

# Event Extraction

Slides were adapted from Ellen Riloff

# Event Information Extraction

Extracting role fillers associated with events.

## Examples

**Terrorism:** perpetrator, victim, target, date, location

**Management succession:** person fired, successor,  
position, organization, date

**Disease outbreaks:** disease, victim, symptoms,  
containment measures

# Sample Text

*perpetrator*

Alleged guerrilla urban commandos launched two

*weapon*

highpower bombs against

*target*

a car dealership in downtown

*location*

San Salvador

*date*

this morning

. A police report said that the

*damage*

attack set the building on fire , but

*injury*

did not result any

casualties.

# Filled Terrorism Event Template

Date	10 January 1990
Location	El Salvador: San Salvador (city)
Event type	bombing
Weapon	“highpower bombs”
Perpetrator individual	“guerrilla urban commandos”
Perpetrator organization	-
Physical target	“car dealership”
Physical target effect	some damage
Human target	-
Human target effect	no injury or death

# Event Extraction vs. Named Entity Recognition

- *Named Entity Recognition* = identifying types of entities
- *Event Extraction* = identifying role relationships associated with events.

Paul Nelson killed John Smith.

Paul Nelson was killed by John Smith.

IBM purchased Microsoft.

IBM was purchased by Microsoft.

IBM was purchased on Tuesday by Microsoft.

# Unstructured vs. Semi-structured Text

*Unstructured text* depends 100% on language understanding.  
*Semi-structured text* has some structure (layout) that can aid in understanding.

## Unstructured Text

Professor John Skvoretz, U. of South Carolina, Columbia, will present a seminar entitled “Embedded Commitment,” on Thursday, May 4th from 4-5:30 in PH 223D.

## Semi-Structured Text

Laura Petite  
Department of Psychology  
McGill University  
  
Thursday, May 4, 1995  
12:00 pm  
Baker Hall 355

# Another Semi-Structured Seminar Announcement

Name: Dr. Jeffrey D. Hermes

Affiliation: Department of AutoImmune Diseases

Research & Biophysical Chemistry Merck Research Laboratories

Title: “MHC Class II: A Target for Specific  
Immunomodulation of the Immune Response”

Host/e-mail: Robert Murphy

Date: Wednesday, May 3, 1995

Time: 3:30 p.m.

Place: Mellon Institute Conference Room

Sponsor: MERCK RESEARCH LABORATORIES

# Supervised Learning for IE

- In the mid-1990s, researchers began to develop methods to automatically create (learn) IE systems.
- Supervised learning requires annotated training data.
- Trade-off: annotating texts vs. manual knowledge engineering
  - weeks vs. months
  - domain experts vs. computational linguists



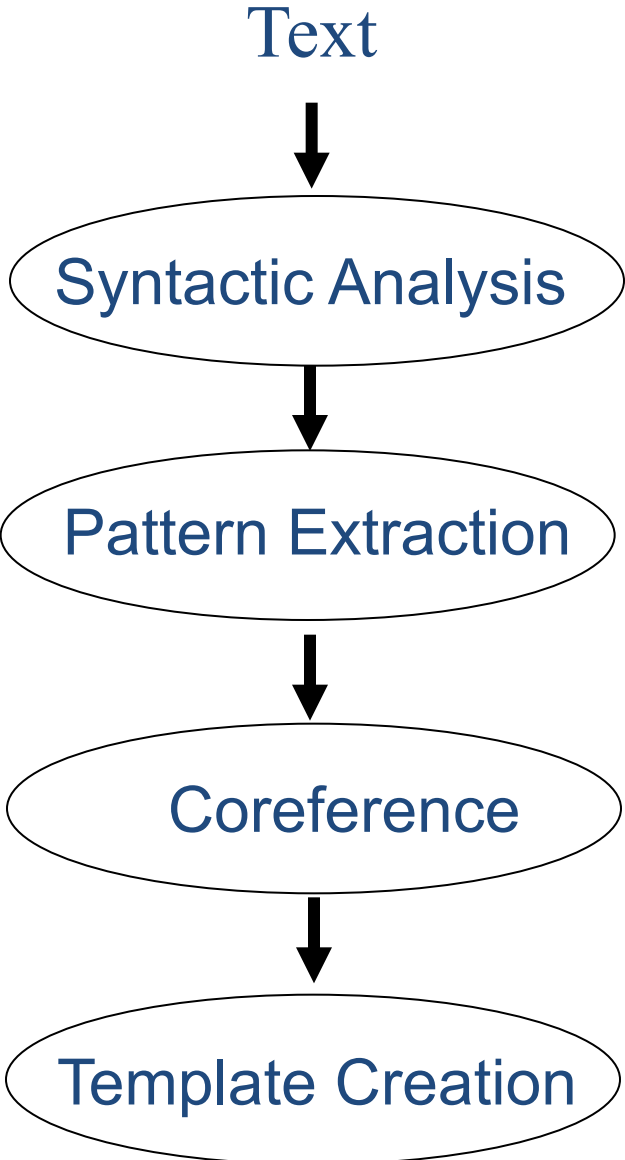
# Patterns/Rules vs. Sequence Tagging

Two general approaches to IE:

*Pattern-based systems* use patterns or rules that are applied to text.

*Sequence tagging models* classify individual tokens as to whether or not they should be extracted.

# Pattern-based Template-Filling Pipeline



IBM fired its CEO.

IBM fired its CEO.  
*Subj VP Dobj*

*Event:* FIRING  
*Agent:* IBM *Fired:* its CEO

IBM fired **its CEO**.  
**John Smith** was let go on Monday.

*Event:* FIRING  
*Agent:* IBM  
*Fired:* John Smith, CEO  
*Date:* Monday

# IE as Sequence Tagging

- A different approach: build a classifier as a sequence tagging model.
- Each document is processed sequentially and each token is labeled as Extraction or Non-Extraction.

Ex: B (beginning), I (inside), or O (outside) tags.

- Features are usually simple: e.g., words, POS tags, orthography, and a small context window of preceding/following words.

# Sequence Tagging Example

*perpetrator*

Alleged guerrilla urban commandos launched two

B I I I O B

*weapon*

*target*

highpower bombs against a car dealership in downtown

I I O B I I O O

*location*

*date*

San Salvador this morning .

B I B I

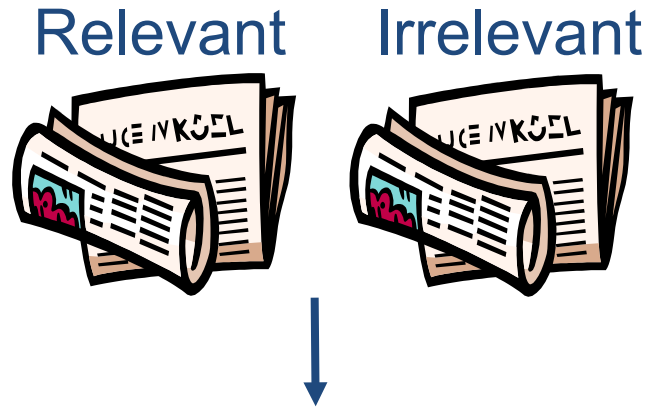
# The Perils of Manual Text Annotation

- Time consuming
- Tedious
- Deceptively tricky
- A new corpus must be annotated for each domain!

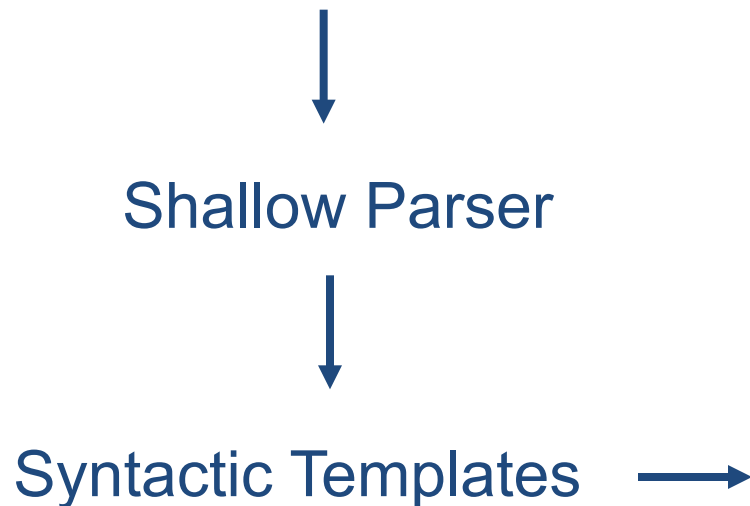
# Weakly Supervised Learning for IE

- Idea: can we train an IE system using only unannotated texts?
- Yes, if we have “preclassified” texts:
  - One pile of relevant texts
  - One pile of irrelevant texts
  - Manual review of ranked patterns
- Much easier than annotating texts!

# AutoSlog-TS [Riloff 96] (Step 1)



[The World Trade Center], [an icon] of [New York City],  
was horrifically attacked on [an otherwise beautiful day]  
in [September 2001] by [Al Qaeda].



Extraction Patterns:  
<subj> was attacked  
icon of <np>  
was attacked on <np>  
was attacked in <np>  
was attacked by <np>

# AutoSlog-TS (Step 2)

Relevant



Irrelevant



Extraction Patterns:

<subj> was attacked

icon of <np>

was attacked on <np>

was attacked in <np>

was attacked by <np>

<u>Extraction Patterns</u>	<u>Freq</u>	<u>Prob</u>
<subj> was attacked	100	.90
icon of <np>	5	.20
was attacked on <np>	80	.79
was attacked in <np>	85	.87
was attacked by <np>	95	.95



# Top Terrorism Extraction Patterns

1. <subject> exploded
2. murder of <np>
3. assassination of <np>
4. <subject> was killed
5. <subject> was kidnapped
6. attack on <np>
7. <subject> was injured
8. exploded in <np>
9. death of <np>
10. <subject> took\_place
11. caused <dobj>
12. claimed <dobj>
13. <subject> was wounded
14. <subject> occurred
15. <subject> was located
16. took\_place on <np>
17. responsibility for <np>
18. occurred on <np>
19. was wounded in <np>
20. destroyed <dobj>
21. <subject> was murdered
22. one of <np>
23. <subject> kidnapped
24. exploded on <np>
25. <subject> died

# Examples of Learned Disease Patterns

outbreak of <np>

<subj> spread

cases of <np>

<subj> was confirmed

outbreaks of <np>

<subj> was transmitted

contracted <dobj>

spread of <np>

<subj> infected

<subj> killed

# Event Keywords

Keywords alone are not as reliable as you might think due to ambiguity, metaphor, and context.

*The comedian bombed at the club ...*

*Parliament exploded in anger about ...*

*Obama was attacked by House Republicans ...*

# Secondary Contexts

*A terrorist arrested by the Salvadoran national police, has been identified as **Ruth Esperanza Aguilar Marroquin**.*

***Oqueli's** body was found next to the body of Guatemalan politician **Gilda Flores**.*

*There were **seven children**, including **four of the Vice President's children**, in the home at the time.*

# Challenges for the Future

- Contextual Effects

*The man took the money and fled.*

→ robbery

*Abilio Diniz is in the hands of a group presumed to be Chilean terrorists.*

→ kidnapping

# Challenges for the Future

- Inference

*He was shot.*

*He was shot to death.*

*He was riddled with machine gun fire.*

*He was shot. His body was found yesterday.*

- Metaphor

*killing two birds with one stone*

*shot in the arm; shot in the dark*