

#### Lexico-Semantic Patterns for Learning Ontology Instances from Text

#### Frederik Hogenboom

fhogenboom@ese.eur.nl

Erasmus University Rotterdam PO Box 1738, NL-3000 DR Rotterdam, the Netherlands

February 7, 2013

# Introduction (1)

- Increasing amount of (digital) data
  - Problem: utilizing extracted information in decision making processes becomes increasingly urgent and difficult:
    - Too much data for manual extraction
    - Yet most data is initially unstructured
    - Data often contains natural language
- Solution: automatically process and interpret information, yet automation is a non-trivial task

# Introduction (2)

- Information Extraction (IE)
  - Multiple sources:
    - News messages
    - Blogs
    - Papers
    - ...
  - Text Mining (TM):
    - Natural Language Processing (NLP)
    - Statistics
    - ...
  - Specific type of information that can be extracted: events

## Events (1)



Steve Jobs resigns from Apple, Cook becomes CEO Nucleosday, Silicon Valley legend Steve Jobs resigned (The Guardian) - Apple's Jobs returned in Apple's Google buys Motorola Mobility for \$12,58 (VentureBeat) - This morning, Google announced that it will buy Motorola Mobility – Moto's mobile device arm – for \$12.5 billion. Google will acquire Motorola Mobility for \$40 per share in cash, a 63 percent premium over the company's Friday closing price. Google says it will run Motorola Mobility as a At separate business. Motorola spun off its business into two divisions last year, Ac Mobility and Solutions (the data and telecom portion), as a response to Google shares were down around 1.5 percent, while Motorola Mobility's stock jumped 57 percent. The company says Motorola Android phones won't \$7.1 declining profits. be receiving any special treatment as a consequence of the deal — but that's a tough nut to swallow, since Google often plays favorites. we soared from

# Events (2)

- Event:
  - Complex combination of relations linked to a set of empirical observations from texts
  - Can be defined as:
    - <subject> <predicate>
       e.g., <Person> <Resigns>
    - <subject> <predicate> <object> e.g., <Company> <Buys> <Company>
- Event extraction could be beneficial to IE systems:
  - Personalized news
  - Risk analysis
  - Monitoring
  - Decision making support

# Events (3)

- Common event domains:
  - Medical
  - Finance
  - Politics
  - Environment

#### **Event Extraction**

- In analogy with the classic distinction within the field of modeling, we distinguish 3 main approaches:
  - Data-driven event extraction:
    - Statistics
    - Machine learning
    - Linear algebra
    - ...
  - Expert knowledge-driven event extraction:
    - Representation & exploitation of expert knowledge
    - Patterns
  - Hybrid event extraction:
    - Combine knowledge and data-driven methods
- Our focus: expert knowledge-driven event extraction through the usage of pattern languages

## **Existing Approaches**

- Various pattern-languages for:
  - News processing frameworks (e.g., PlanetOnto)
  - General purpose frameworks (e.g., CAFETIERE, KIM, etc.)
- Language types:
  - Lexico-syntactic
  - Lexico-semantic
- However:
  - Insufficient use of semantics (if any)
  - Limited syntax
  - Cumbersome in use

#### **Semantics**

- Semantic Web:
  - Collection of technologies that express content meta-data
  - Offers means to help machines understand human-created data on the Web
- Ontologies:
  - Can be used to store domain-specific knowledge in the form of concepts (classes + instances)
  - Also contain inter-concept relations

## Pattern Language (1)

- Basic syntax:
  - LHS :- RHS
  - LHS: subject, predicate, object (optional)
  - RHS: pattern in which subject and object are assigned:
    - Literals (text strings)
    - Lexical categories (nouns, prepositions, verbs, etc.)
    - Orthographic categories (capitalization)
    - Labels (assigning subject and object)
    - Logical operators (and, or, not)
    - Repetition (≥0, ≥1, 0-1, {min,max})
    - Wildcards (skip ≥0 or exactly 1 word)
    - Ontological concepts

### Pattern Language (2)

• Provocation example (lexico-syntactic):

### Pattern Language (3)

• Provocation example (lexico-semantic):

```
($sub, kb:provokes, $obj) :-
  $sub:=([kb:Country] | [kb:Continent] | [kb:Union])
  (!`.' & !`(' & !`)' & !`-'){0,3}
  (kb:toAnger | kb:toAccuse | kb:toInsult |
    kb:toProvoke | kb:toThreaten)
  (!`.' & !`(' & !`)' & !`-'){0,3}
  $obj:=([kb:Country] | [kb:Continent] | [kb:Union])
```

## Implementation (1)

- The Hermes News Portal (HNP) is a stand-alone Java-based news personalization tool
- We have implemented the Hermes Information Extraction Engine (HIEE) within the HNP
- Pipeline-architecture is based on GATE components

### **Implementation (2)**



**Erasmus University Rotterdam** 





## **Implementation (3)**



Hermes Nexy Portal v2.0     Etome [										
Itome       Original graph       Search graph       Beaults       Recommendations       Import news       Rule Editor       Users       Evaluation <ul> <li>Rule Editor</li> <li>Annotation Validation</li> <li>Manual Annotations</li> <li>Evaluation</li> </ul> Rule Editor <ul> <li>Annotation Validation</li> <li>Manual Annotations</li> <li>Evaluation</li> </ul> Rule Editor <ul> <li>Manual Annotations</li> <li>Evaluation</li> </ul> Rule Editor              Annotation Validation              Manual Annotations               Evaluation                 Rule Editor              Rule Stocovery               Person Company CEO             Person Company CEO             Person CED IN Company             Person CED IN Company             Person CED IN Company             Person Wildcard Dercon Recome CEO <li>Company CEO Person             </li> <li>Person State Relation Rule             </li> <li>Rule Name: Company CEO Person             </li> <li>Rule Name: Company CEO Person             </li> <li>Person Company (Licad Recon Recome CEO             </li> <li>Recovery             </li> <li>Person Vildcard CEO, \$obj):=-             </li> <li>Sault Relation Rule             </li> <li>Rule Name: Company CEO Person             </li> <li>Person Company (Licad Recon Recome CEO</li>	Mermes News Portal v2.0									
✓ Rules Editor       ✓ Annotation Validation       ✓ Evaluation         ✓ Rule Groups       Rules         CEO Discovery       Person Company (EO Person         Product Discovery       Person Company (EO Discovery)         Partner Discovery       Person CEO IN Company         Partner Discovery       Person CEO IN Company (EO Person         Person CEO IN Company (EO Person       ✓ Company Wildcard CEO Wildcard Person         Person Discovery       Person CEO IN Company         Partner Discovery       Add ★ Remove         ✓ Rule Name:       Company CEO Person         Company Wildcard CEO Wildcard Person       © create Concept Rule         Rule Name:       Company CEO Person         Rule Name:       Company CEO Person         Rule Description:       © Create Concept Rule         (\$ sub, kb:hasCEO, \$obj) := [kb:Company] ('\'s')       Insert Concept         ', ') ? kb:CEO ', '? \$obj:=[kb:Person]       © Insert ON         © Insert Literal       () Combine         (1) put       Output       (1) Insert Range         ✓ Validate Syntax       Save Rule         ✓ Validate Syntax       Save Rule	Home Original graph Search graph Res	sults Recommendations Import news Rule Editor Users Evaluation								
Rule Groups         ESD Biscovery         Product Discovery         Shares Discovery         Person COD IN Company CEO Person         Person COD IN Company CEO Person         Person Obscovery         Postic Discovery         Partner Discovery         Sales Discovery         Company Let Description:         Editor         Rule Description:         (\$ sub, kb:hasCEO, \$ sobj) := \$ sub:= [kb:Company] ('\'s'           () Combine         () Combine         () Combine         () Combine         () Linsert Range         Validate Syntax         Save Rule	📝 Rules Editor 🖌 Annotation Validation 🔞 Manual Annotations 🗸 Evaluation									
Rules   CEO Discovery   Shares Discovery   Company CEO Person   Person Company CEO   Person CEO IN Company   Person CEO IN Company   Company Midcard CEO Wildcard Ceo Wildcard Ceo Midcard Person   Company Wildcard CEO Wildcard CEO Wildcard Ceo Midcard Person   Company Wildcard CEO Wildcard CEO Wildcard Ceo Midcard Person   Subsidiary Discovery   Partner Discovery   Sales Discovery   President Discovery   Sales Discovery   Company KEO   Person Company CEO Person   © Create Concept Rule   Rule Name:   Company CEO Person   © Create Relation Rule   (\$ sub, kb:hasCEO, \$obj) := \$sub:=[kb:Company] ('\'s')   © Insert Orth   © Insert Range   ✓ Validate Syntax   Save Rule										
Company CEO Person       Company CEO         Product Discovery       Person Company CEO         Port Discovery       Person Company CEO         Partner Discovery       Person Company CEO Person         Company Widcard CEO Widcard CEO Widcard CEO       Company Widcard CEO Widcard CEO         Sales Discovery       Add	Rule Groups Rules									
Preside: Discovery       Person CED IN Company         Person CED IN Company       Person CED OWIdcard CEO Wildcard Company         Company Wildcard CEO Wildcard Person       Company Wildcard CEO Wildcard Person         Parther: Discovery       Subsidiary Discovery         Preside: Discovery       Add ★ Remove         Image: Subsidiary Discovery       Image: Subsidiary Discovery         Parther: Discovery       Subsidiary Discovery         Preside: Discovery       Image: Subsidiary Discovery         Sales Discovery       Rule Name: Company CEO Person         Rule Description:       Image: Subsidiary Discovery         (\$ sub, kb:hasCEO, \$obj) := \$sub:=[kb:Company] ('\'s']         Image: Subsidiary Discovery       Image: Subsidiary Discovery         Sales Discovery       Subsidiary Discovery         Sales Discovery       Subsidiary Discovery         Image: Subsidiary Discovery       Image: Subsidiary Discovery         Image: Subsidiary Discovery       Image: Subsidiary Discovery         Sales Discovery       Sales Discovery         Image: Subsidiary Discovery       Image: Subsidiary Discov	CEO Discovery	<u>^</u>								
Competitor Discovery Person Wildcard CED Wildcard Company Company Wildcard CED Wildcard Cerson Recome CEO Partner Discovery President Discovery Subsidiary Discovery President Discovery Sales Discovery Sales Discovery Fresident Discovery Sales Discovery Company Wildcard CED Wildcard Person Company CEO Person Rule Description: (\$ create Concept Rule Rule Description: (\$ create Relation Rule (\$ sub, kb:hasCEO, \$obj):= \$sub:=[kb:Company] ('\'s'] Insert Concept (\$ sub, kb:hasCEO, \$obj):= \$sub:=[kb:Company] ('\'s'] Dutput (\$ nearer Range V Validate Syntax Save Rule V Validate Syntax Save Rule	Shares Discovery	Person CEO IN Company	=							
Profit Discovery         Loss Discovery         Parther Discovery         Subsidiary Discovery         President Discovery         Sales Discovery         Sales Discovery         Rule Description:         (\$aub, kb:hasCE0, \$obj) :- \$sub:=[kb:Company] ('\'s')         Insert Concept         (\$sub, kb:hasCE0, \$obj) :- \$sub:=[kb:Company] ('\'s')         Insert Concept         Output         Output         Output         Validate Syntax         Save Rule	Competitor Discovery	Person Wildcard CEO Wildcard Company								
Loss Discovery       Partner Discovery         Partner Discovery         President Discovery         Sales Discovery         Sales Discovery         Gales Discovery         Bales Discovery         Gales Discovery         Bales Discovery         Gales Discovery     <	Profit Discovery	Company Wildcard CEO Wildcard Person								
Partner Discovery         Subsidiary Discovery         Sales D	Loss Discovery	Company Wildcard Person Become CEO	*							
Subsidiary Discovery         President Discovery         Sales Discovery         Editor         Rule Name:         Company CEO Person         @ Create Concept Rule         Rule Description:         (\$sub, kb:hasCEO, \$obj) :- \$sub:=[kb:Company] ('\'s']         Insert Concept         (\$sub, kb:hasCEO, \$obj):- \$sub:=[kb:Company] ('\'s']         Insert Concept         ', ') ? kb:CEO ', '? \$obj:=[kb:Person]         Insert Uiteral         () Combine         () Combine         () Combine         () Insert Range         Validate Syntax         Save Rule	Partner Discovery	🕂 Add 🗙 Remove \Lambda Up 🔸 Down								
Fieldor Rule Discovery       Editor         Rule Name:       Company CEO Person         Rule Description:	Subsidiary Discovery President Discovery									
Rule Name:       Company CEO Person       ⓒ Create Concept Rule         Rule Description:       ⓐ Create Relation Rule         (\$sub, kb:hasCEO, \$obj) :- \$sub:=[kb:Company] ('\'s']       ⓐ Insert Concept         ',')?       kb:CEO ','? \$obj:=[kb:Person]       ⓐ Insert Concept         Output       ⓐ Insert Literal       () Combine         () Combine       ⓐ Insert Range         ✓ Validate Syntax       ⑤ Save Rule	Sales Discovery	Editor								
Rule Description:     (\$sub, kb:hasCEO, \$obj) :- \$sub:=[kb:Company] ('\'s']   ',')? kb:CEO ','? \$obj:=[kb:Person]     Insert Concept   Insert OS   Insert Literal   () Combine   () Combine   () Combine   () Save Rule     ✓ Validate Syntax   Save Rule		Rule Name: Company CEO Person	Create Concept Rule							
<pre>\$</pre>		Rule Description:	R Create Relation Rule							
<ul> <li>,',')? kb:CEO ','? \$obj:=[kb:Person]</li> <li>Insert Orth</li> <li>Insert POS</li> <li>Insert Literal</li> <li>() Combine</li> <li>(1) Combine</li> <li>(2) Insert Range</li> <li>✓ Validate Syntax</li> <li>Save Rule</li> </ul>		(\$sub, kb:hasCEO, \$obj) :- \$sub:=[kb:Company] ('\'s'								
<ul> <li>✓ Literal</li> <li>✓ Combine</li> <li>✓ Validate Syntax</li> <li>✓ Save Rule</li> </ul>		Insert Orth								
<ul> <li>✓ Add ✓ Edit Name ★ Remove</li> <li>✓ Up ↓ Down</li> </ul>			Insert POS							
<ul> <li>✓ Insert Literal</li> <li>() Combine</li> <li>(a) Insert Range</li> <li>✓ Validate Syntax</li> <li>☑ Save Rule</li> </ul>										
Cutput C			The sert Literal							
Cutput Cu			() Combine							
✓ Validate Syntax ✓ Validate Syntax Save Rule ✓ Up ↓ Down		Output	{*} Insert Range							
Image: Add / Edit Name X Remove       ↑ Up ↓ Down			🖌 Validate Syntax							
Image: Add / Edit Name X Remove       ↑ Up ↓ Down			Save Rule							
Image: Add / Edit Name X Remove       ↑ Up ↓ Down										
Image: Add / Edit Name     Image: Remove       ↑ Up     ↓ Down										
Add  ✓ Edit Name										
↑ Up ↓ Down	🕂 Add 🥖 Edit Name X Remove									
	🛧 Up 🖖 Down									

## **Implementation (4)**

Hermes News Portal v2.0						• ×			
Home Original graph Search graph Results R	ecommendations Import r	news Rule Editor User	s Evaluation						
📝 Rules Editor 🖌 Annotation Validation 🗔 I	Manual Annotations 🛛 🗸 E	valuation							
Dup!      Perlassify     Keen History for Evaluat	tion Sava Appatations	Close All Appetation							
	aon   🔽 Save Annotations	Ciear Air Annotation	15						
Patterns Found									
Subsidiary Discovery Subject Relation Object Times Found Valid									
Shares Discovery	kh:Cooolo	khthacCEO	kh Eric Schmidt	1		-			
Sales Discovery	kb:Google	kb.hasCEO	kb.Enc_Scrimidt						
Profit Discovery	kb:Comoloft	kb.hasCEO	kb:Rober (A. Iger	1		-			
Product Discovery	kb.samelort	kb.hasCEO	kb.Michel_Guilemot	1					
President Discovery	KD:DP	kD:nasCEO	KD:Tony_hayward	1		_			
Partner Discovery	KD: Alcatel-Lucent	KD:nasCEO	KD:Ben_verwaayen	1		- =			
Loss Discovery	kD:Viacom	kb:hasCEO	kb:Philippe_PDauman	1		_			
Competitor Discovery	KDINIR	kD:hasCEO	kb:Jim_Barry	1		_			
CEO Discovery	kb:Coke	kb:hasCEO	kb:Muhtar_Kent	1		_			
	kb:Picarro	kb:hasCEO	kb:Woelk	1					
	kb:Microsoft	kb:hasCEO	kb:Steven_ABallmer	1					
kb:National_Arts_Council kb:hasCEO kb:Benson_Puah 1									
kb:British_Airways kb:hasCEO kb:Willie_Walsh 1									
kb:Royal_Dutch_Shell kb:hasCEO kb:Peter_Voser 1									
	kb:Merck	kb:hasCEO	kb:Richard_TClark	1					
	kb:BP	kb:hasCEO	kb:Anthony_B_Hayward	1		<b>T</b>			
News Items									
Past Clouds Future of Europe's New	Past Clouds Futu	re of Europe's New	Antitrust Chief						
Wed Mar 03 02:26:22 CET 2010	clouding bis now re	Job of and anciarable . Jallia jubich ha must	menter and carepeants	untries from broaching	rulas on				
	clouding his new ro	ne, in which he musc	keep companies and co	uncries from breaching	rules on				
	Market-ngging and	government subsidies	on for borgoining long o	ssor as antitrust comm	issioner, Neelle				
	Kroes of the Netherlands, had a reputation for bargaining long and hard with powerful people like Steven								
A. Bailiner, chief executive or microsort. At the neight of the financial crisis, she insisted that fulles on									
	like the Dutch bark ING and the Povel take of Scathad ware required to mike the circlinest divertifying								
ince the butten bank into and the koyai Bank or Scotland were required to make significant divestitures.									
Now, the question is whether Mr. Amund will be up to his new job's challenges, including the									
	protectionist tendencies rising in conjunction with unemployment in many European countries and the								
potential for trans-Atlantic tensions, especially as companies like Google Continue to expand globally. His									
spokeswoniari, Ameria Tones, salu mi, Amuria woun not grant interviews so early if this fieldly live-year									
	cerm as competitio	n commissioner, Last i	nonuay, in his hist publi	c appearance in the ro	e, mr. Aimuma				
L									

# **Evaluation (1)**

We compare the performance of lexico-syntactic rules with lexico-semantic rules on two data sets:

President

Revenue

- Economic events (500 news messages):
  - CEO

- Profit
- Product
- Loss
- Shares
   Partner
- Competitor
   Subsidiary
- Political events (100 news messages):

Sanction

Election
 Resignation
 Provocation

Investment
 Help

Visit

Riots

Join

- Collaboration
- Performance is evaluated based on rule creation times and precision, recall, and  $F_1$ -measure.

## **Evaluation (2)**

• Creation times for  $F_1 \ge 0.5$ : 90% improvement

Name	Lex-Syn	Lex-Sem	Name	Lex-Syn	Lex-Sem
CEO	8424	281	Election	1517	232
Product	9428	132	Visit	4238	543
Shares	2403	648	Sanction	4013	419
Competitor	9116	133	Join	3986	297
Profit	1923	416	Resignation	1259	366
Loss	5991	313	Investment	5162	781
Partner	4924	185	Riots	1734	306
Subsidiary	6620	776	Collaboration	1103	137
President	4239	179	Provocation	1428	530
Revenue	5317	498	Help	1987	211
Overall	5839	356	Overall	2643	382

## **Evaluation (3)**

	Lex-Syn			Le	ex-Sem	
Name	Precision	Recall	$F_1$	Precision	Recall	F <sub>1</sub>
CEO	0.5217	0.6000	0.5581	0.8966	0.8667	0.8814
Product	0.6667	0.4118	0.5091	0.8607	0.7721	0.8140
Shares	0.4286	0.6667	0.5218	0.9000	0.8000	0.8471
Competitor	0.5333	0.4800	0.5052	0.7600	0.7600	0.7600
Profit	0.7500	0.4545	0.5660	0.8800	0.6667	0.7586
Loss	0.6471	0.4074	0.5000	0.8125	0.4815	0.6047
Partner	0.3864	0.7391	0.5075	0.8000	0.8696	0.8333
Subsidiary	0.7500	0.3913	0.5143	0.9063	0.6304	0.7436
President	0.4333	0.5909	0.5000	0.6667	0.6364	0.6512
Revenue	0.6429	0.4091	0.5000	0.7143	0.6818	0.6977
Overall	0.5492	0.4935	0.5199	0.8390	0.7414	0.7872

Scores after maximum time (finance)

## **Evaluation (4)**

•	Scores	after	maximum	time	(politics)	
---	--------	-------	---------	------	------------	--

	L	ex-Syn		Le	ex-Sem	
Name	Precision	Recall	F <sub>1</sub>	Precision	Recall	<b>F</b> <sub>1</sub>
Election	0.4615	0.5455	0.5000	1.0000	0.8182	0.9000
Visit	0.6774	0.4038	0.5060	0.7027	0.5000	0.5843
Sanction	0.5263	0.4918	0.5085	0.7857	0.7213	0.7521
Join	0.7222	0.4063	0.5200	0.8125	0.8125	0.8125
Resignation	0.9091	0.3846	0.5405	0.8000	0.9231	0.8572
Investment	0.5208	0.4808	0.5000	0.7778	0.6731	0.7217
Riots	0.5200	0.5200	0.5200	0.7069	0.8200	0.7593
Collaboration	0.4250	0.6538	0.5151	0.7143	0.7692	0.7407
Provocation	0.7727	0.4250	0.5484	0.7857	0.5714	0.6616
Help	0.5510	0.4737	0.5094	0.7541	0.8070	0.7797
Overall	0.5664	0.4694	0.5134	0.7630	0.7164	0.7390

#### Conclusions

- Compared to lexico-syntactic alternatives, our lexicosemantic patterns perform better in terms of precision, recall, F<sub>1</sub>, and creation times
  - Also, our rules are more expressive and easier to use than their lexico-syntactic and lexico-semantic alternatives
  - Future work:
    - Implement language in new IE pipeline
    - Evaluate on big data set
    - Link events to trading algorithms instead of news personalization

#### Questions



