

COMP90042

Workshop Week 08

□ Homework 4

□ Due: Wednesday, May 2

Syllabus

1	Introduction and Preprocessing	Text classification
2	Lexical semantics	Distributional semantics
3	Part of Speech Tagging	Hidden Markov Models
4	Unsupervised Hidden Markov Models	Context-Free Grammars
5	Probabilistic Parsing	Dependency parsing
	<i>Easter holiday break</i>	
6	N-gram language models	Neural language models
7	Information Extraction	Question Answering
8	Topic Models	<i>ANZAC day holiday</i>
9	Information Retrieval -- Boolean search and the vector space model	Indexing and querying in the vector space model, evaluation
10	Index and vocabulary compression	Efficient query processing
11	The Web as a Graph: Page-rank & HITS	Machine Translation (word based)
12	Machine translation (phrase based) and neural encoder-decoder	Subject review

Outline

- Information extraction (IE)
 - Named entity recognition
 - Relation extraction (RE)
- Question answering
 - Parsing
 - QA as IE/RE, QA as IR

Information extraction

- Main goal:

- Turn text into structured data

- Text

- “Brasilia, the Brazilian capital, was founded in 1960.”

- Structured data

- capital(Brazil, Brasilia), founded(Brasilia, 1960)

- Two steps:

- Named Entity Recognition (NER)

- **Find** entities such as “Brasilia” and “1960”

- Use context to **classify** the entities such as “Brasilia” → “LOC”

- Relation Extraction

- Use context to find the relation between “Brasilia” and “1960” (“founded”)

Named entity recognition

- ❑ Hand-written rules
 - ❑ regular expressions
- ❑ Using classifiers
 - ❑ Generative: Naïve Bayes
 - ❑ Discriminative: Maxent models
- ❑ Sequence models
 - ❑ HMMs
 - ❑ CMMs/MEMMs
 - ❑ CRFs

https://web.stanford.edu/class/cs124/lec/Information_Extraction_and_Named_Entity_Recognition.pptx

Named entity recognition (find & classify)

- ❑ Use regular expressions to find & classify
- ❑ Use regular expressions to find, classifiers to classify
- ❑ Use sequence models to find & classify

- ❑ Rules usually give high precision & low recall
 - ❑ Too many ways to express the same thing, hard to cover all
- ❑ Sequence modelling as supervised learning
 - ❑ A more systematic way, but not effective for some cases
- ❑ Combine them in practice

NER - FEATURES

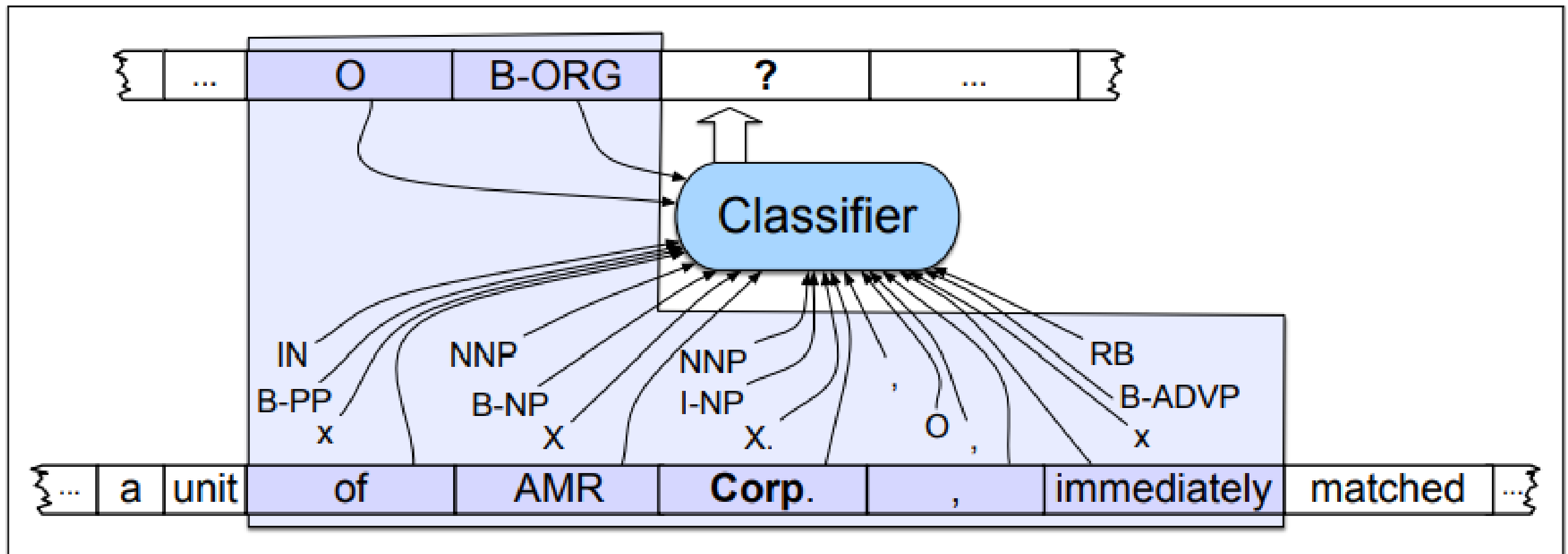


Figure 21.7 Named entity recognition as sequence labeling. The features available to the classifier during training and classification are those in the boxed area.

Encoding classes for sequence labeling

	IO encoding	IOB encoding
Fred	PER	B-PER
showed	O	O
Sue	PER	B-PER
Mengqiu	PER	B-PER
Huang	PER	I-PER
's	O	O
new	O	O
painting	O	O

https://web.stanford.edu/class/cs124/lec/Information_Extraction_and_Named_Entity_Recognition.pptx

Relation extraction - methods

- ❑ If we have access to a fixed relation database:
 - ❑ Rule-based
 - ❑ Supervised
 - ❑ Semi-supervised
 - ❑ Distant supervision
- ❑ If no restrictions on relations:
 - ❑ Unsupervised
 - ❑ Sometimes referred as “OpenIE”

Supervised relation extraction

❑ Binary Relation Association as Binary Classification

Christos Faloutsos conferred with **Ted Senator**, the **KDD 2003 General Chair**.

Person

Person

Role

Person-Role (**Christos Faloutsos**, **KDD 2003 General Chair**) → NO

Person-Role (**Ted Senator**, **KDD 2003 General Chair**) → YES

https://web.stanford.edu/class/cs124/lec/Information_Extraction_and_Named_Entity_Recognition.pptx

Outline

- Information extraction (IE)
 - Named entity recognition
 - Relation extraction (RE)
- Question answering
 - Parsing
 - QA as IE/RE, QA as IR

Parsing

- ☐ Syntactic parsing
- ☐ Dependency parsing
- ☐ Semantic parsing

Parsing

□ Syntactic parsing
(structural relations)

(S (NP Donald Trump)
(VP is
(NP (NP the president)
(PP of
(NP United States))))
.)

□ Dependency parsing



Donald Trump is the president of United States .

□ Semantic parsing
(meaning-based relations)

Donald	Leader
Trump	
is	
the	
president	Leadership
of	Governed
United	
States	
.	

QA as IE/RE

- ❑ iPhone (/ˈaɪfoʊn/ EYE-fohn) is a line of smartphones designed and marketed by Apple Inc. The iPhone line of products use Apple's iOS mobile operating system software. The first-generation iPhone was released on June 29, 2007, and multiple new hardware iterations with new iOS releases have been released since.

Named entity recognition

□ **iPhone** (/ˈaɪfoʊn/ EYE-fohn) is a line of smartphones designed and marketed by **Apple Inc.** The **iPhone** line of products use **Apple's iOS** mobile operating system software. The first-generation **iPhone** was released on **June 29, 2007**, and multiple new hardware iterations with new **iOS** releases have been released since.

Generate a list of relations, knowledge base

□ **iPhone** (/ˈaɪfoʊn/ EYE-fohn) is a line of smartphones designed and marketed by **Apple Inc.** The **iPhone** line of products use **Apple's iOS** mobile operating system software. The first-generation **iPhone** was released on **June 29, 2007**, and multiple new hardware iterations with new **iOS** releases have been released since.

□ “iPhone”, designed and marketed, “Apple Inc.”

□ “iPhone”, use, “Apple’s iOS”

□ “iPhone”, released, “June 29, 2007”

□ ...

Receive a question

☐ Question: When was the first generation iPhone released?

☐ “iPhone”, designed and marketed, “Apple Inc.”

☐ “iPhone”, use, “Apple’s iOS”

☐ “iPhone”, released, “June 29, 2007”

☐ ...

Transform the question into a query

❑ Question: When was the first generation iPhone released?

❑ Query: “iPhone”, released, ?

❑ “iPhone”, designed and marketed, “Apple Inc.”

❑ “iPhone”, use, “Apple’s iOS”

❑ “iPhone”, released, “June 29, 2007”

❑ ...

Find the facts match the query

☐ Question: When was the first generation iPhone released?

☐ Query: “iPhone”, released, ?

☐ “iPhone”, designed and marketed, “Apple Inc.”

☐ “iPhone”, use, “Apple’s iOS”

☐ “iPhone”, released, “June 29, 2007”

☐ ...

Return the answer

☐ Question: When was the first generation iPhone released?

☐ Query: “iPhone”, released, ?

☐ “iPhone”, designed and marketed, “Apple Inc.”

☐ “iPhone”, use, “Apple’s iOS”

☐ “iPhone”, released, “June 29, 2007”

☐ ...

☐ Answer: “June 29, 2007”

QA as IR

- ❑ iPhone (/ˈaɪfoʊn/ EYE-fohn) is a line of smartphones designed and marketed by Apple Inc. The iPhone line of products use Apple's iOS mobile operating system software. The first-generation iPhone was released on June 29, 2007, and multiple new hardware iterations with new iOS releases have been released since.

Process the text into a suitable format

- ❑ iPhone (/ˈaɪfoʊn/ EYE-fohn) is a line of smartphones designed and marketed by Apple Inc. The iPhone line of products use Apple's iOS mobile operating system software. The first-generation iPhone was released on June 29, 2007, and multiple new hardware iterations with new iOS releases have been released since.
- ❑ S1: iPhone (/ˈaɪfoʊn/ EYE-fohn) is a line of smartphones designed and marketed by Apple Inc.
- ❑ S2: The iPhone line of products use Apple's iOS mobile operating system software.
- ❑ S3: The first-generation iPhone was released on June 29, 2007, and multiple new hardware iterations with new iOS releases have been released since.

Receive a question

- ❑ Question: When was the first generation iPhone released?

- ❑ S1: iPhone (/ˈaɪfoʊn/ EYE-fohn) is a line of smartphones designed and marketed by Apple Inc.
- ❑ S2: The iPhone line of products use Apple's iOS mobile operating system software.
- ❑ S3: The first-generation iPhone was released on June 29, 2007, and multiple new hardware iterations with new iOS releases have been released since.

Query formation & answer type detection

- ❑ Question: When was the first generation iPhone released?
- ❑ Query: “iPhone”, “released”, Answer type: “When”
- ❑ S1: iPhone (/ˈaɪfoʊn/ EYE-fohn) is a line of smartphones designed and marketed by Apple Inc.
- ❑ S2: The iPhone line of products use Apple's iOS mobile operating system software.
- ❑ S3: The first-generation iPhone was released on June 29, 2007, and multiple new hardware iterations with new iOS releases have been released since.

Retrieve the most relevant passage(s)

- ❑ Question: When was the first generation iPhone released?
- ❑ Query: “iPhone”, “released”, Answer type: “When”
- ❑ S1: iPhone (/ˈaɪfoʊn/ EYE-fohn) is a line of smartphones designed and marketed by Apple Inc.
- ❑ S2: The iPhone line of products use Apple's iOS mobile operating system software.
- ❑ S3: The first-generation iPhone was released on June 29, 2007, and multiple new hardware iterations with new iOS releases have been released since.

Extract candidate answers & rank & return

- ❑ Question: When was the first generation iPhone released?
- ❑ Query: “iPhone”, “released”, Answer type: “When”
- ❑ S1: iPhone (/ˈaɪfoʊn/ EYE-fohn) is a line of smartphones designed and marketed by Apple Inc.
- ❑ S2: The iPhone line of products use Apple's iOS mobile operating system software.
- ❑ S3: The first-generation iPhone was released on June 29, 2007, and multiple new hardware iterations with new iOS releases have been released since.
- ❑ Answer: “June 29, 2007”