



Bibliothèque nationale
du Luxembourg

chat.eluxemburgensia.lu - AI chatbot

Yves Maurer

Division Informatique et Innovation Numérique
Bibliothèque nationale du Luxembourg



Agenda

- AI @ BnL
- Mini Demo
- From Idea to Prototype
- Implementation
- Production and Communication
- Demo

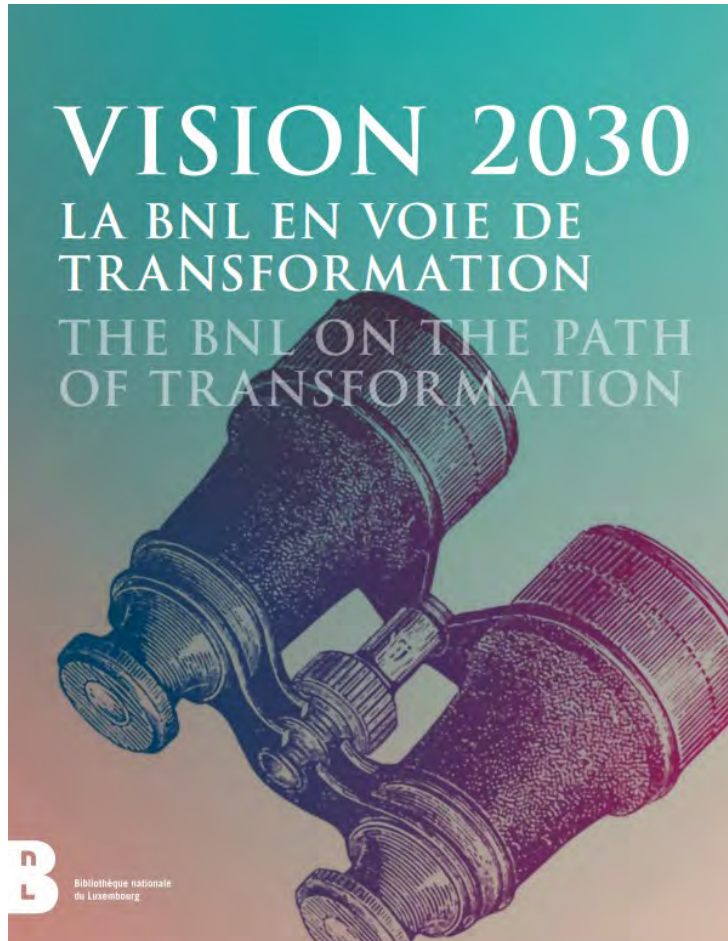


Bibliothèque nationale
du Luxembourg

AI @BnL



Introduction



STRATEGIC PRIORITIES

WE DIGITISE HERITAGE COLLECTIONS

Planned end: 2030

WE ENSURE EASY ACCESS TO THE COLLECTIONS AND SERVICES AND ADAPT THEM TO NEW NEEDS

AI Projects

WE PURSUE STRATEGIC NATIONAL AND INTERNATIONAL PARTNERSHIPS

Regular exchange with other national libraires



Discussion around AI is multifaceted

- **Advancements and Benefits**
AI has made significant technological advancements, enabling automation, personalization, and efficiency in various domains
- **Challenges and Concerns:**
The rapid development of AI has raised ethical, societal, and legal concerns
- **Ethics and Bias**
Ethical considerations in AI are paramount. Biases in AI systems can lead to discrimination, reinforcing existing inequalities.
- **Regulation and Governance**
Governments and institutions are developing regulations to address AI's challenges, including data protection, transparency, and accountability.
- **AI and Creativity**
AI systems can generate creative content, such as art, music, and literature. The debate centers on whether AI can truly be creative and how it impacts human creativity.

-> Potential and concerns . BnL wants to take part in the discussion



Historique

- Quel était le premier... 09/10/2023 13:46
- Quel était le premier... 09/10/2023 13:47
- Même ce qui s'est pa... 09/10/2023 13:48
- Le Luxembourg avait... 09/10/2023 13:49
- Félicité une liste de... 09/10/2023 13:49
- When was the National... 09/10/2023 13:49

La Luxembourg avait-il un projet de construction d'un réacteur nucléaire ?

Selon les articles de journaux locaux, le Luxembourg avait effectivement un projet de construction d'un réacteur nucléaire à Remerschen. Les articles soulignent à plusieurs reprises la possibilité de construire une centrale nucléaire dans la région de Gilscheldersweier (GILSHELDERSWEIER), une entité allemande contrôlée par l'Allemagne. Les articles soulignent également les inquiétudes concernant les risques et les désavantages associés à un tel projet, tels que les problèmes de radiation, l'utilisation de l'eau de la Moselle pour le refroidissement, et la dépendance vis-à-vis de fournisseurs étrangers.

Ma réponse se base sur les articles suivants:

- Luxemburger Wort 41 - 2013-02-09
- Luxemburger Land (f) 41 04 - 2014-07-22
- Luxemburger Land (f) 41 - 2014-07-22
- Kladder (De) 41 - 2014-08-06
- Luxemburger Wort 46 - 2015-06-24
- Luxemburger Wort 47 - 2016-01-27
- Luxemburger Land (f) 48 - 2016-09-24
- Luxemburger Wort 49 - 2016-12-21
- Luxemburger Land (f) 418 - 2017-02-24
- Luxemburger Wort 412 - 2017-11-24
- Luxemburger Wort 412 - 2017-11-24

Écrivez un message en français, allemand ou luxembourgeois. (Wordcount: 23/100)



Mini Demo

B eLuxemburgensia Home Search Chat Rights About Account Sign out

+ New chat

Chat History

- Did Luxembourg have a ...
09/11/2023 11:46
- Wann war der letzte Pe...
25/10/2023 15:42
- Please ignore the stat...
06/10/2023 11:14
- Please write me a pyth...
06/10/2023 11:12
- When and why was the f...
04/10/2023 17:16
- Wofür ist die Rosenzuc...
04/10/2023 15:34
- Quand et pourquoi le d...
04/10/2023 15:30
- Was ist unter dem Maul

Did Luxembourg have a project to build a nuclear reactor?

B Yes, based on the information provided by the newspaper articles, Luxembourg had a project to build a nuclear reactor. This project was discussed and debated in the Luxembourg government and among the public. There were concerns and debates about the risks, disadvantages, and potential impacts of building a nuclear reactor in Luxembourg. However, there is no clear indication in the articles whether the project was actually implemented or not.

My answer is based on these articles:

- Luxemburger Wort #1 • 1973-02-06
- Lëtzebuenger Land (d') #2 #4 • 1974-07-12
- Lëtzebuenger Land (d') #3 • 1974-07-12
- Kéisécker (De) #5 • 1974-10-01
- Luxemburger Wort #6 • 1975-06-28
- Luxemburger Wort #7 • 1976-01-17

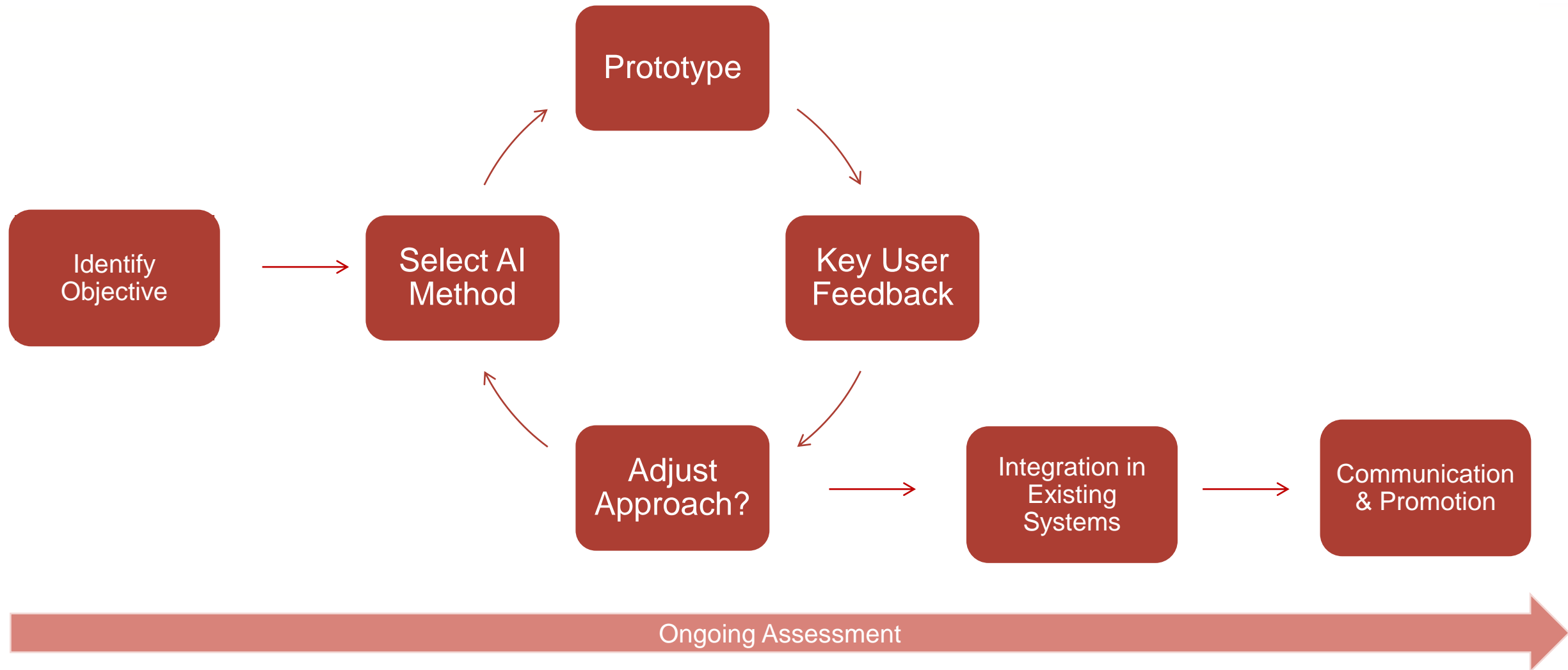
Send a message in French, German, English or Luxembourgish. (Shortcut: ENTER)



Bibliothèque nationale
du Luxembourg

Idea to Prototype

B General AI Project Approach





Chat: Idea to Prototype

Print subscriptions Sign in Search jobs Search Europe edition

Support Guardian Europe
Fund independent journalism with €5 per month
Support us →

The Guardian

News Opinion Sport Culture Lifestyle More

World UK Climate crisis Environment Science Global development Football Tech Business Obituaries

Chatbots

ChatGPT reaches 100 million users two months after launch

Unprecedented take-up may make AI chatbot the fastest-growing consumer internet app ever, analysts say

Dan Milmo and agency
Thu 2 Feb 2023 21.46 CET

f t e



Advertisement

QuickBooks for Small Business
Automate your workflow and see data in real-time





Chat Prototype

- **March 2023:** Is a document chatbot feasible? Yes
 - Chat with Document(s) using OpenAI ChatGPT API and Text Embedding (Kim Sung)*
- **First Prototype uses langchain**
 - Few lines of code
 - Preprogrammed interactions (document queries, chatbot, agents...)

```
1 # ...
2 # Load Documents
3 docs = load_text_documents()
4 # Create Embeddings (x-dimensional vectors)
5 sentence_transformer_ef = embedding_functions.SentenceTransformerEmbeddingFunction(model_name=PARA)
6 # Save to Vector Database
7 vectordb = Chroma.from_documents(docs, sentence_transformer_ef)
8 vectordb.persist()
9 # Call OpenAI API
10 elux_qa = ConversationalRetrievalChain.from_llm(ChatOpenAI(temperature=0, model_name="gpt-3.5-turbo"), vectordb, return_source_documents=True)
11 results = elux_qa({"question": question, "chat_history": ""})
12 # Show Results
13 print(results)
14 # ...
```

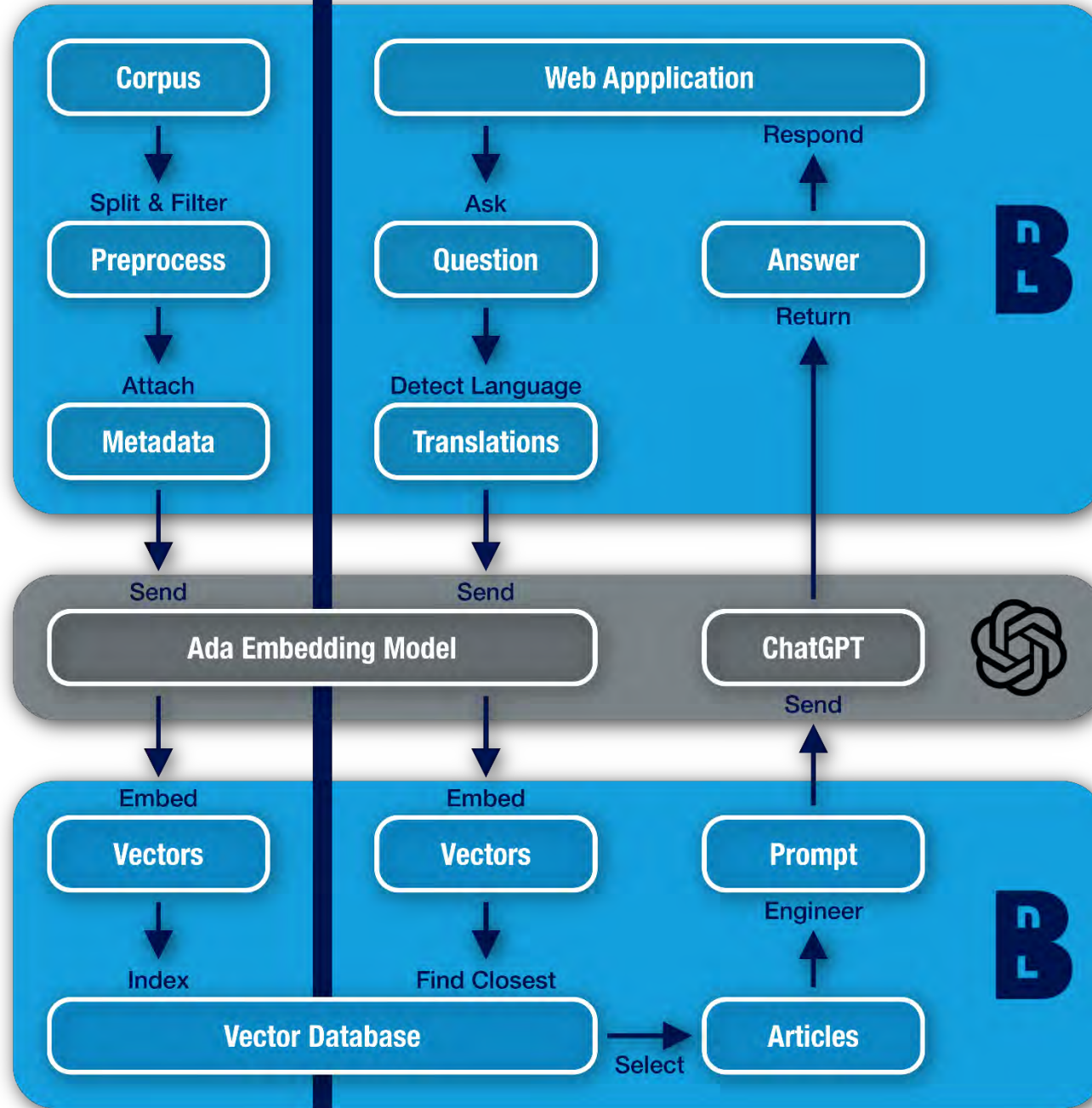
Black Box, In Memory DB, Only English, Cannot Tune, No Document Filtering



Bibliothèque nationale
du Luxembourg

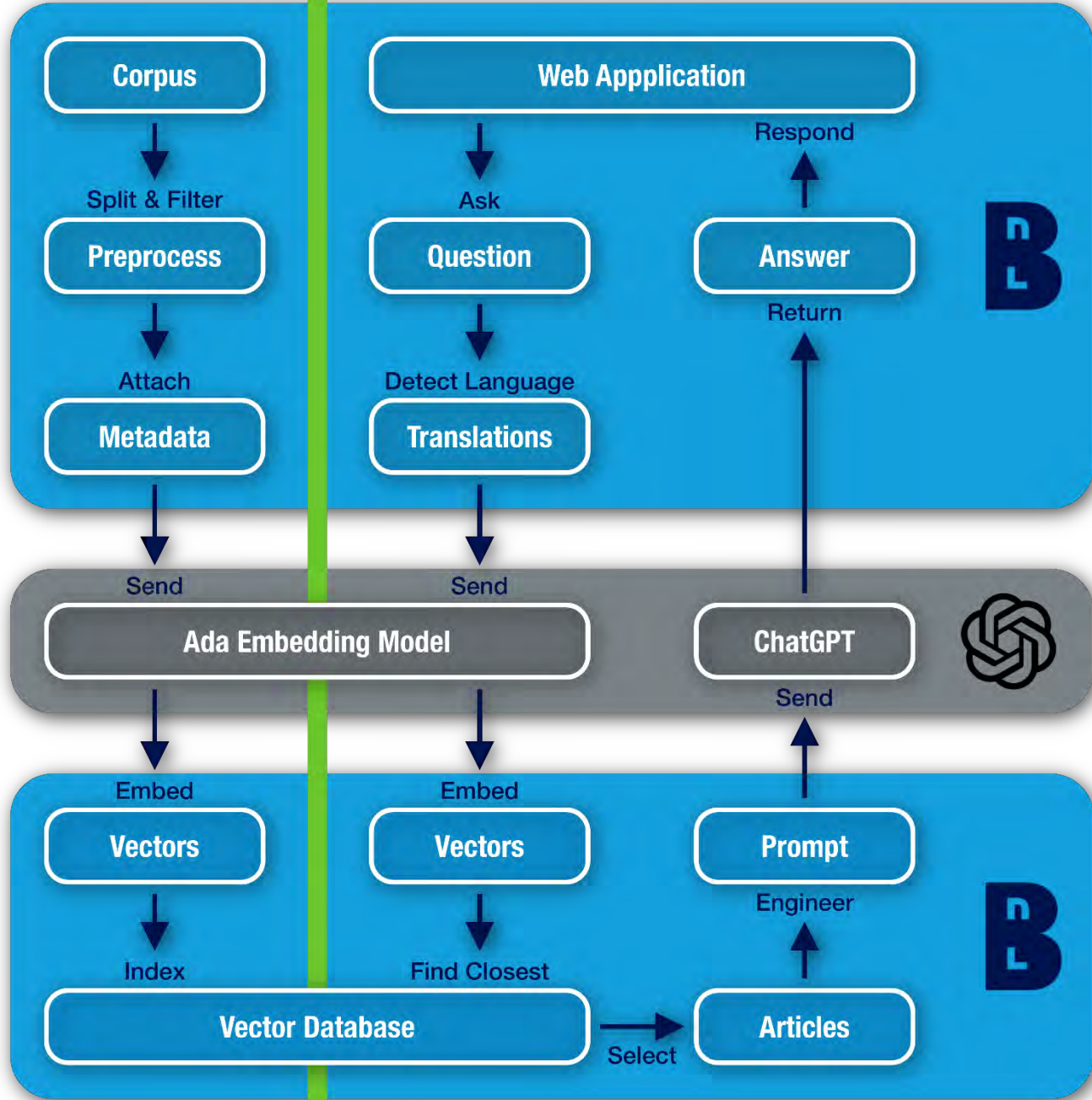
Implementation

PREPROCESSES



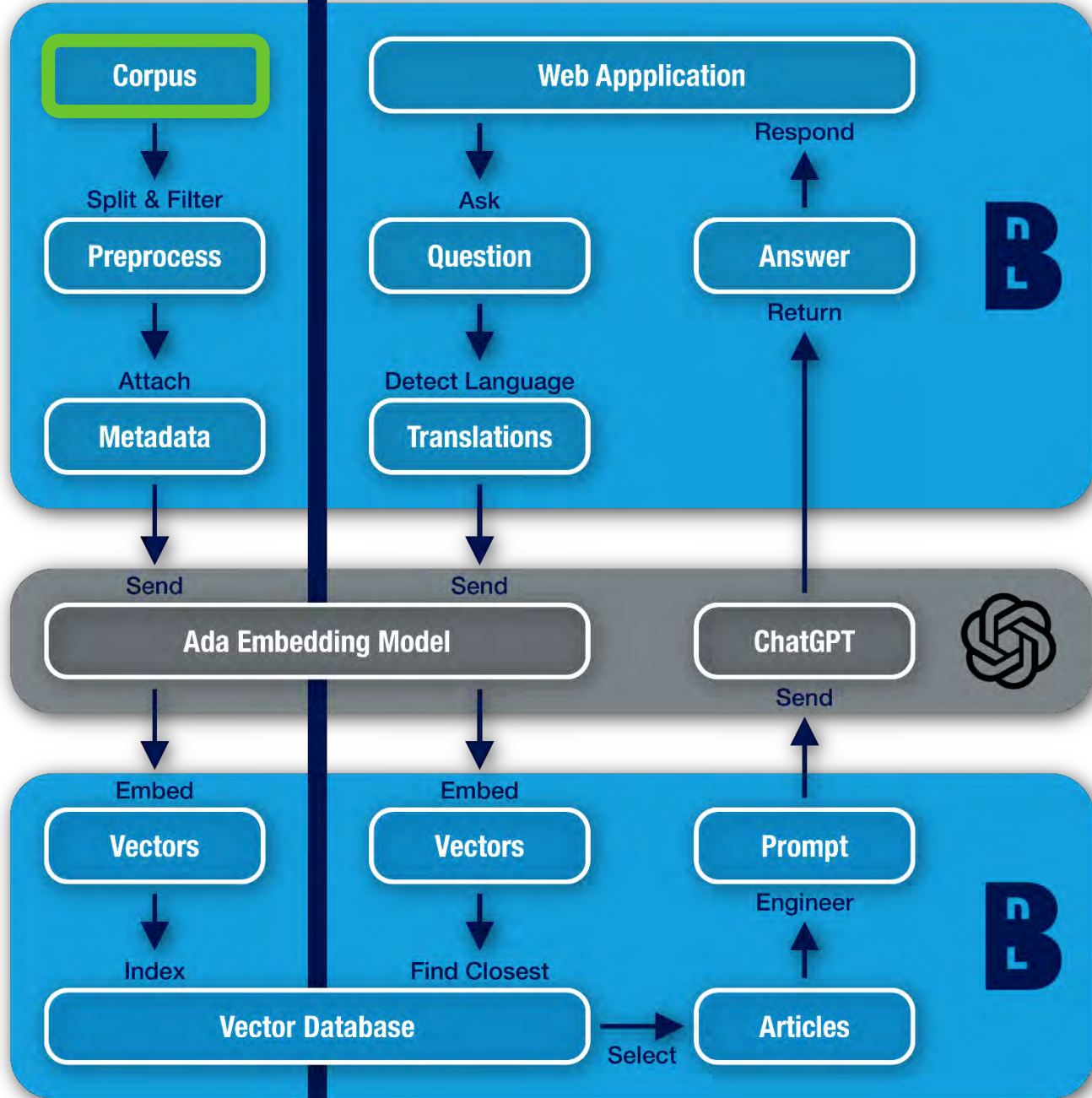
CHAT ENGINE

PREPROCESSES



CHAT ENGINE

PREPROCESSES



CHAT ENGINE

Digitisation since 2003

- Preserve
- Disseminate
- Support research



A-Z : Luxemburger illustrierte Wochenschrift
De 1933 à 1940
Publiée par Franz Stolz et Hubert Clément.



Action féminine (L)
De 1927 à 1940
L'Action Féminine est l'organe homonyme de l'organisation féminine luxembourgeoise "Action féminine" fondée en 1925 et éditée par Cathérine Schleimer-Kill.



Annuaire statistique
De 1955 à 2005
À partir de 1955, *Annuaire statistique* est publié par l'Office de statistique générale, puis par le Service central de la statistique et des études économiques (STATEC).



Arbeiter (Der)
De 1878 à 1882
Le journal *Der Arbeiter* s'adresse avant tout aux ouvriers luxembourgeois.



Arlequin (L)
De 1848 à 1848
Suite à l'abolition de censure de mars 1848, l'avocat Charles Munchen édite le premier journal satirique luxembourgeois.



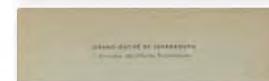
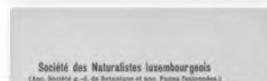
Arme Teufel (Der)
De 1903 à 1929
Journal du Parti social-démocrate luxembourgeois nouvellement fondé en 1903 et édité jusqu'en 1929 par un de ses cofondateurs Jean Schaack-Wirth.



Avenir (L)
De 1868 à 1871
Le journal *L'Avenir* est édité et imprimé par Jean Jons (1828-1893) et paraît entre avril 1868 à septembre 1871.



Bulletin de l'ANCE
De 1978 à 2008
Depuis l'année 1978, l'ANCE (Association Nationale des Communautés Educative www.ance.lu) édite trimestriellement un bulletin de liaison pour les professionnels du secteur social et éducatif.



BL Corpus : Article Segmentation

METS / ALTO

- Public tenders

- Manually tagged

- Table of contents

Luxemburger Wort
für Wahrheit und Recht

Freitag, 24. März 1961 / Jahrgang 114 / C. Ch. Post. 12-01 / Nummer 43 / VERLAG DER ST.-PAULUS-DRUCKEREI / Luxembourg Téléph. 6-3, rue J. Origer 294-41 / Telex 471 / Agence Ech-Ala Téléph. 74, rue de l'Alzette 546-81

Verschärfte Laos-Krise

Präsident Kennedy warnt die Sowjetregierung / Die USA fest entschlossen die Neutralität in Laos zu verteidigen

WASHINGTON, 24. März (UPI) In seiner mit Spannung erwarteten Pressekonferenz hat Präsident Kennedy in der Nacht zum Freitag die Entschlossenheit der Vereinigten Staaten unterstrichen, sich der Unterwerfung von Laos durch von außen geleitete kommunistische Rebellen zu widersetzen, weil sonst die Sicherheit ganz Südostasiens gefährdet sei. Kennedy richtete erneut die dringende Aufforderung an die Sowjetunion, auch ihrerseits für eine sofortige Waffenruhe in Laos einzutreten.

„Alles was wir in Laos wünschen, ist Frieden, nicht Krieg“, sagte Kennedy. Das heiÙe, daß eine Regelung am Konföderat und nicht auf dem Schlachtfeld gefunden werden müsse. Andererseits müÙten sich die USA und ihre Verbündeten in der südostasiatischen Verteidigungsorganisation (SEATO) über Gegenmaßnahmen einig werden, sofern die bewaffneten Angriffe gegen die laotische Regierung nicht aufhören. Wenn in Laos eine friedliche Lösung gefunden werden sollte, müÙten die Kommunisten ihre aus dem Ausland unterstützte bewaffnete Tätigkeit einstellen.

Kennedy, der vor über 400 Journalisten sprach, lehnte es ab, zu bestimmten Siedlungen zu nehmen, nach denen der Flugzeugträger „Midway“ und andere Schiffe von Hongkong in das südchinesische Meer ausgelaufen sind und Bereitchaftsbefehl an Einheiten des Marinekorps im Persen Ozean ergangen, sein soll.

Er sagte, er habe den Sowjets keine Frist für eine Antwort auf die westliche Friedensvorschläge für Laos gestellt. Frieden würde durch den Gang der

gegebenfalls eine internationale Konferenz über Laos vor.

Nach der Darstellung Kennedys verfolgen die USA in ihrer Laos-Politik drei Ziele: 1. MüÙte Laos völlig neutral sein, 2. MüÙte ein Waffenstillstand geschlossen werden und 3. müÙten konstruktive Verhandlungen über alle Streitfragen eingeleitet werden. Niemand, so betonte der Präsident, sollte an der Entschlossenheit der USA zweifeln, ihren Verpflichtungen gegenüber Laos und der SEATO nachzukommen.

An Hand einer großen Karte von Laos demonstrierte Kennedy, wie die Kommunisten einen stetig größer werdenden Teil des hinterindischen Königreiches an sich gerissen hätten, die kommunistischen Putsch-Laos-Streitkräfte hätten von außen wachsende Unterstützung erhalten. Darüber hinaus hätten die Kommunisten den Rebellen ein komplettes, mit schweren Waffen ausgerüstetes Team von Kriegssachverständigen zur Verfügung gestellt, wodurch das „wegenwärtige, schwere Problem“ entstanden sei. Bei der Genfer Indochina-Konferenz sei 1954 die Neutralisierung von Laos einseitig beschlossen worden. Die ersten Jahre hätten Fortschritte in Richtung auf dieses Ziel gebracht, aber die Kommunisten hätten den Versuch, die Macht an sich zu reiÙen, als aufgegeben.

Vor der Pressekonferenz hatte der Präsident die Botschafter der 15 Staaten des Nordatlantikkapitals über die Haltung der USA im Laos-Konflikt unterrichtet lassen.

Der britische Premierminister Harold Macmillan berief noch in der Nacht unmittelbar nach Eintreffen der Berichte über Kennedys Pressekonferenz das ganze britische Kabinett zu einer Besprechung.

Unterstrich Stevenson die Bedeutung eines Ausbildungsprogramms für afrikanisches Fachpersonal, das für die Bewältigung eines Entwicklungsprogramms unerlässlich sei. Er erwähnte ferner den Ausbau von Straßen, Klüften, Wohnungen. Insbesondere seien Straßen wichtig, weil sie die Isolation abgelegener Siedlungen durchbrächen. Straßen allerdings, die nicht an Gewinnen aufbauen dürften, sondern über die Grenzen hinwegführen müÙten.

Die Wiener Besprechungen des luxemburgischen Außenministers

Wien, 24. März (AP) Der luxemburgische Außenminister Eugène Schaus, der sich gegenwärtig zu einem offiziellen Besuch in Wien aufhält, ist gestern von dem österreichischen Bundespräsidenten Dr. Adolf Schärf empfangen worden. Anschließend konferierte der Minister mit Bundeskanzler Raab und Außenminister Dr. Bruno Kreisky.

AnläÙlich des Empfangs beim Bundespräsidenten erhielt der luxemburgische Außenminister eine der höchsten Auszeichnungen, die Österreich zu vergeben hat. Bundespräsident Schärf überreichte Minister Schaus ein Insignien des großen goldenen Zierordens anlässlich für Verdienste um die Republik Österreich.

Bei einem von der österreichischen Regierung zu Ehren des luxemburgischen Außenministers gegebenen Essens betonte Bundeskanzler Raab, daß der Besuch des luxemburgischen Gasten die zwischen den beiden Staaten bestehende Tradition herzlichster Freundschaft und persönlicher Kontakte festsetze.

Außenminister Schaus sagte, Luxemburg werde im Rahmen des möglichen dazu beitragen, daß die wirtschaftliche Zusammenarbeit zwischen den beiden Staaten



Infolge eines Schienensturzes hat Wert in die Schelde über die Dijer getreten und hat mehrere hundert Ha Anzeigend überzochmet. (Dijle-Flotte)

„Verhandeln ist nicht Frieden“

Ferhat Abbas spricht von langen und schwierigen Algerien-Verhandlungen

Tunis/Bern, 24. März. (UPI) Wenige Tage vor Beginn der Verhandlungen zwischen der französischen Regierung und der algerischen Exilregierung warnte die Rede des algerischen Ministerpräsidenten die Exilregierung, daß die Verhandlungen über den Algerienkonflikt nicht zu einem schnellen Ende führen würden. Die Rede des algerischen Ministerpräsidenten wurde von politischen Beobachtern als Zeichen für die Schwierigkeiten der Verhandlungen interpretiert.

Beluxemburgensia Statistics

PAGES
1,609,175

ARTICLES
8,847,159

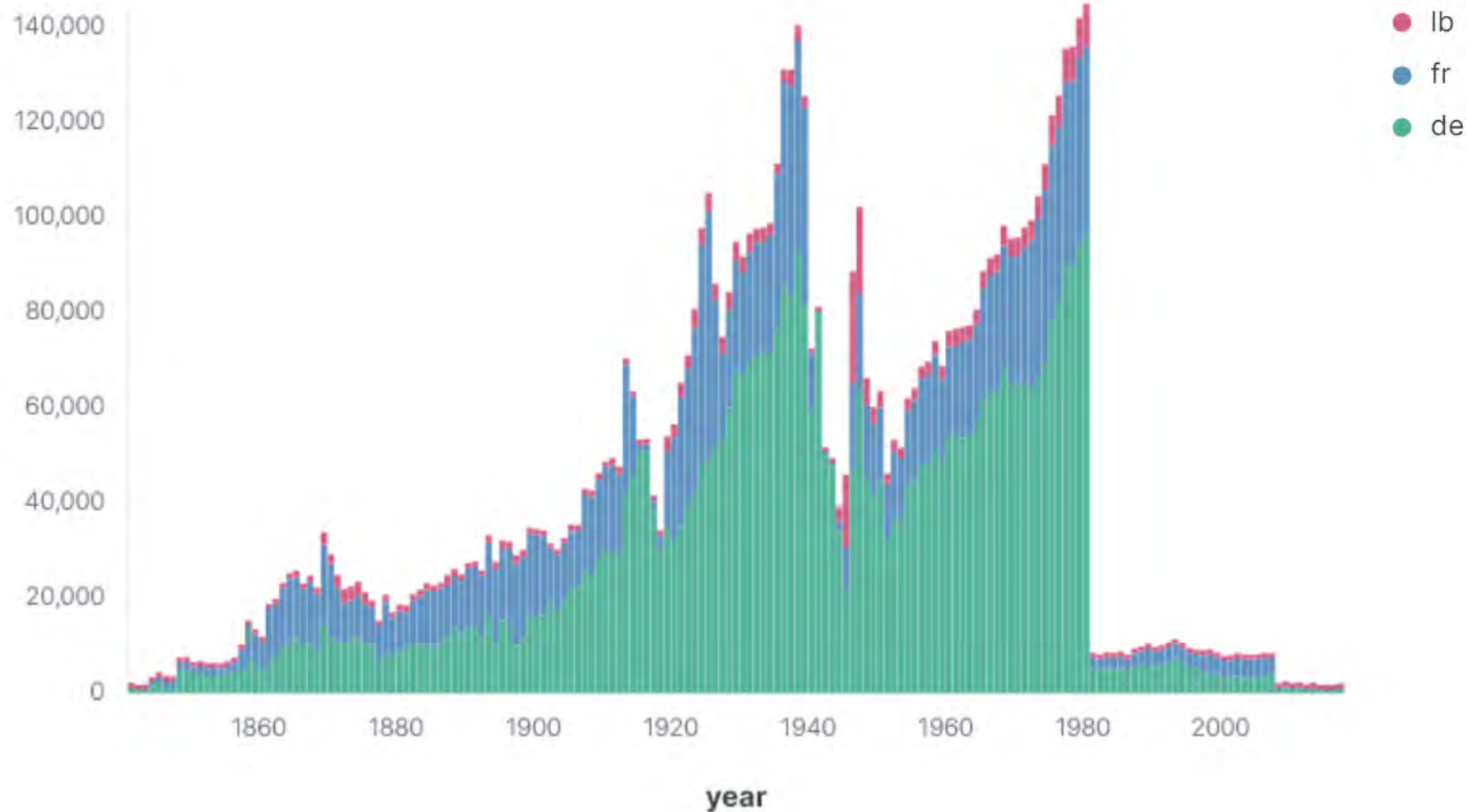
PÉRIODIQUES
113,856

LIVRES
505

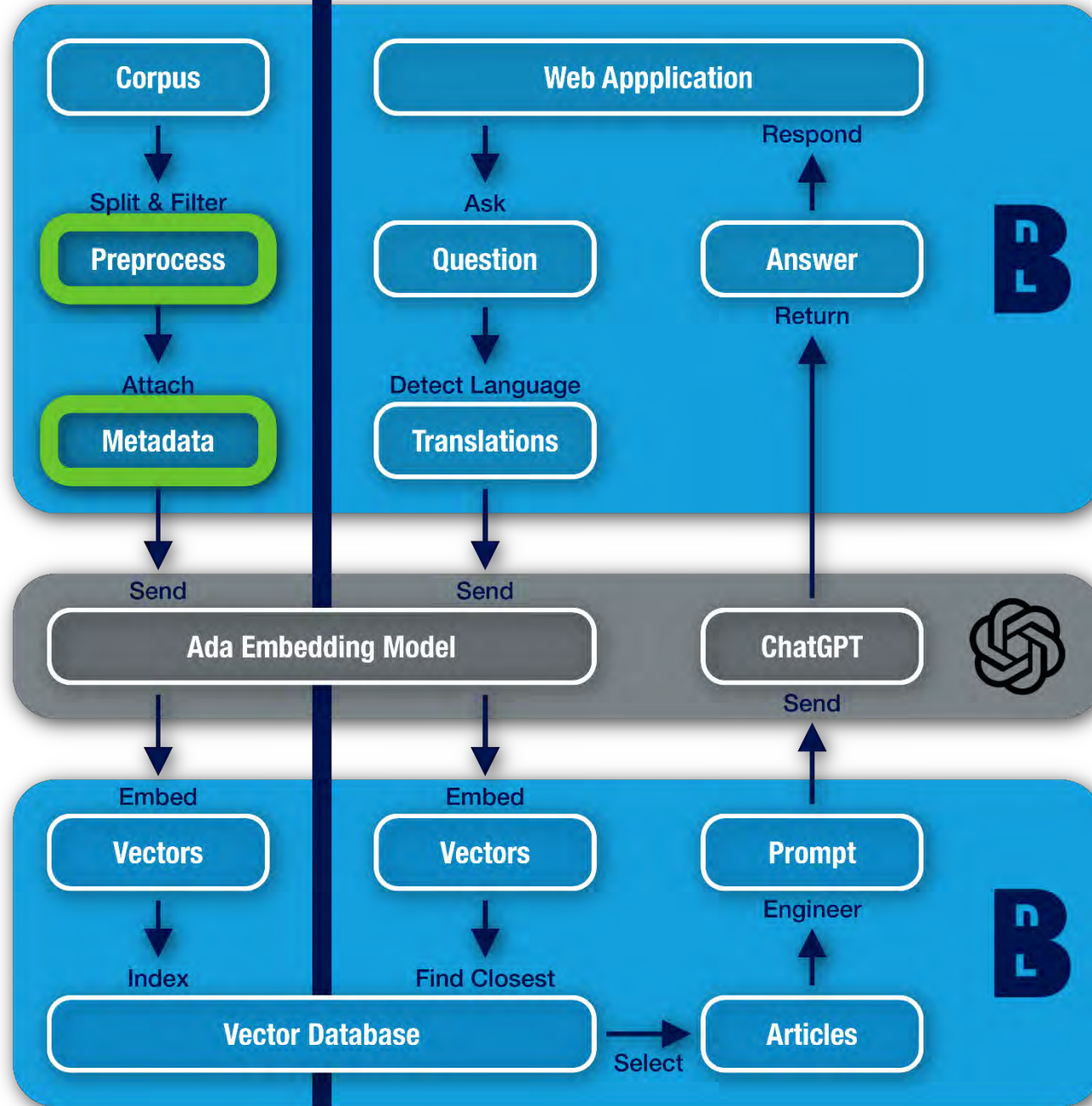
AFFICHES
782

MANUSCRITS
15

CARTES POSTALES
17,421



PREPROCESSES



CHAT ENGINE



Full Article



Max 2.5k Chars (~500 Tokens)



Chunk 1



Chunk 2

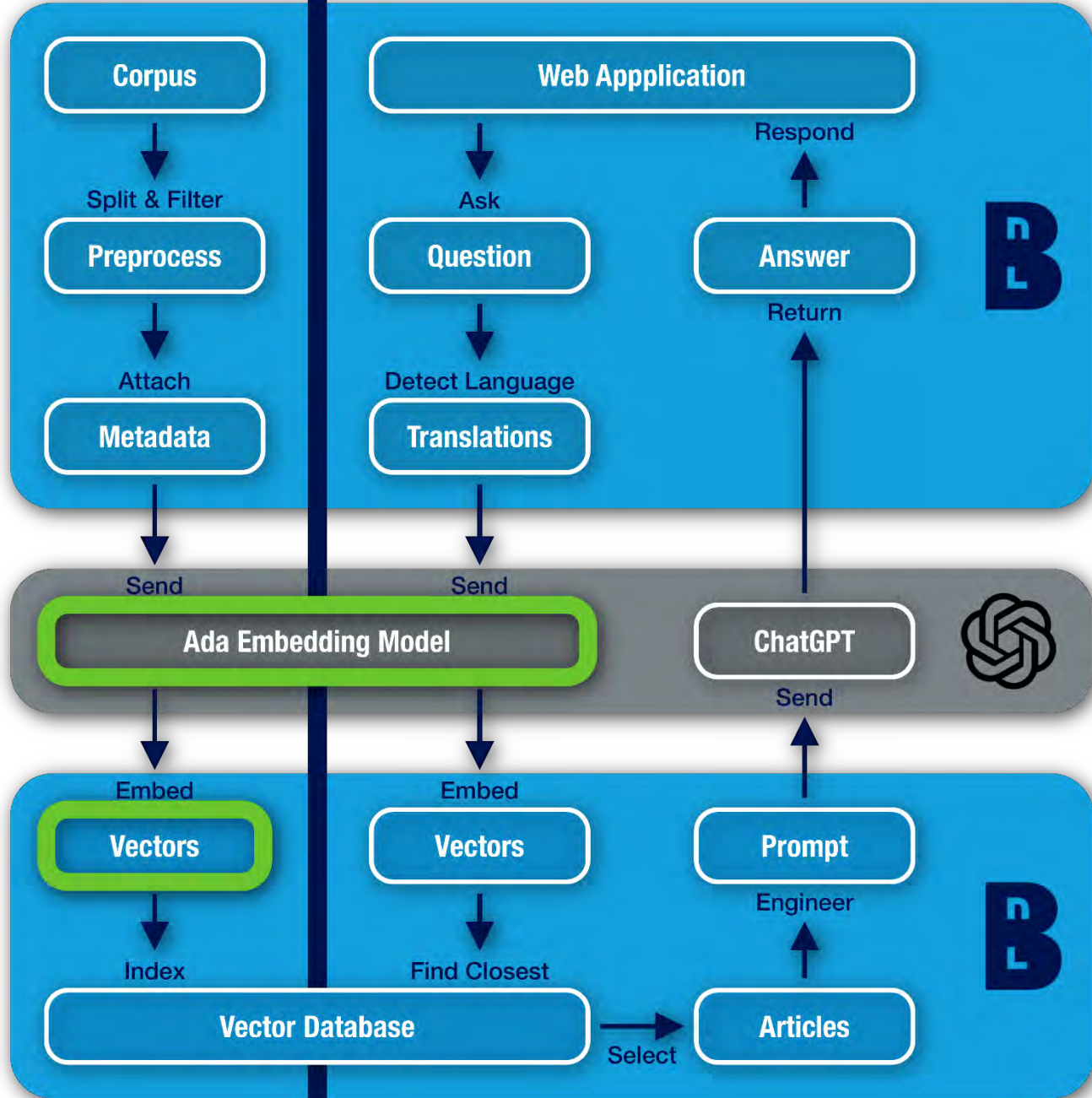


Preprocessing

Final Dataset

- Discarding articles with less than 100 characters
- Newspaper, date, URL, ... associated with every chunk
- 9.5m chunks derived from 8m articles

PREPROCESSES



CHAT ENGINE

horse

Embedding
Model

0.001

-0.032

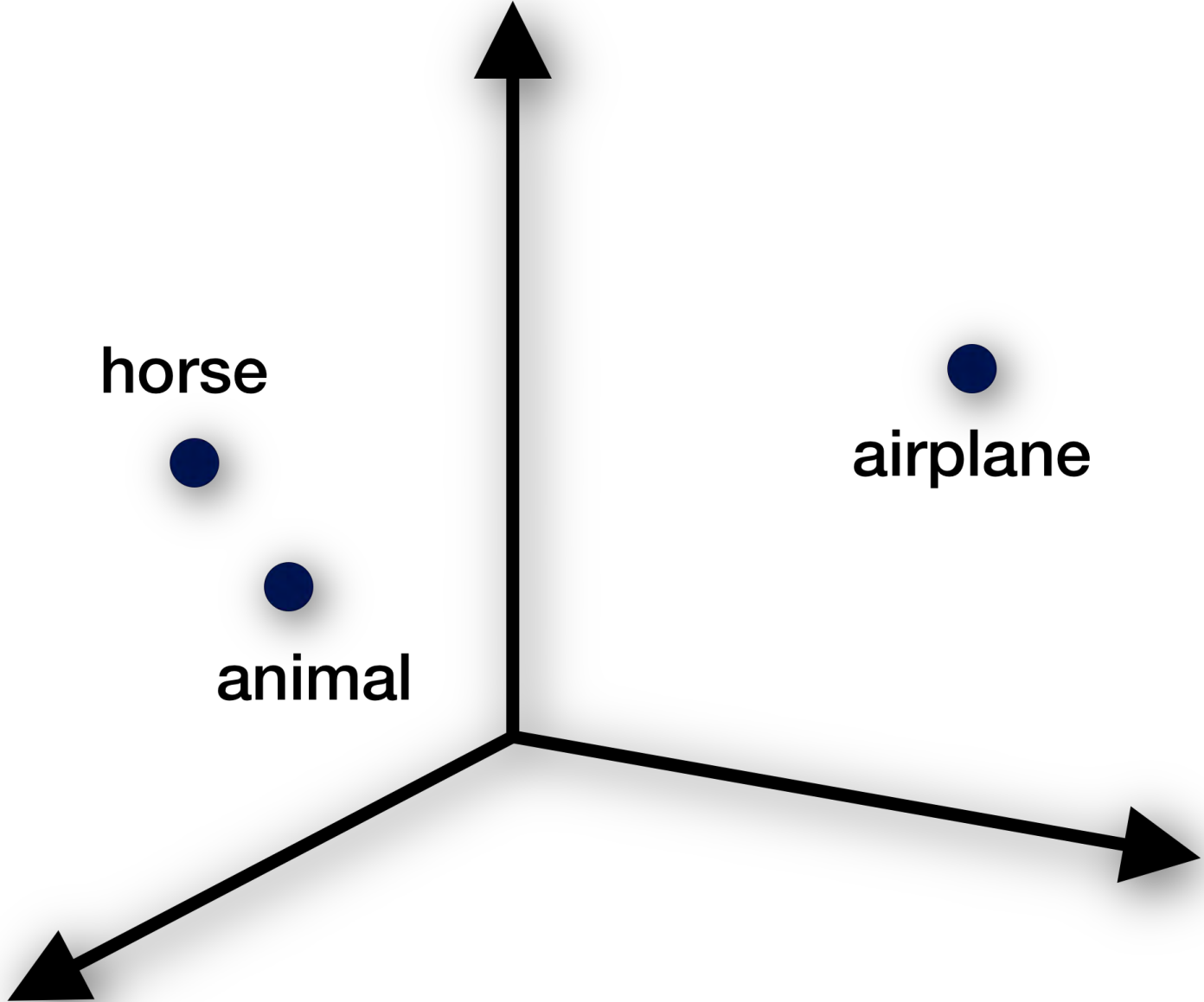
-0.025

0.017

-0.015

...

0.019





Embedding (June 2023)

OpenAI Ada

\$1 / 10m tokens

Unknown LB (multilingual)

1536 dimensions

Cohere

\$4 / 10m tokens

Supports LB (multilingual)

768 dimensions

Paraphrase

Free & open source

No LB (multilingual)

384 dimensions

Symanto

Free & open source

No LB (multilingual)

768 dimensions



Embedding

Manual Evaluation

- Test set with 5k DE and 5k FR chunks (same topic)
- Compute top-5 similar chunks for every model
- Manual evaluation of similarity and overlap
- Analysis of language distribution



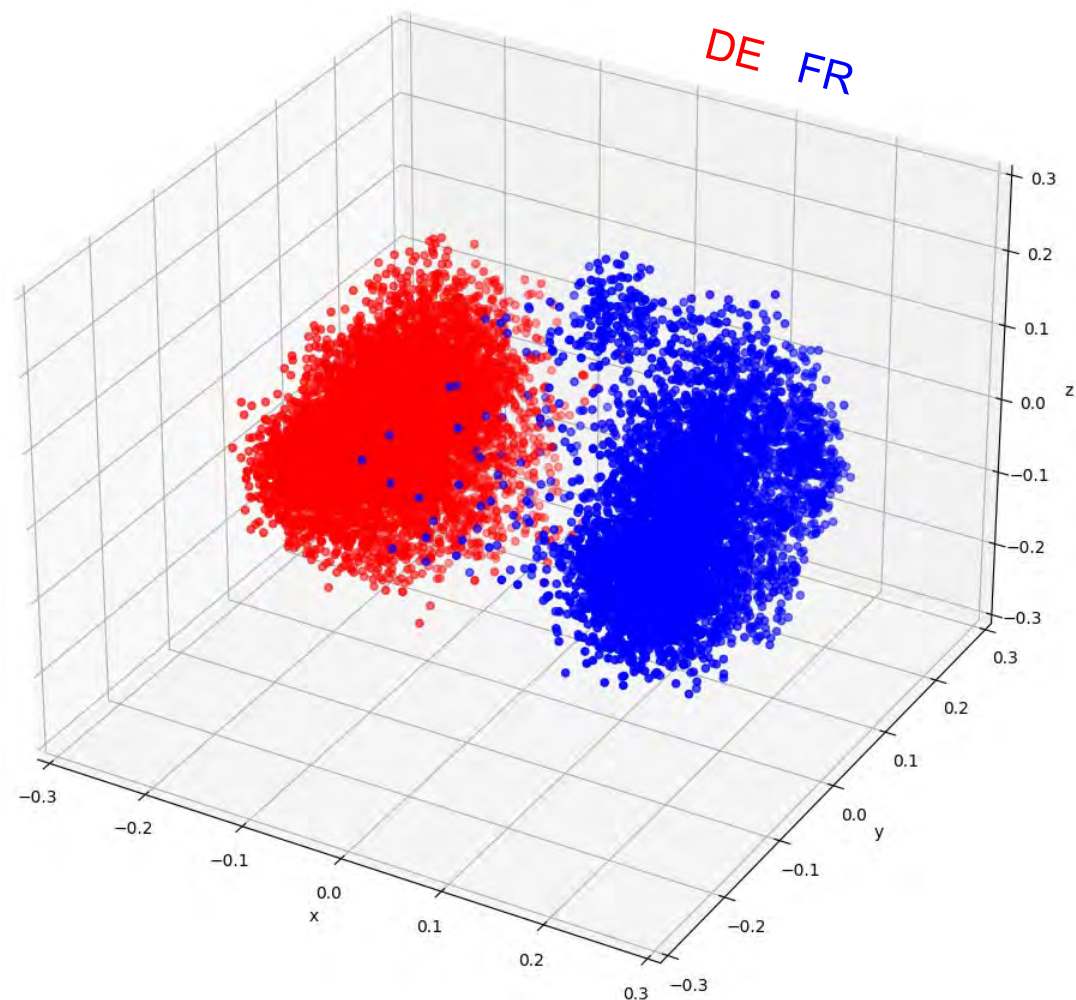
Embedding

Experiment

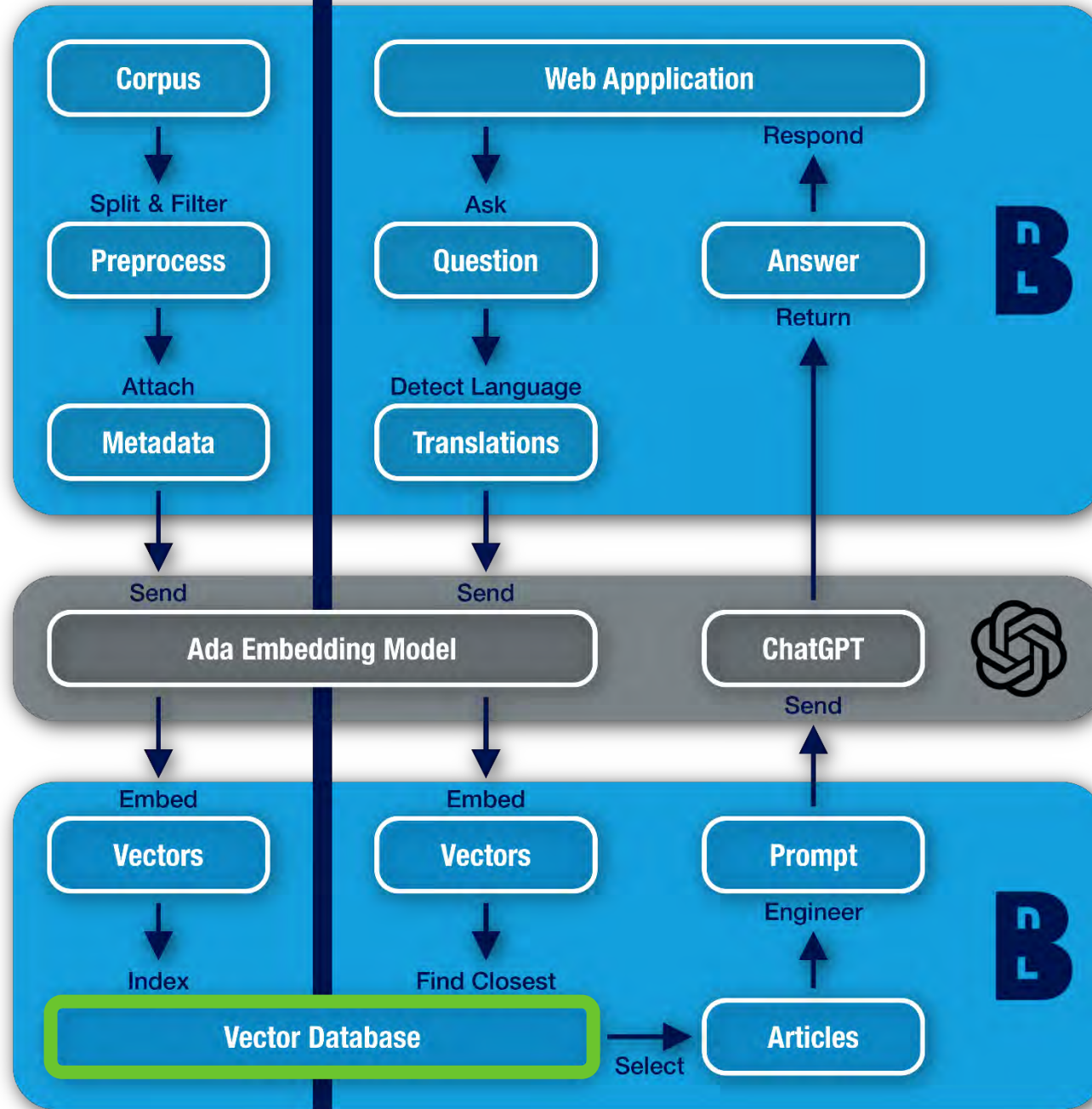
- OpenAI Ada embeddings for test set
- Reduced to 3D using PCA
- Languages divided in space

Solution

- Translation



PREPROCESSES



CHAT ENGINE







Vector Database

- 9.5M vectors, 1536 dim, text, metadata
- Need for a fast vector database with
 - Cosine & dot-product distance
 - Filtering
 - Metadata
- ChromaDB did not work for our scale (in-memory)
- Switch to qdrant (memory & memory-mapped)
- Tests with 2 node cluster
- Tests with elasticsearch

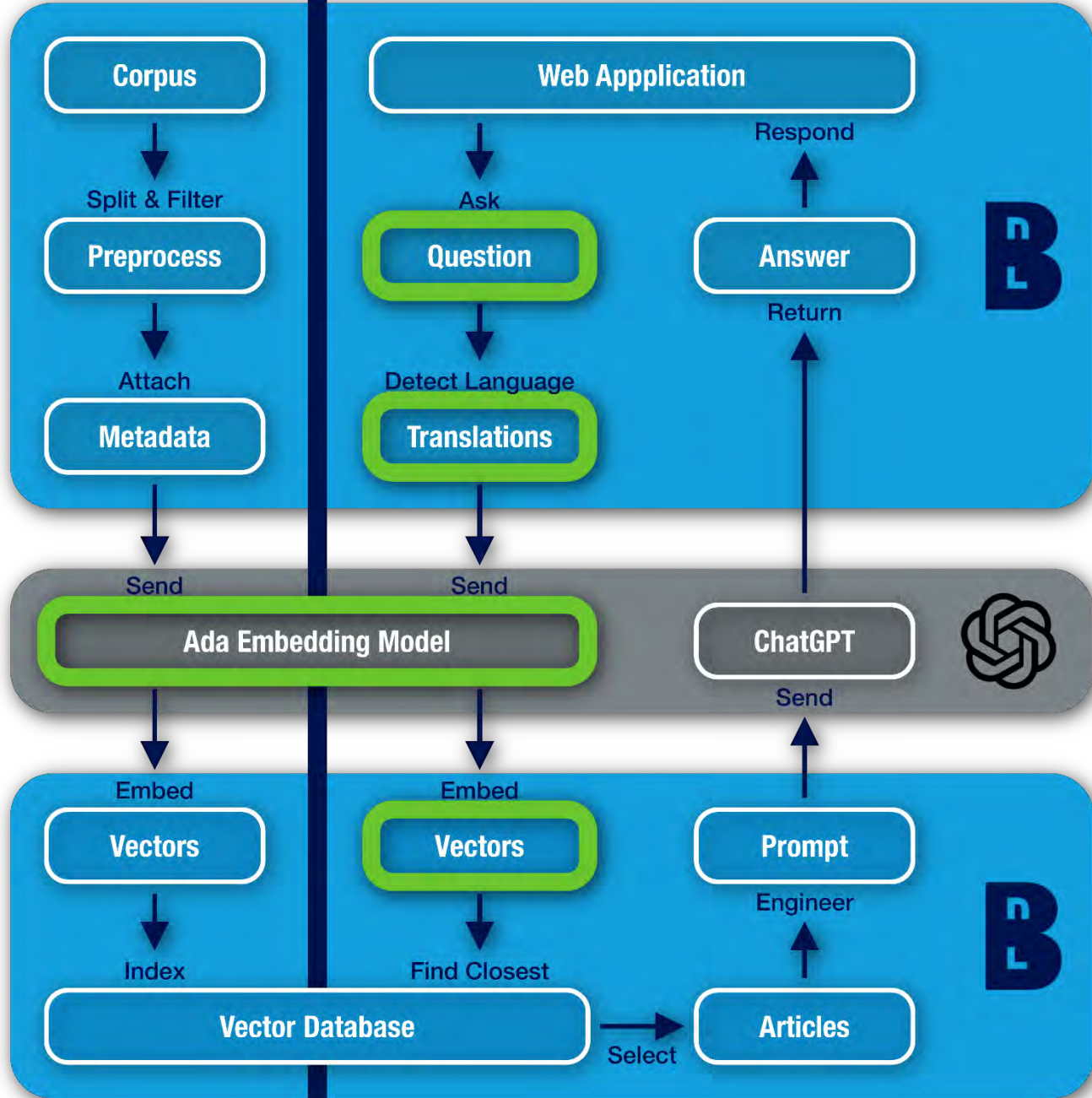




Vector Database

	Server Type	milliseconds
 drant	1 Server (24GB RAM, 8 CPU)	12000
 drant	Cluster of 2 Servers (2x24 GB, 8 CPU)	4200
 elastic	1 Server (24 GB RAM, 8 CPU)	75000
 drant	1 Physical Server (120 GB RAM, 96 CPU)	20

PREPROCESSES



CHAT ENGINE

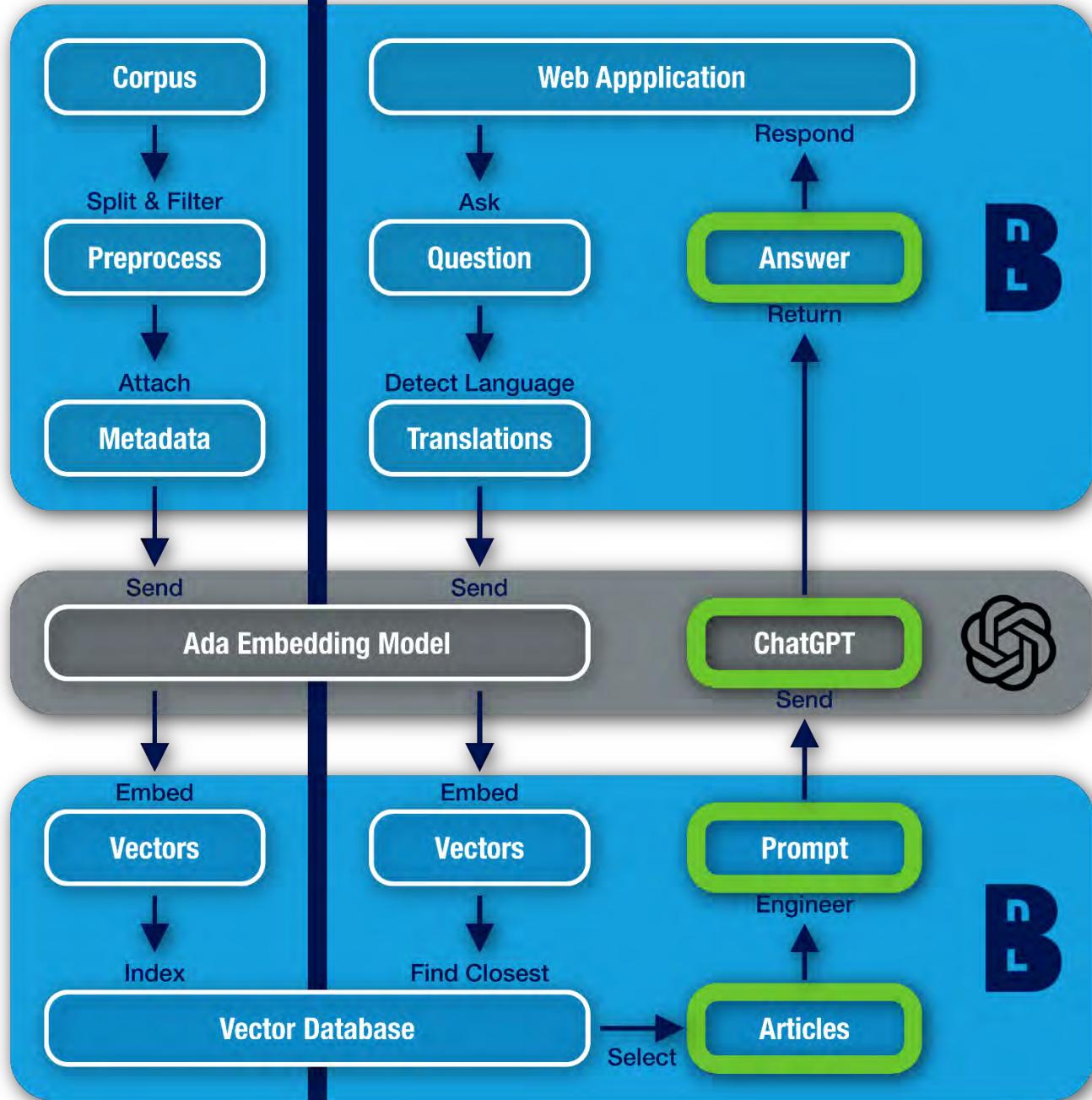


Translation

M2M-100 Meta

- Local and open source
- Requires known source language (langid)
- Can translate between any pair of 100 languages
- Accelerated using GPU
- Target languages: DE and FR

PREPROCESSES



CHAT ENGINE



Prompt Engineering

Benutze nur die Informationen der folgenden Zeitungsartikel:

Artikel 1 vom 20.06.1963:

Arbeiten zur Konstruktion des „Pont Grande-Duchesse Charlotte” ...

Artikel 2 vom 24.10.1966:

Le Pont Grande-Duchesse Charlotte ...

...

Um diese Frage zu beantworten (auf Deutsch):

Wann wurde die Brücke Grand-Duchesse Charlotte gebaut?

Template

Metadata

Chunks

Question



Large Language Models (October 2023)

GPT-3.5 OpenAI

4k / 16k tokens

Some LB (multilingual)

\$15 / \$30 / 10m Tokens

Llama 2 Meta

4k tokens

No LB (multilingual)

Free & open source

Flan-T5 Google

2k tokens

No LB (multilingual)

Free & open source



Bibliothèque nationale
du Luxembourg

Production and Communication



Costs (october 2023)

One time

- Embeddings for 9.5m chunks: €450

Running

- Per user question: €0.02
 - Language model: 99,9%
 - Embedding model: 0,1%
- Daily question limit per account: 30





Acceptable Use

- Usage of OpenAI models → Usage policies from OpenAI
 - Building products for political campaigning or lobbying purposes
 - Automated determinations of eligibility for credit, employment, educational institutions, or public assistance services
 - High risk government decision-making
 - ...
- Queries can be monitored by OpenAI
- Systems must provide a disclaimer to users informing them that AI is being used and of its potential limitations



Communication

Important Messages

- New and unconventional type of service
- Different use case than traditional search bar
- Understanding of the chat engine process
- Usage limitations (e.g. exhaustive search)
- Not perfect, there can be mistakes



Future Updates

Feedback

- Listen to and evaluate user feedback
- Thumbs up/down reactions (96% none / 2% up / 2% down)

Ideas

- Variable amount of reference articles
- Date range to constrain references articles in time
- Improved translation model for *lb*
- Maybe only propose semantic search
- GPT4 Turbo, function calling etc.



Bibliothèque nationale
du Luxembourg

Demo



Bibliothèque nationale
du Luxembourg

Thank you

<https://chat.eluxemburgensia.lu>

Yves Maurer

National Library of Luxembourg

yves.maurer@bnl.etat.lu

@yvesmaurer

@ymaurer@mastodon.top