# Named Entity Recognition and Disambiguation

Prof. Luca Cagliero
Dipartimento di Automatica e Informatica
Politecnico di Torino

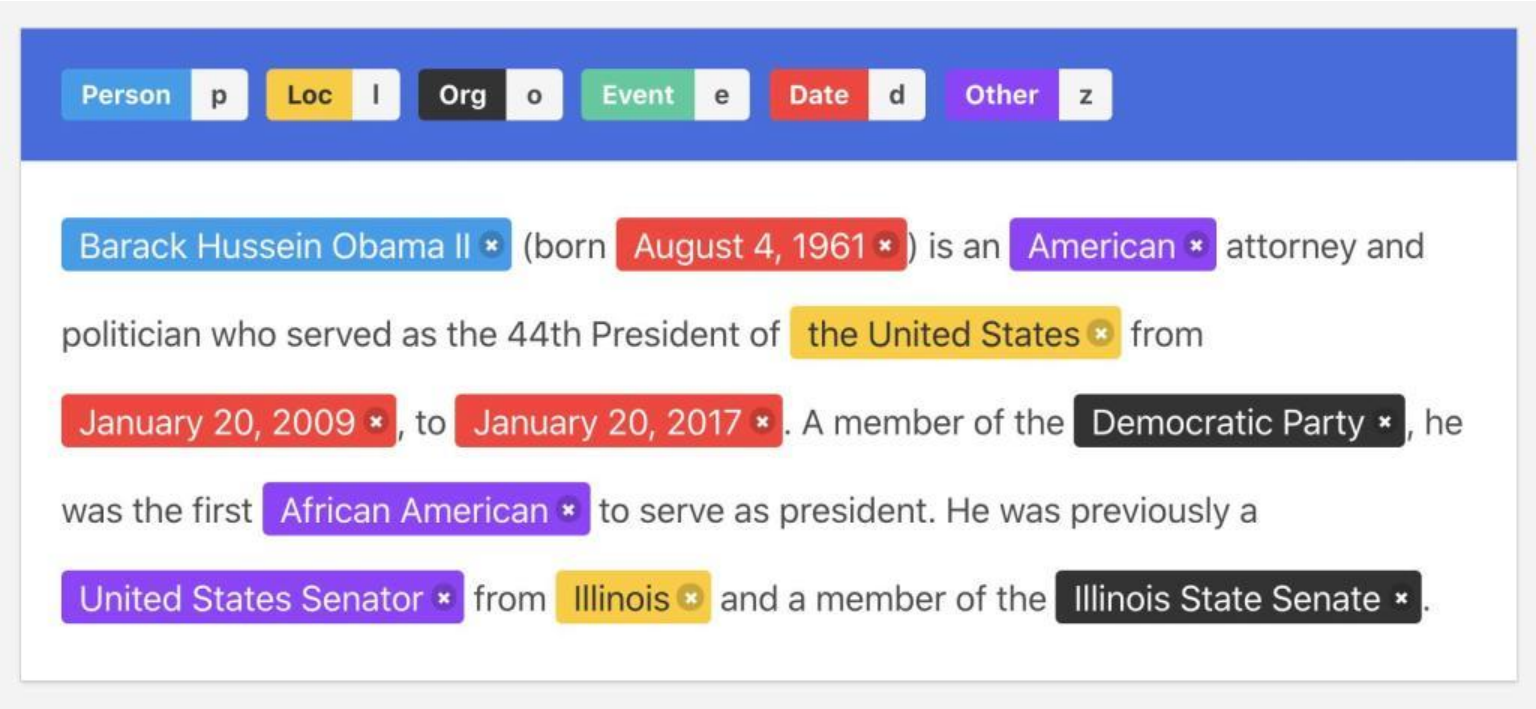

Politecnico di Torino

# Lecture goal

- Definition of NERD
- Application contexts
- Rule-based NER
- Ontology-based NER
- Machine Learning-based NER
- Transformer-based NER

# Preliminary example



# Outline

- Recognition
  - Recognize *Barack Hussain Obama II*
- Disambiguation
  - Distinguish between *Cars (movie) and Cars (vehicles)*
- Resolution
  - Fix typos (e.g., *Obama Barackk*)
  - reconduct to a normal form (e.g., *Barack Hussain Obama II)*

# Outline

- Recognition
  - Recognize *Barack Hussain Obama II*

70%

- Disambiguation
  - Distinguish between *Cars (movie) and Cars (vehicles)*
- Resolution
  - Fix typos (e.g., *Obama Barackk*)
  - reconduct to a normal form (e.g., *Barack Hussain Obama II)*

# Outline

- Recognition
  - Recognize *Barack Hussain Obama II*
- Disambiguation
  - Distinguish between *Cars (movie) and Cars (vehicles)*

20%

- Resolution
  - Fix typos (e.g., *Obama Barackk*)
  - reconduct to a normal form (e.g., *Barack Hussain Obama II)*

# Outline

- Recognition
  - Recognize *Barack Hussain Obama II*
- Disambiguation
  - Distinguish between *Cars (movie) and Cars (vehicles)*
- Resolution
  - Fix typos (e.g., *Obama Barackk*)
  - reconduct to a normal form (e.g., *Barack Hussain Obama II)*

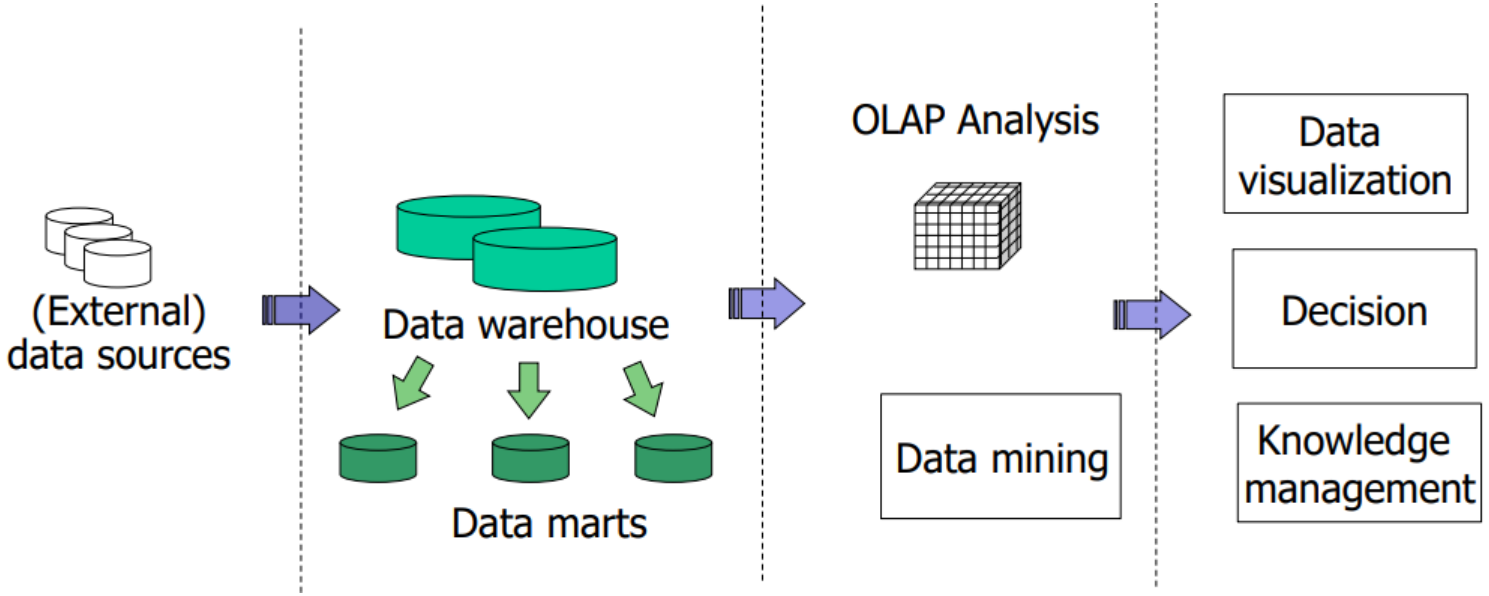
10%

# Goal

- Find semantic references to specific entities of the real world
- Applications
  - Business intelligence
  - Information retrieval
  - Text classification and intent detection
  - Question answering and chatbotting
  - Summarization

# Business Intelligence

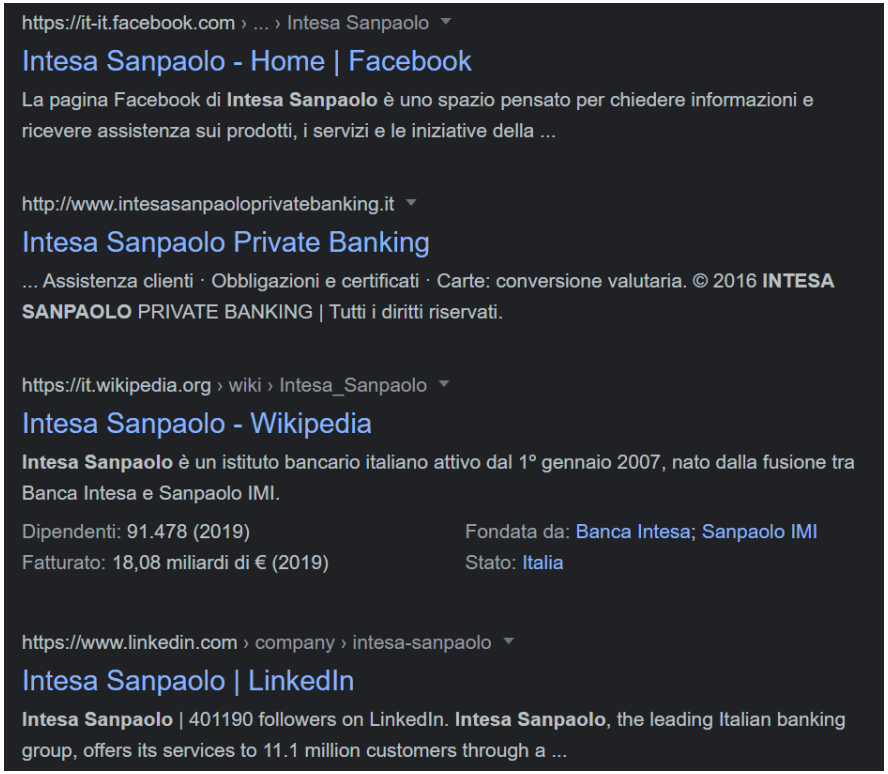
- Provide support in companies to make strategic decisions
- transform company data into actionable information




Data warehouse introduction. Elena Baralis. Politecnico di Torino. https://dbdmg.polito.it/dbdmg_web/index.php/2021/09/13/data-management-and-visualization-2021-22/

# Information Retrieval

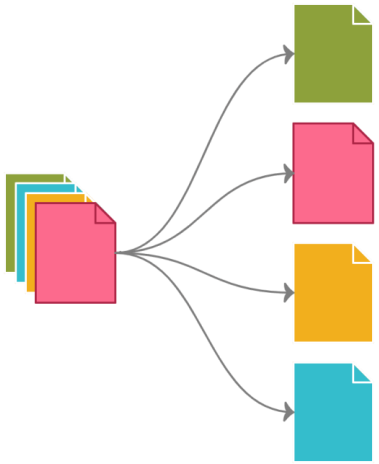
- Software systems that provides access to document collections
  - Full-text search
  - Search based on metadata



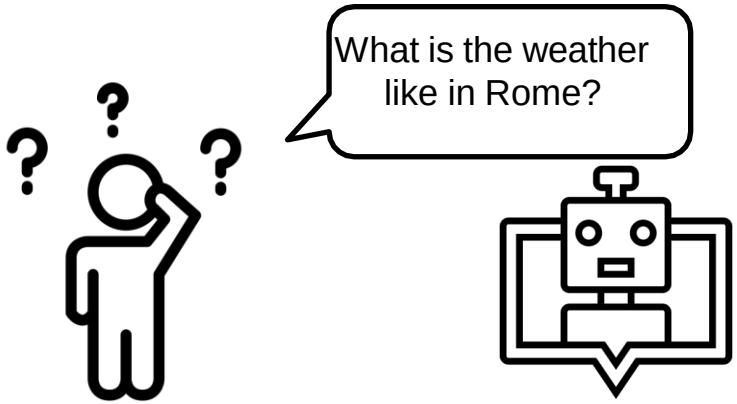

https://www.google.it/search?q=intesa+sanpaolo

# Text classification

- Classify unlabeled pieces of text into one or more predefined classes
  - It requires the availability of a set of annotated documents, i.e., the training set
- Applications
  - Sentiment analysis
  - Topic discovery
  - Language identification
  - Hate speech detection
  - Intent detection



# Intent detection




Icons credit: Creative Mania & Rahmat Dwi Cahyo from the Noun Project

# Intent detection

- The system is unaware of the action users are willing to take
- NER and NED modules are used to understand the entities involved

1 NER module → **Rome**: LOC

2 NED module → **Rome**: /wiki/Rome

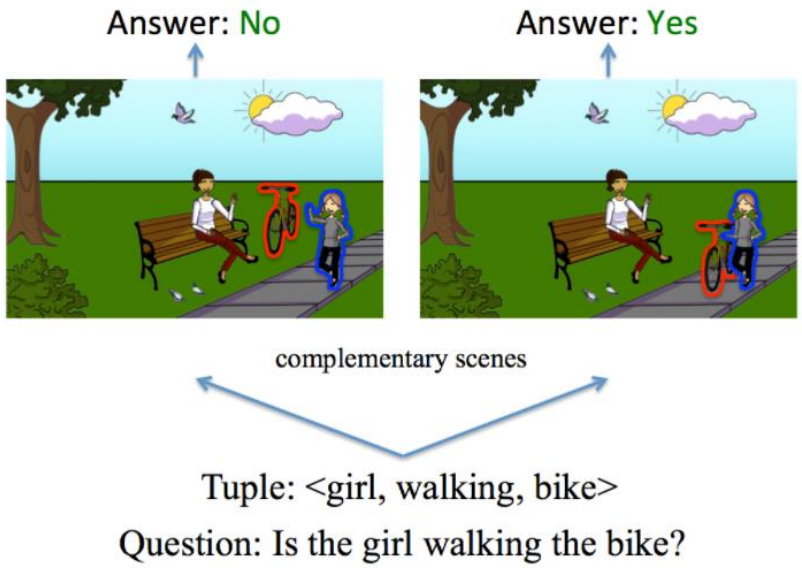
# Question answering

- Provide relevant answers in response to questions posed in natural language
- Main steps
  - Question classification
  - Information Retrieval
  - Answer extraction

flyelite.com (latest access: April 2021)

# Question answering

- Not only text...




https://visualqa.org/ (latest access: April 2021)

# Summarization

- Produce a concise version of a (multimodal) collection that incorporates the most salient content

**Transcript**

today we are going to show you how to make spanish omelet . i 'm going to dice a little bit of peppers here . i 'm not going to use a lot , i 'm going to use very very little . a little bit more then this maybe . you can use red peppers if you like to get a little bit color in your omelet . some people do and some people do n't .... t is the way they make there spanish omelets that is what she says . i loved it , it actually tasted really good . you are going to take the onion also and dice it really small . you do n't want big chunks of onion in there cause it is just pops out of the omelet . so we are going to dice the up also very very small . so we have small pieces of onions and peppers ready to go .

**Summary**

how to cut peppers to make a spanish omelette; get expert tips and advice on making cuban breakfast recipes in this free cooking video .

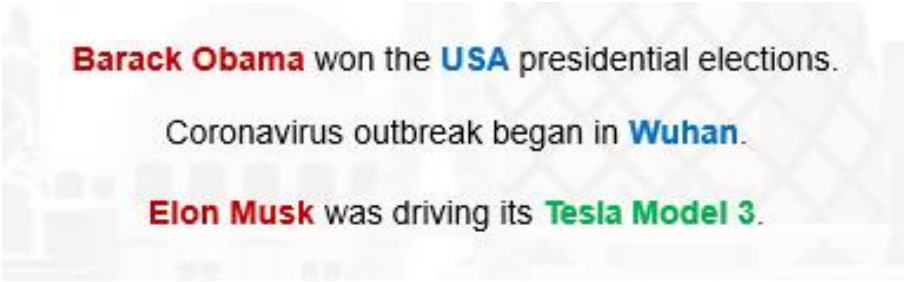
Multimodal Abstractive Summarization for How2 Videos. Shruti Palaskar, Jindrinch Libovicky, Spandana Gella, Florian Metze. https://arxiv.org/abs/1906.07901

# Entity Recognition

- The main purpose is to recognize the entities
- **Entities** consist of any word/multiword forms in a piece of text that refers to **identifiable concepts**
  - E.g., text: *Book a movie ticket*
  - "Movie" is a candidate entity

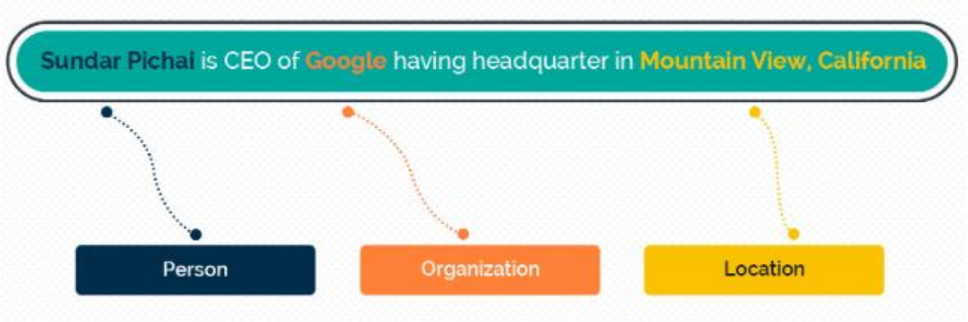
# Named Entities

- Named entities are real-world objects, persons, locations, etc..
- Unlike other entity types, they link to a specific concept in the real world

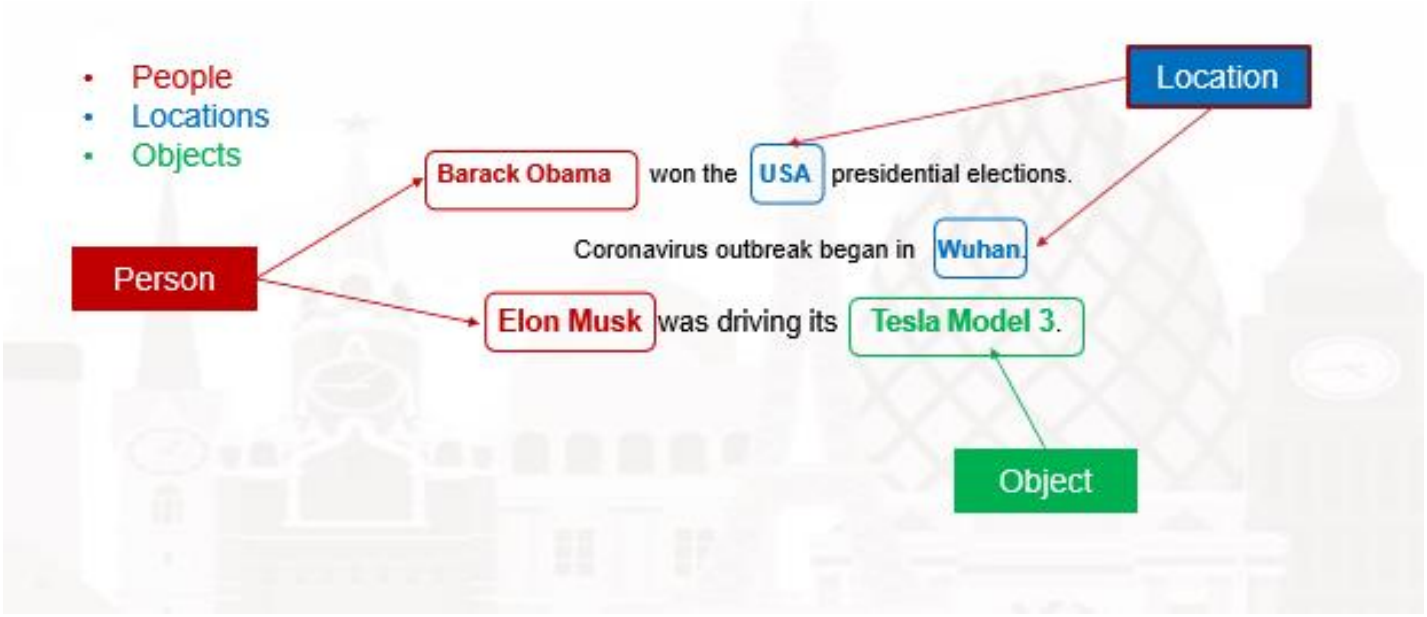


# Named Entity Recognition

- Identify the named entities in the text
  - Define the corresponding text span
- Classify the detected entities in one of the predefined set of classes
  - Assign a category (classification)



# Entity Recognition: example



# Entity recognition – Spacy NER entities

PERSON: People, including fictional.
NORP: Nationalities or religious or political groups.
FAC: Buildings, airports, highways, bridges, etc.
ORG: Companies, agencies, institutions, etc.
GPE: Countries, cities, states.
LOC: Non-GPE locations, mountain ranges, bodies of water.
PRODUCT: Objects, vehicles, foods, etc. (Not services.)
EVENT: Named hurricanes, battles, wars, sports events, etc.
WORK_OF_ART: Titles of books, songs, etc.
LAW: Named documents made into laws.
LANGUAGE: Any named language.
DATE: Absolute or relative dates or periods.
TIME: Times smaller than a day.
PERCENT: Percentage, including ”%“.
MONEY: Monetary values, including unit.
QUANTITY: Measurements, as of weight or distance.
ORDINAL: “first”, “second”, etc.
CARDINAL: Numerals that do not fall under another type.

Spacy NER (https://spacy.io/)

# Entity Recognition – Example with Spacy NER

- Try to guess the recognized entities...

> The Polytechnic University of Turin (Italian: Politecnico di Torino) is a public university based in Turin, Italy.
>
> Established in 1859, it is Italy's oldest technical university.
>
> The university offers several courses in the fields of Engineering, Architecture and Industrial Design.

Spacy NER (https://spacy.io/)

# Entity Recognition – Example with Spacy NER

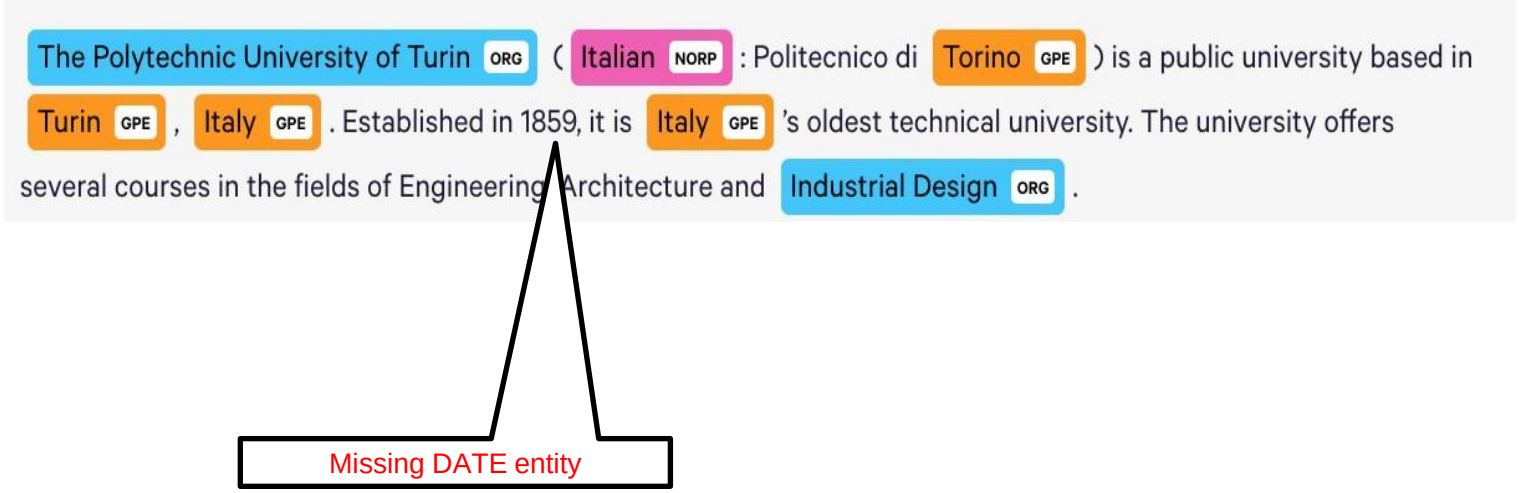
- Good but not great...

# Entity Recognition – Example with Spacy NER

- Good but not great...

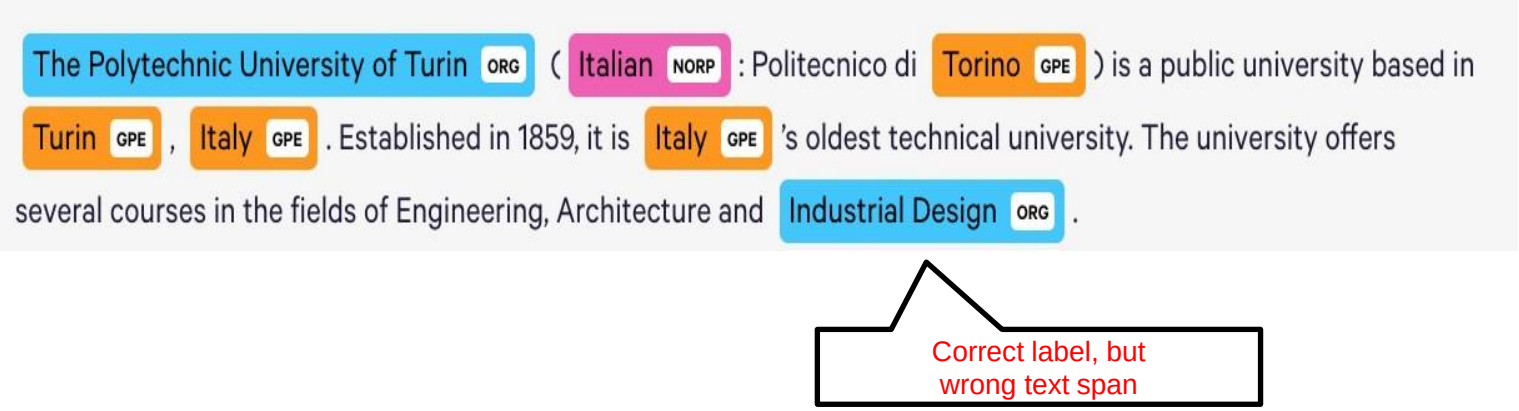
# Entity Recognition – Example with Spacy NER

- Good but not great...



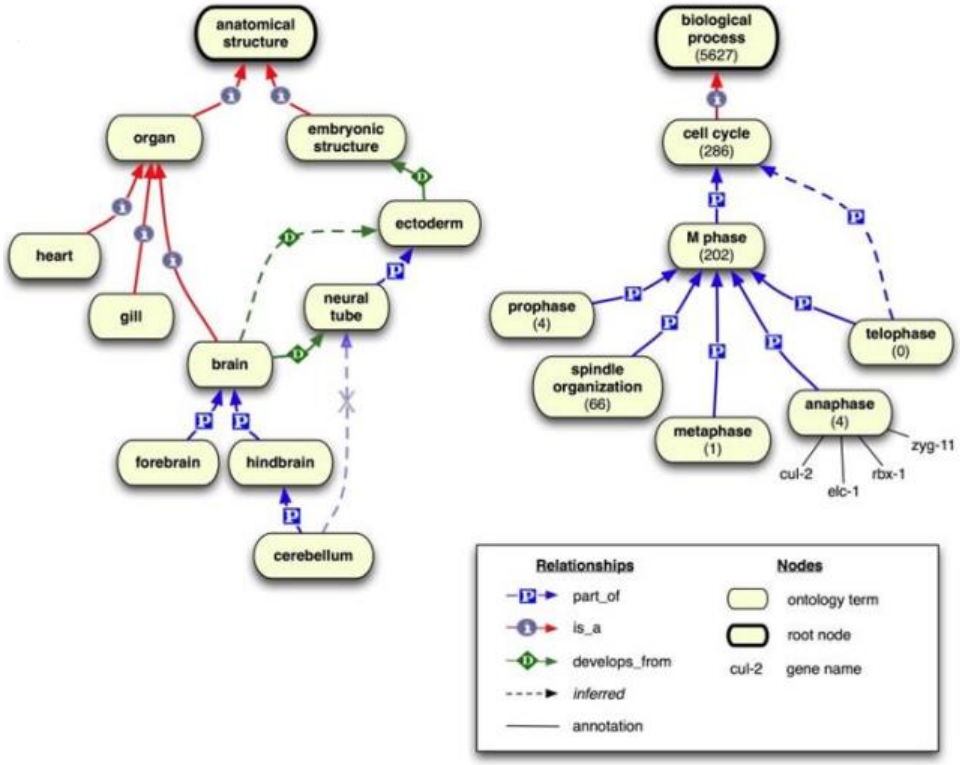
# Entity Recognition – Example with Spacy NER

- Good but not great...



# Semantic models

- Entities are usually stored in (domain-specific) semantic models called **ontologies**

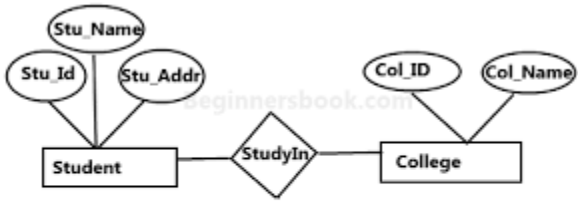


# Semantics in the ICT domain

- Common meaning in semantic technology helps computer systems more accurately interpret what people mean.
- Common meaning enables disparate IT systems – data sources and applications – to interface more efficiently and productively

# Semantic data models

- Semantic data models (SDMs, in short) are high-level semantics-based database descriptions and structuring formalism (database model) for databases
- They describes a database in terms of
  - the kinds of entities that exist in the application environment
  - the classifications and groupings of those entities,
  - the structural interconnections among them

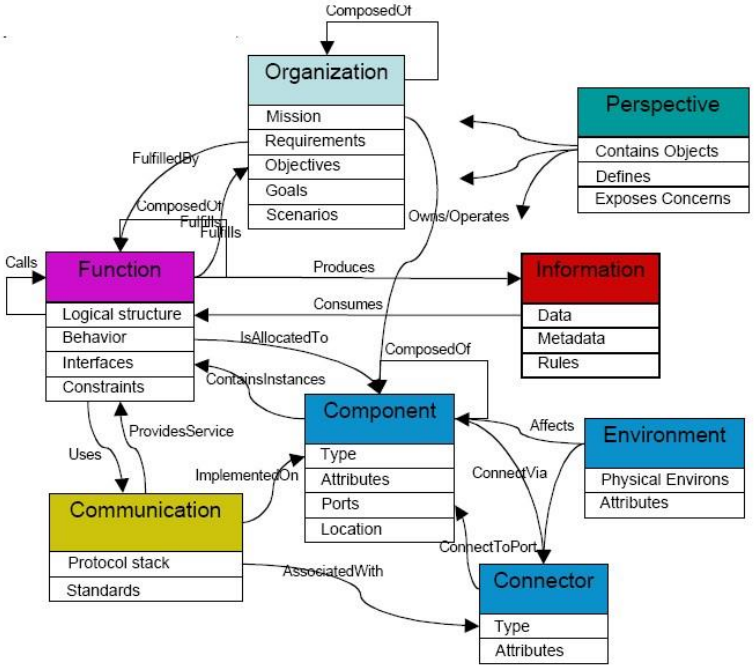



Sample E-R Diagram

Michael Hammer and Dennis McLeod (1978). "The Semantic Data Model: a Modeling Mechanism for Data Base Applications." In: *Proc. ACM SIGMOD Int'l. Conf. on Management of Data*. Austin, Texas, May 31 - June 2, 1978, pp. 26–36.

# Why do we need semantic models?

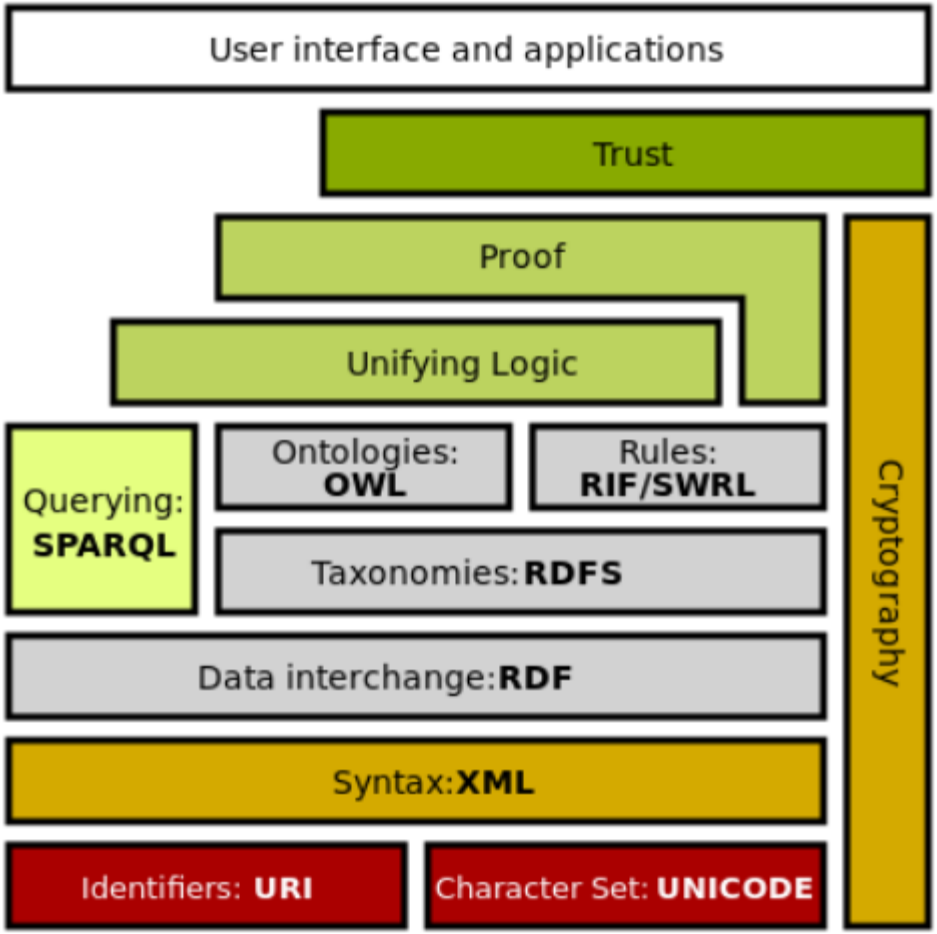
- Semantic models like vocabularies, taxonomies, and ontologies explicitly model semantic relationships
- They are the pillar of traditional NLP approaches
  - Commonly used before the advent of Deep NLP
  - Still used in Deep NLP for addressing particular tasks



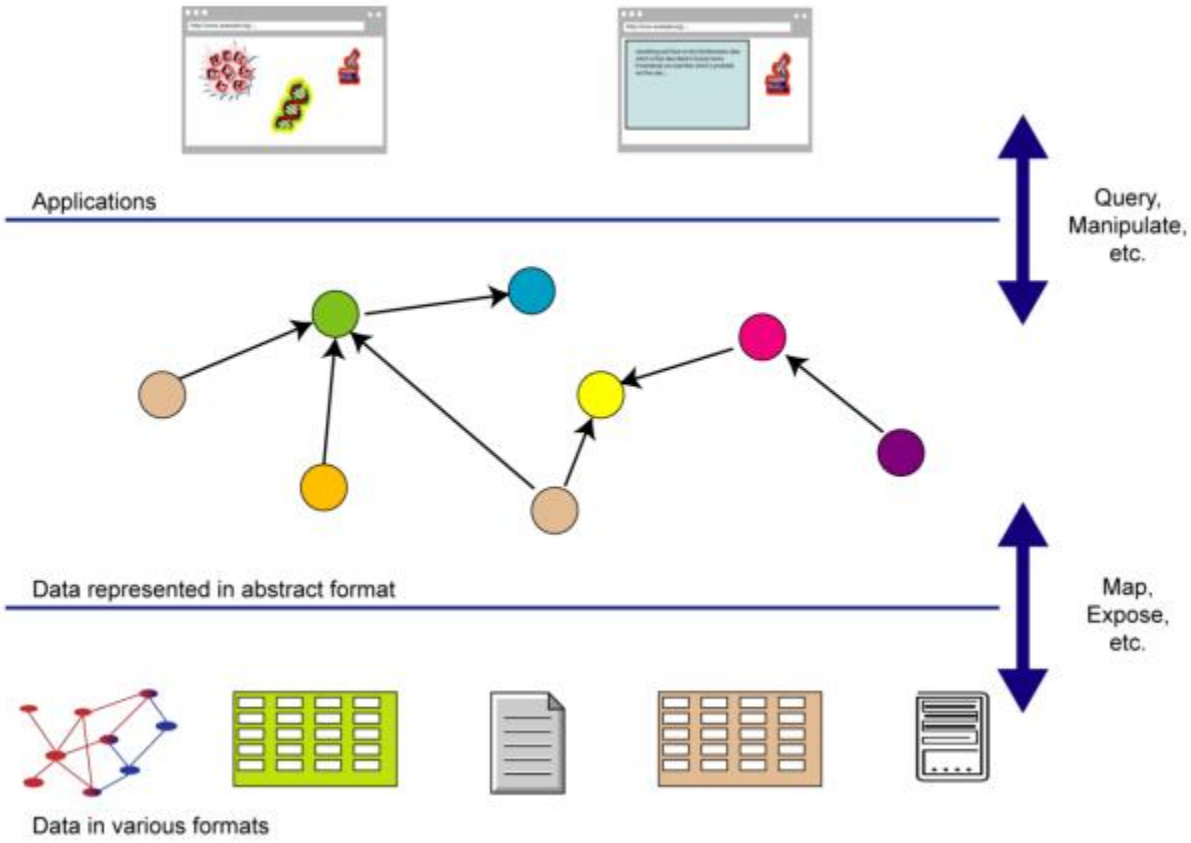
# The Semantic Web

- Extension of the World Wide Web through standards set by the World Wide Web Consortium (W3C)
- Stack of enabling technologies
- Goal: make Internet data machine-readable


# The Semantic Web



# The Semantic Web




01RRDIU - Semantic Web. Fulvio Corno, Laura Farinetti. https://elite.polito.it/ Latest access: July 2021

# Resource Description Framework (RDF)

- A Resource corresponds to a Uniform Resource Identifier
- A RDF is structured in statements
- A statement is a triple <Subject – Predicate – Object>
  - Subject: a resource
  - Predicate: a verb / property / relationship
  - Object: a resource, or a literal string
- RDF/XML
  - XML is a syntax
  - RDF is a data model

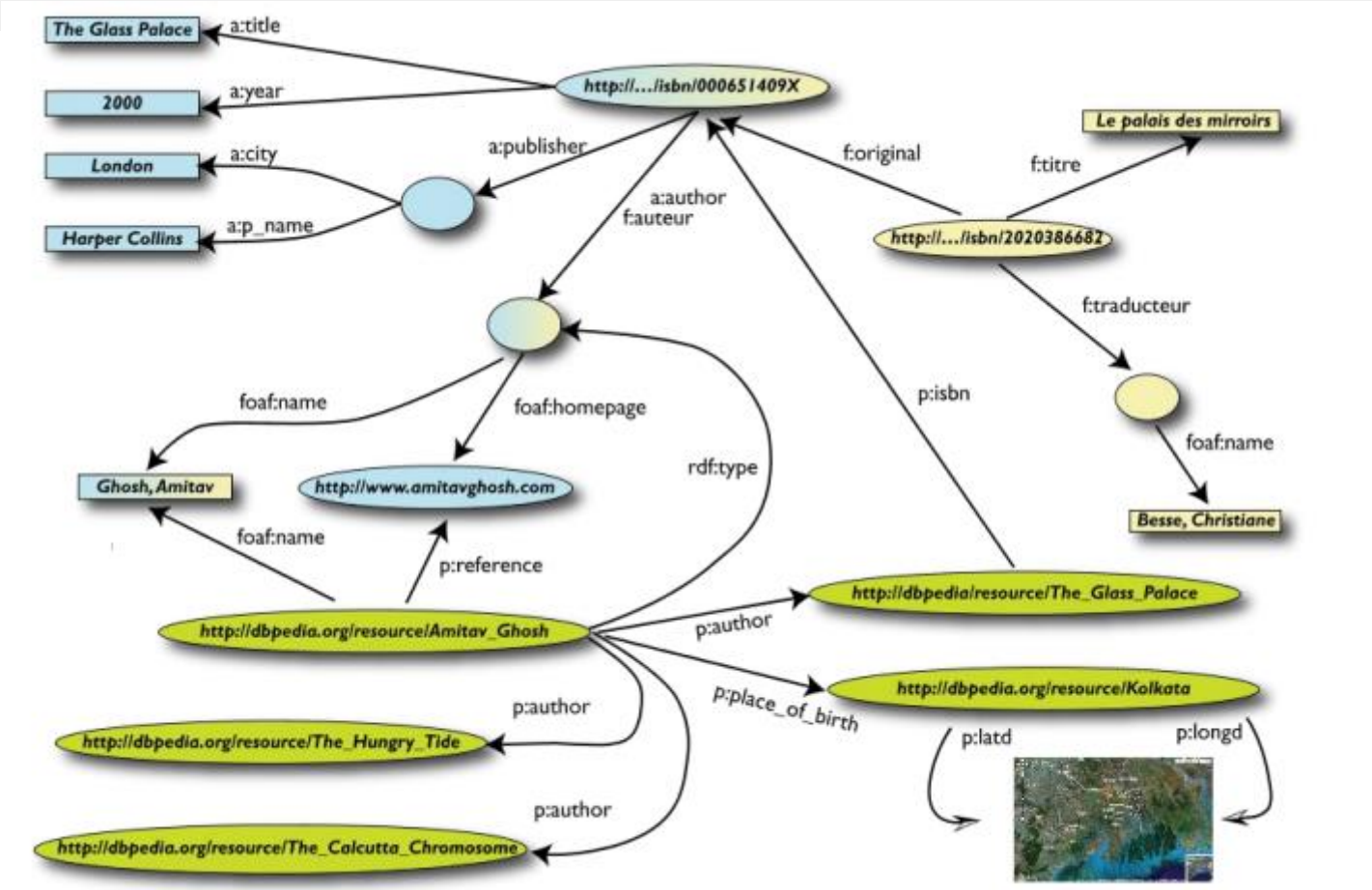
# Resource Description Framework

- RDF is a framework for describing resources on the web
  - E.g., purchased items, documents, pictures, videos
- It is designed to be read and understood by computers
- It is not designed for being displayed to people
- It is usually written in XML

```xml
<?xml version="1.0"?>

<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:si="https://www.w3schools.com/rdf/">

<rdf:Description rdf:about="https://www.w3schools.com">
  <si:title>W3Schools</si:title>
  <si:author>Jan Egil Refsnes</si:author>
</rdf:Description>

</rdf:RDF>
```

https://www.w3schools.com/ Latest access: July 2021

# RDF example


The Glass Palace
a:title
2000
a:year
London
a:city
Harper Collins
a:p_name
http://...isbn/000651409X
a:publisher
a:author
f:auteur
f:original
Le palais des mirroirs
f:titre
http://...isbn/2020386682
f:traducteur
foaf:name
foaf:homepage
Ghosh, Amitav
http://www.amitavghosh.com
foaf:name
p:reference
rdf:type
p:isbn
foaf:name
Besse, Christiane
http://dbpedia.org/resource/Amitav_Ghosh
http://dbpedia/resource/The_Glass_Palace
p:author
p:place_of_birth
http://dbpedia.org/resource/Kolkata
p:author
http://dbpedia.org/resource/The_Hungry_Tide
p:latd
p:longd
p:author
http://dbpedia.org/resource/The_Calcutta_Chromosome

# Ontologies

- Ontologies aim at describing the world without semantic ambiguity
- are used for automatic reasoning as their structure is machine-readable
- Key elements (not exhaustive list)
  - **Classes or concepts** are the nodes of the graph
    - E.g., "portrait" or "painter"
  - **Individuals** are instances of abstract concepts
    - E.g., "The Mona Lisa" or "Leonardo Da Vinci"
  - **Relations** are edges of the graph. They are used to link concepts
  - **Links to descriptors**
    - E.g., dcterms:title links the object with its title

# Ontology components (detailed list)

- Individuals
  - instances or objects (the basic or "ground level" objects)
- Classes
  - sets, collections, concepts, types of objects, or kinds of things
- Attributes
  - aspects, properties, features, characteristics, or parameters that objects (and classes) can have
- Relations
  - ways in which classes and individuals can be related to one another
- Events
  - the changing of attributes or relations

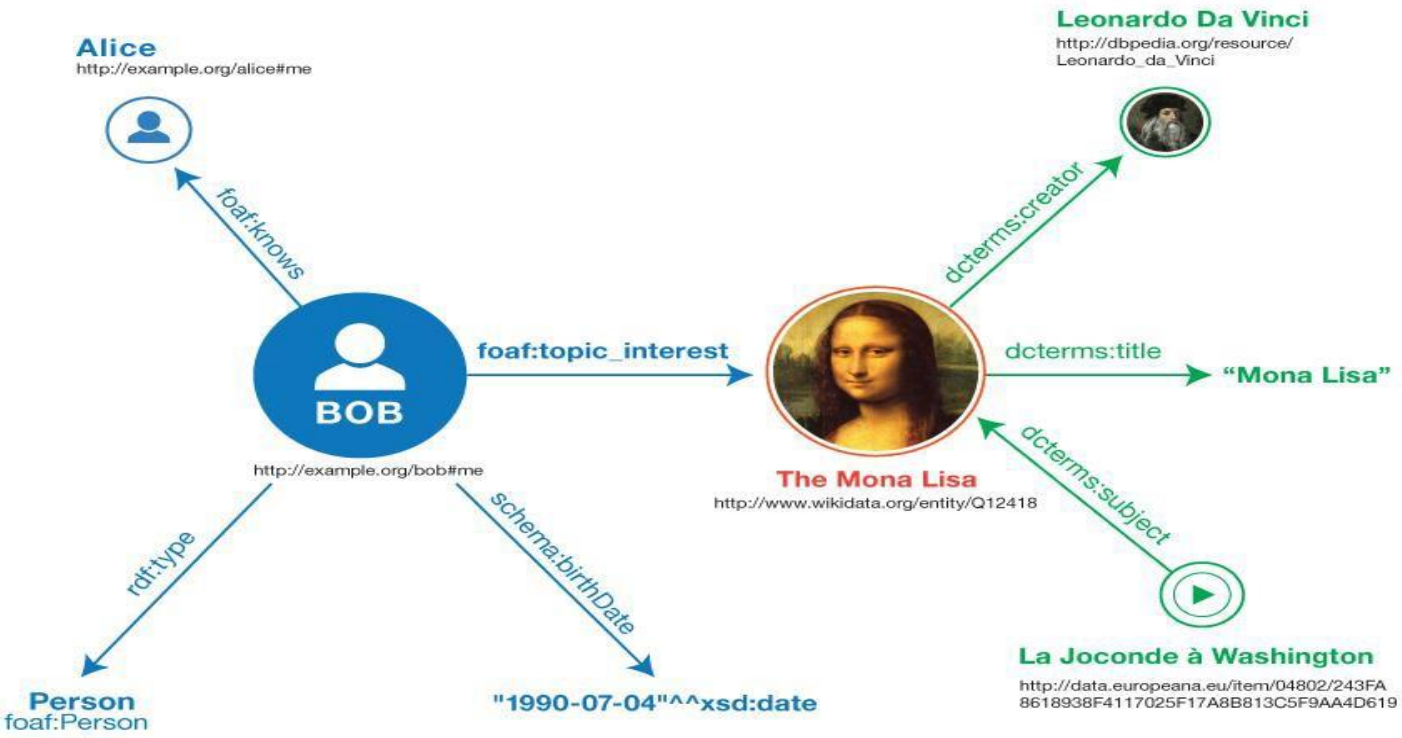
Maureen Donnelly; Giancarlo Guizzardi (2012). Formal Ontology in Information Systems: Proceedings of the Seventh International Conference (FOIS 2012). IOS Press. ISBN 978-1-61499-083-3.

# Ontology components (detailed list)

- Function terms
  - complex structures formed from certain relations that can be used in place of an individual term in a statement
- Restrictions
  - formally stated descriptions of what must be true in order for some assertion to be accepted as input.
- Rules
  - statements in the form of an if-then (antecedent-consequent) sentence that describe the logical inferences that can be drawn from an assertion in a particular form.
- Axioms
  - assertions (including rules) in a logical form that together comprise the overall theory that the ontology describes in its domain of application.
- Events
  - the changing of attributes or relations

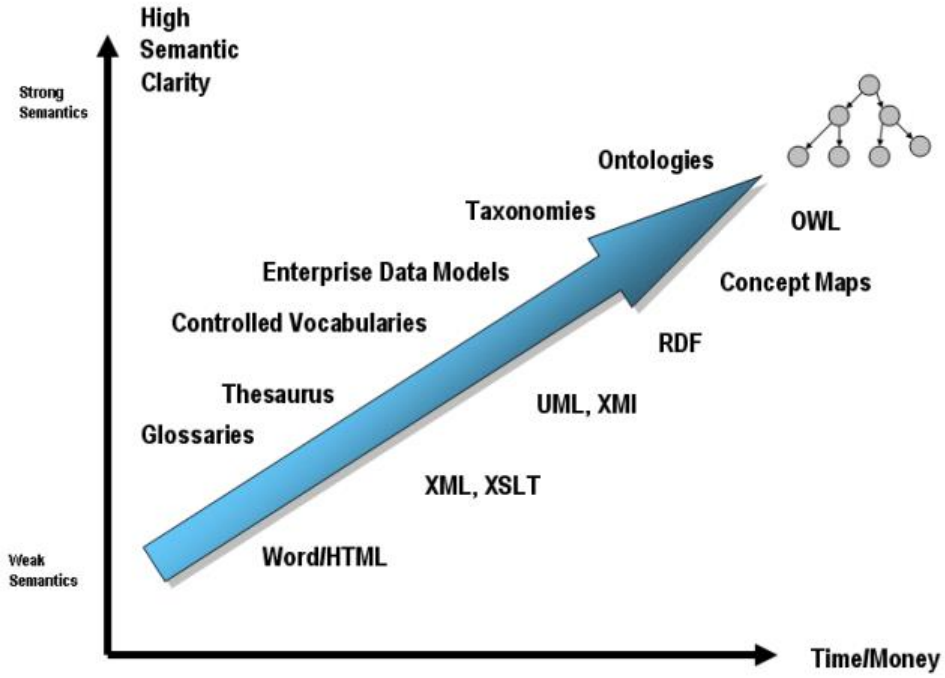
Maureen Donnelly; Giancarlo Guizzardi (2012). Formal Ontology in Information Systems: Proceedings of the Seventh International Conference (FOIS 2012). IOS Press. ISBN 978-1-61499-083-3.

# Ontologies


Alice
http://example.org/alice#me
foaf:knows
BOB
http://example.org/bob#me
foaf:topic_interest
rdf:type
Person
foaf:Person
schema:birthDate
"1990-07-04"^^xsd:date
Leonardo Da Vinci
http://dbpedia.org/resource/Leonardo_da_Vinci
dcterms:creator
The Mona Lisa
http://www.wikidata.org/entity/Q12418
dcterms:title
"Mona Lisa"
dcterms:subject
La Joconde à Washington
http://data.europeana.eu/item/04802/243FA8618938F4117025F17A8B813C5F9AA4D619

# What if ontologies are not available?

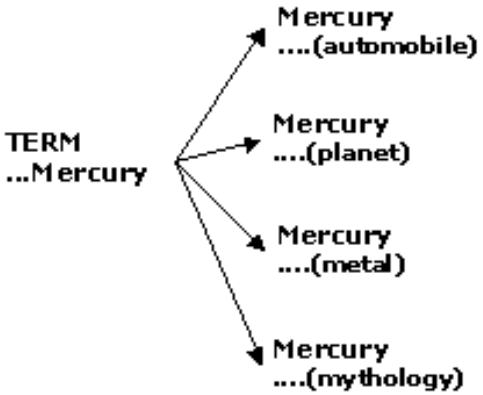
- Simpler semantic models are used




01RRDIU - Semantic Web. Fulvio Corno, Laura Farinetti. https://elite.polito.it/ Latest access: July 2021

# Controlled vocabulary

- List of terms of phrases
  - E.g., indices, subject headings
- Common usage
  - Describe semantically relevant concepts and relationships
  - Prevent linguistic errors and ambiguities in natural language
  - Gather linguistic relations

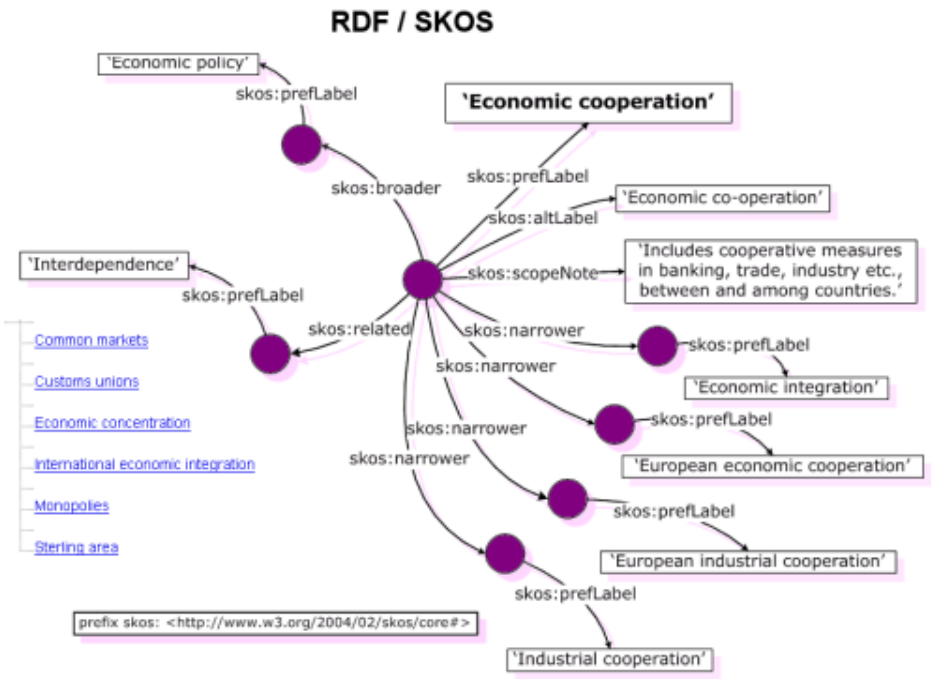


# Taxonomy

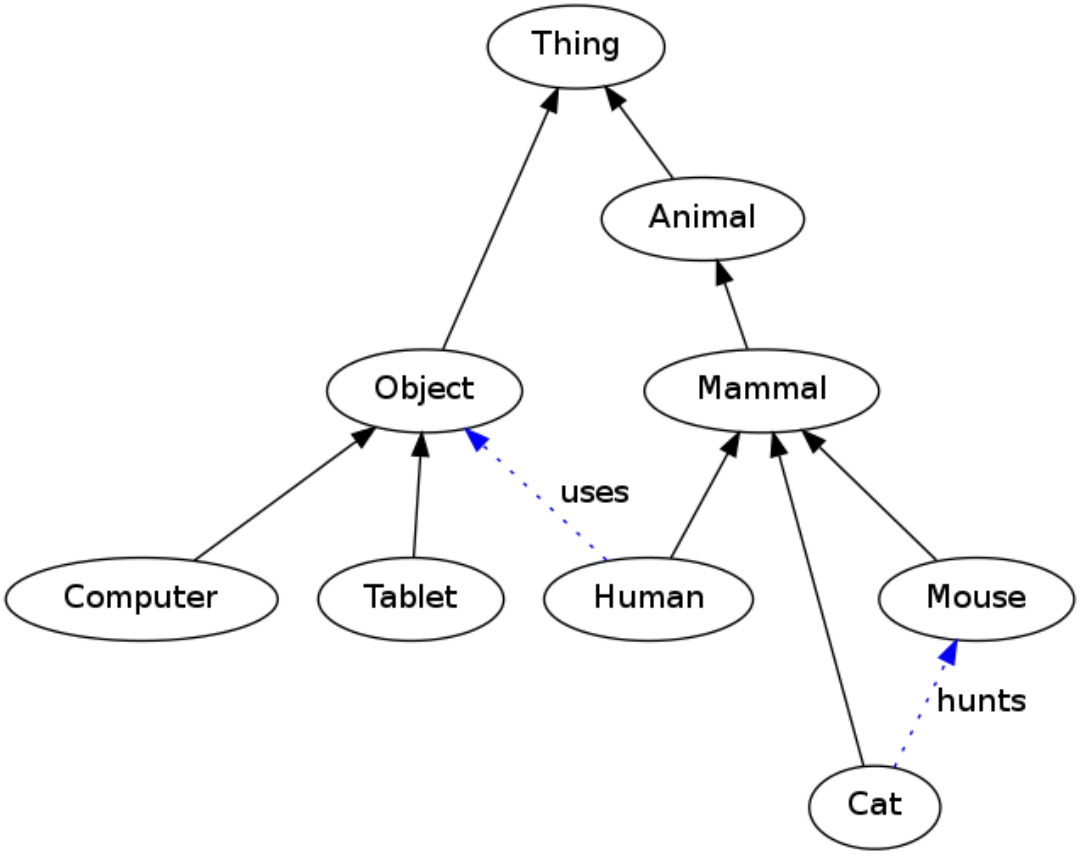
- Hierarchical representation of concepts within a controlled vocabulary
- Specific for a given subject
- Relationships
  - Parent = broader term
  - Child = narrower term

# Thesauri

- Extended taxonomy models
- Specific for a given subject
- Relationships
  - Parent = broader term
  - Child = narrower term



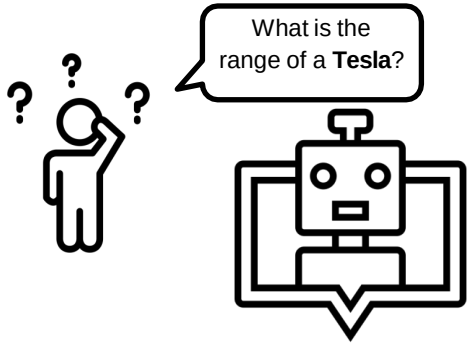
# Ontology example



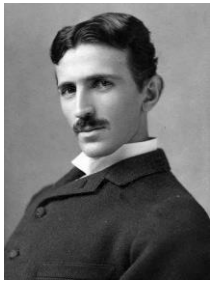

arispe, S., Ranwez, S., Janaqi, S., & Montmain, J. (2017). Semantic Measures for the Comparison of Units of Language, Concepts or Entities from Text and Knowledge Base Analysis.

# Named Entity Disambiguation

- NED entails connecting an entity mention to a concept in a given ontology
- NER only detects the most likely class
  - it does not link to specific concepts



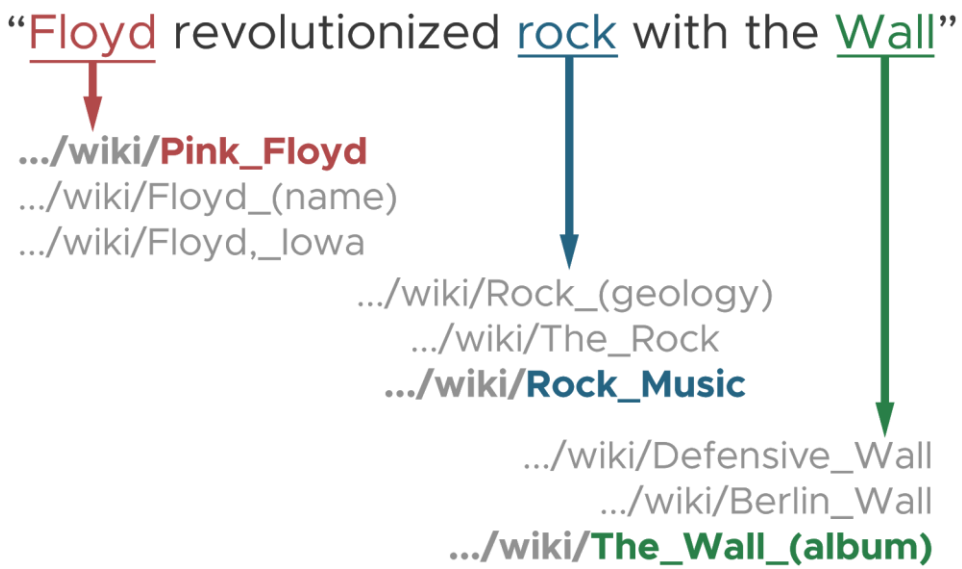

Named Entity Disambiguation

Icons credit: Creative Mania & Rahmat Dwi Cahyo from the Noun Project

# Named Entity Disambiguation

- It is necessary to understand the context of use




https://blogs.oracle.com/ai-and-datascience/post/named-entity-disambiguation-with-knowledge-graphs (latest access: June 2021)

# Approaches to NER

- Knowledge engineering
  - Rule-based NER
  - Ontology-based NER
- Machine Learning-based NER
  - Standard classifiers
    - E.g., NNs, kNN, SVMs, Bayesian classifiers
  - Sequence models
    - RNNs, LSTMs, Transformers

# Approaches to NER

- Knowledge engineering
  - Rule-based NER
  - Ontology-based NER
- Machine Learning-based NER
  - Standard classifiers
    - E.g., NNs, kNN, SVMs, Bayesian classifiers
  - Sequence models
    - RNNs, LSTMs, Transformers

State-of-the-art, but…

# Approaches to NER: comparison

- Knowledge engineering
  - High precision
  - Manually generated rules
  - Requires a small amount of training data
  - Strong domain dependence
  - Expensive development, testing, and updating
- Machine Learning-based NER
  - High recall
  - Requires lots of training data
  - No hand-crafted rules
  - Not dependent on semantic models

# Rule-based NER

- It relies on regular expressions to identify and classify named entities
- Usually tailored to a specific domain
- E.g., identify drug names in the text
- Improve the generalization power if compared with resource-based systems
- They fail in most open-domain classes

# Rule-based NER: example

- Examples of rules suitable for detecting and classifying people:
  - A sequence of words starting with a capital letter, followed by lowercase letters
    - E.g., Barack Obama
  - May contain a prefix title (Dr., Mr., Prof.)
    - E.g., Prof. Enrico Fermi
  - May contain an initial in the middle
    - E.g., George W. Bush
  - May contain a designation indicator prefix
    - E.g., Elon Musk CEO
  - Never include special characters: %/$/=...

Xiang Rneg. ML for Knowledge Extraction and Reasoning. USC Computer Science (CSCI699).

# Rule-based NER: example

- Examples of rules suitable for detecting and classifying people:
  - A sequence of words starting with a capital letter, followed by lowercase letters
    - E.g., Barack Obama
  - May contain a prefix title (Dr., Mr., Prof.)
    - E.g., Prof. Enrico Fermi
  - May contain an initial in the middle
    - E.g., George W. Bush
  - May contain a designation indicator prefix
    - E.g., Elon Musk CEO
  - Never include special characters: %/$/=...

New proper names constantly emerge!

Xiang Rneg. ML for Knowledge Extraction and Reasoning. USC Computer Science (CSCI699).

# Rule-based NER

- Example of regular expression: "^P[a-z][aeiou]"
  - starts with a capital letter "P"
  - It is the first word on a line
  - The second letter is a lower case letter
  - It is exactly three letters long
  - The third letter is a vowel
  - The regular expression would be where

^ indicates the beginning of the string

[a-z] any letter in range a to z

[aeiou] – any vowel

Xiang Rneg. ML for Knowledge Extraction and Reasoning. USC Computer Science (CSCI699).

# Rule-based NER

- Regular expression list
  - \w (word char) any alpha-numeric
  - \d (digit char) any digit
  - \s (space char) any whitespace
  - . (wildcard) anything
  - \b word boundary
  - ^ beginning of string
  - $ end of string
  - ? For 0 or 1 occurrences
  - + for 1 or more occurrences
  - specific range of number of occurrences: {min,max}
    - A{1,5} One to five A's.
    - A{5,} Five or more A's
    - A{5} Exactly five A's

Xiang Rneg. ML for Knowledge Extraction and Reasoning. USC Computer Science (CSCI699).

# Rule-based NER: example

- Regular expression list for a telephone number
  - blocks of digits separated by hyphens

$$\textbf{RegEx} = \textbf{(\textbackslash d+\textbackslash -)+\textbackslash d+}$$

  - matches valid phone numbers like 011-909-7099 and 011-011
  - incorrectly extracts social security numbers 123-45-6789
  - fails to identify numbers like 800.865.1125 and (800)865-CARE

$$\textbf{Improved RegEx} = \textbf{(\textbackslash d\{3\}[-.\textbackslash ()])\{1,2\}[\textbackslash dA-Z]\{4\}}$$

Xiang Rneg. ML for Knowledge Extraction and Reasoning. USC Computer Science (CSCI699).

# Rule-based NER: example

- Regular expression to extract locations
  - Capitalized word + {city, center, river} indicates location
    - E.g., New York city, Central Park
  - Capitalized word + {street, boulevard, avenue} indicates location
    - E.g., Fifth avenue

Xiang Rneg. ML for Knowledge Extraction and Reasoning. USC Computer Science (CSCI699).

# Rule-based NER: example

- Capitalization is useful but...
  - First word of a sentence is capitalized by default
  - Titles are often all capitalized
  - Some languages have peculiar capitalization rules
    - E.g., all nouns in German all capitalized
  - Nested entity names contain non-capital words
    - E.g., University of West Virginia

Xiang Rneg. ML for Knowledge Extraction and Reasoning. USC Computer Science (CSCI699).

# Rule-based NER: challenges

- The same entity can have multiple variants of the same name
  - E.g., Luca Cagliero, Prof. Cagliero, Professor L. Cagliero
- Proper names can be ambiguous
  - E.g., Felice Buonanno, Santa Pazienza, Guido Di Rado, Immacolata Sforza

Xiang Rneg. ML for Knowledge Extraction and Reasoning. USC Computer Science (CSCI699).

# Rule-based NER: challenges

- The same entity can have multiple variants of the same name
  - E.g., Luca Cagliero, Prof. Cagliero, Professor L. Cagliero
- Proper names can be ambiguous
  - E.g., Felice Buonanno, Santa Pazienza, Guido Di Rado, Immacolata Sforza

**Need for ad hoc resolution steps!**

Xiang Rneg. ML for Knowledge Extraction and Reasoning. USC Computer Science (CSCI699).

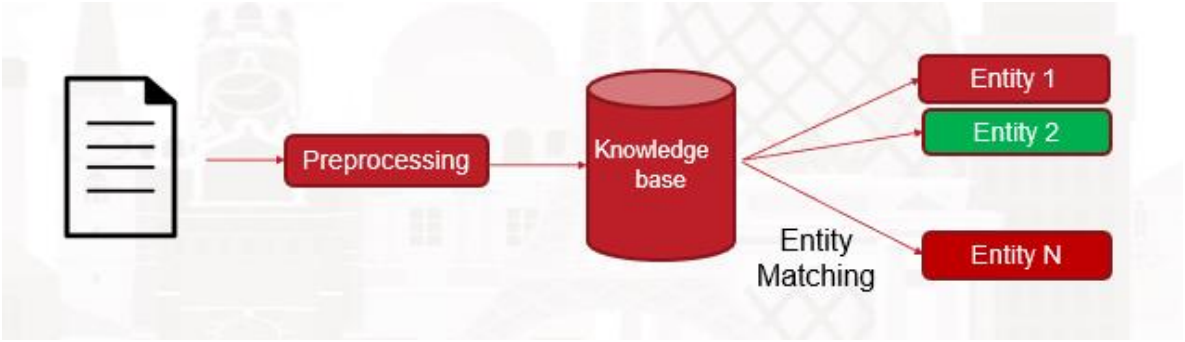
# Enhanced Rule-based NER

- Combine regular expressions with basic NLP preprocessing steps
  - Part-of-speech (POS) tagging
    - Mark each word as a noun, verb, preposition, etc.
  - Syntactic parsing
    - Identify phrases: NP, VP, PP
  - Text-to-audio
    - Identify phonemes
  - Assign semantic word categories
    - E.g., from the WordNet lexical database (https://wordnet.princeton.edu/)
      - KILL: kill, murder, assassinate, strangle, suffocate

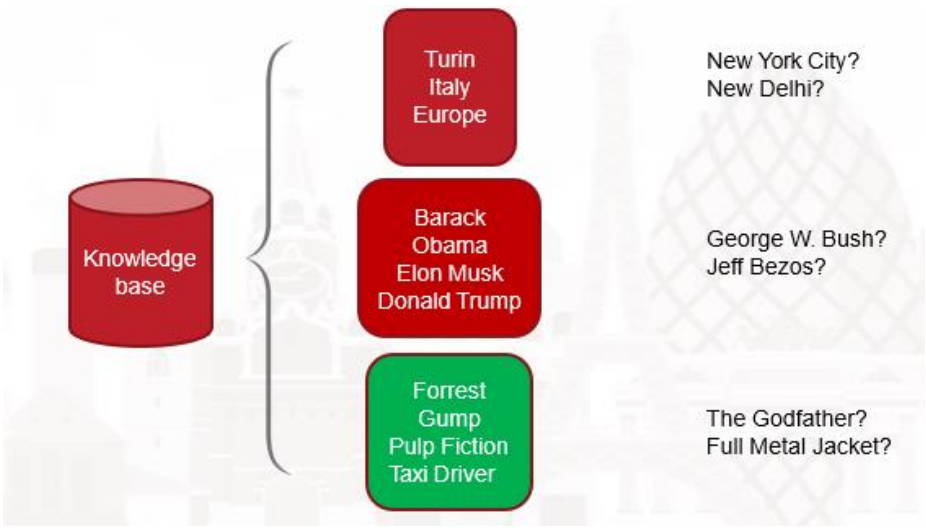
Xiang Rneg. ML for Knowledge Extraction and Reasoning. USC Computer Science (CSCI699).

# Ontology-based NER

- Use existing domain-specific ontologies as resources for discovering specific named entities in text
  - These approaches work well when the ontological resources are exhaustive for the task
  - They fail when some entities are missing

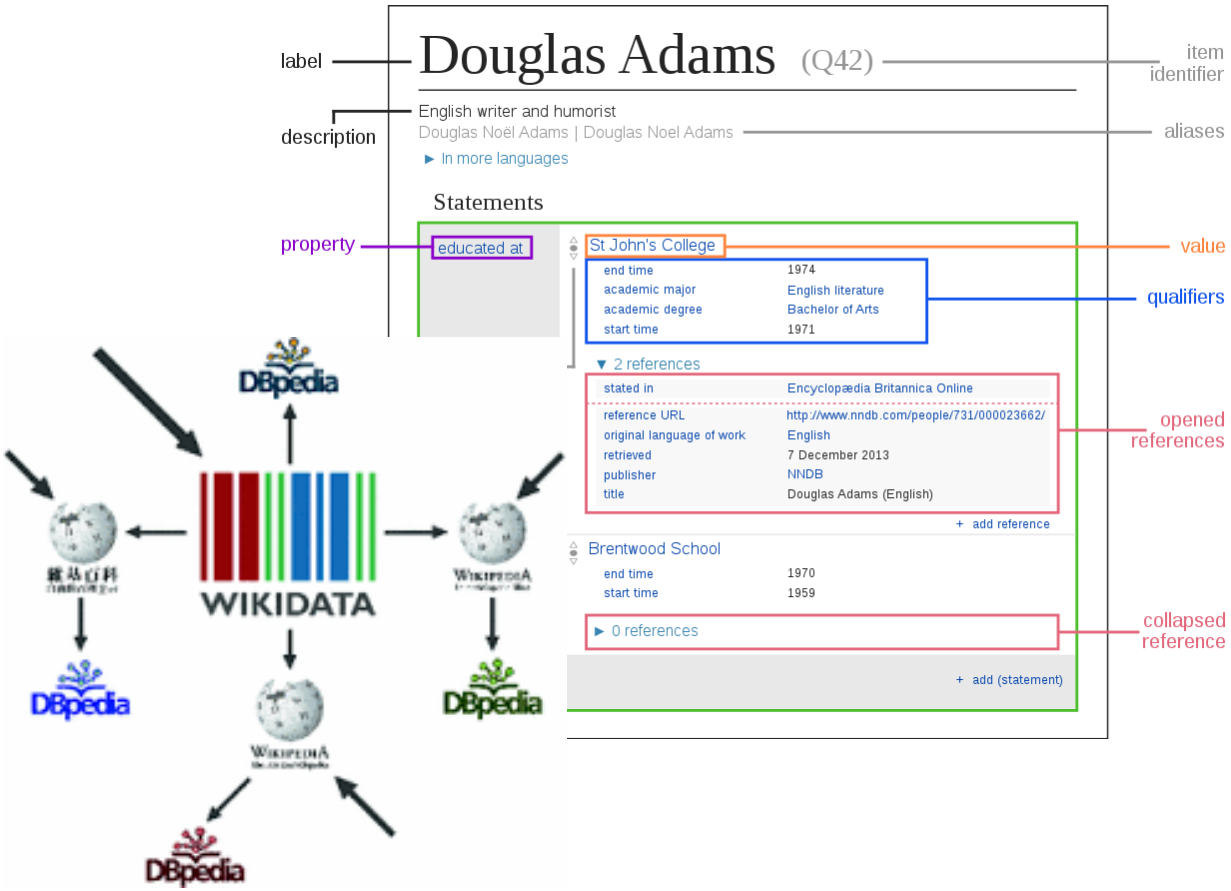


# Ontology-based NER: example



# Examples of ontologies

- Yago
- Dbpedia
- Wikidata
- ...



# Ontology-based NER

- Yago
  - Semantic knowledge base, derived from Wikipedia WordNet and GeoNames
  - More than 10 million entities
    - persons, organizations, cities
  - more than 120 million facts about these entities

Yago: a core of semantic knowledge. FM Suchanek, G Kasneci, G Weikum.
Proceedings of the 16th Proceedings of the 16th international conference on World Wide Web. Pages 697-706 , 2007. ACM.

# Ontology-based NER

- NERD
  - set of mappings established manually between the taxonomies of named entity types
  - NERD core
    - Thing
    - Amount
    - Animal
    - Event
    - Function
    - Location
    - Organization
    - Person
    - Product
    - Time

Giuseppe Rizzo and Raphaël Troncy. NERD: Evaluating Named Entity Recognition Tools in the Web of Data. ISWC'11 Workshop on Web Scale Knowledge Extraction (WEKEX'11), Bonn, Germany, October 23-27, 2011

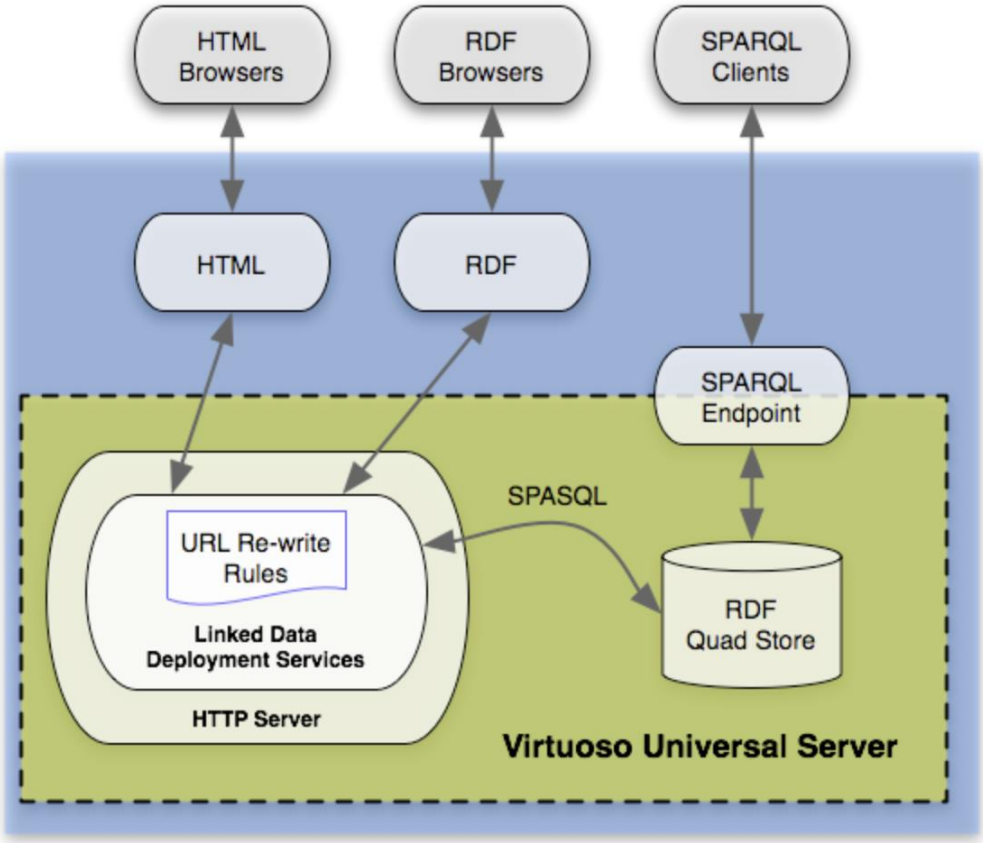
# Ontology-based NER

- DBPedia
  - Crowdsource community effort to extract structure information from Wikipedia
  - NERD core
    - 4.58 million things, out of which 4.22 million are classified in a consistent ontology, including
      - 1,445,000 persons
      - 735,000 places (including 478,000 populated places)
      - 411,000 creative works (including 123,000 music albums, 87,000 films)
      - 241,000 organizations (including 58,000 companies and 49,000 educational institutions)
      - 251,000 species
      - 6,000 diseases

Dimitris Kontokostas, Charalampos Bratsas, Sören Auer, Sebastian Hellmann, Ioannis Antoniou, George Metakides, Internationalization of Linked Data: The case of the Greek DBpedia edition, Web Semantics: Science, Services and Agents on the World Wide Web, Volume 15, September 2012, Pages 51–61, ISSN 1570–8268, 10.1016/j.websem.2012.01.001.

# Ontology-based NER

- DBPedia




Dimitris Kontokostas, Charalampos Bratsas, Sören Auer, Sebastian Hellmann, Ioannis Antoniou, George Metakides, Internationalization of Linked Data: The case of the Greek DBpedia edition, Web Semantics: Science, Services and Agents on the World Wide Web, Volume 15, September 2012, Pages 51–61, ISSN 1570–8268, 10.1016/j.websem.2012.01.001.

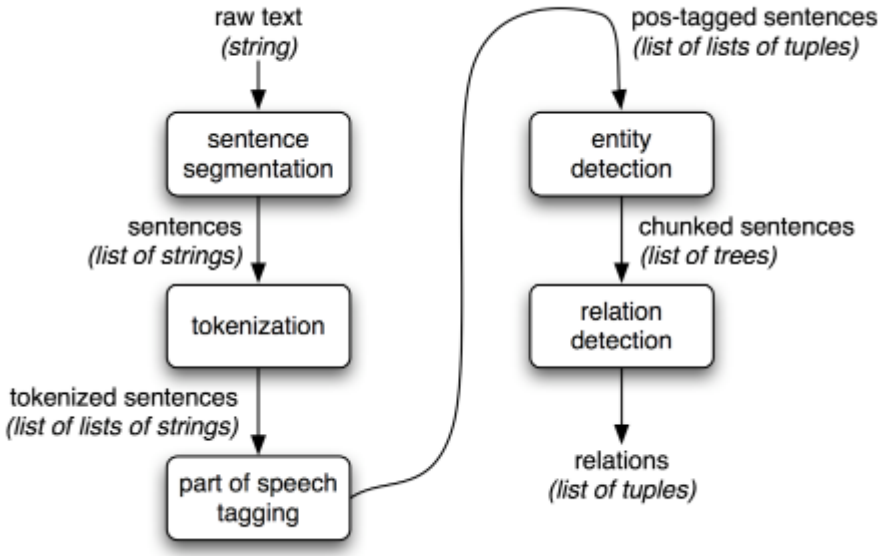
# Ontology-based NER

- Natural Language Toolkit (NLTK)
  - leading platform for building Python programs to work with human language data
  - easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet
  - a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLP libraries, and an active discussion forum.

NTLK website: http://www.nltk.org/

# Ontology-based NER

- Natural Language Toolkit (NLTK)




http://www.nltk.org/book/ch07.html (latest access: June 2021)

# Ontology-based NER

- NER using NLTK
  - It supports the automatic recognition of named entities
    - E.g., Organizations, persons, ...
  - The extraction of relations between them
    - E.g., looking for relations between specified types of named entity (person with the corresponding organization)


http://www.nltk.org/book/ch07.html (latest access: June 2021)

# Ontology-based NER

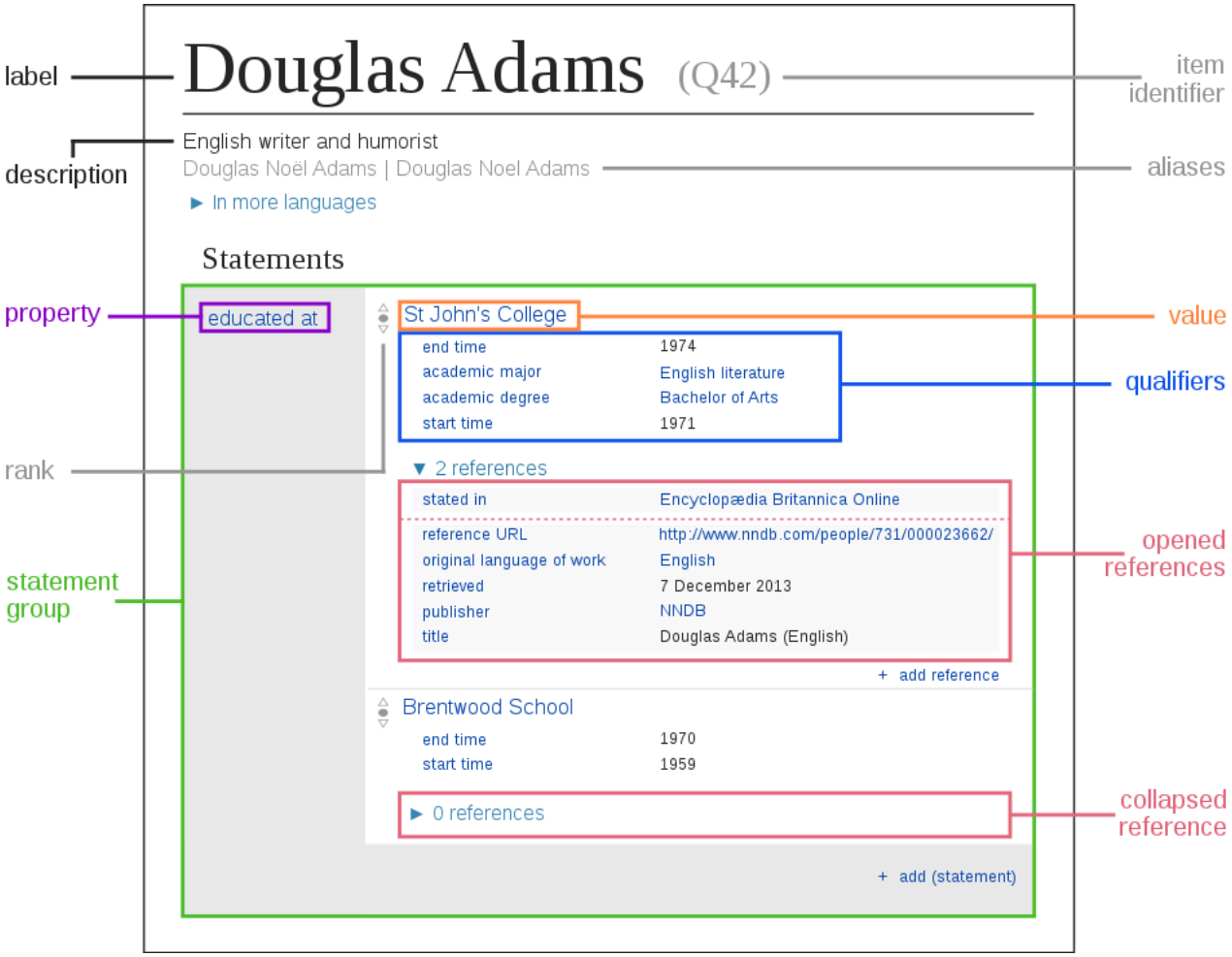
- Wikidata
  - free, open, multilingual knowledge base
  - It provides end-users with access to Wikipedia
  - Wikidata content can be read and edited by both humans and machines
  - Central storage repository
    - Consists of items, each one having an identifier (Q), a label, a description, any number of aliases
  - To access Wikidata
    - Wikidata quert and Reasonator tools
    - Lua Scribunto interface
    - **Wikidata API**

https://wikidata.org (latest access: June 2021)

# Ontology-based NER

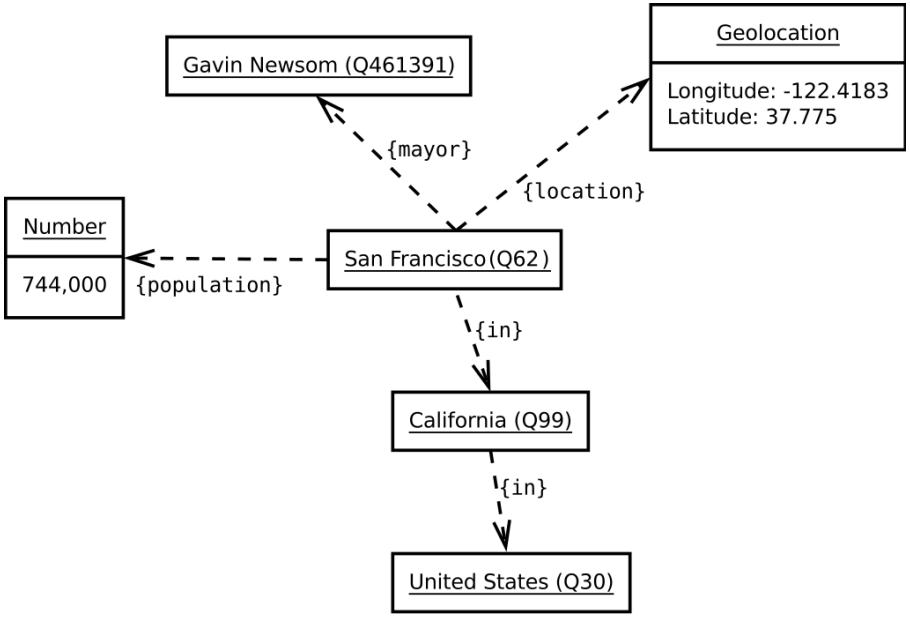
- Wikidata



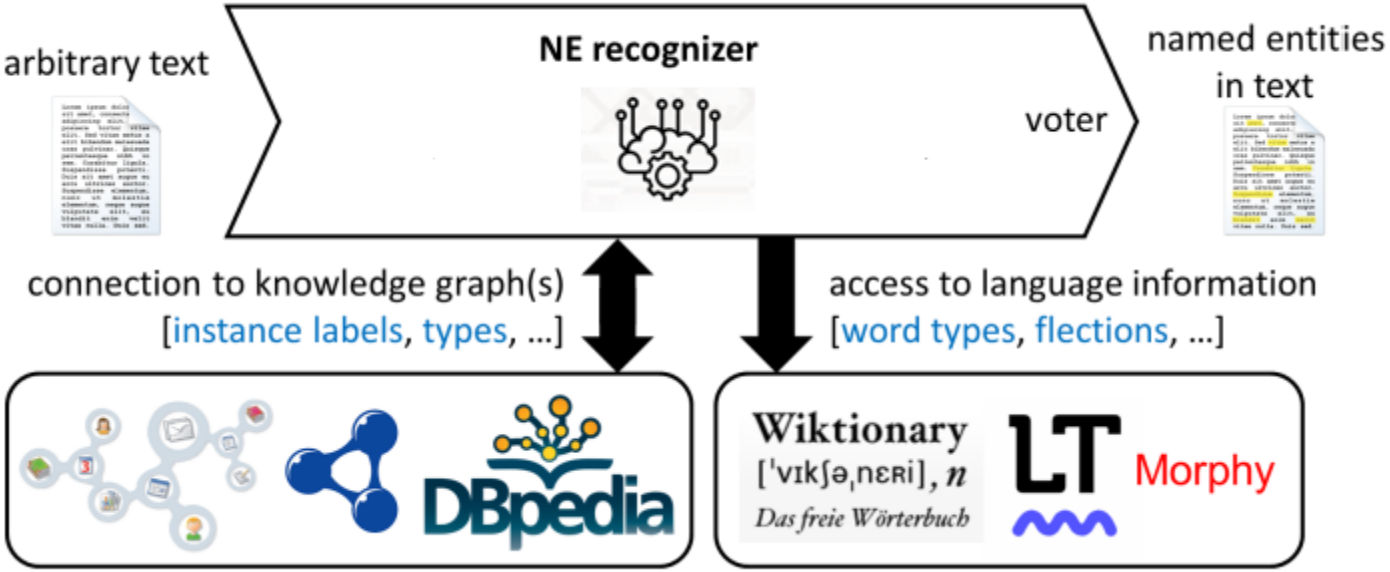

https://wikidata.org (latest access: June 2021)

# Ontology-based NER

- Wikidata
  - Consists of items, each one having an identifier (Q*xy*), a label, a description, any number of aliases
  - A statement describes a detailed characteristic of an item
    - It consists of a property and a value
    - A property can link an item to an external database



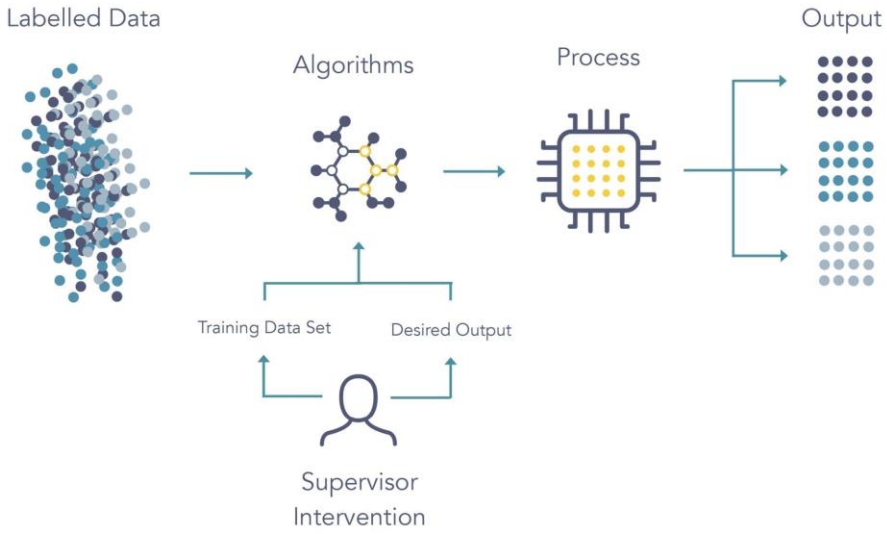
# Ontology-based NER example




Christian Jilek, Markus Schröder, Rudolf Novik, Sven Schwarz, Heiko Maus, Andreas Dengel. Inflection-Tolerant Ontology-Based Named Entity Recognition for Real-Time Applications. LDK 2019: 11:1-11:14

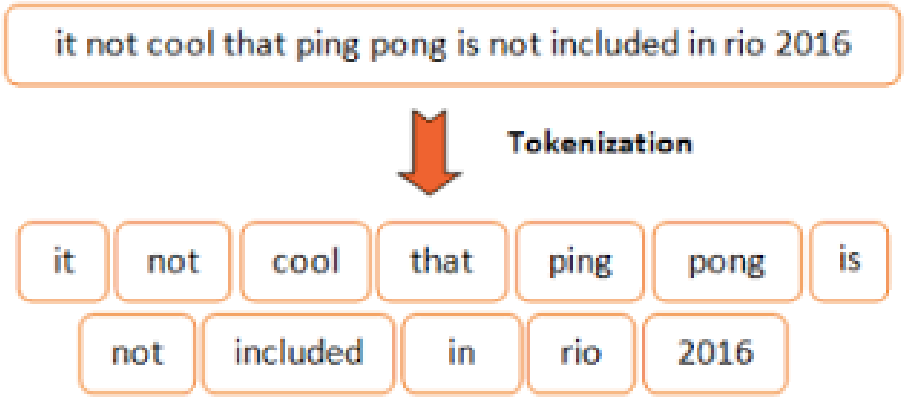
# Machine Learning-based NER

- Apply a Machine Learning pipeline to leverage human-generated annotations
  - Learn a predictive model from annotated data
  - Apply to model to unlabeled data



# Machine Learning-based NER

1) Tokenization



# Machine Learning-based NER

## 2) Human annotation

| Intesa |
|---|
| Sanpaolo |
| S.p.A. |
| Or |
| simply |
| ISP |
| Group |
| … |

Human Annotation

| Intesa | B-ORG |
|---|---|
| Sanpaolo | I-ORG |
| S.p.A. | I-ORG |
| or | O |
| simply | O |
| ISP | B-ORG |
| Group | I-ORG |
| … | … |

# Machine Learning-based NER

- Named Entity Detection
  - B-prefix indicates the beginning of a named entity
  - I-prefix refers to a word inside the named entity
  - O-prefix refers to a word outside the named entity
- Named Entity Classification
  - Person names
  - Organizations
    - E.g., companies, governmental organizations
  - Locations
  - Miscellaneous
    - E.g., sport events

# Machine Learning-based NER

## 3) Feature engineering

| |
|---|
| Intesa |
| Sanpaolo |
| S.p.A. |
| Or |
| simply |
| ISP |
| Group |
| … |

Feature extraction

| Word | POS | Symb. Count |
|---|---|---|
| Intesa | NOUN | 0 |
| Sanpaolo | NOUN | 0 |
| S.p.A. | NOUN | 2 |
| Or | CONJ | 0 |
| simply | ADV | 0 |
| ISP | NOUN | 0 |
| Group | NOUN | 0 |
| … | … | … |

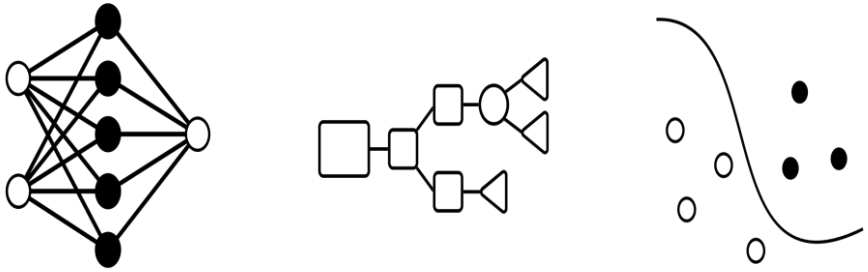
# Machine Learning-based NER

Training dataset

| Word | POS | Symb. Count | Class |
|---|---|---|---|
| Assicurazioni | NOUN | 0 | B-ORG |
| Generali | NOUN | 0 | I-ORG |
| S.p.A. | NOUN | 2 | I-ORG |
| or | CONJ | 0 | O |
| simply | ADV | 0 | O |
| Generali | NOUN | 0 | B-ORG |
| Group | NOUN | 0 | I-ORG |

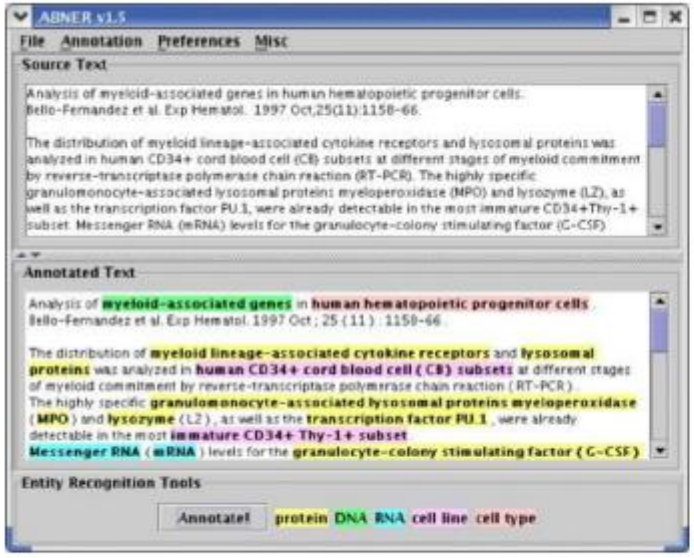
# Machine Learning-based NER

4) Classifier training

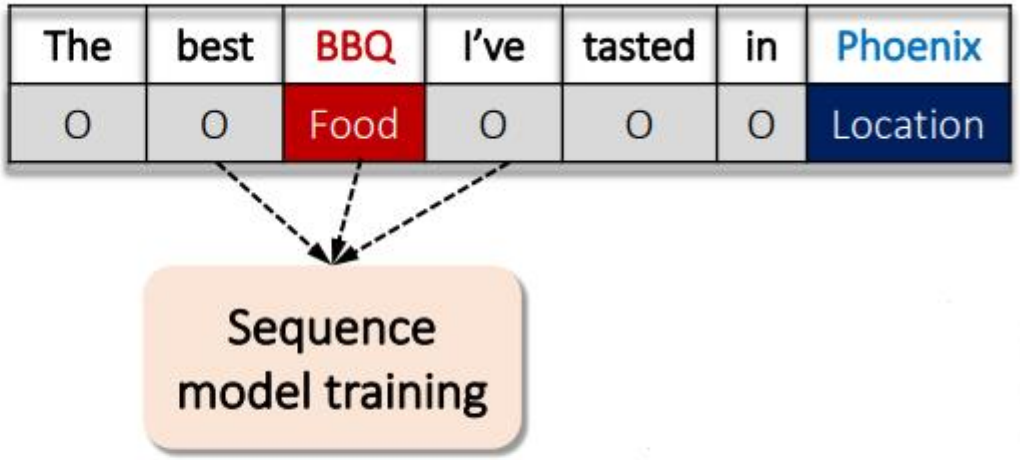


# Sequence labeling for NER

- Classify using sequential models
  - E.g., RNNs, LSTMs



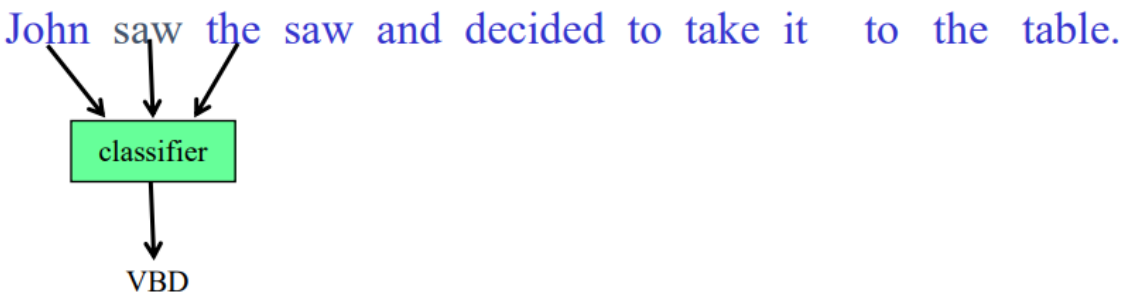

A manual annotation interface




Xiang Rneg. ML for Knowledge Extraction and Reasoning. USC Computer Science (CSCI699).

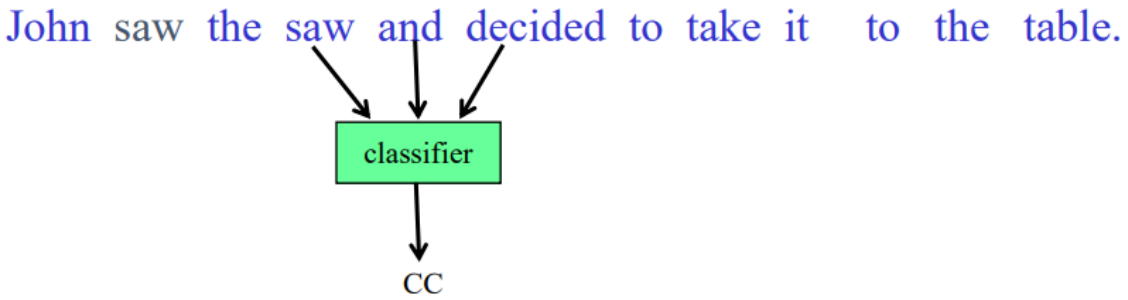
# Sequence labeling for NER

- Classify each token independently but...
  - Use as input features the information about the surrounding tokens
  - Sliding window approach



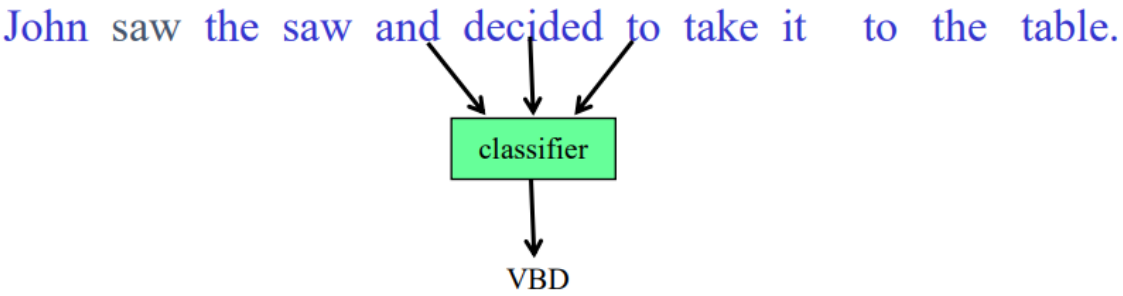
# Sequence labeling for NER

- Classify each token independently but...
  - Use as input features the information about the surrounding tokens
  - Sliding window approach



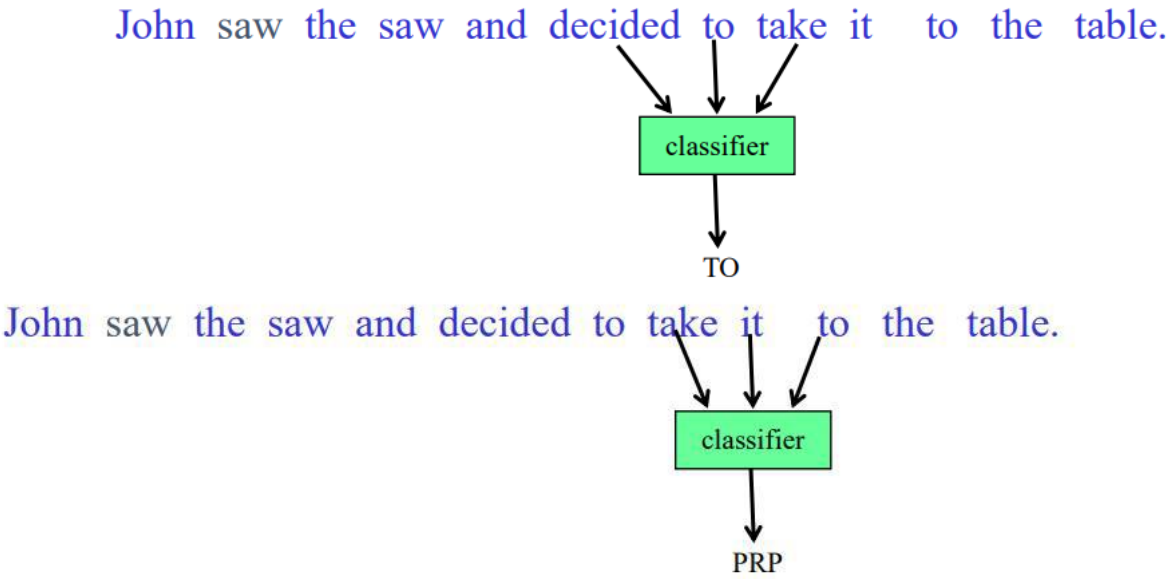
# Sequence labeling for NER

- Classify each token independently but...
  - Use as input features the information about the surrounding tokens
  - Sliding window approach



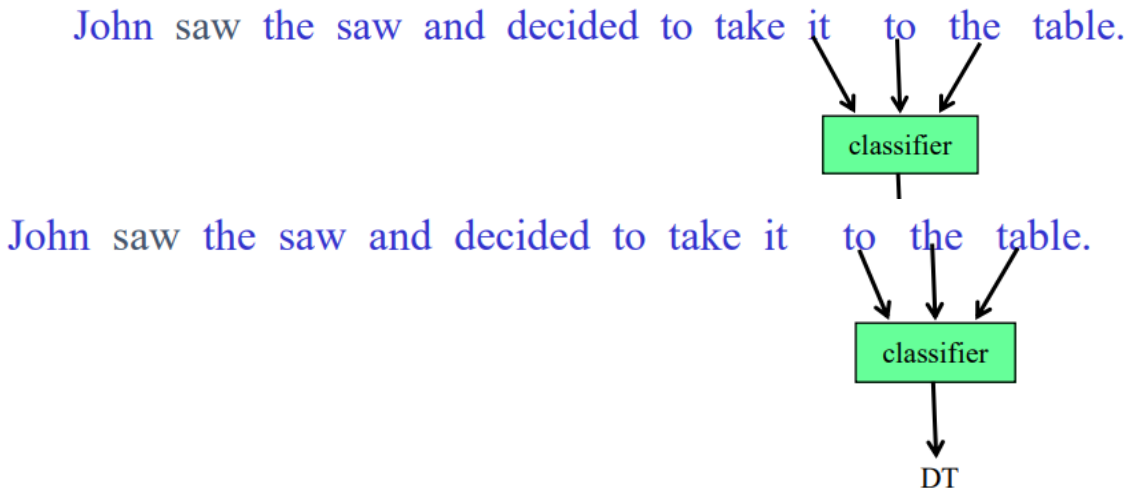
# Sequence labeling for NER

- Classify each token independently but...
  - Use as input features the information about the surrounding tokens
  - Sliding window approach

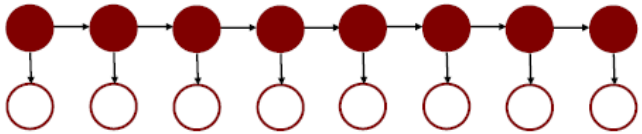


# Sequence labeling for NER

- Classify each token independently but...
  - Use as input features the information about the surrounding tokens
  - Sliding window approach

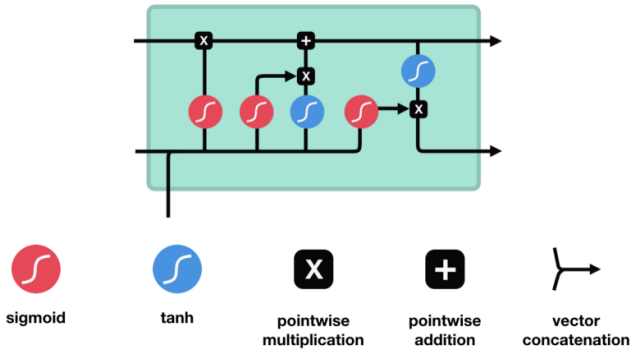


# Sequence labeling: techniques

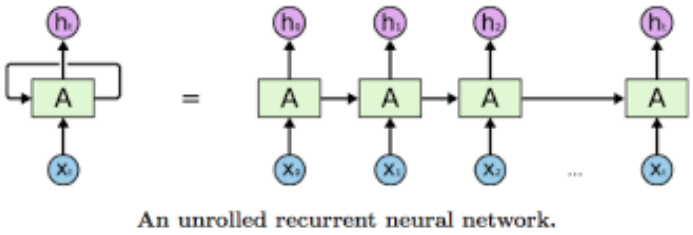


Hidden Markov Models

Conditional Random Forests



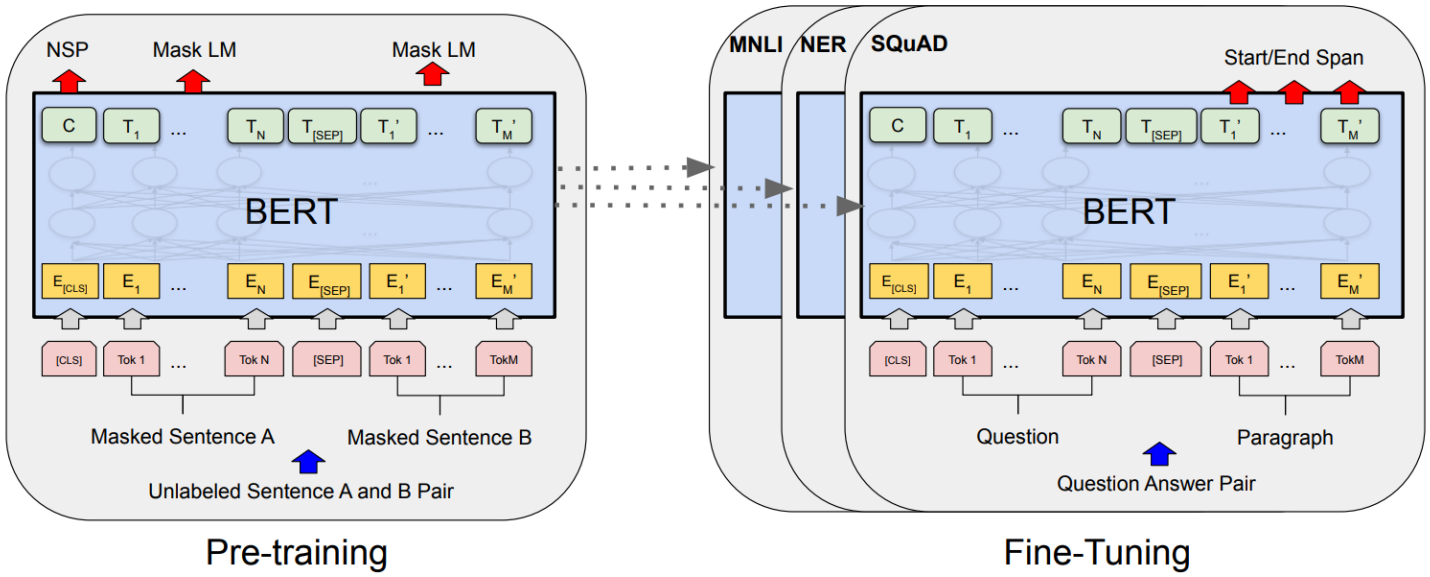

LSTMs




An unrolled recurrent neural network.

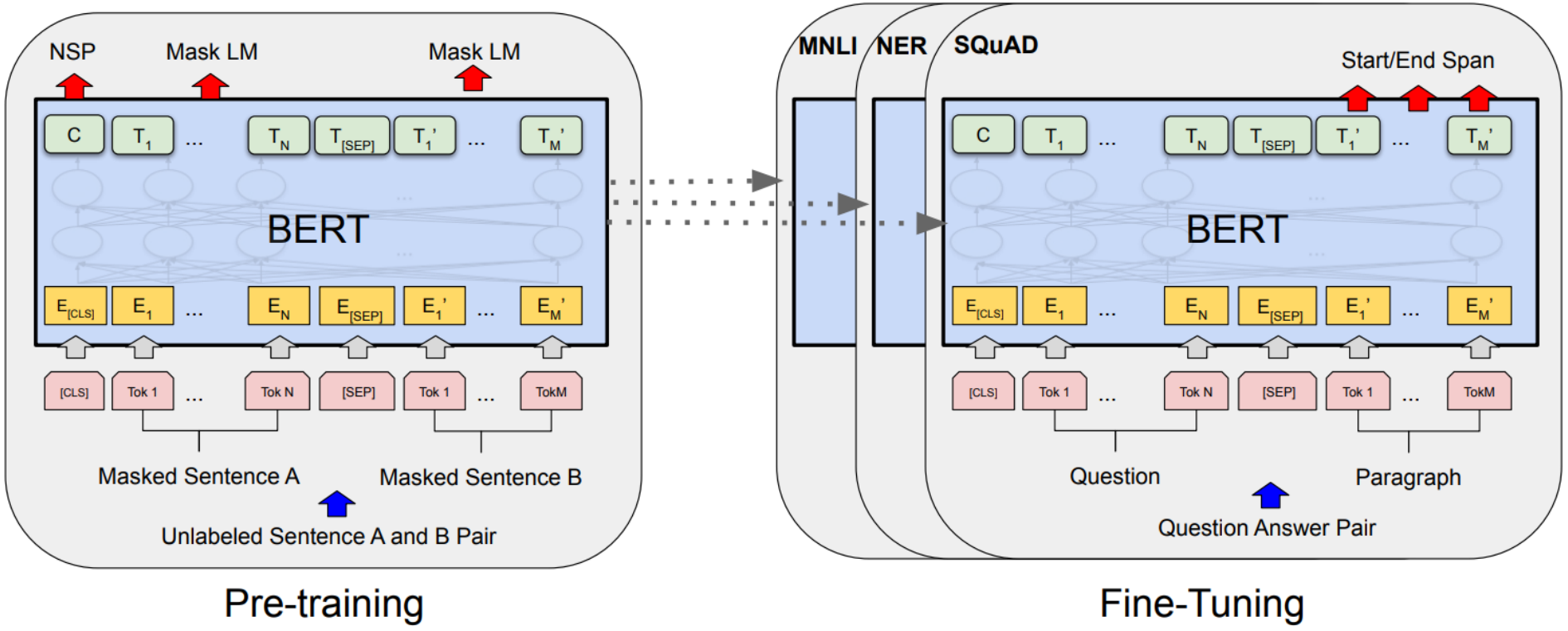
Recurrent Neural Networks

# Transformers for NER

# Transformers for NER

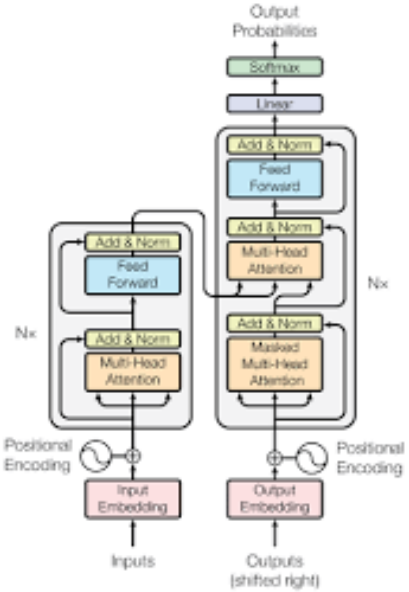
- Leverage attention-based architecture
  - Fine-tune the model for the NER task




BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Jacob Devlin Ming-Wei Chang Kenton Lee Kristina Toutanova. Proceedings of NAACL-HLT 2019.

# Transformers for NER: the LUKE architecture

- **L**anguage **U**nderstanding with **K**nowledge-Based **E**mbeddings
  - Presented at EMNLP'20
  - State-of-the-art approach for NER
- Key idea
  - use the attention mechanism to attend both word and entity relationships
- Based on a pretrained BERT Masked Language Model





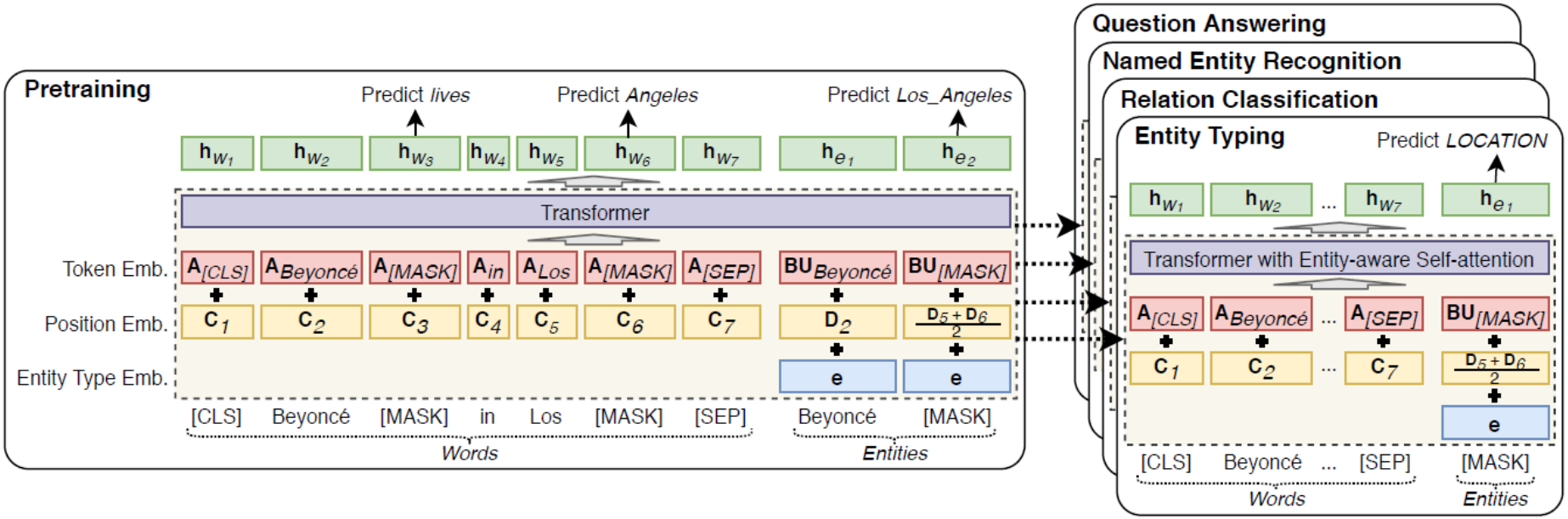
Ikuya Yamada, Akari Asai, Hiroyuki Shindo, Hideaki Takeda, Yuji Matsumoto:
LUKE: Deep Contextualized Entity Representations with Entity-aware Self-attention. EMNLP (1) 2020: 6442-6454

# Transformers for NER: the LUKE architecture

- Pretrained contextualized representations of both words and entities
  - Trained on large-scale entity-annotated corpus obtained from Wikipedia
    - It treats hyperlinks in Wikipedia as entity annotations


Ikuya Yamada, Akari Asai, Hiroyuki Shindo, Hideaki Takeda, Yuji Matsumoto:
LUKE: Deep Contextualized Entity Representations with Entity-aware Self-attention. EMNLP (1) 2020: 6442-6454

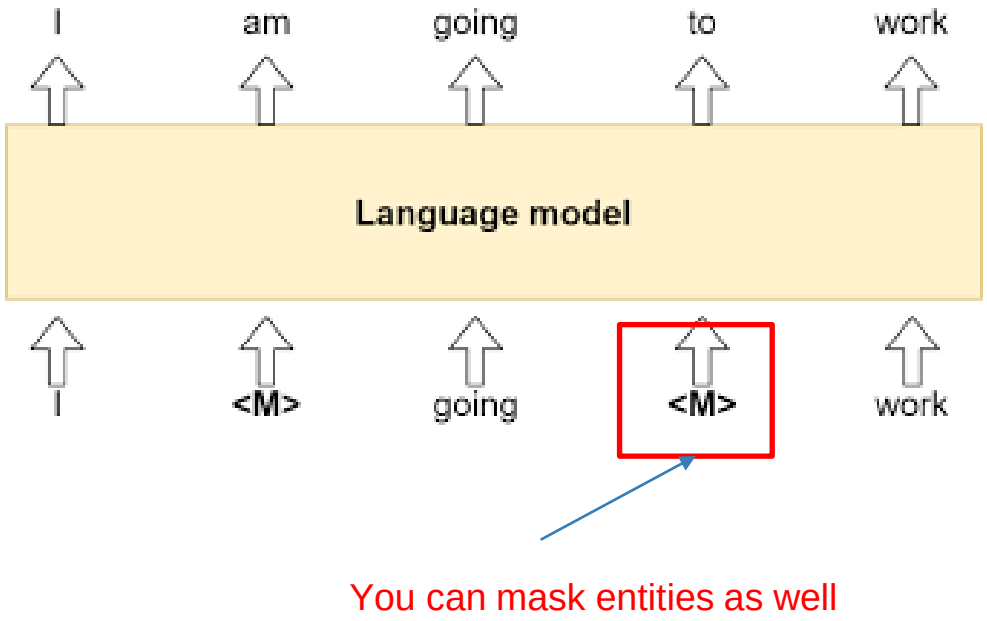
# Transformers for NER: the LUKE architecture




Ikuya Yamada, Akari Asai, Hiroyuki Shindo, Hideaki Takeda, Yuji Matsumoto: LUKE: Deep Contextualized Entity Representations with Entity-aware Self-attention. EMNLP (1) 2020: 6442-6454

# Transformers for NER: the LUKE architecture

- Extended Masked Language Model
  - Entities within the annotated text are randomly masked
  - The trained model predicts the original masked entities



# Transformers for NER: the LUKE architecture

- Entities are treated as independent tokens
- LUKE computes the intermediate and output representations of both word and entity tokens
  - It inherently models the relationships between entities
- Input embeddings
  - Token embedding
    - Representation of the token (either word or entity)
  - Position embedding
    - Position of the token in the word sequence
      - If an entity name contains multiple words, the single embeddings are averaged
  - Entity type embedding
    - Whether the token is an entity or not

Ikuya Yamada, Akari Asai, Hiroyuki Shindo, Hideaki Takeda, Yuji Matsumoto: LUKE: Deep Contextualized Entity Representations with Entity-aware Self-attention. EMNLP (1) 2020: 6442-6454

# Transformers for NER: the LUKE architecture

- Entity-aware attention mechanism
  - It relates tokens (either entities or words) each other based on the attention score between each pair of token
- Input vectors $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k$
- Output vectors $\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_k$

$$\mathbf{y}_i = \sum_{j=1}^{k} \alpha_{ij} \mathbf{V} \mathbf{x}_j$$

$$e_{ij} = \frac{\mathbf{K}\mathbf{x}_j^\top \mathbf{Q}\mathbf{x}_i}{\sqrt{L}}$$

$$\alpha_{ij} = \text{softmax}(e_{ij})$$

$$e_{ij} = \begin{cases} \mathbf{K}\mathbf{x}_j^\top \mathbf{Q}\mathbf{x}_i, & \text{if both } \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ are words} \\ \mathbf{K}\mathbf{x}_j^\top \mathbf{Q}_{w2e}\mathbf{x}_i, & \text{if } \mathbf{x}_i \text{ is word and } \mathbf{x}_j \text{ is entity} \\ \mathbf{K}\mathbf{x}_j^\top \mathbf{Q}_{e2w}\mathbf{x}_i, & \text{if } \mathbf{x}_i \text{ is entity and } \mathbf{x}_j \text{ is word} \\ \mathbf{K}\mathbf{x}_j^\top \mathbf{Q}_{e2e}\mathbf{x}_i, & \text{if both } \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ are entities} \end{cases}$$

$$\hat{\mathbf{y}} = \text{softmax}(\mathbf{B}\mathbf{T}\mathbf{m} + \mathbf{b}_o)$$

$$\mathbf{m} = \text{layer\_norm}\big(\text{gelu}(\mathbf{W}_h \mathbf{h}_e + \mathbf{b}_h)\big)$$

# Additional reading on LUKE

- Ikuya Yamada, Akari Asai, Hiroyuki Shindo, Hideaki Takeda, Yuji Matsumoto. LUKE: Deep Contextualized Entity Representations with Entity-aware Self-attention. EMNLP (1) 2020: 6442-6454
- Please read the paper: https://arxiv.org/pdf/2010.01057.pdf

# Acknowlegdements and copyright license

- Copyright licence
  - Attribution + Noncommercial + NoDerivatives


- Acknowledgements
  - I would like to thank Dr. Moreno La Quatra, who collaborated to the writing and revision of the teaching content
- Affiliation
  - The author and his staff are currently members of the Database and Data Mining Group at Dipartimento di Automatica e Informatica (Politecnico di Torino) and of the SmartData interdepartmental centre
    - https://dbdmg.polito.it
    - https://smartdata.polito.it

Thank you!