Medium

- Query processing
- Knowledge graphs

Query processing

- The front door of the search engine
- In a nutshell
  - Query intent correctly identified and re-written in internal representation
  - Use re-written query to retrieve and match relevant results
  - Use all info to generate SERP
- Lots of annotations and classifiers are used in this step
- NLP components (POS, NER, stemming, keyphrases, geotagging, etc.)
- Performance and matching (not covered in this tutorial)

Query intent

- Broder’s taxonomy
- Web search evolves
- New users and behaviors
- Emerging new types of search intents

Question intent

- Question taxonomy compared to queries
- Multi-faceted question taxonomy
- Detailed methodology
- Agreement analysis and tie-breaking

Type	Examples
Advice	how can I be successful in life? how should I invest my salary?
Attribute	what is pristine edge's real name what is senegal's official language
Calculation	4,146.70+700+11900 1/2 cups in tbsp
Description	what is propylene kit what is oracle vpd functionality
Entity	who replaced ted kennedy in the senate who produced transformers
Language	what is puppy in swahili what is the common name for jade
List	types of aircraft southampton to guernsey types ant poison
Location	where are protists most abundant in humans what is oklahoma's absolute location
Opinion	is donald trump a good president? is ronaldo or messi a better player?
Process	what is needed to get home insurance how to check warranty of sd card sandisk
Quantity	how long is csus transfer orientation cost of an ice cream truck
Reason	why do knees swell up why do lipomas grow back
Resource	python temperature converter code tum mile love reprise lyrics english
Temporal	when do the oscar awards start when does daylight saving time return?
Verification	is tomorrow Monday? is donald trump the 34th president?
Weather	5 day weather forecast for york tybee island weather in march

B. Cambazoglu "An Intent Taxonomy for Questions Asked in Web Search", CHIIR 2021

NFQ intent taxonomy

- Very little understanding of Non-Factoid questions
- Lots of research on factoid QA (SQuAD, MS MARCO) not on NFQ
- Detailed taxonomy creation methodology
- Data set and model available https://github.com/Lurunchik/NF-CATS

Category	Description	Expected Answer Structure	Patterns
INSTRUCTION	You want to understand the **procedure/method** of doing/achieving something.	Instructions/guidelines provided in a step-by-step manner.	How to ...? How can I do ...? What is the process for ...? What is the best way to ...?
REASON	You want to find out **reasons** of/for something.	A list of reasons with evidence.	Why does ...? What is the reason for ...? What causes ...? How come ... happened?
EVIDENCE-BASED	You want to learn about the **features/description/definition** of a concept/idea/object/event.	Wikipedia-like passage describing/defining an event/object or its properties based only on facts.	What is ...? How does/do ... work? What are the properties of ...? What is the meaning of ...? How do you describe ...?
COMPARISON	You want to **compare/contrast** two or more things, understand their differences/similarities.	A list of key differences and/or similarities of something compared to another thing.	How is X ... to/from Y? What are the ... of X over Y? How does X ... against Y?
EXPERIENCE	You want to **get advice** or **recommendations** on a particular topic.	Advantages, disadvantages, and main features of an entity (product, event, person, etc) summarised from personal experiences.	Would you recommend ...? How do you like ...? What do you think about ...? Should I ...?
DEBATE	You want to **debate on a hypothetical question** (is someone right or wrong, is some event perceived positively or negatively?).	Arguments on a debatable topic consisting of different opinions on something supported or weakened by pros and cons of the topic in the question.	Does ... exist? Can ... be successful? Do you think ... are ...? Is ... really a ...?

V. Bolotova et al. “A Non-Factoid Question-Answering Taxonomy” SIGIR 2022

Problems

- Taxonomies and categorization mechanisms
- Better understanding of user intent
- Domain specific cases
- Subjective queries and questions
- Near-duplicate questions

Knowledge graphs

- Knowledge graph (KG) describes objects of interest and connections
- Organizing data as nodes and edges
- Examples
  - Microsoft Satori, Google Knowledge Graph, Amazon Product Graph
  - Knowledge bases (KBs): Yago, Freebase
- Knowledge graph and knowledge base terms are used interchangeably
- Similar ideas have been around for quite some time
  - Ontologies and common sense knowledge
  - Cyc

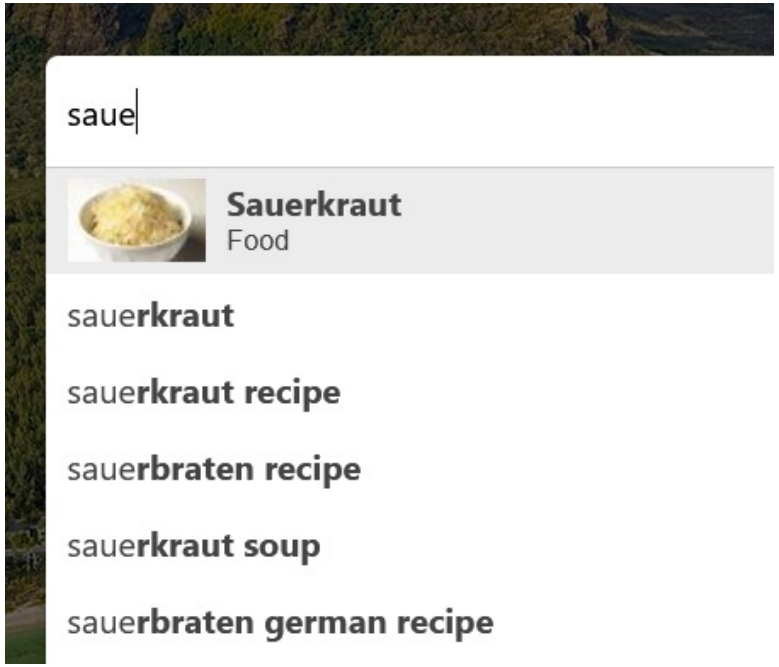
What is it?

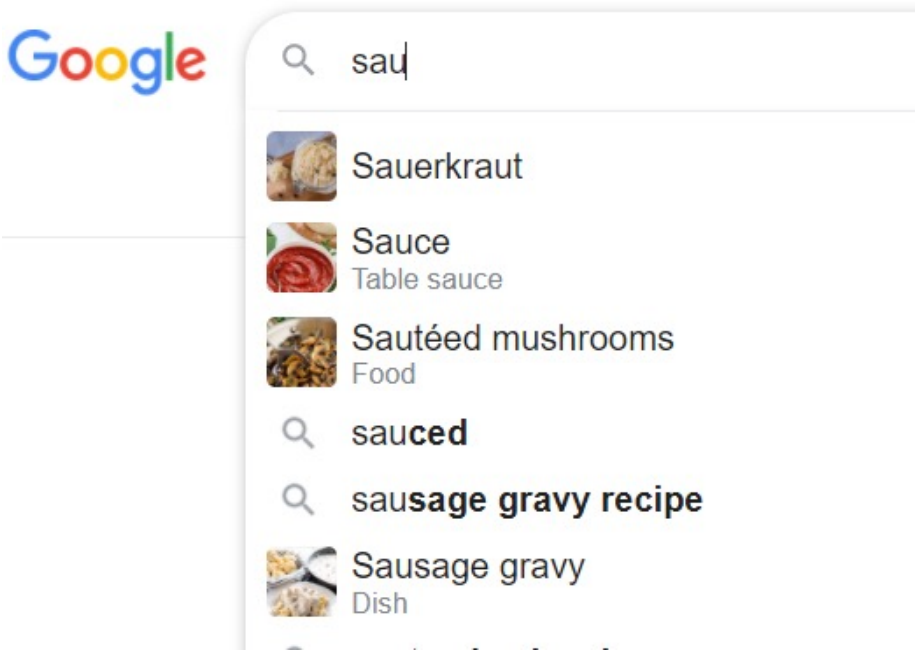
- How to identify nodes and derives edges
- So far, most research on KGs/KBs use Wikipedia as source
  - Benefits: easy to read, easy to parse, Wikipedians
  - Drawbacks: coverage, outdated content, bias
- Why we care?
  - Machine readable facts about a domain
  - Data can be used for different use cases
- More specifically
  - KG is a repository of entities, types and relationships
  - KG defines entities, types, attributes, relations, provenance
  - KG is data
  - KG evolves and needs maintenance

KGs in action

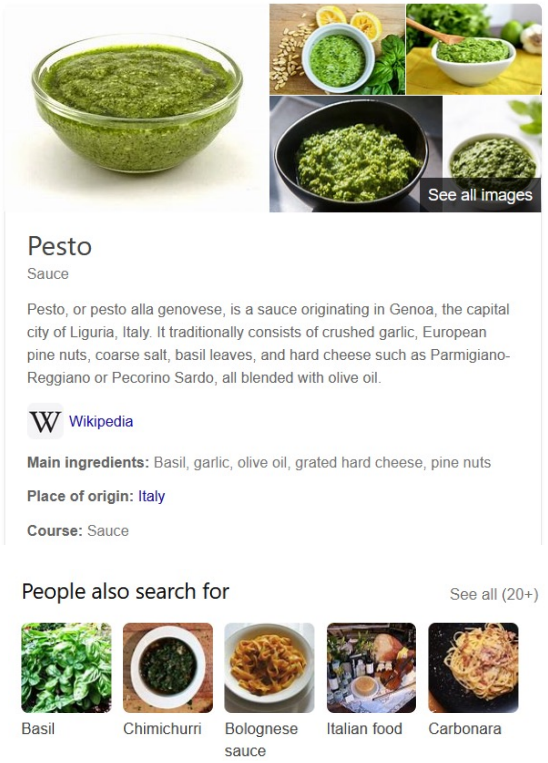
- Semantic search
  - Going beyond 10-blue links
  - Understanding queries and documents
- Document retrieval
  - Expansion
  - Language modeling
- Entity retrieval
- Recommendations
- Question-answering
- Data cleaning

Example - Autocomplete

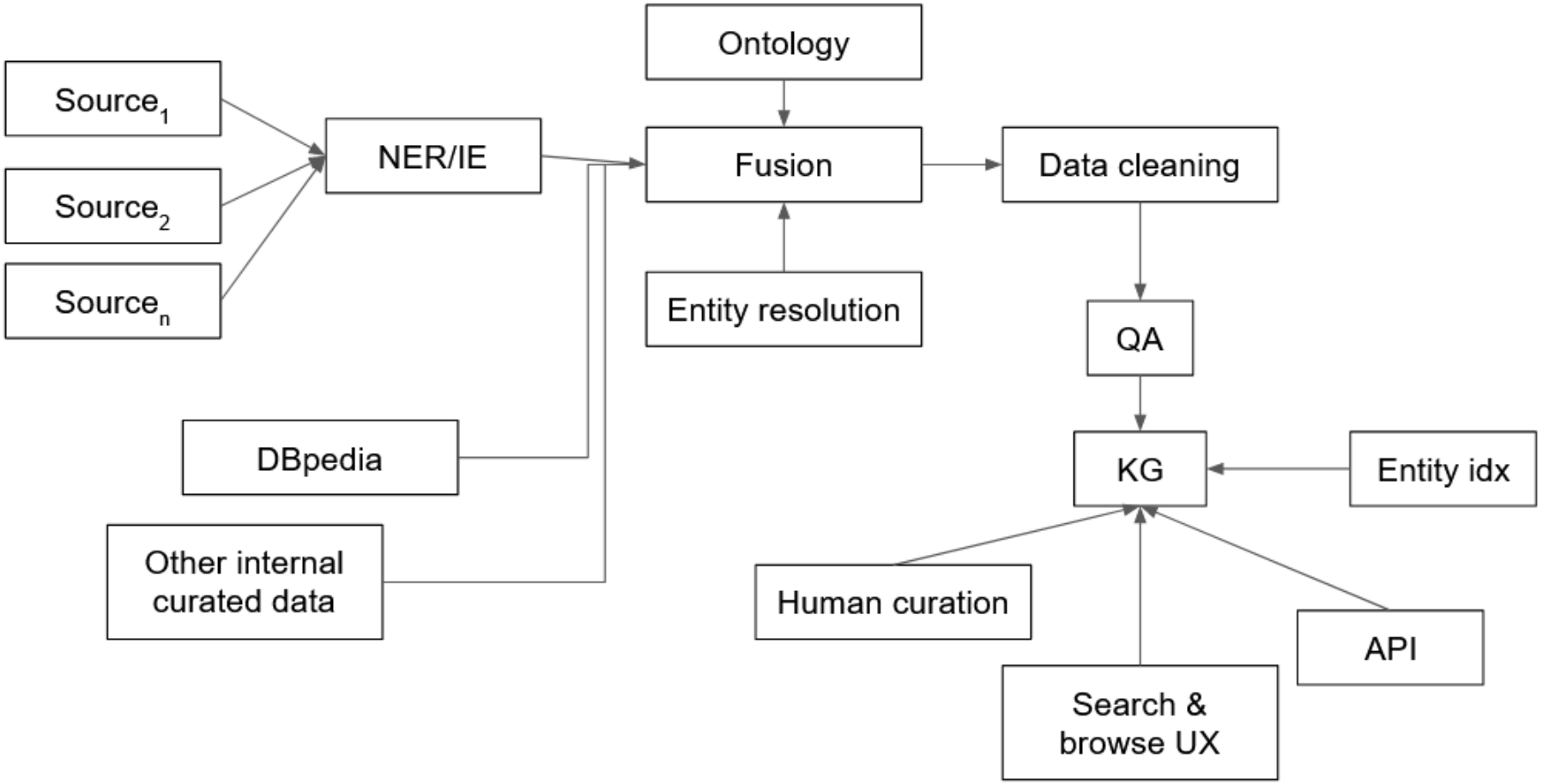




Example - Entity cards



Architecture for KG construction

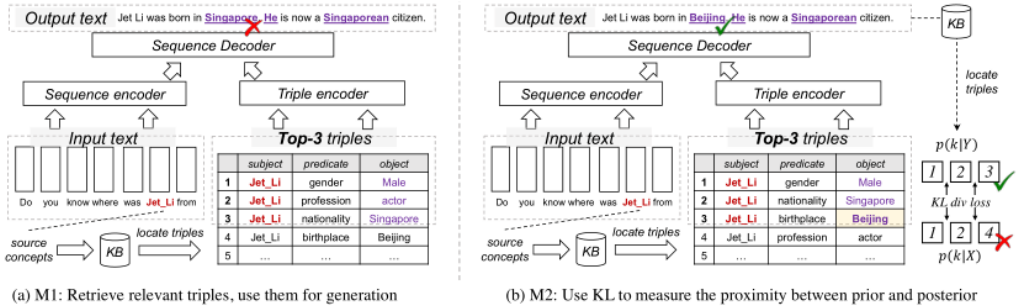


Semantic search

- Understanding information needs
- Query classification
  - Assign a query to one or multiple pre-defined categories
  - Query intent classification (Broder)
- Query annotation
  - Generate semantic markup for a query
  - Query segmentation: group terms into phrases
  - Query tagging (POS, NER)

Knowledge-enhanced generation with KGs

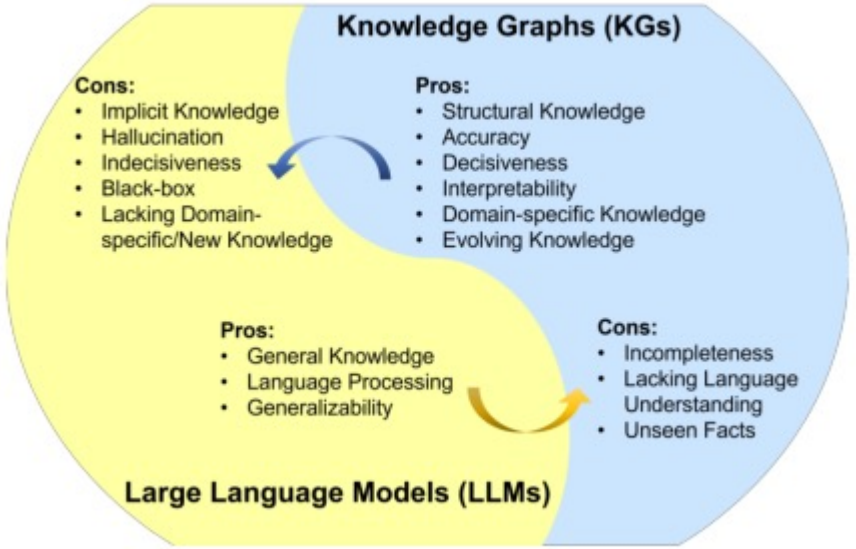
- Design Supervised Tasks around KG
  - Discover the dependencies of elements within a sequence
  - Retrieve relevant triples, then using them for generation
  - Using KL to measure the proximity between prior and posterior distribution
- Selecting KG or facts in a KG



W. Yu et al. "A Survey of Knowledge-Enhanced Text Generation" ACM Computing Surveys 2022

Combining KGs & LLMs

- LLMs lack of factual knowledge
- LLMs memorize and knowledge in a training set
- No interpretability
- KGs are difficult to construct and maintain
- KGs are domain specific



S. Pan et al. "Unifying Large Language Models and Knowledge Graphs: A Roadmap" arxiv.org/abs/2306.08302

KG-enhanced LLMs

- KG-enhanced LLM pre-training
  - Training objective
  - LLM inputs
  - Fusion models
- KG-enhanced LLM inference
  - Dynamic fusion
  - Retrieval augmented
- KG-enhanced LLM interpretability
  - Probing and analysis

LLM-augmented KGs

- Embedding
  - Text encoders
  - Joint text and KG embeddings
- Completion
- Construction
  - Entity discovery
  - Relation extraction
  - Coreference
  - Distilling KGs from LLMs
- KG-to-text generation
- LLM-augmented KG question answering

Problems

- Data quality for KGs
- Use of LLMs for KG construction
- Redundancy and completeness
- Maintenance
- LLM answer validation with KGs