

FIT 5190 Introduction to IT research methods

Type Inference On Noisy Linked Data

Keywords: Type Inference, Bag-of-word, Linked data graph

Name: Chen Jing	ID: 2819****
Name: Pan Luming	ID: 2819****



Contents



01

Part one
What is the
problem?



02

Part two
How to address
the problem?



03

Part three
Conclusions



04

Part four
References



PART 1

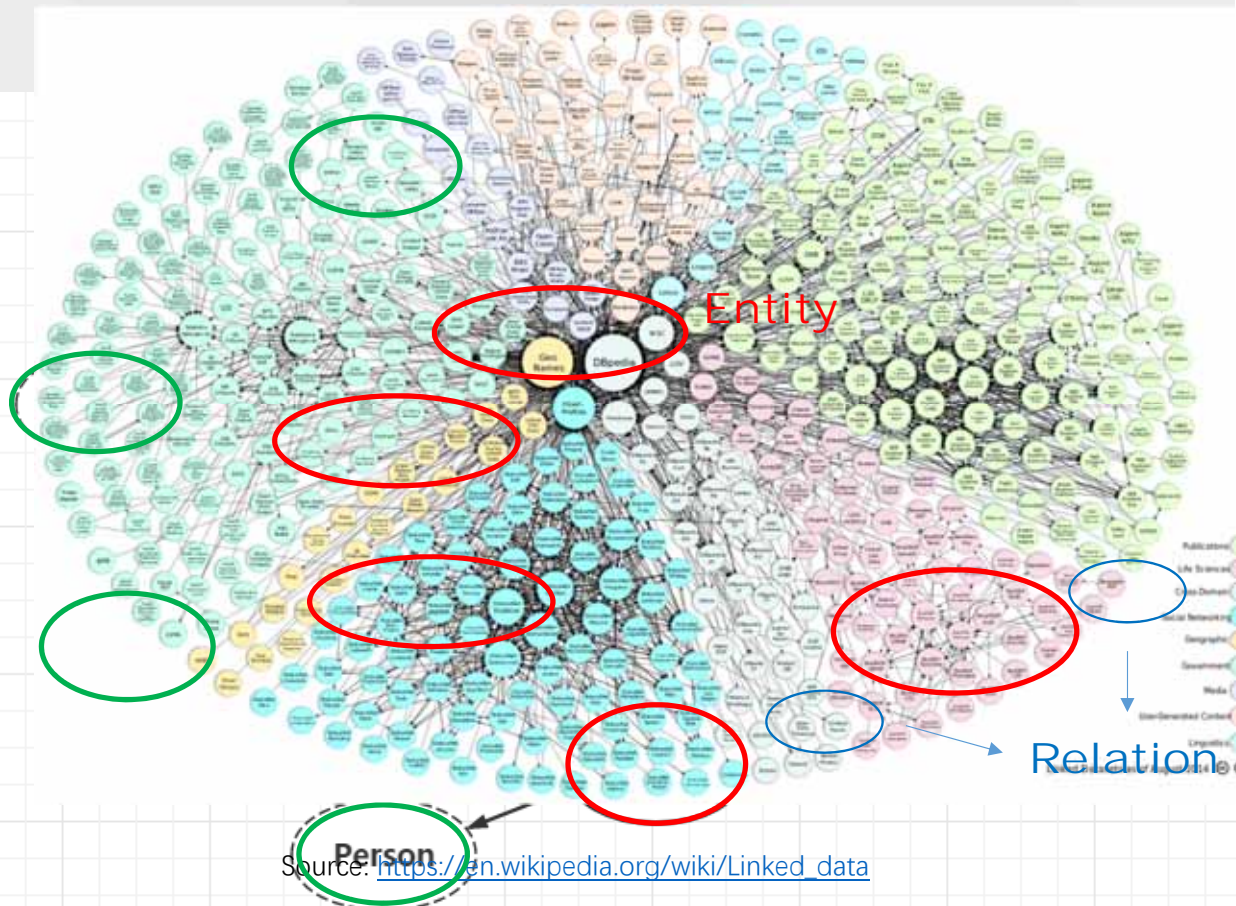
What's the problem?

Speaker: Chen Jing

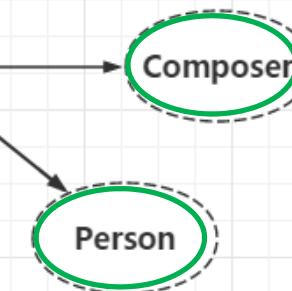
Background and concept

Domain	Relation	Range
Entity	<i>type</i>	Class
Song	<i>writtenby</i>	Composer
Person	<i>haswife</i>	Person
Entity	<i>age</i>	Int
...

Schema table



Linked data bases
for semantic search



Abstract linked data graph of a database

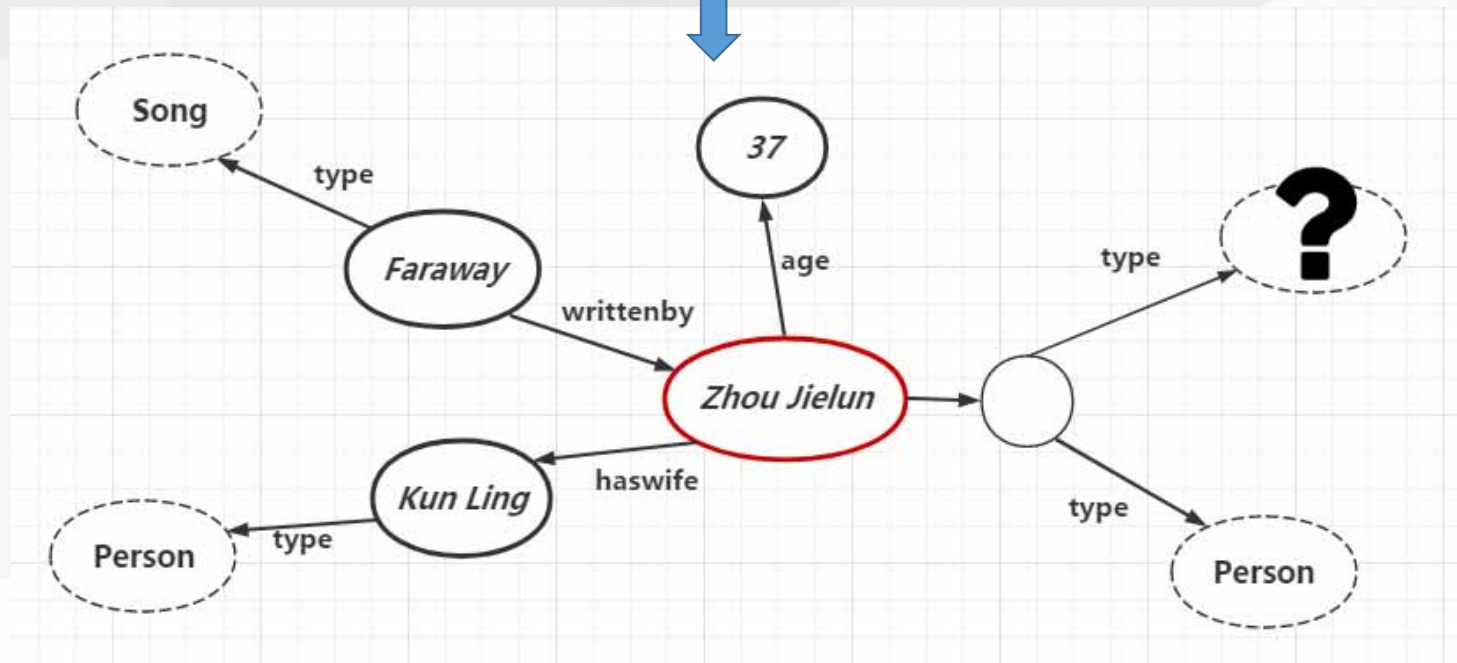
Missing type: A common phenomenon



Crowdsourcing way

Mistake

lead to

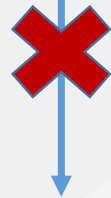


Missing

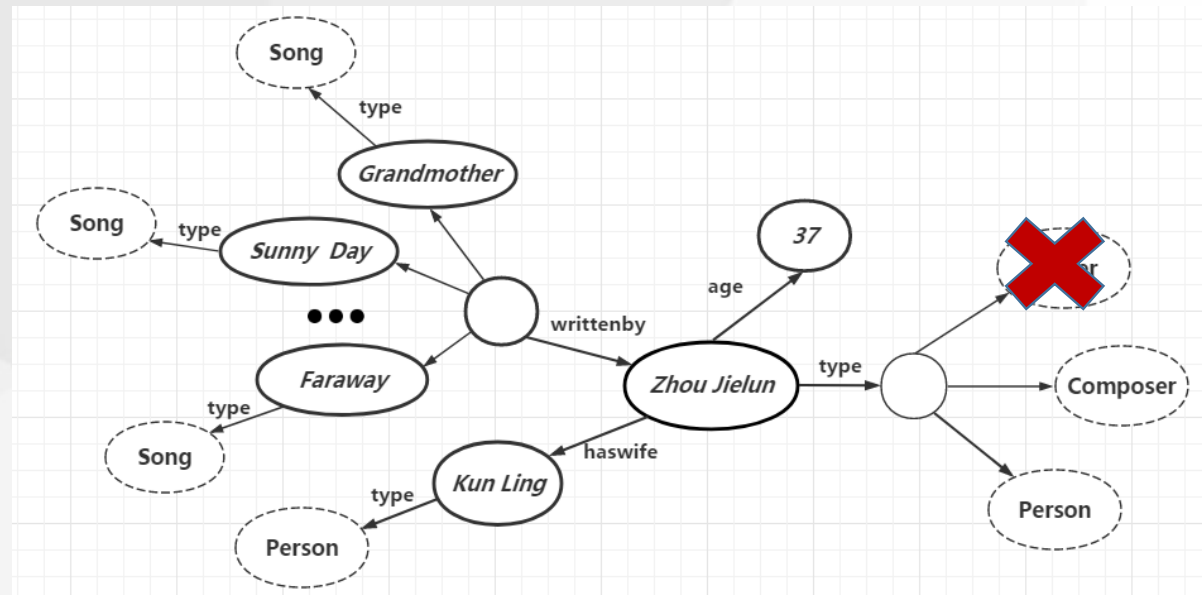
Missing type: Influence semantic search

Sparql: searching for singers whose age is over 30

```
PREFIX job: <http://example.com/job#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?name
WHERE {
  ?person foaf:name ?name.
  ?person a foaf:Person.
  ?person a job:Singer.
  ?person foaf:age ?age.
  FILTER (?age > 30)
}
```



Matching result: *Zhou Jielun*



Predict missing type !



PART 2

How to address the problem?

Speaker: Pan Luming, Chen Jing

Text-classification-based method

A. Why text?

Assumption: Texture information indicates the type

Inspiration: *Tipalo* (Gangemi et al., 2012)

B. Why classification?

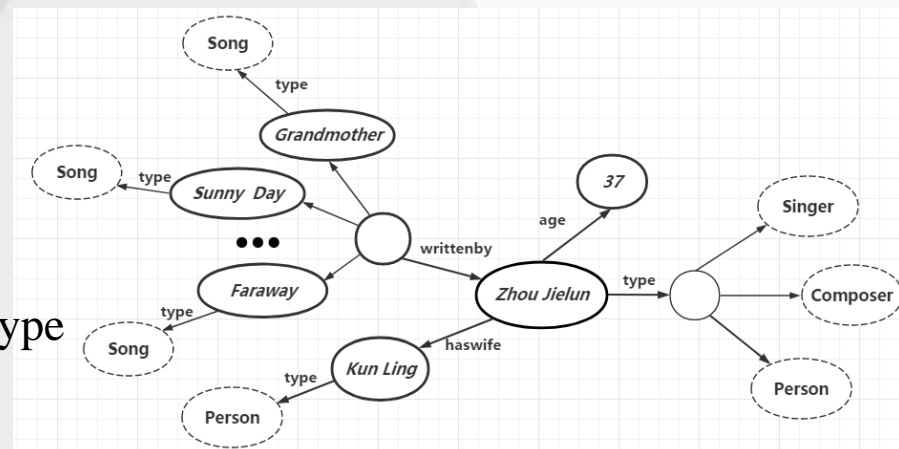
Assumption: Classification method can predict the type

Inspiration: Support vector machine (Sleeman et al., 2015)

C. Additional assumptions?

Assumption: Different relations has different influence on prediction

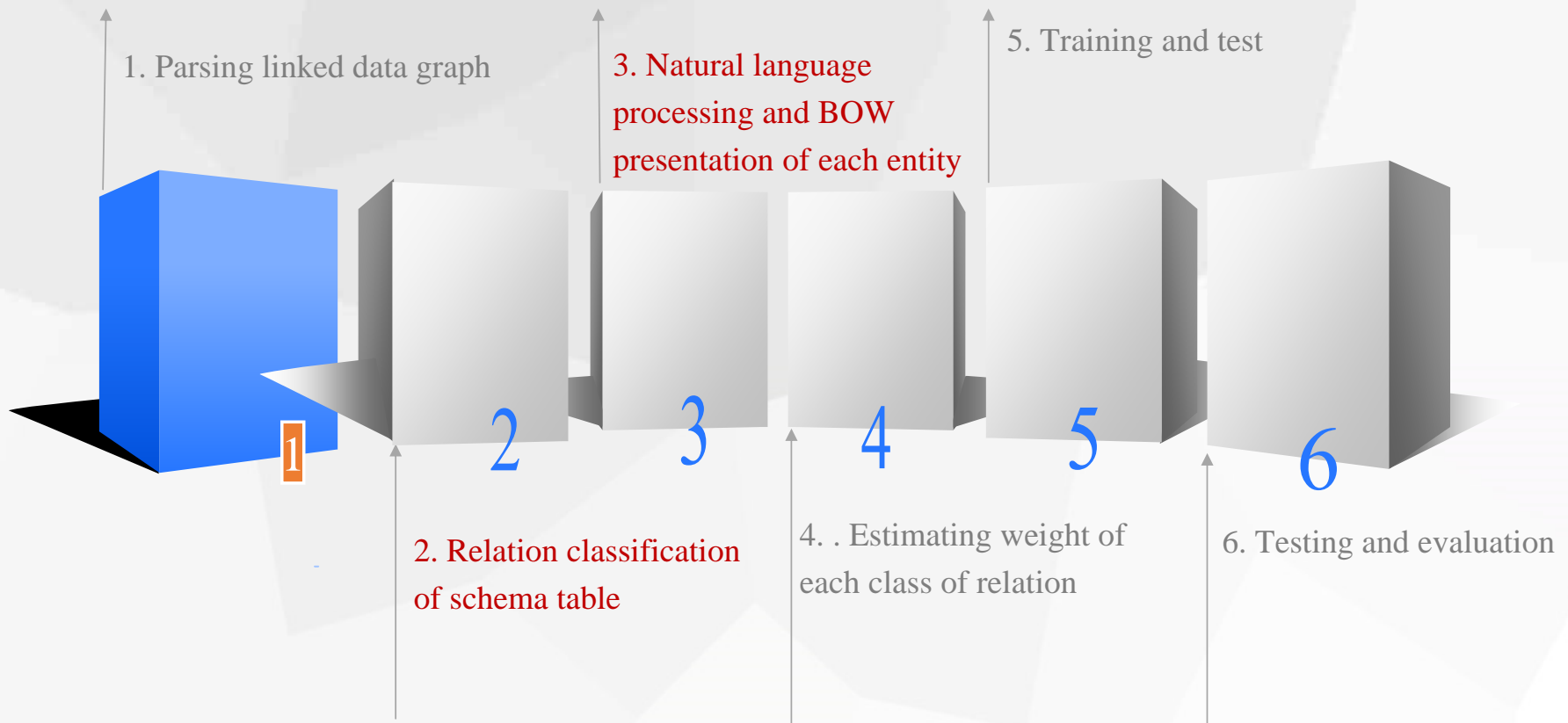
Inspiration: *SD-Type* (Paulheim et al., 2013)



**Bag-of-words
(BOW)
classification**

**Relation
classification**

Procedure(six steps)



Bag-of-words (BOW)

(1) John likes to watch movies. Mary likes movies too.

(2) John also likes to watch football games.

BOW of (1):

```
[  
  "John": 1,  
  "likes": 2,  
  "to": 1,  
  "watch": 1,  
  "movies": 2,  
  "also": 0,  
  "football": 0,  
  "games": 0,  
  "Mary": 1,  
  "too": 1  
]
```

BOW of (2):

```
[  
  "John": 1,  
  "likes": 1,  
  "to": 1,  
  "watch": 1,  
  "movies": 0,  
  "also": 1,  
  "football": 1,  
  "games": 1,  
  "Mary": 0,  
  "too": 0  
]
```

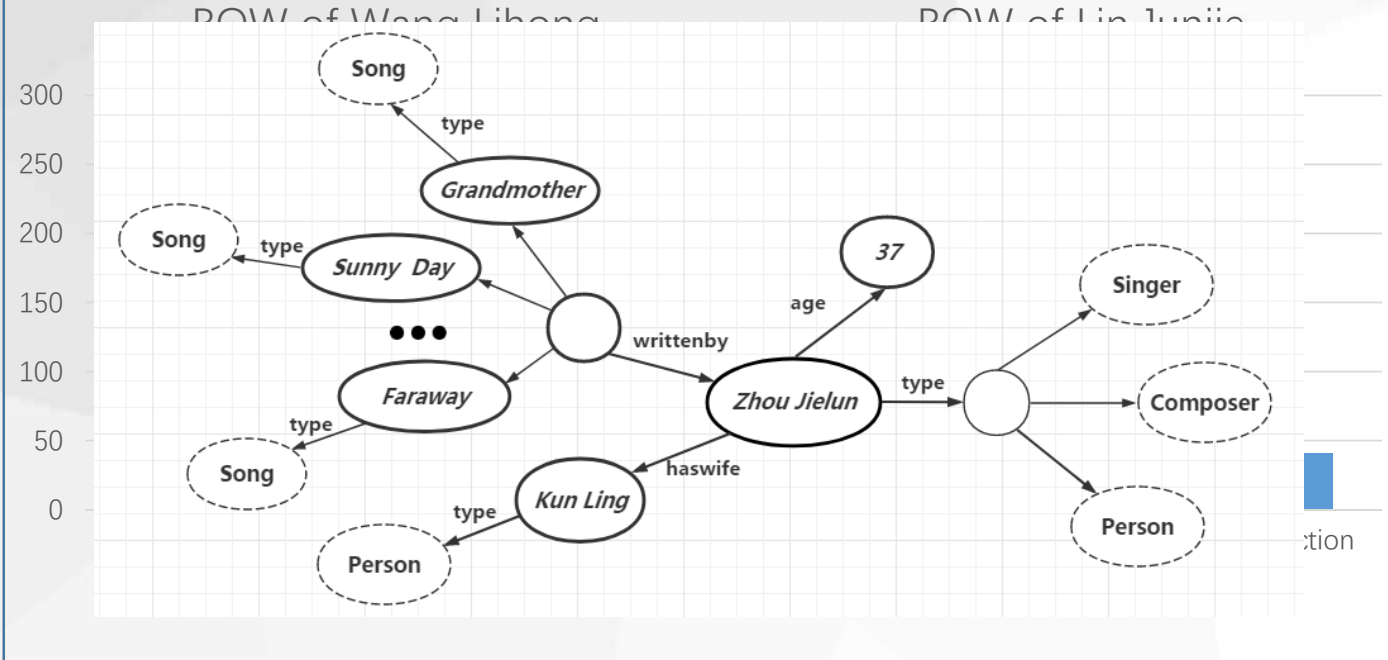
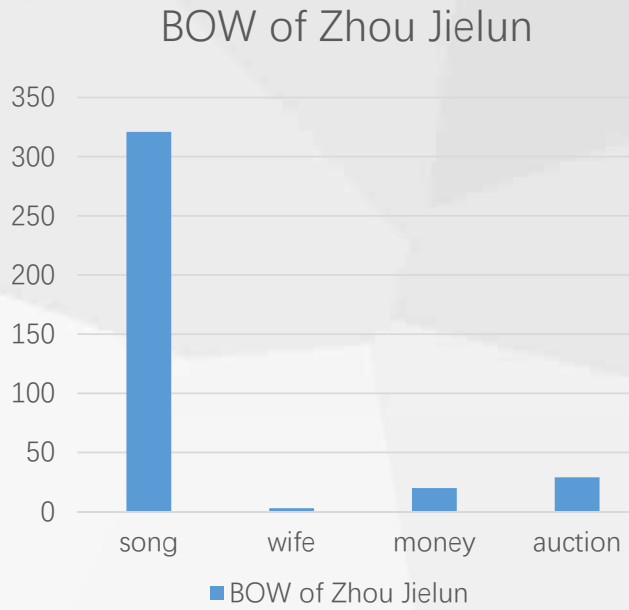
Examples from Wikipedia:

https://en.wikipedia.org/wiki/Bag-of-words_model

Bag-of-word (BOW)

Testing

Training



No_song * **Weight_song_composer** = Precision(song_composer)

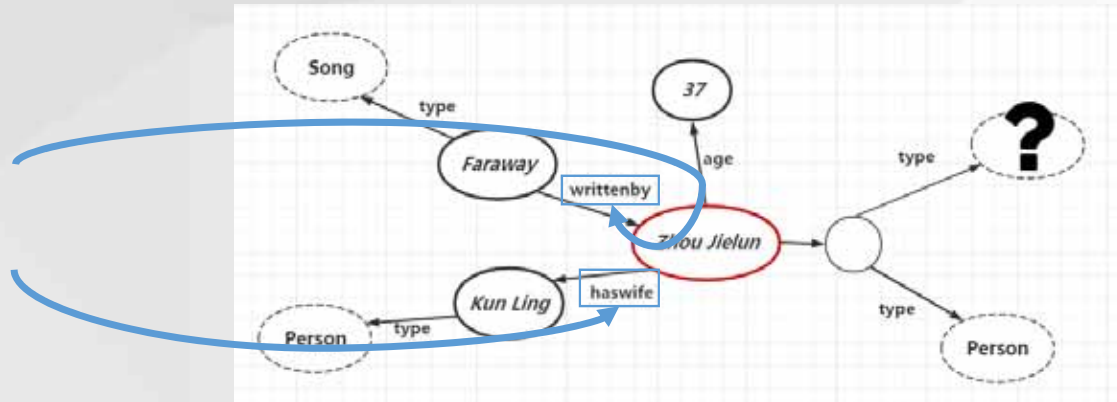
 \succ

No_song * Weight_song_father = Precision(song_father) ❌

Relation classification

A. Based on direction:

1. Incoming relation
2. Outgoing relation



B. Based on domain and range:

1. Homogeneous relation
2. Heterogeneous relation

Domain	Relation	Range	Homogeneous /heterogeneous
Entity	<i>type</i>	class	homogeneous
Song	<i>writtenby</i>	composer	heterogeneous
person	<i>haswife</i>	person	homogeneous
Entity	<i>age</i>	Int	heterogeneous
...	

Relation classification(cont.)

BOW divided

predicted entity:
Zhou Jielun node

in-homogeneous:

in-heterogeneous:
many songs node
(*writtenby*)

out-homogeneous:
his wife node
(*haswife*)

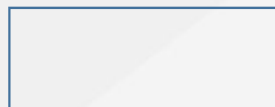
out-heterogeneous:
Age node
(*age*)

**BOW
presentation:**



**Predict type
dependently:**

Person



Composer,
Singer



Millionaire



Father

Target type:

- ① Composer
- ② Millionaire
- ③ Father



PART 3

Conclusion and future work

Lecturer: Chen Jing



Conclusion and future work

A. Conclusion:

1. Consider Linked data structured feature
2. Text-based classification method
3. Solve the type missing problem

B. Improvement and future work

1. Not suitable for huge linked data base (bag-of-word)
2. Tensor decomposition or distributed method

PART 4

References

- Acosta, M., Zaveri, A., Simperl, E., Kontokostas, D., Auer, S., & Lehmann, J. (2013). Crowdsourcing linked data quality assessment. *International Semantic Web Conference*, 260-276.
- Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data-the story so far. *Semantic services, interoperability and web applications: emerging concepts*, 205-227.
- Gangemi, A., Nuzzolese, A., Presutti, V., Draicchio, F., Musetti, A., & Ciancarini, P. (2012). Automatic typing of DBpedia entities. *The Semantic Web-ISWC 2012*, 65-81.
- Ji, Q., Gao, Z., & Huang, Z. (2011). Reasoning with noisy semantic data. *Extended Semantic Web Conference*, 497-502.
- Paulheim, H., & Bizer, C. (2013). Type inference on noisy RDF data. *International Semantic Web Conference*, 510-525.
- Sleeman, J., Finin, T., & Joshi, A. (2015). Entity type recognition for heterogeneous semantic graphs. *AI Magazine*, 36(1), 75-86.

THANKS!
Q & A
