Computational Linguistics

Named Entity Recognition

Parameswari Krishnamurthy

Language Technologies Research Centre
IIIT-Hyderabad

param.krishna@iiit.ac.in

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

Named Entity Recognition (NER)

- **Definition**: A named entity is anything that can be referred to with a proper name: a person, a location, an organization.

Named Entity Recognition (NER)

- **Definition**: A named entity is anything that can be referred to with a proper name: a person, a location, an organization.
- **Task**:Finding spans of text that constitute proper names and tagging their entity type.

Named Entity Recognition (NER)

- **Definition**: A named entity is anything that can be referred to with a proper name: a person, a location, an organization.
- **Task**:Finding spans of text that constitute proper names and tagging their entity type.
- A fundamental task in natural language processing.

Named Entity Recognition (NER)

- **Definition**: A named entity is anything that can be referred to with a proper name: a person, a location, an organization.
- **Task**:Finding spans of text that constitute proper names and tagging their entity type.
- A fundamental task in natural language processing.
- Named entities are typically noun phrases that refer to specific types of individuals, places, or things.

Named Entity Recognition (NER)

- **Definition**: A named entity is anything that can be referred to with a proper name: a person, a location, an organization.
- **Task**:Finding spans of text that constitute proper names and tagging their entity type.
- A fundamental task in natural language processing.
- Named entities are typically noun phrases that refer to specific types of individuals, places, or things.
- However, the term named entity is commonly extended to include things that aren't entities per se, including dates, times, and other kinds of temporal expressions, and even numerical expressions like prices.

Named Entity Recognition (NER)

- **Definition**: A named entity is anything that can be referred to with a proper name: a person, a location, an organization.
- **Task**:Finding spans of text that constitute proper names and tagging their entity type.
- A fundamental task in natural language processing.
- Named entities are typically noun phrases that refer to specific types of individuals, places, or things.
- However, the term named entity is commonly extended to include things that aren't entities per se, including dates, times, and other kinds of temporal expressions, and even numerical expressions like prices.
- **Example of Named Entity in a sentence:**

  Marie Curie was born in **Warsaw, Poland** and later studied at **Sorbonne University**.

Named Entity Recognition (NER)

- **Identifying and classifying named entities** (e.g., person names, organization names, locations) in text.

Named Entity Recognition (NER)

- **Identifying and classifying named entities** (e.g., person names, organization names, locations) in text.
- Helps in extracting valuable information from unstructured text data.

Named Entity Recognition (NER)

- **Identifying and classifying named entities** (e.g., person names, organization names, locations) in text.
- Helps in extracting valuable information from unstructured text data.
- **Common types of named entities**:
  - **Person names** (e.g., John Smith)

Named Entity Recognition (NER)

- **Identifying and classifying named entities** (e.g., person names, organization names, locations) in text.
- Helps in extracting valuable information from unstructured text data.
- **Common types of named entities**:
  - **Person names** (e.g., John Smith)
  - **Organization names** (e.g., Google)

Named Entity Recognition (NER)

- **Identifying and classifying named entities** (e.g., person names, organization names, locations) in text.
- Helps in extracting valuable information from unstructured text data.
- **Common types of named entities**:
  - **Person names** (e.g., John Smith)
  - **Organization names** (e.g., Google)
  - **Location names** (e.g., New York)

Named Entity Recognition (NER)

- **Identifying and classifying named entities** (e.g., person names, organization names, locations) in text.
- Helps in extracting valuable information from unstructured text data.
- **Common types of named entities**:
  - **Person names** (e.g., John Smith)
  - **Organization names** (e.g., Google)
  - **Location names** (e.g., New York)
  - **Date and time expressions** (e.g., January 1, 2022)

Named Entity Recognition (NER)

- **Identifying and classifying named entities** (e.g., person names, organization names, locations) in text.
- Helps in extracting valuable information from unstructured text data.
- **Common types of named entities**:
  - **Person names** (e.g., John Smith)
  - **Organization names** (e.g., Google)
  - **Location names** (e.g., New York)
  - **Date and time expressions** (e.g., January 1, 2022)
  - **Monetary values** (e.g., $100)

Named Entity Recognition (NER)

- **Identifying and classifying named entities** (e.g., person names, organization names, locations) in text.
- Helps in extracting valuable information from unstructured text data.
- **Common types of named entities**:
  - **Person names** (e.g., John Smith)
  - **Organization names** (e.g., Google)
  - **Location names** (e.g., New York)
  - **Date and time expressions** (e.g., January 1, 2022)
  - **Monetary values** (e.g., $100)
  - **Percentages** (e.g., 80%)
- NER is an essential component in various NLP tasks such as information extraction, question answering, and document summarization.

Not NEs & NEs

Comparison of NOT Named Entities and Named Entities:

- **Hotel** & **Taj Hotel**

Not NEs & NEs

Comparison of NOT Named Entities and Named Entities:

- **Hotel** & **Taj Hotel**
- **Flower** & **Rose Flower**

Not NEs & NEs

Comparison of NOT Named Entities and Named Entities:

- **Hotel** & **Taj Hotel**
- **Flower** & **Rose Flower**
- **Beach** & **Kovalam Beach**

Not NEs & NEs

Comparison of NOT Named Entities and Named Entities:

- **Hotel** & **Taj Hotel**
- **Flower** & **Rose Flower**
- **Beach** & **Kovalam Beach**
- **Airport** & **Indira Gandhi International Airport**

Not NEs & NEs

Comparison of NOT Named Entities and Named Entities:

- **Hotel** & **Taj Hotel**
- **Flower** & **Rose Flower**
- **Beach** & **Kovalam Beach**
- **Airport** & **Indira Gandhi International Airport**
- **The School** & **Good Shepherd School**

Not NEs & NEs

Comparison of NOT Named Entities and Named Entities:

- **Hotel** & **Taj Hotel**
- **Flower** & **Rose Flower**
- **Beach** & **Kovalam Beach**
- **Airport** & **Indira Gandhi International Airport**
- **The School** & **Good Shepherd School**
- **Prime Minister** & **Mr. Manmohan Singh**

Illustration

Generic Named Entity Types:

Type	Tag	Sample Categories	Example sentences
People	PER	people, characters	**Turing** is a giant of computer science.
Organization	ORG	companies, sports teams	The **IPCC** warned about the cyclone.
Location	LOC	regions, mountains, seas	**Mt. Sanitas** is in **Sunshine Canyon**.
Geo-Political Entity	GPE	countries, states	**Palo Alto** is raising the fees for parking.

Examples of different generic named entity types.

Output of an NER Tagger:

Citing high fuel prices, [$_{\text{ORG}}$ **United Airlines**] said [$_{\text{TIME}}$ **Friday**] it has increased fares by [$_{\text{MONEY}}$ **$6**] per round trip on flights to some cities also served by lower-cost carriers. [$_{\text{ORG}}$ **American Airlines**], a unit of [$_{\text{ORG}}$ **AMR Corp.**], immediately matched the move, spokesman [$_{\text{PER}}$ **Tim Wagner**] said. [$_{\text{ORG}}$ **United**], a unit of [$_{\text{ORG}}$ **UAL Corp.**], said the increase took effect [$_{\text{TIME}}$ **Thursday**] and applies to most routes where it competes against discount carriers, such as [$_{\text{LOC}}$ **Chicago**] to [$_{\text{LOC}}$ **Dallas**] and [$_{\text{LOC}}$ **Denver**] to [$_{\text{LOC}}$ **San Francisco**].

Description: Example output of a NER tagger

Applications of NER

- **Sentiment analysis:** Identifying sentiment towards specific entities.
  - Example: Determining if reviews mention a product like "iPhone 14" positively or negatively.

Applications of NER

- **Sentiment analysis:** Identifying sentiment towards specific entities.
  - Example: Determining if reviews mention a product like "iPhone 14" positively or negatively.
- **Question answering:** Extracting relevant entities for answering queries.
  - Example: Answering "Who is the CEO of Apple?" by identifying "Tim Cook".

Applications of NER

- **Sentiment analysis:** Identifying sentiment towards specific entities.
  - Example: Determining if reviews mention a product like "iPhone 14" positively or negatively.
- **Question answering:** Extracting relevant entities for answering queries.
  - Example: Answering "Who is the CEO of Apple?" by identifying "Tim Cook".
- **Information extraction:** Building structured data from unstructured text.
  - Example: Extracting company names and financial figures from news articles.

Applications of NER

- **Sentiment analysis:** Identifying sentiment towards specific entities.
  - Example: Determining if reviews mention a product like "iPhone 14" positively or negatively.
- **Question answering:** Extracting relevant entities for answering queries.
  - Example: Answering "Who is the CEO of Apple?" by identifying "Tim Cook".
- **Information extraction:** Building structured data from unstructured text.
  - Example: Extracting company names and financial figures from news articles.
- **Semantic search:** Improving search relevance by understanding entity types.
  - Example: Enhancing search results for "restaurants" by understanding entity types like "Italian" or "vegan".

Applications of NER

- **Sentiment analysis:** Identifying sentiment towards specific entities.
  - Example: Determining if reviews mention a product like "iPhone 14" positively or negatively.
- **Question answering:** Extracting relevant entities for answering queries.
  - Example: Answering "Who is the CEO of Apple?" by identifying "Tim Cook".
- **Information extraction:** Building structured data from unstructured text.
  - Example: Extracting company names and financial figures from news articles.
- **Semantic search:** Improving search relevance by understanding entity types.
  - Example: Enhancing search results for "restaurants" by understanding entity types like "Italian" or "vegan".
- **Content recommendation:** Suggesting related content based on entities.
  - Example: Recommending news articles about "Elon Musk" if a user frequently reads about "SpaceX".

What NER is NOT

- **Event Recognition:**
  - NER focuses on identifying entities, not the events in which they participate.
- **Template Creation:**
  - NER does not generate templates for documents or texts.
- **Coreference or Entity Linking:**
  - NER does not handle coreference resolution or linking entities across texts.
  - These processes are often part of a broader Information Extraction (IE) system.
- **Simple Text Matching:**
  - NER is not just about matching text strings with pre-defined name lists.
  - It involves recognizing entities based on their contextual usage.
- **NER is Not an Easy Task!**

BIO Tagging for NER

Tagging Scheme:

- **B:** Beginning of entity
- **I:** Inside entity
- **O:** Outside any entity

BIO Tagging for NER

Tagging Scheme:

- **B:** Beginning of entity
- **I:** Inside entity
- **O:** Outside any entity

Example:

- [$_{\text{PER}}$ Jane Villanueva] of [$_{\text{ORG}}$ United Airlines]
- Jane (B-PER) Villanueva (I-PER) of (O) United (B-ORG) Airlines (I-ORG)

BIO Tagging Variants

IO Tagging:

- **I:** Inside entity
- **O:** Outside entity

BIO Tagging Variants

IO Tagging:

- **I:** Inside entity
- **O:** Outside entity

BIOES Tagging:

- **B:** Beginning of multi-token entity
- **I:** Inside multi-token entity
- **O:** Outside any entity
- **E:** End of multi-token entity
- **S:** Single-token entity

Illustration

- **The text:**

[$_{PER}$ **Jane Villanueva**] of [$_{ORG}$ **United**] , a unit of [$_{ORG}$ **United Airlines Holding**] , said the fare applies to the [$_{LOC}$ **Chicago**] route.

Illustration

- **The text:**

[$_{PER}$ **Jane Villanueva**] of [$_{ORG}$ **United**] , a unit of [$_{ORG}$ **United Airlines Holding**] , said the fare applies to the [$_{LOC}$ **Chicago**] route.

- **Levels of BIO Tagging:**

Words	IO Label	BIO Label	BIOES Label
Jane	I-PER	B-PER	B-PER
Villanueva	I-PER	I-PER	E-PER
of	O	O	O
United	I-ORG	B-ORG	B-ORG
Airlines	I-ORG	I-ORG	I-ORG
Holding	I-ORG	I-ORG	E-ORG
discussed	O	O	O
the	O	O	O
Chicago	I-LOC	B-LOC	S-LOC
route	O	O	O
.	O	O	O

Examples of NER Tagsets

- **ACE Tagset (Automatic Content Extraction):**
  - Hierarchical structure
  - Categories include entities like Person, Organization, Location, and more

Examples of NER Tagsets

- **ACE Tagset (Automatic Content Extraction):**
  - Hierarchical structure
  - Categories include entities like Person, Organization, Location, and more
- **CLIA Tagset:**
  - Hierarchical structure similar to ACE
  - Developed for specific domains:
    - Tourism
    - Health

Examples of NER Tagsets

- **ACE Tagset (Automatic Content Extraction):**
  - Hierarchical structure
  - Categories include entities like Person, Organization, Location, and more
- **CLIA Tagset:**
  - Hierarchical structure similar to ACE
  - Developed for specific domains:
    - Tourism
    - Health
- **ENAMEX:**
  - Tags for named entities
  - Categories include Person (PER), Organization (ORG), Location (LOC), etc.

Examples of NER Tagsets

- **ACE Tagset (Automatic Content Extraction):**
  - Hierarchical structure
  - Categories include entities like Person, Organization, Location, and more
- **CLIA Tagset:**
  - Hierarchical structure similar to ACE
  - Developed for specific domains:
    - Tourism
    - Health
- **ENAMEX:**
  - Tags for named entities
  - Categories include Person (PER), Organization (ORG), Location (LOC), etc.
- **NUMEX:**
  - Tags for numerical expressions
  - Includes dates, times, and quantities

Examples of NER Tagsets

- **ACE Tagset (Automatic Content Extraction):**
  - Hierarchical structure
  - Categories include entities like Person, Organization, Location, and more
- **CLIA Tagset:**
  - Hierarchical structure similar to ACE
  - Developed for specific domains:
    - Tourism
    - Health
- **ENAMEX:**
  - Tags for named entities
  - Categories include Person (PER), Organization (ORG), Location (LOC), etc.
- **NUMEX:**
  - Tags for numerical expressions
  - Includes dates, times, and quantities
- **TIMEX:**
  - Tags for temporal expressions
  - Includes dates, times, durations

Example

TAGSET

- **ENAMEX**
  - Person
    - Individual
      - Family name
      - Title
    - Group
  - Organization
    - Government
    - Public/private company
    - Religious
    - Non-government
      - Political Party
      - Para military
      - Charitable
      - Association
    - GPE (Geo-political Social Entity)
    - Media
  - Location
    - Place
      - District
      - City
      - State
      - Nation
      - Continent
    - Address
    - Water-bodies
    - Landscapes
    - Celestial Bodies
  - Manmade
    - » Religious Places
    - » Roads/Highways
    - » Museum
    - » Theme parks/Parks/Gardens
    - » Monuments
- Facilities
  - Hospitals
- Institutes
- Library
  - Hotel/Restaurants/Lodges
  - Plant/Factories
  - Police Station/Fire Services
  - Public Comfort Stations
  - Airports
  - Ports
  - Bus-Stations
- Locomotives
- Artifacts
  - Implements
  - Ammunition
  - Paintings
  - Sculptures
  - Cloths
  - Gems & Stones
- Entertainment
  - Dance
  - Music
  - Drama/Cinema
  - Sports
  - Events/Exhibitions/Conferences
- Cuisine's
- Animals
- Plants

Sequence Labeling & Standard Algorithms for NER

- A sequence labeler is trained to label each token in a text with tags that indicate the presence (or absence) of particular kinds of named entities.

Sequence Labeling & Standard Algorithms for NER

- A sequence labeler is trained to label each token in a text with tags that indicate the presence (or absence) of particular kinds of named entities.
- Standard Algorithms for NER:
  - **Hidden Markov Models (HMM):** Statistical models that predict sequences of states.

Sequence Labeling & Standard Algorithms for NER

- A sequence labeler is trained to label each token in a text with tags that indicate the presence (or absence) of particular kinds of named entities.
- Standard Algorithms for NER:
  - **Hidden Markov Models (HMM):** Statistical models that predict sequences of states.
  - **Conditional Random Fields (CRF) / Maximum Entropy Markov Models (MEMM):** Advanced statistical models for sequence labeling.

Sequence Labeling & Standard Algorithms for NER

- A sequence labeler is trained to label each token in a text with tags that indicate the presence (or absence) of particular kinds of named entities.
- Standard Algorithms for NER:
  - **Hidden Markov Models (HMM):** Statistical models that predict sequences of states.
  - **Conditional Random Fields (CRF) / Maximum Entropy Markov Models (MEMM):** Advanced statistical models for sequence labeling.
  - **Supervised Machine Learning:** Given a human-labeled training set of text annotated with tags.

Sequence Labeling & Standard Algorithms for NER

- A sequence labeler is trained to label each token in a text with tags that indicate the presence (or absence) of particular kinds of named entities.
- Standard Algorithms for NER:
  - **Hidden Markov Models (HMM):** Statistical models that predict sequences of states.
  - **Conditional Random Fields (CRF) / Maximum Entropy Markov Models (MEMM):** Advanced statistical models for sequence labeling.
  - **Supervised Machine Learning:** Given a human-labeled training set of text annotated with tags.
  - **Neural Sequence Models:** Recurrent Neural Networks (RNNs) or Transformers that learn from data representations.

Sequence Labeling & Standard Algorithms for NER

- A sequence labeler is trained to label each token in a text with tags that indicate the presence (or absence) of particular kinds of named entities.
- Standard Algorithms for NER:
  - **Hidden Markov Models (HMM):** Statistical models that predict sequences of states.
  - **Conditional Random Fields (CRF) / Maximum Entropy Markov Models (MEMM):** Advanced statistical models for sequence labeling.
  - **Supervised Machine Learning:** Given a human-labeled training set of text annotated with tags.
  - **Neural Sequence Models:** Recurrent Neural Networks (RNNs) or Transformers that learn from data representations.
  - **Large Language Models (e.g., BERT):** Pre-trained models fine-tuned for specific NER tasks.

Challenges in NER Tagging

- **Segmentation Ambiguity:** Unlike part-of-speech tagging where each word gets one tag, NER involves identifying and labeling spans of text. This segmentation is challenging due to ambiguity in defining entity boundaries.

Challenges in NER Tagging

- **Segmentation Ambiguity:** Unlike part-of-speech tagging where each word gets one tag, NER involves identifying and labeling spans of text. This segmentation is challenging due to ambiguity in defining entity boundaries.
- **Determining Entity Boundaries:** It is necessary to decide what constitutes an entity and where the boundaries lie. Many words in a text will not be named entities, complicating the identification process.

Challenges in NER Tagging

- **Segmentation Ambiguity:** Unlike part-of-speech tagging where each word gets one tag, NER involves identifying and labeling spans of text. This segmentation is challenging due to ambiguity in defining entity boundaries.
- **Determining Entity Boundaries:** It is necessary to decide what constitutes an entity and where the boundaries lie. Many words in a text will not be named entities, complicating the identification process.
- **Type Ambiguity:** Distinguishing between different types of entities can be difficult. Entities may overlap or belong to multiple categories, adding complexity to the tagging process.

Challenges in NER Tagging

- **Segmentation Ambiguity:** Unlike part-of-speech tagging where each word gets one tag, NER involves identifying and labeling spans of text. This segmentation is challenging due to ambiguity in defining entity boundaries.
- **Determining Entity Boundaries:** It is necessary to decide what constitutes an entity and where the boundaries lie. Many words in a text will not be named entities, complicating the identification process.
- **Type Ambiguity:** Distinguishing between different types of entities can be difficult. Entities may overlap or belong to multiple categories, adding complexity to the tagging process.
- **Category Definitions and Metonymy:** Entities that overlap or span multiple categories.

Challenges in NER Tagging

- **Segmentation Ambiguity:** Unlike part-of-speech tagging where each word gets one tag, NER involves identifying and labeling spans of text. This segmentation is challenging due to ambiguity in defining entity boundaries.

- **Determining Entity Boundaries:** It is necessary to decide what constitutes an entity and where the boundaries lie. Many words in a text will not be named entities, complicating the identification process.

- **Type Ambiguity:** Distinguishing between different types of entities can be difficult. Entities may overlap or belong to multiple categories, adding complexity to the tagging process.

- **Category Definitions and Metonymy:** Entities that overlap or span multiple categories.
  - Category definitions are intuitively quite clear, but there are many grey areas.
  - Many of these grey areas are caused by metonymy:
    - Person vs. Artefact
    - Organisation vs. Location
    - Company vs. Artefact
    - Location vs. Organisation

Ambiguity Types as Challenges in NER

- **Type Ambiguity:** Different entities might have overlapping or ambiguous classifications.

Ambiguity Types as Challenges in NER

- **Type Ambiguity:** Different entities might have overlapping or ambiguous classifications.
  - Example: **Apple** can refer to a company or a fruit.
  - Text: Apple Inc. is known for its apple products.

Ambiguity Types as Challenges in NER

- **Type Ambiguity:** Different entities might have overlapping or ambiguous classifications.
  - Example: **Apple** can refer to a company or a fruit.
  - Text: Apple Inc. is known for its apple products.
- **Boundary Ambiguity:** Difficulties in determining the exact boundaries of an entity.
  - Example: New York can be part of a larger entity.
  - Text: New York City is a major city in New York State.

Ambiguity Types as Challenges in NER

- **Type Ambiguity:** Different entities might have overlapping or ambiguous classifications.
    - Example: **Apple** can refer to a company or a fruit.
    - Text: Apple Inc. is known for its apple products.
- **Boundary Ambiguity:** Difficulties in determining the exact boundaries of an entity.
    - Example: New York can be part of a larger entity.
    - Text: New York City is a major city in New York State.
- **Overlapping Entities:** Entities that overlap or span multiple categories.
    - Example: Barack Obama as a person and President Obama as a title.
    - Text: Barack Obama was the President of the United States.
- **A More Realistic Example:**

> [$_{\text{PER}}$ Washington] was born into slavery on the farm of James Burroughs.
> [$_{\text{ORG}}$ Washington] went up 2 games to 1 in the four-game series.
> Blair arrived in [$_{\text{LOC}}$ Washington] for what may well be his last state visit.
> In June, [$_{\text{GPE}}$ Washington] passed a primary seatbelt law.

Challenges in Indian Language NER

- **Diverse Language Families:**
  - Indian languages belong to several language families: Indo-Aryan, Dravidian, Tibeto-Burman,Austro-Asiatic and other Isolates.

Challenges in Indian Language NER

- **Diverse Language Families:**
  - Indian languages belong to several language families: Indo-Aryan, Dravidian, Tibeto-Burman,Austro-Asiatic and other Isolates.
- **Morphologically Rich:**
  - Many Indian languages are morphologically rich and agglutinative.

Challenges in Indian Language NER

- **Diverse Language Families:**
  - Indian languages belong to several language families: Indo-Aryan, Dravidian, Tibeto-Burman,Austro-Asiatic and other Isolates.
- **Morphologically Rich:**
  - Many Indian languages are morphologically rich and agglutinative.
- **No Capitalization Feature:**
  - Unlike English, Indian languages lack capitalization as a feature.

Challenges in Indian Language NER

- **Diverse Language Families:**
  - Indian languages belong to several language families: Indo-Aryan, Dravidian, Tibeto-Burman,Austro-Asiatic and other Isolates.
- **Morphologically Rich:**
  - Many Indian languages are morphologically rich and agglutinative.
- **No Capitalization Feature:**
  - Unlike English, Indian languages lack capitalization as a feature.
- **Ambiguity:**
  - Ambiguity between common and proper nouns.
  - Example: "Roja" means Rose flower but is also a person's name.

Challenges in Indian Language NER Contd.

- **Spell Variations:**
  - Web data shows different spellings of the same entity. .

Challenges in Indian Language NER Contd.

- **Spell Variations:**
  - Web data shows different spellings of the same entity. .
- **Less Resources:**
  - Many Indian languages are less resourced.
  - Limited automated tools for preprocessing tasks like Part-of-Speech tagging and chunking.
  - Tools that do exist often have lower performance. .

Challenges in Indian Language NER Contd.

- **Spell Variations:**
  - Web data shows different spellings of the same entity. .
- **Less Resources:**
  - Many Indian languages are less resourced.
  - Limited automated tools for preprocessing tasks like Part-of-Speech tagging and chunking.
  - Tools that do exist often have lower performance. .
- **Lack of Annotated Data:**
  - Few efforts in developing NER systems for Indian languages.

Challenges in Indian Language NER Contd.

- **Spell Variations:**
  - Web data shows different spellings of the same entity. .
- **Less Resources:**
  - Many Indian languages are less resourced.
  - Limited automated tools for preprocessing tasks like Part-of-Speech tagging and chunking.
  - Tools that do exist often have lower performance. .
- **Lack of Annotated Data:**
  - Few efforts in developing NER systems for Indian languages.
  - Scarcity of easily accessible NE-annotated corpora in the community.

Evaluation of NER Systems

- **Metrics:**
  - **Precision:**
    Correctly identified entities / Total identified entities
  - **Recall:**
    Correctly identified entities / Total actual entities
  - **F1-score:**
    Harmonic mean of precision and recall
- **Modern Metrics:**
  - **Exact Match Ratio:** Measures the proportion of entities that are correctly identified with exact matches.
  - **Entity-Level F1-score:** Evaluates precision, recall, and F1-score at the entity level rather than the token level.
- **Challenges in Evaluation:**
  - Importance of consistent annotation guidelines
  - Partial matches (e.g., "President Obama" vs. "Obama")
  - Cross-domain evaluation: Testing on different text genres
  - Cross-lingual evaluation: Assessing performance across languages