Overview of FASTUS

Feiyu Xu
DFKI LT-Lab
Two Approaches to Building IE Systems
[Appelt & Israel, 99]

- **Knowledge Engineering Approach**
  - Grammars are constructed by hand
  - Domain patterns are discovered by a human expert through introspection and inspection of a corpus
  - Much laborious tuning

- **Automatically Trainable Systems**
  - Use statistical methods when possible
  - Learn rules from annotated corpora, e.g.,
    - statistical name recognizer
  - Learn rules from interaction with user, e.g.,
    - learning template filler rules
Knowledge Engineering
[Appelt & Israel, 99]

- **Advantages**
  - With skill and experience, good performing systems are not conceptually hard to develop
  - The best performing systems have been hand crafted

- **Disadvantages**
  - Very laborious development process
  - Some changes to specifications can be hard to accommodate, e.g.,
    - Name recognition rules based upper and lower cases
  - Required experts may not be available
Trainable Systems
[Appelt & Israel, 99]

Advantages
- Domain portability is relatively straightforward
- System experts is not required for customization
- “Data driven” rule acquisition ensures full coverage of examples

Disadvantages
- 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
When Works Best for Knowledge based IE? [Appelt & Israel, 99]

- Resources (e.g. lexicons, lists) are available
- Rule writers are available
- Training data scarce or expensive to obtain
- A mixture of knowledge based and machine learning based approach is also possible!
FASTUS
[Hobbs et al. 96]

- Acronym for *Finite State Automaton Text Understanding System*
- Developed in SRI International, Menlo Park, California
- Joined MUC-4 (92), MUC-5 (93), MUC-6 (95), MUC-7 (97)
- English and Japanese

Inspired by

- the performance in MUC-3 that the group at the University of Massachusetts got out of a simple system [Lehnert et al., 1991]
- Pereira’s work on finite-state approximations of grammars [Pereira, 1990]

- Works as a set of cascaded and nondeterministic finite-state automata
Overview of the FASTUS Architecture
[Hobbs et al. 96]

Recognizing Phrases

- Complex Words
  - Text
  - joint venture, set up, Bridgstone Sports Taiwan Co
  - noun groups: a local concern
  - verb groups: had set up
- Basic Phrases
- Complex Phrases
  - appositive attachment: Peter, the director
  - pp attachment: production of iron and wood
- Domain Event Patterns
- Merging Structures

Source: Feiyu Xu

Sommersemester 2001
Example of Phrase Recognition
[Hobbs et al. 96]

Salvadoran President-elect Alfredo Cristiani condemned the terrorist killing of Attorney General Roberto Garcia Alvarado and accused the Farabundo Marti National Liberation Front (FMLN) of crime

Noun Group: Salvadoran President-elect
Name: Alfredo Cristiani
Verb Group: condemned
Noun Group: the terrorist killing
Preposition: of
Noun Group: Attorney General
Name: Roberto Garcia Alvarado
Conjunction: and
Verb Group: accused
Noun Group: the Farabundo Marti National Liberation Front (FMLN)
Preposition: of
Noun Group: the crime
Example of Domain Event Pattern Recognition
[Hobbs et al. 96]

[Salvadoran President-elect Afredo Cristiani]² condemned [the terrorist killing of Attorney General Roberto Garcia Alvarado]¹ and [accused the Farabundo Marti National Liberation Front (FMLN) of crime]²

Two patterns are recognized:
1. <Perpetrator> killing of <HumanTarget>
2. <GovtOfficial> accused <PerpOrg> of <Incident>

Two incident structures are constructed:

Incident: KILLING
Perpetrator: „terrorist“
Confidence: __
Human Target: „Roberto Garcia Alvarado“

Incident: KILLING
Perpetrator: FMLN
Confidence: Accused by Authorities
Human Target: __
“pseudo-syntax” [Hobbs et al. 96]

- The material between the end of the subject noun group and the beginning of the main verb group must be read over, for example,

  ✓ Read over prepositional phrases and relative clauses

  1. Subject {Preposition NounGroup}* VerbGroup
  2. Subject Relpro {NounGroup | Other}* VerbGroup

  The mayor, who was kidnapped yesterday, was found dead today.

- Conjoined verb phrase, skipping over the first conjunct and associate the subject with the verb group in the second conjunct

  Subject VerbGroup {NounGroup|Other}* Conjunction VerbGroup
Problem of “pseudo-syntax” 
[Hobbs et al. 96]

- Same semantic content can be realized in different forms.
  - GM manufactures cars.
  - Cars are manufactured by GM.
  - ... GM, which manufactures cars ...
  - ... cars, which are manufactured by GM ...
  - ... cars manufactured by GM ...
  - GM is to manufacture cars.
  - Cars are to be manufactured by GM.
  - GM is a car manufacturer.

- Question: How many rules are needed to extract all relevant patterns? Why not using a linguistic theory?
  - Performance vs. Competence
  - TACITUS: 36 hours to process 100 Messages
  - FASTUS: 12 minutes to process 100 messages
Example of Template Merging
[Hobbs et al. 96]

Salvadoran President-elect Alfredo Cristiani condemned the terrorist killing of Attorney General Roberto Garcia Alvarado and accused the Farabundo Marti National Liberation Front (FMLN) of crime

Two incident structures are merged as:

Incident: KILLING
Perpetrator: „terrorist“
Confidence: __
Human Target: „Roberto Garcia Alvarado"

Incident: KILLING
Perpetrator: FMLN
Confidence: Accused by Authorities
Human Target: __

Incident: KILLING
Perpetrator: FMLN
Confidence: Accused by Authorities
Human Target: „Roberto Garcia Alvarado"
Evaluation of FASTUS
[Hobbs et al. 96]

<table>
<thead>
<tr>
<th>EVALUATION\TE</th>
<th>MUC-4 (English)</th>
<th>MUC-5 (English)</th>
<th>MUC-5 (Japanese)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RECALL</td>
<td>44</td>
<td>34</td>
<td>34</td>
</tr>
<tr>
<td>PRECISION</td>
<td>55</td>
<td>56</td>
<td>62</td>
</tr>
</tbody>
</table>