



Data Science Convention 2024

April 11, 2024
Norris Conference Center
Houston, TX

Knowledge Recovery: How Large Language Models Help Surface Knowledge Hidden In Vast Repositories

Joshua Eckroth, PhD
Chief Architect
i2k Connect, Inc



Data Science Convention 2024

April 11, 2024
Norris Conference Center
Houston, TX

Goal: Talk to your Data

- Basic idea:
 - Accept a query from the user *in natural language*
 - Figure out what they're asking for
 - Find the answer(s)
 - Structured & unstructured sources; very large corpora
 - Deliver or even explain the answer back to the user
 - Handle possible follow-up queries



Data Science Convention 2024

April 11, 2024
Norris Conference Center
Houston, TX

what is the total depth of 15/9 F-7?

The total depth of 15/9 F-7 is 1083 m MD / 1077 m TVD ([Final Well Report for Well 15/9-F-7](#)).

what is the TD of 15/9 F-7?

The Total Depth (TD) of well 15/9-F-7 is 1083 m MD / 1077 m TVD. ([Final Well Report for Well 15/9-F-7](#)).



Use of Large Language Models

- Distant Past (no LLMs):
 - Templates to recognize “intent” to extract relevant facts/entities from query
- Recent Past:
 - Early LLMs (e.g., BERT) to recognize query intent and extract entities
 - Must be trained for the kinds of queries and entities that are expected
- Present:
 - SOTA LLMs (e.g., GPT-4, Claude 3) that are “instructed”
 - Typically are not trained for the task; given examples (zero-shot, one-shot, ...)



Retrieval-Augmented Generation

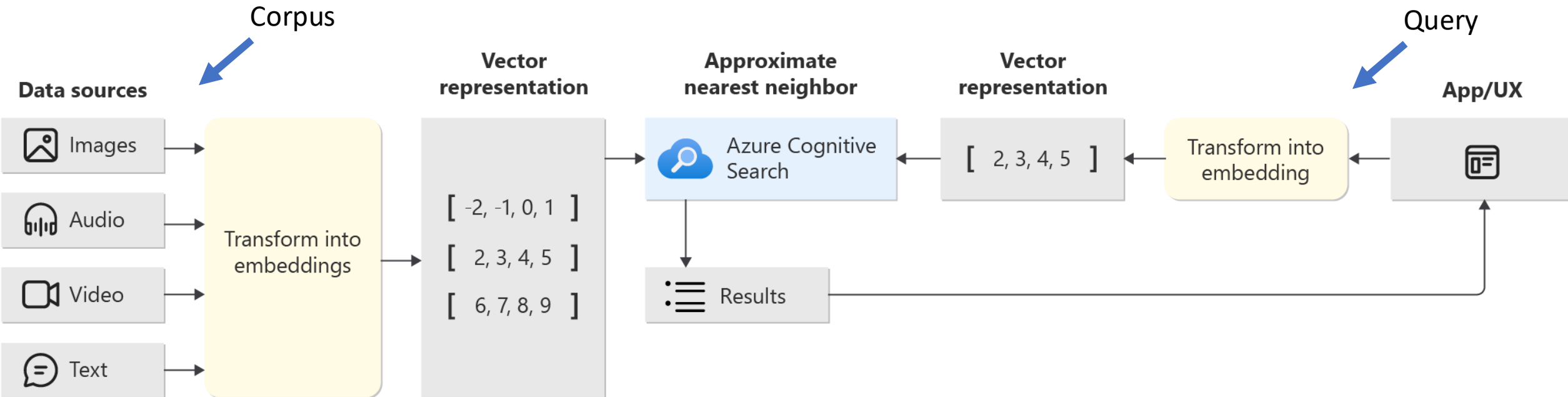
- *Just Retrieval:*
 - Typical search like SharePoint.
- *Just Generation:*
 - Talking to ChatGPT.
- *Retrieval-Augmented Generation:*
 - Finding documents/pages/passages/tables/etc. in corpus based on the user's query, *then* feeding this + the user's query to an LLM so it can answer the question given some knowledge sources.



Challenge: Finding, not Reading

- Shockingly, if you find the right text/table/etc., and give it to GPT-4 or Claude 3, plus the user's query, the answer is typically produced.
 - SOTA LLMs mostly eliminate the “reading” step.
- So the challenge is finding the right knowledge in the corpus.
 - Every LLM has limited “context” and the corpus may be huge.
- Public QA systems (Copilot on web, ChatGPT, etc.) benefit from redundant knowledge availability.
 - But we often see a “needle in a haystack” challenge – one fact from one paragraph from one document in a corpus of millions.

Vector/Embedded/Semantic Search





Data Science Convention 2024

April 11, 2024
Norris Conference Center
Houston, TX

RAG-based Startups (subset)



Trieve San Francisco, CA, USA

Infrastructure for search teams building retrieval and RAG



W24

B2B

INFRASTRUCTURE



SciPhi

SciPhi is the best way to build retrieval for AI



W24

B2B



Reducto San Francisco, CA, USA

Unlocking data behind complex documents



W24

B2B

ENGINEERING, PRODUCT AND DESIGN



Superpowered AI Salt Lake City, UT, USA

API for Retrieval Augmented Generation



S22

B2B



Trellis

AI-powered Snowflake for unstructured data



W24

B2B



Buster San Francisco, CA, USA

Infrastructure to connect LLMs and databases



W24

B2B

ENGINEERING, PRODUCT AND DESIGN



Data Science Convention 2024

April 11, 2024
Norris Conference Center
Houston, TX



in the volve field, what is the total depth of well 15/9 F-7?

Show drafts



I don't have access to specific well data like the total depth of well 15/9 F-7 in the Volve field. However, some resources like research papers and oil and gas company websites might publish well information like this.

Here are some resources you can explore to find the total depth of well 15/9 F-7:

- **Technical papers:** Search for papers related to the Volve field that mention well 15/9 F-7. You can find these papers on engineering research databases or through scholarly search engines.
- **Oil and gas company websites:** If the Volve field is still operational, the company that operates it might have information about the wells on their website, although this is less likely since the field was shut down in 2016.

Check completed • No results found • [Understand the results](#)

Google Gemini



Data Science Convention 2024

April 11, 2024
Norris Conference Center
Houston, TX



You

what is the total depth of 15/9 f-7?



Copilot



Your personal and company data are protected

The total depth of the expression **15/9 f-7** depends on the context in which it is used. Let's break it down:

1. **15/9**: This fraction simplifies to **5/3**.
2. **f**: The variable "f" represents an unknown value or function. Without further context, we cannot determine its specific value.
3. **-7**: Subtract 7 from the result of the previous steps.

To calculate the total depth, we need more information about the value of "f." If you have additional context or specific values for "f," please provide them, and I'll be happy to assist further! 🤖

Learn more

1

[omnicalculator.com](#)

2

[mathalino.com](#)

3

[shotkit.com](#)



1 of 30 responses

Microsoft Copilot (web)



Data Science Convention 2024

April 11, 2024
Norris Conference Center
Houston, TX

Microsoft Copilot (web)



You

in the Volve field, what is the total depth of well 15/9 F-7?



Copilot



Your personal and company data are protected

The **Volve oilfield** is situated in Block 15/9 in the southern part of the Norwegian North Sea, at a **water depth of around 80 meters**¹². Now, let's delve into the specifics of **well 15/9 F-7**:

- **Well 15/9 F-7** is one of the wells in the Volve oilfield. It was designed for **water production** and targeted the shallow **Utsira formation**. The completion type for this well is **7" monobore perforated liner**³.
- The **water depth** (or air gap) for this well is approximately **91 meters above mean sea level (MSL)** or **54.9 meters below rotary table (RT)**⁴.
- The **bottom hole sampling depth** for well 15/9 F-7 is **3748 meters measured depth (MD) below rotary kelly bushing (RKB)**⁵.

If you have any further questions or need additional details, feel free to ask! 🗣️🌊

Learn more ¹ [offshore-technology.com](https://www.offshore-technology.com) ² norskpetroleum.no ³ discovervolve.com ⁴ discovervolve.com ⁵ discovervolve.com +1 more



1 of 30 responses



Data Science Convention 2024

April 11, 2024
Norris Conference Center
Houston, TX



Verba

The Golden RAGtriever

+ Add Documents

▲ CohereEmbedd...

📄 WindowRetrie...

🗨️ GPT4Generator

Weaviate

Documentation

75 Volve 15_9_F-7 Dril

43 Volve 15_9_F-7 Dril

71 Volve 15_9_F-7 Dril

42 15-9-f-7_nr_001_MWD

67 15-9-f-7_nr_001_MWD

43 15-9-f-7_nr_001_MWD

45 15-9-f-7_nr_001_MWD

45 15-9-f-7_nr_001_MWD

51 15-9-f-7_nr_001_MWD

42 15-9-f-7_nr

🐶 RAGtriever Chat

Online v0.3.1

Powered by Weaviate ❤️

what is the total depth of 15/9 F-7 ?

📄 Copy

The provided context does not contain a direct answer to the query 'what is the total depth of 15/9 F-7'. However, it does contain various pieces of information that could be related to well depths and measurements. The context includes several references to depths and measurements taken at different stages of the drilling process, but it does not clearly state the total depth of the well 15/9 F-7.

To determine the total depth, one would typically look for the final measured depth (MD) or total vertical depth (TVD) at the end of the drilling report. This information is usually found in the well completion report or the final well report section that summarizes the deepest point reached by the drill bit.

Since the provided context does not explicitly mention the total depth, I cannot provide a definitive answer. If you have access to the final well report or a section that specifies the total depth,

What is a vector database?

Documentation Volve 15_9_F-7 Drilling and Completion EOWR.pdf

3120000 15 100 | 19:00:00 handling 1 1 | | /3009200/ | . Blockage in cement | 04:00:00 line. RIG-E388
General 483708 195000 : 3120000 Downtime 15 | i v » e 01.10.2007 3 Malfunct f the (3560 handling equ/sys |
4s6142 | Pé's L 5180000 | 780000 | jother | | . | ! | .10. | | Malfunction of the : B O it ol e B 1 1™ S00I
Mot s il B kit RIG-E347 Pipe i : 05.10.2007 % H { Malfunction og 9 : 08:30:00 ! } ;f;ilmg equ/sys 527091 |
LD 650000 | 3120000 5,0 R e e ERT Gnir | — T R e NN B f G e | | RIG-E342 Drillfloor | b handl system k. B
[Total [4217500 https://partners.statoil.com/sites/78 7fd9ce-3096-4aa9-a6al-9b57e5d2a94a/FWR
reporting/Document library/FWR_NO 15_9-F-7[1] Drilling and Completion.doc FINAL WELL REPORT Doc no
StatoHydro Drilling and Completion 3 5 oE 36 Licence no: PL046BS zD;ot; T Oe" no o Well: NO 15/9-F-7 | 34
Time distribution Length Actual Section Start time m hrs | days NO 15/9-F-7 36" 13.06.2007 18:30 0,0 30 01
NO 15/9-F-7 36" 14.09.2007 21:00 640 900 38| 690 29| 1445| 60| 969 NO 15/9-F-7 17 172" 20.09.200721:30 |
7060 | 1520 | 63 | 830 | 35| 1928 | 80| 843 NO 15

/9-F-7 12 1/4" 28.09.2007 22:15 630 | 1045| 44| 85| 34| 1263| 53| 863 NO 15/9-F-7 COMPL_WPREP_|
04.10.2007 04:30 T 0| o o i) T O P Y B W e NO 15/9-F-7 COMPL_SCRUN_| 05.10.2007 04:00 PN el D B e o e
D) NO 15/9-F-7 COMPL_TBGRN_| 08.10.2007 12:45 200 EAA 0N1 8| 32,00 13| 0 62,510 1 2,6/[E100,0 Sum 5015 |
209 | 3390 141| 6319 264 https://partners.statoil.com/sites/787fd9ce-3096-4aa9-a6al-9b57e5d2a94a/FWR
reporting/Document library/FWR_NO 15_9-F-

7[1] Drilling and Completion.doc FINAL WELL REPORT Drilling and Completion Licence no: PL046BS \Well: NO
15/9-F-7 – StatoilHydro Date Rev no [16 of 56 £009-07-15 0 1 Time distribution 15/8-F-7 36" 15/9-F-7 36"
158F-717 112" 158-F-7121/4" Section 15/9-F-7 COMPL_WPREP 15/9-F-7 COMPL_SCRUN 15/9-F-7 COMPL_TBGRN Total 0o
The graph above is based on the following details: 50 100 150 Days 200 250 mUptime 0 Quality time @ Downtime



Microsoft Copilot (Azure)

where is the Volve field?



The Volve field is located in the southern part of the North Sea, approximately eight kilometres north of Sleipner Øst ¹. It is a relatively small oil discovery, situated on a structural high within the Sleipner area, and has geological similarities with the neighboring structures Loke and Sleipner Øst ^{2 3}.

✓ 3 references

1 15-9-F-14_nr_003_FWR_MUD.pdf - Part 1

2 15_9_F_15_Recommendation_to_Drill.pdf - Part 1

3 DRILLING_PROGRAMME_1.PDF - Part 1



Microsoft Copilot (Azure) – hybrid search

what is the total depth of 15/9 f-7?



The requested information is not available in the retrieved data. Please try another query or topic.



Data Science Convention 2024

April 11, 2024
Norris Conference Center
Houston, TX

LlamaIndex

```
(venv) [josh@nostromo llamaindex]$ llamaindex-cli rag --verbose -q "what is the total depth of 15/9 F-7?"
Saving/Loading from persist_dir: /tmp/llama_index/rag_cli
> Running module query with input:
query_str: What is the total depth of 15/9 F-7?

> Running module retriever with input:
input: What is the total depth of 15/9 F-7?

> Running module summarizer with input:
query_str: What is the total depth of 15/9 F-7?
nodes: [NodeWithScore(node=TextNode(id_='529daa57-b457-45be-bdef-453260263042', embedding=None, metadata={'page_label': '73', 'file_name': 'FWR MUD F-15_ABC_nr_004.pdf', 'file_path': '/data/josh/corpus/files...

The total depth of 15/9 F-7 is not provided in the context information.
```



Data Science Convention 2024

April 11, 2024
Norris Conference Center
Houston, TX

LlamaIndex

```
(venv) [josh@nostromo llamaindex]$ llamaindex-cli rag --verbose -q "what is the TD of 15/9 F-7?"

Saving/Loading from persist_dir: /tmp/llama_index/rag_cli
> Running module query with input:
query_str: What is the time difference between 15/9 F-7?

> Running module retriever with input:
input: What is the time difference between 15/9 F-7?

> Running module summarizer with input:
query_str: What is the time difference between 15/9 F-7?
nodes: [NodeWithScore(node=TextNode(id_='3cdd5a68-093d-483e-9b54-89f5d18a5307', embedding=None, metadata={'page_label': '9', 'file_name': 'FWR_Completion_NO 15-9-F-15 D.pdf', 'file_path': '/data/josh/corpus/...

The time difference between 15/9 F-7 is 13 days.
```

what is the total depth of 15/9 F-7?

The total depth of 15/9 F-7 is 1083 m MD / 1077 m TVD ([Final Well Report for Well 15/9-F-7](#)).

- We use an agent architecture composed of some domain-general and domain-specific search agents.
- Each agent gives some evidence towards the answer.
- Example agents: geography, geology, wells/basins/formations/fields/..., SPE Discipline, dates, file types, keywords, embeddings, etc.

Final Well Report for Well 15/9-F-7

[Open in new window](#)

Ask a question related to this document...

| | |
|--------------------------------------|---------------------------------------|
| Prospect | Voive |
| Licence | PL 046BS |
| <u>Template centre coordinates:</u> | |
| Latitude | 58°26'29,807" N |
| Longitude | 01°53'14,929" E |
| UTM Co-ordinates | 6 478 563,523 m N 435 050,021 m E |
| <u>Slot centre coordinates N-2H:</u> | |
| Latitude | 58° 26' 29,869" N |
| Longitude | 01° 53' 14,858" E |
| UTM Co-ordinates | 6 478 565,478 m N 435 048,9071 m E |
| RKB to MSL | 54,9m |
| Water depth | 91mMSL |
| Spud date | 17 th September 2007 |
| TD reached | 3 rd October 2007 |
| Total Depth | 1083 m MD / 1077 m TVD |
| Total Drilling Days | 17 |
| Final Well Status | Completed for water production |




Data Science Convention 2024

April 11, 2024
Norris Conference Center
Houston, TX

what is the total depth of 15/9 F-7?

The total depth of 15/9 F-7 is 1083 m MD / 1077 m TVD ([Final Well Report for Well 15/9-F-7](#)).

| Document | Score |
|--|-------|
| Individual Well Drilling and Completion Program - Well 15/9-F-7 - Volve | 1.00 |
| 15_9_F7 Well Log: BHPR-GR-MECH-MWD | 0.71 |
| Final Well Report for Well 15/9-F-7  | 0.70 |
| Individual Well Drilling & Completion Program for Well 15/9-F-7 at Volve | 0.70 |
| Recommendation to Drill: Volve Field, Well NO 15/9-F-7 and NO 15/9-F-9 | 0.67 |
| 15/9-F-7 Well DLIS File: Raw GR and REMP MWD Logs | 0.53 |



Data Science Convention 2024

April 11, 2024
Norris Conference Center
Houston, TX

what is the total depth of 15/9 F-7?

The total depth of 15/9 F-7 is 1083 m MD / 1077 m TVD ([Final Well Report for Well 15/9-F-7](#)).

- Identified query: what is the total depth of 15/9 F-7?
- Extracted keywords for keywords_queries: ['total depth', '15/9 F-7', 'F-7']
- Search result for keywords ('total depth',) from queries: 20 docs
- Search result for keywords ('15/9 F-7',) from queries: 20 docs
- Search result for keywords ('F-7',) from queries: 20 docs
- Search result for keywords ('total depth', '15/9 F-7') from queries: 15 docs
- Search result for keywords ('total depth', 'F-7') from queries: 20 docs
- Search result for keywords ('15/9 F-7', 'F-7') from queries: 20 docs
- Classified taxnodes_queries as {'SPE Disciplines|Well Drilling|Drilling Operations'}
- Search result for taxnode SPE Disciplines|Well Drilling|Drilling Operations from queries: 20 docs
- Search result for embedding from prompt: 20 docs
- Search result for generated query {'q': 'title:"15/9 F-7" OR pagetext:"15/9 F-7" AND pagetext:"total depth"', 'filters': ['taxnodes:"SPE Disciplines|Well Drilling|Drilling Operations"']} from queries: 15 docs
- Answer found in snippets



Data Science Convention 2024

April 11, 2024
Norris Conference Center
Houston, TX

Conversation Tracking

How many intervals of Well 15/9-F-1 C were perforated on April 10, 2014?

Two intervals of Well 15/9-F-1 C were perforated on April 10, 2014. ([Daily Drilling Report for Wellbore 15/9-F-1 C - April 10, 2014](#))

what depths?

The intervals perforated in Well 15/9-F-1 C on April 10, 2014, were at:

- From 3973 m to 3991 m
- From 3936 m to 3954 m ([Daily Drilling Report for Well 15/9-F-1 C: Perforation and Pressure Testing Operations](#))



Other kinds of knowledge in the corpus

tell me about BSEE inspections in the US title code 30

To learn about BSEE inspections in the US under Title 30 of the U.S. Code, the most relevant documents are "PART 250 - OIL AND GAS AND SULPHUR OPERATIONS IN THE OUTER CONTINENTAL SHELF" (document: [PART 250 - OIL AND GAS AND SULPHUR OPERATIONS IN THE OUTER CONTINENTAL SHELF](#)) and "PART 842 - FEDERAL INSPECTIONS AND MONITORING" (document: [PART 842 - FEDERAL INSPECTIONS AND MONITORING](#)). These documents are directly related to regulatory frameworks and inspections pertaining to oil, gas, and sulphur operations, as well as federal inspection and monitoring activities, under the jurisdiction specified in Title 30.



Data Science Convention 2024

April 11, 2024
Norris Conference Center
Houston, TX

what's the definition/limit for sour gas?

Natural gas that contains more than 4 ppmv of hydrogen sulphide (H_2S) is commonly referred to as "sour" gas.
([Sour Gas: Challenges, Safety Training, and Sweetening Processes](#))

what can be used to treat it?

To treat sour gas, various processes and solvents can be used, depending on the conditions and requirements. The processes and solvents include:

- Regenerative chemical solvents like alkanolamines (primary, secondary, or tertiary amines such as Monoethanolamine (MEA), Diethanolamine (DEA), and Methyldiethanolamine (MDEA)) for sour gas sweetening ([Sour Gas Sweetening: Overview, Process Equipment, and Solvent Selection](#)) [4] .
- The Claus process for recovering elemental sulfur from gaseous hydrogen sulfide ([Sour Gas: Challenges, Safety Training, and Sweetening Processes](#)).
- The GT-SPOC process, a Claus type process with a catalytic section for sulfide to sulfur conversion ([Sour Gas: Challenges, Safety Training, and Sweetening Processes](#)).



Should the LLM be Domain-Tuned?

- We are working towards building an LLM that's finetuned specifically for the Energy domain.
- How this helps:
 - Better understanding of the query: keywords, well names, units of measure, etc.
 - Better retrieval using more accurate embeddings.
 - Better generation of the answer given the knowledge context.
- How we will do it:
 - Train a foundation LLM (e.g., Mixtral or Llama-2) with questions and answers appropriate for the domain.
 - I.e., guide the LLM to mimic expected behavior as seen in the training set.



Data Science Convention 2024

April 11, 2024
Norris Conference Center
Houston, TX

Rapid Advances

- This technology is highly dynamic.
- What is commonplace now was not possible last year.
 - Particularly in terms of LLM understanding, generation, and embeddings.
- As of now, effective RAG or “knowledge recovery” depends on a mixture of horizontal and vertical solutions.