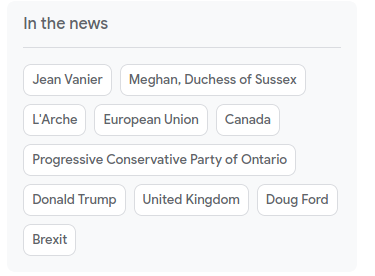
**Information Extraction**

**Information** can mean multiple things, such as key events, people, or relationships between people, places, or organizations

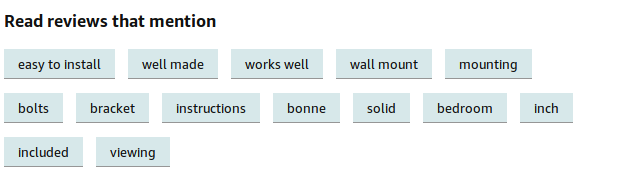
IE is to extract “knowledge” from text, and each of these tasks provides different information

Applications

1. *Tagging news and other content*



1. Keyword*or*keyphrase extraction (KPE)

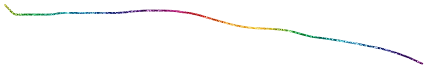


*Chatbots*

A chatbot needs to understand the user’s question in order to generate/retrieve a correct response. For example, consider the question, “What are the best cafes around the Eiffel Tower?” The chatbot needs to understand that “Eiffel Tower” and “cafe” are locations, then identify cafes within a certain distance of the Eiffel Tower

*Applications in social media*

Traffic updates and disaster relief efforts, based on tweets





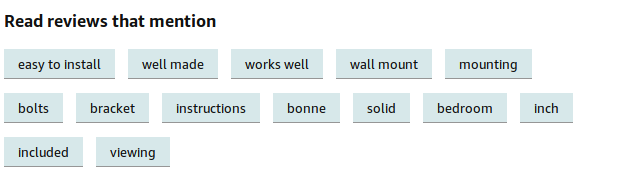


KPE- Key Phrase Extraction- Identifying key phrases like Buyback, Stock prices

NER- Apple is not a fruit but it’s a company

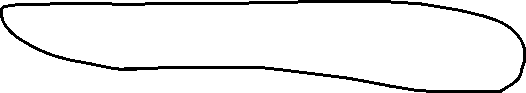
KPE- Keyword and phrase extraction, as the name indicates, is the IE task concerned with extracting important words and phrases that capture the gist of the text from a given text document

With more than 100 of reviews Amazon- you cant go through all the reviews-

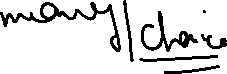


**Named Entity Recognition**

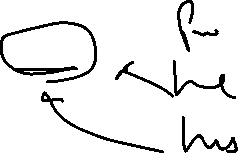
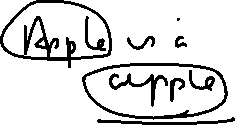
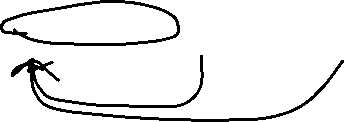
Definition -NER refers to the IE task of identifying the entities in a document. Entities are typically names of persons, locations, and organizations, and other specialized strings, such as money expressions, dates, products, names/numbers of laws or articles,

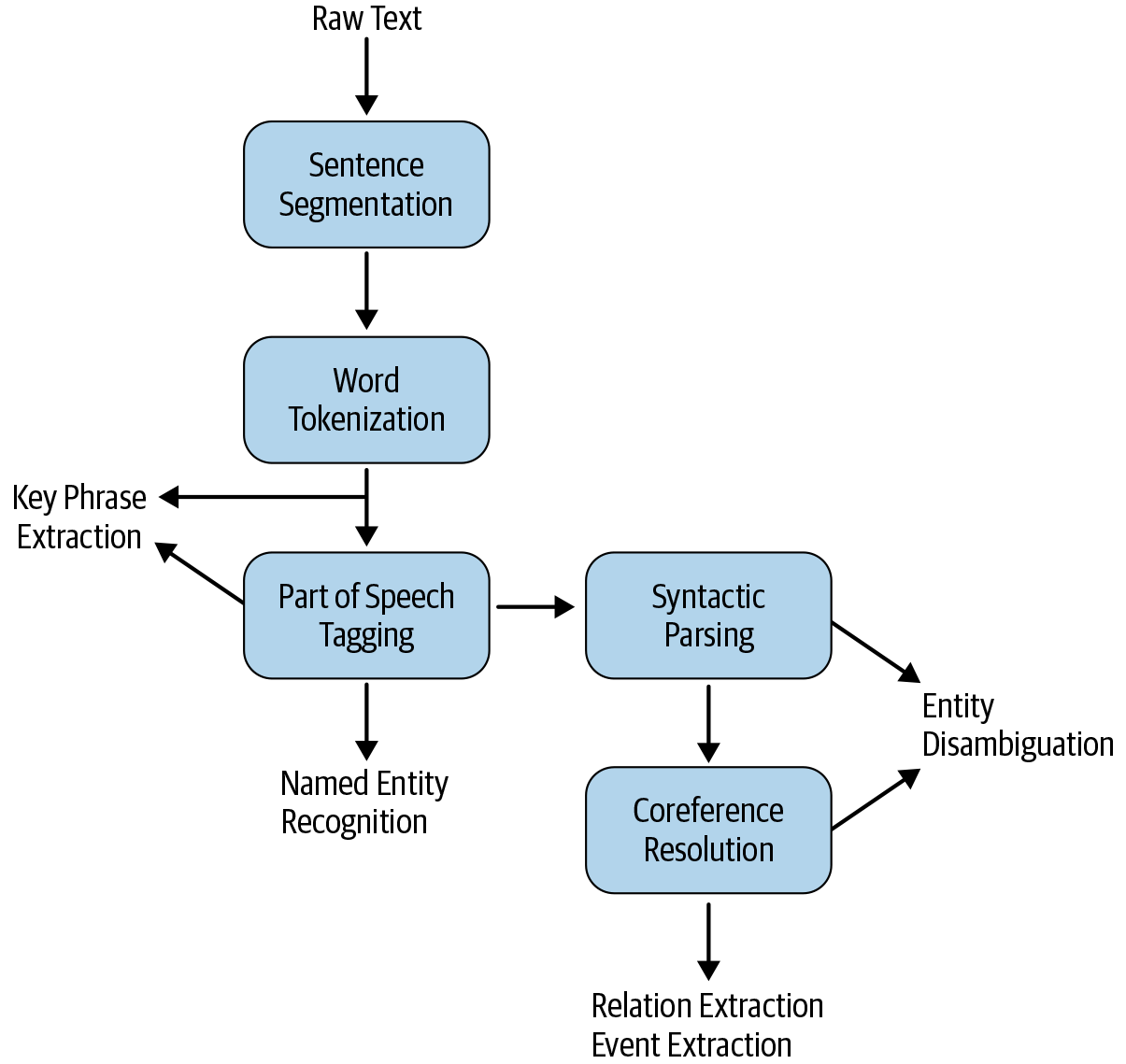


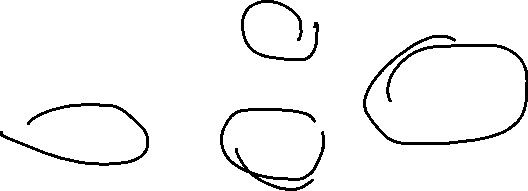
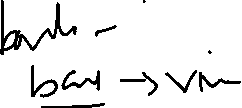
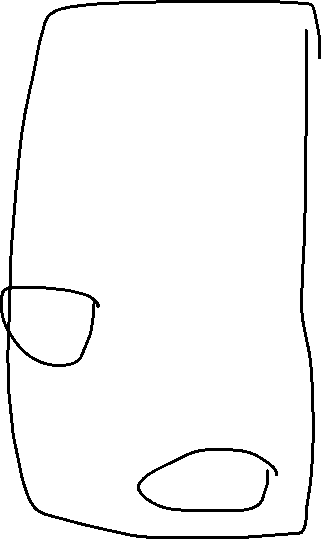
Coreference Resolution

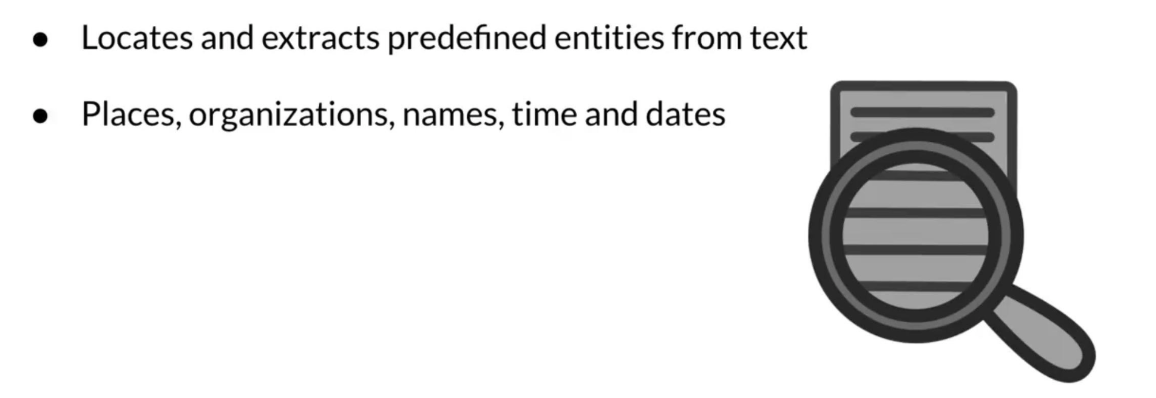


John is visited school today, he is not doing well and his stomach is upset

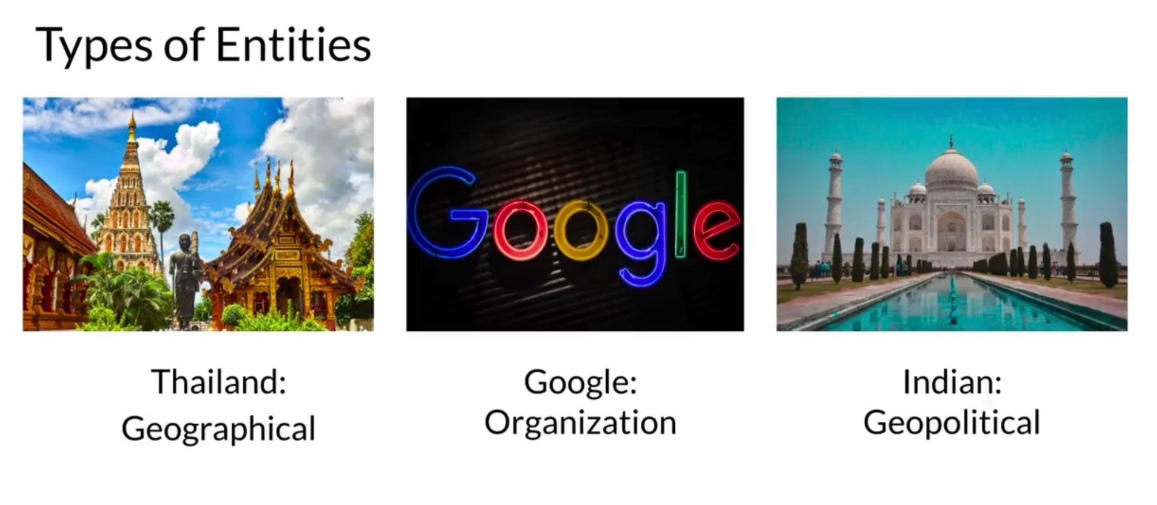




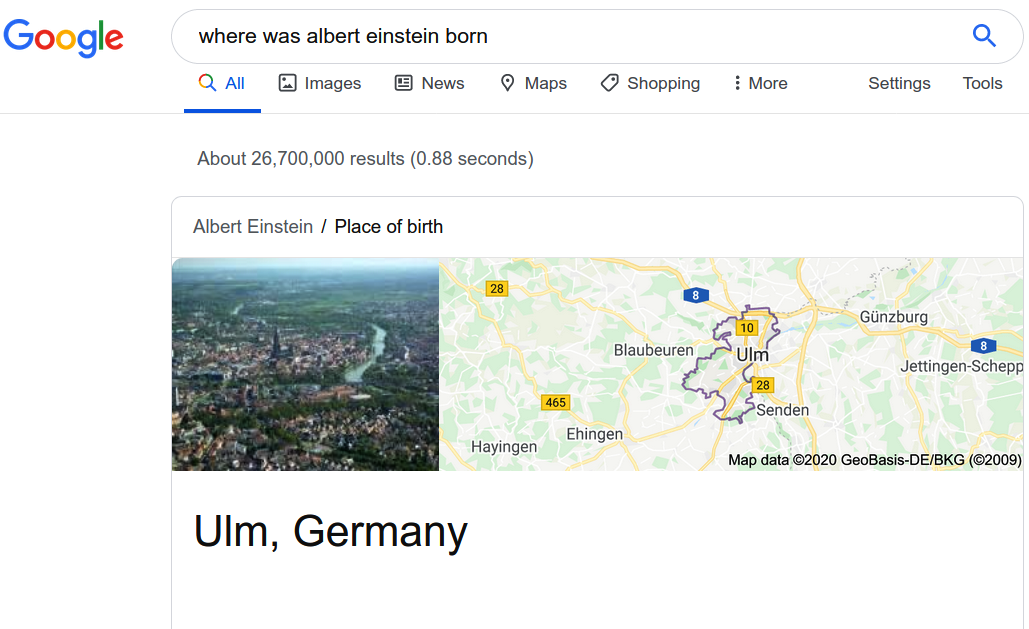




Examples



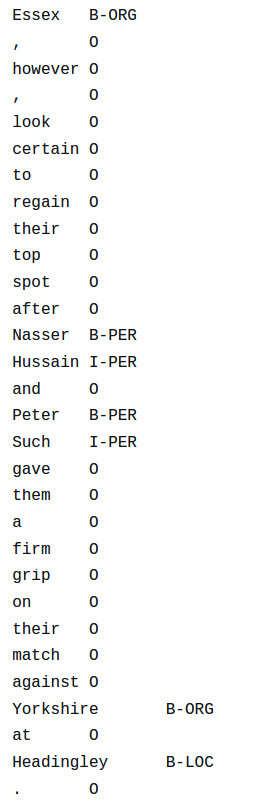




Building a NER System

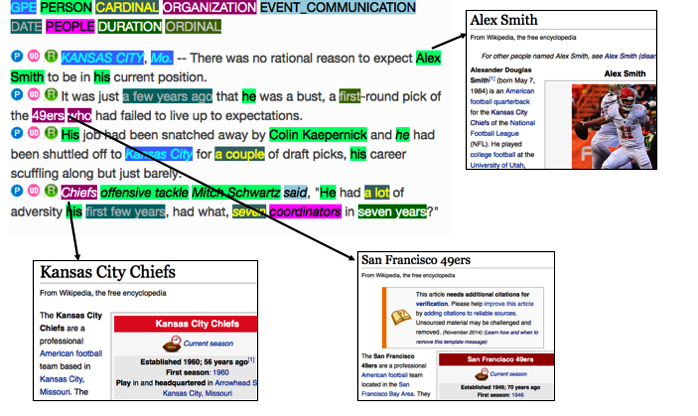
**Gazetteer**-Large collection of person/organization/location names that are the most relevant to our company

B indicates the beginning of an entity; I, inside an entity, indicates when entities comprise more than one word; and O, other, indicates non-entities



Named Entity Disambiguation and Linking

NED-assigning a unique identity to entities mentioned in the text.



“Lincoln drives a Lincoln Aviator and lives on Lincoln Way.”

* first Lincoln is a person
* second one is a vehicle
* third is a location

assigning a unique identity to entities mentioned in the text. It’s also the first step in moving toward more sophisticated tasks to address the scenario mentioned above by identifying relationships between entities. NER and NED together are known as named entity linking (NEL).

