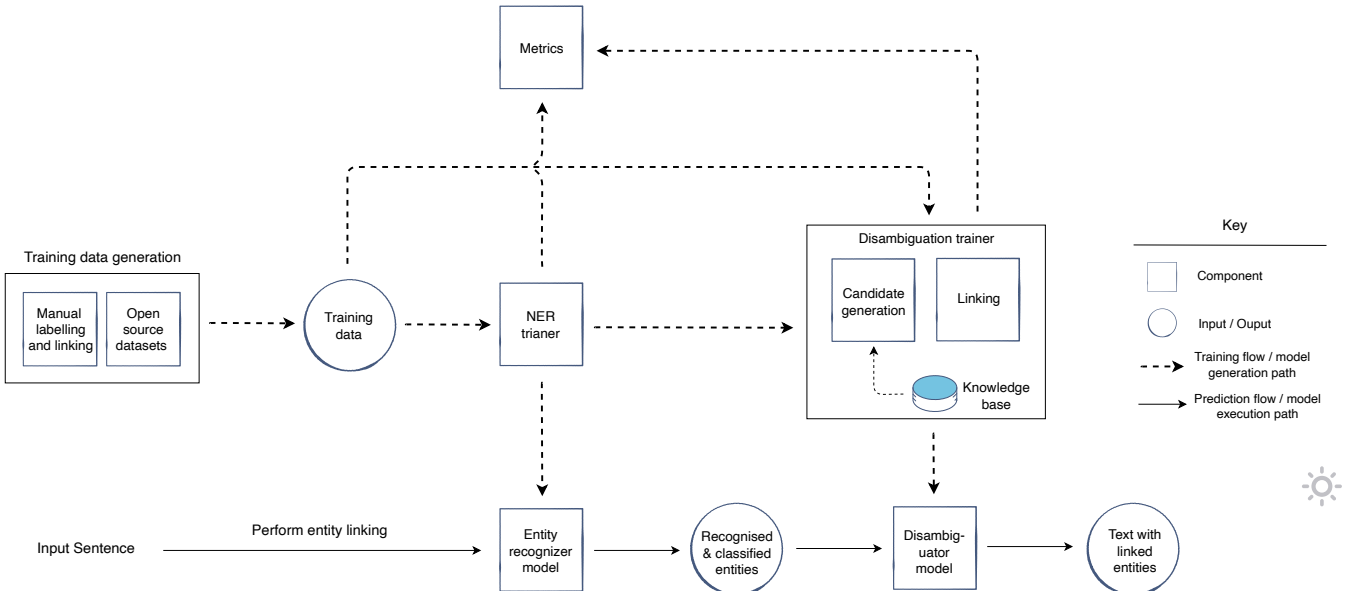
Architectural Components

Let's go over the architectural components of the entity linking unit.

We'll cover the following

The architectural components diagram for entity linking is shown below. It consists of two paths:

1. Model generation path (training flow)
2. Model execution path (prediction flow)



Architectural diagram for entity linking

> 📝 The terms *entity mention* and *recognized entity* are used interchangeably.

Model generation path#

Model generation is responsible for training models for entity linking task. Let's look at the components of this path.

Training data generation#

You will begin by gathering training data for entity linking through open-source datasets and manual labelling/linking of text. These methods will be detailed in the next lesson.

You will pass the training data to the named entity recognition (NER) model trainer and the named entity disambiguation (NED) model trainer.

NER#

NER is responsible for building a machine learning model to recognize entities, such as a person, organization, etc., for a given input text.

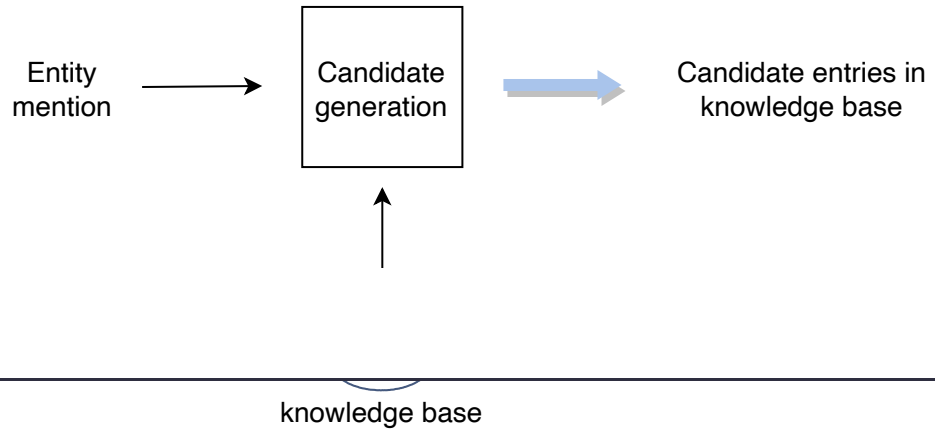
NED#

The disambiguation component will receive the output of the NER for linking. The disambiguation process has two phases:

1. Candidate generation

   The first phase of disambiguation is candidate generation. It finds potential matches for the entity mentions, by reducing the size of the knowledge base to a smaller subset of *candidate documents/entities*. This

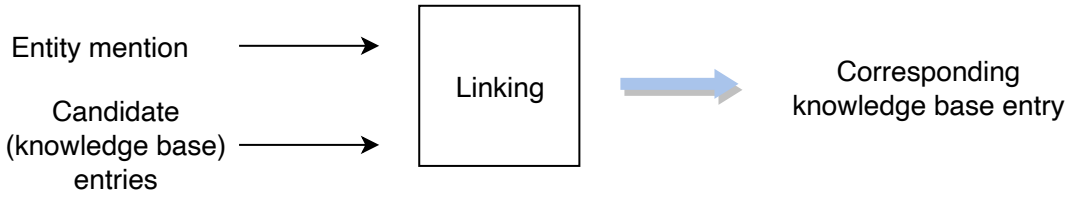
knowledge base to a smaller subset of *candidate documents/entities*. This saves us from running the linking model on the entire knowledge base for each entity mention.



Candidate generation

2. Linking

   The second phase of disambiguation is linking. Here, you will select the exact corresponding entry in the knowledge base for each recognized entity. The linking model runs on only the candidate entries for each mention.



Linking

Metrics#

As explained in the previous lesson, the metrics component will measure the performance of:

1. NER component separately
2. NED component separately

3. Entity linking as a whole

> 📝 The arrow going from disambiguation to the metrics component shows that the metrics component will not only measure the quality of the disambiguation but the quality of the whole entity linking system.

Model execution path#

The model execution path is very straightforward. It begins with an input sentence that is fed to the NER component. NER identifies the entity mentions in the sentence, along with their types, and sends this information to the NED component. This component then links each entity mention to its corresponding entity in the knowledge base (if it exists). Like this, entity linking will be performed for a given piece of text.

← Back
Metrics

Next →
Training Data Generation

Mark as Completed

Report an Issue