



Problem Statement

Let's get acquainted with the task of building an entity linking system.

We'll cover the following



- Introduction
- Applications
- Problem statement
- Interview questions

Introduction#

Named entity linking (NEL) is the process of detecting and linking entity mentions in a given text to corresponding entities in a target knowledge base.

There are two parts to entity linking:

1. *Named-entity recognition*


Named-entity recognition (NER) detects and classifies potential **named entities** in the text into predefined categories such as a person, organization, location, medical code, time expression, etc. (multi-class prediction).

2. *Disambiguation*

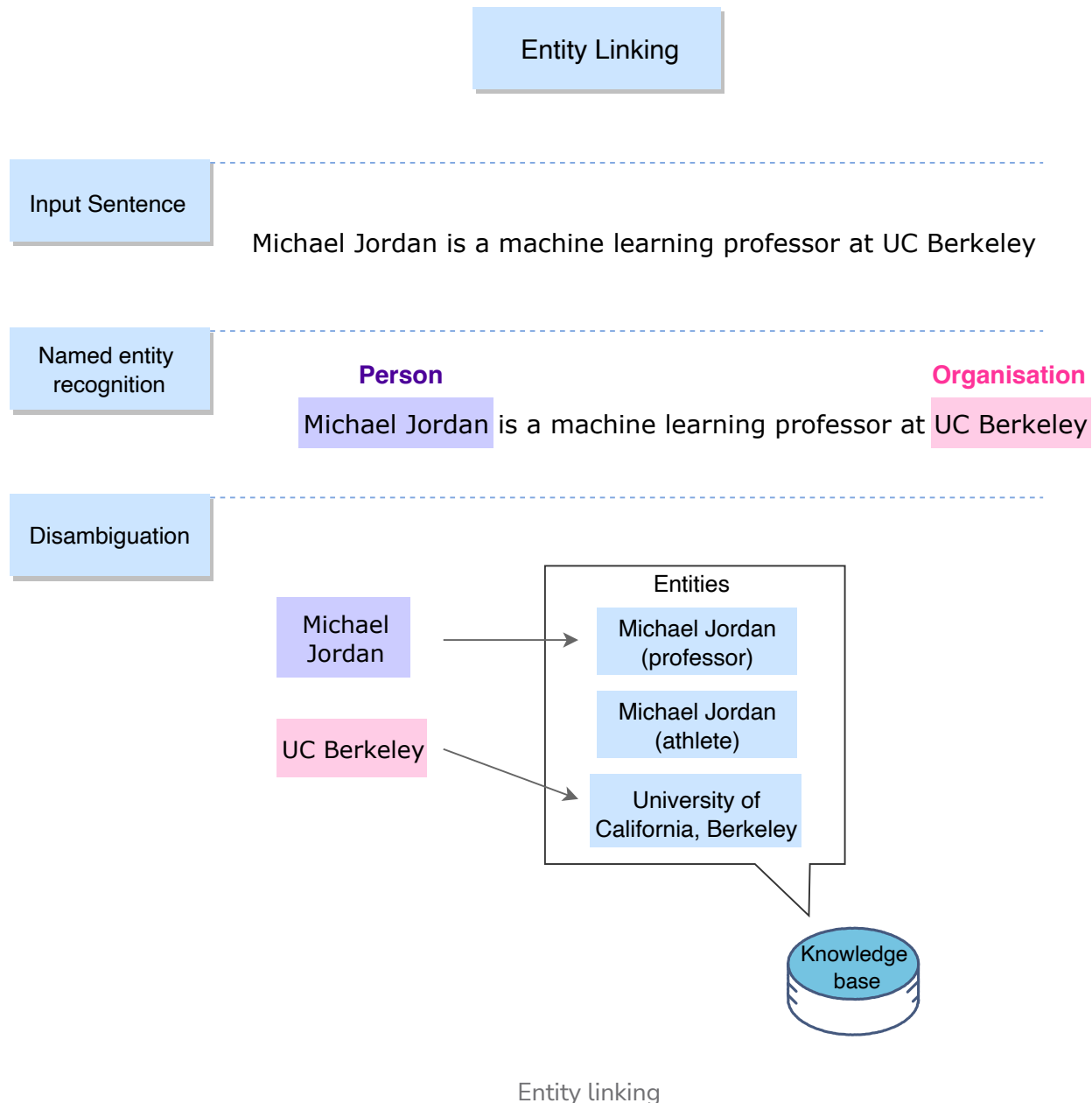
Next, disambiguation disambiguates each detected entity by linking it to its corresponding entity in the knowledge base.





 The target knowledge base depends on the application, but for generic systems, a common choice is Wikidata or DBpedia.

Let's see entity linking in action in the following example.



The sentence/text says, “Michael Jordan is a machine learning professor at UC Berkeley.” First NER detects and classifies the named entities Michael Jordan and UC Berkeley as person and organisation.



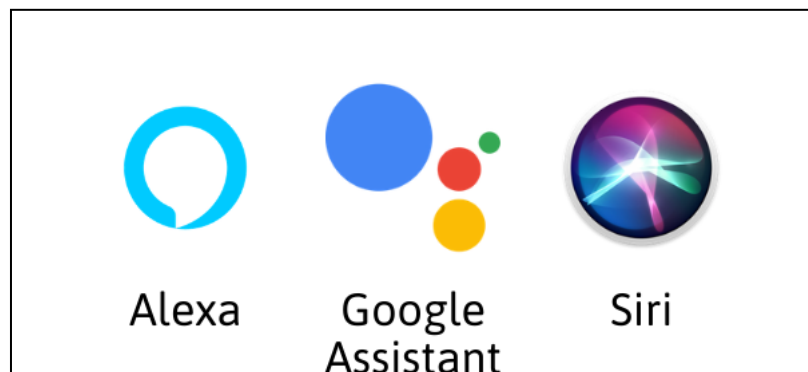


Then disambiguation takes place. Assume that there are two ‘Michael Jordan’ entities in the given knowledge base, the UC Berkeley professor and the athlete. Michael Jordan in the text is linked to the *professor at the University of California, Berkeley* entity in the knowledge base (that the text is referring to). Similarly, UC Berkeley in the text is linked to the *University of California* entity in the knowledge base.

Applications#

Entity linking has applications in many natural language processing tasks. The use cases can be broadly categorized as information retrieval, information extraction and building knowledge graphs, which in turn can be used in many systems, such as:

- Semantic search
- Content analysis
- Question answering systems/chatbots/virtual assistants, like Alexa, Siri, and Google assistant



NEL applications. Image source: www.alexeko.com

All of the above-mentioned applications require a high-level representation of the text, in which concepts relevant to the application are separated from

the text and other non-meaningful data.

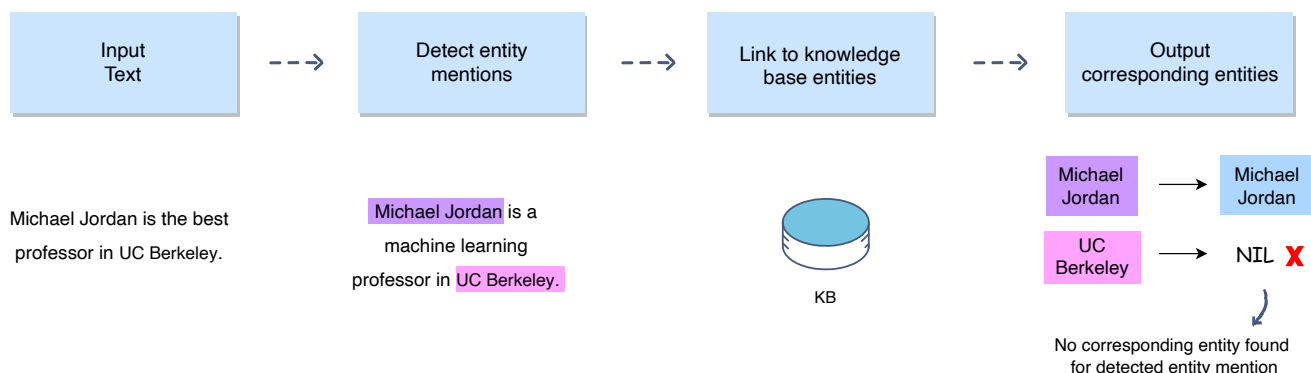


Now that we have introduced named entity linking and its applications, we are ready to look at the problem statement.

Problem statement#

The interviewer has asked you to design an entity linking system that:

- Identifies potential named entity mentions in the text.
- Searches for possible corresponding entities in the target knowledge base for disambiguation.
- Returns either the best candidate corresponding entity or nil.



Design a named entity linking system

The problem statement translates to the following machine learning problem:

"Given a text and knowledge base, find all the entity mentions in the text(Recognize) and then link them to the corresponding correct entry in the knowledge base(Disambiguate)."

Interview questions#



These are some of the questions that an interviewer can put forth.



1. How would you build an entity recognizer system?
2. How would you build a disambiguation system?
3. Given a piece of text, how would you extract all persons, countries, and businesses mentioned in it?
4. How would you measure the performance of a disambiguator/entity recognizer/entity linker?
5. Given multiple disambiguators/recognizers/linkers, how would you figure out which is the best one?

[← Back](#)[Next →](#)

Modeling

Metrics

[Mark as Completed](#)[Report an Issue](#)