



Generating Templates of Entity Summaries with an Entity-Aspect Model and Pattern Mining

Peng Li, Jing Jiang, Yinglin Wang Shanghai Jiao Tong University Singapore Management University

An Example Entity Summary



Main page
Contents
Featured content
Current events
Random article

- Interaction
 About Wikipedia
 Community portal
 Recent changes
 Contact Wikipedia
 Donate to Wikipedia
 Help
- Toolbox
- Print/export
- Languages
 Bân-lâm-gú
 Afrikaans
- *Alemannisch

العربية⊕

*Aragonés

Albert Einstein

From Wikipedia, the free encyclopedia

"Einstein" redirects here. For other uses, see Einstein (disambiguation).

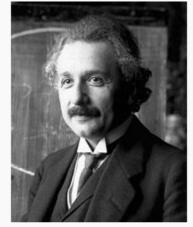
Albert Einstein (pronounced /'ælbert 'aɪnstaɪn/; German: ['albet 'aɪnʃtaɪn] (🌓 listen); 14 March 1879 – 18 April 1955) was a theoretical physicist, philosopher and author who is widely regarded as one of the most influential and best known scientists and intellectuals of all time. He is often regarded as the father of modern physics. [3] He received the 1921 Nobel Prize in Physics "for his services to Theoretical Physics, and especially for his discovery of the law of the photoelectric effect." [4]

His many contributions to physics include the special and general theories of relativity, the founding of relativistic cosmology, the first post-Newtonian expansion, the explanation the perihelion precession of Mercury, the prediction of the deflection of light by gravity (gravitational lensing), the first fluctuation dissipation theorem which explained the Brownian motion of molecules, the photon theory and the wave-particle duality, the quantum theory of atomic motion in solids, the zero-point energy concept, the semiclassical version of the Schrödinger equation, and the quantum theory of a monatomic gas which predicted Bose–Einstein condensation.

Einstein published more than 300 scientific and over 150 non-scientific works; he additionally wrote and commentated prolifically on various philosophical and political subjects.^[5] His great intelligence and originality has made the word "Einstein" synonymous with genius.^[6]

Contents [hide]

Albert Einstein



Albert Einstein, 1921

Born 14 March 1879

Ulm, Kingdom of Württemberg, German

Empire

Died 18 April 1955 (aged 76)

Princeton, New Jersey, USA

Resting Grounds of the Institute for Advanced place Study, Princeton, New Jersey.





An Example Entity Summary

Albert Einstein (pronounced /ˈælbərt ˈaɪnstaɪn/; German ˈaɪnʃtaɪn] (📢 listen); 14 March 1879 – 18 April 1955) was ... received ... Nobel Prize ...

physicist, philosopher and author who is widely regarded as a most influential and best known scientists and intellectuals of the me. He is often

His ... contributions ... include ... He received the 1921 Nobel tical Physics, and especially for

his discovery the law of the photoelectric effect."[4]

His many contributions to physics include the special and general theories of relativity, the founding of relativistic cosmology, the first post-Newtonian expansion, the explanation the perihelion precession of Mercury, the prediction of the deflection of light by gravity (gravitational lensing), the first fluctuation dissipation theorem which explained the Brownian motion of molecules, the photon the theory of atomic motion in semiclassical version of the deflected Bose-Einstein condensation.

Einstein published more than 300 scientific and over 150 non-scientific works; he additionally wrote and commentated prolifically on various philosophical and political subjects. [5] His great intelligence and originality has made the word "Einstein" synonymous with genius. [6]



Entity Summaries in the Same Category



 Interaction About Wikipedia Community portal Recent changes Contact Wikipedia Donate to Wikipedia

▶ Toolbox

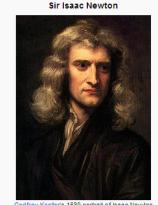
Isaac Newton

From Wikipedia, the free encyclopedia

Sir Isaac Newton FRS (4 January 1643 - 31 March 1727 [OS: 25 December 1642 - 20 March 1726])[1] was an English physicist, mathematician, astronomer, natural philosopher, alchemist, and theologian who is considered by many scholars and members of the general public to be one of the most influential people in human history. His 1687 publication of the Philosophiæ Naturalis Principia Mathematica (usually called the Principia) is considered to be among the most influential books in the history of science, laying the groundwork for most of classical mechanics. In this work. Newton described universal gravitation and the three laws of motion which dominated the scientific view of the physical universe for the next three centuries. Newton showed that the motions of objects on Earth and of celestial bodies are governed by the same set of natural laws by demonstrating the consistency between Kepler's laws of planetary motion and his theory of gravitation, thus removing the last doubts

Newton built the first practical reflecting telescope [7] and developed a theory of colour based on the observation that a prism decomposes white light into the

about heliocentrism and advancing the scientific revolution.



Godfrey Kneller's 1689 portrait of Isaac Newton (aged 46)

[OS: 25 December 1642][1]

Albert Einstein

Albert Finstein 1921

14 March 1879

Kingdom of Württemberg, German

oril 1955 (aged 76) ceton, New Jersey, USA

unds of the Institute for Advanced ly, Princeton, New Jersey

WikipediA The Free Encyclopedia

Main page Contents Featured content Current events Random article

 Interaction About Wikipedia Community portal Recent changes Contact Wikipedia Donate to Wikinedia

▶ Toolbox

▶ Print/export

 Languages Afrikaans Azərbaycan Bân-lâm-qú Bosanski

Stephen Hawking

From Wikipedia, the free encyclopedia

Stephen William Hawking, CH, CBE, FRS, FRSA (born 8 January 1942)[1] is a British theoretical physicist, whose scientific career spans over forty years. His books and public appearances have made him an academic celebrity and he is an Honorary Fellow of the Royal Society of Arts. [2] a lifetime member of the Pontifical Academy of Sciences. [3] and in 2009 was awarded the Presidential Medal of Freedom, the highest civilian award in the

Hawking was the Lucasian Professor of Mathematics at the University of Cambridge for thirty years, taking up the post in 1979 and retiring on 1 October 2009. [5][6] He is also a Fellow of Gonville and Caius College, Cambridge and a Distinguished Research Chair at the Perimeter Institute for Theoretical Physics in Waterloo, Ontario. [7] He is known for his contributions to the fields of cosmology and quantum gravity, especially in the context of black holes. He has also achieved success with works of popular science in which he discusses his own theories and cosmology in general; these include the runaway best seller A Brief History of Time, which stayed on the British Sunday Times bestsellers list for a record-breaking 237 weeks.[8][9]

Hawking's key scientific works to date have included providing, with Roger Penrose, theorems regarding gravitational singularities in the framework of general relativity, and the theoretical prediction that black holes should emit radiation, which is today known as Hawking radiation (or sometimes as Bekenstein-Hawking radiation).^[10]

Hawking has a neuro-muscular dystrophy that is related to amyotrophic lateral sclerosis, a condition that has progressed over the years and has left him almost completely paralysed.

Contents (hide)

1 Early life and education

Institutions

Stephen William Hawking

Q w

lopedia

Stephen Hawking at NASA in 1999 Stephen William Hawking 8 January 1942 (age 68) Oxford, England, UK Residence England UK

Nationality Fields

Applied mathematician Theoretical physicist University of Cambridge

Niels Bohr ctober 1885 - 18 sh physicist who ons to understanding n mechanics, for Prize in Physics in laborated with many entury at his was part of a team Manhattan Project. and in 1912 and grew up to be an 375 also received en described as one ists of the 20th

Born

Died

Niels Henrik David Bohr 7 October 1885 Copenhagen, Denmark X 18 November 1962 (age

SINGAPORE MANAGEMENT UNIVERSITY

Shanghai Jiao Tong University

Entity Summaries in the Same Category

	Aspect	Representative Sentence Patterns				
	Education	X received his PhD from University X studied under X earned his in physics from University of	<u> </u>			
WikipediA	Awards	X was awarded the medal in X won the award X received the Nobel Prize in physics in				
The Free Encyclopedia Main page Contents Featured content Current events Random article Interaction About Wikipedia Community portal Recent changes Contact Wikipedia Donate to Wikipedia Help Toolbox Print/export	Career	X was director X was the head of X worked for	rg, German A Advanced			
	Contributions	X made contributions to X is best known for work on X is noted for	ey.			
Languages Afrikaans العربية Azerbaycan काला Bân-lâm-gú Bosanski Български	theoretical prediction that black holes should emi Hawking radiation (of Hawking has a neuro a condition that has paralysed. Co 1 Early life and educt at Jiao Tong University	task is to automatically generate such entity summary templates.	J			

Why Is It Useful?

- Better organizes information units
 - Wikipedia infoboxes
- Provides a structured template for humans to create new entity summaries
- Facilitates automatic entity summary generation





Born Niels Henrik David Bohr

7 October 1885 Copenhagen, Denmark

Died 18 November 1962 (aged 77)

Copenhagen, Denmark

Nationality Denmark
Fields Physics

Institutions University of Copenhagen

University of Cambridge University of Manchester

Alma mater University of Copenhagen

Doctoral advisor

r Christian Christiansen

Other academic advisors

J. J. Thomson Ernest Rutherford

Doctoral students Hen

Hendrik Anthony Kramers

Known for Copenhagen interpretation

Complementarity





Outline

- Task and motivation
- Related work
- Our approach
 - Sentence clustering
 - Pattern mining
- Evaluation
- Conclusions





Related Work

- Filatova et al. (2006) "Automatic creation of domain templates"
 - Patterns must contain a non-auxiliary verb
 - Patterns are not clustered into aspects
 - Pattern slots are identified through heuristics
- Sauper & Barzilay (2009) "Automatically generating Wikipedia articles"
 - Long, comprehensive articles
- LDA extensions
 - Chemudugunta et al. (2007)
 - Titov & McDonald (2008)
 - Daume III & Marcu (2006), Haghi & Vanderwende (2009)



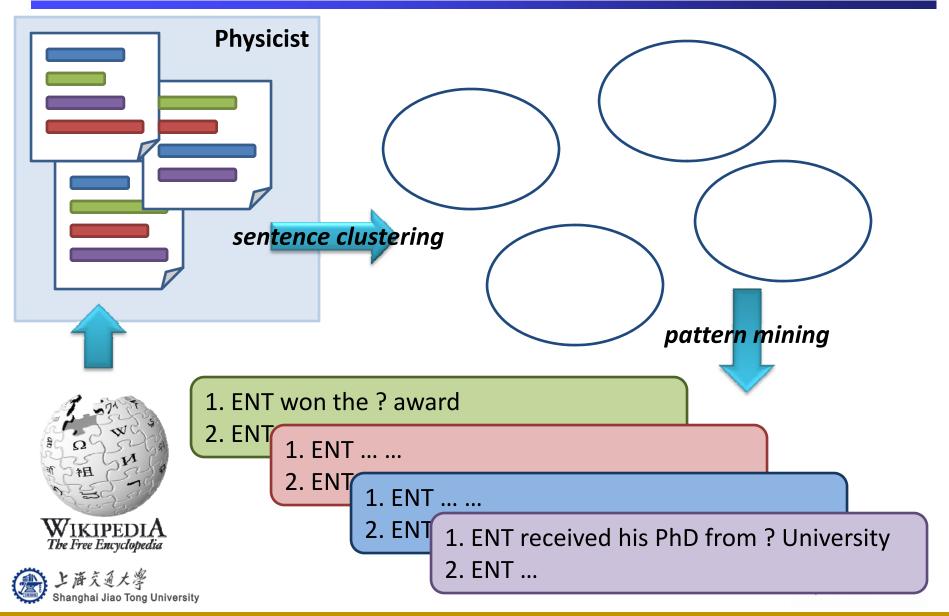
Outline

- Task and motivation
- Related work
- Our approach
 - Sentence clustering
 - Pattern mining
- Evaluation
- Conclusions





Overview of Our Approach



Outline

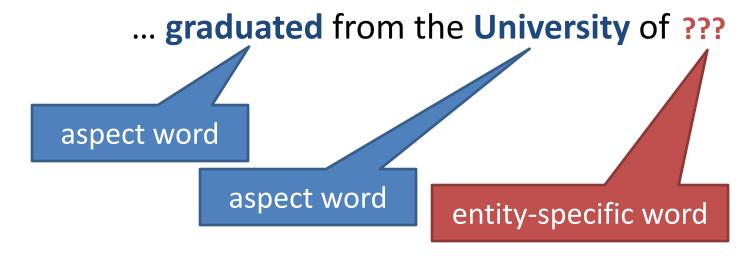
- Task and motivation
- Related work
- Our approach
 - Sentence clustering
 - Pattern mining
- Evaluation
- Conclusions





Sentence Clustering

- To group sentences related to the same aspect together
- To distinguish between aspect words and entity-specific words







... ... Venturi was a professor of physics at the University of Modena

... ... He was a professor of physics at the University of Chicago until 1982





```
... ... Venturi was a professor of physics at the University of Modena ... ...
```

... ... He was a professor of physics at the University of Chicago until 1982





... ... <u>Venturi</u> was a professor of physics at the University of <u>Modena</u>

venturi modena

Venturi

... ... He was a professor of physics at the University of Chicago until 1982

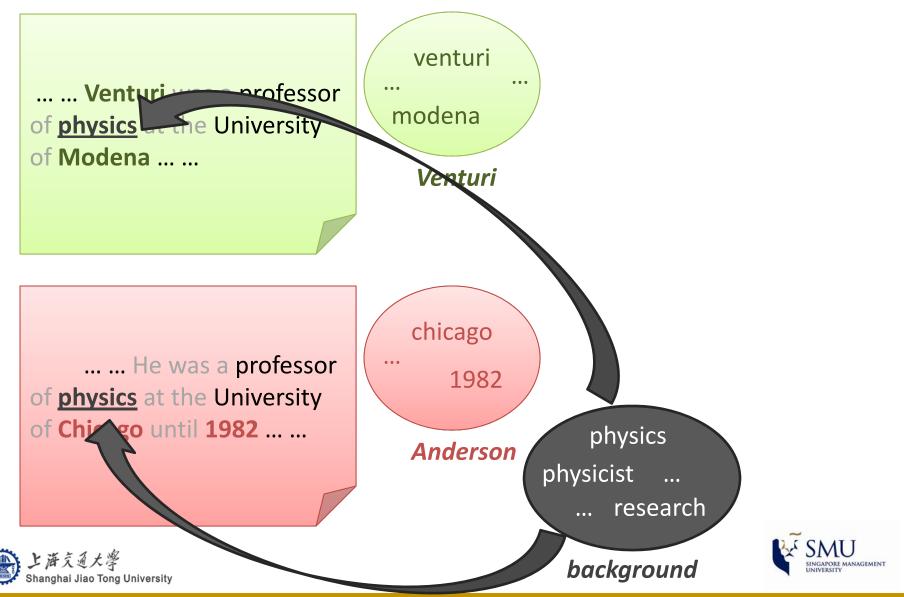
chicago

1982

Anderson







... ... <u>Venturi</u> was a <u>professor</u> of <u>physics</u> at the <u>University</u> of <u>Modena</u>

venturi modena

Venturi

professor institute ... university

affiliation

of <u>physics</u> at the <u>University</u> of <u>Chicago</u> until <u>1982</u>

chicago

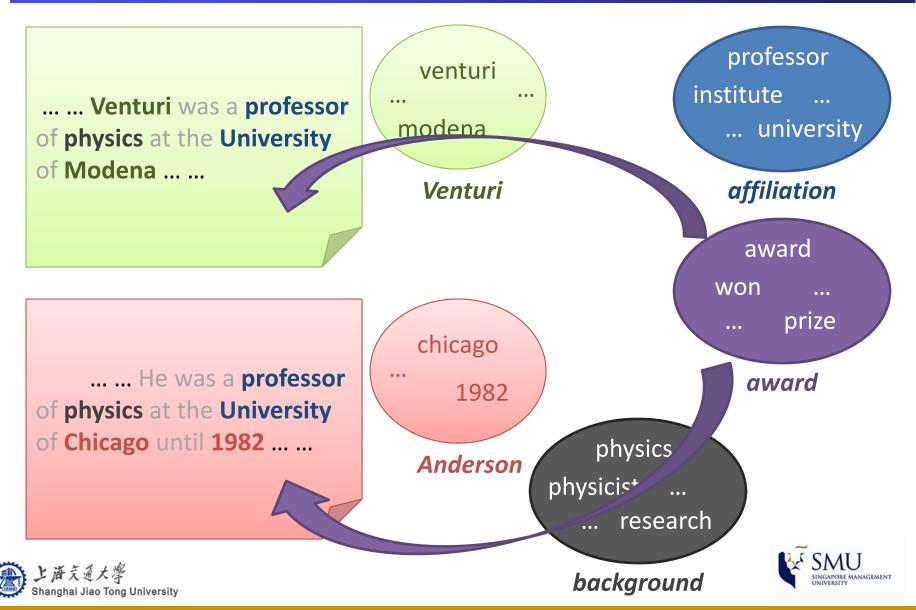
1982

Anderson

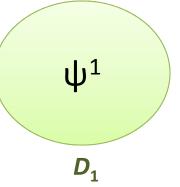
physics
physicist ...
... research

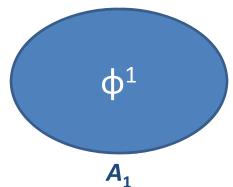
background





... ... Venturi was a professor of physics at the University of Modena





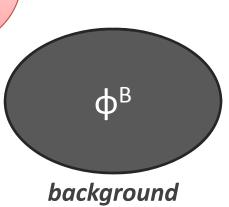
...... He was a professor of physics at the University of Chicago until 1982



Φ2

 A_2

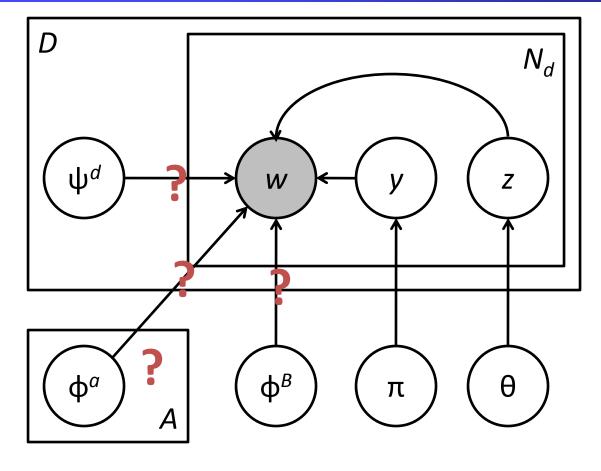
 D_2







Entity-Aspect Model

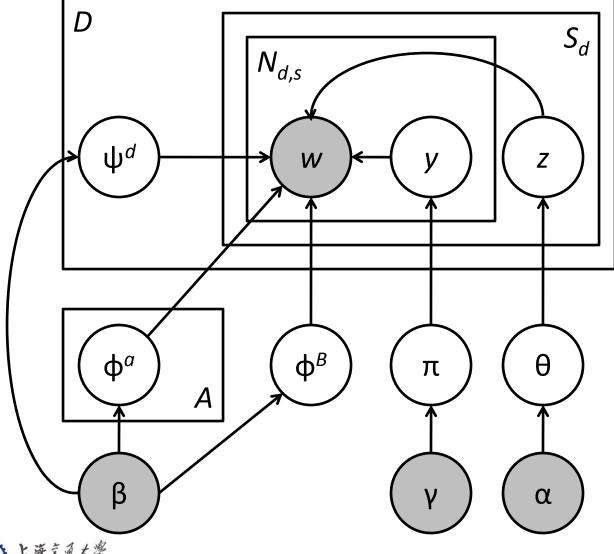


 $y \in \{1, 2, 3\}$ $z \in \{1, ..., A\}$





Entity-Aspect Model



 $y \in \{1, 2, 3\}$

 $z \in \{1, ..., A\}$



Model Inference

Gibbs sampling

$$p(z_{d,s} = a | \boldsymbol{z}_{\neg\{d,s\}}, \boldsymbol{y}, \boldsymbol{w})$$

$$\propto \frac{C_{(a)}^{\mathcal{A}} + \alpha}{C_{(\cdot)}^{\mathcal{A}} + A\alpha} \cdot \frac{\prod_{v=1}^{V} \prod_{i=0}^{E_{(v)}} (C_{(v)}^{a} + i + \beta)}{\prod_{i=0}^{E_{(\cdot)}} (C_{(\cdot)}^{a} + i + V\beta)}$$

$$p(y_{d,s,n} = 1 | \boldsymbol{z}, \boldsymbol{y}_{\neg\{d,s,n\}}) \propto \frac{C_{(1)}^{\pi} + \gamma}{C_{(\cdot)}^{\pi} + 3\gamma} \cdot \frac{C_{(w_{d,s,n})}^{\mathcal{B}} + \beta}{C_{(\cdot)}^{\mathcal{B}} + V\beta}$$
$$p(y_{d,s,n} = 2 | \boldsymbol{z}, \boldsymbol{y}_{\neg\{d,s,n\}}) \propto \frac{C_{(2)}^{\pi} + \gamma}{C_{(\cdot)}^{\pi} + 3\gamma} \cdot \frac{C_{(w_{d,s,n})}^{d} + \beta}{C_{(\cdot)}^{d} + V\beta}$$
$$p(y_{d,s,n} = 3 | \boldsymbol{z}, \boldsymbol{y}_{\neg\{d,s,n\}}) \propto \frac{C_{(3)}^{\pi} + \gamma}{C_{(\cdot)}^{\pi} + 3\gamma} \cdot \frac{C_{(w_{d,s,n})}^{a} + \beta}{C_{(\cdot)}^{a} + V\beta}$$





Clustered Sentences

Venturi/D was/S a/S professor/A of/S physics/B at/S the/S
University/A of/S Modena/D ./S
He/S was/S a/S professor/A of/S physics/B at/S the/S University/A of/S Chicago/D until/S 1982/D ./S

S: stop word

B: background word

A: aspect word

D: document word





Outline

- Task and motivation
- Related work
- Our approach
 - Sentence clustering
 - Pattern mining
- Evaluation
- Conclusions





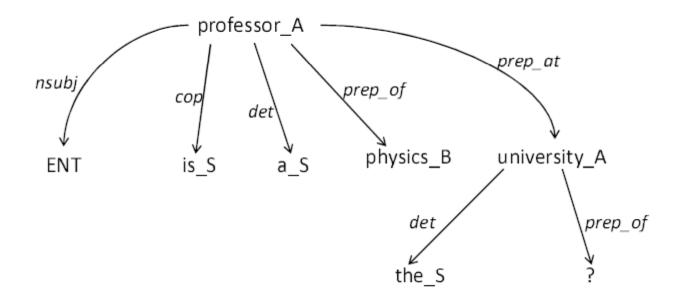
Pattern Mining

- Use heuristics to locate subject entities
 - Title of the Wikipedia article
 - Top 3 frequent subject noun phrases in the article
- Generate labeled dependency parse trees
 - Replace document words with "?"
- Mine frequent subtree patterns
- Prune patterns
 - Remove patterns with no subject entity
 - Remove patterns with no aspect word
- Convert subtree patterns to sentence patterns





A Labeled Dependency Parse Tree







Sample Aspects and Patterns

Aspect	Sample Sentence Patterns
1	X received his PhD from University X studied under X earned his in physics from University of
2	X was awarded the medal in X won the award X received the Nobel Prize in physics in
3	X was director X was the head of X worked for
4	X made contributions to X is best known for work on X is noted for





Outline

- Task and motivation
- Related work
- Our approach
 - Sentence clustering
 - Pattern mining
- Evaluation
- Conclusions





Data Set

• Wikipedia articles from 5 categories

Category	Documents	Sentences	Avg Sent/Doc
US Actress	407	1721	4
Physicist	697	4238	6
US CEO	179	1040	5
US Company	375	2477	6
Restaurant	152	1195	7





Quantitative Evaluation

- Sentence patterns
 - Manually judged whether each sentence pattern is meaningful for the given entity category
 - Allows us to compute precision, recall and f1
 - Baseline 1: pattern mining without entity-aspect model
 - Baseline 2: verb-based pruning (Filatova et al. 2006)
 on top of BL-1
- Aspect clusters
 - Manually grouped meaningful patterns into clusters
 - Allows us to compute purity





Quality of Sentence Patterns

		Category				
Method		US Actress	Physicist	US CEO	US Company	Restaurant
BL-1	precision	0.714	0.695	0.778	0.622	0.706
	recall	0.545	0.300	0.367	0.425	0.361
	f1	0.618	0.419	0.499	0.505	0.478
BL-2	precision	0.845	0.767	0.829	0.809	1.000
	recall	0.260	0.096	0.127	0.167	0.188
	f1	0.397	0.17	0.220	0.276	0.316
Ours	precision	0.544	0.607	0.586	0.450	0.560
	recall	0.710	0.785	0.712	0.618	0.701
	f1	0.616	0.684	0.643	0.520	0.624

- No consideration of aspect clusters.
- High precision of BL-1 and BL-2: They use a higher frequency threshold to select frequent sentence patterns.
- Low recall of BL-2: Many meaningful sentence patterns do not contain a non-auxiliary verb.
- BL-1 and BL-2 do not generate template slots.





Quality of Aspect Clusters

Category	B	Purity
US Actress	4	0.626
Physicist	6	0.714
US CEO	4	0.674
US Company	4	0.614
Restaurant	3	0.587





Sample Aspects and Their Representative Words

Method	Sample Aspects		
	1	2	3
Our	university	prize	academy
entity-	received	nobel	sciences
aspect	ph.d.	physics	member
model	college	awarded	national
	degree	medal	society
Standard	physics	nobel	physics
LDA	american	prize	institute
	professor	physicist	research
	received	awarded	member
	university	john	sciences
K-means	physics	physicist	physics
	university	american	academy
	institute	physics	sciences
	work	university	university
	research	nobel	new





Conclusions

- We proposed an LDA-based entity-aspect model to simultaneously cluster sentences and label words
- We used pattern mining to identify sentence patterns with slots
- We empirically evaluated our method and showed its effectiveness



Future Directions

- Use linguistic knowledge to further prune sentence patterns and improve their readability
- Automatic aspect labeling
- Automatic entity summary generation





THANK YOU!



