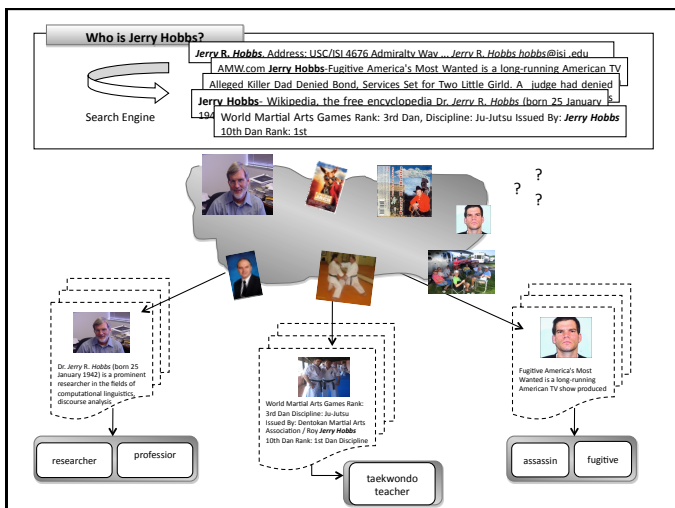


CS544: Information Extraction

January 18, 2011

Zornitsa Kozareva
USC/ISI
Marina del Rey, CA
kozareva@isi.edu
www.isi.edu/~kozareva



Named Entity Recognition and Classification

<PER>Prof. Jerry Hobbs</PER> will teach CS544 during <DATE>February</DATE>.
<PER>Jerry Hobbs</PER> killed his daughter in <LOC>Ohio</LOC>.
<ORG>Hobbs corporation</ORG> bought <ORG>FbK</ORG>.

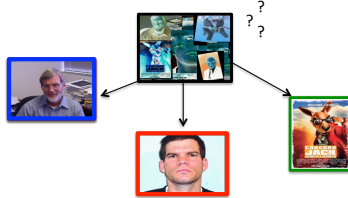
• Identify mentions in text and classify them into a predefined set of categories of interest:

- Person Names: Prof. Jerry Hobbs, Jerry Hobbs
- Organizations: Hobbs corporation, FbK
- Locations: Ohio
- Date and time expressions: February
- E-mail: mkg@gmail.com
- Web address: www.usc.edu
- Names of drugs: paracetamol
- Names of ships: Queen Mary
- ...

Named Entity Discrimination

Prof. Jerry Hobbs will teach CS544 during February.
Jerry Hobbs killed his daughter in Ohio.
Hobbs corporation bought FbK.

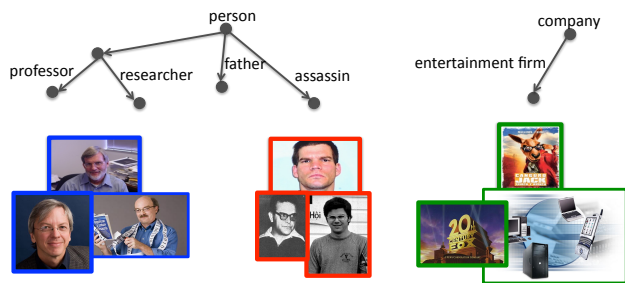
- Discover the underlying meaning of a proper name



- The number of entities and their meaning is unknown

Knowledge Harvesting

- Given the name Jerry Hobbs, learn automatically from the Web semantic classes and members similar to it



Information Extraction

What is “Information Extraction”?

- Goal: identify specific pieces of information from the content of unstructured or semi-structured textual documents.
- Input:
 - scenario of extraction (template schema to be filled)
 - document collection
- Output:
 - a set of instantiated templates

A bomb went off this morning near a power tower in San Salvador leaving a large part of the city without energy, but no casualties have been reported.

According to unofficial sources, the bomb-allegedly detonated by urban guerrilla commandos blew up a power tower in the north western part of San Salvador at 0650.

Incident type:
Date:
Time:
Location:
Perpetrator:
Physical target:
Human target:
Effect on physical target:
Effect on human target:
Instrument:

A **bomb** went off this morning near a **power tower** in **San Salvador** leaving a large part of the city without energy, but **no casualties** have been reported.











According to unofficial sources, the bomb-allegedly detonated by **urban guerrilla commandos** **blew up** a power tower in the north western part of San Salvador at **0650**.

Incident type:	bombing
Date:	January 18, 2011
Time:	0650
Location:	San Salvador (city)
Perpetrator:	urban guerrilla commandos
Physical target:	power tower
Human target:	-
Effect on physical target:	destroyed
Effect on human target:	no injury or death
Instrument:	bomb

MUC

- Message Understanding Conference (MUC) was an annual competition for IE systems funded by DARPA
 - MUC-1 (1987) and MUC-2 (1989)
 - Messages about **naval operations**
 - MUC-3 (1991) and MUC-4 (1992)
 - News articles about **terrorist attacks**
 - MUC-5 (1993)
 - News articles about **joint ventures** and **microelectronics**
 - MUC-6 (1995)
 - News articles about **management changes**
 - MUC-7 (1997)
 - News articles about **space vehicle** and **missile launches**

Today

Google squared						
http://www.google.com/squared						
Item Name	Image	Description	Date Of Birth	Nationality	Date Of Death	Died
Franz Boas		Franz Boas (July 9, 1858 – December 21, 1942) was a German American anthropologist and a pioneer of modern linguistics.	July 9, 1858	American	December 21, 1942	22-Dec-1942
Edward Sapir		Edward Sapir (pronounced /sa-pair/), (January 26, 1884 – February 4, 1939) was a prominent American linguist.	January 26, 1884	American	February 4, 1939	4-Feb-1939
Leonard Bloomfield		Leonard Bloomfield (April 1, 1887 – April 18, 1949) was an American linguist who led the development of structural linguistics in the United States.	April 1, 1887	American	April 18, 1949	April 18, 1949
Jonathan Edwards		David Brainerd – Jonathan Edwards William Tyndale: A biography – David Daniell.	October 5, 1703	American	March 22, 1758	March 22, 1758
Benjamin Lee Whorf		BENJAMIN LEWIS FORT LUTHER PROJECT: Benjamin Lee Whorf (April 24, 1897 in Winthrop, Massachusetts – July 26, 1941) was an American linguist.	April 24, 1897	American	July 26, 1941	July 26, 1941
Donna Jo Napoli		Donna Jo Napoli writes for all ages, from picture books through young adult books. Her awards include: 1999 American Library Association Award for Best Children's Book.	02/28/1948	American		
Joseph Greenberg		Joseph Harold Greenberg (May 28, 1915 – May 7, 2001) was a prominent and controversial American linguist and anthropologist.	May 28, 1915	American	May 7, 2001	May 7, 2001
Mark Baker		Parker W&S District: Mark Baker, OMI, Inc. City of Rio Rancho, 2003. Kenny Lewis Fort Lupton Project.	Jun 17, 1954	United States		1 possible value
Samuel Martin		Martin, Samuel (1976). A Reference Grammar of Japanese. Yale University Press. 47. Nadathur, Chinnappa, and Joseph, Aravind M.	1714	2 possible values	4 possible values	4 possible values
Hans Kurath		Hans Kurath (13 December 1891–2 January 1962) was an American linguist of Austrian origin.	1891-12-13	3 possible values	2 possible values	2 possible values

Today

factual						
http://www.factual.com/						
Video Games & Cheats						
Submitted by factual, 10 October 11, 2009, more						
Video Game	Platform	Publisher	Developer	Genre	Theme	
3 on 3 NHL Arcade	Xbox 360	EA Sports	EA Canada	Sports	Hockey	
4x4 Evo 2	Xbox	2K Games	Terminal Reality, Inc.	Sports; Driving/Racing		
24: The Game	PlayStation 2	2K Games	SCE Studio Cambridge	Action	Espionage	
25 to Life	PlayStation 2	Eidos Interactive	Avalanche Software LLC; Ritt	Action; Shooter	Crime	
50 Cent: Bulletproof	Xbox	Vivendi Universal	Genuine Games Ltd.	Action; Shooter	Crime	
1942: Joint Strike	Xbox 360	Backbone Entertainment		Coin-Op Classics		
Abuse	PC	Alive Mediasoft; Bungie Studi	Crack dot Com	Action; Shooter	Sci-Fi; Horror	
AC/DC Live: Rock Band Track Pac	Xbox 360	Electronic Arts				
Ace Combat 5: The Unsung War	PlayStation 2	Namco	Namco Bandai Games Inc.	Action; Simulation; Flight Sim	Modern Military	
Ace Combat X: Skies of Deception	PlayStation Portable	Namco Games	Namco Bandai Games Inc.	Action; Simulation; Flight Sim	Modern Military	
Ace Combat Zero: The Belkan War	PlayStation 2	Namco	Namco Bandai Games Inc.	Action; Simulation; Flight Sim	Modern Military	
Activision Anthology	PlayStation 2	Activision	Activision	Action; Strategy; Sports; Driv		
Act of War: Direct Action	PC	GFI Russia; Atari, Inc.	Eugen Systems	Strategy; Real-Time Strategy	Modern Military	
Advance Guardian Heroes	Game Boy Advance	TUBI Soft	Treasure	Action; Fighting		
Advance Wars: Days of Ruin	Nintendo DS	Nintendo	Intelligent Systems Co., Ltd.	Strategy	Post-Apocalyp	
Advance Wars: Dual Strike	Nintendo DS	Nintendo	Intelligent Systems Co., Ltd.	Strategy	Modern Military	
Advent Rising	PC	Majesco Entertainment	GlyphX Games, LLC	Action; Shooter; Adventure	Sci-Fi	
Advent Rising	Xbox	Majesco Entertainment	GlyphX Games, LLC	Action; Shooter; Adventure	Sci-Fi	
Aegis Wing	Xbox 360	Microsoft Game Studios				
Aeon Flux	Xbox	Majesco Sales Inc.				
Afro Samurai	Xbox 360	Namco	Namco Bandai Games Inc.	Platformer; Action	Martial Arts; An	
Age of Booty	Xbox 360	Certain Affinity				
Age of Empires III	PC	Microsoft Game Studios; Mac	Ensemble Studios	Strategy; Real-Time Strategy	Alternate Histor	
Age of Empires: The Age of Kings	Nintendo DS	Majesco Sales Inc.	Backbone Entertainment	Strategy	Fantasy	

Other Applications

- Job postings
- Seminar announcements
- Conference call for papers
- Company information
- Apartment rental adds
- Social event announcements
- USC alert system
- ...

Example is adapted from Raymond Mooney

Subject: US-TN-SOFTWARE PROGRAMMER
Date: 17 Nov 2009 17:37:29 GMT
Organization: Reference.Com Posting Service
Message-ID: <56nigp\$mrs@bilbo.reference.com>

SOFTWARE PROGRAMMER

Position available for Software Programmer experienced in generating software for PC-Based Voice Mail systems. Experienced in C Programming. Must be familiar with communicating with and controlling voice cards; preferable Dialogic, however, experience with others such as Rhetorix and Natural Microsystems is okay.

Prefer 5 years or more experience with PC Based Voice Mail, but will consider as little as 2 years. Need to find a Senior level person who can come on board and pick up code with very little training. Present Operating System is DOS. May go to OS-2 or UNIX in future.

Please reply to:
Kim Anderson
AdNET
(901) 458-2888 fax
kimander@memphisonline.com

computer_science_job_template

id:
title:
salary:
company:
recruiter:
state:
city:
country:
language:
platform:
application:
area:
req_years_experience:
desired_years_experience:
req_degree:
desired_degree:
post_date:

Example is adapted from Raymond Mooney

Subject: **US-TN-SOFTWARE PROGRAMMER**
Date: **17 Nov 2009** 17:37:29 GMT
Organization: Reference.Com Posting Service
Message-ID: <56nigp\$mrs@bilbo.reference.com>

SOFTWARE PROGRAMMER

Position available for Software Programmer experienced in generating software for PC-Based **Voice Mail** systems. Experienced in **C** Programming. Must be familiar with communicating with and controlling voice cards; preferable Dialogic, however, experience with others such as Rhetorix and Natural Microsystems is okay.

Prefer **5** years or more experience with **PC** Based **Voice Mail**, but will consider as little as **2** years. Need to find a Senior level person who can come on board and pick up code with very little training. Present Operating System is **DOS**. May go to **OS-2** or **UNIX** in future.

Please reply to:
Kim Anderson
AdNET
(901) 458-2888 fax
kimander@memphisonline.com

computer_science_job_template

id: **56nigp\$mrs@bilbo.reference.com**
title: **SOFTWARE PROGRAMMER**
salary:
company:
recruiter:
state: **TN**
city:
country: **US**
language: **C**
platform: **PC \ DOS \ OS-2 \ UNIX**
application:
area: **Voice Mail**
req_years_experience: **2**
desired_years_experience: **5**
req_degree:
desired_degree:
post_date: **17 Nov 2009**

Two general approaches to IE

Pattern-Based Systems



Use patterns and rules which are applied to text.

Machine Learning



Use sequence tagging models to classify individual tokens as to whether or not they should be extracted.

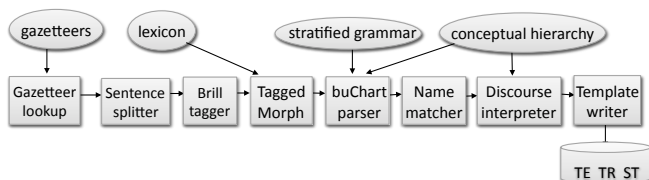
15

A **bomb** went off this morning near a **power tower** in **San Salvador** leaving a large part of the city without energy, but **no casualties** have been reported.

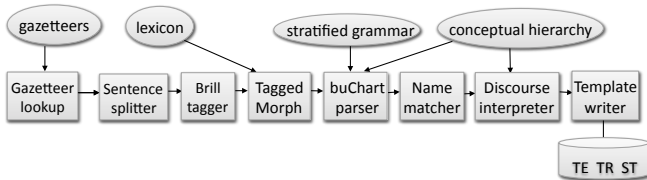
According to unofficial sources, the bomb-allegedly detonated by **urban guerrilla commandos** **blew up** a power tower in the north western part of San Salvador at **0650**.

Incident type:	bombing
Date:	January 18, 2011
Time:	0650
Location:	San Salvador (city)
Perpetrator:	urban guerrilla commandos
Physical target:	power tower
Human target:	-
Effect on physical target:	destroyed
Effect on human target:	no injury or death
Instrument:	bomb

LaSIE Information Extraction System



LaSIE Information Extraction System

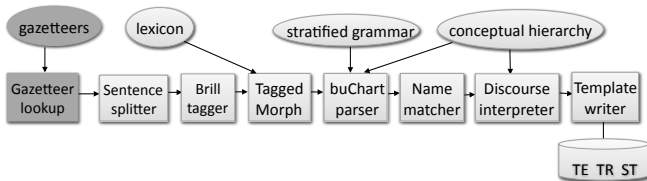


Tokenization - identify word boundaries in text

- white spaces indicate token boundaries
- full stops indicate sentences boundaries

(not always true for example, **1. September; Nov. 1998**)

LaSIE Information Extraction System

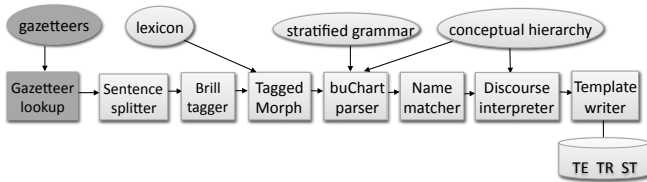


Gazetteer Lookup – recognize phrases and keywords related to named entities which were previously stored in its lists (gazetteers)

A **bomb** went off this morning near a **power tower** in **San Salvador** leaving a large part of the city without energy , but **no casualties** have been reported .

According to unofficial sources , the bomb-allegedly detonated by **urban guerrilla commandos** **blew up** a power tower in the north western part of **San Salvador** at 0650 .

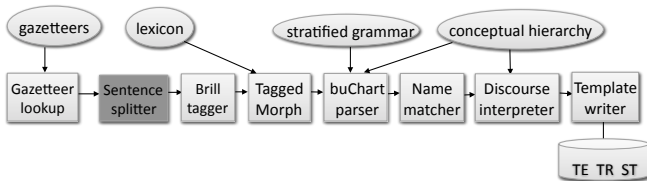
LaSIE Information Extraction System



Gazetteer Lookup – recognize phrases and keywords related to named entities which were previously stored in its lists (gazetteers)

- advantage – simple, fast, language independent as one just has to create the lists
- disadvantage – collection and maintenance is time consuming, cannot deal with name variants, cannot resolve ambiguity

LaSIE Information Extraction System



Sentence splitter - given a text, returns a list of strings where each element is a sentence.

- uses a set of rules like the occurrences of “.”, “?” and “!” are indicators of sentence delimiters

(not so simple, the “.” in “B. Clinton” or “U.S.” does not have this role)

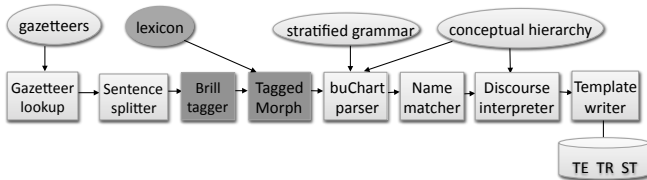
Sentence1:

A **bomb** went off this morning near a **power tower** in **San Salvador** leaving a large part of the city without energy , but **no casualties** have been reported .

Sentence2:

According to unofficial sources, the bomb-allegedly detonated by **urban guerrilla commandos blew up** a power tower in the north western part of **San Salvador** at **0650**.

LaSIE Information Extraction System



Part-of-speech tagging – identify and mark up the words in a text with the corresponding part of speech such as noun, verb, adjective, adverb etc.

Why do we care?

According_to-adv unofficial-adj source[s]-n , the-det bomb-n allegedly-adv detonate[ed]-v by-prep urban-adj guerrilla-n commando[s]-n blow_up-v a-det power_tower-n in-prep the-det northwestern-adj part-n of-prep San Salvador-loc at-prep 0650-time

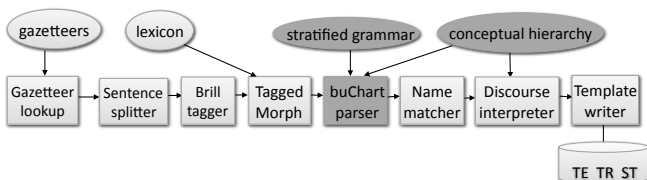
Sentence1:

A **bomb** went off this morning near a **power tower** in **San Salvador** leaving a large part of the city without energy , but **no casualties** have been reported .

Sentence2:

According to unofficial sources, the bomb-allegedly detonated by **urban guerrilla commandos** **blew up** a power tower in the north western part of **San Salvador** at **0650** .

LaSIE Information Extraction System

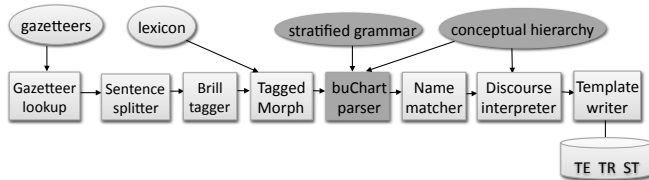


Syntactico-semantic interpretation

- bottom-up chart parser
- cascade of NERC grammars (eg. aircraft, person, money, time)

According_to-adv unofficial-adj source[s]-n , the-det bomb-n allegedly-adv detonate[ed]-v by-prep urban-adj guerrilla-n commando[s]-n blow_up-v a-det power_tower-n in-prep the-det **northwestern part of** San Salvador-loc at-prep **0650-time**
NE2 NE1

LaSIE Information Extraction System

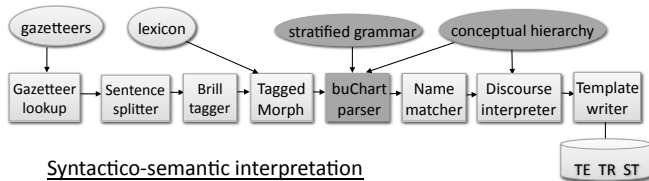


Syntactico-semantic interpretation

- cascade of partial grammars (NPs, PPs, complex NP, VPs, complex VPs, RelClauses, Sentence)

S(According_to-adv NP(unofficial-adj source[s]-n) , NP(the-det bomb-n) allegedly-adv VP(detonate[ed]-v) PP(by-prep NP(urban-adj guerrilla-n commando[s]-n)) - VP(blew_up-v) NP(a-det power_tower-n) PP(in-prep NP (the-det **NE1-loc**) PP(at-prep NP(**NE2-time**)))

LaSIE Information Extraction System



Syntactico-semantic interpretation

- bottom-up chart parser
- cascade of NERC grammars (eg. aircraft, person, money, time)
- cascade of partial grammars (NPs, PPs, complex NP, VPs, complex VPs, RelClauses, Sentence)
- logic form

Event(E1), detonate(E1,Y,X), urban_guerrilla_comando(X), bomb(Y)

Event(E2), blow_up(E2,Y,Z), power_tower(Z), location_of(Z,**NE1**), time_of(E2,**NE2**)

According to unofficial sources, the bomb-allegedly detonated by **urban guerrilla commandos** **blew up** a power tower in the **north western part of San Salvador** at **0650**.

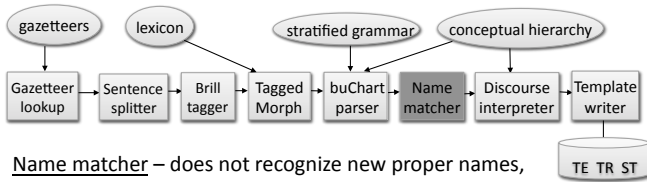
NE1

NE2

Event(E1), detonate(E1,Y,X), urban_guerrilla_comando(X), bomb(Y)

Event(E2), blow_up(E2,Y,Z), power_tower(Z), location_of(Z,**NE1**), time_of(E2,**NE2**)

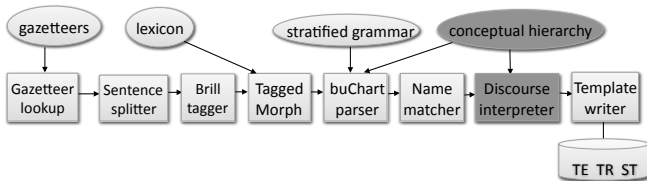
LaSIE Information Extraction System



Name matcher – does not recognize new proper names, just adds identity relations between those found by the parser

- first token of the name matches the second name
“Pepsi Cola” equals “Pepsi”
- one of the names is an acronym of the other
“ISI” is equivalent to “Information Sciences Institute”
- one name is a reversal of the other
“Defense Department” equals “Department of Defense”
- one name consists of concatenated contractions of the other
“Pan America” equals “Pan Am”

LaSIE Information Extraction System



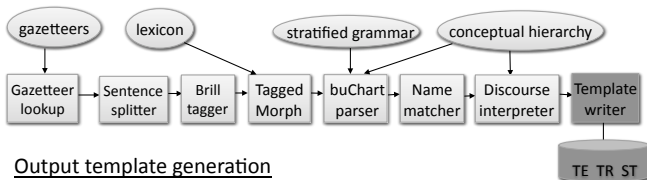
Discourse interpreter – translates the semantic representations produced by the parser into

- representation of instances, their ontological classes and attributes
- coreference resolution

Event(E1, detonate(E1,Y,X), urban_guerrilla_comando(X), bomb(Y),
 Event(E2, blow_up(E2,Y,Z), power_tower(Z), location_of(Z,NE1), time_of(E2,NE2))

implies bombing event
 isa destroy
 implies location of event

LaSIE Information Extraction System



Output template generation

- procedure that writes the templates in the desired format

Incident type:	bombing
Date:	March 11, 2010
Time:	0650
Location:	San Salvador (city)
Perpetrator:	urban guerrilla commandos
Physical target:	power tower
Human target:	-
Effect on physical target:	destroyed
Effect on human target:	no injury or death
Instrument:	bomb

How well does this work¹?

- Evaluate system's performance on independent manually-annotated test data which was not used during system development
- IE systems are typically evaluated in terms of Precision (P) and Recall (R)

$$P = \frac{\text{correctly extracted facts}}{\text{extracted facts}}$$

$$R = \frac{\text{correctly extracted facts}}{\text{correct facts}}$$

$$F_1 = \frac{2PR}{P+R}$$

¹http://www.nlpir.nist.gov/related_projects/muc/proceedings/st_score_report.html

LaSIE in MUC-6

Task	Precision	Recall
Named Entity	.94	.84
Co-reference resolution	.71	.51
Template Elements	.74	.66
Scenario Templates	.73	.37

LaSIE Named Entity

- Results for the Named Entity task over 30 texts
- Each setting indicates the contribution of LaSIE's components

No.	Setting	Precision	Recall
1	Gazetteer Look Up	.74	.37
2	1 + Parsing	.93	.80
3	2 + Name matching	.93	.88
4	3 + Discourse interpretation	.93	.89

Can I test an existing IE system?



— General Architecture for Text Engineering

ANNIE Demo

<http://services.gate.ac.uk/annie/index.jsp>

ANNIE is one of many Information Extraction systems that have been developed using GATE. It uses finite state algorithms and the JAPE language. This demo shows ANNIE recognising entities in texts.

Note: this demo uses a default set of components and IE resources; your mileage may vary! Also, complex HTML structures may prevent the system from being able to analyse the text they contain. The system does name recognition; see the [IE User Guide](#) for details of other forms of IE, and issues of domain-specificity and porting. [Contact us](#) about our [cross-domain](#), [multi-genre](#) systems.

To use ANNIE, enter a URL in the box below. Select the types of entities that you would like to mark. GATE will then retrieve the document and extract the required information. This process may take a few seconds.

Enter a URL:

- ☒ Person
- ☒ Location
- ☒ Organization
- ☒ Date
- ☒ Address
- ☒ Money
- ☒ Percent

Run ANNIE

California Prius incident probed; GM offers criticized

BY JEFFREY HAYES
FREE PRESS WASHINGTON STAFF

Comments (3) | Recommend (2) | 0 | Print | E-mail | Letter to the editor | Share |

WASHINGTON — As Toyota sought to contain the fallout from a California sudden-acceleration case caught on camera, its dealers accused General Motors of offering predatory incentives using federal money.

The Japanese automaker and the National Highway Traffic Safety Administration dispatched investigators to San Diego to analyze the 2008 Toyota Prius belonging to James Sikes, 61. Sikes called the California Highway Patrol on Monday evening reporting that his Prius was accelerating out of his control, hitting speeds of up to 94 m.p.h.

"I pushed the gas pedal to pass a car and it did something kind of funny," Sikes said at a news conference. "It jumped and it just stuck there."

An officer pulled alongside the Prius, and over a loudspeaker told Sikes to pull the emergency brake and press the regular brake hard. Sikes was able to slow the car and shut it off after 20 minutes.

The incident happened a few miles from the site of the Santee crash last August that spurred Toyota to recall 4.4 million vehicles for mechanical defects that could lead to sudden acceleration — including Sikes Prius, which was covered by the floor-mat recall.

But Sikes said he had taken his Prius to his Toyota dealer and was told it wasn't covered. Toyota said in a statement that the repairs for the Prius had not yet been sent to dealers. The automaker had told owners to remove the driver's-side floor mats until the repairs could be made, but Sikes had left the floor mat in his vehicle.

Toyota reiterated Tuesday that the Prius was under recall. It had said last year that it could take months for all of the recalled vehicles to be fixed.

In another sign of the pressure facing Toyota, its national dealer panel accused GM of using "taxpayer dollars to fund ... a nationwide predatory advertising campaign." Shortly after Toyota began its recalls last month, GM began incentives for Toyota owners that now include zero-percent financing and up to \$1,000 cash back.

"It is outrageous that GM is using our taxpayer dollars against us, making me and other Toyota dealers pay to undermine our own businesses," said Paul Atkinson, president of Toyota's U.S. dealer council.

GM's move was matched by other automakers. Last week, Toyota launched an incentive

WASHINGTON — As Toyota sought to contain the fallout from a California sudden-acceleration case caught on camera, its dealers accused General Motors of offering predatory incentives using federal money.

The Japanese automaker and the National Highway Traffic Safety Administration dispatched investigators to San Diego to analyze the 2008 Toyota Prius belonging to James Sikes, 61. Sikes called the California Highway Patrol on Monday evening reporting that his Prius was accelerating out of his control, hitting speeds of up to 94 m.p.h.

"I pushed the gas pedal to pass a car and it did something kind of funny," Sikes said at a news conference. "It jumped and it just stuck there."

An officer pulled alongside the Prius, and over a loudspeaker told Sikes to pull the emergency brake and press the regular brake hard. Sikes was able to slow the car and shut it off after 20 minutes.

The incident happened a few miles from the site of the Santee crash last August that spurred Toyota to recall 4.4 million vehicles for mechanical defects that could lead to sudden acceleration — including Sikes Prius, which was covered by the floor-mat recall.

But Sikes said he had taken his Prius to his Toyota dealer and was told it wasn't covered. Toyota said in a statement that the repairs for the Prius had not yet been sent to dealers. The automaker had told owners to remove the driver's-side floor mats until the repairs could be made, but Sikes had left the floor mat in his vehicle.

Toyota reiterated Tuesday that the Prius was under recall. It had said last year that it could take months for all of the recalled vehicles to be fixed.

In another sign of the pressure facing Toyota, its national dealer panel accused GM of using "taxpayer dollars to fund ... a nationwide predatory advertising campaign." Shortly after Toyota began its recalls last month, GM began incentives for Toyota owners that now include zero-percent financing and up to \$1,000 cash back.

"It is outrageous that GM is using our taxpayer dollars against us, making me and other Toyota dealers pay to undermine our own businesses," said Paul Atkinson, president of Toyota's U.S. dealer council.

GM's move was matched by other automakers. Last week, Toyota launched an incentive campaign of its own after a 5% drop in February sales.

"We understand why Toyota dealers would be frustrated, but they know better," said GM spokesman Kerry

Rule-based IE: Pros and Cons



PROS:

- + clearly understood technology
- + hand-written rules are relatively precise
- + people can write rules with a reasonable amount of training

CONS:

- rules need to be written by hand
- requires experienced grammar developers
- difficult to port to a different domain

39

Can we automatically learn IE?

- In the mid-1990's supervised IE systems were created.
- Supervised learning requires annotated training data.
- Trade-off: annotating texts vs. manual knowledge engineering
 - weeks vs. months
 - domain experts vs. computational linguists

Annotating Texts for IE

date
location
perpetrator
target
weapon
damage
injury

Alleged guerilla urban commandos launched
two highpower bombs against a car dealership
in downtown San Salvador this morning . A
police report said that the attack set the building
on fire but did not result any casualties .

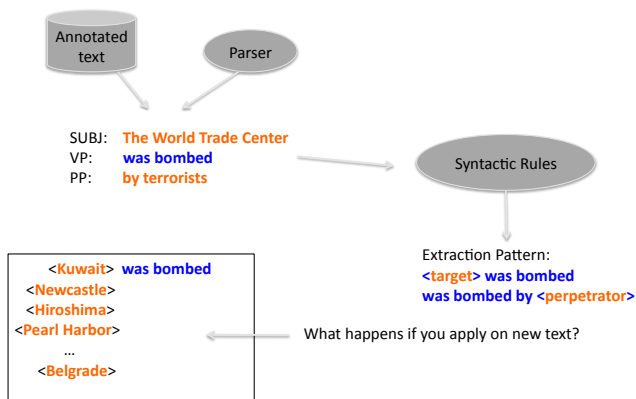
Annotating Texts for IE

Alleged ^{perpetrator} guerilla urban commandos launched
two ^{weapon} highpower bombs against a ^{target} car dealership
in downtown ^{location} San Salvador ^{date} this morning . A
police report said that the attack ^{damage} set the building
^{injury} on fire but did not result any casualties .

Pattern Learning Algorithms

- A variety of different pattern/rule representations have been developed, but very commonly:
 - IE systems learn patterns by beginning with highly specialized patterns and iteratively generalizing them.
 - rule-learning stops when a set of patterns has been generated to sufficiently “cover” the training examples.

AutoSlog [Riloff 1993]



Examples of learned patterns with AutoSlog:

<subject> passive-vp	<target> was bombed
<subject> active-vp	<perpetrator> bombed
<subject> active-vp dobj	<perpetrator> threw dynamite
<subject> active-vp infinitive	<perpetrator> tried to kill
<subject> passive-vp infinitive	<perpetrator> was hired to kill
<subject> auxiliary dobj	<victim> was fatality

active-vp <dobj>	bombed <target>
infinitive <dobj>	to kill <victim>
active-vp infinitive <dobj>	tried to kill <victim>

passive-vp infinitive <dobj>	was hired to kill <victim>
subject auxiliary <dobj>	fatality was <victim>
passive-vp prep <np>	was killed by <perpetrator>
active-vp prep <np>	exploded in <target>
infinitive prep <np>	to kill with <weapon>
noun prep <np>	assassination of <victim>
