Difference Between Text Mining and NLP

School of Information Studies
Syracuse University

Text Mining and NLP

NLP: deep linguistic analysis

- A typical NLP package includes:
  - Part-of-speech tagger
  - Parser
  - Named-entity recognizer
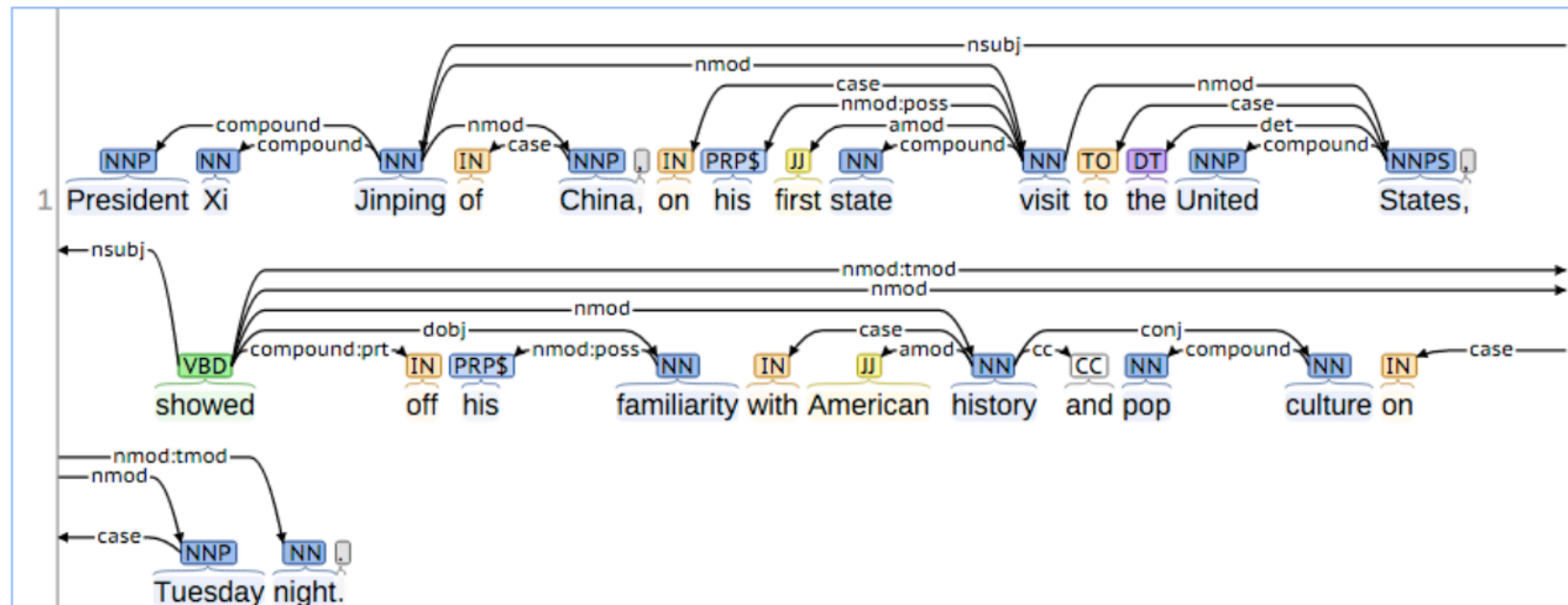 - Co-reference resolution system

School of Information Studies
Syracuse University

Named Entity Recognition:

1 President Xi [Person] Jinping of China [Loc], on his first [ORDINAL] state visit to the United States [Location], showed off his familiarity with American [Misc] history and pop culture on Tuesday [Date] night [Time].

Coreference:

1 President Xi Jinping of China, on his first state visit to the United States, showed off his [M] familiarity with American history and pop culture on Tuesday night. [Mention] ----Coref----

Basic Dependencies:

School of Information Studies
Syracuse University

Text Mining and NLP

NLP

- Deep linguistic analysis
- May take a long time to analyze large text collections

Text mining

- Use shallow features, usually N-grams, to analyze large text collections quickly
- Sometimes use deep NLP features like PoS tags or dependencies for feature engineering

School of Information Studies
Syracuse University

Text Mining and Information Retrieval

Information retrieval (IR): given a query, find relevant documents

Text mining can help information retrieval

- Clustering retrieval results for focused navigation
- Categorize documents for more precise results (e.g., find documents with certain readability levels)

Techniques from IR are used in text mining

- Example: TFIDF weighting

School of Information Studies
Syracuse University

School of Information Studies
Syracuse University