



DSDA 310 - FALL 2025

# Case Study 2 - Company Intelligence Agent

*Presented by Julie Bach and Samuel Doherty*

# DATA PROCESSING



## DATA PROCESSING

- Document extraction
- Text cleaning
- Text chunking

## APPLICATION LAYER

- TF-IDF model
- Groq LLM
- Streamlit interface

## DOCUMENTS

- Delta Airline's 2024 Form 10-K (SEC Filing)
- Delta Airline's 2024 ESG Report
- Major Shareholders Summary (Yahoo Finance)

## ASSISTANT

- Company overview
- Business segments & revenue sources
- Risks
- ESG Goals
- Stakeholder information

# How Are Texts Extracted?



yahoo!  
finance

PDFPlumber

OCR

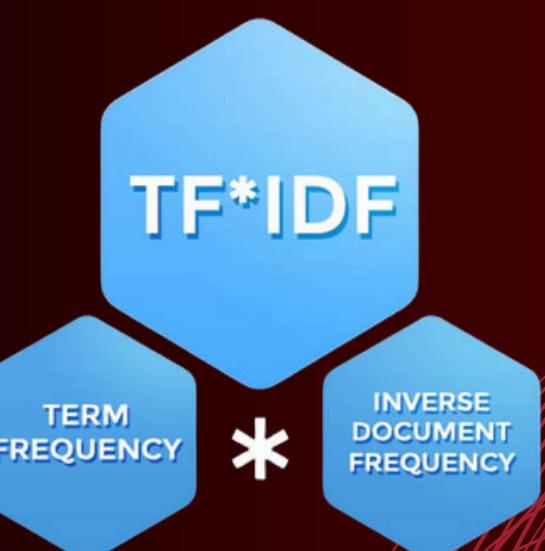
PDFPlumber

The ground or stays  
iverse is vast, and you  
s also beautiful. You a  
nthing bigger than yo  
t of something that ma  
most of your time. Take  
a blog post. Make a

The ground or stays  
iverse is vast, and you  
s also beautiful. You a  
nthing bigger than yo  
t of something that ma  
most of your time. Take  
a blog post. Make a

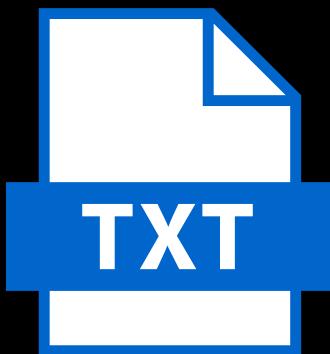
# KNIFE

# KNIFE



## READY

## How Are Texts Cleaned?



- Remove headers/ footers
- Fix hyphen, broken words, & extra whitespace
- Normalize punctuation & line breaks
- Repair PDF artifacts

# How Are Texts Chunked?

600

Chunk Size

100

Overlap

169

Total created chunks

chunk_id	company	source_file	chunk_text	word_count
0	chunk_001	delta annual_report_2024	UNITED STATES SECURITIES AND EXCHANGE COMMISSION Washington, D.C. 20549 FORM 10-K ANNUAL REPORT PURSUANT TO SECTION 13 OR 15(d) OF THE SECURITIES EXCHANGE ACT OF 1934 iS} For the fiscal year ended...	600
1	chunk_002	delta annual_report_2024	statements of the registrant included in the filing reflect the correction of an error to previously issued financial statements 0 Indicate by check mark whether any of those error corrections are...	600
2	chunk_003	delta annual_report_2024	actual results to differ materially from historical experience or our present expectations. Known material risk factors applicable to Delta are described in "Risk Factors Relating to Delta" and "R...	600
3	chunk_004	delta annual_report_2024	six continents. Our domestic network is centered around core hubs in Atlanta, Detroit, Minneapolis-St. Paul and Salt Lake City. Core hubs have strong local passenger share, a high penetration of c...	600
4	chunk_005	delta annual_report_2024	consumer brand, including: + Recognized as the 2024 Airline of the Year by aviation publication Air Transport World. + Named Best U.S. Airline and topped five categories in the Forbes Travel Guide...	600
5	chunk_006	delta annual_report_2024	mileage credits ("miles") when traveling on Delta, Delta Connection and our partner airlines. Miles may also be earned by using certain services offered by program partners, such as credit card, r...	600
6	chunk_007	delta annual_report_2024	Consolidated Financial Statements for additional information about our equity investments. Each of our joint venture or cooperation arrangements provides for joint commercial cooperation with the ...	600
7	chunk_008	delta annual_report_2024	the other member airlines, providing opportunities to increase connecting traffic while offering enhanced customer service through reciprocal codesharing and loyalty program participation, airport...	600
8	chunk_009	delta annual_report_2024	from around the world. With agreements to service multiple next-generation aircraft engines, Delta TechOps is positioned as a leading global service provider for state-of-the-art, more sustainable...	600
9	chunk_010	delta annual_report_2024	work with airport partners to drive fuel savings by limiting the use of aircraft Auxiliary Power Units ("APUs") during ground operations. In 2024, we installed 30 new preconditioned air units at o...	600

# Chunks

	chunk_id	company	source_file
0	chunk_001	delta	annual_report_2024
1	chunk_002	delta	annual_report_2024
2	chunk_003	delta	annual_report_2024
3	chunk_004	delta	annual_report_2024
4	chunk_005	delta	annual_report_2024
5	chunk_006	delta	annual_report_2024

## Query

"What are the expected future benefit payments in 2025?"

## TF-IDF weighting system

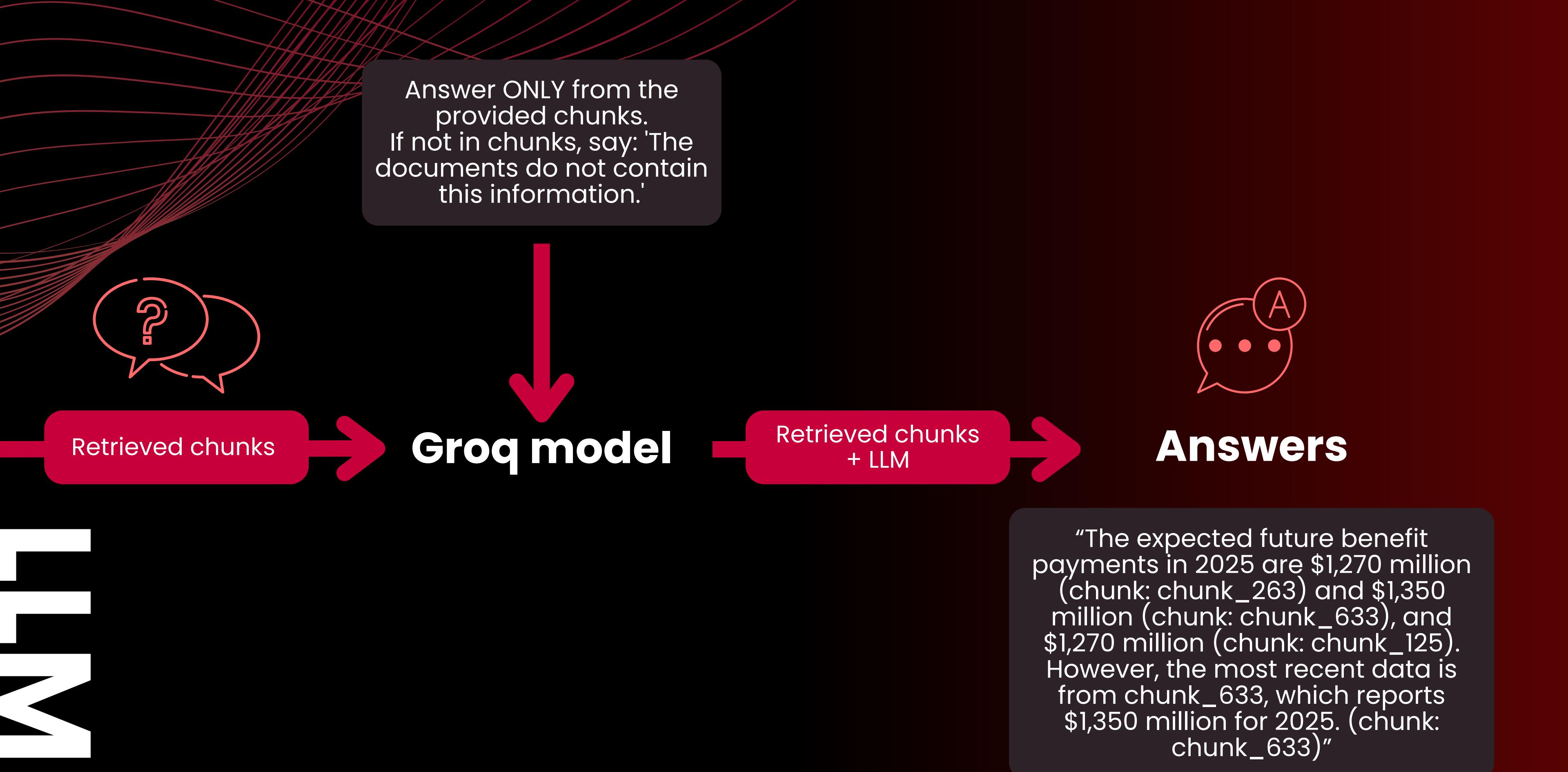
$$f_{t,d} \cdot \log \frac{N}{n_t}$$

Most relevant chunks are returned.

### Top Chunks:

	chunk_id	score
392	chunk_393	0.358145
262	chunk_263	0.352271
632	chunk_633	0.348160
124	chunk_125	0.346644
516	chunk_517	0.315630

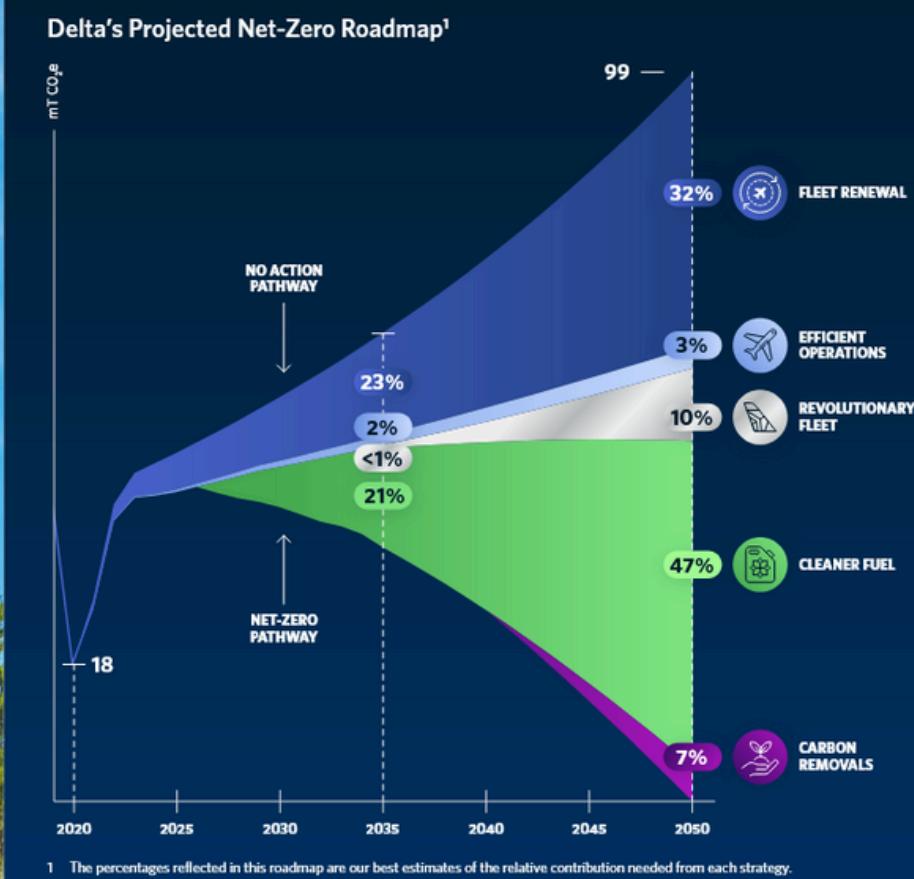
RETRIEVE



# ANSWER

# **RETRIEVAL DEMO & LLM ANSWER**

# LLM Fact-Check



“Sustainability is good for business, our people and the planet, and as we enter our centennial year, we’re as focused as ever on driving long-term impact as we aspire to net-zero emissions by 2050.”



**AMELIA DELUCA**  
Chief Sustainability Officer

Our reputation and brand could also be adversely impacted by, among other things, failure to make progress toward and achieve our environmental sustainability goals, as well as public pressure from investors or policy groups to change our policies or negative public perception of the environmental impact of air travel. For example, we have established ambitious goals to reduce our greenhouse gas emissions. Achieving these ambitious goals will require significant capital investment from manufacturers and other stakeholders, as we are unable to achieve these goals using our existing fleet, current technologies and available fuel sources. We are continuing to develop our climate strategy and transition plan; however, our ability to execute on such a plan is subject to substantial risks and uncertainties, as it is dependent on the actions of governments and third parties and will require, among other things, significant capital investment, including from third parties, research and development from manufacturers and other stakeholders, along with government policies and incentives to reduce the cost, and incent production, of SAF and other technologies that are not presently in existence or available at scale. Significant damage to our reputation and brand could have a material adverse effect on our business and financial results, including as a result of litigation related to any of these matters.

## Delta's DEI Strategy

DRIVING SOCIAL IMPACT FOR EVERYONE, EVERYWHERE TO THRIVE



### ACTIVELY VALUING DIVERSITY

We actively value diversity by hiring highly qualified candidates with diverse backgrounds and skills and supporting and uplifting our employees, customers, communities and business partners that reflect and respect the world we serve

### BOLDLY PURSUING EQUITY

We boldly pursue equity, with an emphasis on the workforce, community and supply chain, creating fair and open access to opportunities for all.

### CONSCIOUSLY PROMOTING INCLUSION

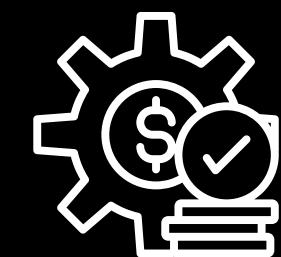
We consciously promote inclusion, building on Delta's culture of listening and engagement to foster greater understanding and a sense of belonging for all.

# Regex-Assisted Preview

Text File  
Pattern  
Window

Print  
Snippets

Found 4 matches for pattern: 'cargo revenue'  
[1] ...a fixed dollar or percentage division of ,  
[2] ...mber of SkyTeam Cargo, an international ai  
[3] ...ts our estimate of revenue expected to be  
[4] ... of revenue expected to be recognized in 1



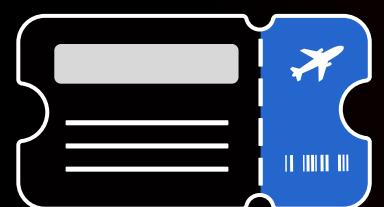
*Operating  
Revenue*



*Cargo  
Revenue*



*Total Business  
Segments*



*Total  
Customers*

# Manual Verification

Text Scan  
Locate values

Add  
KPI



*ESG  
Goals*



*YoY Revenue  
Growth*



*Total Fuel  
Savings*

# STREAMLIT DASHBOARD

# LIMITATIONS



**01** Accuracy in Text Extraction

**02** Manual KPI Extraction

**03** Chunking Strategies

**04** Limited Groq API Calls

**THANKS FOR LISTENING!**  
**QUESTIONS?**