Artificial Intelligence (CS303)

Lecture 9-annex: Knowledge Graph

Hints for this lecture

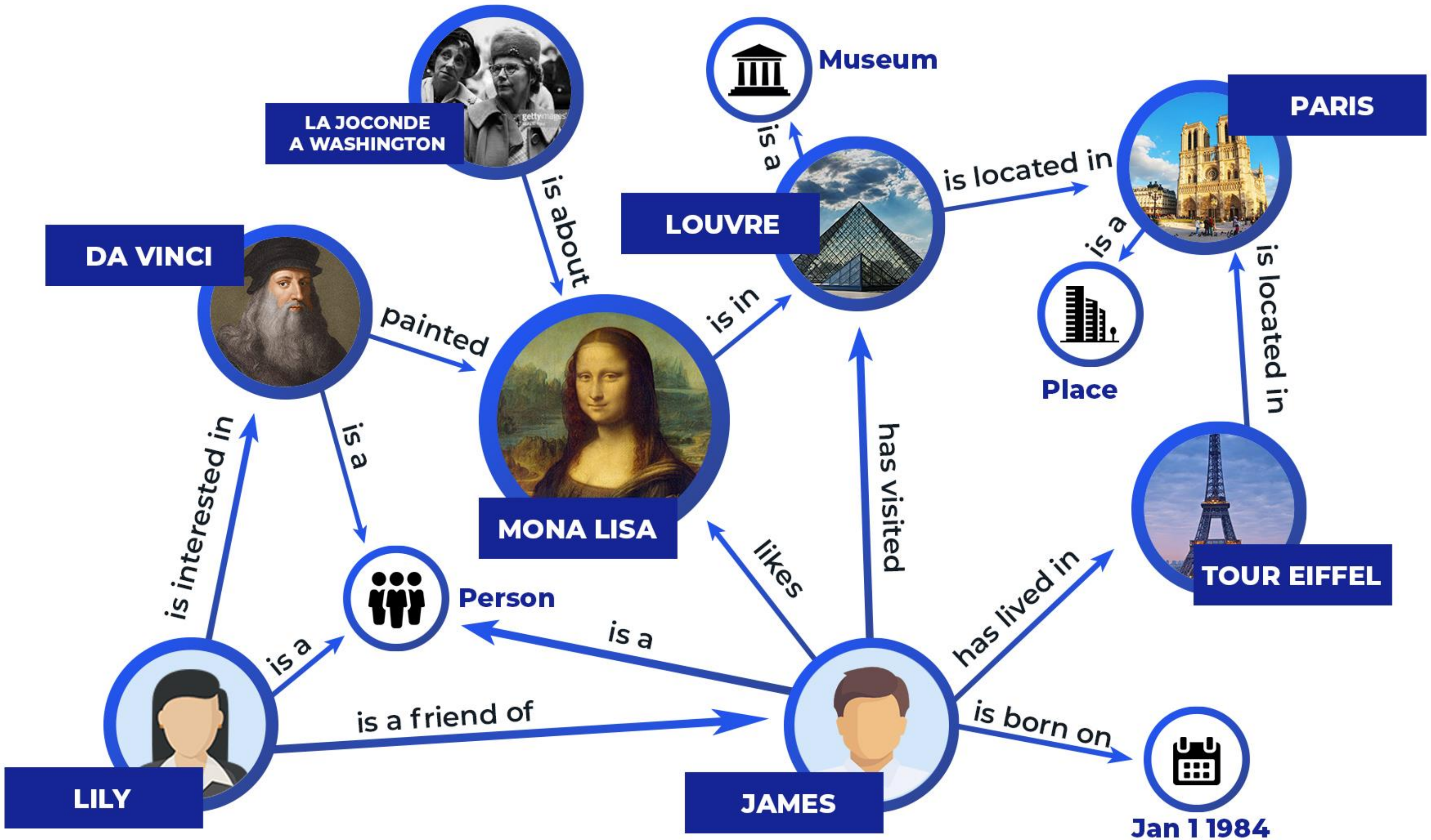
- **Use GRAPH to represent sth. that is less general, but still useful.**

Outline

- Overview of knowledge graph(KG)
- How to construct KG
- How to manage KG
- How to apply KG
- Reference

I. Overview of knowledge graph

What is knowledge graph?



Definition

- Semantic Network:
  - It uses vertexes and edges to describe knowledge graphically.
  - Basic elements:
    - Vertex: entities and concepts
    - Edges: relations and properties
- Knowledge graph: *Large-scale* semantic network

Classification of KG

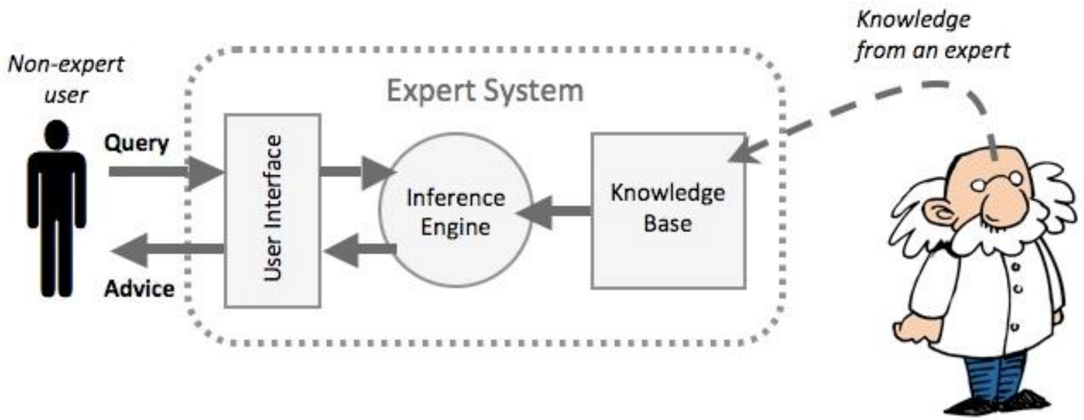
- **Knowledge graphs are classified by the type of knowledge in KGs:**
  - **Domain-specific Knowledge Graph(DKG)**
  - **General-purpose Knowledge Graph(GKG)**

		DKG	**GKG**
Knowledge Representation	Breadth	Narrow	Wide
	Depth	Profound	Superficial
	Fine-grained	√	×
Knowledge Acquisition	Demand of quality	Harsh	High
	Reliance on experts	Heavy	Light
	Highly automated	×	√
Knowledge Application	Chain of reasoning	Long	Short
	Application complexity	Sophisticated	Simple

Several typical KGs

Knowledge Graph	Constructed by	DKG or GKG	Feature	Scale	Construction method	Language	Type of knowledge
Cyc	Cycorp Inc.	GKG	1. Encode millions of general knowledge by human. 2. It's utilized in intelligent inference.	7m assertions 630k concepts 38k relations	Artificial	English	Common sense
WordNet	Princeton University	GKG	A synset is a basic element.	150k words 110k synsets 200k relations	Artificial	English	Lexical
ConceptNet	MIT	GKG	Multilingual knowledge base	8m entities 21m relations	Automatic	Multilingual	Common sense
GeoNames	Geonames.org	DKG	Multilingual geographic location information	25m entities	Semi-automatic	Multilingual	Geographic
CN-Dbpedia	Fudan University	GKG	Live updated Complete data/service interface	16m entities 0.22bn relations	Automatic	Chinese	Encyclopedic

Evolutoin of KG



Drawbacks

1. Relying heavily on participation of human being
2. Unable to describe implicit knowledge(Such as, how to make delicious dish)
3. Cannot update knowledge on time, etc.

Massive data and advanced computing power provide the conditions of KG's birth

Knowledge Graph

First proposed by Google in 2012

II. How to construct KG

How to construct KG

- Knowledge representation
  - Graph
  - Triple

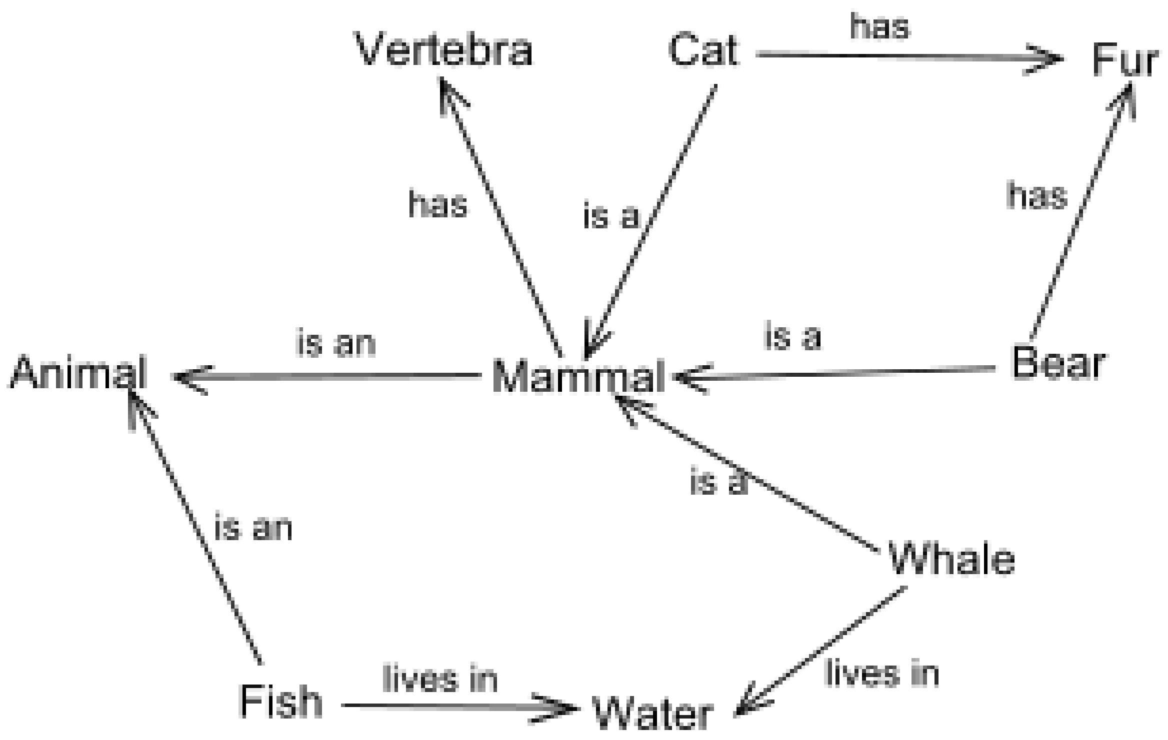
- Knowledge acquisition
  - Entity recognition
  - Relation extraction

Knowledge representation

- Graph
- Triple

Knowledge representation-Graph

- **Graph: G = G(V,E), V is a set of vetxes and E is a set of edges.**



Knowledge representation-Triple

- **RDF: RDF is a standard model for data interchange on the Web. Using this simple model, it allows structured and semi-structured data to be mixed, exposed, and shared across different applications.**
- **Triple based on RDF**
  - **A triple consists of three elements: subject, predicate and object**
  - **Form of triple:**
    - **<Subject, Predicate, Object>**
    - **<Subject, Property, Property value>**
  - **Examples:**
    - **<Aristotle, influencedBy, Plato>**
    - **<Boethius, placeOfDeath, Pavia>**

Knowledge acquisition

- Entity recognition (Vertex)
- Relation extraction (Edge)

Knowledge acquisition-Entity recognition

- **The first step of construct a knowledge graph is to acquire entities and this step contains two part:**
    - **Part1: Extract as many high-quality vocabularies as possible from source documents.**
        - **Domain phrase mining**
        - **Synonym mining**
        - **Abbreviation mining**
    - **Part2: Select targeted enities from these vocabularies for KG.**
        - **Entity recognition**

Knowledge acquisition-Entity recognition

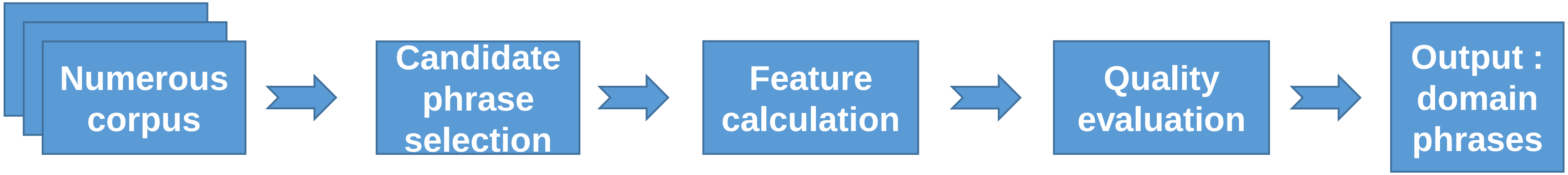
Domain Phrase Mining

Knowledge acquisition-Entity recognition

- **Domain phrase mining**

Knowledge acquisition-Entity recognition

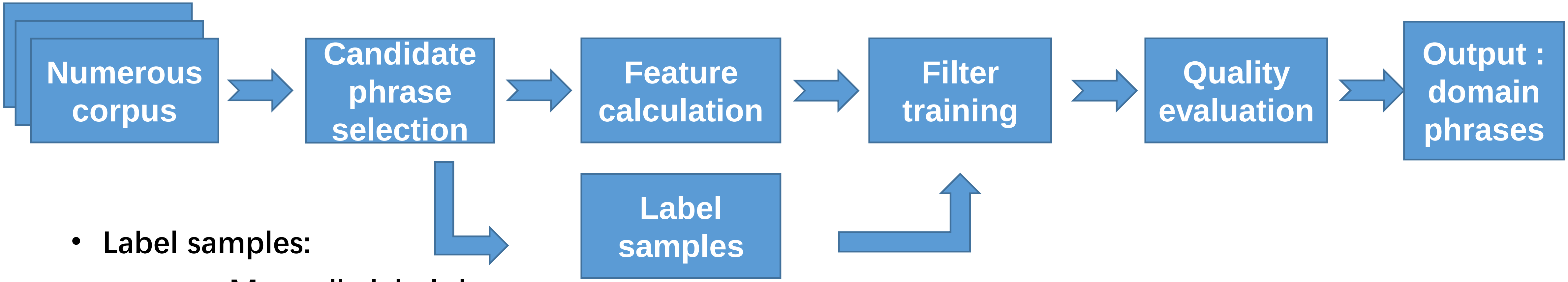
- **Unsupervised domain phrase mining:**



- **1. Candidate phrase selection: Use N-gram to find the most frequent phrases**
- **2. Feature calculation: Calculate indexes like** *TF-IDF, PMI,* **etc, for each phrase**
- **3. Quality evaluation: Combine those indexes to get a score**
- **4. Output the phrases in order of the score**

Knowledge acquisition-Entity recognition

- Supervised domain phrase mining:



- Label samples:
  - Manually label data
  - Remote supervision and annotation, using online knowledge bases (Baidu Baike, Wikipedia, etc.) as the source of high-quality phrases because high-quality phrases should be an entry in the online knowledge base
- Filter trainning: Train a binary classifier, using statistical indexes as the corresponding features

Knowledge acquisition-Entity recognition

- **Indexes for supervised/unsupervised learning:**
  - **TF-IDF: Mining phrases that can effectively represent the characteristics of a document**
  - **C-value: Utilize the relationship between the phrase and its parent to mine high-quality phrases**
  - **NC-value: On the basis of C-value, the context is further considered to mine high-quality phrases**
  - **PMI: Mining phrases that appear together regularly**
  - **Entropy: Mining phrases whose neighbor words are abundant**

Knowledge acquisition-Entity recognition

- **TF-IDF**

For a corpus of documents D:

- Term frequency (TF): $P(w|d)$
- Inverse document frequency (IDF): $\log\left(\frac{|D|}{|\{d \in D | w \in d\}|}\right) \qquad (\log(0) = 0)$
- TF-IDF: TF $\times$ IDF

Knowledge acquisition-Entity recognition

- **C-value**

$$C\text{-}value(a) = \begin{cases} log_2|a| \cdot f(a) & a \text{ is not nested}, \\ log_2|a|(f(a) - \frac{1}{P(T_a)} \sum_{b \epsilon T_a} f(b)) & otherwise \end{cases}$$

- *a* **represents a word. '***a is not nested'* **means '***a has no parent phrase***'.**
- **C-value always has rewards for longer phrases, and it is generally believed that longer phrases are more likely high-quality phrases in the field**
- **This equation eliminates the deviation caused by the repetition statistics of the parent phrase when counting the frequency.** *Ta* **is the collection of all parent phrase of** *a*

Knowledge acquisition-Entity recognition

- **NC-value**

$$\text{NC}-\text{value}(u) = 0.8\text{C}-\text{value}(u) + 0.2\sum_{b\in C_u} f_u(b)\text{weight}(b)$$

$$\text{weight}(b) = \frac{t(b)}{n}$$

- $f_u(b)$ **refers to the number of occurences of** *b* **as the context of** *u*
- *weight(b)* **is the weight to measure the importance of** *b*. *t(b)* **refers to the number of co-occurences between top k% candidate phrases and word** *b* **and** *n* **regerts to the number of occurences of top k% candidate.**

Knowledge acquisition-Entity recognition

- **PMI(Pointwise Mutual Information)**

$$\mathrm{PMI}(u_l, u_r) = \log \frac{p(u)}{p(u_l)p(u_r)}$$

- **Assuming the phrase u is composed of u_l and u_r, the greater the PMI of u_l and u_r is, the more likely u is a combination of u_l and u_r.**
- **$f(u)$ indicates the probability that the phrase u appears completely, $p(u_l)$ and $p(u_r)$ indicate probability of independent appearance of u_l and u_r.**
- **The smallest PMI value of all possible split pair (u_l, u_r) indicates the PMI value of the phrase u.**

Knowledge acquisition-Entity recognition

- **Entropy**
  - **A qualified phrase should have a abundant set of left and right neighbors**
  - **Entropy evaluate the abundance of a phrase**

$$H(u) = -\sum_{x \in \chi} p(x)\log p(x)$$

- *p(x)* **is the probability of a certain left neighbor (right neighbor) word,** χ **is the set of all left neighbor (right neighbor) characters of** u**.**
- **The larger** *H(u)* **is, more abundant the set of** *u***'s neighbors is.**
- **We choose the smaller of the two** *H(u)* **to measure the quality of the phrase.**

Knowledge acquisition-Entity recognition

Synonym mining

Knowledge acquisition-Entity recognition

- Synonym indicates words that have the same or similar meaning.
- Methods to mine synonyms:
  - Search synonyms in dictionaries, web dictionaries and encyclopedia entries
  - Design a pattern to match synonyms in documents
    - High accuracy, low recall rate
  - Bootstrap
    - This method starts from some seed samples or predefined patterns, and continuously learns new expression patterns of synonyms in the text from the corpus, thereby improving the recall rate.
  - Other methods could refer to papers:
    - https://arxiv.org/abs/1301.3781
    - http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.677.4175&rep=rep1&type=pdf
    - https://ieeexplore.ieee.org/abstract/document/5963679

Knowledge acquisition-Entity recognition

Abbreviation extraction

Knowledge acquisition-Entity recognition

- Abbreviation extraction
  - Abbreviation contians contractions, crasis and acronyms.
  - The process of extraction
    - Obtain abbreviation pairs by extraction based on pattern
    - Refine those pairs
  - Other methods to get abbreviation pairs(word, abbreviation of this word)
    - Prediction based on pattern, CRF or deep learning

Knowledge acquisition-Entity recognition

Named Entity Rcognition(NER)

Knowledge acquisition-Entity recognition

- **NER** refers to the process of locating the boundaries of named entities in the text and classifying them into a set of predefined type

- l_s, $l_e \in \{1, N\}$, l_s represents the starting position of the entity in the sentence and
- l_e represents the ending position of the entity in the sentence.
- t indicates the type of this entity.

Knowledge acquisition-Entity recognition

- **Traditional methods of NER**
  - **Based on pattern, dictionaries and online knowledge base**
  - **Based on supervised learning**
  - **Based on semi-supervised learning**
- **NER method based on deep learning**

Knowledge acquisition-Entity recognition

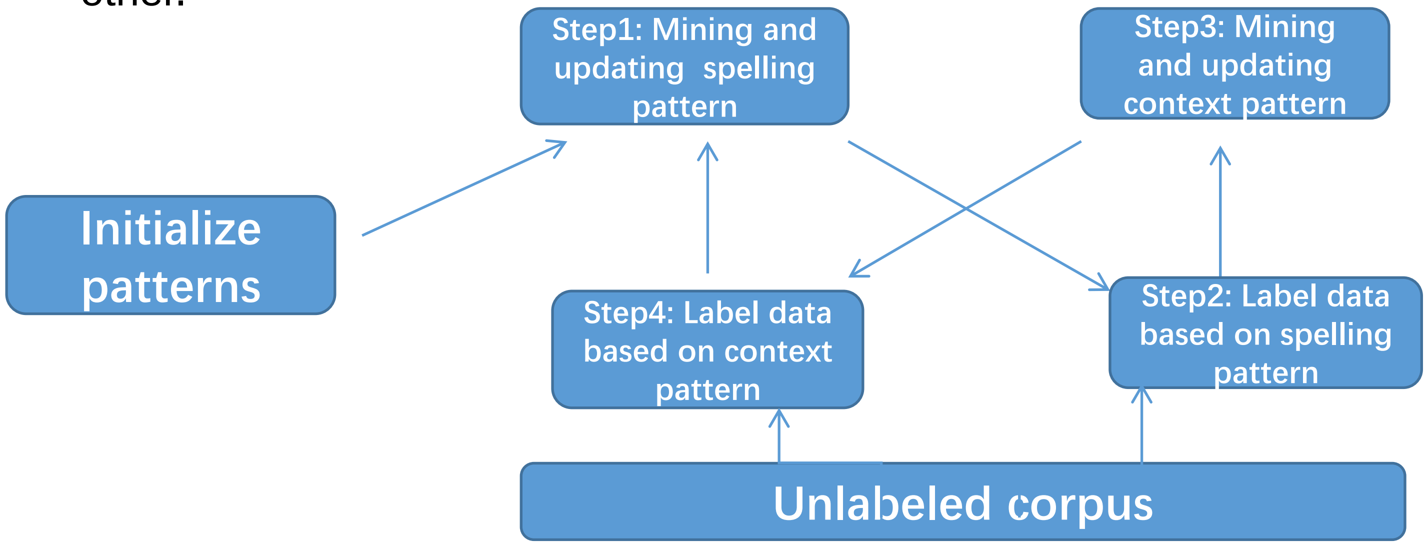
- **Methods based on pattern, dictionaries and online knowledge base**
    - Rule-based entity recognition systems often need to use entity dictionaries. When the dictionaries are exhaustive, the system will perform well, but the dictionaries are usually incomplete, which leads to a low recall rate of the system.
    - Relying on linguistic experts to manually construct rules, each rule is given a weight. When the rules conflict, the rule with the highest weight is selected

Knowledge acquisition-Entity recognition

- Methods based on supervised learning
  - NER is modeled as a sequence labeling problem in this method
  - Models for common sequence labeling problems:
    - HMM(Hidden Markov Model)
    - CRF
  - The method based on supervised learning learns entity labeling patterns in text from large-scale sequence labeling samples, and then uses this pattern to label new sentences.

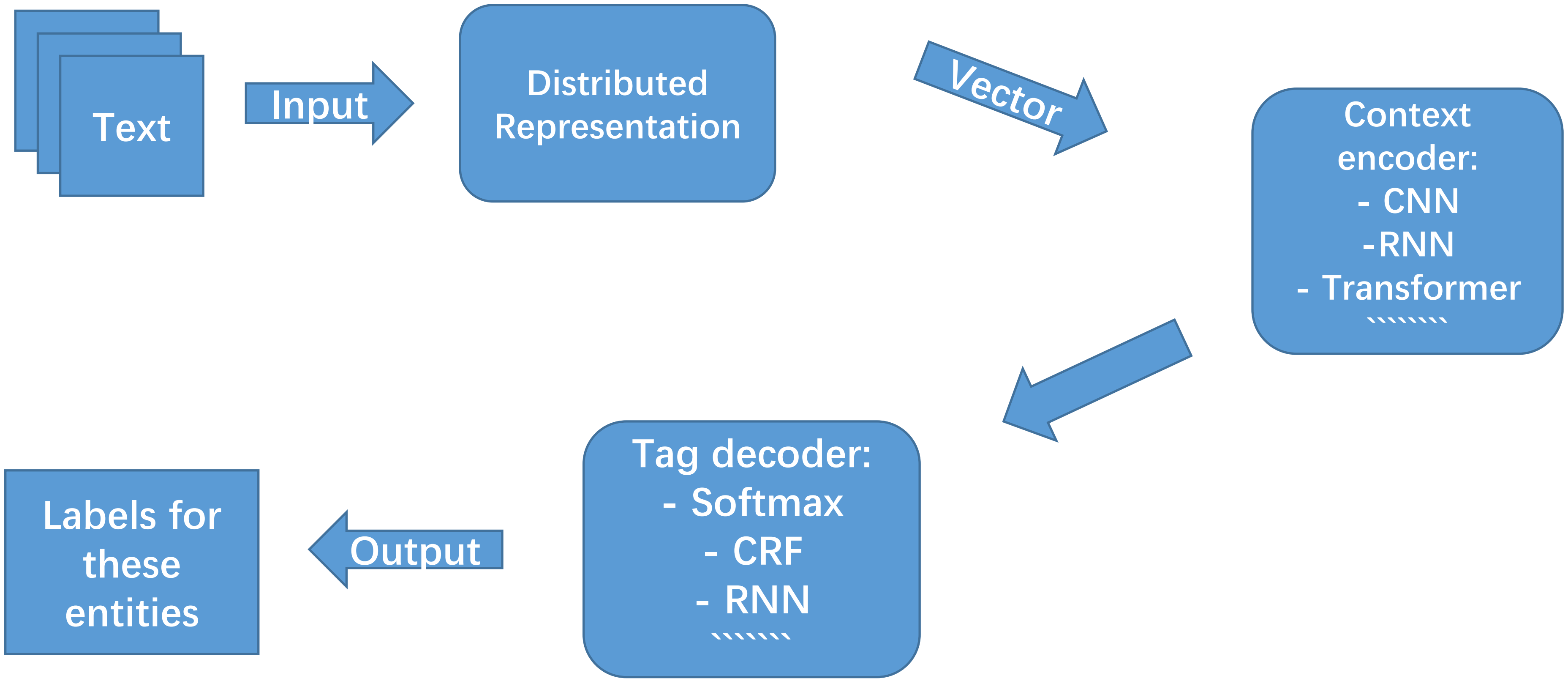
Knowledge acquisition-Entity recognition

- Methods based on semi-supervised learning
  - The first type of semi-supervised learning method is Bootstrap
    - This method usually start with a small amount of labeled data, a large amount of unlabeled data, and the initial group hypothesis or classifier.
    - Then, iteratively generate more labeled data, until a certain threshold is reached
  - Another is proposed by M.Collins and Singer[4], called Co-training
    - This method aims at learning two patterns for NER and use one to provide weak supervision for the other.



Knowledge acquisition-Entity recognition

- **Methods based on deep learning**



Knowledge acquisition-Relation extraction

Relation extraction

Knowledge acquisition-Relation extraction

- **Relation extraction:** extract the relationship(like <subject, predicate, object>) between entities and entities from unstructured text.
- **Methods:**
    - **Extraction based on patterns**
    - **Extraction based on learning**
    - **Evaluation method and metric**

Knowledge acquisition-Relation extraction

- **Methods:**
  - **Extraction based on patterns**
  - **Extraction based on learning**
    - **Extraction based on supervised learning**
    - **Extraction based on distance supervised learning**
    - **Extraction based on deep learning**
    - **Evaluation**

Knowledge acquisition-Relation extraction

- **Extraction based on patterns:**
  - **Character-based extraction:**
    - Construct regular expressions for string matching
    - It's only suitable for fixed description, highly templated content
  - **Grammar-based extraction:**
    - This method describes the extraction pattern by introducing the grammatical information contained in the text, relax the rigor of the pattern by adding grammatical constraints, and improve the expressive ability.
    - It's still easy to make mistake in extraction
  - **Semantic-based extraction:**
    - Utilize semantic elements (such as concepts) to express the scope of pattern's adaptation more accurately and improve the accuracy of this pattern.
    - This method can reduce the possibility of semantic drift but heavily relies on completeness of concept graph

Knowledge acquisition-Relation extraction

- **Extraction based on supervised learning:**
  - Applicable to situations where large-scale labeled data exists
  - Manual labeling is expensive and difficult to generalize to large-scale tasks

Knowledge acquisition-Relation extraction

- **Extraction based on distance supervised learning:**
  - **Basic assumption:** Given a triple <s, r, o>, any sentence containing <s, o> describes the relationship between the pair of entities
  - **Steps:**
    - Obtain as many entity pairs as possible from knowledge base(Like Freebase)
    - For each entity pair, use entity links to extract a set of sentences mentioning the entity pair from the large-scale text, and labe each sentence with the corresponding relationship
    - The sentence set containing the entity pair and the relationship type label constitutes the relation extraction's training data set
    - Use this data to train the models

Knowledge acquisition-Relation extraction

- **Extraction based on deep learning:**
  - **Compared with traditional learning methods, deep learning-based methods can reduce the cost of manually designing features in traditional relational classification modelsand make use of some invisible features**
  - **Deep neural network models require a large amount of labeled data and are often used in conjunction with remote supervised learning**

Knowledge acquisition-Relation extraction

- **Evaluation:**

Total Population n = a number		Predicted Class: False (0)	Predicted Class: True (1)
Actual Class	False (0)	TN True Negative	FP False Positive
	True (1)	FN False Negative	TP True Positive

$$\text{Precision} = \frac{tp}{tp + fp}$$

$$F_1 = \left(\frac{2}{\text{recall}^{-1} + \text{precision}^{-1}}\right) = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

$$\text{Recall} = \frac{tp}{tp + fn}$$

$$\text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn}$$

III. How to manage KG

How to manage KG

- **Data modeling&Storage**
- **Query&Retrieval**

Data modeling

- **Triple:**
  - It's unable to express complex semantics including multiple relationships, spatiotemporal information, and multimodal information
- Expand triples to **quintuples** to represent temporal and spatial knowledge
- **Multimodal knowledge:**
  - Model multi-modal resources (such as pictures) as an entity, and define relationships such as "picture resource links", so that they can be expressed as a triplet

Data modeling

- **Graph:**
  - **Directed graph:** In the triples, entities are modeled as nodes and relationships are modeled as edges. The knowledge graph represented by the triples can be naturally converted into a directed graph
  - **Attribute graph:** On the basis of a directed graph, each edge and node has a type label. For example, Plato’s label is an idealist philosopher, and each edge and node has a set of attributes
  - **Tree:** In fact, it is a graph without loops, which is mainly used to define semantic relations, such as the upper and lower relations of synonyms. It is used more in concept maps (directed acyclic graphs)

Data modeling

- **Graph:**
    - **Directed graph with weight and probability:** Increase the ability of probabilistic modeling. For each ***isA*** relationship, the frequency of the relationship observed in the corpus can be added to quantify the credibility of the relationship.
    - **Heterogeneous Information Network**: The same graph contains different types of points and edges (for example, some edges are directed edges, and some are undirected edges), which is usually used when using knowledge graphs to model other information networks (such as social network data)

Storage

- **Storage based on relational model**
  - **Storage method based on three lists**

Subject	**Predicate**	**Object**
Socrates	student	Plato
Socrates	birthTime	469 BC
《The Republic》	isA	book
Idealist philosopher	subclassOf	philosopher

  - **Other methods based on attibute table, vertical table and full index**
- **Storage based on graph model**

Storage

- **Storage based on graph model**
  - **Adjacency list:** Each entity corresponds to a row, storing information related to that entity

Node	Adjacency list
Plato	(isA idealist philosopher), (birthPlace Athens)......
Socrates	(isA philosopher), (birthTime 469 BC)......
......	......

  - **Adjacency matrix:** Use a matrix of *subject number * relation number * object number* size to store knowledge
  - **comparison**: The adjacency matrix is expensive, but the query is faster (especially the query that expresses the subgraph)

Query&Retrieval

- **SPARQL**: A query language for RDF(Resource Description Frame work) data
- Like SQL, SPARQL is a declarative structured language
- SPARQL provides a complete set of query operators, including selecting, sorting, aggregating operators and so on. There is no need to declare additional schema definitions

IV. How to apply KG

How to apply KG

- **Language cognition based on KG**
- **How KG facilitate search and recommendation**
- **Q&A based on KG**

Language cognition based on KG

- **What does understanding natural language mean to machine?**
  - It means the process of forming the corresponding internal representation after the machine accepts natural language input.
- **Why is it difficult for machine to understand natural language?**
  - Human's ability of understanding language is based on their cognition and this kind of cognition derives from a huge amount of background knowledge.
- **How KG make it possible for machine?**
  - KG could link text in natural language to entities, concepts, relation and substructure and provide a kind of representation for machine.

How KG facilitate search and recommendation

- KG helps to **perfect the portrait of the object**. Knowledge graph can enrich and enhance the description of users and various Internet resources. An accurate and comprehensive portrait is the prerequisite for accurate query matching (for search) and item matching users (for recommendation)
- KG can **discover the semantic relationship** between queries (users) and answers (items). (Semantic relation ship: “Mountaineers - Love climbing - Climbing equipment is needed - including Trekking poles”)
- KG **provides interpretable basis for search and recommendation**.
- KG **provides a cognitive framework for users' information exploration**.

Q&A based on KG

- **Techniques of QA system**:
  - Reading comprehension: searching answer from one document or several documents.
  - KBQA(knowledge base QA): searching answer from one huge triple table
  - RQA(Relational base QA): searching answer from multiple tables
- **Why KBQA is more advanced**?
  - Provide background for semantic understanding
  - Enable preliminary reasoning skill

V. Reference

Reference

- [1] Mikolov T, Chen K, Corrado G, et al. Efficient Estimation of Word Representations in VectorSpace[J]. Computer Science, 2013.
- [2] Hu F, Shao Z, RUan T. Self-Supervised Synonym Extraction from Web[J]. J. Inf. Sci. Eng., 2015,31(3): 1133-1148.
- [3] Cheng T, Lauw H W, Paparizos S. Entity synonyms for structured web search[J]. IEEE transactions on knowledge and data engineering, 2012, 24(10): 1862-1875.
- [4] Michael Collins and Yoram Singer. Unsupervised models for named entity classification. In proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora, 1999.
- [5] Xiao Yanghua. Knowledge Graph-Concepts and Techniques. Publishing House of Electronics Industry, Beijing, 2020.