

Explass: Exploring Associations between Entities via Top-K Ontological Patterns and Facets

Gong Cheng, Yanan Zhang, Yuzhong Qu

Websoft Research Group

State Key Laboratory for Novel Software Technology

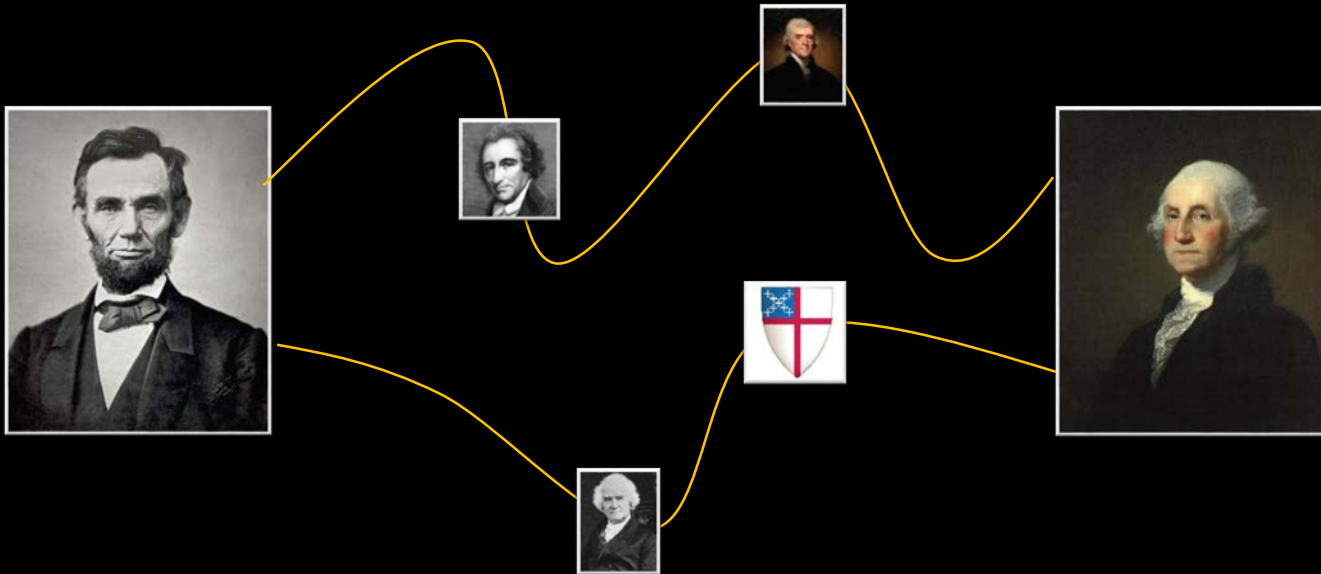
Nanjing University, China



Websoft

EXPLASS

Association search



Association search



air pollution



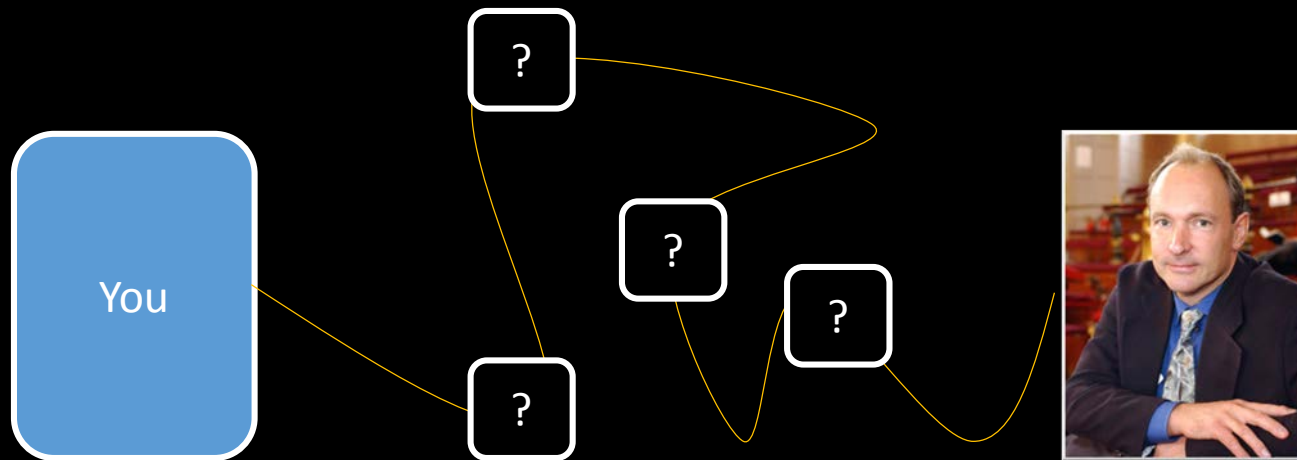
autism

?

?

?

Association search



Association search on the Web of documents

[The Alice and Bob After Dinner Speech - downlode.org](#)

[downlode.org](#) › [Etext Library](#) ▼

Firstly and most importantly they are concerned with the private lives of two people called **Alice** and **Bob**. In theory papers, whenever a coding theorist wants to ...

[Bob & Carol & Ted & Alice \(1969\) - IMDb](#)

[www.imdb.com/title/tt0064100/](#) ▼ [Internet Movie Database](#) ▼

★★★★★ Rating: 6.6/10 - 3,299 votes

Directed by Paul Mazursky. With Natalie Wood, Robert Culp, Elliott Gould, Dyan Cannon. Documentary film-maker **Bob** Saunders and his wife **Carol** attend a ...

[Alice and Bob in Wonderland | Perimeter Institute](#)

[www.perimeterinstitute.ca/.../ali...](#) ▼ [Perimeter Institute for Theoretical Physics](#) ▼

Alice and **Bob** in Wonderland: Can we travel through time? Wednesday Mar 28, 2012. (PDF). Collection/Series: **Alice** and **Bob** in Wonderland ...

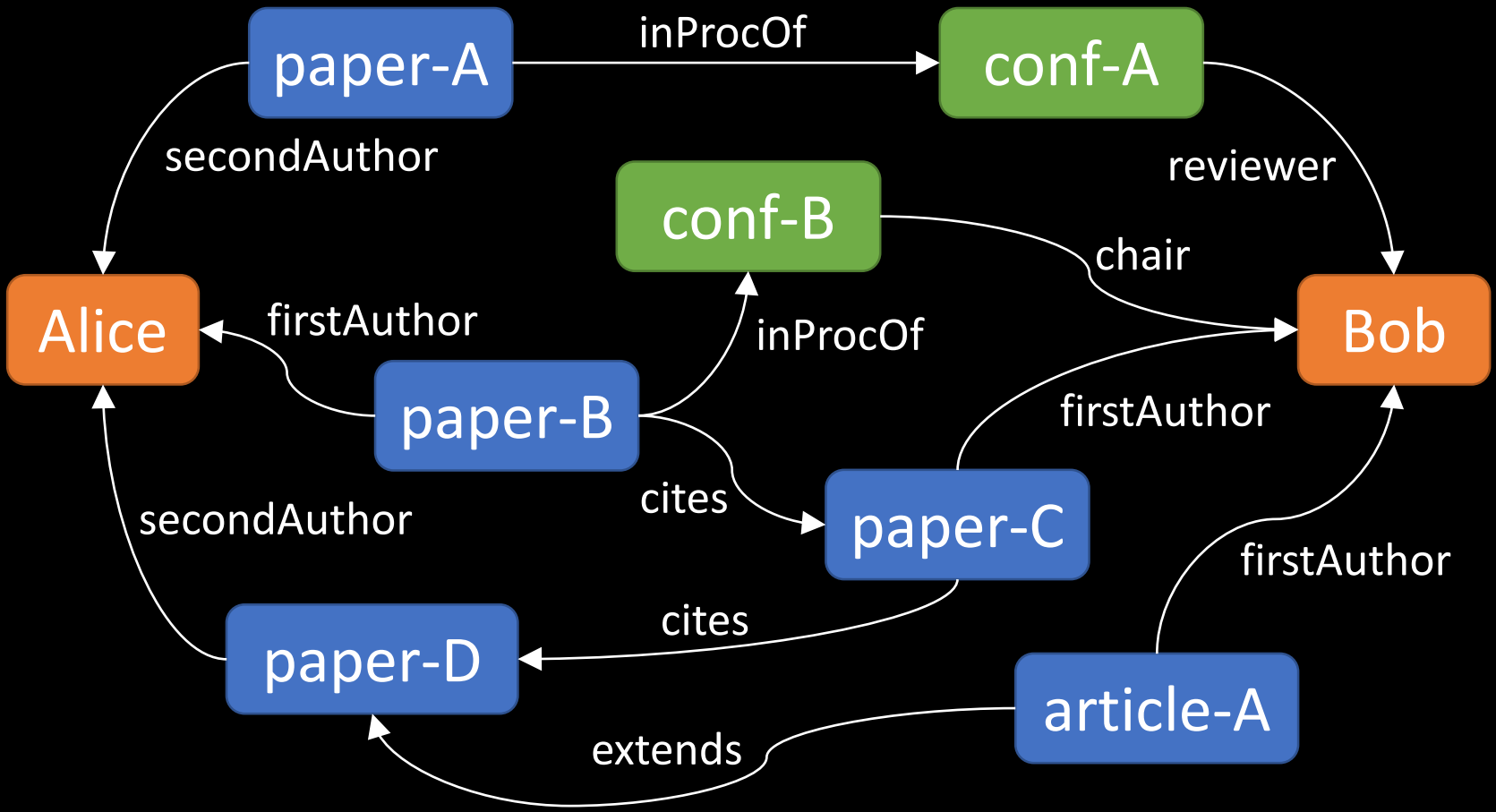
[terminology - Who are Alice and Bob? - Stack Overflow](#)

[stackoverflow.com/questions/2805183/who-are-alice-and-bob](#) ▼

May 10, 2010 - The names **Alice** and **Bob** are commonly used placeholder names for archetypal characters in fields such as cryptography and physics. The names ...

associations hidden in text

Association search on an entity-relation graph

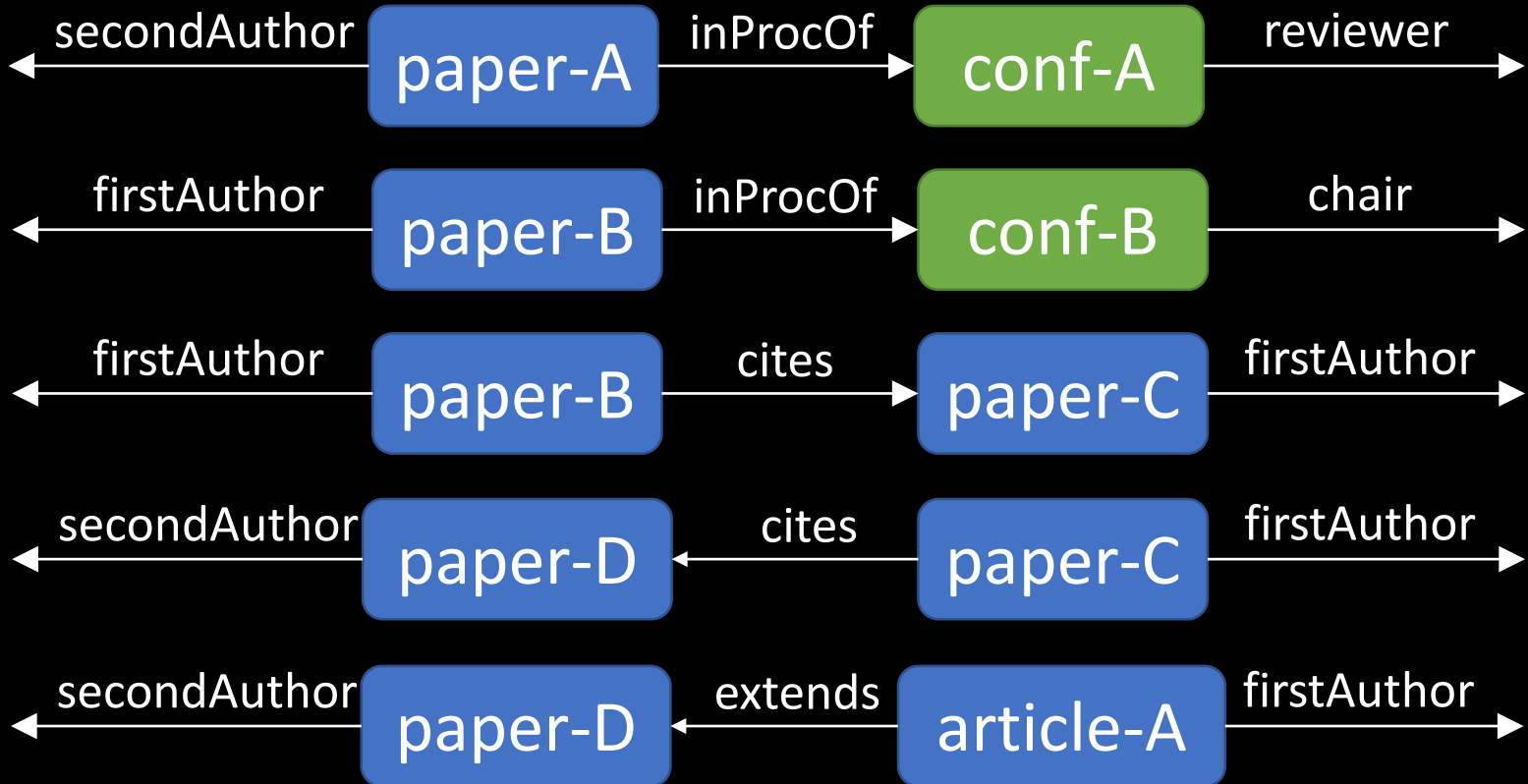


associations exposed as graph

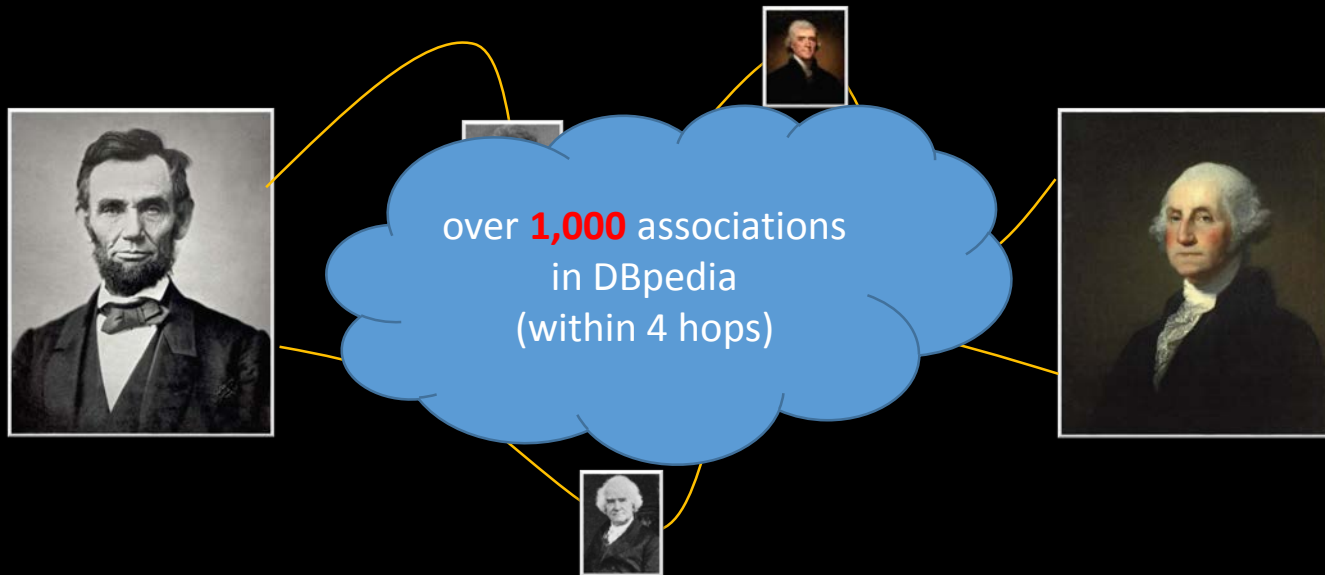
association = path

Alice

Bob



Challenge

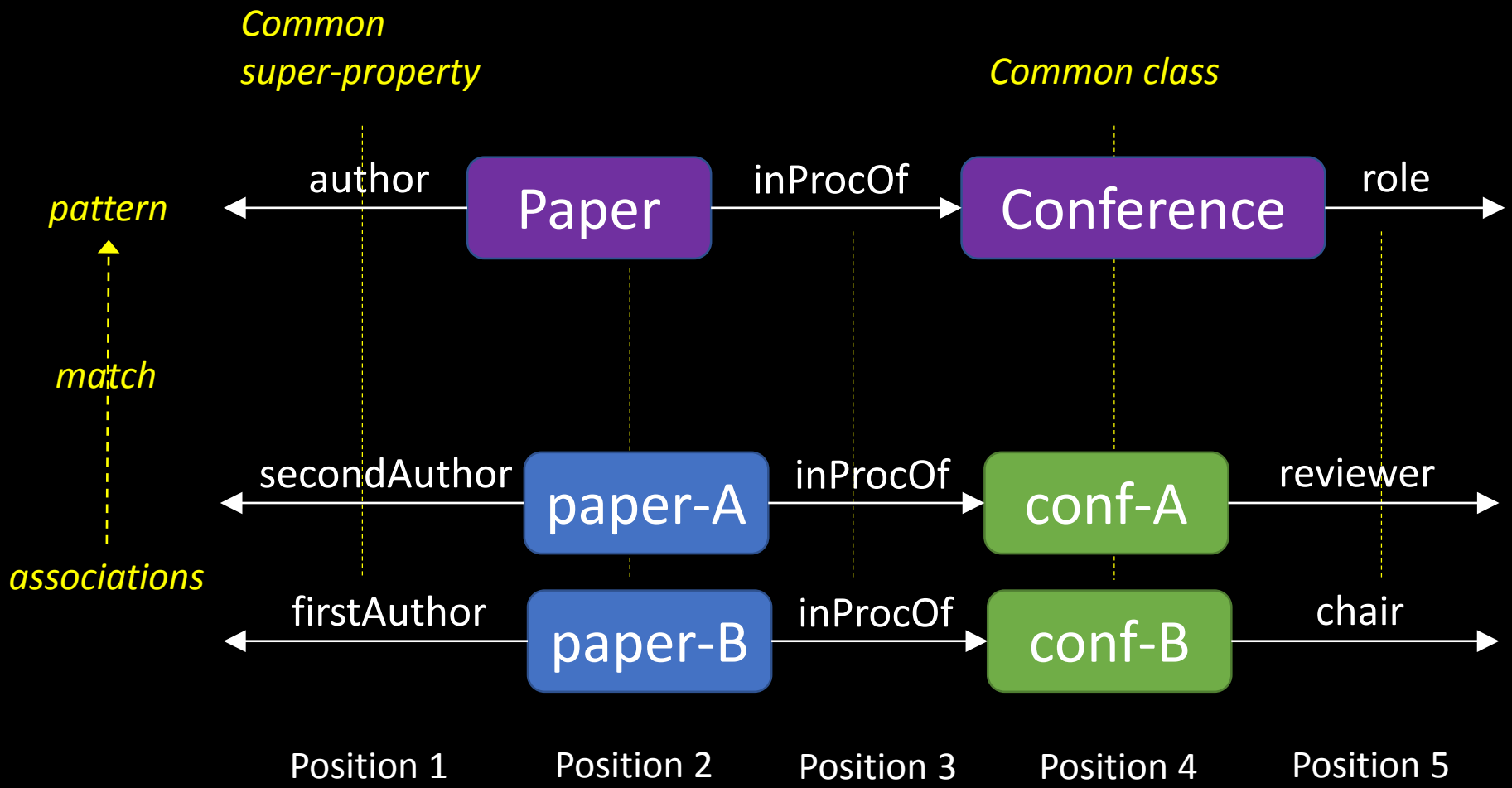


How to explore them?

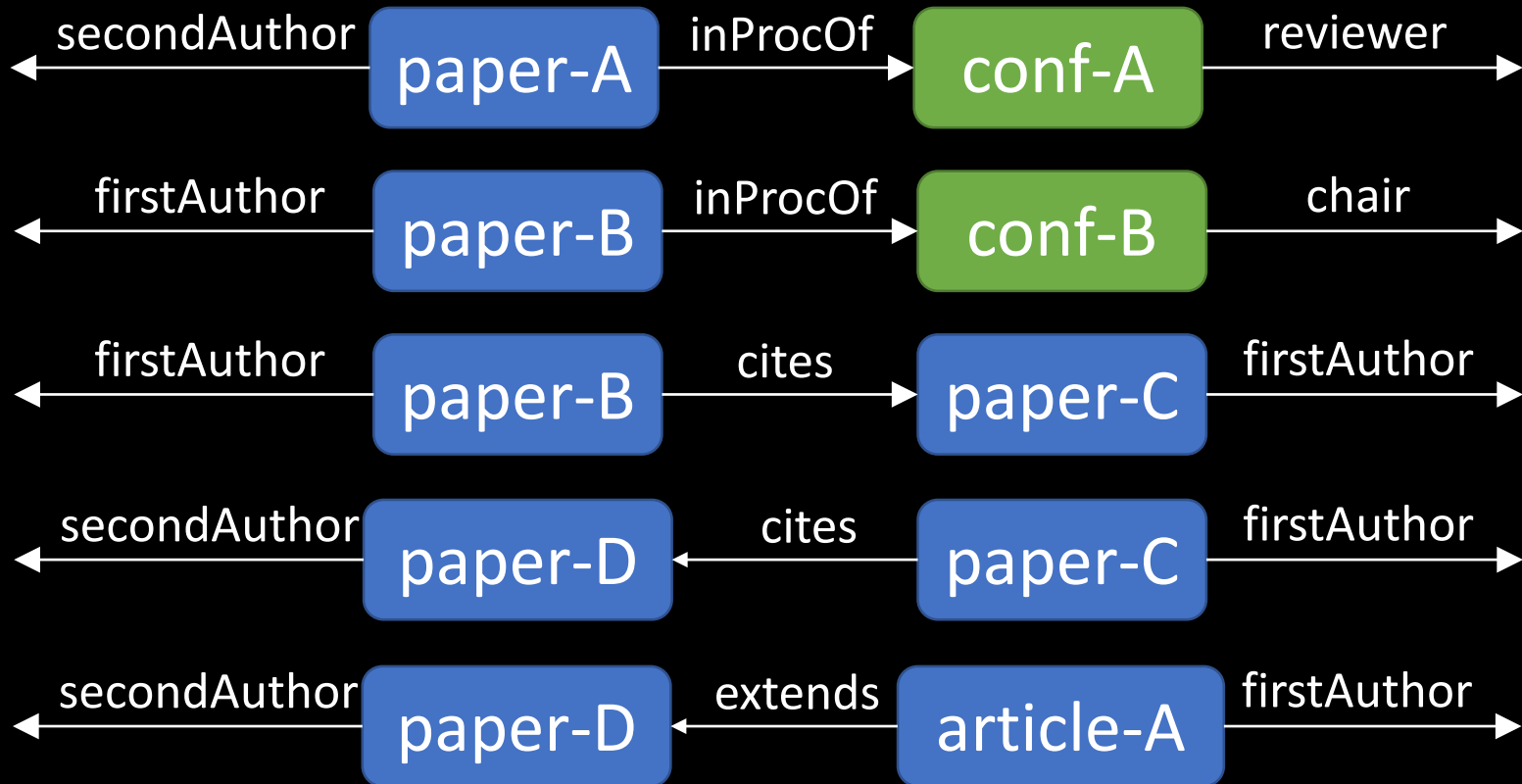
Exploration methods (1)

- Clustering
- Facets

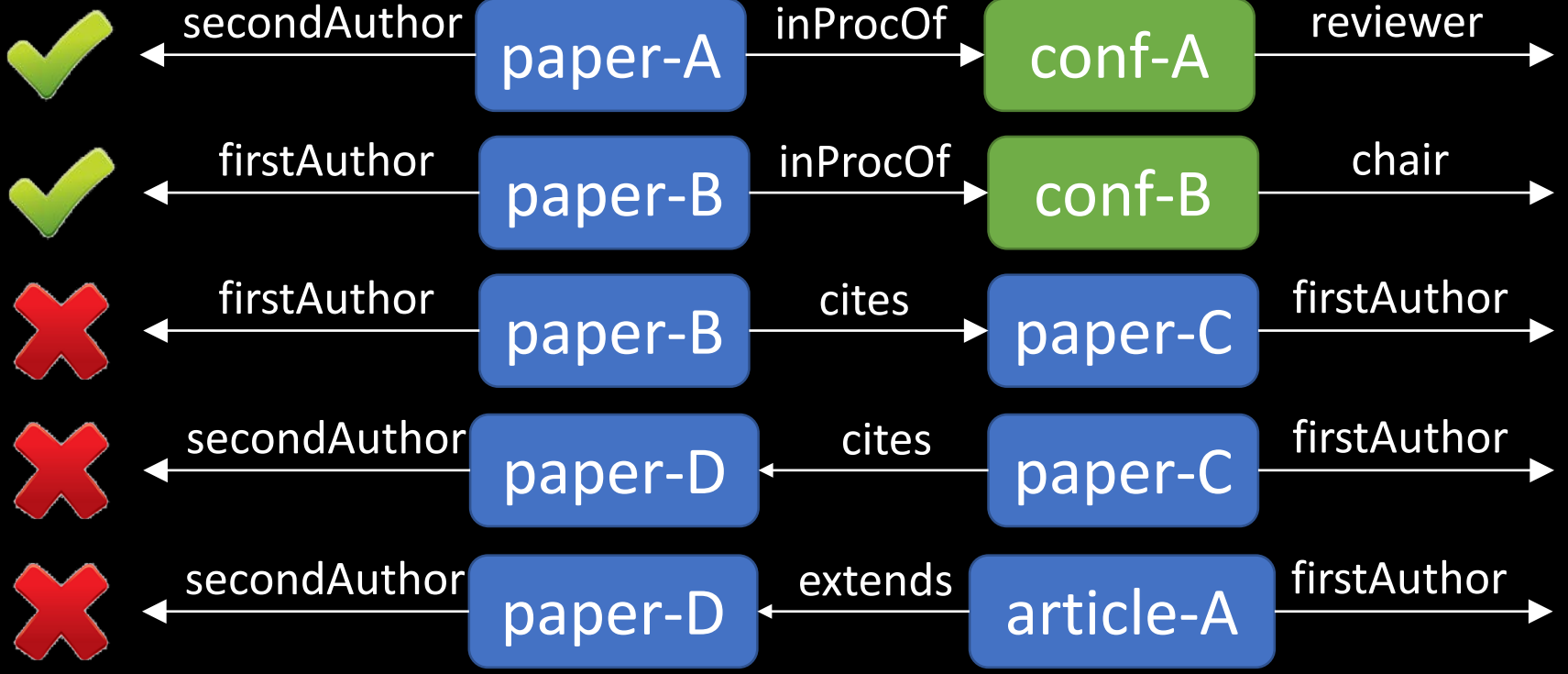
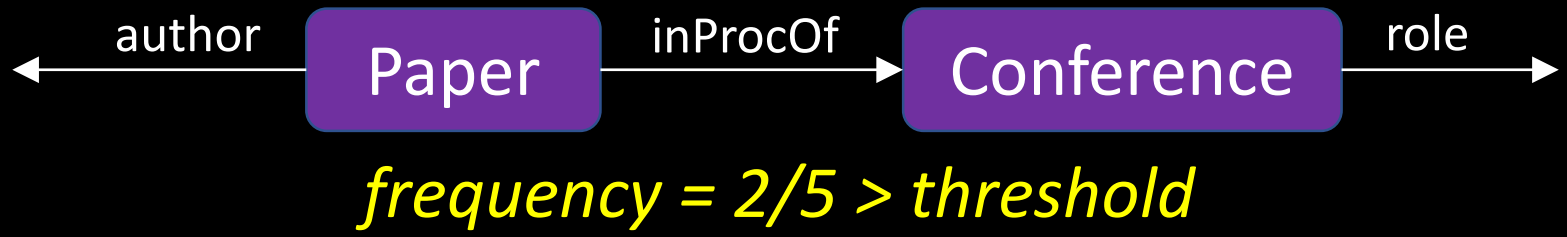
cluster = pattern



Problem: To recommend k patterns

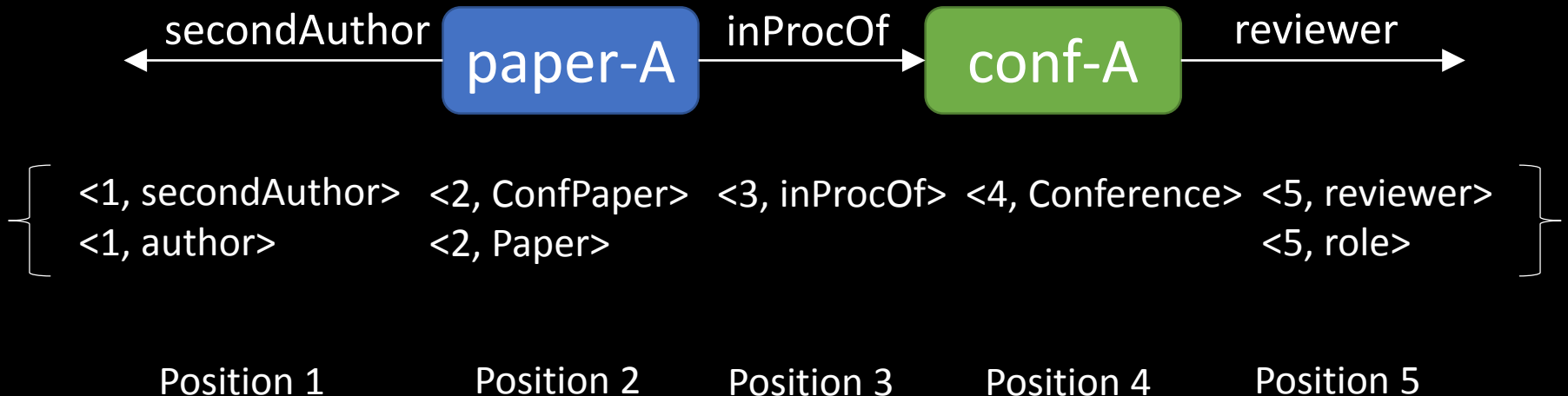


Step 1: Mining all significant patterns



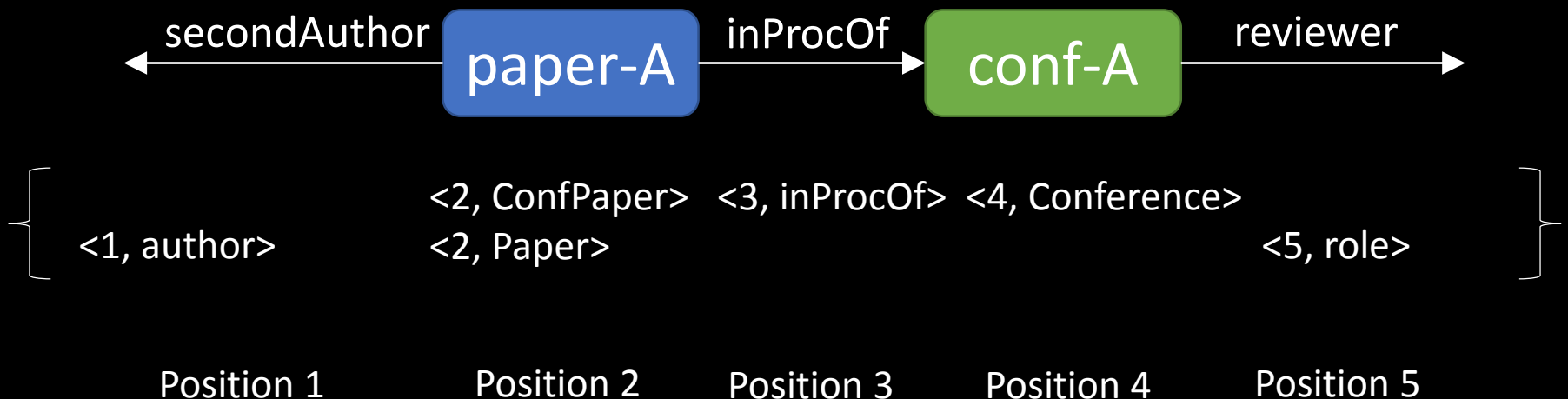
Formulated as frequent itemset mining

1. transaction = association
item = <position, class> or <position, property>
2. Mining frequent itemsets
3. itemset \rightarrow pattern



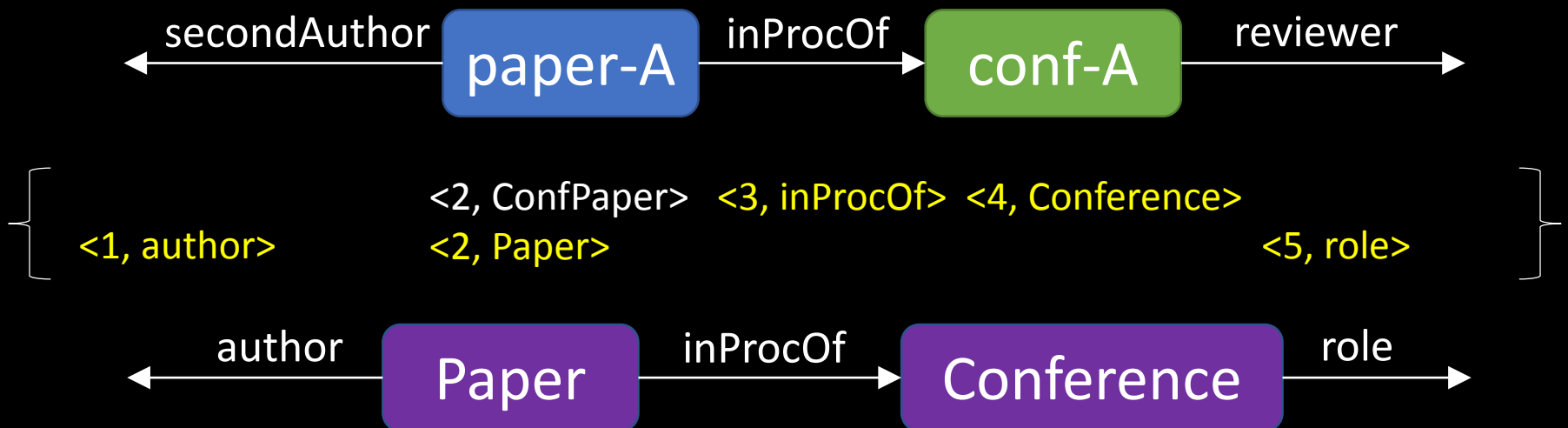
Formulated as frequent itemset mining

1. transaction = association
item = $\langle \text{position}, \text{class} \rangle$ or $\langle \text{position}, \text{property} \rangle$
2. Mining frequent itemsets
3. itemset \rightarrow pattern



Formulated as frequent itemset mining

1. transaction = association
item = $\langle \text{position, class} \rangle$ or $\langle \text{position, property} \rangle$
2. Mining frequent itemsets
3. itemset \rightarrow pattern



Step 2: Finding k frequent, informative, and small-overlapping patterns

- Frequency (as previous)
- Informativeness
- Overlap

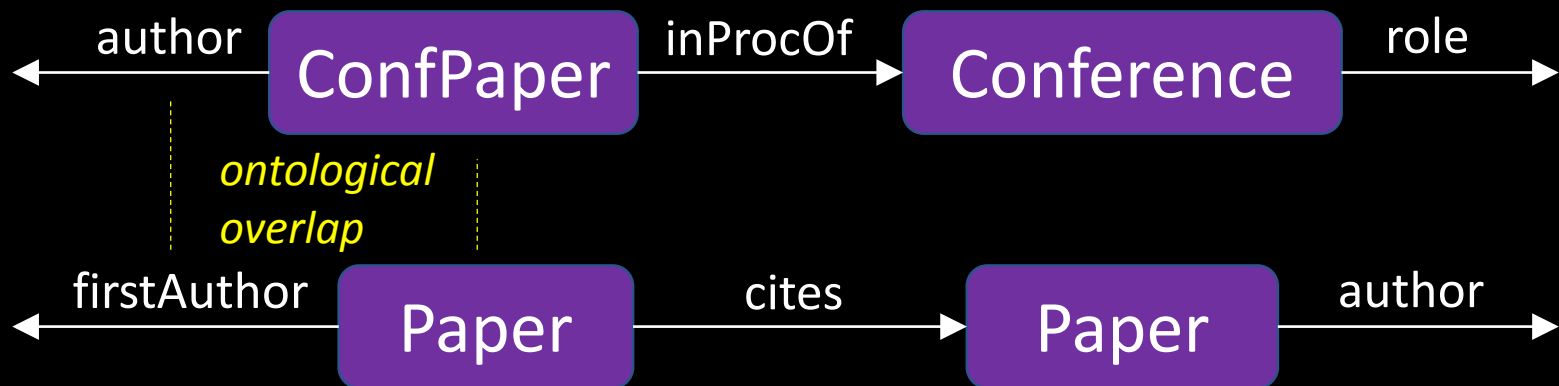
Step 2: Finding k frequent, informative, and small-overlapping patterns

- Frequency (as previous)
- **Informativeness**
 - informativeness of a class = self-information of its occurrence (more informative = having fewer instances)
e.g. ConfPaper > Paper
 - informativeness of a property = entropy of its values (more Informative = having more diverse values)
e.g. is-author-of > nationality
- Overlap



Step 2: Finding k frequent, informative, and small-overlapping patterns

- Frequency (as previous)
- Informativeness
- Overlap
 - Ontological overlap: holding subClassOf/subPropertyOf relations
 - Contextual overlap: matched by common associations in the results

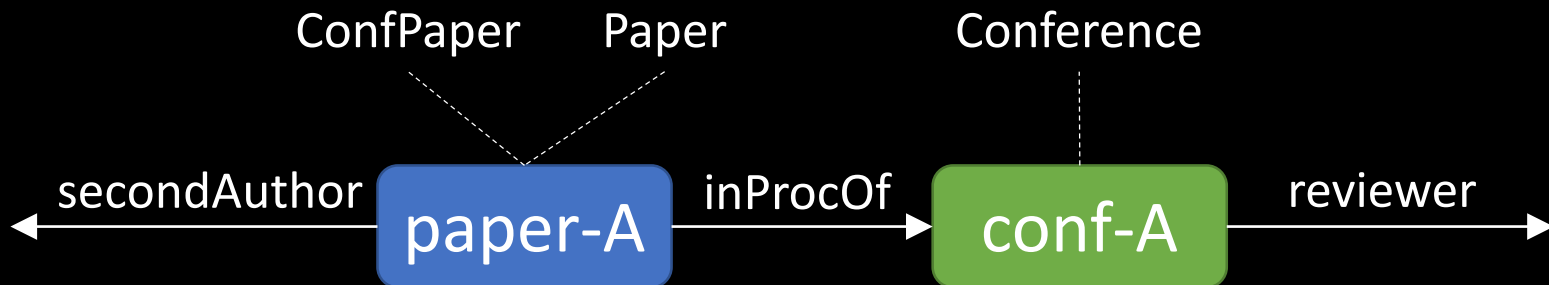


Formulated as multidimensional 0-1 knapsack

- Find k patterns that
 - maximize frequency*Informativeness (goal)
 - and not share considerably large overlap (constraints)
- Solved by a greedy algorithm

Exploration methods (2)

- Clustering
- Facets
 - facet values = classes of entities and properties appearing in associations in the results
 - Problem: To recommend k facet values (solved in a similar way)



Demo based on DBpedia

ws.nju.edu.cn/explass

EXPLASS Between and [Explore](#)

Current filters: [Clear all](#)

Found 748 associations between Abraham Lincoln and George Washington after filtering.

Pass by a type of entity

- Person (748)
- Country (744)
- AdministrativeArea (48)
- Book (25)
- City (9)

Pass by a type of link

- president (575)
- deathPlace (394)
- birthPlace (366)
- country (318)
- child (173)

Other results (83)

- Abraham Lincoln **influenced** Thomas Paine **influenced** Thomas Jefferson **related** George Washington (2)
- Abraham Lincoln **president** some OfficeHolder **religion** Episcopal Church (United States) **religion** George Washington (2)
- Abraham Lincoln **president** George M. Dallas **religion** Episcopal Church (United States) **religion** George Washington
- Abraham Lincoln **president** Salmon P. Chase **religion** Episcopal Church (United States) **religion** George Washington
- Abraham Lincoln **related** some Person **related** United States **related** some Person **related** George Washington (661)
- Abraham Lincoln **successor** James Buchanan **deathPlace** United States **country** George Washington
- Abraham Lincoln **birthPlace** Hodgenville, Kentucky **country** United States **country** Henry Knox **president** George Washington
- Abraham Lincoln **birthPlace** Hodgenville, Kentucky **country** United States **deathPlace** Timothy Pickering **president** George Washington

Demo based on DBpedia

ws.nju.edu.cn/explass

EXPLASS Between and [Explore](#)

Current filters: [Clear all](#)

Found 748 associations between Abraham Lincoln and George Washington after filtering.

Pass by a type of entity

- Person (748)
- Country (744)
- Administrative
- Book (25)
- City (9)

Pass by a type of link

- president (575)
- deathPlace (39)
- birthPlace (366)
- country (318)
- child (173)

facet values (classes)

facet values (properties)

Abraham Lincoln **influenced** Thomas Paine **influenced** Thomas Jefferson **related** George Washington (2)

Abraham Lincoln **president** some OfficeHolder **religion** Episcopal Church (United States) **religion** George Washington (2)

Abraham Lincoln **president** George M. Dallas **religion** Episcopal Church (United States) **religion** George Washington

Abraham Lincoln **president** Salmon P. Chase **religion** Episcopal Church (United States) **religion** George Washington

Abraham Lincoln **related** some Person **related** United States **related** some Person **related** George Washington (661)

Other results (83)

Abraham Lincoln **successor** James Buchanan **deathPlace** United States **country** George Washington

Abraham Lincoln **birthPlace** Hodgenville, Kentucky **country** United States **country** Henry Knox **president** George Washington

Abraham Lincoln **birthPlace** Hodgenville, Kentucky **country** United States **deathPlace** Timothy Pickering **president** George Washington

Demo based on DBpedia

ws.nju.edu.cn/explass

EXPLASS Between and [Explore](#)

Current filters: [Clear all](#)

Found 748 associations between Abraham Lincoln and George Washington after filtering.

Pass by a type of entity

- Person (748)
- Country (744)
- AdministrativeArea (48)
- Book (25)
- City (9)

Pass by a type of link

- president (575)
- deathPlace (394)
- birthPlace (366)
- country (318)
- child (173)

a collapsed pattern

an expanded pattern

associations not matching any pattern above

Abraham Lincoln **influenced** Thomas Paine **influenced** Thomas Jefferson **related** George Washington (2)

Abraham Lincoln **president** some OfficeHolder **religion** Episcopal Church (United States) **religion** George Washington (2)

Abraham Lincoln **president** George M. Dallas **religion** Episcopal Church (United States) **religion** George Washington (2)

Abraham Lincoln **president** Salmon P. Chase **religion** Episcopal Church (United States) **religion** George Washington (2)

Abraham Lincoln **related** some Person **related** United States **related** some Person **related** George Washington (661)

Other results (83)

Abraham Lincoln **successor** James Buchanan **deathPlace** United States **country** United States

Abraham Lincoln **birthPlace** Hodgenville, Kentucky **country** United States **country** United States

Abraham Lincoln **birthPlace** Hodgenville, Kentucky **country** United States **deathPlace** Timothy Pickering **president** George Washington

User study

- 26 association exploration tasks over DBpedia
 - Derived from QALD queries and “People also search for”
 - Example: Suppose you will write an article about the associations between **Abraham Lincoln** and **George Washington**. Use the given system to explore their associations and identify several themes to discuss in the article.
- 20 subjects
- 3 approaches
 - Expass: clustering + facets
 - RelClus: clustering into a hierarchy of patterns
 - RF: facets only (similar to RelFinder)

from QALD

Abraham Lincoln

16th U.S. President

Abraham Lincoln was the 16th president of the United States, serving from March 1861 until his assassination in April 1865. [Wikipedia](#)

Born: February 12, 1809, Hodgenville, Kentucky, United States

Height: 6' 4" (1.93 m)

Assassinated: April 15, 1865, Washington, D.C., United States

Spouse: Mary Todd Lincoln (m. 1842–1865)

Children: William Wallace Lincoln, Robert Todd Lincoln, Edward Baker Lincoln, Tad Lincoln

Vice presidents: Hannibal Hamlin (1861–1865), Andrew Johnson (1865)

People also search for

[View 15+ more](#)



George Washington



Mary Todd Lincoln
Spouse



William Wallace Lincoln
Son



John F. Kennedy



John Wilkes Booth

Post-task questionnaire results

Question	Response: Mean (SD)			$F(2, 38)$ (p -value)	LSD post-hoc ($p < 0.05$)
	Explass	RelClus	RF		
Q1: The system helped me get an overview of all the information.	4.25 (0.85)	3.80 (0.77)	3.05 (0.94)	14.989 (0.000)	Explass, RelClus > RF
Q2: The system helped me easily find information relevant to this task.	4.30 (0.57)	3.25 (0.79)	3.15 (0.99)	18.769 (0.000)	Explass > RelClus, RF
Q3: The system helped me easily compare and synthesize all kinds of relevant information.	4.00 (0.86)	3.25 (0.85)	2.60 (0.99)	14.901 (0.000)	Explass > RelClus > RF
Q4: The system provided me with much support for carrying out this task.	4.10 (0.72)	3.45 (0.94)	2.85 (0.88)	16.172 (0.000)	Explass > RelClus > RF
Q5: The system provided me with sufficient support for carrying out this task. ^a	3.85 (0.88)	3.20 (1.11)	2.65 (0.75)	11.636 (0.000)	Explass > RelClus, RF

Usability scores (SUS)

Explclass	Mean (SD)		$F(2, 38)$ (p -value)	LSD post-hoc ($p < 0.05$)
	RelClus	RF		
76.13 (12.53)	68.00 (17.93)	62.75 (14.93)	9.062 (0.001)	Explclass > RelClus, RF

User behavior

Operation	Explass	RelClus	RF
Refocusing by <u>expanding or collapsing a pattern</u>	9.55	19.60	n/a
Refining the search by a pattern filter or canceling it	0.35	n/a	n/a
<u>Refining the search by a facet value</u> filter or canceling it	5.35	n/a	9.60

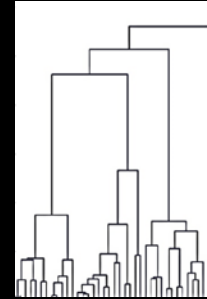
Conclusion

1. Provide patterns wisely.

- To avoid deep, complicated hierarchy
- To avoid very general, almost meaningless concepts

2. Combine patterns and facets wisely.

- Patterns as meaningful summaries of results
- Facets as filters for refining the search



Pass by a type of entity

Person (718)

Country

Address

Book (25)

City (9)

Filters

Pass by a type of link

Abraham Lincoln **influenced** Thomas Paine **influenced** Thomas Jefferson **related** George Washington (2)

Abraham Lincoln **president** some Person **related** George Washington (2)

Abraham Lincoln **president** George Washington **related** George Washington

Abraham Lincoln **president** Salmon P. Chase **religion** Episcopal Church (United States) **religion** George Washington

Abraham Lincoln **related** some Person **related** United States **related** some Person **related** George Washington (661)

Summaries of results

Future work

- Performance optimization
 - (online) path finding
 - (online) frequent itemset mining
- Exploring associations between several entities or, a data set

Questions?

EXPLASS

