Language Technology I 2013/2014

Relation Extraction Exercises

1. What is relation extraction? (Slide 7)

Given an unstructured text, a relation extraction (RE) tool should be able to automatically recognize and extract relations among the relevant entities or concepts that are salient to the user's needs

2. ***What are linguistic mentions? What are linguistic entities? (Slide 26)***

- Linguistic Mention
  - A particular linguistic phrase
  - Denotes a particular entity, relation, or event
    - A noun phrase, name, or possessive pronoun
    - A verb, nominalization, compound nominal, or other linguistic construct relating other linguistic mentions
- Linguistic Entity
  - Equivalence class of mentions with same meaning
    - Coreferring noun phrases
    - Relations and events derived from different mentions, but conveying the same meaning

3. What are the parameters of the IE Real-Work tasks? (slide 41)

Parameters of IE Real-World Tasks

- Document structure
  - Free text
  - Semi-structured
  - Structured
- Linguistic annotation
  - Shallow NLP
  - Deep NLP
- Complexity and specificity of relation
  - Unary
  - N-ary
- Depth of extraction
  - Recognition
  - Classification
  - Semantic role labelling
- Degree of automation
  - Semi-automatic
  - Supervised
  - Semi-Supervised
  - Minimally-Supervised
  - Distant Supervision
  - Unsupervised
- Human interaction/contribution
- Data properties
  - Domain relevance
  - Redundancy
  - Connectivity
- Evaluation/validation
  - With/without gold standard
  - Performance: recall & precision
  - Interaction among parameters

4. ***What are the motivations of machine learning for information extraction? (Slide 39)***

Answers:

- 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

5. *Please calculate Precision and Recall of the following information extraction task.*

$$\text{Precision} = \frac{\textit{correct answers extracted by the system}}{\textit{all answers extracted by the system}}$$

$$\text{Recall} = \frac{\textit{correct answers extracted by the system}}{\textit{total possible answers}}$$

Text

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

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

3. **"We can be open source. We love the concept of shared source," said Bill Veghte, a Microsoft VP. "That's a super-important shift for us in terms of code access.“**

4. **Richard Stallman, founder of the Free Software Foundation, countered saying…**

Possible Answers distributed in the different paragraphs

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

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

3. **"We can be open source. We love the concept of shared source," said Bill Veghte, a Microsoft VP. "That's a super-important shift for us in terms of code access.“**

4. **Richard Stallman, founder of the Free Software Foundation, countered saying…**

In the following, we list all possible names and their types in each paragraph

1. **Microsoft Corporation (organization name), CEO (title/position name), Bill Gates (person name)**
2. **Microsoft (organization name), Gates (person name)**
3. **Bill Veghte (person name), Microsoft (organization name), VP (title/position name)**
4. **Richard Stallman (person name), founder (title/position name), Free Software Foundation (organization name)**

Answers exacted by the System

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

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

3. **"We can be open source. We love the concept of shared source," said Bill Veghte, a Microsoft VP. "That's a super-important shift for us in terms of code access.“**

4. **Richard Stallman, founder of the Free Software Foundation, countered saying…**

In the following, we list all the extracted names and their types in each paragraph

1. **Microsoft Corporation (organization name), CEO (title/position name), Bill Gates (person name), Orwellian (person name)**
2. **Windows (person name)**
3. **Bill Veghte (person name), Microsoft VP (person name)**
4. **Richard Stallman (person name), founder (title/position name), Free Software Foundation (organization name)**

Please calculate for each name type the precision and recall values of the named recognition task of the system.

Name Types	Precision	Recall
Person Name		
Organization Name		
Title/Position Name		

Right Answer:

Name Types	Precision	Recall
Person Name	50%	75%
Organization Name	100%	50%
Title/Position Name	100%	66.7%

6. *Information Extraction in the Management Succession Domain*

Text for Information Extraction

1. **Cash-strapped Figgie International Inc. eliminated its quarterly dividend and received a temporary cash infusion from a new lender.**

2. **The Willoughby, Ohio, industrial company also named a longtime outside director as a new vice chairman, although it stopped short of bringing in a total outsider to bolster management.**

3. **Meanwhile, Figgie signaled that its 1993 operating losses and year-end adjustments would be greater than previously expected when reported in two to three weeks.**

4. **The company's Class A share fell $1.875, or 15%, to $10.875 in Nasdaq Stock Market trading.**

5. **The conglomerate said it received a new $40 million, one-year renewable loan from CIT Group, which is secured by receivables. That eliminated the need to ask its other banks, which have extended a $150 million credit line, for more funds. "With CIT's support, we have addressed our near-term cash flow needs while we continue to put into place the elements of a more comprehensive deleveraging of the balance sheet," Chairman Harry E. Figgie Jr. said in a statement.**

6. **Nevertheless, Figgie sought to calm its syndicate of bankers at a meeting yesterday. The bankers expressed concerns about the company's financial plight and the need for new management, according to a person who attended the meeting. A company spokesman confirmed that those issues were discussed, but said they were "not representative of the whole meeting."**

7. **Wall Street was hoping for stronger outside management to help Figgie. Instead, the company named a director, 66-year-old Walter M. Vannoy, who has been on the board since 1981. Although the current vice chairman, Harry E. Figgie III, 40, will continue to hold that title, Mr. Vannoy will be second in command, the company said.**

8. **Mr. Vannoy formerly served as vice chairman of McDermott International Inc. and as president and chief operating officer of Babcock & Wilcox.**

9. **The company said suspension of the six-cents-a-share quarterly dividend, which would have been payable in March on both Class A and Class B common stock, will save it about $4 million a year. "The board felt the dividend suspension was prudent until we effect a profitability turnaround."**

10. **The actions, coupled with the company's previously announced plan to sell its Rawlings Sporting Goods division and two other divisions, are the first phase of a turnaround plan to be disclosed within the next two weeks, Figgie said. Among other things, the company wants to lop $200 million off its $450 million of debt this year.**

Tasks

a. Named entity recognition
Please extract all names of the following types from the above text and specify the paragraph number, where the name occurs.

i. Person names, e.g., **Walter M. Vannoy (7)**
ii. Location names, e.g., **Ohio (2),**
iii. Company/organization names, e.g., **Figgie International Inc. (1)**
iv. Position names, e.g., **director (2), vice chairman (2)**

b. Relation recognition
Please extract the person-position and person-company relations from the above texts

i. Person-Position Relationships: a relationship between a person and his position in a company

Person	Position

ii. Person-Company Relationships: the affiliation of a specific person

Person	Company

c. Filling templates
Please try to fill the following database records by extracting corresponding information from the above text. The database record corresponds to a management succession event, which contains four attributes

i. Person In: person, who is named for the position
ii. Person Out: person, who resigned from the position
iii. Position: position, which needs a personnel change
iv. Company: company or organization, where the personnel changes take place
v. Time: when the personnel change takes place

Person IN	Person OUT	Position	Company	Time

d. Identify linguistic patterns as relation extraction rules for Person_In, Person_Out, Person_Company

Answers

a. Named entity recognition
Please extract all names of the following types from the above text and specify the paragraph number, where the name occurs.

i. Person names, e.g., **Walter M. Vannoy (7)**
ii. Location names, e.g., **Ohio (2),**
iii. Company/organization names, e.g., **Figgie International Inc. (1)**
iv. Position names, e.g., **director (2), vice chairman (2)**

b. Relation recognition
Please extract the person-position and person-company relations from the above texts

i. Person-Position Relationships: a relationship between a person and his position in a company

Person	Position
Harry E. Figgie Jr. (5)	**Chairman (5)**
Walter M. Vannoy (7)	**Director (7)**
Harry E. Figgie III (7)	**vice chairman (7)**
Mr. Vannoy (8)	**vice chairman (8)**
Mr. Vannoy (8)	**president and chief operating officer (8)**

ii. Person-Company Relationships: the affiliation of a specific person

Person	Company
Harry E. Figgie Jr. (5)	**CIT Group (5)**
Walter M. Vannoy (7)	**Figgie (7)**
Harry E. Figgie III (7)	**Figgie (7)**
Mr. Vannoy (7)	**Figgie (7)**
Mr. Vannoy (8)	**McDermott International Inc. (8)**
Mr. Vannoy (8)	**Babcock & Wilcox (8)**

b. Filling templates
Please try to fill the following database records by extracting corresponding information from the above text. The database record corresponds to a management succession event, which contains four attributes

i. Person In: person, who is named for the position
ii. Person Out: person, who resigned from the position
iii. Position: position, which needs a personnel change
iv. Company: company or organization, where the personnel changes take place
v. Time: when the personnel change takes place

Right Answer:

Person IN	Person OUT	Position	Company	Time
Walter M. Vannoy (7)		**Director (7)**	**Figgie (7)**	
	Mr. Vannoy (8)	**vice chairman (8)**	**McDermott International Inc. (8)**	
	Mr. Vannoy (8)	**president and chief operating officer (8)**	**Babcock & Wilcox (8)**	

c. Identify linguistic patterns as relation extraction rules for Person_In, Person_Out,

Right Answer:

1. Person_In:
<Company> ***named*** *<Person: Person_In> as <Position>*
2. Person_Out:
<Person: Person_Out> ***formerly served as*** *<Position> of <Company>*