Homeland Security

CERATOPS

Center for Extraction and Summarization of Events and Opinions in Text

Janyce Wiebe, U. Pittsburgh

Claire Cardie, Cornell U.

Ellen Riloff, U. Utah

Overview

- Text analysis to support a broad range of knowledge discovery tasks
- Automatic annotators that assign semantic and conceptual labels to words, phrases, and documents
- Automatically extracting, summarizing and tracking information about events and opinions

Technological Goals

Rapidly re-trainable, robust components for:

1. Information extraction of facts and entities related to events from text
2. Extraction of opinions and motivations expressed in text
3. Tracking, linking, and summarizing events and opinions and their progressions over time

Information Extraction for Events

Extracting facts and entity relations associated with events of interest.

Terrorist incidents: perpetrators, victims, physical targets, weapons, date, location

Disease outbreaks: disease, organisms, victim, symptoms, location, country, date, containment measures

Keywords and named entity recognition are not sufficient.

Troops were vaccinated against anthrax, cholera, …

Researchers have discovered how anthrax toxin destroys cells and rapidly causes death ...

Fact Extraction: Example

After a brief lull, the avian flu is on the march again through Fraser Valley poultry farms.

The Canadian Food Inspection Agency says ongoing surveillance efforts have led to the detection of bird flu on 36 commercial premises.

The agency says it is continuing depopulation efforts on infected farms on a priority basis.

OUTBREAK
Disease: avian flu / bird flu
Victims: poultry
Location: Fraser Valley poultry farms / 36 commercial premises
Country: Canada
Status: confirmed
Containment: depopulation

Text

3 chickens died from avian flu.

Syntactic Analysis

3 chickens died from avian flu.
SUBJ *VP* *PP*

Semantic Extraction

Fact: DEATH
Victim: 3 chickens **Disease:** avian flu

Coreference Resolution

3 chickens died from avian flu.
The birds were found in Canada.

Relation and Event Analysis

Event: Outbreak
Victim: 3 chickens / the birds
Disease: bird flu
Country: Canada

Motivation for Event IE Systems

- Rapid semantic processing of large volumes of unstructured text
- Automatic merging of facts and entity relationships across sets of documents
- Automatic population of large databases with factual information from many text sources

Extraction of Opinions

- Knowledge discovery involves not only facts, but also opinions, sentiments, and motivations

Subjectivity: opinions, motivations, sentiments

- Information Extraction of
  - NL expressions
  - Components
  - Properties

Angolans are terrified of the Marburg virus

Source **Attitude** **Target**

Negative Emotion **Intensity: High**

Direct subjective annotation
Source: **Angolans**
Polarity: **negative**
Attitude: **emotion**
Intensity: **high**
Target: **Marburg virus**
●●●

Distinguish Objective from Subjective Language

- Speculations, hyperbole, emotions
  - The Chilean president admitted that “the US had been attacked” because “their initial measure was hasty. It amounted to using a tank to kill a flea.”
  - Alverenga accused the FMLN of being directly responsible for the attack

Reflect ***argument credibility***

Cause ***false hits*** *in IE systems*

Extract Components and Properties

Direct subjective annotation
Source:
Polarity: negative
Attitude:
Intensity: high
Target:
●●●

E.g., are people extremely afraid?

Extract Components and Properties

Direct subjective annotation
Source:
Polarity:
Attitude:
Intensity:
Target:
●●●

The industry is scared and so, even if they do find an ornamental carp with KHV, they will keep it secret

Recognize motivations

Predict actions

Extract Components and Properties

Direct subjective annotation
Source:
Polarity:
Attitude:
Intensity:
Target: Ban on British beef
•••

Brugere-Picoux backs the decision to ban British Beef

Search *for opinions* ***about*** *particular named targets*

Extract Components and Properties

Direct subjective annotation
Source: Brugere-Picoux
Polarity:
Attitude:
Intensity:
Target:
●●●

Brugere-Picoux backs the decision to ban British Beef

Search *for opinions* ***held by*** *particular named sources*

Extract Components and Properties

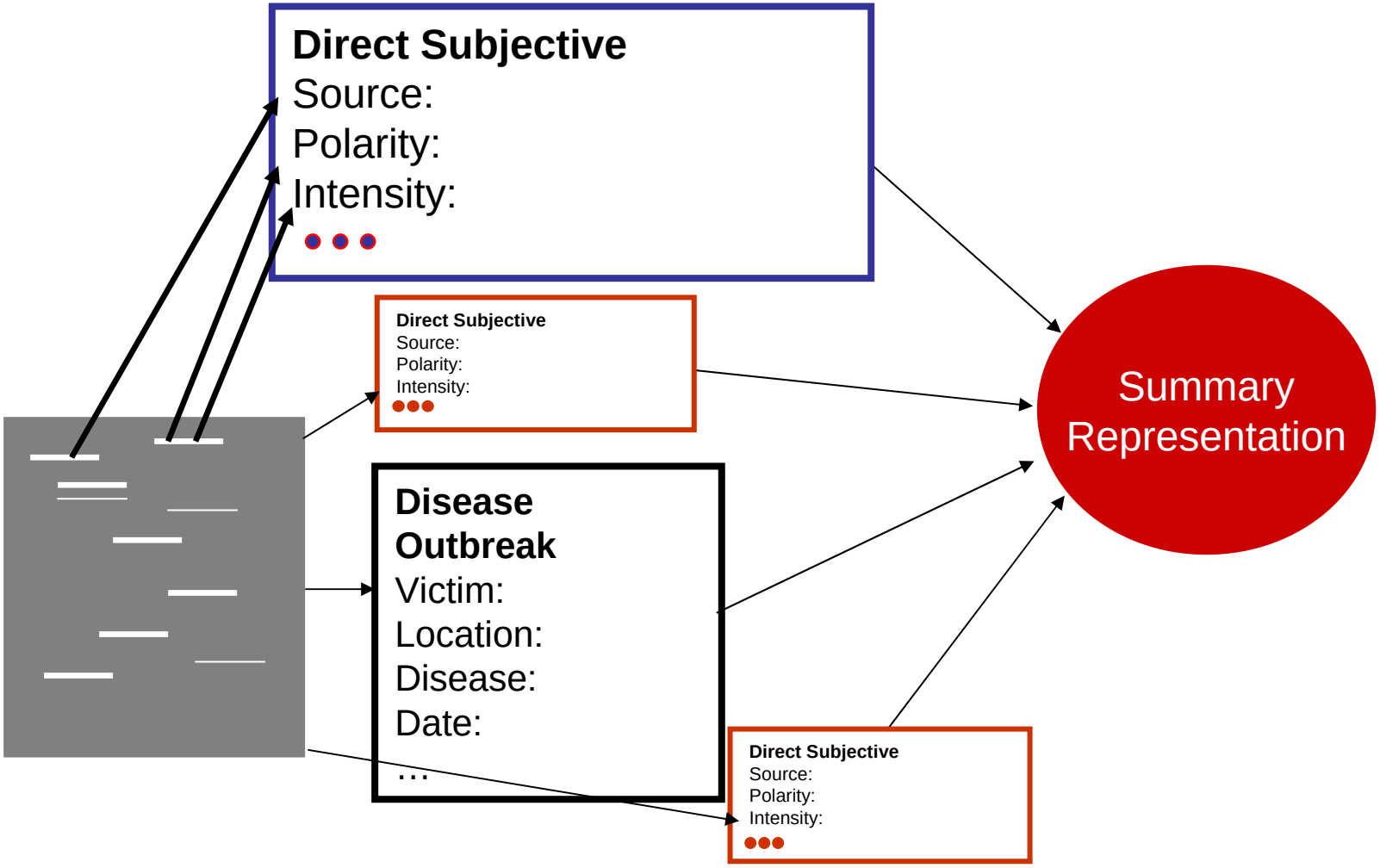
Direct subjective annotation
Source:
Polarity:
Attitude:
Intensity:
Target:
●●●

Direct subjective annotation
Source:
Polarity:
Attitude:
Intensity:
Target:
●●●

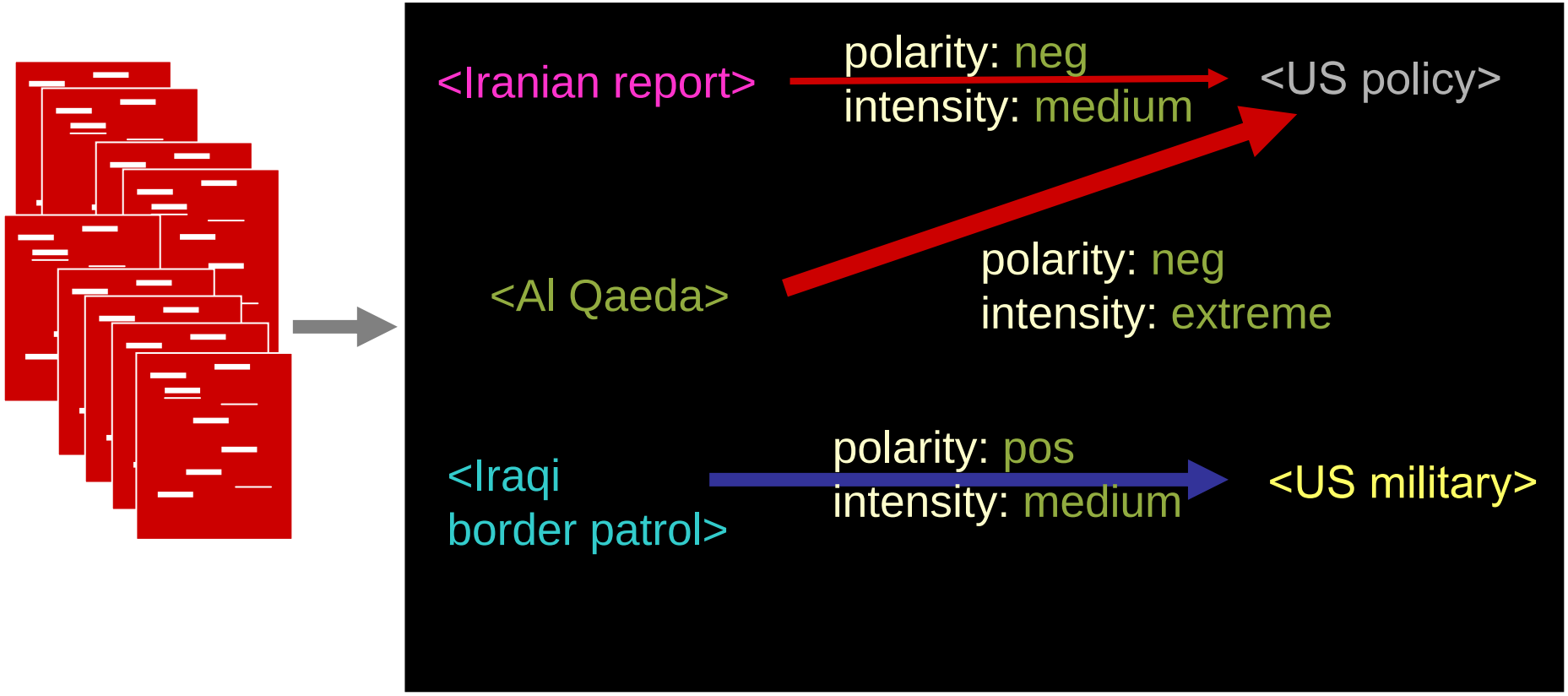
Government officials are locked in a bitter dispute with fish breeders

E.g., adversarial relationships

Summarization of Opinions + Events



Graph-Based Representation



Motivation for the Summaries

- Quickly determine the **opinions of** a person, organization, community, region, etc.
- Quickly determine the **opinions toward** a person, organization, issue, event, …
  - Across an entire document
  - Across multiple documents
  - Over time
- Reveal relationships and identify cliques and communities of interest
- Complement work in social network analysis

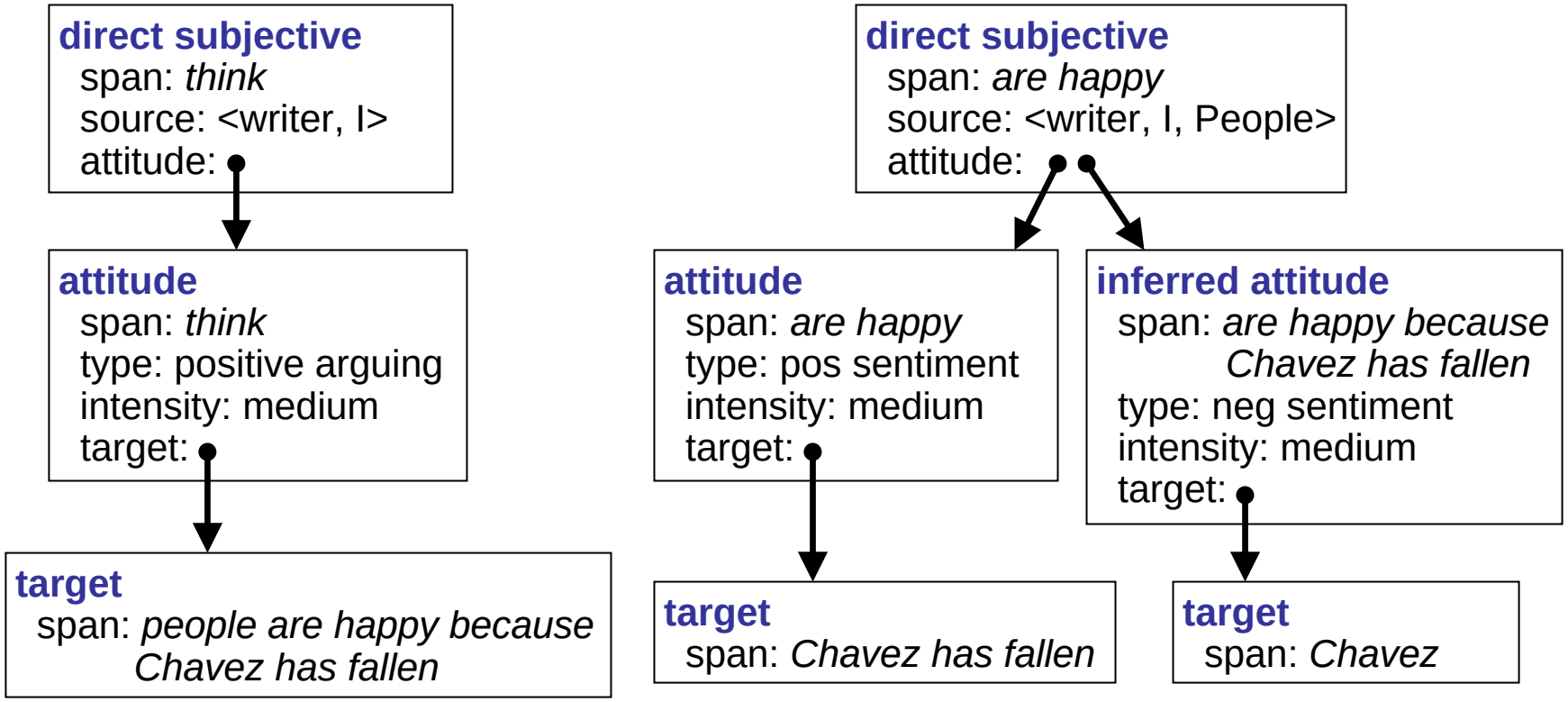
Motivation for the Summaries

- Track user queries over time
- Support the **discovery of patterns** of change or similarity
  - User hypothesizes patterns
    - E.g. *Opinion holder = known 9/11 terrorist; time-frame = sep 00 - sep 01*
  - System generates graphs w.r.t. pattern across selected time periods

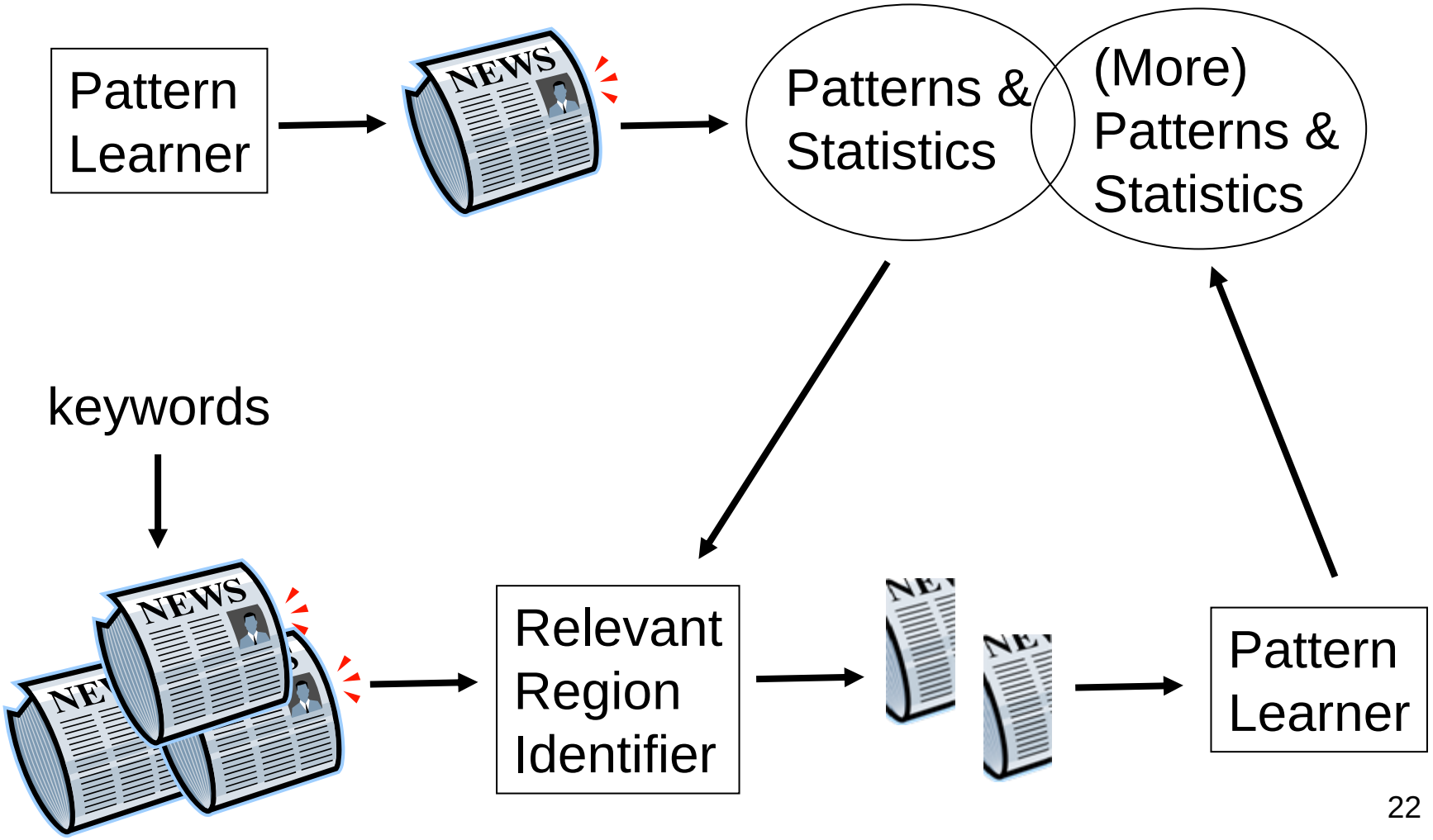
Methods and Techniques

Representation Schemes

I think people are happy because Chavez has fallen.

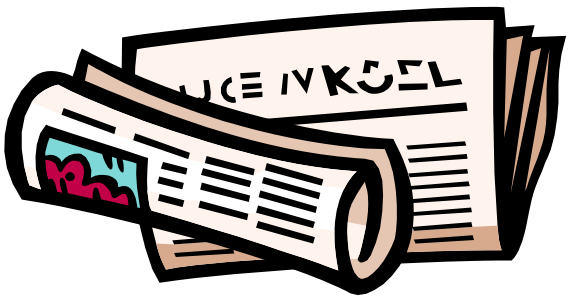


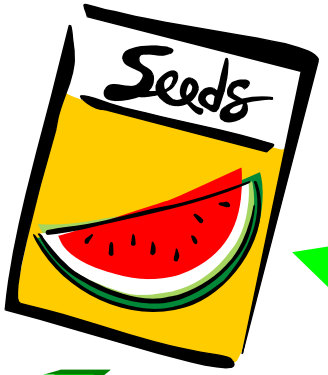
IE Pattern Bootstrapping Process



Semantic Dictionary Bootstrapping

Unannotated Texts





Ex: anthrax, ebola, cholera, flu, plague

Best Extraction Pattern(s)

Ex: outbreak of <NP>

Extractions (Nouns)

Ex: smallpox, tularemia, botulism

Semantic Learning Case Study

- Input to Basilisk: 10 common disease names
- Of the top 200 words hypothesized to be diseases: 89 were already in the UMLS metathesaurus (32,000 names of diseases and organisms), but 111 were not! Including:

adenomatosis	flu	h5n1
tularaemia	kawasaki	h7n3
tularamia	mad-cow-disease	ev71
diarrhoea	smut	yf
diphtheriae	pertussis	jyf
enterovirus-71	pleuro-pneumonia	nvcjd
fibropapillomas	polioencephalomyelitis	pepmv
gastroeneteritis	poliovirus	wsmv

Learning Subjective Phrases

- Using Information Extraction techniques

Extractions

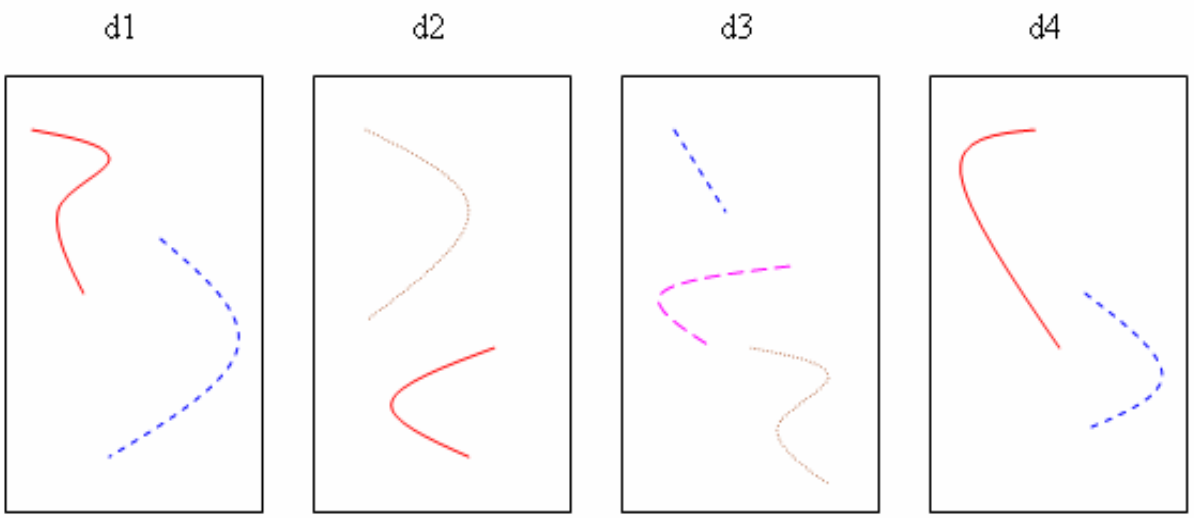
expressed <dobj>	condolences, hope, grief, views, worries
indicative of <np>	compromise, desire, thinking
inject <dobj>	vitality, hatred
reaffirmed <dobj>	resolve, position, commitment
voiced <dobj>	outrage, support, skepticism, opposition, gratitude, indignation
show of <np>	support, strength, goodwill, solidarity
<subj> was shared	anxiety, view, niceties, feeling

Subjective Expressions as IE Patterns

PATTERN	*FREQ*	*P(Subj | Pattern)*
<subj> asked	128	0.63
<subj> was asked	11	1.00
<subj> was expected	45	0.42
was expected from <np>	5	1.00
<subj> put	187	0.67
<subj> put end	10	0.90
<subj> talk	28	0.71
talk of <np>	10	0.90
<subj> is talk	5	1.00
<subj> is fact	38	1.00
fact is <dobj>	12	1.00

Coreference Resolution

- Links entities, events, and opinions within and across documents



Chain1: U.S. State Dept.

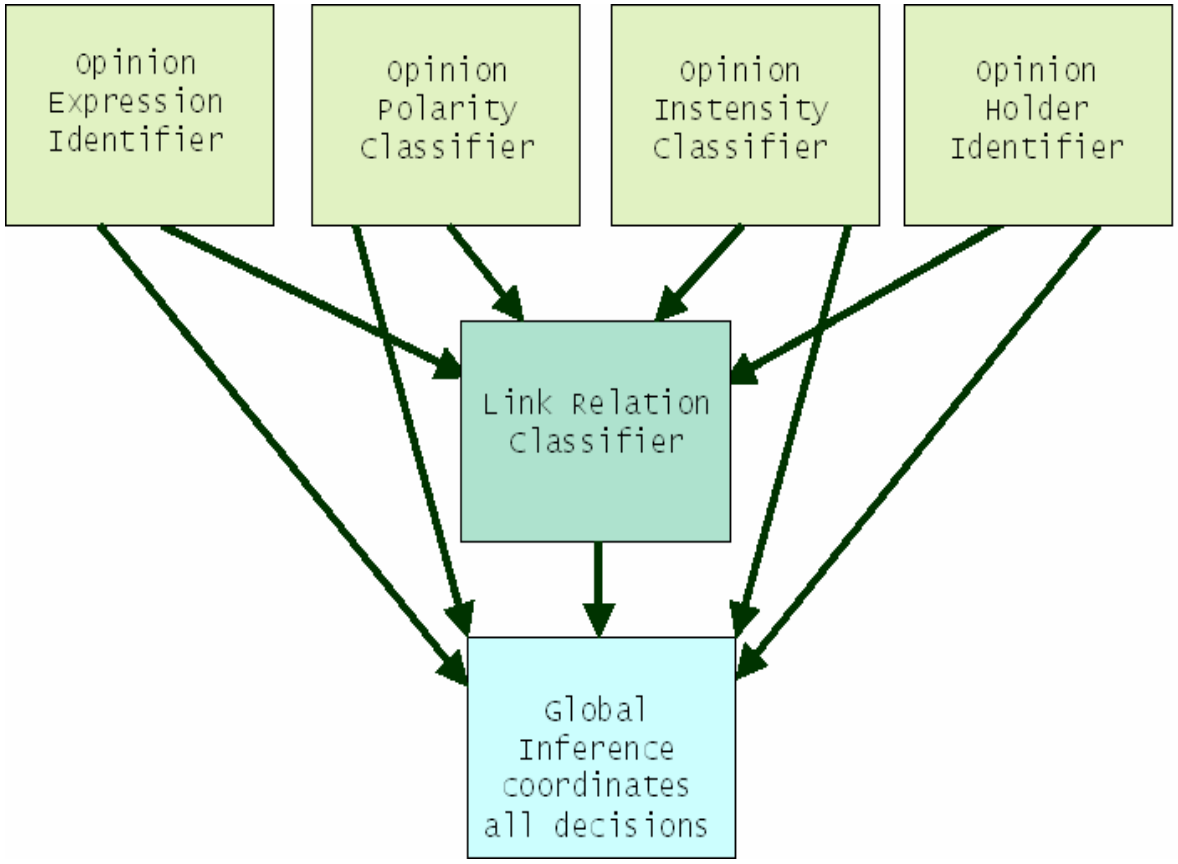
Chain2: President Bush

Chain3: NIH

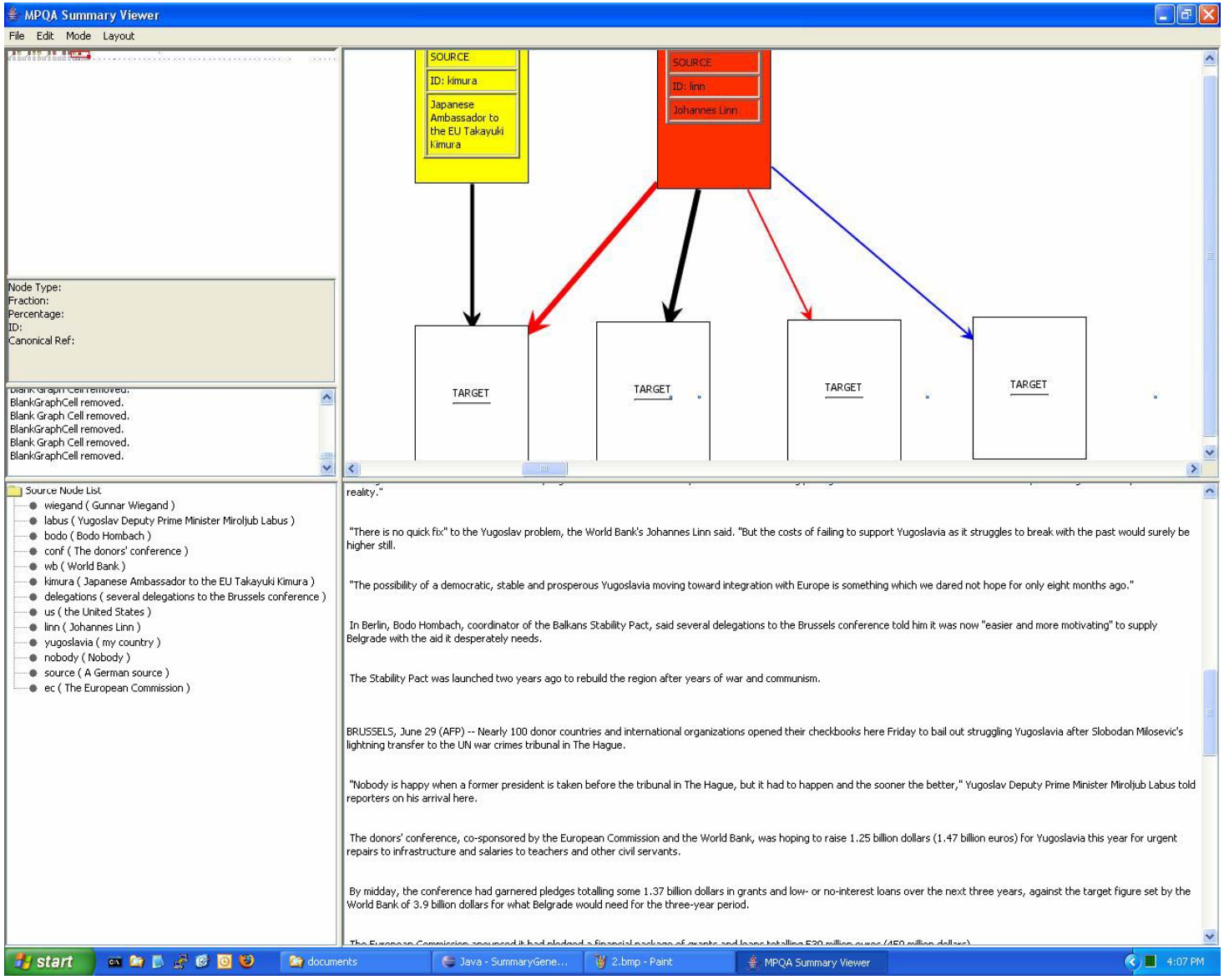
Chain4: Inspector General

New Methods for Identifying Relationships

- Jointly learn multiple components of opinion frames



Domain Experts Direct the Discovery Process



Conclusions

Rapidly re-trainable, robust components for

1. Information extraction of facts and entities
2. Extraction of opinions
3. Tracking, linking, and summarizing events and opinions and their progressions over time