中国科学院大学
University of Chinese Academy of Sciences

Identifying Mappings among Knowledge Graphs by Formal Concept Analysis

Guowei Chen and Songmao Zhang

Academy of Mathematics & Systems Science, Chinese Academy of Sciences

Beijing, China

勤笃求真 数系天地

OM-2019, ISWC 2019 Workshop
Auckland, New Zealand
October 26, 2019

中国科学院大学
University of Chinese Academy of Sciences

Table of Contents

An Introduction to Formal Concept Analysis

University of Chinese Academy of Sciences

Introduction

Formal Concept Analysis (FCA) is a well-developed mathematical model for clustering individuals and structuring concepts. (Wille 1982; Ganter and Wille 1999)

- Conceptual data analysis.
- Knowledge processing.


Bernhard Ganter and Rudolf Wille. *Formal Concept Analysis - Mathematical Foundations*. Springer, 1999.


Rudolf Wille. “Restructuring Lattice Theory: An Approach Based on Hierarchies of Concepts”. In: *Ordered Sets*. Springer, 1982, pp. 445–470.

中国科学院大学
University of Chinese Academy of Sciences

Formal Concept Analysis - Formal Context

Definition

A **formal context** is a triple $\mathbb{K} := (G, M, I)$, where G is a set of **objects**, M is a set of **attributes**, and I is a relation between G and M.

$(g, m) \in I$ reads "object g has attribute m" represented by "×" in the table cell.

University of Chinese Academy of Sciences

Formal Concept Analysis - Formal Context

Definition

A **formal context** is a triple $\mathbb{K} := (G, M, I)$, where G is a set of **objects**, M is a set of **attributes**, and I is a relation between G and M.

Objects \ Attributes	Asgardian	Avenger	Female	Human	Infinity Stones User	Male	Scientist	Villain
Black Widow		×	×	×				
Captain America		×		×		×		
Hela	×		×					×
Hulk		×		×	×	×	×	
Iron Man		×		×	×	×	×	
Thanos					×	×		×
Thor	×	×				×		

$(g, m) \in I$ reads "object g has attribute m" represented by "×" in the table cell.

University of Chinese Academy of Sciences

Formal Concept Analysis - Derivation Operators

Definition

For a subset of objects $A \subseteq G$ and a subset of attributes $B \subseteq M$, two **derivation operators** are defined as follows:

- $A^{\uparrow} = \{m \in M \mid \forall g \in A.\ (g, m) \in I\}$
- $B^{\downarrow} = \{g \in G \mid \forall m \in B.\ (g, m) \in I\}$

For example,

University of Chinese Academy of Sciences

Formal Concept Analysis - Derivation Operators

Objects \ Attributes	Asgardian	Avenger	Female	Human	Infinity Stones User	Male	Scientist	Villain
Black Widow		×	×	×				
Captain America		×		×		×		
Hela	×		×					×
Hulk		×		×	×	×	×	
Iron Man		×		×	×	×	×	
Thanos					×	×		×
Thor	×	×				×		

Example

Given $A = \{\text{Hulk}, \text{Iron Man}\}$. Each object in A has Avenger, Human, Infinity Stones User, Male and Scientist. Then
$A^{\uparrow} = \{\text{Avenger}, \text{Human}, \text{Infinity Stones User}, \text{Male}, \text{Scientist}\}$.

University of Chinese Academy of Sciences

Formal Concept Analysis - Derivation Operators

Objects \ Attributes	Asgardian	Avenger	Female	Human	Infinity Stones User	Male	Scientist	Villain
Black Widow		×	×	×				
Captain America		×		×		×		
Hela	×		×					×
Hulk		×		×	×	×	×	
Iron Man		×		×	×	×	×	
Thanos					×	×		×
Thor	×	×				×		

Example

Given $B = \{\text{Avenger}, \text{Human}, \text{Male}\}$. Each attribute in B is common to Captain America, Hulk and Iron Man. Then,
$B^{\downarrow} = \{\text{Captain America}, \text{Hulk}, \text{Iron Man}\}$.

University of Chinese Academy of Sciences

Formal Concept Analysis - Formal Concept

Definition

A **formal concept** of context $\mathbb{K}$ is a pair (A, B), where $A \subseteq G$, $B \subseteq M$ such that $A = B^{\downarrow}$ and $B = A^{\uparrow}$.

University of Chinese Academy of Sciences

Formal Concept Analysis - Formal Concept

Definition

A **formal concept** of context $\mathbb{K}$ is a pair (A, B), where $A \subseteq G$, $B \subseteq M$ such that $A = B^{\downarrow}$ and $B = A^{\uparrow}$.

Objects \ Attributes	Asgardian	Avenger	Female	Human	Infinity Stones User	Male	Scientist	Villain
Black Widow		×	×	×				
Captain America		×		×		×		
Hela	×		×					×
Hulk		×		×	×	×	×	
Iron Man		×		×	×	×	×	
Thanos					×	×		×
Thor	×	×				×		

Example

$(\{\text{Captain America}, \text{Hulk}, \text{Iron Man}\}, \{\text{Avenger}, \text{Human}, \text{Male}\})$

中国科学院大学
University of Chinese Academy of Sciences

Formal Concept Analysis - Concept Lattice

Definition

The subconcept-superconcept relation is formalized by

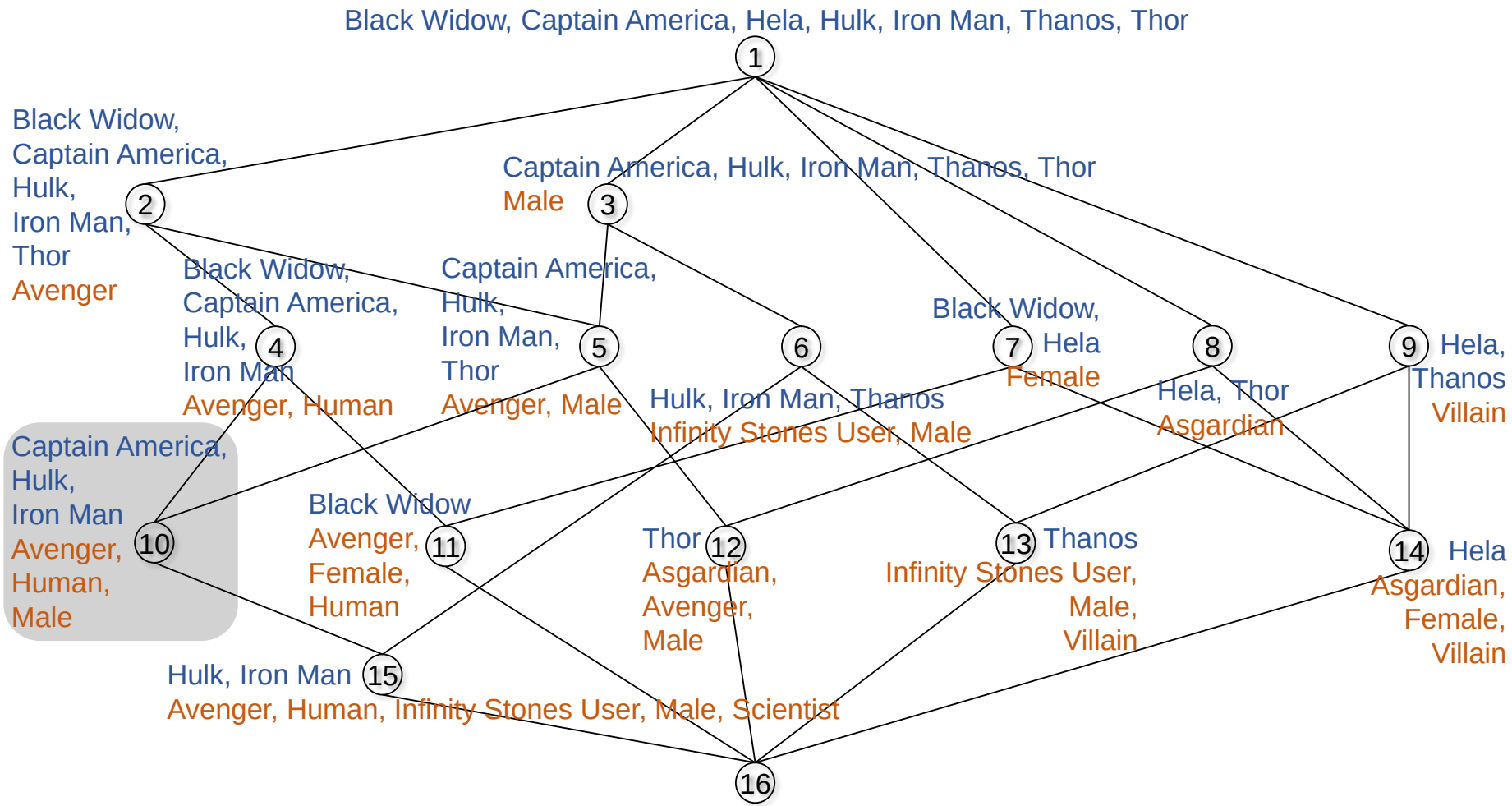
$$(A_1, B_1) \preceq (A_2, B_2) :\Leftrightarrow A_1 \subseteq A_2 (\Leftrightarrow B_1 \supseteq B_2)$$

All formal concepts of $\mathbb{K}$ ordered by relation $\preceq$ is called the **concept lattice** of $\mathbb{K}$.

For example,

University of Chinese Academy of Sciences

Formal Concept Analysis - Concept Lattice



Our Method

University of Chinese Academy of Sciences

Our Method

1. Identifying lexical anchors from a token-based formal context.
   - For classes, properties and instances.
   - Using the tokens from strings of labels/names.
2. Identifying additional property mappings from a structural formal context.
   - For properties.
   - Using the structure of the (subject, predicate, object) triples and lexical anchors.

中国科学院大学
University of Chinese Academy of Sciences

The Knowledge Graph Track for OAEI 2018

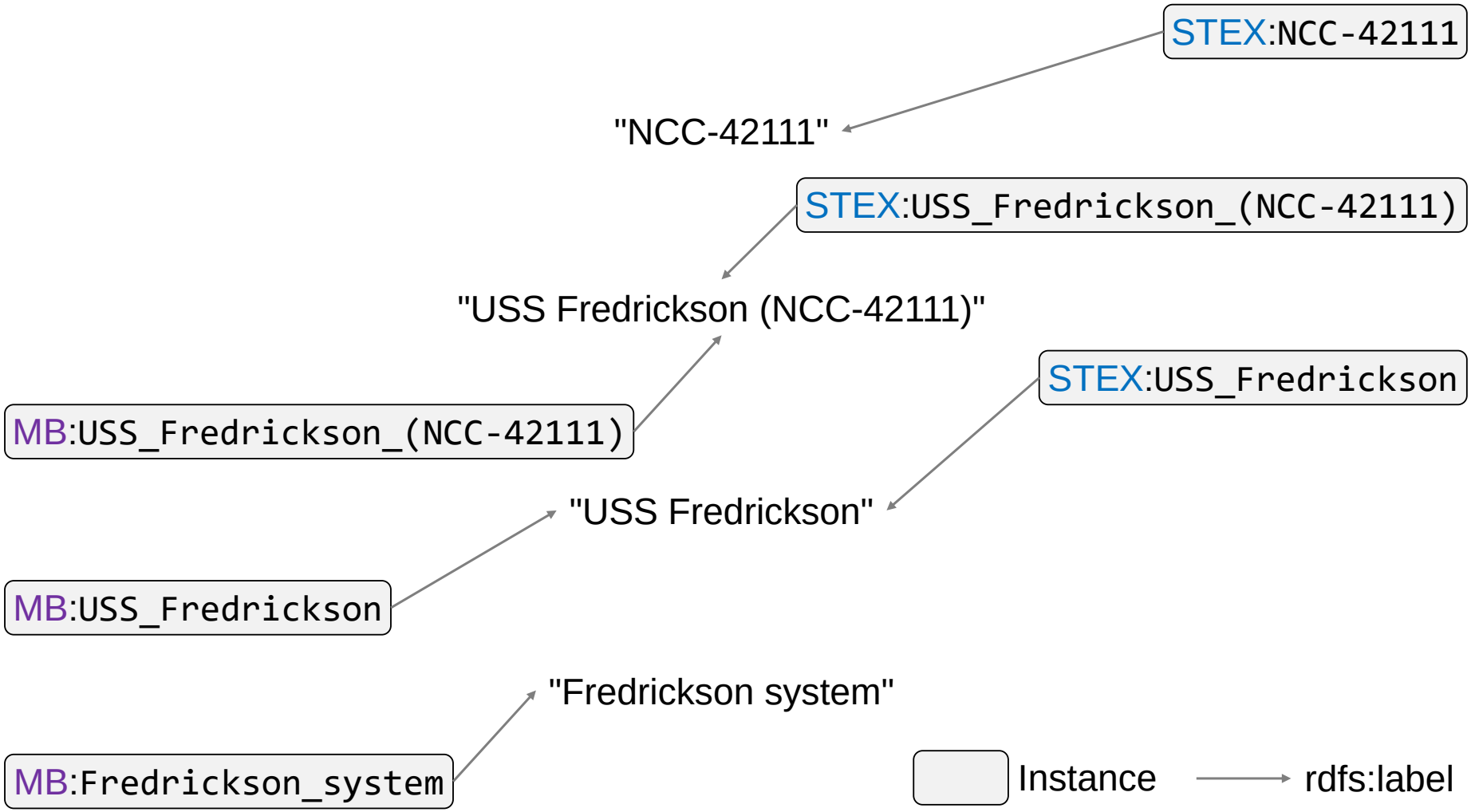
Description

Source	Hub	# Instances	# Properties	# Classes
RuneScape Wiki	Games	200,605	1,998	106
Old School RuneScape Wiki	Games	38,563	488	53
DarkScape Wiki	Games	19,623	686	65
Marvel Database	Comics	56,464	99	2
Hey Kids Comics Wiki	Comics	158,234	1,925	181
DC Database	Comics	128,495	177	5
Memory Alpha	TV	63,240	326	0
Star Trek Expanded Universe	TV	17,659	201	3
Memory Beta	Books	63,223	413	11

For convenience sake, MA stands for Memory Alpha, MB stands for Memory Beta and STEX stands for Star Trek Expanded Universe.

Identifying Lexical Anchors

中国科学院大学
University of Chinese Academy of Sciences



University of Chinese Academy of Sciences

Lexical Anchors - Token-based Formal Context

	uss	fredrickson	ncc	42111	system
uss fredrickson	×	×			
uss fredrickson (ncc-42111)	×	×	×	×	
ncc-42111			×	×	
fredrickson system		×			×

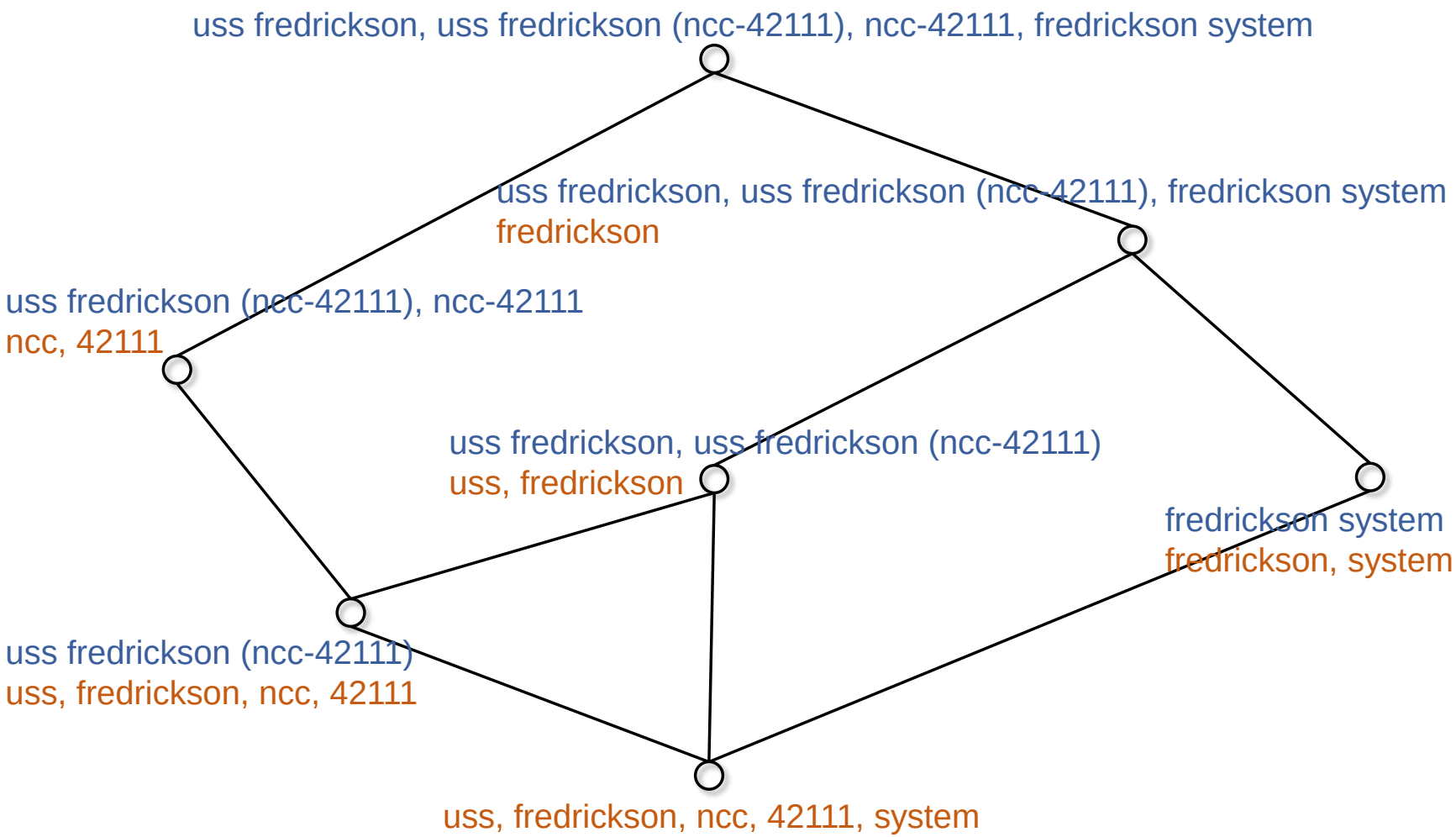
中国科学院大学
University of Chinese Academy of Sciences

Lexical Anchors - Token-based Formal Context

	uss	fredrickson	ncc	42111	system
uss fredrickson	×	×			
uss fredrickson (ncc-42111)	×	×	×	×	
ncc-42111			×	×	
fredrickson system		×			×

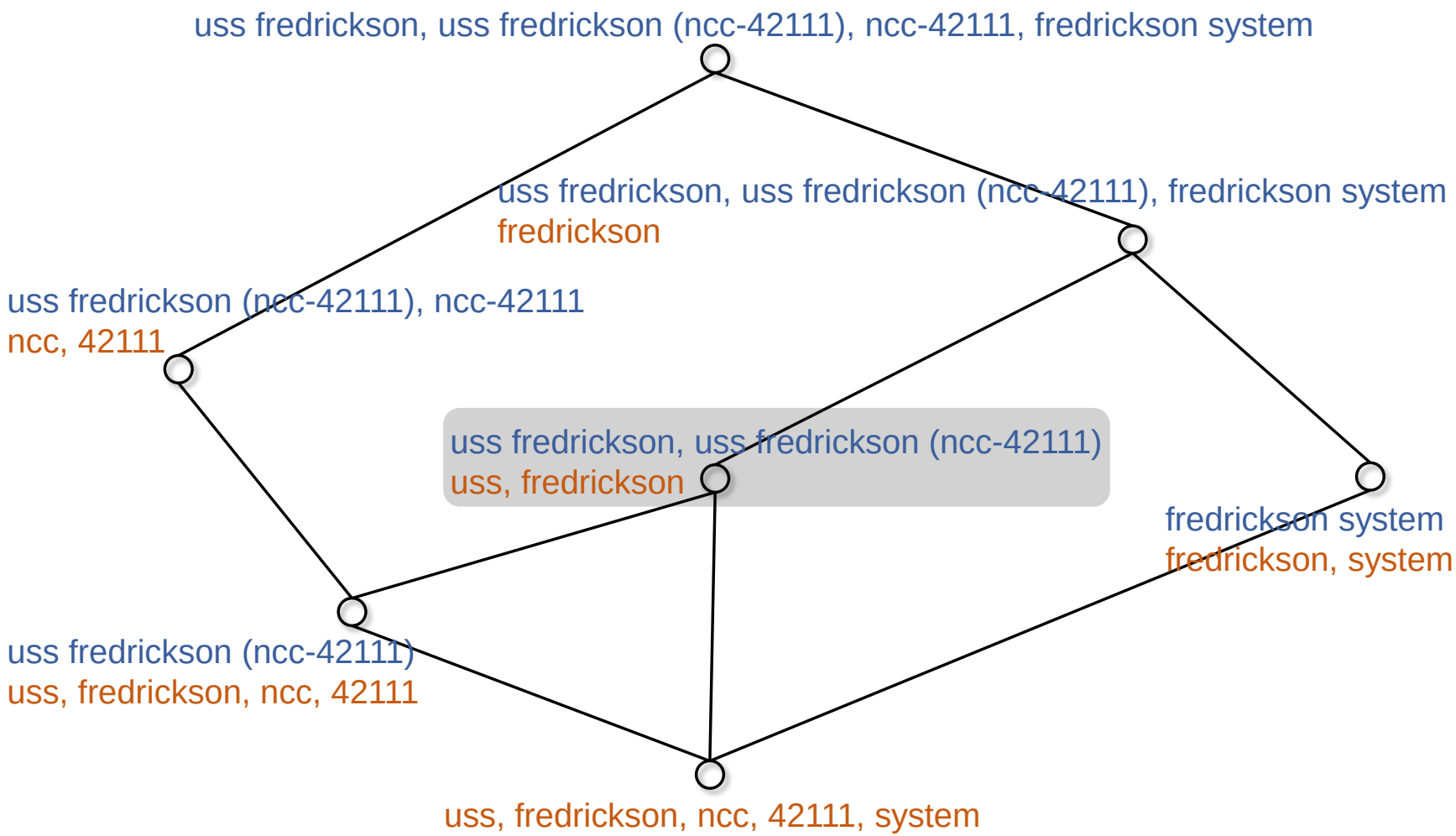
中国科学院大学
University of Chinese Academy of Sciences

Lexical Anchors - Token-based Concept Lattice



University of Chinese Academy of Sciences

Lexical Anchors - Token-based Concept Lattice



University of Chinese Academy of Sciences

Lexical Anchors

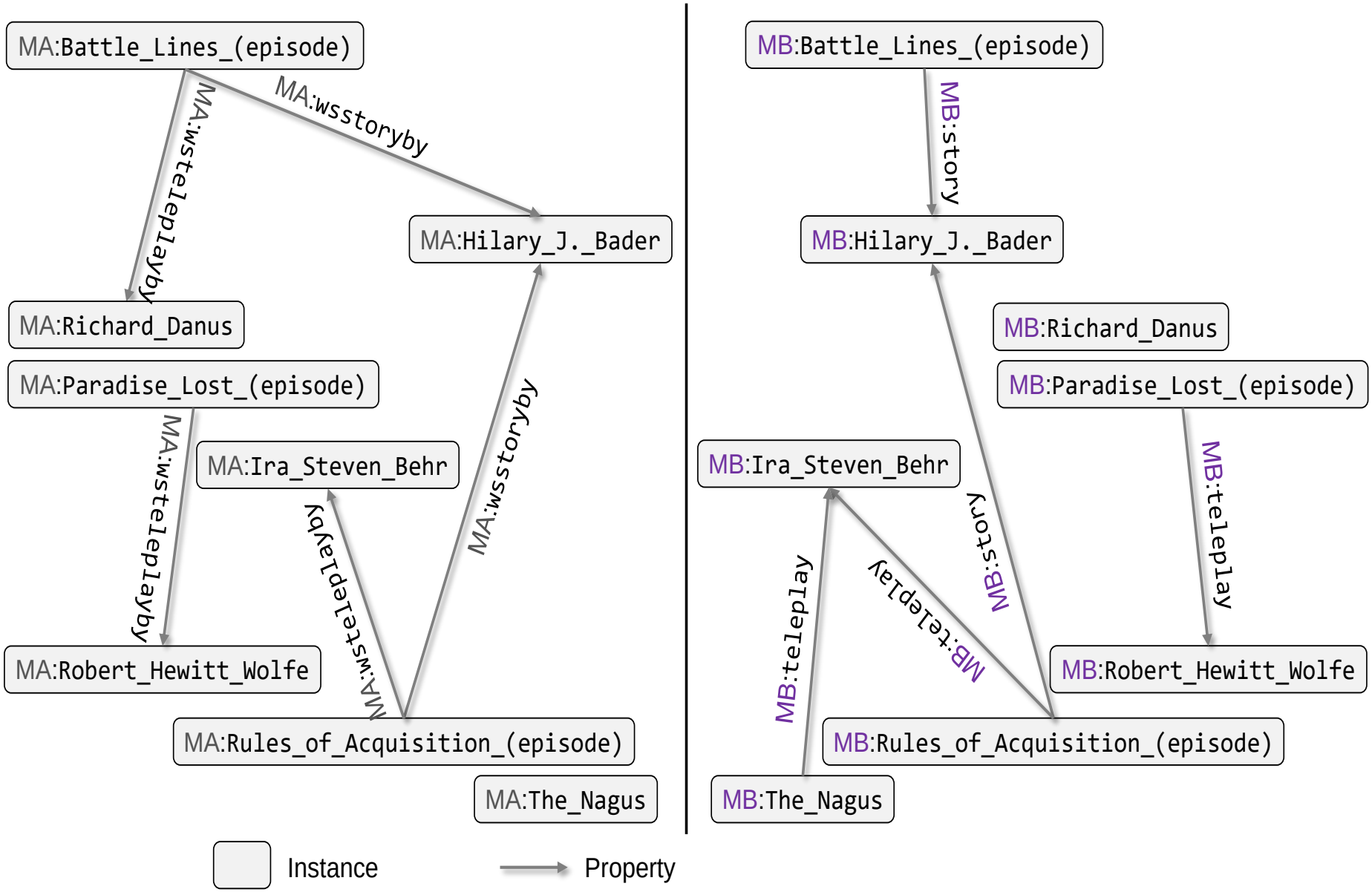
String	Instance
uss fredrickson	MB:USS_Fredrickson STEX:USS_Fredrickson
uss fredrickson (ncc-42111)	MB:USS_Fredrickson_(NCC-42111) STEX:USS_Fredrickson_(NCC-42111)

Four instance mappings can be extracted as follows:

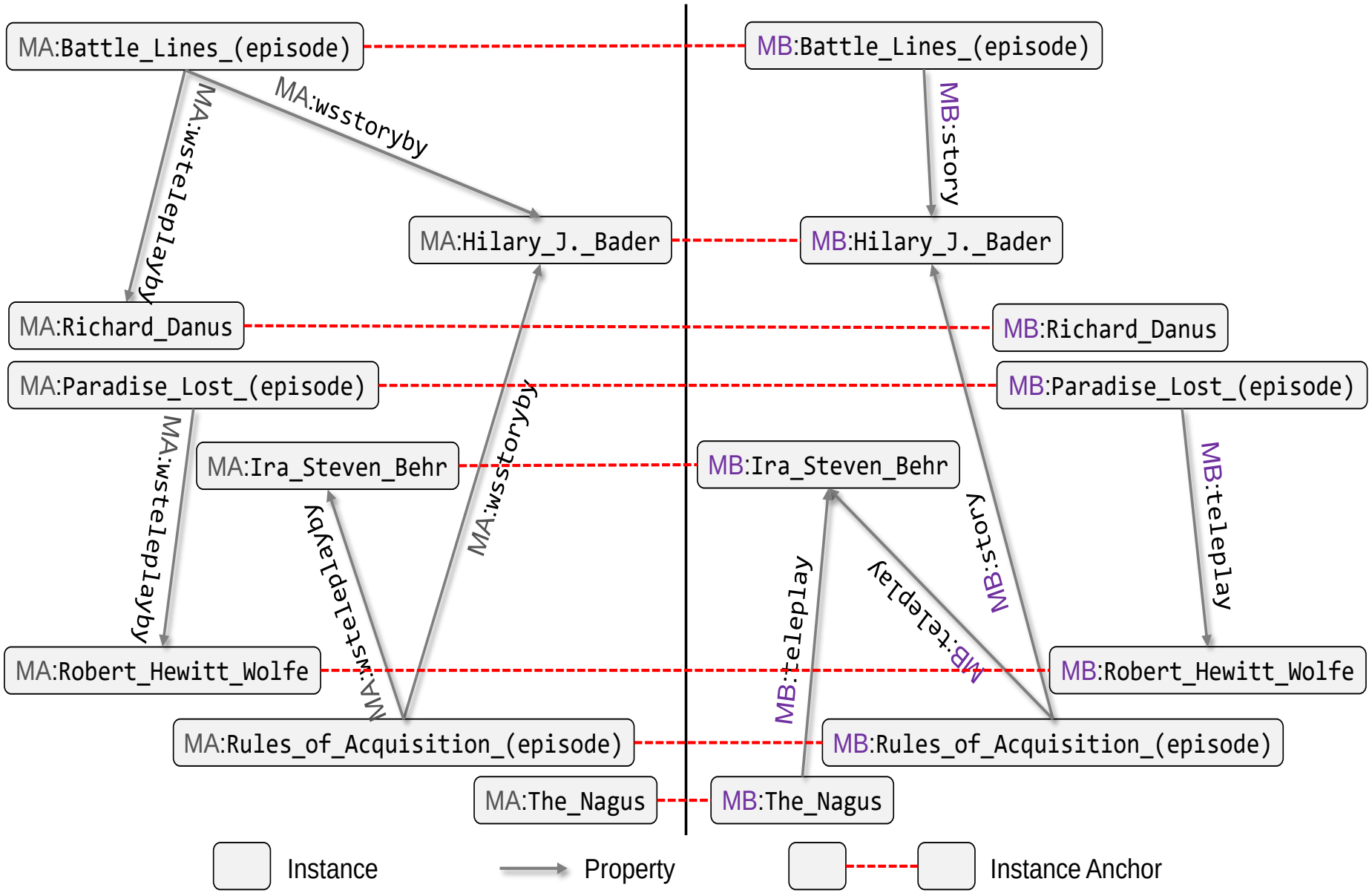
⟨MB:USS_Fredrickson_(NCC-42111), STEX:USS_Fredrickson_(NCC-42111)⟩
⟨MB:USS_Fredrickson_(NCC-42111), STEX:USS_Fredrickson⟩
⟨MB:USS_Fredrickson, STEX:USS_Fredrickson_(NCC-42111)⟩
⟨MB:USS_Fredrickson, STEX:USS_Fredrickson⟩

Identifying Structural Mappings

University of Chinese Academy of Sciences



University of Chinese Academy of Sciences



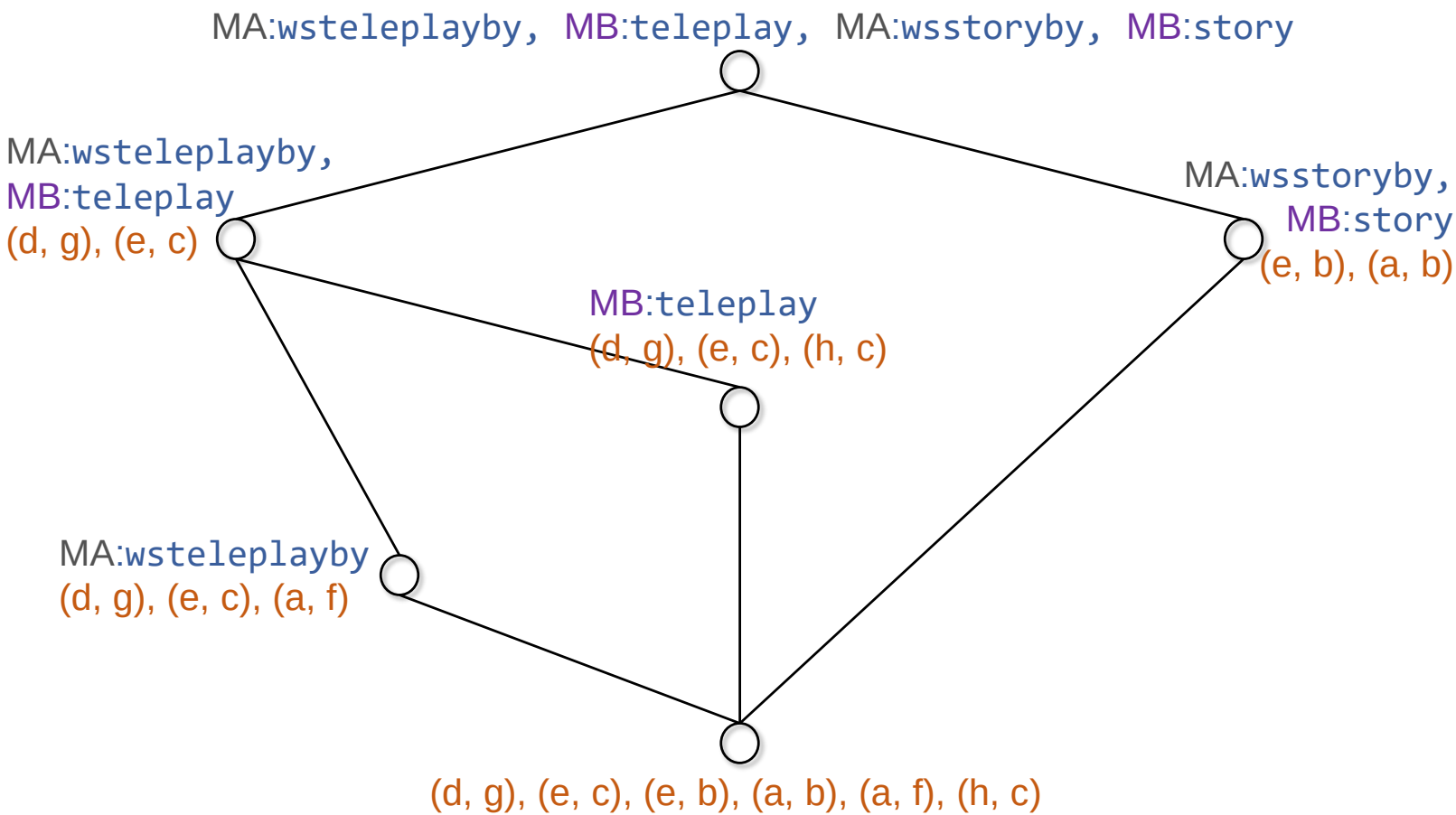
University of Chinese Academy of Sciences

	(d, g)	(e, c)	(e, b)	(a, b)	(a, f)	(h, c)
MA:wsteleplayby	×	×			×	
MB:teleplay	×	×				×
MA:wsstoryby			×	×		
MB:story			×	×		

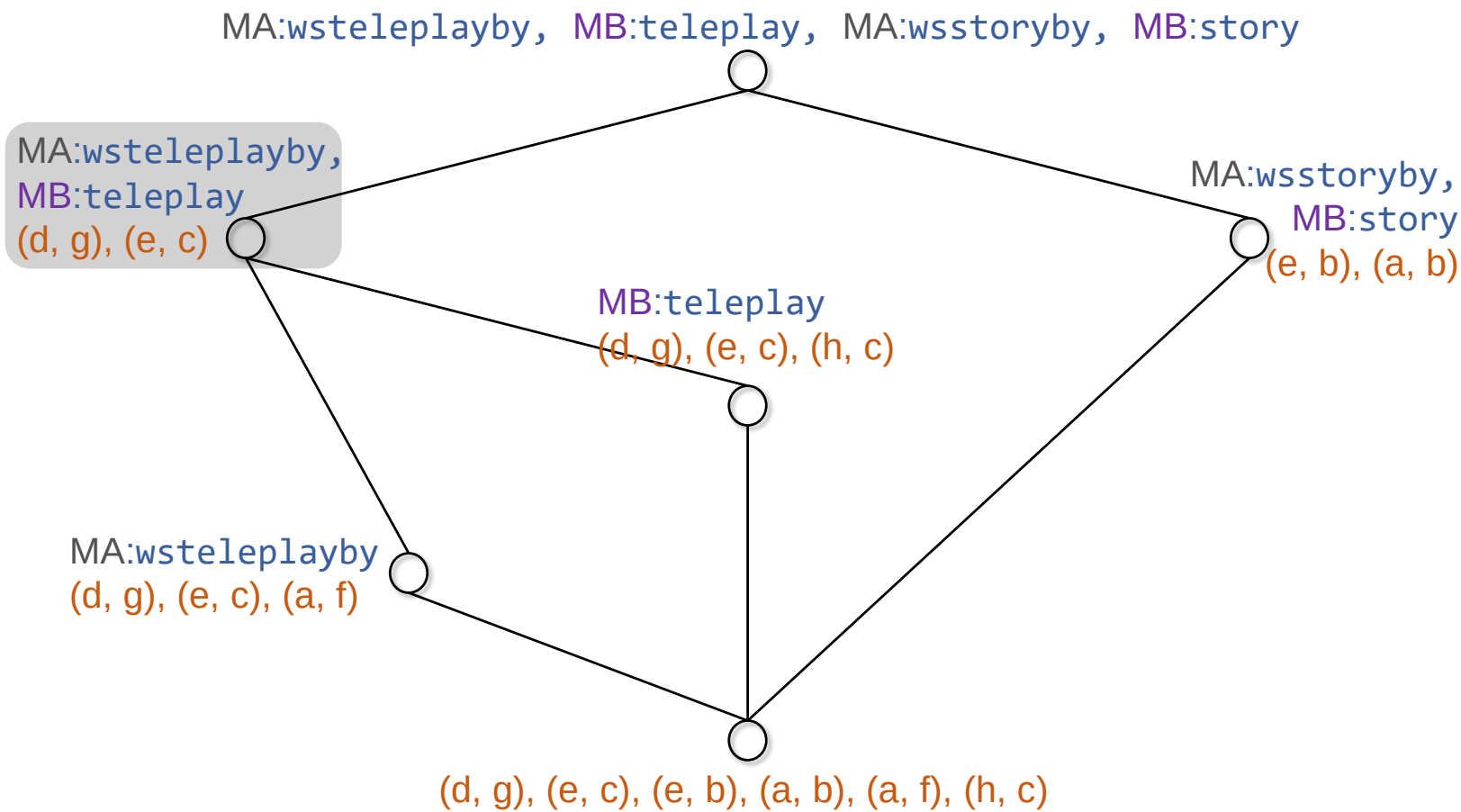
a = ⟨MA:Battle_Lines_(episode), MB:Battle_Lines_(episode)⟩
b = ⟨MA:Hilary_J._Bader, MB:Hilary_J._Bader⟩
c = ⟨MA:Ira_Steven_Behr, MB:Ira_Steven_Behr⟩
d = ⟨MA:Paradise_Lost_(episode), MB:Paradise_Lost_(episode)⟩
e = ⟨MA:Rules_of_Acquisition_(episode), MB:Rules_of_Acquisition_(episode)⟩
f = ⟨MA:Richard_Danus, MB:Richard_Danus⟩
g = ⟨MA:Robert_Hewitt_Wolfe, MB:Robert_Hewitt_Wolfe⟩
h = ⟨MA:The_Nagus, MB:The_Nagus⟩

e.g. MA:Paradise_Lost_(episode) $\xrightarrow{\text{MA:wsteleplayby}}$ MA:Robert_Hewitt_Wolfe

University of Chinese Academy of Sciences

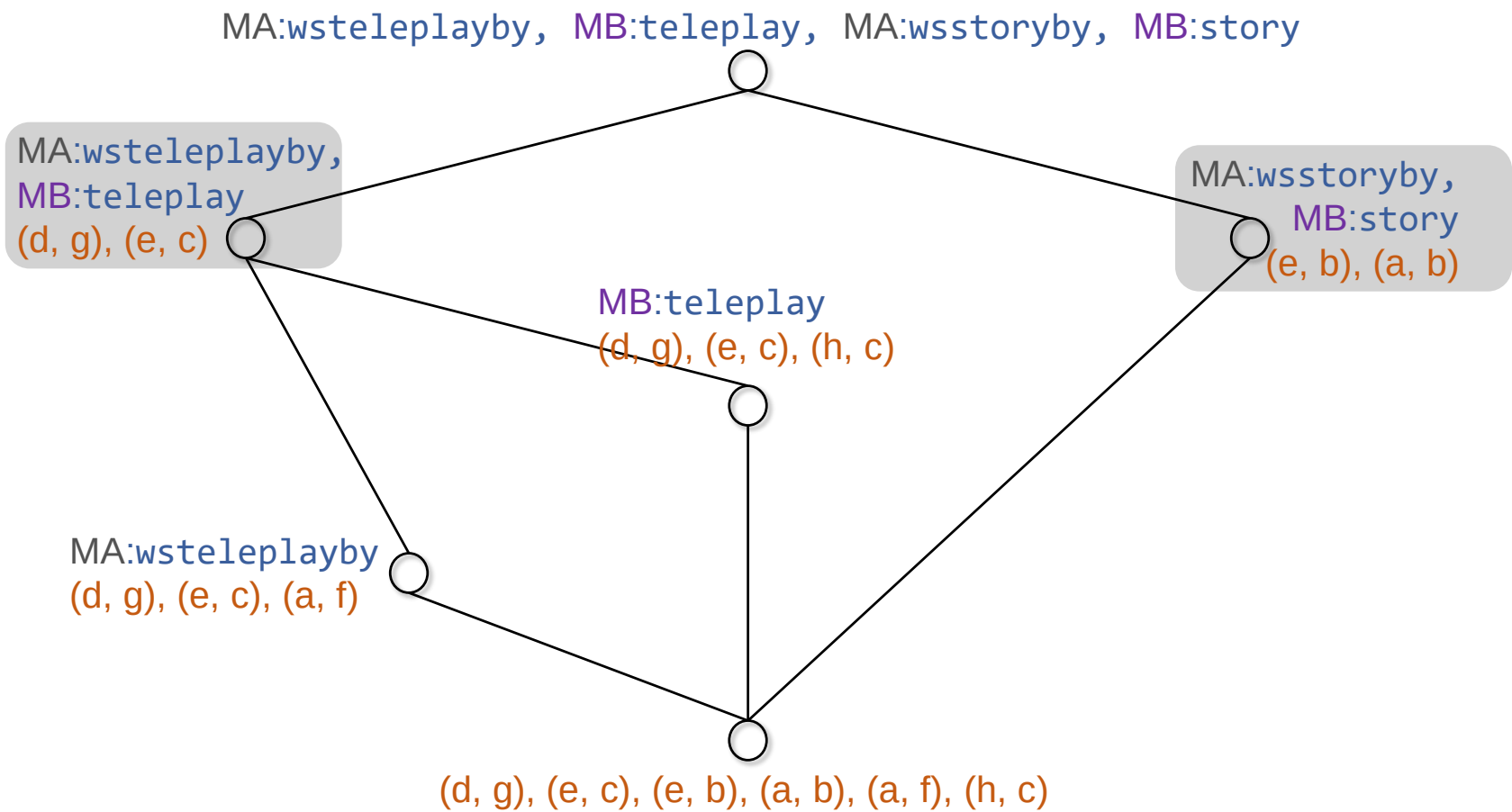


University of Chinese Academy of Sciences



⟨MA:wsteleplayby, MB:teleplay⟩

University of Chinese Academy of Sciences



⟨MA:wsteleplayby, MB:teleplay⟩
⟨MA:wsstoryby, MB:story⟩

中国科学院大学
University of Chinese Academy of Sciences

Evaluation on Knowledge Graph track for OAEI 2018

中国科学院大学
University of Chinese Academy of Sciences

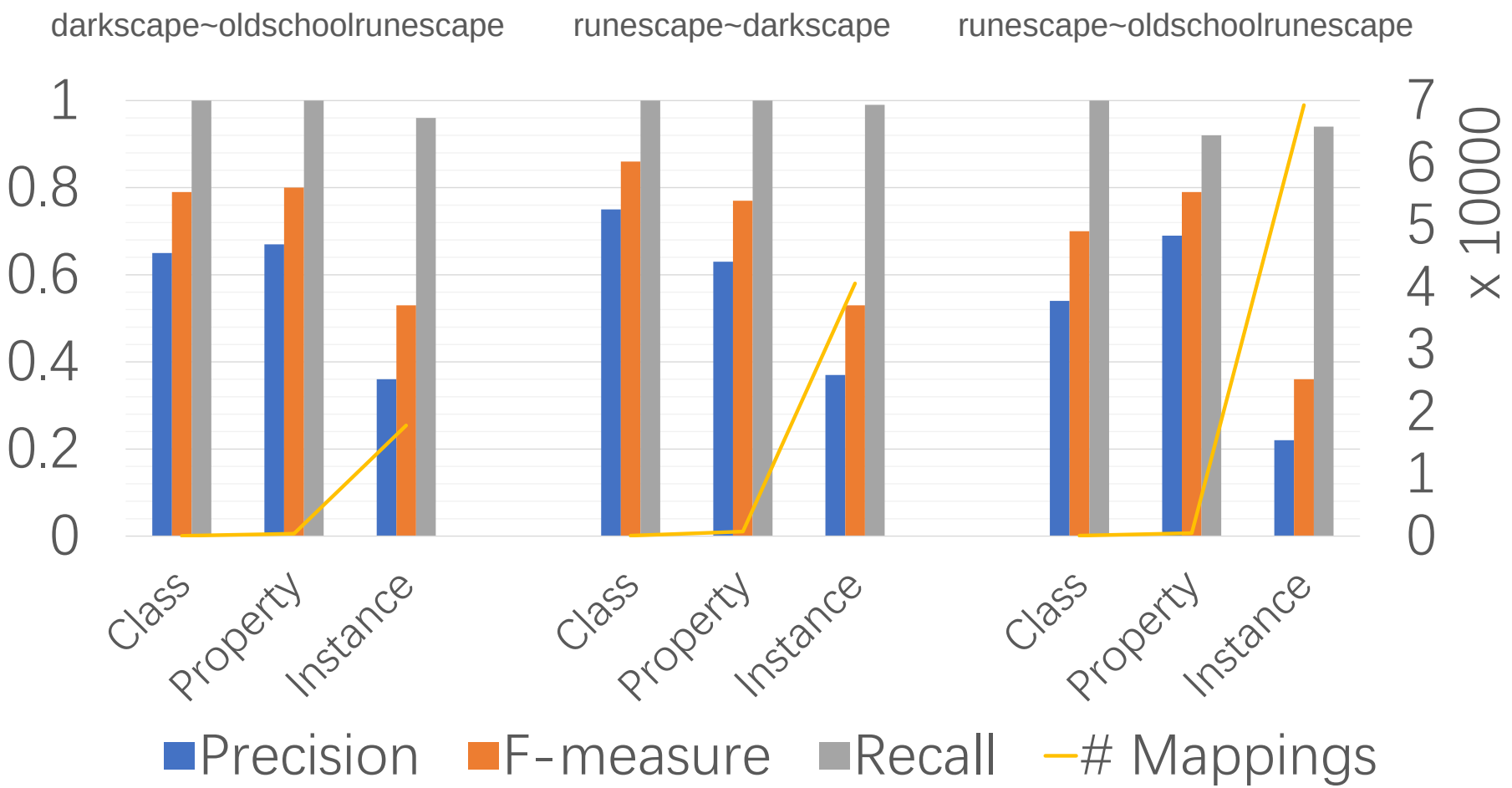
Evaluation - Knowledge Graph Track

Description

Source	Hub	# Instances	# Properties	# Classes
RuneScape Wiki	Games	200,605	1,998	106
Old School RuneScape Wiki	Games	38,563	488	53
DarkScape Wiki	Games	19,623	686	65
Marvel Database	Comics	56,464	99	2
Hey Kids Comics Wiki	Comics	158,234	1,925	181
DC Database	Comics	128,495	177	5
Memory Alpha	TV	63,240	326	0
Star Trek Expanded Universe	TV	17,659	201	3
Memory Beta	Books	63,223	413	11

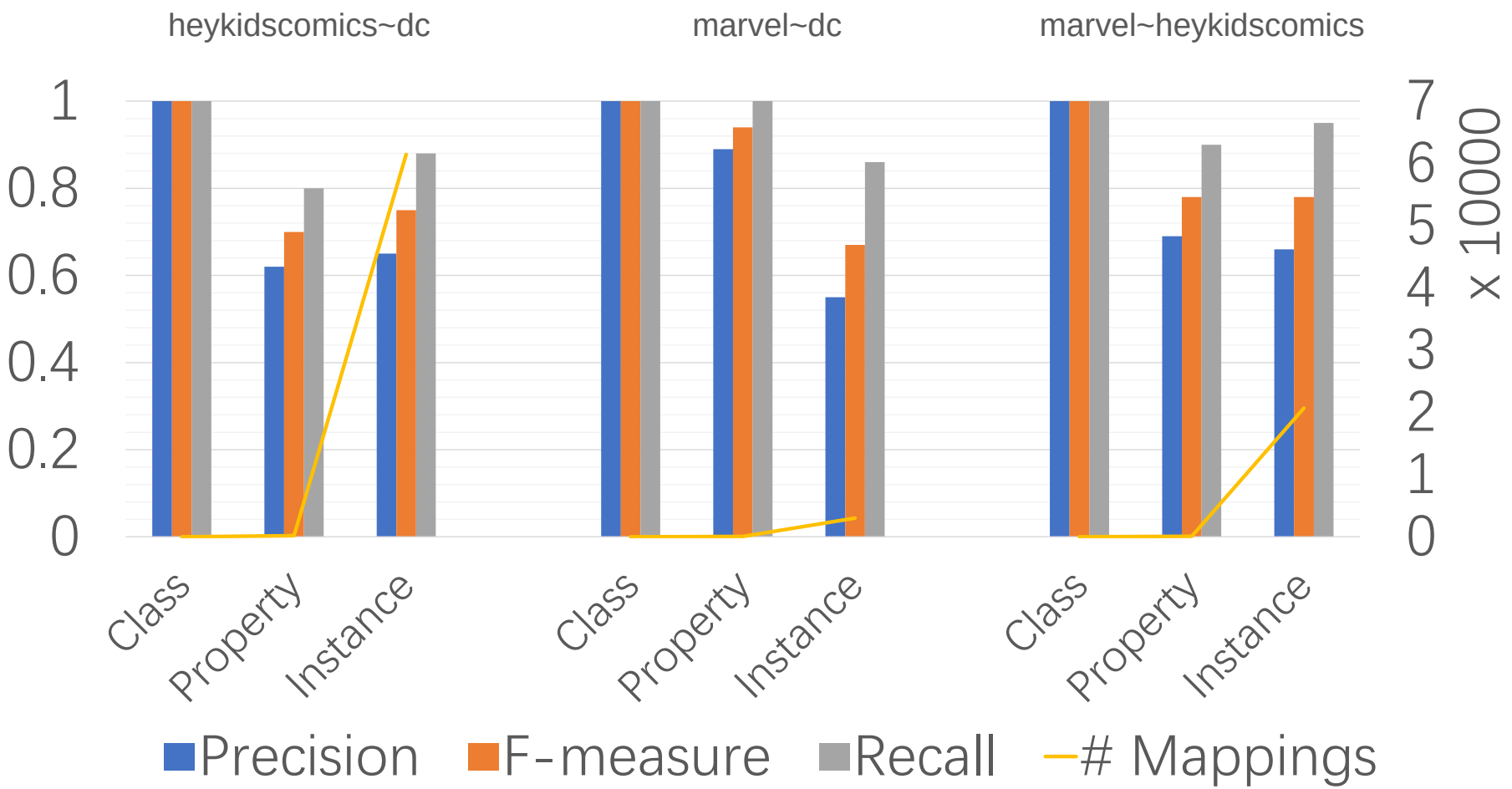
University of Chinese Academy of Sciences

Evaluation on Lexical Matching - Games



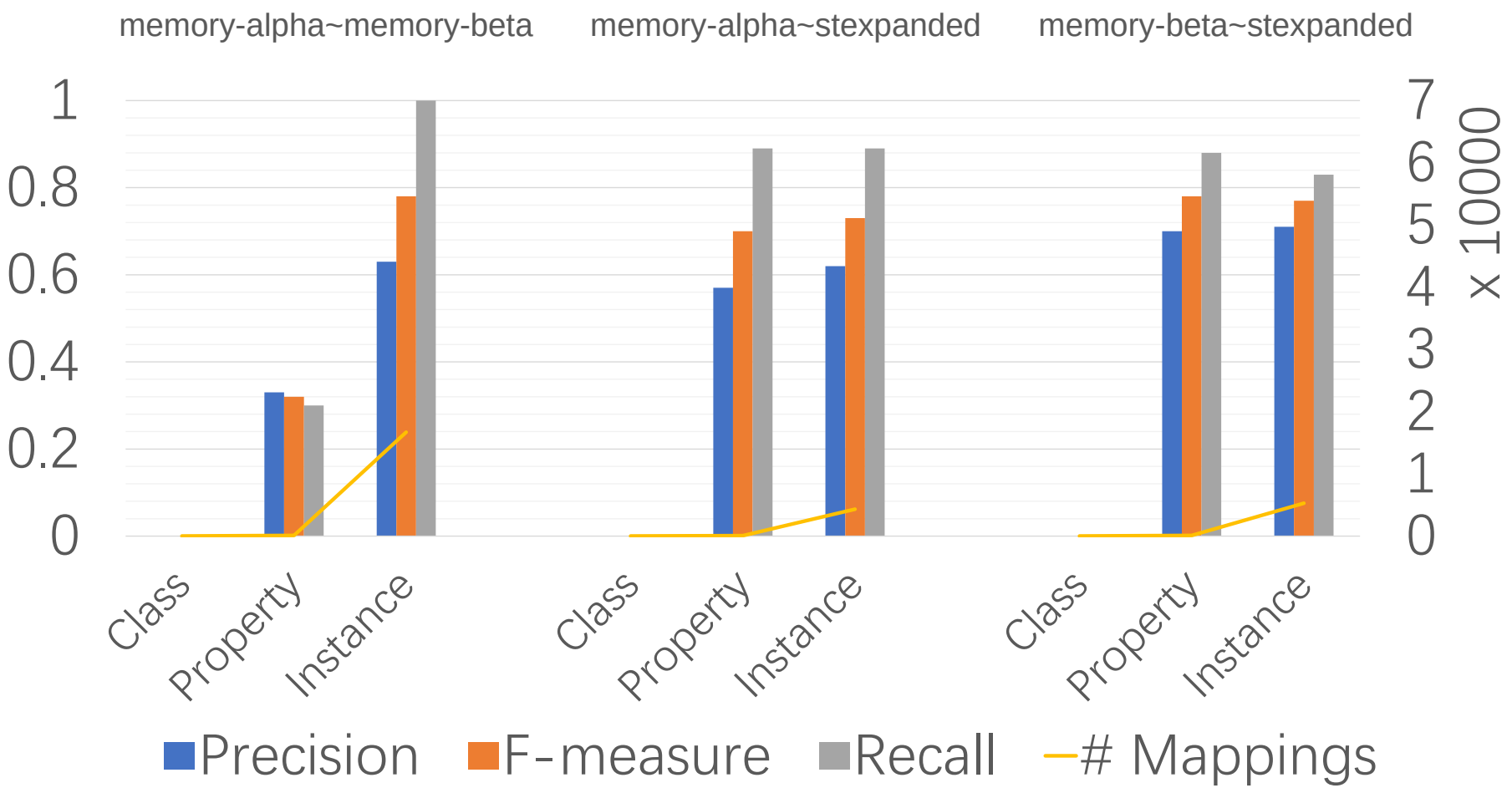
University of Chinese Academy of Sciences

Evaluation on Lexical Matching - Comics



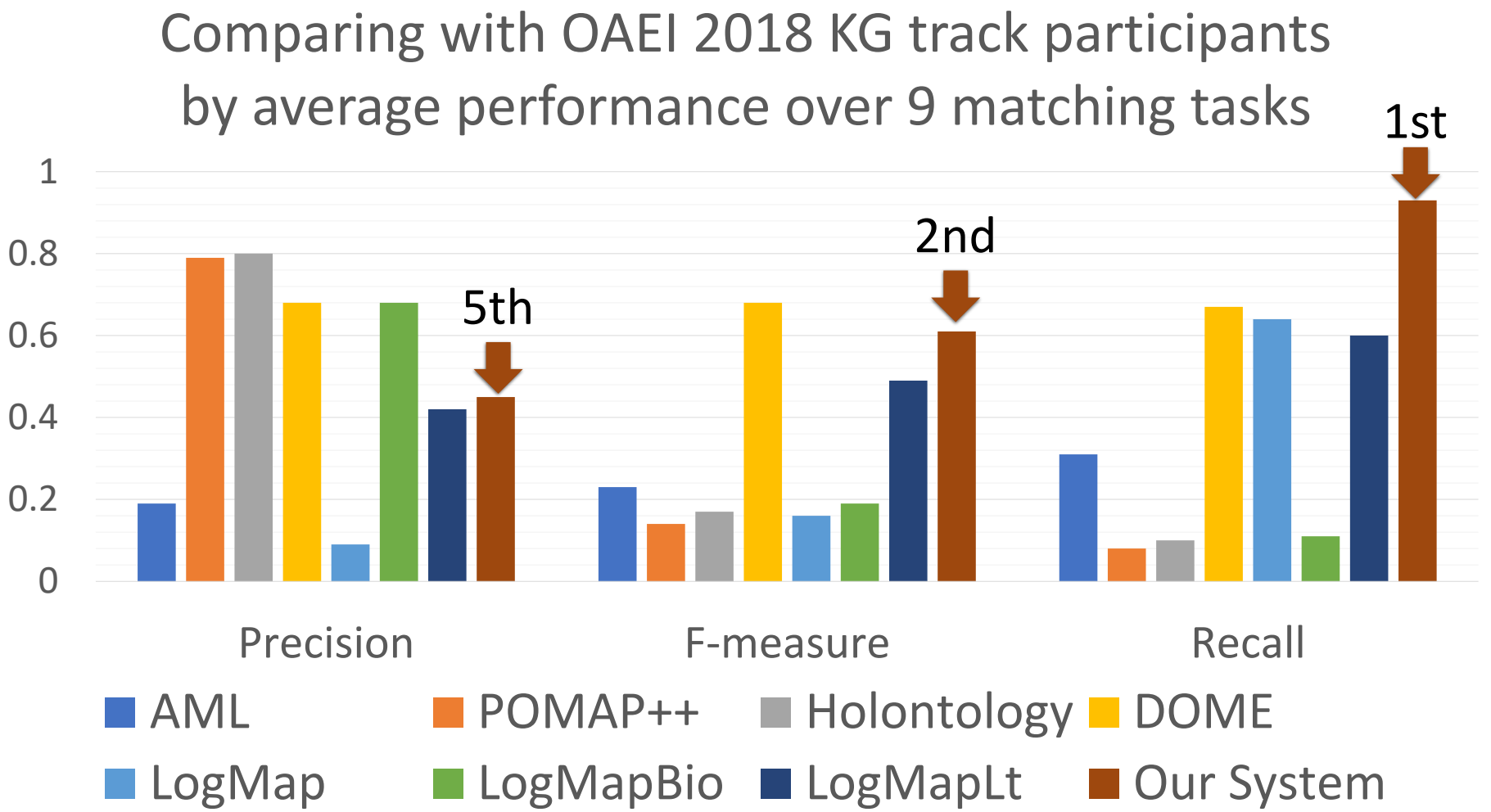
University of Chinese Academy of Sciences

Evaluation on Lexical Matching - TV & Books



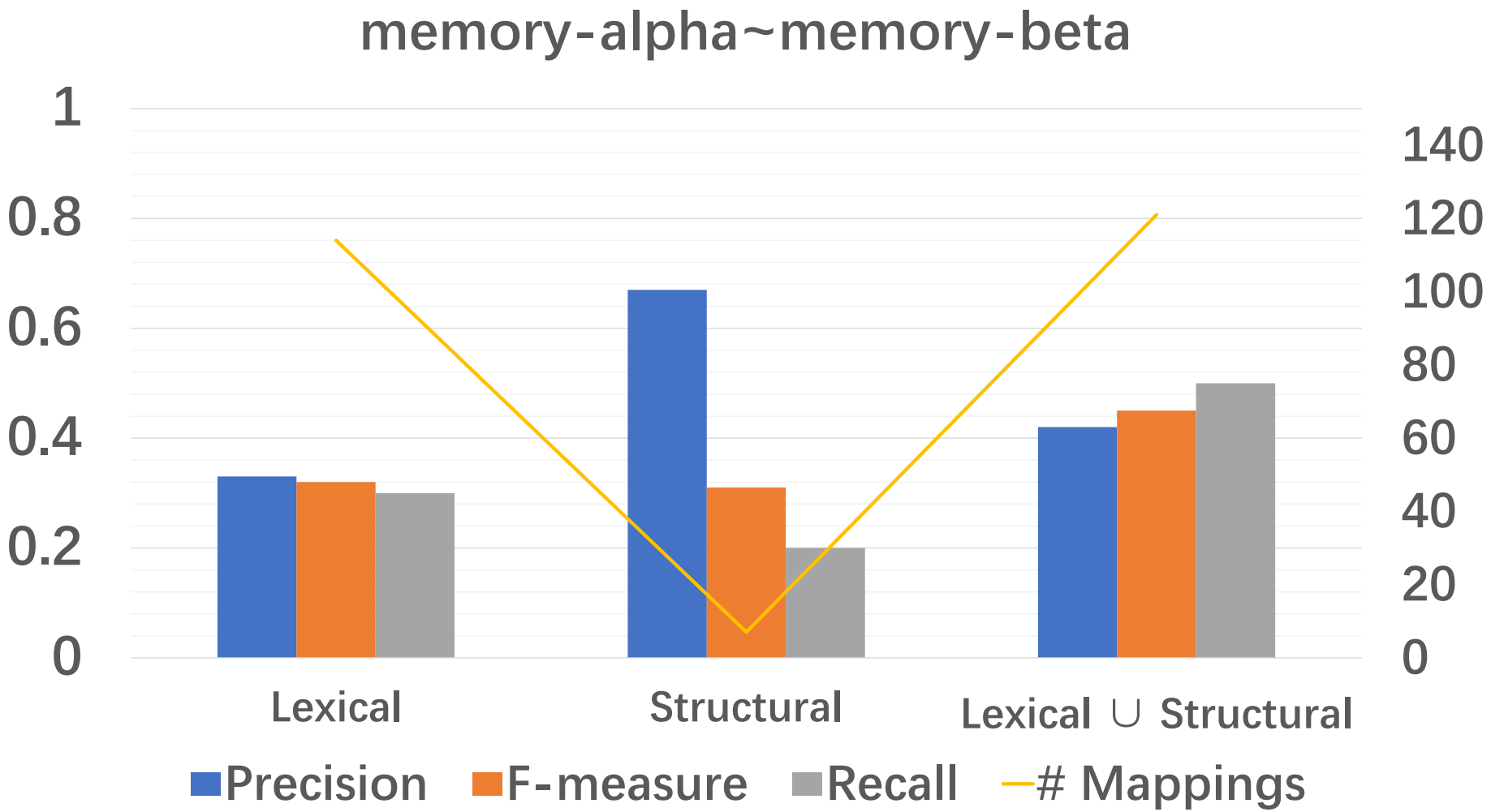
University of Chinese Academy of Sciences

Comparison - OAEI 2018



中国科学院大学
University of Chinese Academy of Sciences

Evaluation - Structural Property Matching



University of Chinese Academy of Sciences

Additional Property Mappings

Solely Identified Structurally

	Additional property mappings	
Those in the gold standard	⟨MA:relative,	MB:otherRelatives⟩
	⟨MA:wsteleplayby,	MB:teleplay⟩
Those not in the gold standard	⟨MA:wsstoryby,	MB:story⟩
	⟨MA:prev,	MB:before⟩
	⟨MA:next,	MB:after⟩
	⟨MA:abreadby,	MB:narrator⟩
	⟨MA:relative,	MB:grandparents⟩

中国科学院大学
University of Chinese Academy of Sciences

Results on the Knowledge Graph track for OAEI 2019

中国科学院大学
University of Chinese Academy of Sciences

Results for OAEI 2019 - Knowledge Graph Track

System	#	Class			
		Size	Precision	Recall	F-measure
baselineAltLabel	5	16.4	1.00	0.59	0.74
baselineLabel	5	16.4	1.00	0.59	0.74
Wiktionary	5	21.4	1.00	0.67	0.80
DOME	4	22.5	0.74	0.53	0.62
AML	4	27.5	0.78	0.61	0.69
LogMapLt	4	23.0	0.80	0.43	0.56
LogMapKG	5	26.0	0.95	0.76	0.84
FCAMap-KG	5	18.6	1.00	0.70	0.82

http://oaei.ontologymatching.org/2019/results/knowledgegraph/

中国科学院大学
University of Chinese Academy of Sciences

Results for OAEI 2019 - Knowledge Graph Track

System	#	Property			
		Size	Precision	Recall	F-measure
baselineAltLabel	5	47.8	0.99	0.66	0.79
baselineLabel	5	47.8	0.99	0.66	0.79
Wiktionary	5	75.8	0.97	0.98	0.98
DOME	4	75.5	0.79	0.75	0.77
AML	4	58.2	0.72	0.49	0.59
LogMapLt	4	0.0	0.00	0.00	0.00
LogMapKG	5	0.0	0.00	0.00	0.00
FCAMap-KG	5	69.0	1.00	0.96	0.98

http://oaei.ontologymatching.org/2019/results/knowledgegraph/

中国科学院大学
University of Chinese Academy of Sciences

Results for OAEI 2019 - Knowledge Graph Track

System	#	Instance			
		Size	Precision	Recall	F-measure
baselineAltLabel	5	4674.2	0.89	0.80	0.84
baselineLabel	5	3641.2	0.95	0.71	0.81
Wiktionary	5	3483.6	0.91	0.70	0.79
DOME	4	4895.2	0.74	0.67	0.70
AML	4	7529.8	0.72	0.69	0.71
LogMapLt	4	6653.8	0.73	0.62	0.67
LogMapKG	5	29190.4	0.40	0.86	0.54
FCAMap-KG	5	4530.6	0.90	0.79	0.84

http://oaei.ontologymatching.org/2019/results/knowledgegraph/

中国科学院大学
University of Chinese Academy of Sciences

Results for OAEI 2019 - Knowledge Graph Track

System	#	Overall			
		Size	Precision	Recall	F-measure
baselineAltLabel	5	4739.0	0.89	0.80	0.84
baselineLabel	5	3706.0	0.95	0.71	0.81
Wiktionary	5	3581.8	0.91	0.71	0.80
DOME	4	4994.8	0.74	0.67	0.70
AML	4	7615.5	0.72	0.69	0.70
LogMapLt	4	6676.8	0.73	0.61	0.66
LogMapKG	5	29216.4	0.40	0.84	0.54
FCAMap-KG	5	4792.6	0.91	0.79	0.85

http://oaei.ontologymatching.org/2019/results/knowledgegraph/

Conclusions

中国科学院大学
University of Chinese Academy of Sciences

FCA-based Systems - Ontology Matching

Systems	Objects	Attributes	External resources
FCA-Merge (Stumme and Maedche 2001)	Documents	Classes of ontologies	Textual documents from the Web
(Souza and Davis 2004; Zhao, Wang, and Halang 2006; Xu, Wu, and Chen 2010)	Classes of ontologies	Terms in thesaurus	Domain thesuarus
FCA-OntMerge (Guan-yu and Shu-peng 2010)	Classes of ontologies	Properties of ontologies	None
FCA-Map (Zhao and Zhang 2016; Zhao et al. 2018)	Names/labels/synonyms and classes of ontologies	Tokens and relations/axioms among classes	Domain thesuarus

University of Chinese Academy of Sciences

Conclusions

- This work reveals the advantage of our FCA-based approach to knowledge graph matching.
- The lexical matching part already achieved high recall in class, property, and instance matching for *OAEI 2018 Knowledge Graph* track.
- The structural matching part using a property-based lattice is able to identify additional property mappings.

University of Chinese Academy of Sciences

Future Work

- An alternative validation strategy so as to ensure the quality of mappings.
- An iterative framework so as to perform class, property, and instance matching in an augmented way until no further matches can be found.
- A formal context constructed from multiple sources so as to apply it to three or more knowledge graphs matching directly.
- This work is ongoing and the source code is open. (Chen 2019)

Guowei Chen. *FCA-Map source code*. 2019. URL: https://github.com/icgw/FCA-Map (visited on 10/26/2019).

Thanks for your attention!