Caravaggio: A Multimodal Data Integration Framework for Complex Question Answering

Gerardo Vitagliano, Ziyu Zhang, Ferdinand Kossman, Michael Cafarella
MIT CSAIL

Problem Definition

We are interested in an AI system capable of answering questions that require complex combinations of multimodal entries drawn from a database as evidences. The database may contain data in a variety of formats and modalities.

Format

- Structured (tabular)
- Semi-structured
- Unstructured

Modality

- Text
- Image
- Tables (numeric)
- Time Series
- Domain-specific modes
- Audio/video

Previous Work

- Existing complex question answering (QA) work focus on the scenario where the data corpus is text only and simple paragraph/sentence-based segmentation strategies work well.
- Existing multimodal data processing work focus on the search/retrieval or single document comprehension use-cases where the task is simple and well-formatted.
- Complex multimodal QA calls for capabilities beyond existing methods.

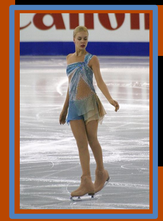
Example

During the 2015-16 figure skating season, what is the name of the blonde-haired contestant in ladies' events who also competed in the Junior Worlds event?

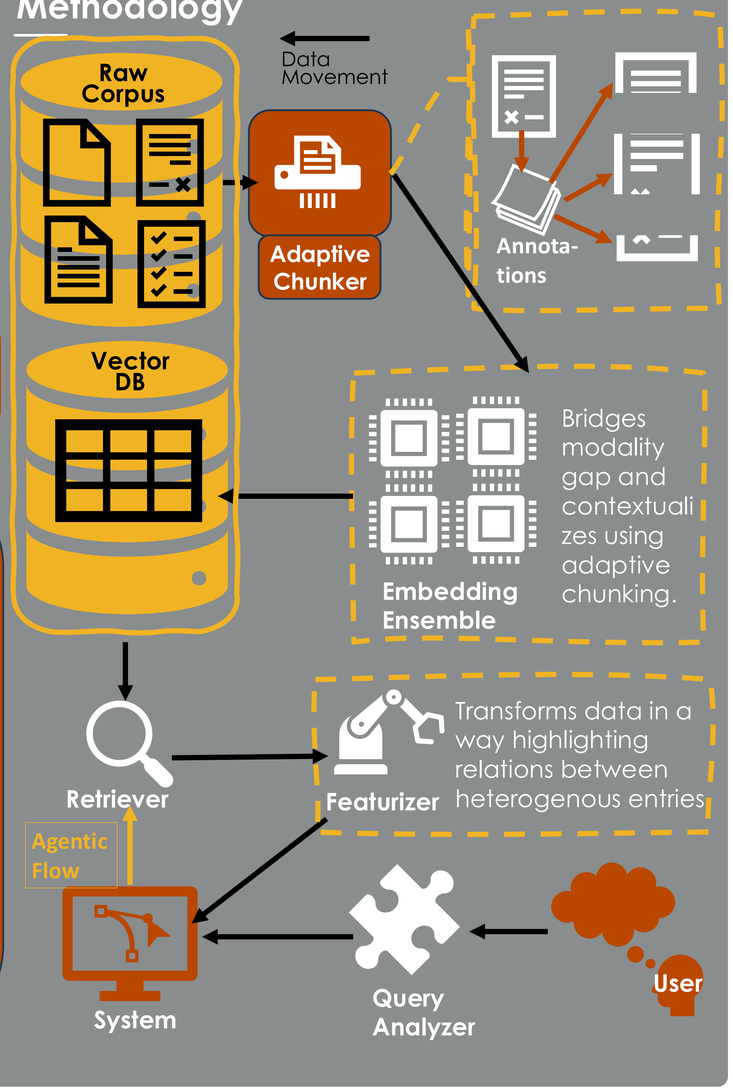
Maria Sotskova

Knowledge Corpus

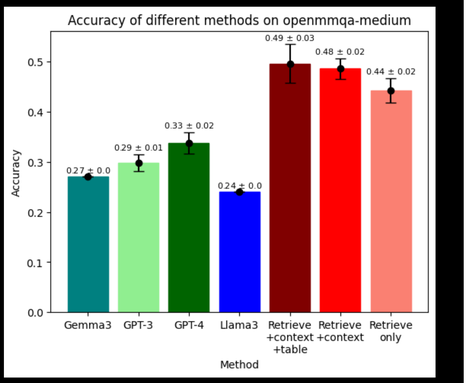
```
Mirai Nagasu,13,193.86,Four Continents
Marin Honda,14,192.98,Junior Worlds
Maria Sotskova,15,188.72,Junior Worlds
Yuna Shiraiwa,16,186.80,JGP Spain
Alisa Fedichkina,17,186.38,JGP Spain
Yuka Nagai,18,185.93,Bavarian Open
```



Methodology



Preliminary Results



Next Steps

- Continue to improve the adaptive chunker and featurizer
- Based on state-of-the-art unifying IR methods, develop better solutions for the embedding ensemble and query analyzer to further improve performance
- More extensive comparison with previous works on real-world use-cases

Acknowledgements

We thank Sage Bionetworks for their collaboration with us on the use-case. We are grateful for support from the DARPA AKSEM project (Award HR00112220042), the ARPA-H Biomedical Data Fabric project.