



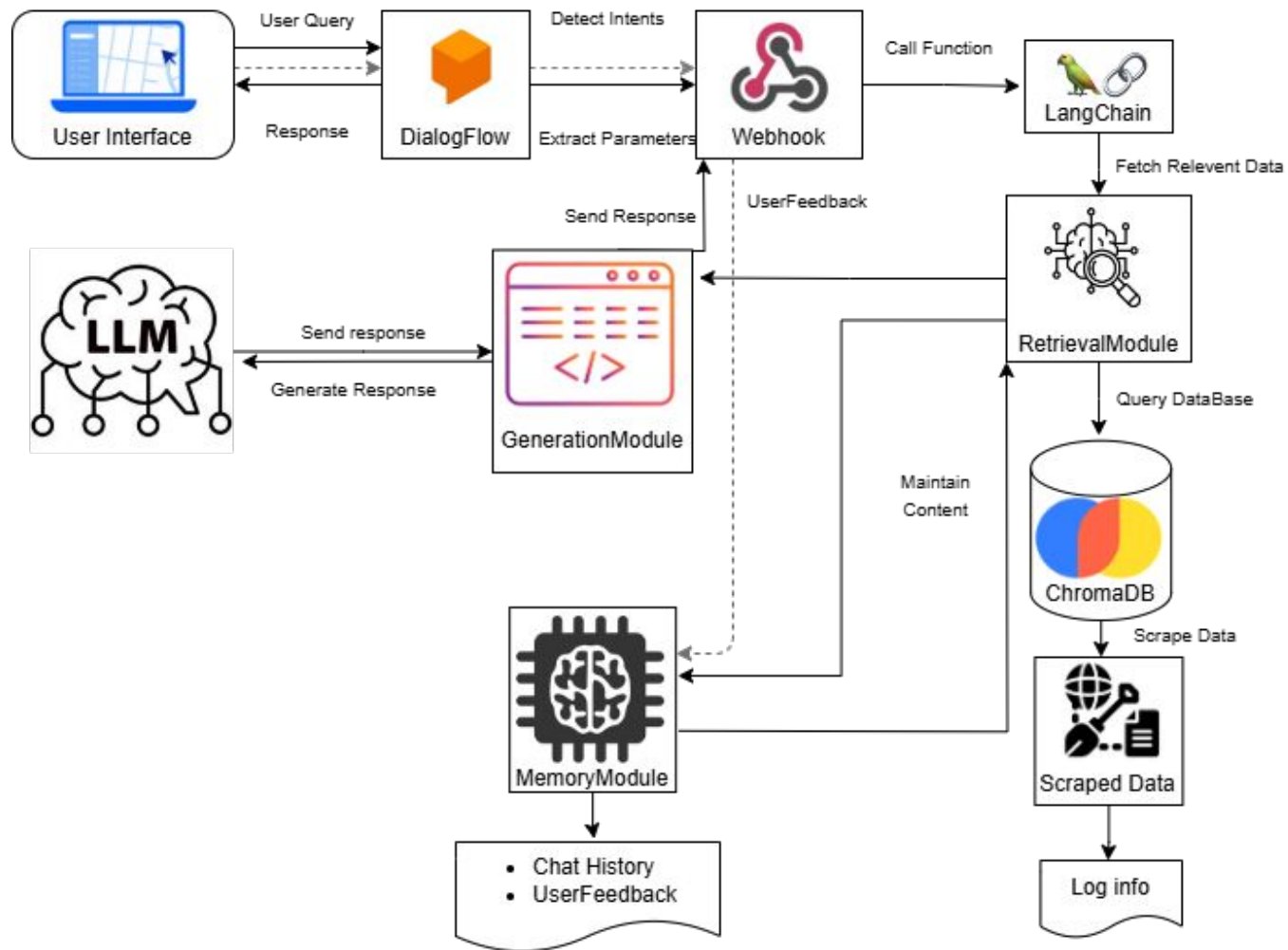
Retriever Bot

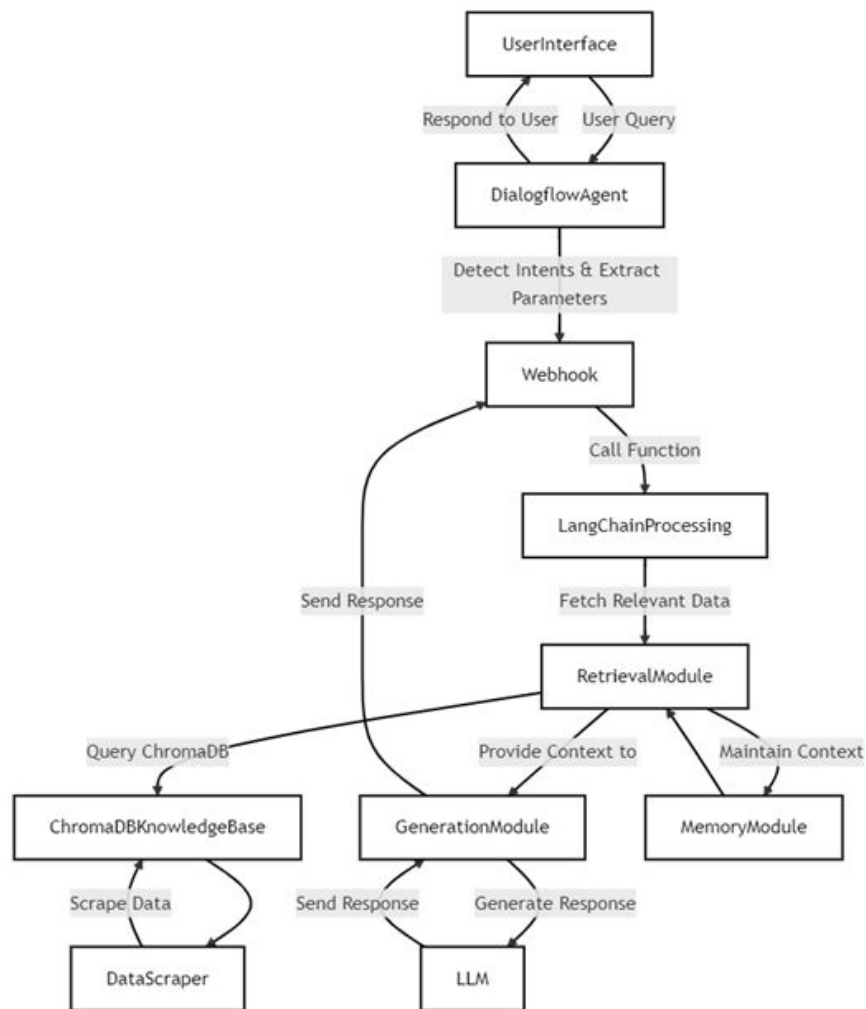


Presented by: Almas Fathimah, Surya Tejaswi Yerramsetty, Yugala Kakarla

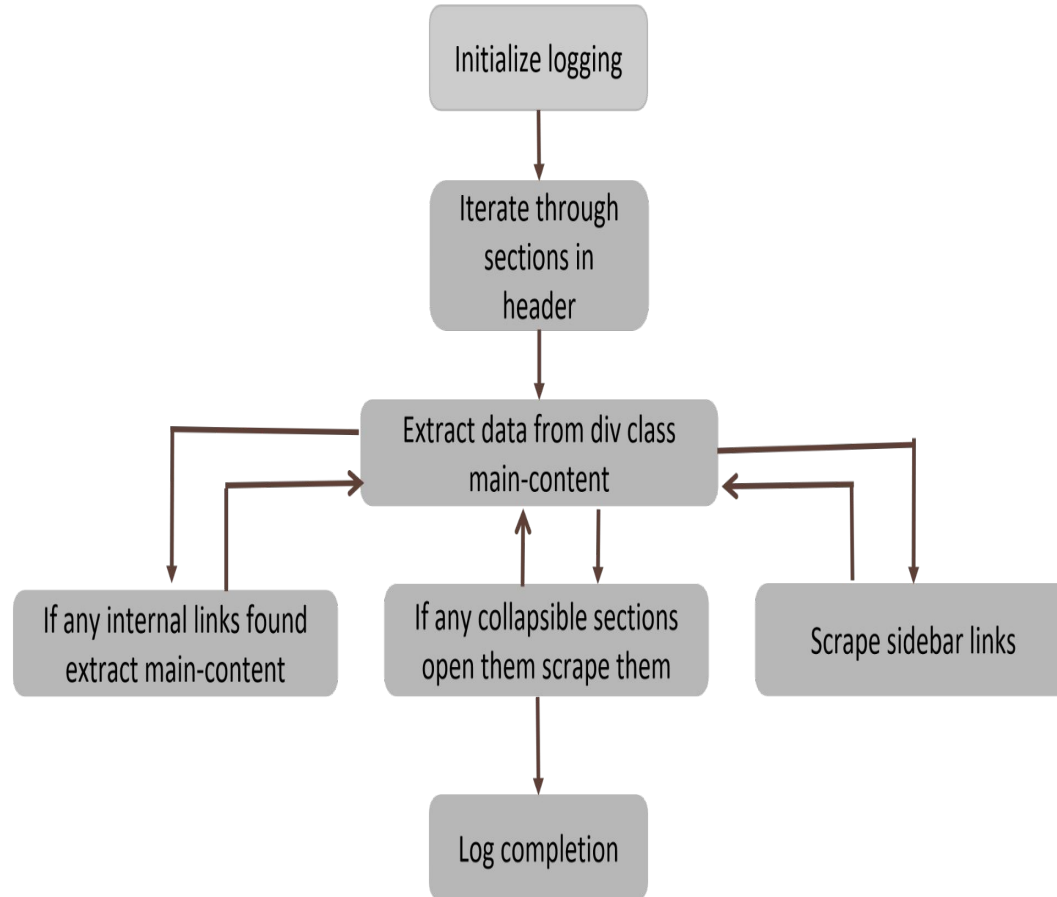
PROBLEM STATEMENT

- Students face difficulties accessing accurate and timely information about coursework, faculty, and registration processes in the UMBC Data Science program.
- Traditional support methods (emails and calls) are inefficient and time-consuming.
- International students require precise guidance for complex processes like CPT and OPT registration.
- The lack of an efficient, interactive, and domain-specific solution hinders user satisfaction and operational efficiency.
- There is a need for an AI-powered chatbot to provide instant, personalized, and context-aware responses, reducing administrative workloads and improving the student experience.





DATA EXTRACTION PIPELINE



Automating tasks with Jupyter Notebook Scheduler

Why Scheduling? The content in the website may change periodically but requires most recent information in order to provide answer.

Jupyter Notebook Job: Using Jupyter Lab's Scheduler to run scraping tasks at set intervals, like every Sunday at 12 AM.

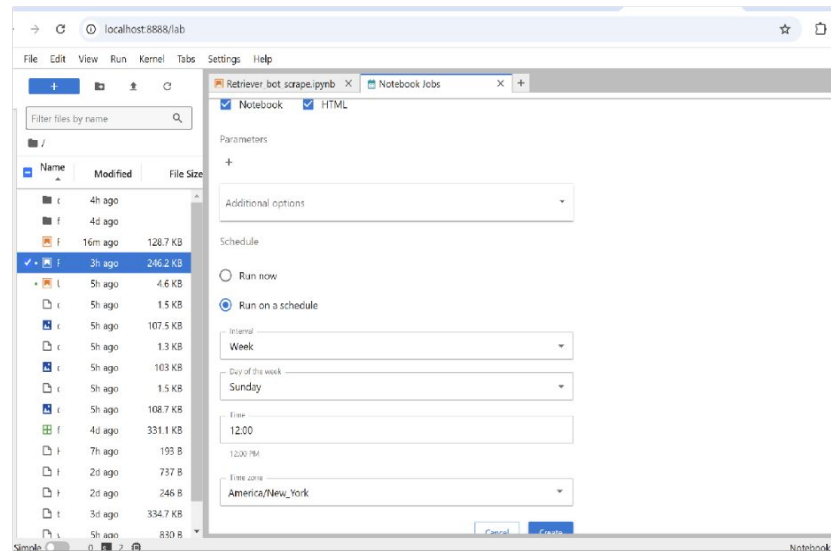
Flexibility: Add parameters for handling multiple websites or datasets. (param like interval,time-zone)

Logs & Monitoring: Access job history and logs to verify execution or debug issues.

Key Benefits:

Consistency: Ensures timely and automated data collection.

Simplicity: Manage scheduling and execution within Jupyter Lab.



Text Splitter

- Chunk Size: 1000 characters for manageable text segments.
- Chunk Overlap: 100 characters to maintain context across chunks.
- Chunk Content: Includes actual page content and metadata
- Meta Data

Link: Source URL.

Section: Document section.

Title: Content title.

Advantages:

Retains context and flow.

Enables efficient content organization with metadata.

Suitable for scalable applications like information retrieval and summarization.

Splitter

Recursive character textsplitter

Chunk Size

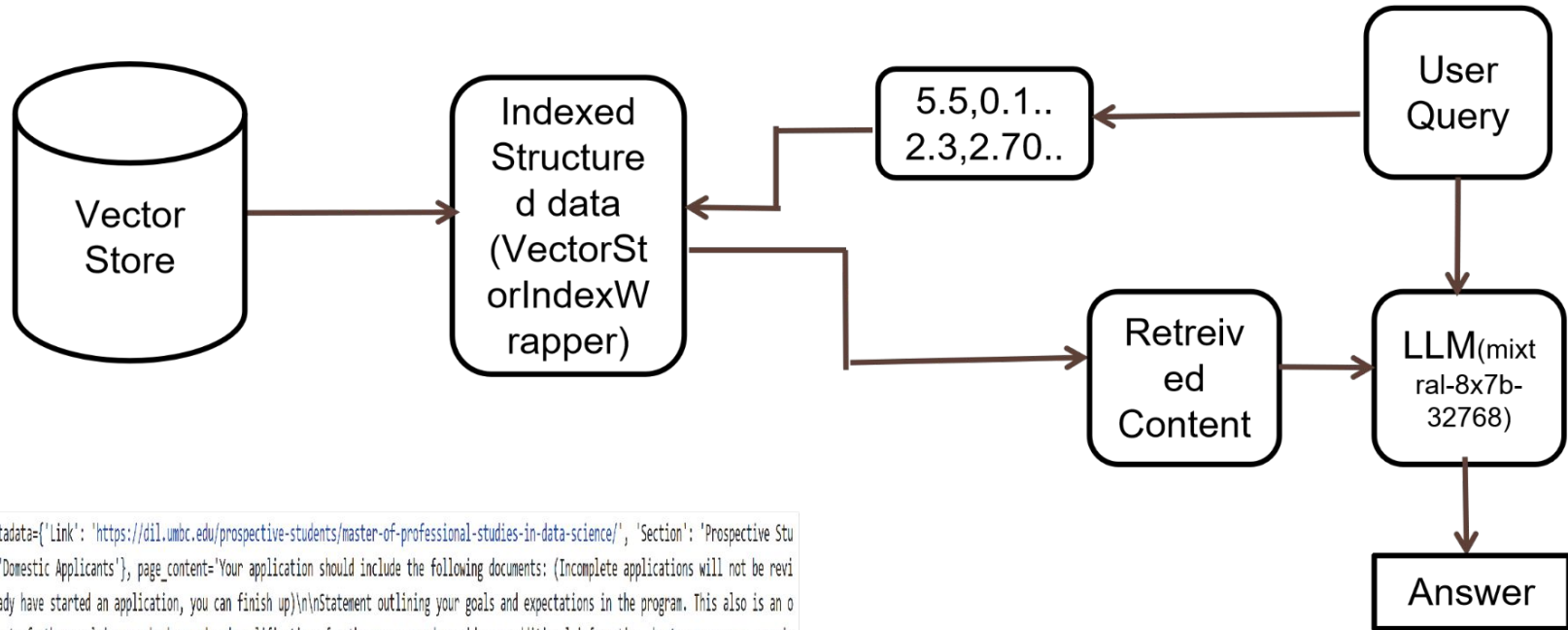
1000

Chunk Overlap

100

Chunk1 1-1000 [Pg Content] [Meta-Data: Section: Link: Title:]	Chunk2 900-1900 [Pg Content] [Meta-Data: Section: Link: Title:]	Chunk3 1800-2800 [Pg Content] [Meta-Data: Section: Link: Title:]	Chunk4 2700-3700 [Pg Content] [Meta-Data: Section: Link: Title:]
---	---	--	--

Information Retrieval using Vector Stores



;', Document(metadata={'Link': 'https://dil.umbc.edu/prospective-students/master-of-professional-studies-in-data-science/', 'Section': 'Prospective Students', 'Title': 'Domestic Applicants'}, page_content='Your application should include the following documents: (Incomplete applications will not be reviewed. If you already have started an application, you can finish up)\n\nStatement outlining your goals and expectations in the program. This also is an opportunity for you to further explain your background and qualifications for the program and provide any additional information about you or your experience which will help with the evaluation of your application.\n\nCurrent resume (When applying online, upload your resume to the graduate school application.)\n\nTranscript(s) from each college university attended\n\nLocal Students: Completed application including residency form (Residency form determines in-state tuition eligibility)\n\nInternational Students: TOEFL, IELTS or PTE test score\n\nNon-refundable application fee (50 online)\n\nHow do I apply?\n\nBy submitting an online application. Here are the specific steps:'); Document(metadata={'Link': 'https://dil.umbc.edu/prospective-students/international-student

```
[22]: chroma_index.query("How early should i begin my opt application process",llm=llm)
```

```
[22]: "Based on the information provided, you should begin the OPT application process by requesting the OPT I-20 from the ISSS office early. After receiving the OPT I-20, you can then submit your complete OPT application to USCIS. It's important to note that OPT applications must be submitted to USCIS no earlier than 90 days before your program end date and no later than 60 days after your program end date. Therefore, you should aim to have your OPT I-20 ready and your application complete within this timeframe to ensure that your application is not denied. Working with the ISSS office early in the process will help you prepare and submit your application on time."
```



Retriever_Bot

POWERED BY  Dialogflow


How can I apply for CPT?

To apply for Curricular Practical Training (CPT), you will need to log in to your ISSS portal, click on "Search Forms," and select the relevant request - CPT. Make sure to read and follow all directions provided in the application. Additionally, CPT often requires enrollment in the Career Center's "Internship Success Practicum (PRAC)." You must upload the email confirmation of your enrollment to meet this requirement. Remember not to submit your application until your PRAC is approved and you are able to upload the approval email. Your academic advisor's support is also

Ask something...



Retriever_Bot

POWERED BY  Dialogflow

what are different pathways I can choose as a Data Science student?

As a Data Science student, you can choose from various pathways such as Cybersecurity, Economics/Econometrics, Healthcare Analytics, Policy Analysis, Project Management, Management Sciences, Aging Studies, Advanced Computing and Analytics, Bioinformatics, and Clinical Informatics. Each pathway offers specialized courses tailored to specific industries or fields of study. You can find more information about these pathways and the courses offered in each pathway on our pathways page. Was this

Ask something...



Retriever_Bot

POWERED BY  Dialogflow

what is the syllabus for Data 602?

| Week | Topic Description | |
—| 1 | Introduction to machine learning and data analysis | 2 | Supervised vs. unsupervised learning | 3 | Decision trees and logistic regression | 4 | Linear discriminant analysis | 5 | Linear and non-linear regression | 6 | Support vector machines | 7 | Neural networks introduction | 8 | Bagging, Boosting, Ensemble Methods, and Random Forests | 9 | Experiment Design, Model Selection, Productizing Models | 10 | Unsupervised learning: agglomerative, divisive, k-means, DBSCAN | 11 | Dimensionality reduction, regularization, PCA, t-SNE, Deep Learning

Ask something...



Retriever_Bot

POWERED BY  Dialogflow

Networks | 13 | Introduction to Reinforcement Learning | 14 | Introduction to Reinforcement Learning | 15 | Final Exam and Project presentations |

any prerequisite I should take for this?

Yes, if you are familiar with data analysis and machine learning or if you have an undergraduate degree in Computer Science or Information Systems, you can take DATA 602 concurrently with DATA 601. However, if you do not meet these criteria, it is recommended to take DATA 602 after completing DATA 601.

Ask something...



Raw interaction log

0% Select Administrator: Command Prompt - ngrok http 5000

ngrok

Sign up to try new private endpoints <https://ngrok.com/new-features-update?ref=private>

Session Status

online

Account syerram1@umbc.edu (Plan: Free)

Version 3.18.4

Region United States (us)

Latency 19ms

Web Interface <http://127.0.0.1:4040>

Forwarding [\[redacted\]](#) -> [http://\[redacted\]](#)

Connections

t1	opn	rt1	rt5	p50	p90
99	0	0.00	0.00	5.03	9.01

HTTP Requests

16:16:59.090 EST POST /webhook 200 OK

16:16:31.013 EST POST /webhook 200 OK

16:13:12.233 EST POST /webhook

Developed a feedback system where user feedback is tied to specific conversation contexts (intent and chat history) and stored in a JSON file to analyze and improve the bot's performance over time based on user interactions.

Feedback Prompting:

Prompts the user for feedback using affirmative or negative responses (e.g., "Yes" or "No")

- when the user's intent changes (e.g., switching from "Get_course_info" to "Get_Research_info").

- If the intent remains the same, feedback is triggered after 5 interactions to assess the quality of responses for prolonged conversations.

Handling Feedback Responses:

- If the user provides feedback:

- Saves the feedback and conversation history

```
context = f'You are a helpful assistant that organizes syllabus information into a clear and concise tabular format. Based on  
Construct the prompt for the LLM  
prompt = f'''  
You are a helpful assistant that organizes syllabus information into a clear and concise tabular format. Based on  
  
Context: {context}  
Question: {query}  
Format the answer as follows:  
| Week | Topic Description |  
|-----|-----|  
| 1 | Topic details |  
| 2 | Topic details |  
...  
Ensure that:  
1. The information is concise and well-organized.  
2. If a course does not have sufficient details in the context, indicate it with "Details not available."  
Answer:  
'''
```

```
elif intent_name == "Get_CPT_OPT_info":  
    prompt = f'''You need to answer the user's question about CPT or OPT information.  
Give friendly responses.  
Use the conversation history to provide a consistent response.  
Conversation History: {chat_history}  
Context: {context}  
Question: {question}'''
```

Conclusion

How effective is the Retriever Bot in improving access to specific information about coursework, faculty, and other program details, as measured by user satisfaction and reduced information search time?

- Reduced information search time significantly by leveraging a retrieval-augmented generation (RAG) system.
- User satisfaction was enhanced, as evidenced by positive feedback on the clarity and accuracy of responses.
- Integration with ChromaDB allowed precise filtering and retrieval based on metadata, ensuring contextually relevant answers.

How can conversational AI enhance CPT and OPT registration guidance for international students, and how helpful will it be in improving user satisfaction?

- The bot provides accurate, multi-step guidance for CPT and OPT registration, covering timelines, eligibility, and documentation.
- Context-aware interactions were enabled by LangChain's memory integration, which allowed users to ask follow-up questions seamlessly.

How accurately can the chatbot respond to diverse student inquiries using up-to-date collected data from the UMBC Data Science website, and how well does this meet students' needs?

Regular web scraping ensures the knowledge base remains current and relevant.

Contextual embeddings generated using state-of-the-art models (e.g., GPT-3.5 with RAG) resulted in high accuracy for user queries.

Extended Achievements

- **FEEDBACK**

- Collecting user feedback ensures that the bot is aligned with user needs, leading to a more engaging and satisfying experience.
- Timely prompts for feedback after specific intents or prolonged interactions ensure the bot remains user-focused.
- Feedback helps refine conversation flow, improve intent recognition. It informs adjustments to the generation model or knowledge base as needed.

- **Meta Data Filtering**

- This ensures that the bot retrieves only the most relevant information for a query. Improving the precision
- Filters out irrelevant or redundant data, focusing only on what matches the user's query intent. Improves Contextual Relevance.

- **Research Lab Centers**

- Addition of this entity expands the bot's capability in answering the research areas of faculty.