Domain Adaptive IE: Learning Template Filling Rules

Günter Neumann & Feiyu Xu

{neumann, feiyu}@dfki.de

Language Technology-Lab DFKI, Saarbrücken

Motivations

- Porting to new domains or applications is expensive
- Current technology requires IE experts
 - Expertise difficult to find on the market
 - SME cannot afford IE experts
- Machine learning approaches
 - Domain portability is relatively straightforward
 - System expertise is not required for customization
 - "Data driven" rule acquisition ensures full coverage of examples

Problems

- Training data may not exist, and may be very expensive to acquire
- Large volume of training data may be required
- Changes to specifications may require reannotation of large quantities of training data
- Understanding and control of a domain adaptive system is not always easy for non-experts

Parameters

- Document structure
 - Free text
 - Semi-structured
 - Structured
- Richness of the annotation
 - Shallow NLP
 - Deep NLP
- Complexity of the template filling rules
 - Single slot
 - Multi slot
- Amount of data

- Degree of automation
 - Semi-automatic
 - Supervised
 - Semi-Supervised
 - Unsupervised
- Human interaction/contribution
- Evaluation/validation
 - during learning loop
 - Performance: recall and precision

Learning Methods for Template Filling Rules

- Inductive learning
- Statistical methods
- Bootstrapping techniques
- Active learning

Documents

- Unstructured (Free) Text
 - Regular sentences and paragraphs
 - Linguistic techniques, e.g., NLP
- Structured Text
 - Itemized information
 - Uniform syntactic clues, e.g., table understanding
- Semi-structured Text
 - Ungrammatical, telegraphic (e.g., missing attributes, multi-value attributes, ...)
 - Specialized programs, e.g., wrappers

"Information Extraction" From Free Text

October 14, 2002, 4:00 a.m. PT

For years, <u>Microsoft Corporation CEO</u> <u>Bill Gates</u> railed against the economic philosophy of open-source software with Orwellian fervor, denouncing its communal licensing as a "cancer" that stifled technological innovation.

Today, <u>Microsoft</u> claims to "love" the open-source concept, by which software code is made public to encourage improvement and development by outside programmers. <u>Gates</u> himself says <u>Microsoft</u> will gladly disclose its crown jewels--the coveted code behind the Windows operating system--to select customers.

"We can be open source. We love the concept of shared source," said <u>Bill</u> <u>Veghte</u>, a <u>Microsoft VP</u>. "That's a superimportant shift for us in terms of code access."

<u>Richard Stallman</u>, <u>founder</u> of the <u>Free</u> <u>Software Foundation</u>, countered saying...



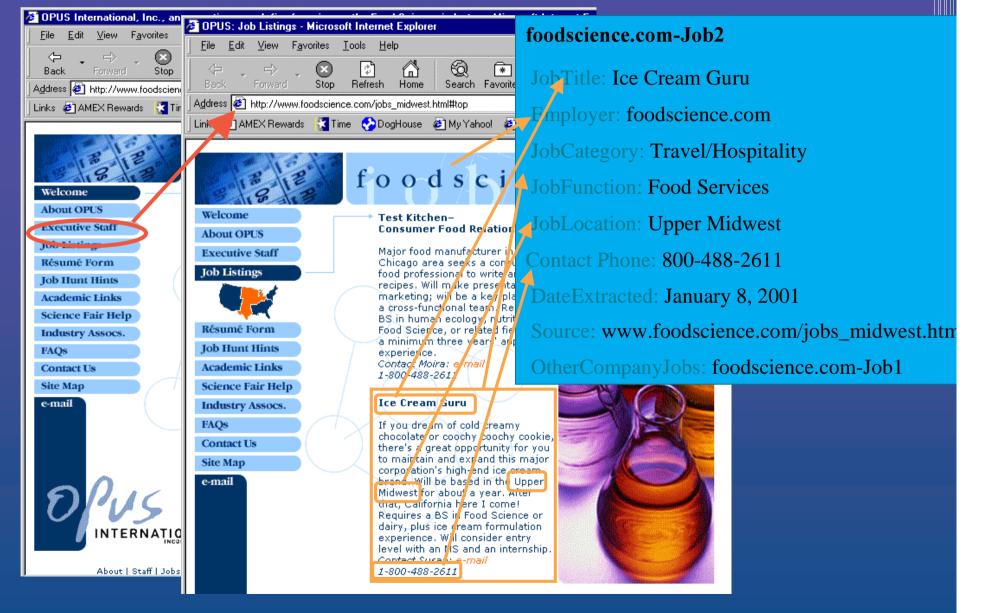
NAME	TITLE	ORGANIZATION
Bill Gates	CEO	Microsoft
BillVeghte	VP	Microsoft
RichardStallm	an founder	Free Soft

IE from Research Papers

Bed: yew Fyorites look jew Bed: + + @ @ @ @ @ Search gravenes @ instary @ @ @ @ Abetract: his provides of Commensurable Abduction Models for Semantic Interpretation (1990) (Correct) (5 citations) Pater Nong Abdert Wilensky Unkersty of California, Berkeley Computer. Thitsenth International Conference on Computational Linguistics, Volume 3 Context States I Attain and Edwards (1990), and Kg and Mooney (1990). These three models add the ingottant property of commensurability is a desirable property, and there is a clear need for a very to compete allemate explanations, a greeter doel researce is nodel measure is not enough to account for all types of processing. We present on the song action results is and continue (Index) Context and account currency that can be compared and combined White commensurability is a desirable property, and there is a clear need for a very to compete allemate explanations, a prepresent is a sing acount ensure is not enough to account for all types of processing. We present of the providens for the abduction and combined (Index) to account for all types of processing. We present of the providens of the area of the sing other measure is not enough to account for all types of processing. We present of the providens for the abduction and combined (Index) compared and combined (Index) Context algebraic measures in the sing account measure is not enough to account for all types of processing. We present of the providens for the abduction and the providens for the abduction and acounts of great results. The providence measures is not enough to account for all types of processing. We present of the providens for the abduction and the providence. The providence measures is not enough to account for all types of processing. We present of the providens for the abduction and account for all types of processing. We present and Wilensky (1990) for creaters and for all types of processing. We present and the providence of the accurestice the anomatice the anomatic	🗿 A Critical Evaluation of Commensurable Abduction Models for Semantic Interpretation - Peter, Wi - Microsoft Internet Explorer p	- O ×
Address @ Thtp://dicesser.ninec.com/peter900tical.thm Image: United States	Eile Edit View Favorites Tools Help	*
A Critical Evaluation of Commensurable Abduction Models for Semantic Interpretation (1990) (Correct) (5 cliations) Peter Norig Robert Wilensky University of California, Berkeley Computer Thatsenth Intermational Conference on Computational Linguistics, Volume 3 NCC Researchindex Bookmark Context Related Charles By EP DYU Image Update Help From: norvig com/resume (more) Home: R.Wilensky HEStarch (Correct) Astract: this paper we critically evaluate three recent shokactive interpretation models, those of Charnisk and Ooldman (1989), Hobbs, Stickel, Mattin and Edwards (1983), and Ng and Mooney (1990). These three models add the important property of commensusability: al types of sorvidence are represented in a common currency that can be compared and combined While commensurability is a desirable property, and there is a clear meed for a wey to compare alternate explanations, it appears that a single colar messure is not enough to account for all types of processing. We present other problems for the solutions update frames tensitive solutions. (Update) Context clicking spurious Interpretations of greater depters. Table 1: Empiried Results Comparing Coherence and costs as probabilities, opecifically within the context of using adductions for text Interpretations, are discussed in Norvig and Wilensky (1990). The use of solutions update in a Hybrid MT System - Gawron (1999) (Correct) Abduction and Alismatch in a Hybrid MT System - Gawron (1999) (Correct) Abduction and Alismatch in a Hybrid MT System - Gawron (1999) (Correct) Abduction and Alismatch in a Hybrid MT System - Gawron (1999) (Correct) Abduction and Alismatch in a Hybrid MT System - Gawron (1999) (Correct) Abduction and Alismatch in a Hybrid MT System - Gawron (1999) (Correct) Abduction and Alismatch in a Hybrid MT System - Gawron (1999) (Correct) Abduction and Alismatch in a Hybrid MT System - Gawron (1999) (Correct) Abduction and Alismatch in a Hybrid MT System - Gawron (1999) (Correct) Abduction and Alismatch in a Hybrid	→ Back + → - ② ③ ☆ ③Search → Favorites ③History 🖏 - → 🗃	
A Critical Evaluation of Comments able Addition Models for Semantic Interpretation (1990) (Correct) (5 citations) Peter Norig Robert Wilensky University of California, Berkeley Computer Thatsenth Intermational Conference on Computational Linguistics, Volume 3 NCC Rescearchindex Bookmark Context Related Cinter summary Cinter summary Astract: this paper we critically evaluate three recent abdictive interpretation models, those of Charnisk and Ooldman (1989), Hobbs, Stickel, Matin and Edwards (1983), and Ng and Moorey (1990). These three models add the inpotenti property of comments articles (1989), Hobbs, Stickel, Matin and Edwards (1983), and Ng and Moorey (1990). These three models add the inpotenti property of comments articles (1989), Hobbs, Stickel, Matin and Edwards (1983), and Ng and Moorey (1990). These three models add the inpotenti property of commensus addity: all types of evaluates expresented in a common currency that can be compared and combined. While commensus addity is a destable property, and there is a clear mesh of the abductive approach, and come tentative solutions. (Update) Context clight modifications of the one given in [Ng and Moorey. (1990) The new definition remedies the samely reported in [Forvig and Wilensky, 1990] of eccessionally preferring spuries interpretations of greater deptes. Table 1: Empiried Results Comparing Coherence and , costs as probabilities, opecifically within the context of using abductions for text interpretations. (Judates) Context clight modification of the one given in [Ng and Moorey (1999) (Correct) Abduction and Alismatch in a Hybrid MT System - Gawron (1999) (Correct) Abduction and Alismatch in a Hybrid MT System - Gawron (1999) (Correct) Abduction and Alismatch in a Hybrid MT System - Gawron (1999) (Correct) Abduction and Alismatch in a Hybrid MT System - Gawron (1999) (Correct) Abduction and Alismatch in Abduction Translation - Cawron (1999) (Correct) Abduction and Alismatch in Abduction an Adduction and Alismatch in Machine	Address 🙆 http://citeseer.nj.nec.com/peter90critical.html	Links »
Abstract: this paper we enticelly evaluate three recent abductive interpretation models, those of Charniak and Oddman (1989); Hobbs, Stickel, Mattin and Edwards (1983); and Ng and Mooney (1990). These three models add the inportant property of commensusability: all types of evidence are represented in a common currency that can be compared and combined. While commensusability is a desirable property, and there is a clear meed for a way to compare alternate explanations, it appears that a single scalar measure is not enough to account for all types of processing. We present other problems for the abductive approach, and some tentative solutions. (Update) Context of citations to this paper: More , (Areak slight modification of the one given in [Ng and Mooney, 1990] The new definition remedies the anomaly reported in [Norvig and Wilensky, 1990] of occasionally preferring spurious interpretations of greater depths. Table 1: Empiric all Results Comparing Coherence and , costs as probabilities, specifically within the context of using abduction for text interpretation, are discussed in Norvig and Wilensky (1990). The use of abduction in disambiguation is discussed in Kay et al 1990) We will assume the following: [3] a. Only literals Clied by: More Translation Mismatch in a Hybrid MT System - Gawron (1999) (Correct) Abduction and Mismatch in Machine Translation - Gawron (1999) (Correct) Interpretation as Abduction - Hobbs, Stickel, Appelt, Martin (1990) (Correct) Active bibliography (related documents): More All 0.1: Critiquing: Effective Decision Support in Time-Critical Domains - Gertner (1995) (Correct) 0.1: Decision Analytic Networks in Arthicial Intelligence - Matzkevich, Abramson (1995) (Correct) 0.1: Decision Analytic Networks of Drevious Line(1902) (Correct)	Semantic Interpretation (1990) (Correct) (5 citations) Peter Norvig Robert Wilensky University of California, Berkeley Computer Thirteenth International Conference on Computational Linguistics, Volume 3 From: norvig.com/resume (more) From: norvig.com/resume (more)	
Abstract: this paper we critically evaluate three recent abductive interpretation models, those of Charniak and Goldman (1989), Hobbs, Stickel, Martin and Edwards (1983), and Ng and Mooney (1990). These three models add the important property of commensuability: all types of evidence are represented in a common currency that can be compared and combined. While commensuability is a desirable property, and there is a clear meed for a way to compare alternate explanations, it appears that a single scalar measure is not enough to account for all types of processing. We present other problems for the abductive approach, and some tentative solutions. (<u>Ipdate</u>) Context of citations to this paper: <u>More</u> (hreak slight modification of the one given in [Ng and Mooney, 1990] The new definition remedies the anomaly reported in [Norvig and Wilensky, 1990] of eccasionally preferring spurious interpretations of greater depths. Table 1: Empirical Results Comparing Coherence and costs as probabilities, specifically within the context of using abduction for text interpretation, are discussed in Norvig and Wilensky (1990). The use of abduction in disambiguation is discussed in Kay et al. 1990) We will assume the following: [3] a. Only literals Clied by: More <u>Translation Mismatch in a Hybrid MT System - Gawron (1999) (Correct)</u> Abduction and Mismatch in Machine Translation - Gawron (1999) (Correct) Interpretation as Abduction - Hobbs, Stickel, Appelt, Martin (1990) (Correct) Active bibliography (related documents): More All 0.1: Critical prepinty (related documents): More All 0.2: Correct) 0.1: Decision Analytic Networks in Artificial Intelligence - Matzkevich, Abramson (1995) (Correct) 0.1: Decision Analytic Networks of DeveloretsDoleme Lin (1902). (Correct)	Male this atticle.	est)
 (hreak slight modification of the one given in [Ng and Mooney, 1990] The new definition remedies the anomaly reported in [Norvig and Wilensky, 1990] of occasionally preferring spurious interpretations of greater depths. Table 1: Empirical Results Comparing Coherence and costs as probabilities, specifically within the context of using adduction for text interpretation, are discussed in Norvig and Wilensky (1990). The use of abduction in disambiguation is discussed in Kay et al 1990) We will assume the following: 13) a Only literals Cited by: More Translation Mismatch in a Hybrid MT System - Gawron (1999) (Correct) Abduction and Mismatch in Machine Translation - Gawron (1999) (Correct) Interpretation as Abduction - Hobbs, Stickel, Appelt, Martin (1990) (Correct) Active bibliography (related documents): More All 0.1 Critical Domains - Gertner (1995) (Correct) 0.2 Correct) 0.3 Correct) 0.4 Decision Analytic Networks in Artificial Intelligence - Matzkevich, Abramson (1995) (Correct) 0.4 Decision Analytic Networks in Artificial Intelligence - Matzkevich, Abramson (1995) (Correct) 0.5 A Decision Analytic Networks of Decisions - Delever U in (1992). (Correct) 0.5 A Decision Analytic Networks of Decisions - Delever U in (1992). (Correct) 0.5 A Decision Analytic Networks of Decisions - Delever U in (1992). (Correct) 0.5 A Decision Analytic Networks of Decisions - Delever U in (1992). (Correct) 0.5 A Decision Analytic Networks of Decisions - Delever U in (1992). (Correct) 0.5 A Decision Analytic Networks of Decisions - Delever U in (1992). (Correct) 0.5 A Decision Analytic Networks of Decisions - Delever U in (1992). (Correct) 0.5 A Decision Analytic Networks in Artificial Intelligence - Matzkevich, Abramson (1995). (Cor	(1989); Hobbs, Stickel, Martin and Edwards (1983); and Ng and Mooney (1990). These three models add the important property of commensurability: all types of evidence are represented in a common currency that can be compared and combined. While commensurability is a desirable property, and there is a clear need for way to compare alternate explanations, it appears that a single scalar measure is not enough to account for all types of processing. We present other problems for	ra
occasionally preferring spurious interpretations of greater depths. Table 1: Empirical Results Comparing Coherence and costs as probabilities, specifically within the context of using abduction for text interpretation, are discussed in Norvig and Wilensky (1990). The use of abduction in disambiguation is discussed in Kay et al. 1990) We will assume the following: 13) a. Only literals Cited by: More Translation Mismatch in a Hybrid MT System - Gawron (1999) (Correct) Abduction and Mismatch in Machine Translation - Gawron (1999) (Correct) Interpretation as Abduction - Hobbs, Stickel, Appelt, Martin (1990) (Correct) Active bibliography (related documents): More All 0.1: Critiquing: Effective Decision Support in Time-Critical Domains - Gertner (1995) (Correct) 0.1: Decision Analytic Networks in Artificial Intelligence - Matzkevich, Abramson (1995) (Correct) 0.1: A Decksholiditie Networks in Artificial Intelligence - Matzkevich, Abramson (1995) (Correct)	Context of citations to this paper: More	
Cited by: More Translation Mismatch in a Hybrid MT System - Gawron (1999) (Correct) Abduction and Mismatch in Machine Translation - Gawron (1999) (Correct) Interpretation as Abduction - Hobbs, Stickel, Appelt, Martin (1990) (Correct) Active bibliography (related documents): More All 0.1: Critiquing: Effective Decision Support in Time-Critical Domains - Gertner (1995) (Correct) 0.1: Decision Analytic Networks in Artificial Intelligence - Matzkevich, Abramson (1995) (Correct) 0.1: A Deckeholdistic Networks of Decisions - Deleges Lin (1992). (Correct)	occasionally preferring spurious interpretations of greater depths. Table 1: Empirical Results Comparing Coherence and	
Translation Mismatch in a Hybrid MT System - Gawron (1999) (Correct) Abduction and Mismatch in Machine Translation - Gawron (1999) (Correct) Interpretation as Abduction - Hobbs, Stickel, Appelt, Martin (1990) (Correct) Active bibliography (related documents): More All 0.1: Critiquing: Effective Decision Support in Time-Critical Domains - Gertner (1995) (Correct) 0.1: Decision Analytic Networks in Artificial Intelligence - Matzkevich, Abramson (1995) (Correct) 0.1: A Deckeholistic Networks of Decisions - Deleges Lin (1992). (Correct)	abduction in disambiguation is discussed in Kay et al. 1990) We will assume the following: 13) a. Only literals	
Abduction and Mismatch in Machine Translation - Gawron (1999) (Correct) Interpretation as Abduction - Hobbs, Stickel, Appelt, Martin (1990) (Correct) Active bibliography (related documents): More All 0.1: Critiquing: Effective Decision Support in Time-Critical Domains - Gertner (1995) (Correct) 0.1: Decision Analytic Networks in Artificial Intelligence - Matzkevich, Abramson (1995) (Correct) 0.1: A Deckeholdistic Networks of Decisions - Deleges Lin (1992) (Correct)		
Interpretation as Abduction - Hobbs, Stickel, Appelt, Martin (1990) (Correct) Active bibliography (related documents): More All 0.1: Critiquing: Effective Decision Support in Time-Critical Domains - Gertner (1995) (Correct) 0.1: Decision Analytic Networks in Artificial Intelligence - Matzkevich, Abramson (1995) (Correct) 0.1: A Deckeholdistic Networks of Decisions - Deleges Lin (1992) (Correct)		
Active bibliography (related documents): More All 0.1: Critiquing: Effective Decision Support in Time-Critical Domains - Gertner (1995) (Correct) 0.1: Decision Analytic Networks in Artificial Intelligence - Matzkevich, Abramson (1995) (Correct) 0.1: A Dechebulistic Networks of Decisions - Deleges Line (1992). (Correct)		
 0.1: Critiquing: Effective Decision Support in Time-Critical Domains - Gertner (1995) (Correct) 0.1: Decision Analytic Networks in Artificial Intelligence - Matzkevich, Abramson (1995) (Correct) 1: A Dechebulistic Network of Decisions - Deleges Lie (1992). (Correct) 	merpretation as reduction - 110005, Sticker, report, Intarian (1550) (Sources)	
0.1: Decision Analytic Networks in Artificial Intelligence - Matzkevich, Abramson (1995) (Correct)		
a 1. A Deshahilistia Naturale of Deadiastas Doleana Lin (1992) (Correct)		
		-

Extracting Job Openings from the Web:

Semi-Structured Data



Outline

- Free text
 - Supervised and semi-automatic
 - AutoSlog
 - Semi-Supervised
 - AutoSlog-TS
 - Unsupervised
 - ExDisco
- Semi-structured and unstructured text
 - NLP-based wrapping techniques
 - RAPIER



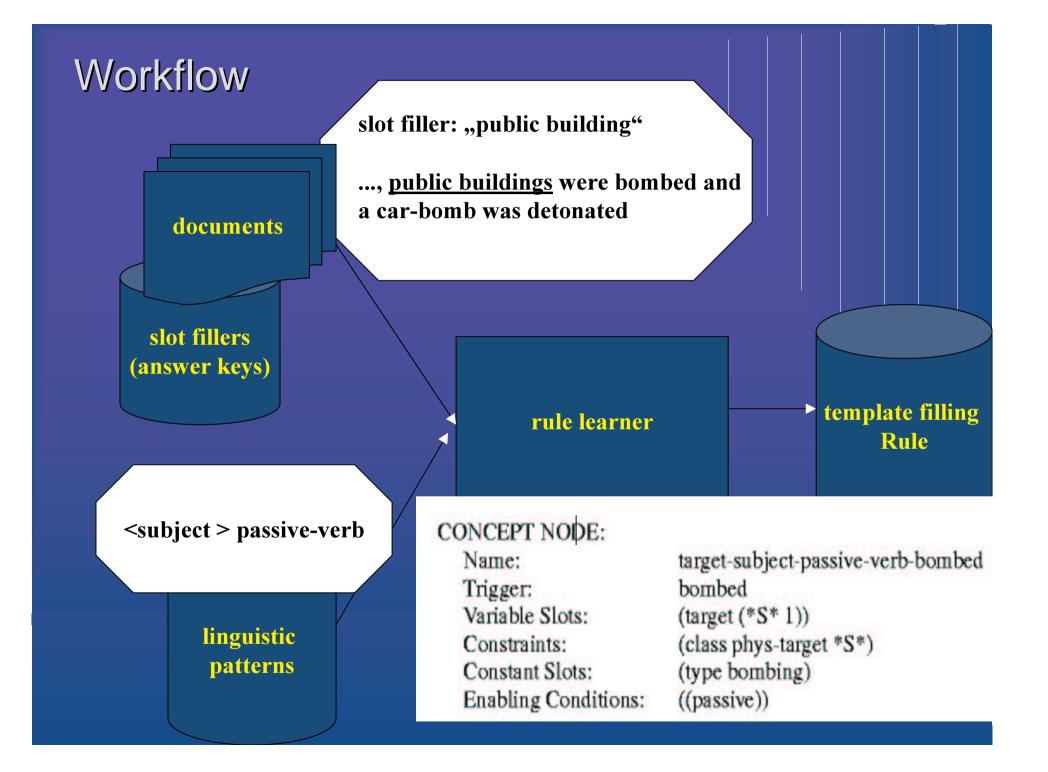
NLP-based Supervised Approaches

Input is an annotated corpus

- Documents with associated templates
- A parser
 - Chunk parser
 - Full sentence parser
- Learning the mapping rules
 - From linguistic constructions to template fillers

AutoSlog (1993)

- Extracting a concept dictionary for template filling
- Full sentence parser
- One slot filler rules
- Domain adaptation performance
 - Before AutoSlog: hand-crafted dictionary
 - two highly skilled graduate students
 - 1500 person-hours
 - AutoSlog:
 - A dictionary for the terrorist domain: 5 person hours
 - 98% performance achievement of the hand-crafted dictionary



Linguistic Patterns

Linguistic Pattern

Example

<subject> passive-verb <subject> active-verb <subject> verb infinitive <subject> auxiliary noun

passive-verb <dobj>¹ active-verb <dobj> infinitive <dobj> verb infinitive <dobj> gerund <dobj> noun auxiliary <dobj>

noun prep <np> active-verb prep <np> <victim> was <u>murdered</u> <perpetrator> <u>bombed</u> <perpetrator> attempted to <u>kill</u> <victim> was <u>victim</u>

<u>killed</u> <victim> <u>bombed</u> <target> to <u>kill</u> <victim> threatened to <u>attack</u> <target> <u>killing</u> <victim> fatality was <victim>

<u>bomb</u> against <target> killed with <instrument> Slot filler: "gilberto molasco" patricio rodriguez, and 17-year-old andres argueta, son of Sentence: (they took 2-year-old gilberto molasco, son of Id: DEV-MUC4-1192 emimesto argueta.)

CONCEPT NODE

Name: Trigger: Variable Slots: Constraints: Constant Slots: Enabling Conditions:

victim-active-verb-dobj-took took (victim (*DOBJ* 1)) (class victim *DOBJ*) (type kidnapping) ((active))

A bad concept node definition

Error Sources

- A sentence contains the answer key string but does not contain the event
- The sentence parser delivers wrong results
- A heuristic proposes a wrong conceptual anchor

Training Data

- MUC-4 corpus
- 1500 texts
- 1258 answer keys
- 4780 string fillers
- 1237 concept node definition
- Human in loop for validation to filter out bad and wrong definitions: 5 hours
- 450 concept nodes left after human review

System/Test Set	Recall	Precision	F-measure
MUC-4/TST3	46	56	50.51
AutoSlog/TST3	43	56	48.65
MUC-4/TST4	44	40	41.90
AutoSlog/TST4	39	45	41.79

Comparative Results

Summary

Advantages

- Semi-automatic
- Less human effort

- Disadvantages
 - Human interaction
 - Still very naive approach
 - Need a big amount of annotation
 - Domain adaptation bottelneck is shifted to human annotation
 - No generation of rules
 - One slot filling rule
 - No mechanism for filtering out bad rules

NLP-based ML Approaches

- LIEP (Huffman, 1995)
- PALKA (Kim & Moldovan, 1995)
- HASTEN (Krupka, 1995)
- CRYSTAL (Soderland et al., 1995)

CRYSTAL [1995]

The Parliament building was bombed by Carlos.

Concept type: BUILDING BOMBING

SUBJECT:	Classes include: Terms include: Extract:	<pre>< <physicaltarget> BUILDING target</physicaltarget></pre>
VERB:	Root: Mode:	BOMB passive
PREPOS-PHRASE:	Preposition: Classes include: Extract:	BY <personname> perpetrator name</personname>

PALKA [1995]

The Parliament building was bombed by *Carlos*.

FP-structure = N	MeaningFrame + Phrasa	lPattern
Meaning Frame:	(BOMBING agent:	ANIMATE
_	target:	PHYS-OBJ
	instrument:	PHYS-OBJ
	effect:	STATE)

Phrasal Pattern: ((PHYS-OBJ) was bombed by (PERP))

FP-structure:

(BOMBING target: PHYS-OBJ agent: PERP pattern: ((target) was bombed by (agent))

LIEP [1995]

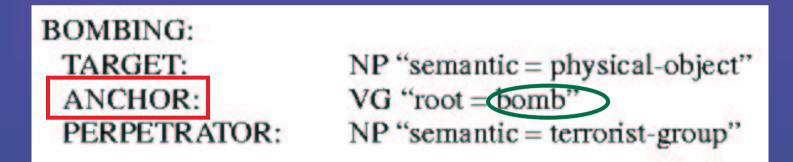
The Parliament building was bombed by *Carlos*.

TARGET-was-bombed-by-PERPETRATOR: noun-group(TRGT, head(isa(physical-target))), noun-group(PERP, head(isa(perpetrator))) verb-group(VG, type(passive), head(bombed)) preposition(PREP, head(by))

subject(TRGT, VG),
post-verbal-prep(VG, PREP),
prep-object(PREP, PERP)
⇒ bombing-event(BE, target(TRGT), agent(PERP))

HASTEN [1995]

The Parliament building was bombed by Carlos.



Egraphs

(SemanticLabel, StructuralElement)

Semi-Supervised Approaches

AutoSlog TS [Riloff, 1996]

 Input: pre-classified documents (relevant vs. irrelevant)
 NLP as preprocessing: full parser for detecting subject-vobject relationships

•Principle

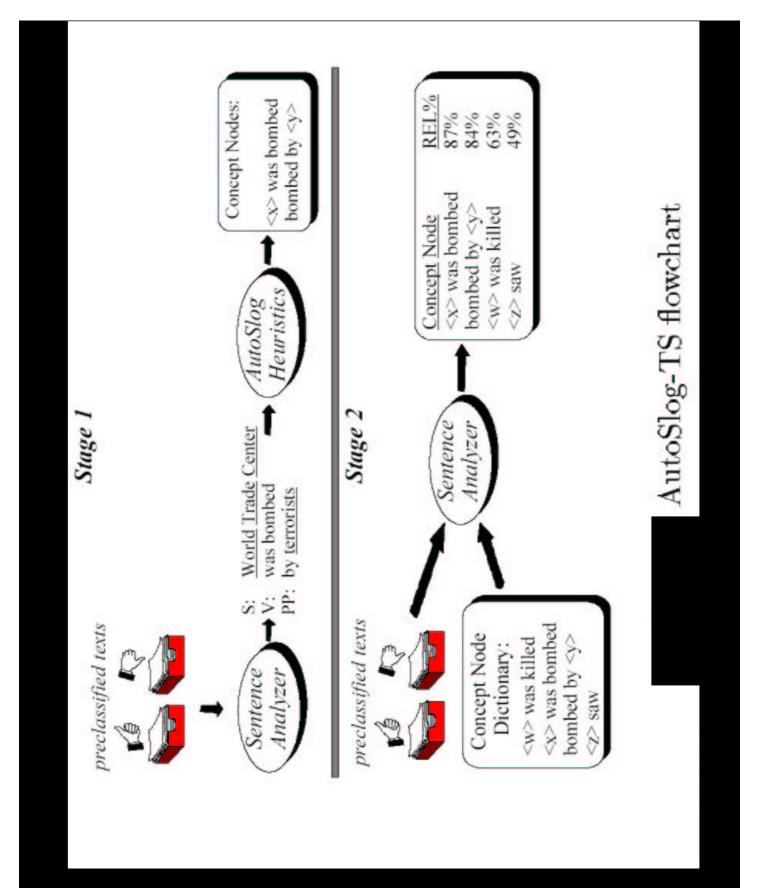
•Relevant patterns are patterns occuring more often in the relevant documents

•Output: ranked patterns, but not classified, namely, only the left hand side of a template filling rule

•The dictionary construction process consists of two stages:

pattern generation and

- statistical filtering
- •Manual review of the results



Pattern Extraction

The sentence analyzer produces a syntactic analysis for each sentence and identified noun phrases. For each noun phrase, the heuristic rules generate a pattern to extract noun phrase.

<subject> bombed

Relevance Filtering

- the whole text corpus will be processed a second time using the extracted patterns obtained by stage 1.
- Then each pattern will be assigned with a relevance rate based on its occurring frequency in the relevant documents relatively to its occurrence in the total corpus.
- A preferred pattern is the one which occurs more often in the relevant documents.

Statistical Filtering Relevance Rate:

 $Pr(relevant text \setminus text contains case frame_{i}) = \frac{rel - freq_{i}}{total - freq_{i}}$ $rel-freq_{i} \text{ number of instances of } case - frame_{i} \text{ in the relevant documents}$ $total-freq_{i} \text{ total number of instances of } case - frame_{i}$

Ranking Function:

score_i = relevance rate_i * log₂ (frequency_i)
Pr < 0,5negatively correlated with the domain</pre>

"Тор"

- 1. <subj> exploded
- 2. murder of <np>
- 3. assassination of <np>
- 4. <subj> was killed
- 5. <subj> was kidnapped
- 6. attack on <np>
- 7. <subj> was injured
- 8. exploded in <np>
- 9. death of $\langle np \rangle$
- 10. <subj> took_place
- 11. caused <dobj>
- 12. claimed <dobj>
- 13. <subj> was wounded

- 14. <subj> occurred
- 15. <subj> was located
- 16. took_place on <np>
- 17. responsibility for <np>
- 18. occurred on $\langle np \rangle$
- 19. was wounded in <np>
- 20. destroyed <dobj>
- 21. <subj> was murdered
- 22. one of $\langle np \rangle$
- 23. <subj> kidnapped
- 24. exploded on <np>
- 25. <subj> died

The Top 25 Extraction Patterns

Empirical Results

•1500 MUC-4 texts

•50% are relevant.

•In stage 1, 32,345 unique extraction patterns.

 A user reviewed the top 1970 patterns in about 85 minutes and kept the best 210 patterns.

•Evaluation

•AutoSlog and AutoSlog-TS systems return comparable performance.

Conclusion

• Advantages

- Pioneer approach to automatic learning of extraction patterns
- Reduce the manual annotation
- Disadvantages
 - Ranking function is too dependent on the occurrence of a pattern, relevant patterns with low frequency can not float to the top
 - Only patterns, not classification

Unsupervised

ExDisco (Yangarber 2001)

Seed

- Bootstrapping
- Duality/Density Principle for validation of each iteration

Input

- a corpus of unclassified and unannotated documents
- a seed of patterns, e.g.,

subject(company)-verb(appoint)-object(person)

NLP as Preprocessing

- full parser for detecting subject-v-object relationships
 - NE recognition
 - FDG formalism (Tapannaien & Järvinen, 1997)

Duality/Density Principle (boostrapping)

• Density:

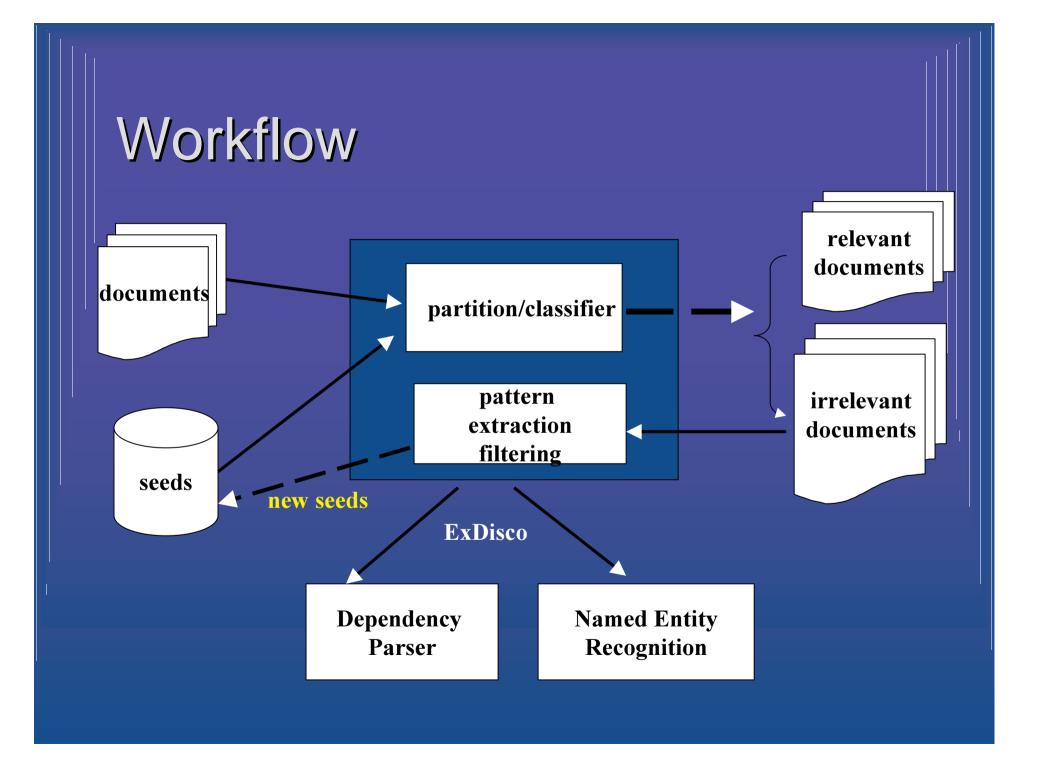
• Relevant documents contain more relevant patterns

• Duality:

- documents that are relevant to the scenario are strong indicators of good patterns
- good patterns are indicators of relevant documents

Algorithm

- Given:
 - a large corpus of un-annotated and un-classified documents
 - a trusted set of scenario patterns, initially chosen ad hoc by the user, the seed. Normally is the seed relatively small, two or three
 - (possibly empty) set of concept classes
- Partition
 - applying seed to the documents and divide them into relevant and irrelevant documents
- Search for new candidate patterns:
 - automatic convert each sentence into a set of candidate patterns.
 - choose those patterns which are strongly distributed in the relevant documents
 - Find new concepts
- User feedback
- Repeat



Pattern Ranking

Score(P)=|H∩R| IH|

Evaluation of Event Extraction

Pattern Base	Recall	Precision	F
Seed	27	74	39.58
ExDisco	52	72	60.16
Union	57	73	63.56
Manual-MUC	47	70	56.40
Manual-NOW	56	75	64.04

ExDisco

- Advantages
 - Unsupervised
 - Multi-slot template filler rules

Disadvantages

- Only subject-verb-object patterns, local patterns are ignored
- No generalization of pattern rules (see inductive learning)
- Collocations are not taken into account, e.g., PN take responsibility of Company

• Evaluation methods

- Event extraction: integration of patterns into IE system and test recall and precision
- Qualitative observation: manual evaluation
- Document filtering: using ExDisco as document classifier and document retrieval system

Relational learning and Inductive Logic Programming (ILP)

 Allow induction over structured examples that can include first-order logical representations and unbounded data structures

Semi-Structured and Un-Structured Documents

RAPIER [Califf, 1998]

- Uses relational learning to construct unbounded patternmatch rules, given a database of texts and filled templates
- Primarily consists of a bottom-up search
- Employs limited syntactic and semantic information
- Learn rules for the complete IE task

Filled template of RAPIER

Posting from Newsgroup

Telecommunications. SOLARIS Systems Administrator. 38-44K. Immediate need

Leading telecommunications firm in need of an energetic individual to fill the following position in the Atlanta office:

SOLARIS SYSTEMS ADMINISTRATOR Salary: 38-44K with full benefits Location: Atlanta Georgia, no relocation assistance provided

Filled Template

computer_science_job
title: SOLARIS Systems Administrator
salary: 38-44K
state: Georgia
city: Atlanta
platform: SOLARIS
area: telecommunications

Figure 1: Sample Message and Filled Template

RAPIER's rule representation

Indexed by template name and slot name
Consists of three parts:

A pre-filler pattern
Filler pattern (matches the actual slot)
Post-filler

Pattern

- Pattern item: matches exactly one word
- Pattern list: has a maximum length N and matches 0...N words.
- Must satisfy a set of constraints
 - 1. Specific word, POS, Semantic class
 - 2. Disjunctive lists

RAPIER Rule

ORIGINAL DOCUMENT: AI. C Programmer. 38-44K. Leading Alfirm in need of an energetic individual to fill the following position: EXTRACTED DATA: computer-science-job title: C Programmer salary: 38-44K area: Al

AREA extraction pattern:Pre-filler pattern:word: leadingFiller pattern:list: len: 2tags: [nn, nns]Post-filler pattern:word: [firm, company]

RAPIER'S Learning Algorithm

- Begins with a most specific definition and compresses it by replacing with more general ones
- Attempts to compress the rules for each slot
- Preferring more specific rules

Implementation

- Least general generalization (LGG)
- Starts with rules containing only generalizations of the filler patterns
- Employs top-down beam search for pre and post fillers
- Rules are ordered using an information gain metric and weighted by the size of the rule (preferring smaller rules)

Example

Located in Atlanta, Georgia. Offices in Kansas City, Missouri

Pre-filler: Filler: Post-filler: 1) word: located 1) word: atlanta 1) word: . tag: vbn tag: nnp tag: . 2) word: in 2) word: georgia tag: in tag: nnp 3) word: . tag: . and Pre-filler: Filler: Post-filler: 1) word: offices 1) word: kansas 1) word: , tag: nnp tag: nns tag: . 2) word: in 2) word: city word: missouri tag: in tag: nnp tag: nnp 3) word: . tag: .

Example (cont)

Pre-filler:	Filler:	Post-filler:
	1) list: max len word: {atlan	gth: 2 ita, kansas, city}
and	tag: nnp	
Pre-filler:	Filler:	Post-filler:
	1) list: max len	gth: 2
	tag: nnp	

Pre-filler:	Filler:	Post-filler:
1) word: in tag: in	1) list: max length: word: {atlanta, kansas, city} tag: nnp	2 1) word: , tag: ,
and		
Pre-filler:	Filler:	Post-filler:
1) word: in	1) list: max length:	2 1) word: ,

tag: in tag: nnp tag: ,

Example (cont)

Final best rule:

Pre-filler: Filler: Post-filler: 1) word: in 1) list: max length: 2 1) word: , tag: in tag: nnp tag: , 2) tag: nnp semantic: state

Experimental Evaluation

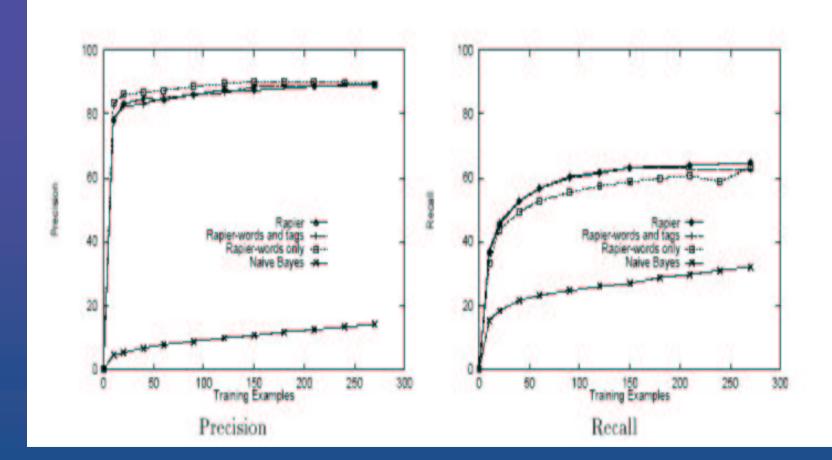
- A set of 300 computer-related job posting from austin.jobs
- A set of 485 seminar announcements from CMU.
- Three different versions of RAPIER were tested

1.words, POS tags, semantic classes

2. words, POS tags

3. words

Performance on job postings



Results for seminar announcement task

System	stime		etime		loc		speaker	
	Prec	Rec	Prec	Rec	Prec	Rec	Prec	Rec
RAPIER	93.9	92.9	95.8	94.6	91.0	60.5	80.9	39.4
RAP-WT	96.5	95.3	94.9	94.4	91.0	61.5	79.0	40.0
RAP-W	96.5	95,9	96.8	96.6	90.0	54.8	76.9	29.1
NAIBAY	98.2	98.2	49.5	95.7	57.3	58.8	34.5	25.6
SRV	98.6	98.4	67.3	92.6	74.5	70.1	54.4	58.4
WHISK	86.2	100.0	85.0	87.2	83.6	55.4	52.6	11.1
WH-PR	96.2	100.0	89.5	87.2	93.8	36.1	0.0	0.0

Conclusion

- Pros
 - Have the potential to help automate the development process of IE systems.
 - Work well in locating specific data in newsgroup messages
 - Identify potential slot fillers and their surrounding context with limited syntactic and semantic information
 - Learn rules from relatively small sets of examples in some specific domain

Cons

- single slot
- regular expression
- Unknown performances for more complicated situations