Entity Notes

Methods for connecting and assigning attributes to organizational entities and individuals.

# LLMs and KGs

Language models and Knowledge Graph-based semantic processing vary notably not only in formats but in basic properties of expected inputs and derived results. Typical data store and indexing technologies have distinct architectures, as vector, graph, or relational databases or formatted files.

Meanwhile, exported chat logs, meeting transcripts, survey data, Social Media posts, diagrams and publications institute rather specific content structures. The identifiers for participants and accounts, as representative of individuals, have strict rules, formats, and privacy requirements in each information silo, repeating infrequently. The designators for organizational entities mutate, as do their participation levels and roles. Membership rosters update frequently.

We will provide a practical methodology, compatible with standards and legacy formats, to address robust construction and flexible interchange of mappings of identifies and attribute labels between technology-specific data stores and information silo formats used to support diverse community environments and entity contexts.

A sample implementation will support a cross-section of LLM and KG processing steps by mapping identifiers, designators, and labels from several data sources immediately relevant to the Singularity NET ecosystem.

# Knowledge Graph

Model for Knowledge Graph Reasoning [Ultra](https://www.marktechpost.com/2023/10/30/meet-ultra-a-pre-trained-foundation-model-for-knowledge-graph-reasoning-that-works-on-any-graph-and-outperforms-supervised-sota-models-on-50-graphs)

What is a (Knowledge) Graph https://www.youtube.com/watch?v=NZN2jVyRKVc&list=PLar5iR7mhb4dJHDSjmeo6W7HomHBSZf9t

https://iccl.inf.tu-dresden.de/web/KG2020/en

Turing <https://www.turing.ac.uk/research/interest-groups/knowledge-graphs>

<https://github.com/turing-knowledge-graphs/teaching/tree/main/bcu>

https://www.youtube.com/@AshleighFaith

<https://blog.google/products/search/introducing-knowledge-graph-things-not>

automatically annotating DBpedia https://www.dbpedia-spotlight.org/api

https://github.com/dbpedia-spotlight/dbpedia-spotlight/wiki

Semantic search https://scholarai.io/knowledge-graphs-vector-embeddings-and-semantic-search

Open Work <https://backdroplabs.notion.site/Open-Work-Graph-1504d7227e0340f98f334cfc36e901b6>

<https://enterprise-knowledge.com/synergizing-knowledge-graphs-with-large-language-models-llms>

#### Neo4j

<https://neo4j.com/developer-blog/knowledge-graphs-llms-multi-hop-question-answering>

LLM KG https://neo4j.com/blog/unifying-llm-knowledge-graph

RAG KG https://go.neo4j.com/WBR-240423-Intro-to-RAG\_Registration.html

Schema https://www.schemaapp.com/schema-markup/how-to-leverage-your-schema-org-knowledge-graph-for-llms-like-chatgpt

### RDF

https://www.w3.org/2001/sw/wiki/Category:RDF\_Generator

https://google.github.io/digitalbuildings/tools/rdf\_generator

### Entity Matching Resolution

https://towardsdatascience.com/is-this-you-entity-matching-in-the-modern-data-stack-with-large-language-models-19a730373b26

https://github.com/megagonlabs/ditto

https://github.com/zinggAI/zingg

https://github.com/TeeanRonson/EntityResolution

https://github.com/Valires/Awesome-Entity-Resolution

Name Resolution <https://github.com/TranslatorSRI/NameResolution>

### 

### Dictionary APIs

https://medium.com/@martin.breuss/finding-a-useful-dictionary-api-52084a01503d

https://www.jsonapi.co/public-api/Free%20Dictionary

https://rapidapi.com/blog/dictionary-apis/

### LLM Custom Vocabulary

https://emaggiori.com/chatgpt-vocabulary

keyword boosting https://docs.symbl.ai/docs/asr-custom-vocabulary

### Out of Vocabulary (OOV)

https://www.chatgptguide.ai/2024/03/01/what-is-out-of-vocabulary-oov-words-llms-explained

Non-string https://medium.com/@murongweibo2/how-to-make-llm-agent-recognize-non-string-variables-9bc92d6bdb8d

Controlled vocabularies https://www.loc.gov/librarians/controlled-vocabularies

https://www.controlledvocabulary.com/examples.html

### Frequency Lists

https://www.routledge.com/A-Frequency-Dictionary-of-Contemporary-American-English-Word-Sketches-Collocates-and-Thematic-Lists/Davies-Gardner/p/book/9780415490634

Word frequency effect https://en.wikipedia.org/wiki/Word\_frequency\_effect

4K Frequency https://simple.wikipedia.org/wiki/Wikipedia:Word\_frequency

https://www.teacher-joe.com/chat-gpt-for-language-learning/chatgpt-english-vocabulary

SAT https://blog.prepscholar.com/sat-vocabulary-words

SEO stop words list https://cseo.com/blog/seo-stop-words

<https://github.com/openai/tiktoken>

### Ontology and Schema

https://genai.stackexchange.com/questions/641/match-llm-output-to-fixed-ontology

https://ontologforum.com/index.php/OntologySummit2024

Person Schema.org https://schema.org/Person

Terminology Registries https://bartoc.org/registries

https://en.wikipedia.org/wiki/Microdata\_(HTML)

### Fine Tuning

https://platform.openai.com/docs/api-reference/fine-tuning/object

https://openai.com/index/introducing-improvements-to-the-fine-tuning-api-and-expanding-our-custom-models-program

Data https://cookbook.openai.com/examples/chat\_finetuning\_data\_prep

https://www.finetuna.io

<https://towardsdatascience.com/dive-into-lora-adapters-38f4da488ede>

## RAG

knowledge graph KG Neo4j demo https://github.com/leannchen86/graph-rag-ai-assistant

RAG https://stackoverflow.blog/2023/10/18/retrieval-augmented-generation-keeping-llms-relevant-and-current

Retrieval RAG https://fullstackretrieval.com

https://cookbook.openai.com/examples/rag\_with\_graph\_db

Chunk https://chunkviz.up.railway.app

5 Levels Of Text Splitting https://www.youtube.com/watch?v=8OJC21T2SL4

https://docs.bito.ai/bitos-ai-stack/llm-tokens#id-5-strategies-to-beat-token-limits

Semantic chunking https://medium.com/@sthanikamsanthosh1994/enhancing-rag-efficiency-the-power-of-semantic-chunking-844f9cfbdd0b

retrieval through completion api https://community.openai.com/t/using-retrieval-through-completion-api/592995

influence when RAG fetches https://community.openai.com/t/can-i-control-or-influence-when-rag-fetches-external-data/678366

RAG Really Dead https://www.youtube.com/watch?v=zgP6aWRQBb0

RAG And Long Context Lance Martin LangChain https://docs.google.com/presentation/d/1mJUiPBdtf58NfuSEQ7pVSEQ2Oqmek7F1i4gBwR6JDss

RAG From Scratch https://www.youtube.com/watch?v=wd7TZ4w1mSw

https://github.com/langchain-ai/rag-from-scratch/blob/main/rag\_from\_scratch\_1\_to\_4.ipynb

Cohere Retrieval Augmented Generation https://cohere.com/blog/command-r

Vectara https://vectara.com

HHEM https://github.com/vectara/example-notebooks/blob/main/notebooks/using-hhem-with-RAG.ipynb

RAG 4 Lines Of Code https://www.clarifai.com/blog/clarifai-release-10.1

Berkeley EECS Teaching Assistant https://events.berkeley.edu/eecs/event/237627-eecs-teaching-demo-building-an-ai-teaching

Domain Specific RAG RAFT https://gorilla.cs.berkeley.edu/blogs/9\_raft.html

## AI Agents

List https://darkvisitors.com/agents

Framework https://www.crewai.io

DAO AI agents https://medium.com/@idrees535/decentralized-intelligence-the-intersection-of-daos-and-autonomous-agents-f2c713abb0fb

Agents https://medium.com/@learn-simplified/why-entire-ai-field-is-headed-towards-ai-agents-a268ac9661ed

AI https://www.kiwitech.com/blog/web3-and-ai-a-collaboration-and-innovation-crossroads/

AI Crypto AI token https://cointelegraph.com/magazine/moral-responsibility-how-blockchain-can-improve-ai-trustworthiness

AI-powered Token https://tokenengineering.net/event/iup\_gv97y

AI HR hiring interview https://oku.trade/about.html

### Identity in Web3

https://twitter.com/shreyjaineth/status/1562603221650411521

Identity https://www.disco.xyz

Decentralized identity storage https://github.com/trustfractal/protocol

KG DF4 Proposals

Semantic Web and Knowledge Engineering

<https://scholarai.io/knowledge-graphs-vector-embeddings-and-semantic-search>

## 

## Notes:

Ambassador Program Sentiment Analysis [Report May 2024](https://docs.google.com/document/d/1oPzXTi8ziYXB-P1FjF-IUDVLNJyzggtJmAtd6QoRO7A)