

Evidence Contextualization and Counterfactual Attribution For ConvQA over Heterogeneous Data with RAG Systems

Rishiraj Saha Roy, Joel Schlotthauer, Chris Hinze,
Andreas Foltyn, Luzian Hahn, and Fabian Küch

{rishiraj.saha.roy, joel.Schlotthauer, chris.hinze, andreas.foltyn, luzian.hahn, fabian.kuech}@iis.fraunhofer.de

RAG over enterprise wikis is challenging

2024-10-02 Meeting Notes

Today we will talk about the progress of the project on retrieval augmented generation.

Agenda

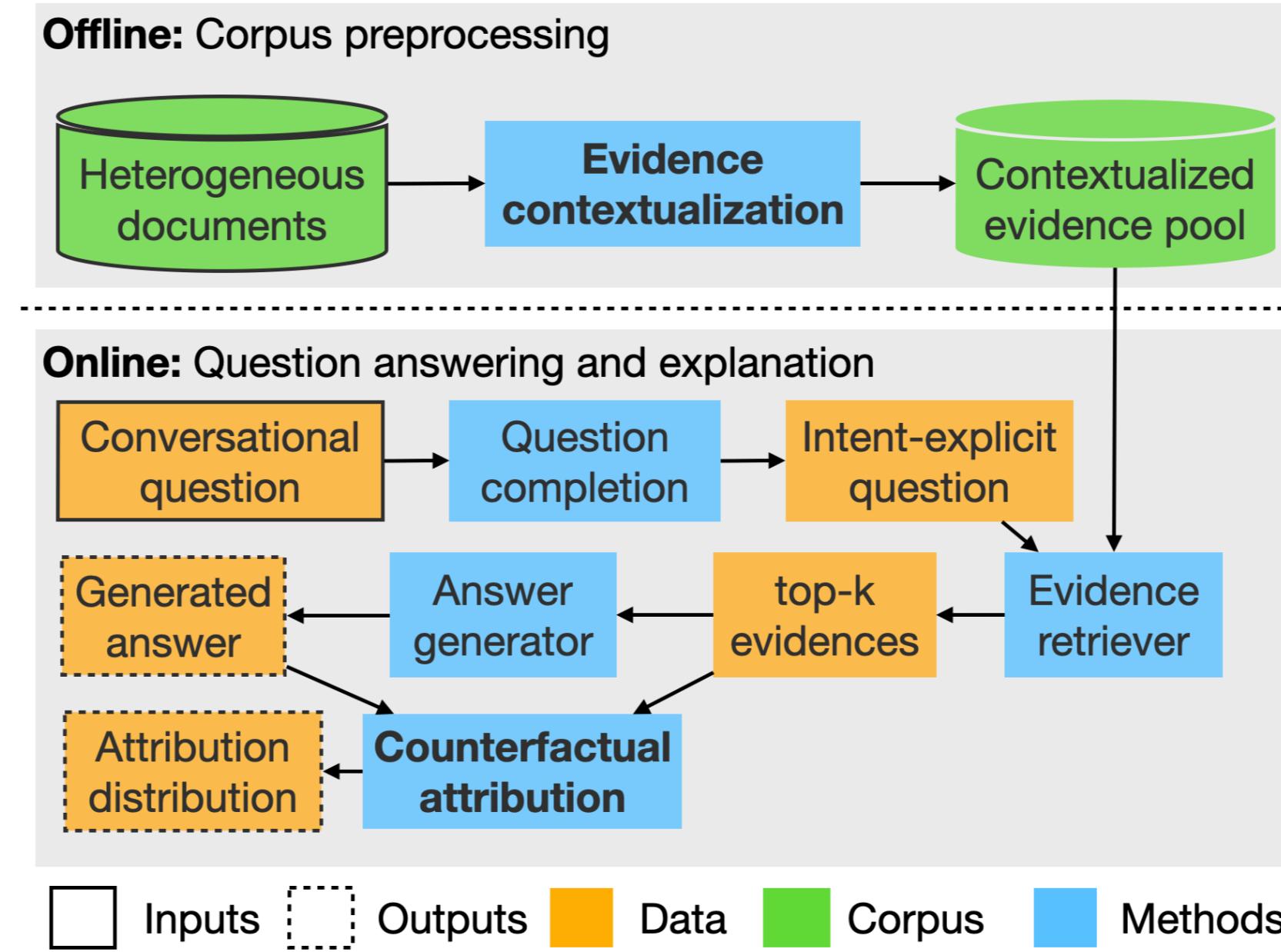
- We'll first do a basic round of RAG team updates in this month's meeting
- Everyone will report what has been done, and the to-dos

Member	Task	Action items	Time needed	Notes
Bob	Basic FE and BE	Follow-up q in UI	3 days	Currently manual
Alice	Similarity function	Fine-tune with gpt4o*	1 week	Now w/ embed cos
Trudy	Verbalizations	Batch configs*	6 hours	Running superbly

* Alice and Trudy to fix long-standing embedding error with openxt strings

Question: todo for alice in oct rag meeting?

We propose RAGONITE, and ...

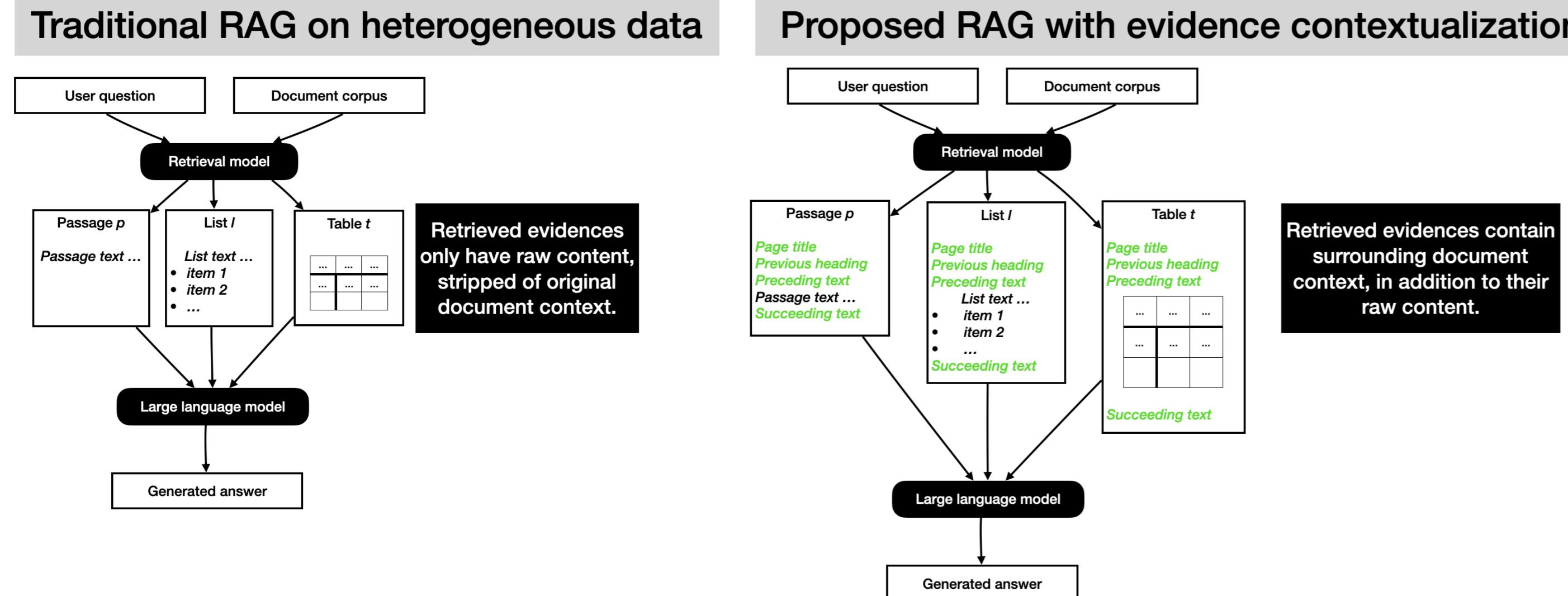


... a new benchmark: ConfQuestions

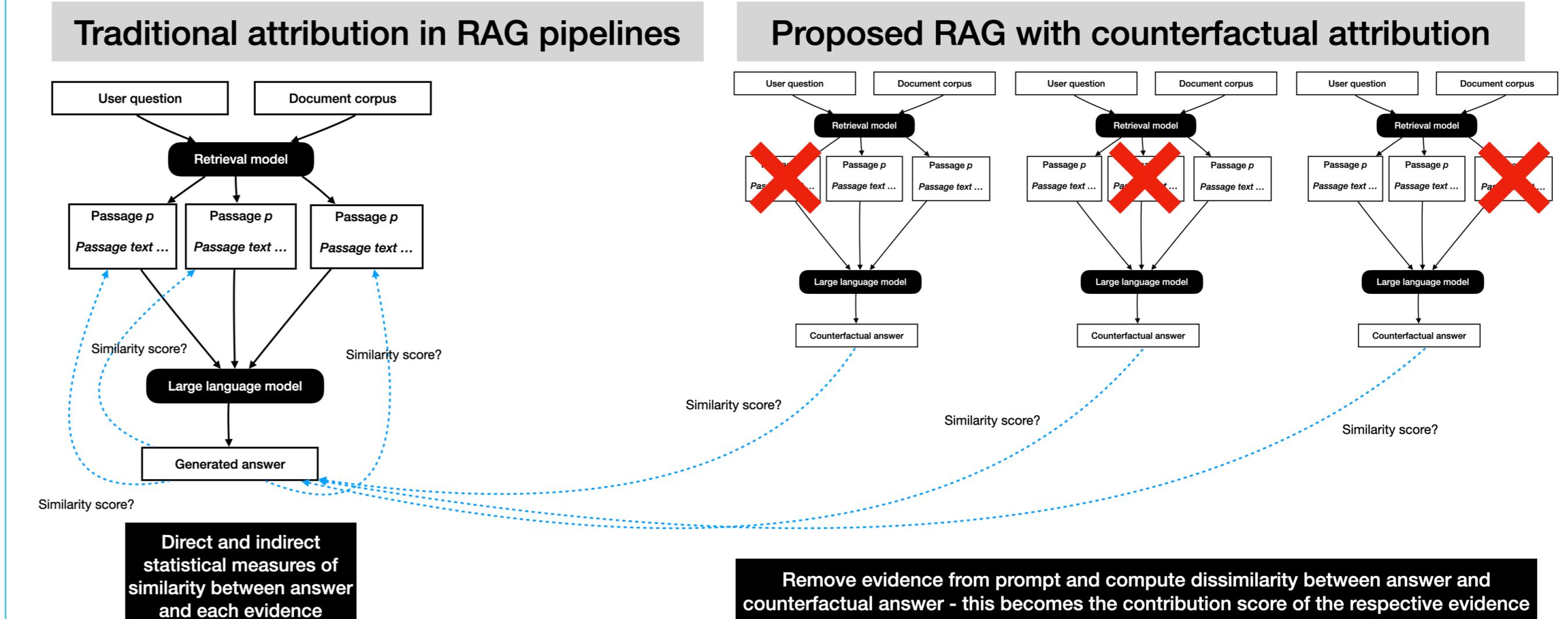
ConfQuestions is suitable for evaluating

- ✓ Conversational question answering
 - ✓ Complex question answering
 - ✓ RAG on heterogeneous enterprise documents
- ConfQuestions contains**
- ✓ 50 human conversations with 300 questions
 - ✓ Corresponding human-generated completed questions
 - ✓ Gold URL and answers: entities, phrases, lists, passages
 - ✓ Each question in English and German
 - ✓ Metadata: Answer-source, Complexity type
 - ✓ 215 Confluence pages as corpus

Idea sketch: Evidence contextualization



Idea sketch: Counterfactual attribution



RAGONITE screenshot

RAGONITE

User question: What was the BIOS and Build versions used for Dell Optiplex 7040 in the OpenXT 9.0 measurement tests?
User response: Q. What was the BIOS and Build versions used for Dell Optiplex 7040 in the OpenXT 9.0 measurement tests?
Ragonite response: The BIOS version used for Dell Optiplex 7040 in the OpenXT 9.0 measurement tests was 1.14.0, and the build versions were 6662 and 6571 [Sources 1, 2, 7].
User response: And what about TPM?
Ragonite response: Q. What was the TPM version used for Dell Optiplex 7040 in the OpenXT 9.0 measurement tests?
User response: The TPM version used for Dell Optiplex 7040 in the OpenXT 9.0 measurement tests was 2.0 [Sources 1, 2, 3].
Ragonite response: The Dell Optiplex 9010 passed the following OpenXT 9.0 tests for the legacy system:
1. OTA to OpenXT 9.0.2 (\$6768) [Sources 1, 2]
2. OpenXT 9.0.2-pre (\$6768) [Sources 1, 2]
User response: Which machines were tested with both legacy and EFI for OXT 9.0.2?
Ragonite response: Answer explanation with counterfactual attribution:
1. OTA to OpenXT 9.0.2 to cluster 1 [Evidence 1, 2]
2. OpenXT 9.0.2-pre to cluster 2 [Evidence 3, 4]
3. OpenXT 9.0.2-pre to cluster 3 [Evidence 5, 6]
4. Attributed 4.5% to cluster 4 [Evidence 7]
5. Attributed 0.3% to cluster 5 [Evidence 8]
6. Attributed 0.2% to cluster 6 [Evidence 9]
7. Attributed 0.2% to cluster 7 [Evidence 10]
User response: Ask RAGONITE something about Confluence OpenXT.
Ragonite response: Coming back to OXT 9.0, did the Optiplex 7060 pass the UEFI OTA upgrade 8.0.1 to 9.0.0 test?
User response: 1. Question input box
Ragonite response: 12. Follow-up question suggestion

ConfQuestions excerpt

```
"conv_id": "1",
"turns": [
  {
    "turn_id": "1",
    "q_type": "complex",
    "q_en": "What was the BIOS and Build versions used for Dell Optiplex 7040 in the OpenXT 9.0 measurement tests?",
    "q_de": "Welches BIOS und welche Build Versionen wurden für den Dell Optiplex 7040 im Zuge der OpenXT 9.0 measurement tests eingesetzt?",
    "completed_q_en": "What was the BIOS and Build versions used for Dell Optiplex 7040 in the OpenXT 9.0 measurement tests?",
    "completed_q_de": "Welches BIOS und welche Build Versionen wurden für den Dell Optiplex 7040 im Zuge der OpenXT 9.0 measurement tests eingesetzt?",
    "a_url": "https://openxt.atlassian.net/wiki/spaces/TEST/pages/761823271/OpenXT+9.0+Measurement+Test",
    "a_source": "table",
    "a": "Build version: 6662 and BIOS version: 1.14.0"
  },
  {
    "turn_id": "2",
    "q_type": "complex",
    "q_en": "And what about TPM?",
    "q_de": "Und was ist mit TPM?",
    "completed_q_en": "What was the TPM version used for Dell Optiplex 7040 in the OpenXT 9.0 measurement tests?",
    "completed_q_de": "Welche TPM Version wurde für den Dell Optiplex 7040 im Zuge der OpenXT 9.0 measurement tests genutzt?",
    "a_url": [
      "https://openxt.atlassian.net/wiki/spaces/TEST/pages/761823271/OpenXT+9.0+Measurement+Test"
    ],
    "a_source": "table",
    "a": "Version 2.0"
  },
  {
    "turn_id": "3",
    "q_type": "complex",
    "q_en": "Which OpenXT 9.0.2 tests did Dell 9010 pass for the legacy system?",
    "q_de": "Welche OpenXT 9.0.2 Tests hat der Dell 9010 für veraltete (legacy) Systeme bestanden?",
    "completed_q_en": "Which OpenXT 9.0.2 measurement tests did Dell Optiplex 9010 pass for the legacy system?",
    "completed_q_de": "Welche OpenXT 9.0.2 Tests hat der Dell 9010 für veraltete (legacy) Systeme bestanden?",
    "a_url": [
      "https://openxt.atlassian.net/wiki/spaces/TEST/pages/761823271/OpenXT+9.0+Measurement+Test"
    ],
    "a_source": "table",
    "a": "Version 2.0"
  }
]
```

Evaluation results

Retrieval P@1 for CONTEXTUALIZATION methods						Accuracy for ATTRIBUTION methods		
None	w/ Page title	w/ Preceding heading	w/ Preceding text	w/ Succeeding text	w/ All context	Standard	CF	CF with Clusters
0.398	0.483	0.448	0.453	0.460	0.528	0.772	0.791	0.799
Answer relevance for CONTEXTUALIZATION methods								
None	w/ Page title	w/ Preceding heading	w/ Preceding text	w/ Succeeding text	w/ All context	Standard	CF	CF with Clusters
0.388	0.477	0.435	0.445	0.404	0.529			

Take-home message: Contextualizing evidence with surrounding document text, and explaining answers with counterfactual attribution, are vital RAG add-ons



Scan the QR code for all info on project... ☺