SEMI-AUTOMATIC RULE CONSTRUCTION FOR SEMANTIC LINKING OF RELATION ARGUMENTS

Janez Starc
Dunja Mladenić
Introduction

- Constructing pattern rules
  - Lexical patterns

<table>
<thead>
<tr>
<th>Lexical tokens</th>
<th>Federer</th>
<th>won</th>
<th>5</th>
<th>trophies</th>
<th>in</th>
<th>Asia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part-of-speech tags</td>
<td>NNP</td>
<td>VBD</td>
<td>CD</td>
<td>NNS</td>
<td>IN</td>
<td>NNP</td>
</tr>
<tr>
<td>Named entities</td>
<td>B-PER</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>B-LOC</td>
</tr>
<tr>
<td>Generalized tokens</td>
<td>[Person]</td>
<td>won</td>
<td>[Number]</td>
<td>trophies</td>
<td>in</td>
<td>[Location]</td>
</tr>
</tbody>
</table>

- Logical patterns
  \[
  (\text{numberOfTrophies} \ [\text{Person}] \ [\text{Number}] \ [\text{Location}])
  \rightarrow
  (\text{numberOfTrophies} \ \text{RogerFederer} \ 5 \ \text{Asia})
  \]

- Functions useful for nesting
  \[
  (\text{JuiceFn} \ \text{Orange-Fruit}) = \text{orange juice}
  \]
Role in organization - example

Roy Blake, chief researcher and former president of Tech Lab
Role in organization - example

Roy Blake, chief researcher and former president of Tech Lab

chief [Role]

(ChiefFn [Role])

(ChiefFn Researcher)
Role in organization - example

Roy Blake, chief researcher and former president of Tech Lab

chief [Role]  
(ChiefFn [Role])

former [Role]  
(FormerFn [Role])

(ChiefFn Researcher)  =  [Role]  
(FormerFn President)  =  [Role]
Role in organization - example

Roy Blake, chief researcher and former president of Tech Lab
Role in organization - example

Roy Blake, chief researcher and former president of Tech Lab

[Role] and [Role]

(Intersection [Role] [Role])

(Intersection (ChiefFn Researcher) (FormerFn Researcher))

= [Role]
Role in organization - example

Roy Blake, chief researcher and former president of Tech Lab
Roy Blake, chief researcher and former president of Tech Lab

(Person, Role) of (Organization)

(rolenInOrganization [Person] [Organization] [Role])

(rolenInOrganization RoyBlake TechLab
  (Intersection (ChiefFn Researcher) (FormerFn President)))
Architecture

Corpus

Relation extraction

Initial relations

Unlinked arguments

Applying rules

Generalization

Pattern Recommendation

Pattern rule construction

Rule store

Seed entity rules

User

lexical patterns

focus on target argument

new unlinked arguments

source

entity rules

new entity rules

all rules

added at the beginning

most frequent lexical patterns

constructs

corpus

Pattern & entity rules
Initial relation extraction

- Applying pattern:

  \([Person], [Role] \text{ of } [Organization]\)

- former chairman
- manager
- president and publisher
- who is now president
- who has composed the scores for most
Architecture

Corpus

Relation extraction

Initial relations

Unlinked arguments

Generalization

Pattern Recommendation

Pattern rule construction

Applying rules

Rule store
Pattern & entity rules

Seed entity rules

User

lexical patterns

focus on target argument

new unextracted arguments

source

entity rules

new entity rules

most frequent lexical patterns

constructs

new pattern rules

added at the beginning

all rules
Generalization

chief researcher and former president

Generalization:
Named-entity generalization
Entity rules (seed entity rules from Freebase)

“engineer”

chief [Role] and former [Role]
Architecture

Corpus

Relation extraction

Initial relations

Unlinked arguments

Focus on target argument

Generalization

lexical patterns

Pattern Recommendation

Pattern rule construction

Most frequent lexical patterns

constructs

Pattern rule construction

Rule store

Pattern & entity rules

Seed entity rules

Applying rules

all rules

new entity rules

new unextracted arguments

source

Entity rules

new pattern rules

User

Added at the beginning

All rules
## Pattern recommendation

<table>
<thead>
<tr>
<th>LEXICAL PATTERN</th>
<th>FREQUENCY</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Role]</td>
<td>7149</td>
</tr>
<tr>
<td>the [Role]</td>
<td>1439</td>
</tr>
<tr>
<td>a [Role]</td>
<td>911</td>
</tr>
<tr>
<td>[Role] and [Role]</td>
<td>870</td>
</tr>
<tr>
<td>who is [Role]</td>
<td>344</td>
</tr>
<tr>
<td>former [Role]</td>
<td>245</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>[Role] and the [Role]</td>
<td>30</td>
</tr>
</tbody>
</table>

ENTITY RULES
Architecture

Corpus

Relation extraction

Initial relations

Unlinked arguments

Generalization

Pattern Recommendation

Pattern rule construction

Applying rules

Rule store

Seed entity rules

Lexical patterns

Focus on target argument

Most frequent lexical patterns

Entity rules

New entity rules

New pattern rules

Constructs

New unextracted arguments

Source

All rules

Added at the beginning

User
Pattern rule construction

<table>
<thead>
<tr>
<th>LEXICAL PATTERN</th>
<th>FREQUENCY</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Role]</td>
<td>7149</td>
</tr>
<tr>
<td>the [Role]</td>
<td>1439</td>
</tr>
<tr>
<td>a [Role]</td>
<td>911</td>
</tr>
<tr>
<td>[Role] and [Role]</td>
<td>870</td>
</tr>
<tr>
<td>who is [Role]</td>
<td>344</td>
</tr>
<tr>
<td>former [Role]</td>
<td>245</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>[Role] and the [Role]</td>
<td>30</td>
</tr>
</tbody>
</table>

USER SELECTS (FormerFn [Role])
Architecture

- Corpus
- Relation extraction
- Initial relations
- Unlinked arguments
- Applying rules
  - Rule store
    - Pattern & entity rules
  - Seed entity rules
- Generalization
- Pattern Recommendation
- Pattern rule construction
- New unextracted arguments
- Focus on target argument
- Source
- Entity rules
- New entity rules
- All rules
- User
- Constructs
- New pattern rules
- Added at the beginning
- Lexical patterns
- Most frequent lexical patterns
- Seed entity rules
Applying pattern rules

former president

former [Role] → (FormerFn [Role])

(FormerFn President)

= ENTITY RULE
Applying pattern rules

former president

former [Role] → (FormerFn [Role])

(FormerFn President)

= ENTITY RULE

chief researcher and former president

chief researcher = [Role]
former president = [Role]

[Role] and [Role]

= UNLINKED ARGUMENTS
Architecture

- Corpus
  - Relation extraction
    - Initial relations
      - Unlinked arguments
        - focus on target argument
        - new unextracted arguments
  - Applying rules
    - all rules
      - new entity rules
      - new pattern rules
      - added at the beginning

- Generalization
  - lexical patterns
  - source

- Pattern Recommendation
  - most frequent lexical patterns
  - constructs

- Pattern rule construction
  - entity rules
  - new entity rules

- Rule store
  - Pattern & entity rules
  - User
  - Seed entity rules
## Next iteration

<table>
<thead>
<tr>
<th>LEXICAL PATTERN</th>
<th>OLD FREQUENCY</th>
<th>NEW FREQUENCY</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Role]</td>
<td>7149</td>
<td>7732</td>
</tr>
<tr>
<td>the [Role]</td>
<td>1439</td>
<td></td>
</tr>
<tr>
<td>[Role] and [Role]</td>
<td>870</td>
<td>968</td>
</tr>
<tr>
<td>a [Role]</td>
<td>911</td>
<td>911</td>
</tr>
<tr>
<td>who is [Role]</td>
<td>344</td>
<td>359</td>
</tr>
<tr>
<td>former [Role]</td>
<td>245</td>
<td>256</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>[Role] and the [Role]</td>
<td>30</td>
<td></td>
</tr>
</tbody>
</table>

Selected on previous iteration
## Next iteration

<table>
<thead>
<tr>
<th>LEXICAL PATTERN</th>
<th>OLD FREQUENCY</th>
<th>NEW FREQUENCY</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Role]</td>
<td>7149</td>
<td>7732</td>
</tr>
<tr>
<td>the [Role]</td>
<td>1439</td>
<td></td>
</tr>
<tr>
<td>[Role] and [Role]</td>
<td>870</td>
<td>968</td>
</tr>
<tr>
<td>a [Role]</td>
<td>911</td>
<td>911</td>
</tr>
<tr>
<td>who is [Role]</td>
<td>344</td>
<td>359</td>
</tr>
<tr>
<td>former [Role]</td>
<td>245</td>
<td>256</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td>...</td>
</tr>
<tr>
<td>[Role] and the [Role]</td>
<td>30</td>
<td></td>
</tr>
</tbody>
</table>

Selected on previous iteration

Merge automatically
Evaluation

- 1,300,000 general news articles
- 31 pattern rules created
- 14,247 completely linked roles
- 300 initial arguments left out for testing
- 100% precision, 84% recall
Conclusion and future work

• Suggesting frequent lexical patterns for linking relation arguments
• Constructing small number of rules that cover a large portion of text

• Future work
  • Automatic decomposition of text (lexical patterns)
  • Adding complexity → Wikipedia (first paragraph of person articles)
QUESTIONS ?