Generative AI and Symbolic Knowledge Representations
LLMs, Knowledge and Reasoning
3

Damir Cavar & Billy Dickson

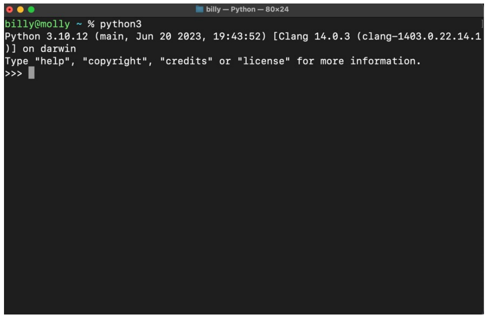
ESSLLI 2024

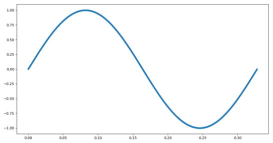
July 2024

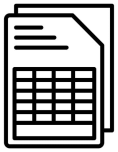
Continuation

- LLMs and NLP
- Code examples:
  - Vectorization and embeddings

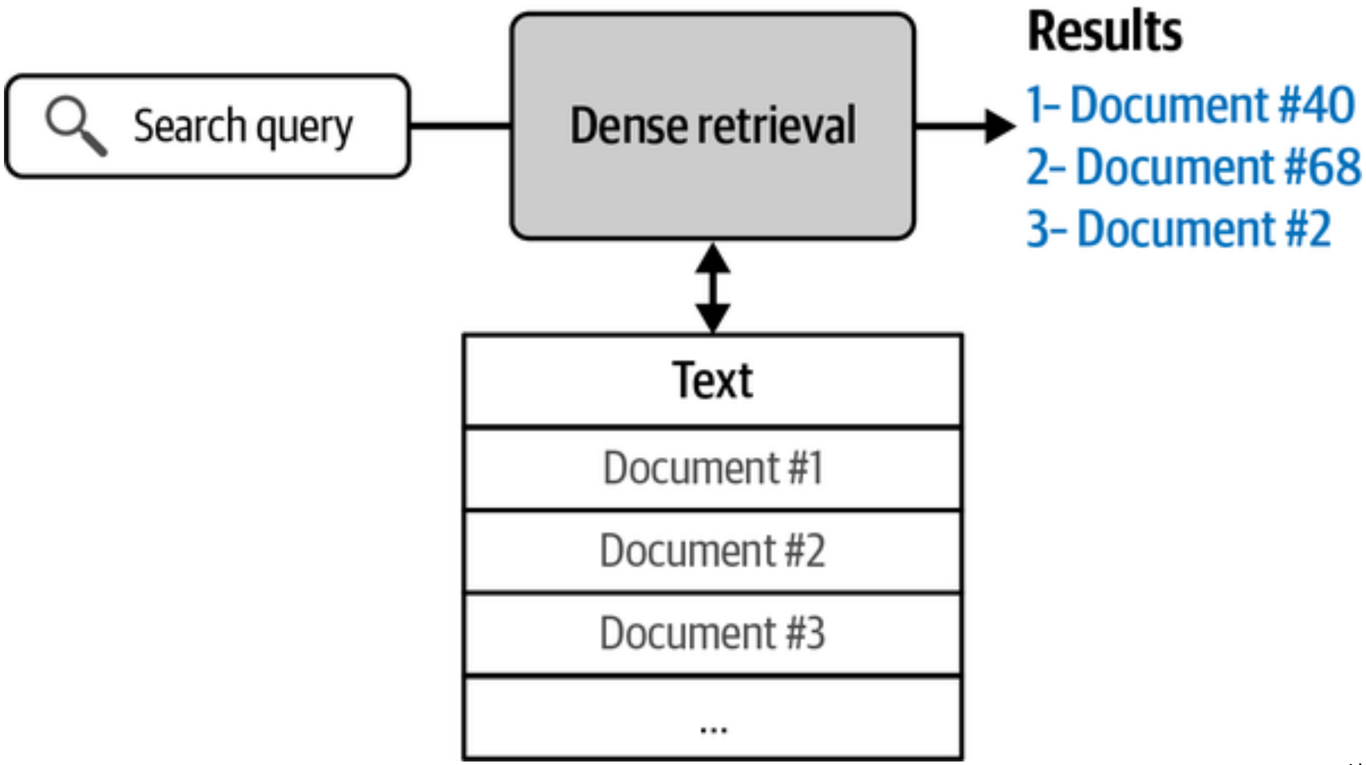
LLMs and Tools



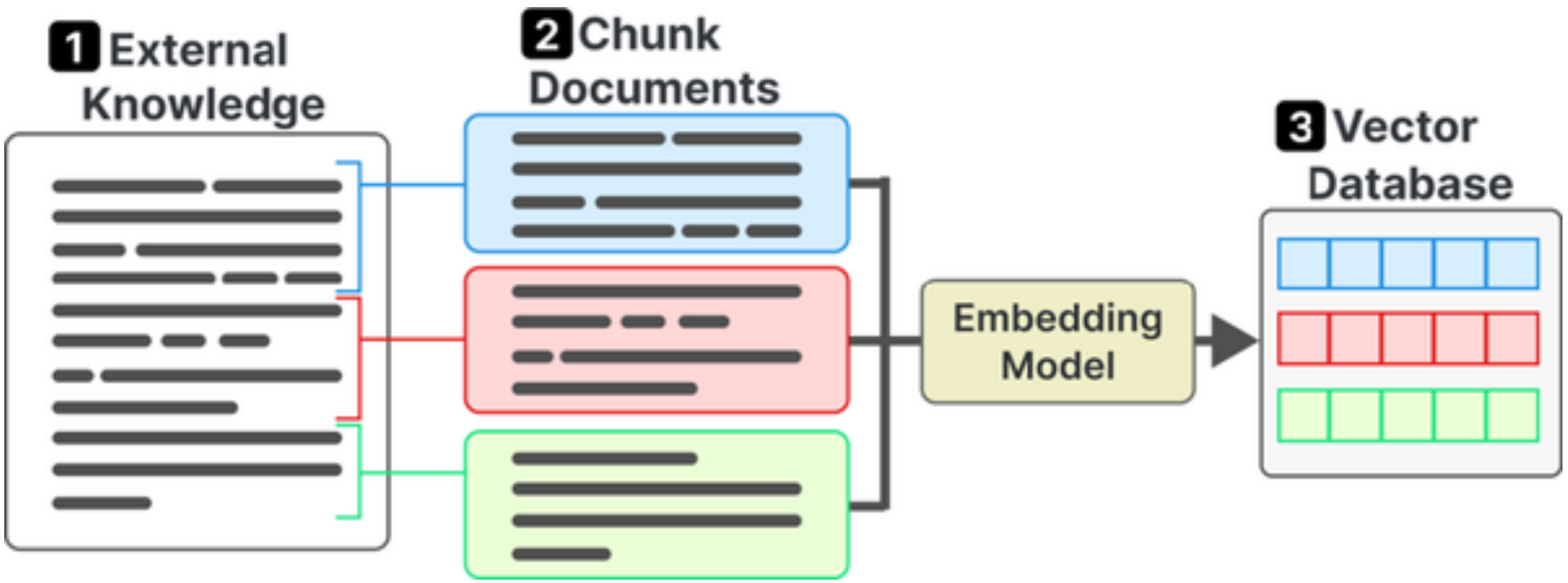




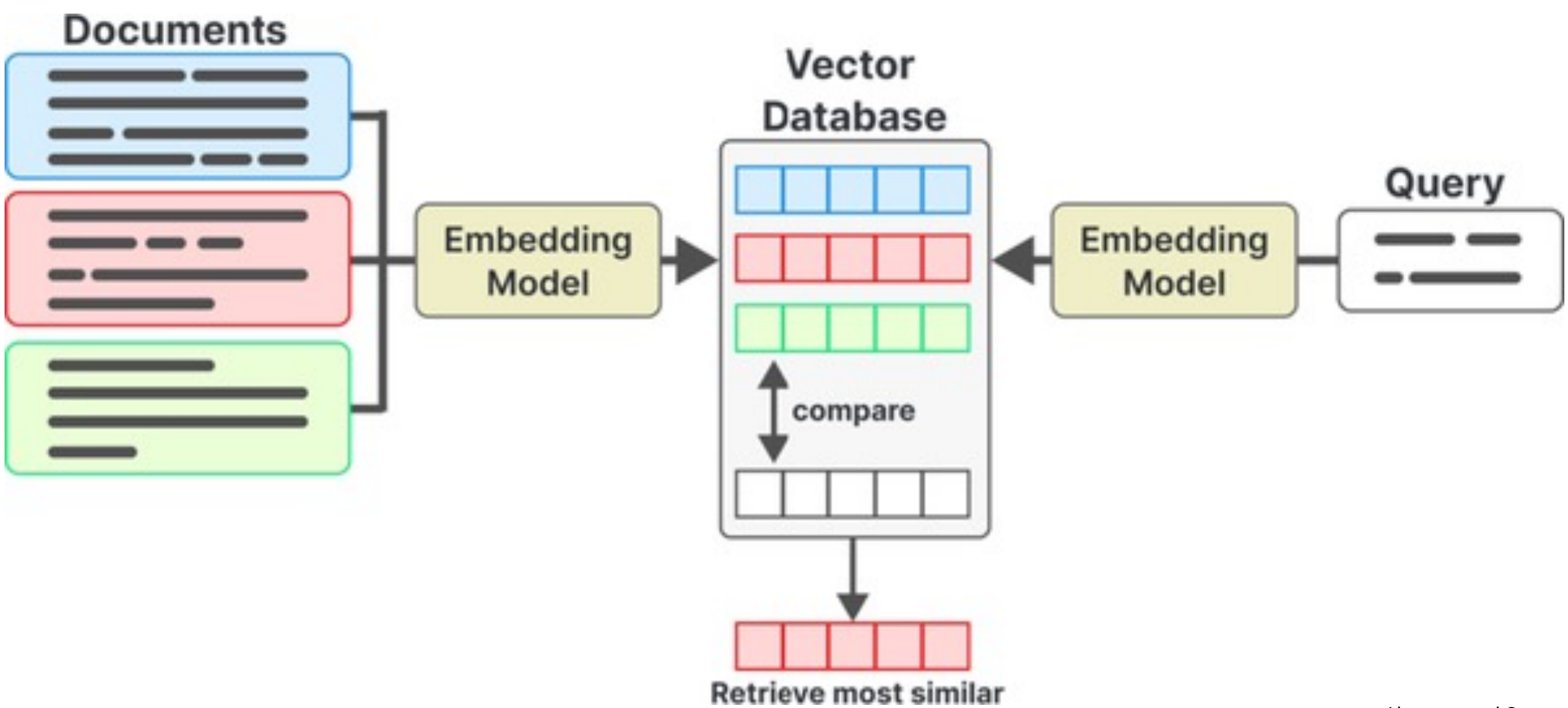
Dense Retrieval



Hands on Large Language Models - Alammar and Grootendorst (2024)

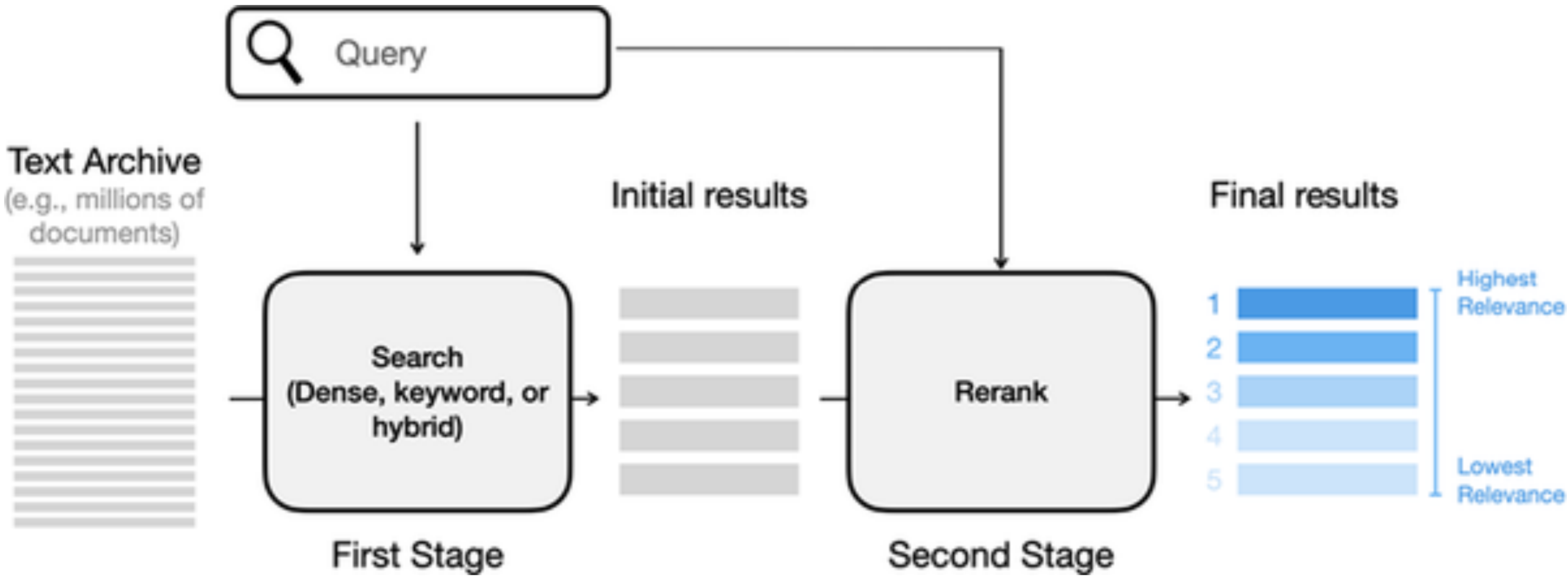


Hands on Large Language Models - Alammar and Grootendorst (2024)

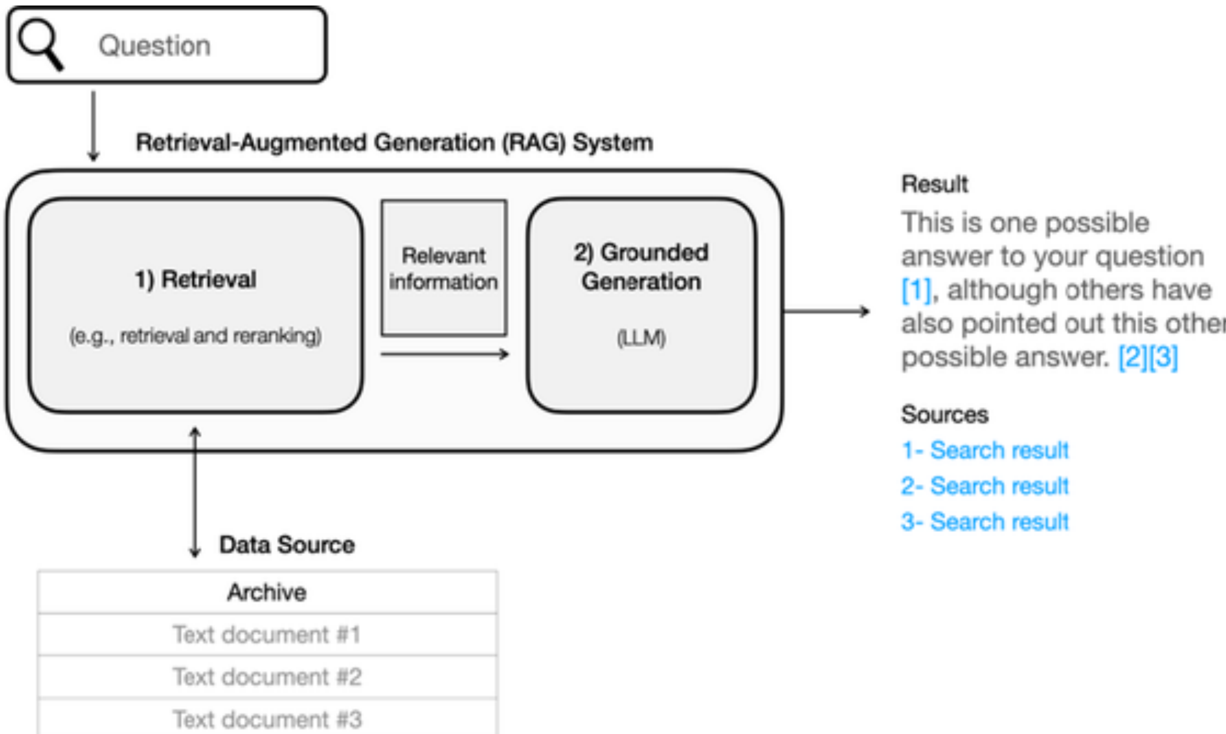


Hands on Large Language Models - Alammar and Grootendorst (2024)

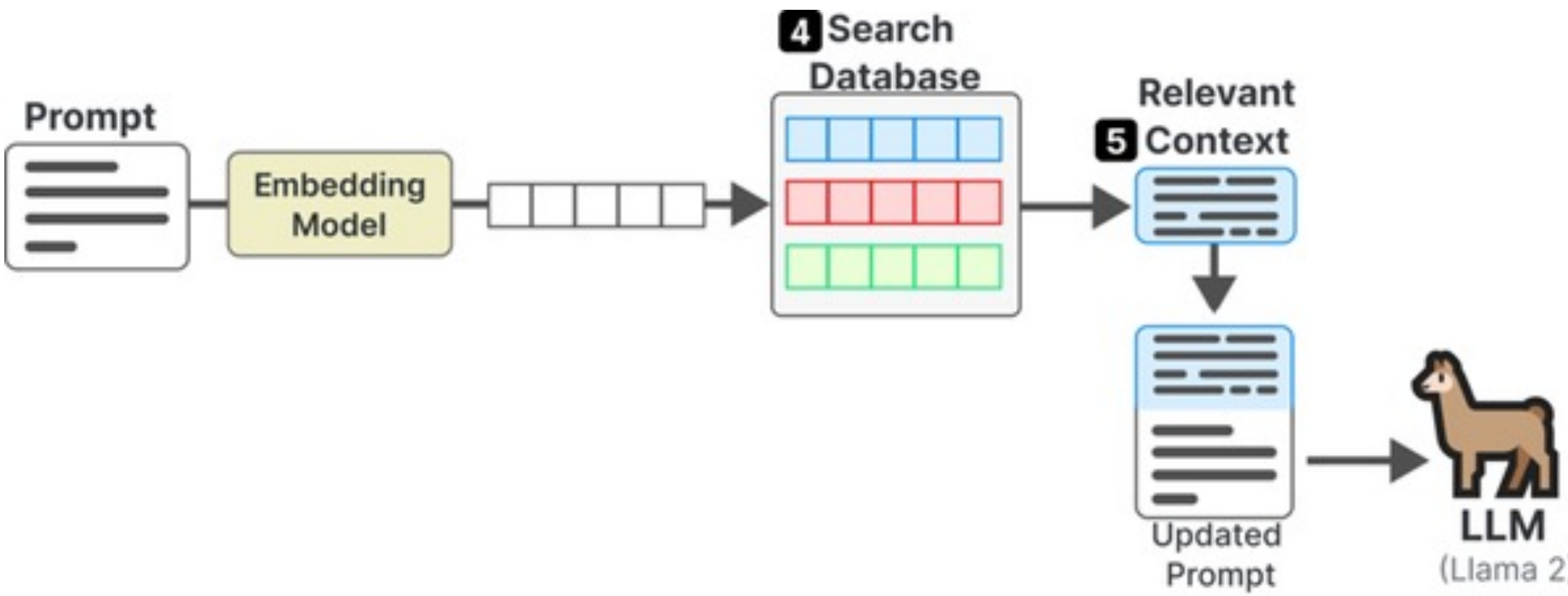
Reranking



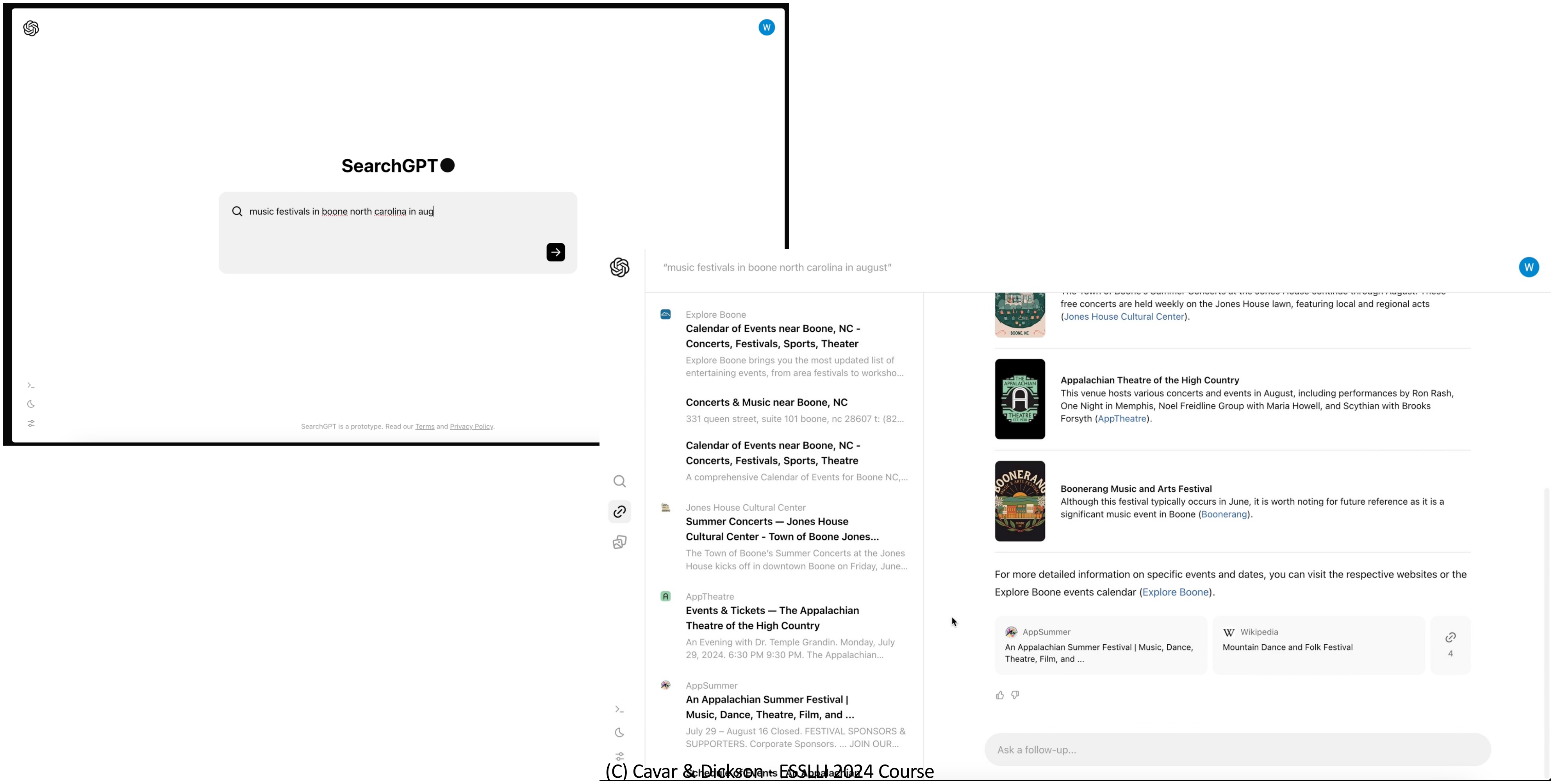
Hands on Large Language Models - Alammar and Grootendorst (2024)



Hands on Large Language Models - Alammar and Grootendorst (2024)



Hands on Large Language Models - Alammar and Grootendorst (2024)


SearchGPT
music festivals in boone north carolina in aug
SearchGPT is a prototype. Read our Terms and Privacy Policy.
"music festivals in boone north carolina in august"
Explore Boone
Calendar of Events near Boone, NC - Concerts, Festivals, Sports, Theater
Explore Boone brings you the most updated list of entertaining events, from area festivals to worksho...
Concerts & Music near Boone, NC
331 queen street, suite 101 boone, nc 28607 t: (82...
Calendar of Events near Boone, NC - Concerts, Festivals, Sports, Theater
A comprehensive Calendar of Events for Boone NC,...
Jones House Cultural Center
Summer Concerts — Jones House Cultural Center - Town of Boone Jones...
The Town of Boone's Summer Concerts at the Jones House kicks off in downtown Boone on Friday, June...
AppTheatre
Events & Tickets — The Appalachian Theatre of the High Country
An Evening with Dr. Temple Grandin. Monday, July 29, 2024. 6:30 PM 9:30 PM. The Appalachian...
AppSummer
An Appalachian Summer Festival | Music, Dance, Theatre, Film, and ...
July 29 – August 16 Closed. FESTIVAL SPONSORS & SUPPORTERS. Corporate Sponsors. ... JOIN OUR...
free concerts are held weekly on the Jones House lawn, featuring local and regional acts (Jones House Cultural Center).
Appalachian Theatre of the High Country
This venue hosts various concerts and events in August, including performances by Ron Rash, One Night in Memphis, Noel Freidline Group with Maria Howell, and Scythian with Brooks Forsyth (AppTheatre).
Boonerang Music and Arts Festival
Although this festival typically occurs in June, it is worth noting for future reference as it is a significant music event in Boone (Boonerang).
For more detailed information on specific events and dates, you can visit the respective websites or the Explore Boone events calendar (Explore Boone).
AppSummer
An Appalachian Summer Festival | Music, Dance, Theatre, Film, and ...
Wikipedia
Mountain Dance and Folk Festival
4
Ask a follow-up...

Knowledge

- Passive knowledge
  - Canned text
    - Question/Query + Response/Answer
    - Response:
      - Text
      - Code
      - Image
  - Indexing response for a given query?
    - Query: vector
    - Response index: vector
    - Match: vector similarity, approximate distance in semantic space, find K Nearest Neighbors…

Symbolic Knowledge Representations

- Knowledge Graphs
- Ontologies
- Reasoning

Goals

- Build the models, infrastructure, technologies to
  - engineer AI systems with advanced computational semantics and pragmatics capabilities
    - Description Logic
    - Temporal Logic
    - Event Semantics
  - Build data sets with semantic annotations (multi-modal: text, speech, image) for training and evaluation of ML models

- Application domains:
  - Discourse Models
  - Text forensics:
    - Fake news, propaganda, and deception detection
  - Domain specific:
    - Medical, Business, …

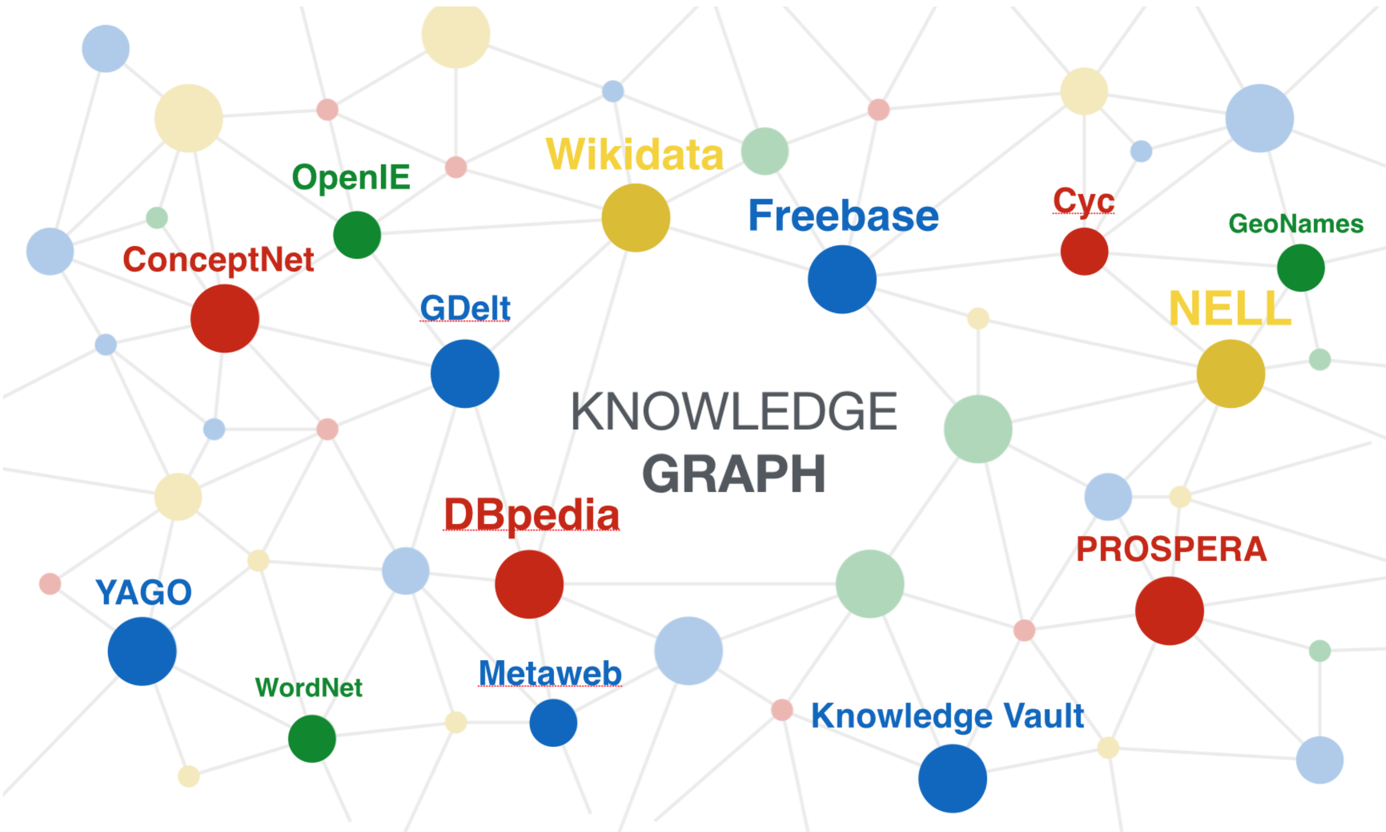
Knowledge and Computing

- Model World Knowledge - Linguistically Motivated Computational Processing
  - Meaning of words
  - Meaning of speech acts (including Presuppositions, Implicatures, Events, Temporal Logic)
  - Computational Models of World Knowledge / Common Sense in phenomena like Binding, Anaphora Resolution, Reasoning
    - “Take the knife, cut the lime in half, and put it down.” it = knife
    - “Take the knife, cut the lime in half, and squeeze it.” it = lime

Objectives

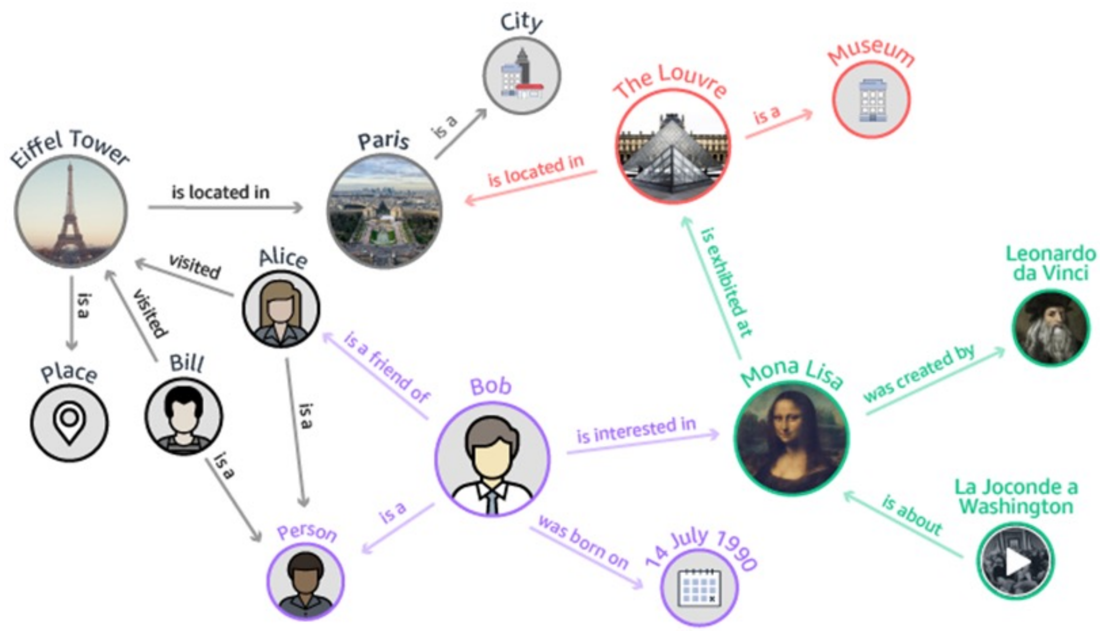
- Multi-Modality
  - Language, Speech, Vision, Sensory Information, …
- Deep Semantic and Pragmatic Processing
  - High precision
  - Deep insights
- High Performance Computation for Big Data Analysis
  - Scalability (Speed, Memory)
  - Efficient Parallel Algorithms

Large Knowledge Graphs



10/2/24 16

Knowledge Graphs



 17

Knowledge Representations

- Model (or Ontology) and Individuals (or Assertions of Facts)

- Model: Web Ontology Language (OWL) as a restrictive model/ontology specifying
  - Concepts (or entity types)
  - Relations (between entities, instances of concepts or entity types)
  - Properties of Concepts and Relations

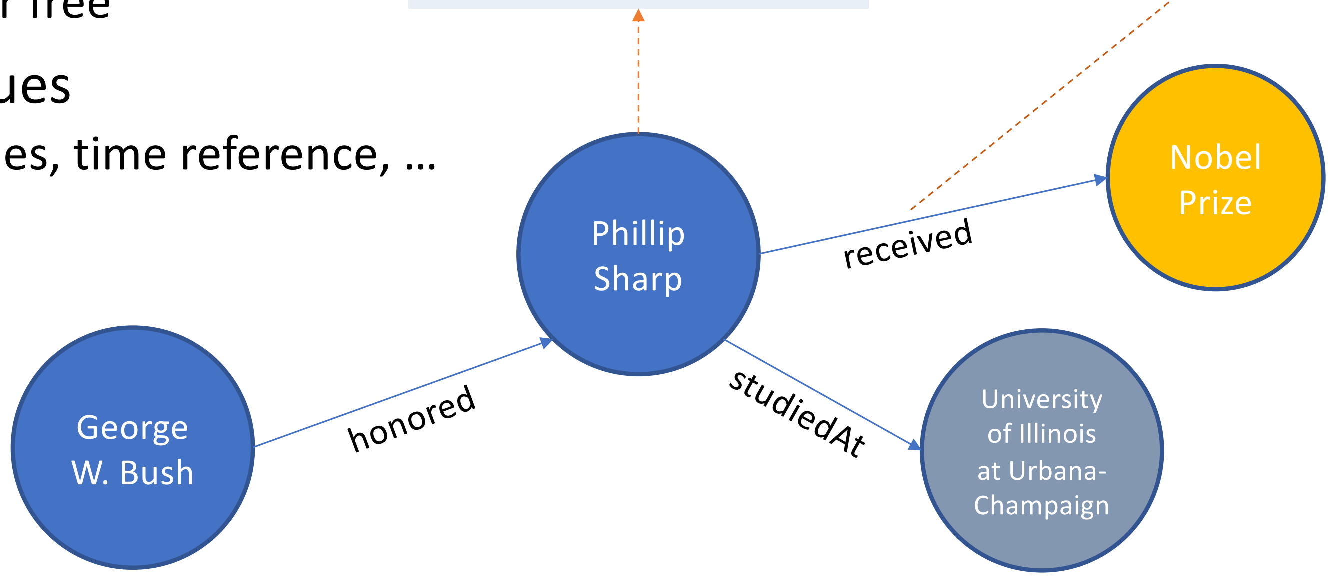
- Individuals
  - Concrete instances of concepts and relations

Knowledge Graphs

- Concepts and Relations
  - Mostly unconstrained
  - Domain specific or free
- Attributes and Values
  - encoding properties, time reference, …

Attribute	Value
Birth place	Falmouth, KY
DOB	06/06/1944
gender	male
…	

Attribute	Value
year	1993
…	



19

Focus

- Knowledge Representations
  - Description Logic
    - Probabilistic
    - Dynamic
  - Events
    - Temporal Logic
      - Event sequences and durations
    - Resource Logic
      - Transfer of resources
      - Intentionality

Knowledge Graphs

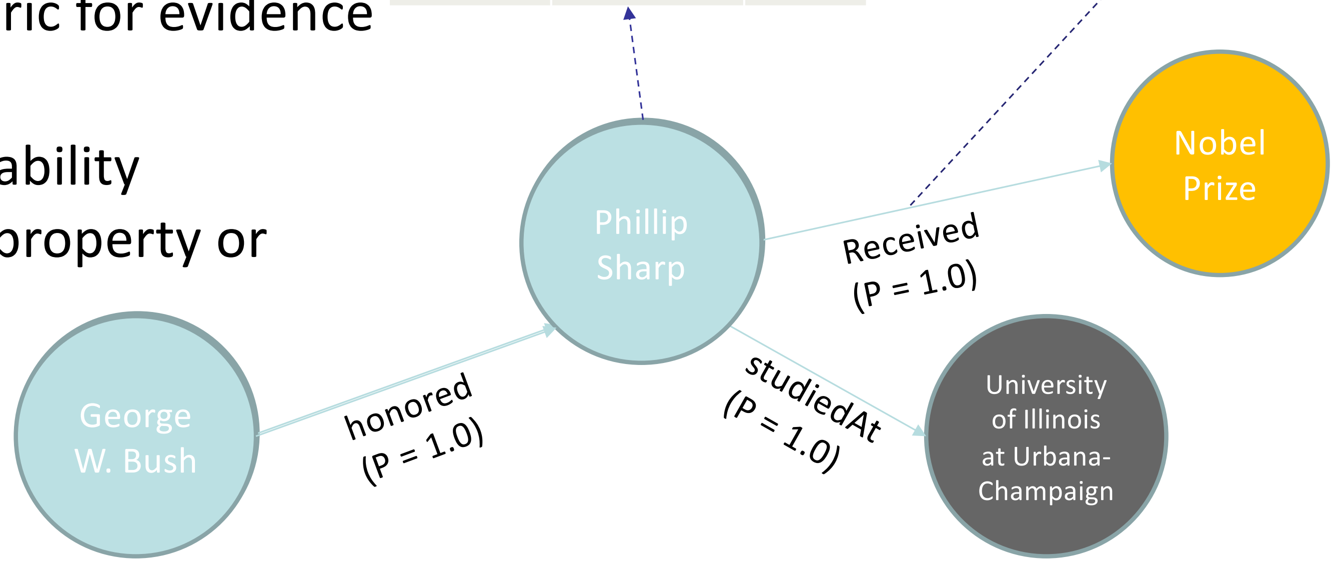
- Formally:
  - Sets of triples:
    - Subject Predicate→ Object
  - Quad notation:
    - Sets of triples grouped by a 4^{th} element
      - Graph ID
      - Temporal ID
      - Event ID
      - …

Knowledge Representations

- Confidence Scores
  - Qualitative metric for source
  - Quantitative metric for evidence
- Scoring
  - Sources wrt. Reliability
  - Counts of entity property or relation

Attribute	Value	P(x)
Birth place	Falmouth, KY	0.9
DOB	06/06/1944	1.0
gender	male	1.0
…		

Attribute	Value	P(x)
year	1993	0.8
…		



Knowledge Graphs

- No computation or interpretation of logic equations (e.g., no access to universal or existential quantifiers)
- Direct mapping of knowledge from language input limited to Description Logic

- Description of Knowledge
  - Directed Graph: encoding concept, events, domain specific knowledge…
  - Attribute-Value encoded features like size and shape, but also event time references (start, end, duration), etc.

- Reasoning: OWL & Reasoner, Common Graph Algs.
- Prediction: Links, Class prediction, etc.
- Machine Learning of concepts and concept properties: node or edge embedding

7/24/24 23

Reasoning

- Validation of assertions
- Entailment computation
- Link prediction
- Entity classification
- Question Answering and Dialog

- Domains:
  – Medical, FinTech, Legal, Cybersecurity, …

Resources

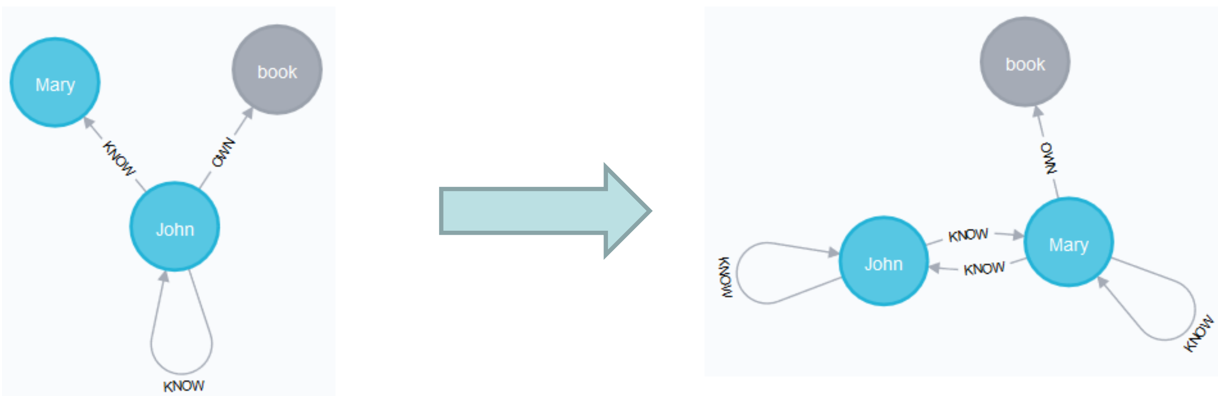
- General Knowledge – Knowledge Graphs
  - DBpedia
  - YAGO
  - ConceptGraph
  - …
- Domain Specific Knowledge Graphs
  - Unified Medical Language System (UMLS)
  - USDA Food Database
  - CyGraph
  - …

Knowledge Graphs

- Static: Concept and Relation Properties
  - Even when dynamically growing or changing
- Problem to encode events or procedures
  - *Mary gave John a book*.
    - Event as a state change / transformation:
      - Mary owns a book, John does not → John owns a book, Mary does not
  - Peter was fetching his daughter from school.
    - Intermediate states:
      - Peter is at home, daughter at school → Peter is at school, daughter at school → Peter is at home, daughter at home

Events and Resources

- *John gave Mary a book.*



- Events as sequences of graphs and graph transformations
  - Encoding of Intentionality

Temporal Relations

- Sequencing of events or sub-events
  - Wash the veggies, chop them, fry them.
    - Presentation and Temporal event sequence: 1 2 3
  - Before you fry the veggies, wash and chop them.
    - Presentation sequence: 3 1 2
    - Temporal event sequence: 1 2 3
- Duration of events
  - Clear reference: “for 30 minutes”
  - Common sense

Temporal Relations

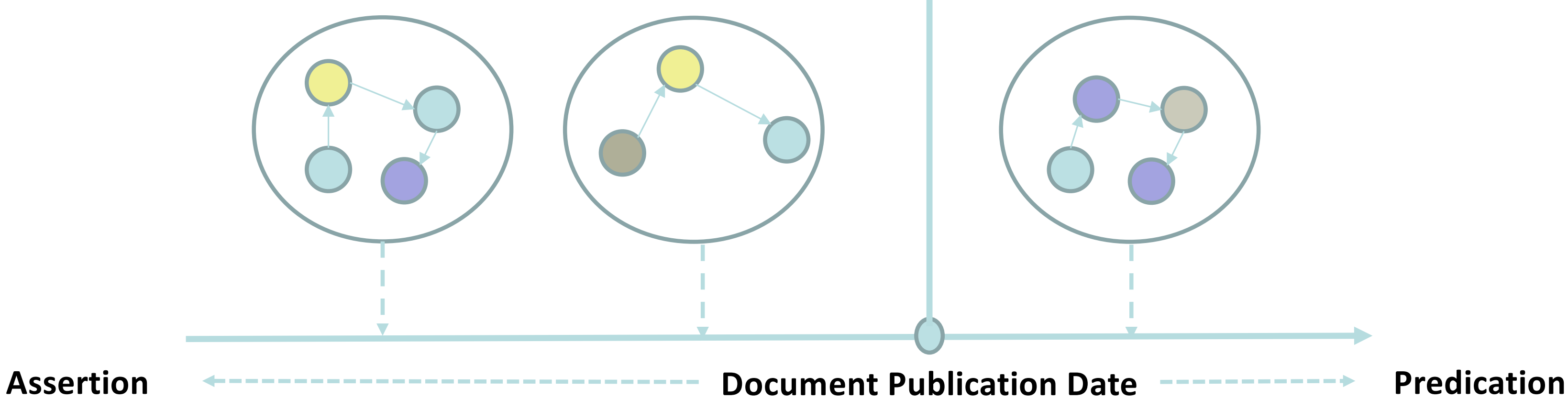
- Duration of events
- Unfolding over time
  - Events relate to time
  - States are points in time

- Temporal sequencing relates to
  - Causal reasoning

 29

Event Graphs

- Arrangement of sub-events along time axis
  - Approximation of duration
  - Identification of geo-location
  - Linked entities and relations



Graphlets of Events and States

- Graphlets
  - Sub-graphs using a point in time
  - Temporal sequencing of events in time = graphlets

- Technically
  - Graphs and Graphlets use temporal timestamps as properties
    - Temporal reference
    - Event ID or variable

Temporal Scope

- Simple temporal relations
  - Past tense: Tim Cook bought Google.
    - Assumptions: factive, true event
  - Future tense: Tim Cook will buy Google.
    - Assumptions: non-factive, hypothetical
- Complex relations: temporal scope
  - Reuters reported that

    Tim Cook bought Google

  - Reuters will report that

 32

Temporal Scope

- Sub-event triples:
  - Tim Cook buy→ Google
  - Event-ID is→ 43829

- Dominating event triples:
  - Reuters report→ eventX
  - eventX hasID→ 43829
  - eventTime …
  - speakerTime …
  - referenceTime …

NLP Extensions

- Implicatures:
  - John to Peter: I bought the blue car.
    - John and Peter talked about cars earlier.
    - There should be a set with at least one more car the John could have bought, but did not, and
    - None of the cars in the set is blue.
  - Clues: Definiteness of NP via the, and specificity of NP
- Presuppositions:
  - John fed his cat this morning.
  - Assumptions:
    - John owns/has a cat/pet.
    - John owned cat-food this morning.
  - Clues: Possessive pronoun as modifier of Direct Object.

Pragmatics and Implicatures

- Input: *Mary is the sister of Tracy.*
  → **Mary -> gender: female**
  → **Mary hasSibling Tracy**, **Tracy hasSibling Mary**
  …
- Input: *Donald Trump met Vladimir Putin.*
  → **Vladimir Putin met Donald Trump**
- Input: *I regret that I drank a coffee.*
  → Claimed to be a fact / true: **I drank coffee.**
- Input: *Tom parked his car in the Atwater garage.*
  → **Tom knows how to operate a car.**
  → **Tom was in possession of a car.**
  …
- And many more…
- Knowledge of language (universal and particular) is necessary to generate these presuppositions and implicatures.

Pragmatics and World Knowledge

- Scenario in Medical
  - Medication Metformin found in pocket of unconscious person:
    → *Patient might have (Type-2) diabetes.*

- Domain specific knowledge (Knowledge Graph) is necessary to resolve the drug name and reason about the diagnosis.

Predicates

- Veridicality
  - Factive predicates: *know*, *regret*, *realize*, *notice*, …
    - *I regret that* … (X did something to Y)
    - Complements are assumed to be true
  - Non-factive predicates: *believe*, *think*, *claim*, …
    - *I believe that* … (X did something to Y)
    - Complements cannot be assumed to be true
  - Counter-factive predicates: *pretend*, …
    - *John pretends that he is ill.*
    - Complement cannot be true: *John is not ill*

- Question:
  - Cross-linguistic similarity = universal properties related to factivity

Predicates

- Predicative Properties
  - Functional
    - X → hasBirthMother → Sue & X → hasBirthMother → Susan
    - Implies: Sue & Susan are identical entities
  - Inverse functional
    - Same as Functional, just for the inverse predicate: isBirthMotherOf
  - Transitive
    - X → hasAncestor → Y → hasAncestor Z
    - Implies: X → hasAncestor → Z
  - Symmetry
    - X → met → Y
    - Implies: Y → met → X
  - Asymmetry
    - X → isChildOf → Y
  - Reflexive
    - X → knows → Y
    - Implies: X → knows → X
  - Irreflexive
    - Any relation where X → relation → Y implies X cannot be Y

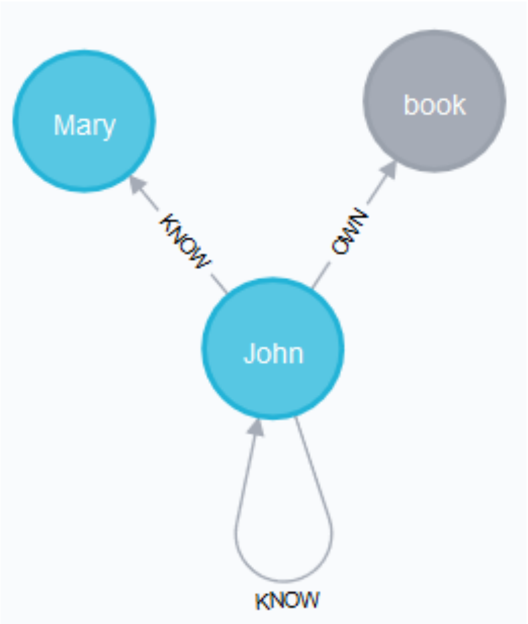
Predicates

- Functional
  - X → hasBirthMother → Sue & X → hasBirthMother → Susan
  - Implies: Sue & Susan are identical entities
- Inverse functional
  - Same as Functional, just for the inverse predicate: isBirthMotherOf
- Transitive
  - X → hasAncestor → Y → hasAncestor Z
  - Implies: X → hasAncestor → Z
- Symmetry
  - X → met → Y
  - Implies: Y → met → X
- Asymmetry
  - X → isChildOf → Y
- Reflexive
  - X → knows → Y
  - Implies: X → knows → X
- Irreflexive
  - Any relation where X → relation → Y implies X cannot be Y

- Strong cross-linguistic similarity
  - Universal?

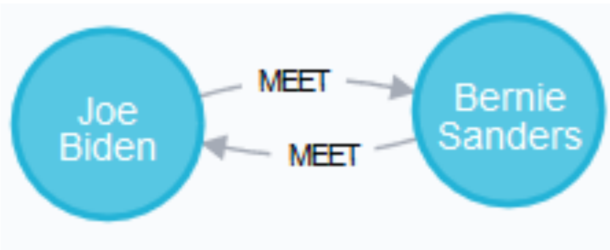
Predicates

- Reflexive
  - X → knows → Y
  - Implies: X → knows → X



Predicates

- Symmetry or Reciprocity
  - X → met → Y
  - Implies: Y → met → X



- Also:
  - Bernie and Joe hugged each other.

Predicates

- Veridicality
  - Factive predicates: *know*, *regret*, *realize*, *notice*, …
    - *I regret that* … (X did something to Y)
    - Complements are assumed to be true
  - Non-factive predicates: *believe*, *think*, *claim*, …
    - *I believe that* … (X did something to Y)
    - Complements cannot be assumed to be true
  - Counter-factive predicates: *pretend*, …
    - *John pretends that he is ill.*
    - Complement cannot be true: *John is not ill*

- Question:
  - Cross-linguistic similarity = universal properties related to factivity

Typing and Predicates

- Verb Frames and Predicate Properties
  - Type information for arguments
  - Differentiation between
    - Modifiers: for properties of entities and relations in DL graphs
    - Arguments: core entities and relation links in DL graphs

Semantic Mapping and Reasoning

- Type of Predicative Arguments: Typing
  - Named Entity Recognition
  - Closes possible Hypernym in a Taxonomy or Ontology of isA relations

- Identity of entity: Linking
  - Named Entity Recognition
  - Link to unique identifier of entity in some knowledge representation, Ontology, Wikipedia, Knowledge Graph

- Issues: Ambiguity

Graph Extraction and Linking

- Graph generation sample:
  - NLP pipelines
  - Graph extraction
  - Linking (conceptualization, language independent representation)

- Goal:
  - Extract predicate-argument tuples
  - Type the entities (e.g. NER, ontology lookup, Knowledge Graph linking)
  - Dynamically expand the knowledge graph and track weights (probabilities)

Linking Strategy

- 1 to *n* relation between entity and entities in Knowledge Graphs
- Disambiguation via Geometrical Similarity
  - Text representation as average vector of word embeddings (e.g. Numberbatch, Glove, Bert)
  - Graph as average vector of concept and edge embeddings
- Maximization of context word prediction for each linking candidate
- KGs:
  - YAGO, ConceptNet, MS Word Graph, etc.

 46

Linguistic Bias

- Pragmatic effects in language use data (e.g. Sperber & Wilson's Relevance Theory, Grice's Maxims)
  - People communicate facts and information that is relevant, new, exciting

  - Observation:
    - Exciting information: purple carrots

Less exciting information: orange carrots

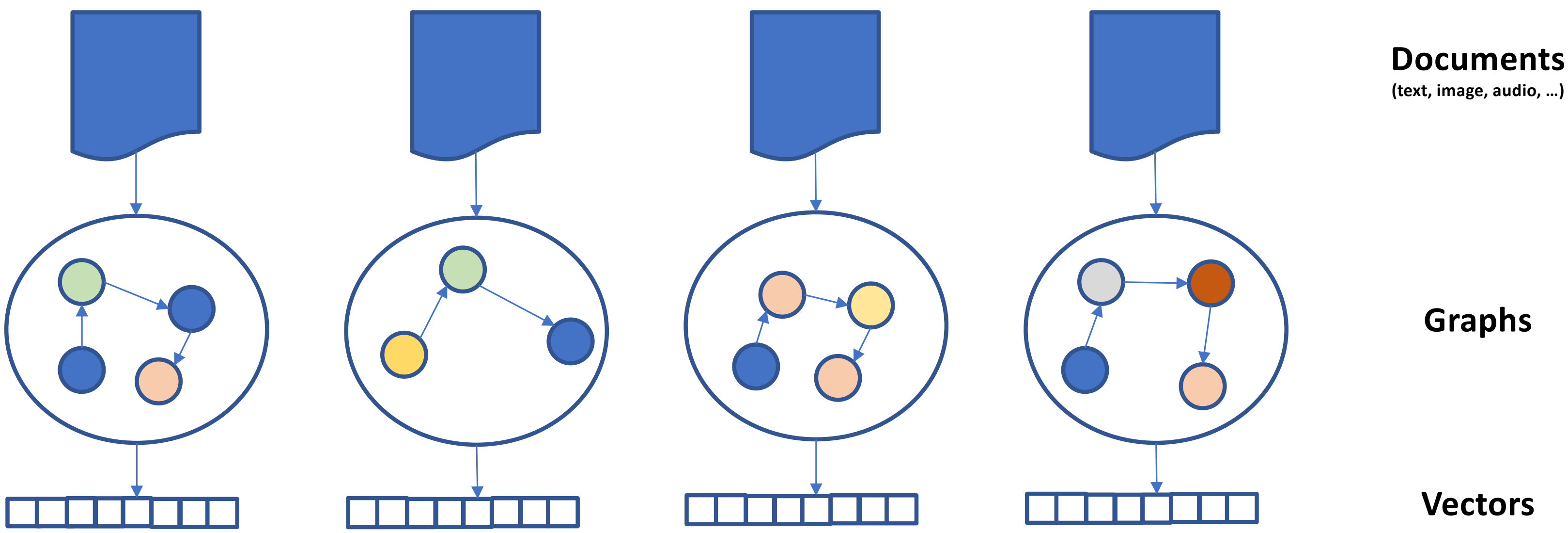
Solution

- Multi-modal information input to knowledge representation
  - Language input (speech and text)
  - Information in images
  - Haptic information
  - Secondary information: sound it makes, properties when shaking, tossing, etc.

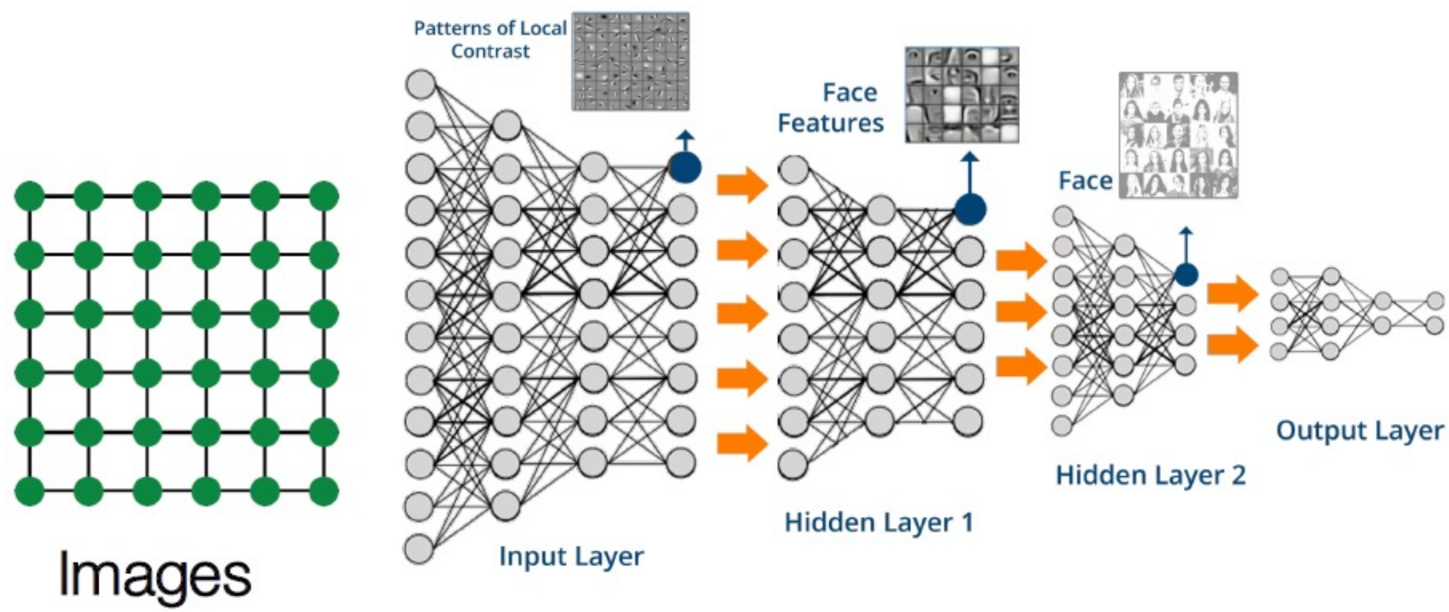
- Graphs generated using:
  - General or common sense knowledge
  - Domain specific knowledge
  - Semantic restrictions over graphs: ontologies, taxonomies

Document Graphs

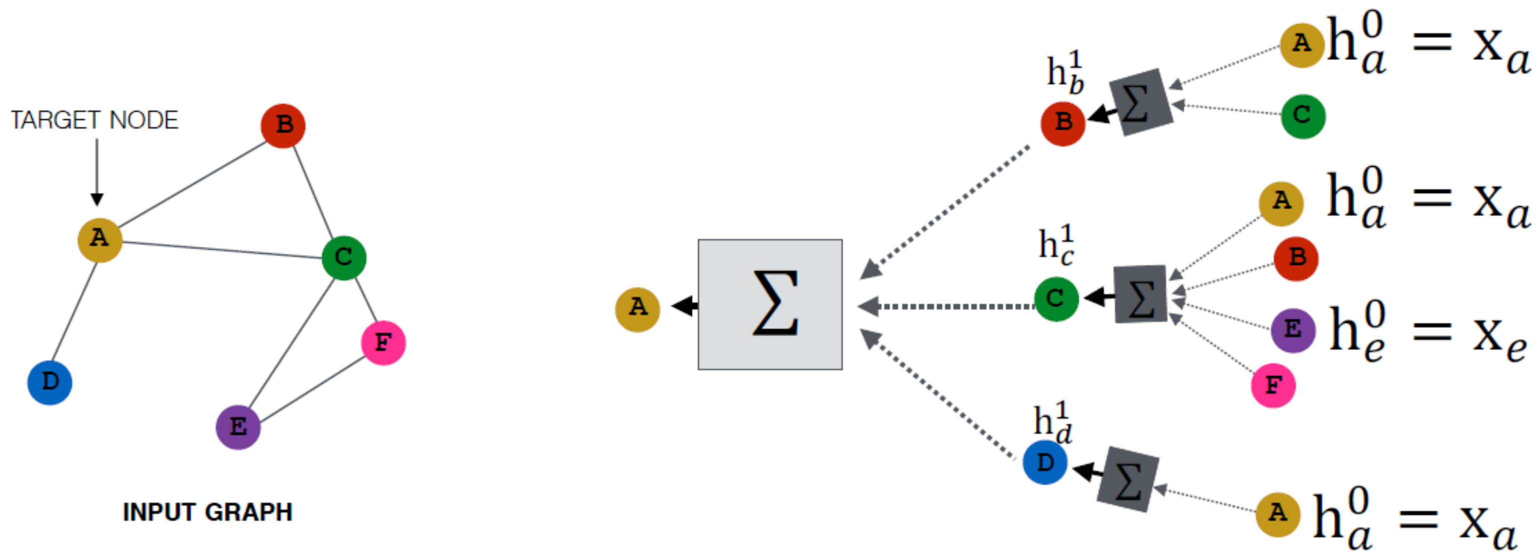
- Concept/Knowledge Graph Document Representation



Graph Neural Network Models



Graph Neural Network Models



$$h_v^{(l+1)} = \sigma(W_l \sum_{u \in N(v)} \frac{h_u^{(l)}}{|N(v)|} + B_l h_v^{(l)})$$

References

- Semantic Communication Enhanced by Knowledge Graph Representation Learning
  - https://arxiv.org/abs/2407.19338

Protégé and OWL

- Download Protégé and
  - Define a simple OWL ontology
  - Reason over some asserted facts / individuals