traveling to destinations that offer family-friendly activities and attractions, such as amusement parks, nature reserves, interactive museums, or beach resorts. Look for places that cater to children's interests and provide a mix of fun and educational experiences to make the trip enjoyable for the whole family.", "**reference_answer**": "Der Vulcania Themenpark in der Auvergne ist \u00fcberraschend informativ und unterhaltend. Das Science Museum in London ist einen Besuch wert.", "**bleu**": 0}

Evaluierung des RAG-Systems: Bleu

3. "What is springtime in Paris like?": { **"result_answer"**: "Springtime in Paris is characterized by blooming chestnut trees, blossoming flowers in parks and gardens, and the emergence of cafe terraces. The weather is described as lovely, making it an ideal time to visit the city.", **"reference_answer"**: "Springtime in Paris is a lovely season with Cafe terraces opening and parks and chestnut trees blossoming.",

"bleu": 0.08202649282493453 }

4. "What is the capital of Lithuania?": { **"result_answer"**: "Vilnius", **"reference_answer"**: "Vilnius",

"bleu": 1.821831989445342e-231 }

5. "Welche Fremdsprachen werden in Tschechien gesprochen?": { **"result_answer"**: "Slowakisch, Russisch, Englisch und Deutsch werden in Tschechien gesprochen.", **"reference_answer"**: "In Tschechien werden Slowakisch, Russisch, Englisch und Deutsch als Fremdsprachen gesprochen.",

"bleu": 0.40194809583392677 }

Evaluierung des RAG-Systems: Bleu

6. "What are the best beaches?": { "**result_answer**": "Europe's most stunning beaches to explore in 2024.", "**reference_answer**": "Great beaches can be found in the South of France close to Nice and also in Ireland.", "**bleu**": 4.6284422396541125e-232 }

7. "What to visit in Hungary?": { "**result_answer**": "You can visit various culinary events and festivals in Hungary, such as the Aprikosenfest in Borsod-Aba\u00faj-Zempl\u00e9n, the Lecs\u00f3fesztiv\u00e1l in Balatonkenese, the Gulaschfest in Szolnok, and the Paprikafest in Kalocsa. Additionally, you can explore the different national holidays and cultural celebrations that take place throughout the year in Hungary.", "**reference_answer**": "The capital Budapest is the political, economic and cultural centre with cosmopolitan flair. Other places worth visiting include Debrecen, Gy\u00f6r, Miskolc, Esztergom and Sopron. There are also events such as the apricot feast and the Gulyas feast.", "**bleu**": 1.7113028512335047e-78 }

Evaluierung des RAG-Systems: Bleu

Übersicht der Evaluierung mit Bleu

1	"bleu"	0.18449721823522283
2	"bleu"	0
3	"bleu"	0.08202649282493453
4	"bleu"	1.821831989445342e-231
5	"bleu"	0.40194809583392677
6	"bleu"	4.6284422396541125e-232
7	"bleu"	1.7113028512335047e-78

Evaluierung des RAG-Systems: Meteor

Meteor

1. "Where to travel with children?": {"**result_answer**": "Head to the surprisingly educative Vulcania theme park in Auvergne if you're traveling with children.", "**reference_answer**": "The Vulcania theme park in Auvergne is surprisingly educational and entertaining. The Science Museum in London is worth a visit.", "**meteor**": 0.3565891472868217}
2. "Wohin mit Kindern reisen?": {"**result_answer**": "Consider traveling to destinations that offer family-friendly activities and attractions, such as amusement parks, nature reserves, interactive museums, or beach resorts. Look for places that cater to children's interests and provide a mix of fun and educational experiences to make the trip enjoyable for the whole family.", "**reference_answer**": "Der Vulcania Themenpark in der Auvergne ist \u00fcberraschend informativ und unterhaltend. Das Science Museum in London ist einen Besuch wert.", "**meteor**": 0.05952380952380953}

Evaluierung des RAG-Systems: Meteor

3. "What is springtime in Paris like?": { "**result_answer**": "Springtime in Paris is characterized by blooming chestnut trees, blossoming flowers in parks and gardens, and the emergence of cafe terraces. The weather is described as lovely, making it an ideal time to visit the city.", "**reference_answer**": "Springtime in Paris is a lovely season with Cafe terraces opening and parks and chestnut trees blossoming.", "**meteor**": **0.5497536945812808** }
4. "What is the capital of Lithuania?": { "**result_answer**": "Vilnius", "**reference_answer**": "Vilnius", "**meteor**": **0.5** }
5. "Welche Fremdsprachen werden in Tschechien gesprochen?": { "**result_answer**": "Slowakisch, Russisch, Englisch und Deutsch werden in Tschechien gesprochen.", "**reference_answer**": "In Tschechien werden Slowakisch, Russisch, Englisch und Deutsch als Fremdsprachen gesprochen.", "**meteor**": **0.8534621578099838** }

Evaluierung des RAG-Systems: Meteor

6. "What are the best beaches?": {"**result_answer**": "Europe's most stunning beaches to explore in 2024.", "**reference_answer**": "Great beaches can be found in the South of France close to Nice and also in Ireland.", "**meteor**": **0.11627906976744186**}

7. "What to visit in Hungary?": {"**result_answer**": "You can visit various culinary events and festivals in Hungary, such as the Aprikosenfest in Borsod-Aba\u00faj-Zempl\u00e9n, the Lecs\u00f3fesztiv\u00e1l in Balatonkenese, the Gulaschfest in Szolnok, and the Paprikafest in Kalocsa. Additionally, you can explore the different national holidays and cultural celebrations that take place throughout the year in Hungary.", "**reference_answer**": "The capital Budapest is the political, economic and cultural centre with cosmopolitan flair. Other places worth visiting include Debrecen, Gy\u00f6r, Miskolc, Esztergom and Sopron. There are also events such as the apricot feast and the Gulyas feast.", "**meteor**": **0.24218265350621276**}

Evaluierung des RAG-Systems: Meteor

Übersicht der Evaluierung mit Meteor

1	"meteor"	0.3565891472868217
2	"meteor"	0.05952380952380953
3	"meteor"	0.5497536945812808
4	"meteor"	0.5
5	"meteor"	0.8534621578099838
6	"meteor"	0.11627906976744186
7	"meteor"	0.24218265350621276

Zusammenfassung

- TravelBuddy in Europe ist ein vollständiges RAG-System
- Größeren Aufwand verursachte die Installation des Modells, insbesondere des Retrievers, obwohl er open source und die Nutzung zum Teil gegen Bezahlung erfolgt ist.
- Die Zusammenstellung geeigneter Komponenten und die Evaluierung des Retrievers mit NDCG war der schwierigste Teil.
- Die Sprachgrenze bereitet Schwierigkeiten.
- Die Bleu-Metrik ist eher geeignet zur Auswertung längerer Passagen.

Quellen (Auswahl)

- <https://colab.research.google.com/drive/1ByiIaBqCwbH9QXJOQWqOfUdsq4LEFq-g?usp=sharing>
- https://docs.llamaindex.ai/en/stable/module_guides/indexing/vector_store_index/
- https://plainenglish.io/community/evaluating-nlp-models-a-comprehensive-guide-to-rouge-bleu-meteor-and-bertscore-metrics-d0f1b1#google_vignette
- <https://llama.meta.com/llama-downloads/>
- <https://arize.com/blog/ragas-how-to-evaluate-rag-pipeline-phoenix>
- <https://docs.llamaindex.ai/en/stable/understanding/>
- <https://ai.meta.com/blog/5-steps-to-getting-started-with-llama-2/>

Vielen Dank für eure Aufmerksamkeit!