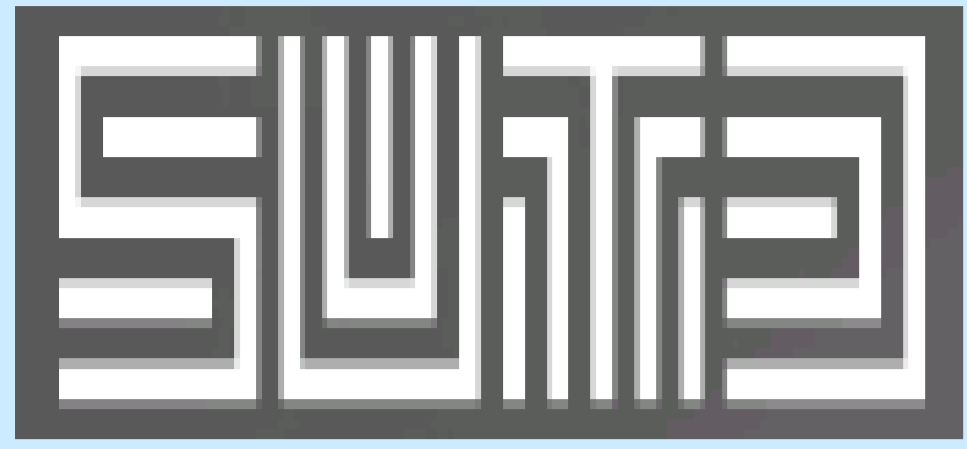


# ZORE: A Syntax-based System for Chinese Open Relation Extraction



Likun Qiu and Yue Zhang

Singapore University of Technology and Design, Singapore

## Motivation and goal:

- Motivation: Existing ORE techniques pay little attention to give semantic analysis to extracted relations, which is the advantage of traditional IE.
- Goal: to extract relations by using syntactic dependency patterns, while associating them with explicit semantic information

Relation	奥巴马 总统	毕业	于	哈佛 法学院
Semantic Pattern:	nsubj-NR(Af)	Pred(毕业)	prep-于	pobj-NN(Di)
Syntactic Pattern:	nsubj-NR(A)	Pred(毕业)	prep-于	pobj-NN(A)
Semantic Signature:	Af			Di
Predicate and Arguments:	argument 1	predicate phrase		argument 2
Base NPs:	base NP 1			base NP 2
Sentence:	奥巴马 总统	毕业	于	哈佛 法学院
	Obama President	graduated	from	Harvard Law School

1

## Semantic Tagging by Double Propagation

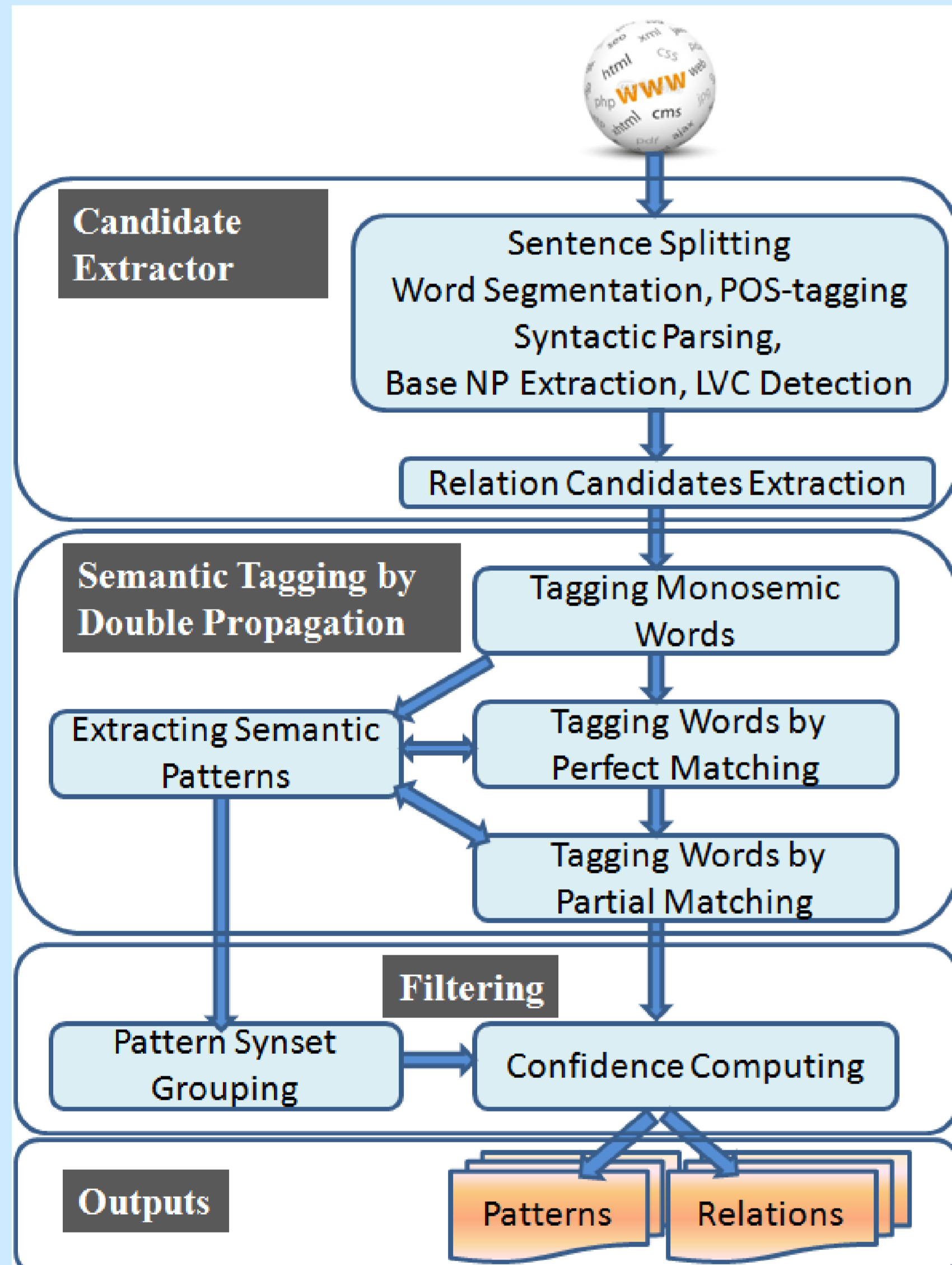
- Step1: Tagging Monosemic Arguments:
  - ✓ Take the semantic category of the head word of a base NP as the semantic category of the base NP
  - ✓ Only monosemic words are tagged
  - ✓ Tagged relation candidate are generalized into semantic patterns  $Set^{SemPat}$ , consisting of syntactic patterns and semantic signatures
- Step2: Tagging by Perfect Pattern Matching
  - ✓ Acquire a set of possible semantic categories according to the characters of a head word.
  - ✓ Acquire possible semantic signatures of untagged relation candidates and match patterns in  $Set^{SemPat}$ . The matched pattern with highest frequency is taken as the final pattern for the current relation, and so the relation is semantically tagged.
- Step3: Tagging by Partial Pattern Matching
  - ✓ Tag the ambiguous and unknown words by partial matching rather than perfect matching of the whole semantic pattern.
  - ✓ This can be treated as a back-off of the last step.

4

## Architecture of ZORE

Three components:

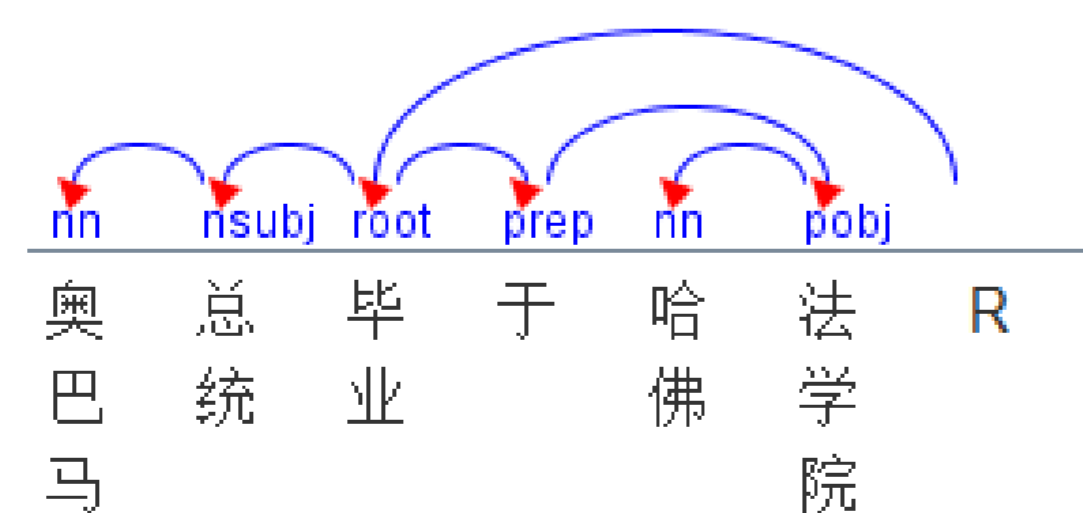
- (1) consumes input text and outputs a set of relation candidates.
- (2) tags relations and extracts semantic patterns by a double propagation algorithm.
- (3) Group patterns into synsets, and filter relations.



2

## Dependency parsing

Parsing result of the example sentence in Figure 1, in Stanford dependencies.



## Relations candidates

**Common and dummy LVC relations:** the predicate phrase of the relation is an LVC (e.g., a light verb and a nominal object)

Verb	Noun
进行 (do) (*)	发行 (distribution), 分析 (analysis), 收集 (collection),
有 (have) (*)	影响 (effect), 贡献 (contribution), 兴趣 (interest),
产生 (generate) (**)	影响 (effect), 兴趣 (interest), 怀疑 (doubt),
造成 (cause) (**)	影响 (effect), 破坏 (destruction), 伤害 (harm),
表示 (express) (**)	满意 (satisfaction), 欢迎 (welcome), 尊重 (respect),
展开 (launch) (**)	调查 (investigation), 攻击 (attack), 攻势 (offensive),

**Verb relations:** a verb acts as the predicate phrase (see the instance in Figure 1)

## Relative-clause relations

毕业 (graduate) 于 (from) 哈佛 (Harvard) 法学院 (Law School) 的 (de, an auxiliary word) 奥巴马 (Obama) 总统 (president)

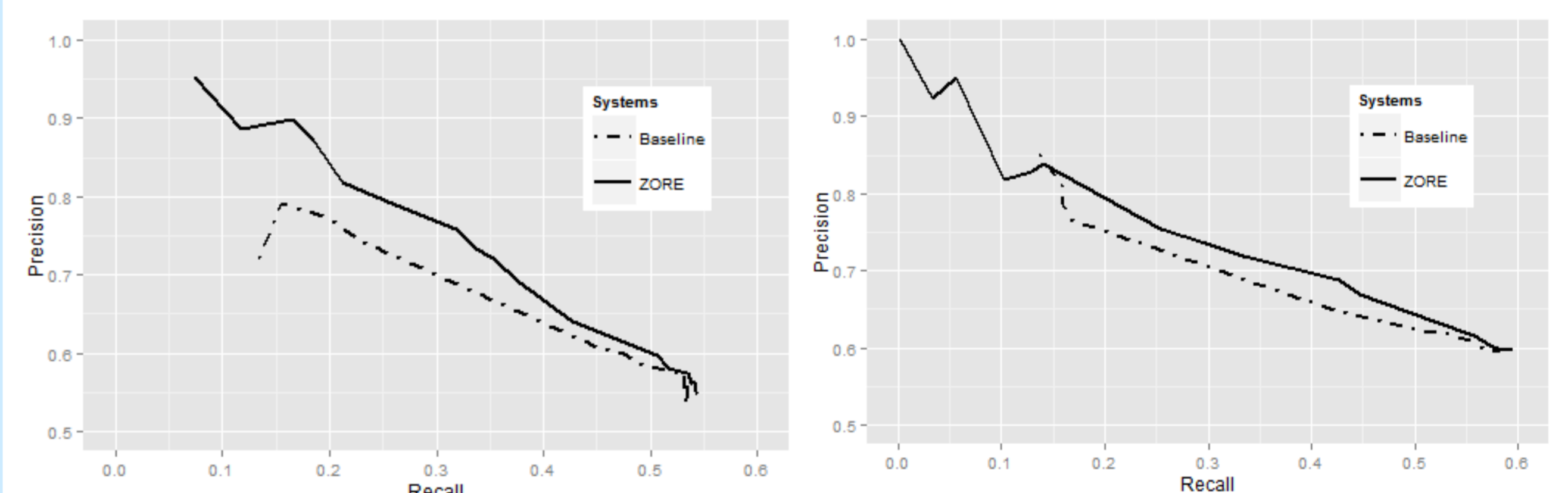
3

## Experiments

- Annotated relation datasets

Dataset	Source	#Sen	#Rel
Wiki-500	Chinese Wikipedia	500	561
Sina-500	Sina News	500	707

- Performance on Wiki and news



Sentences	Step 1			Step 2			Step 3		
	P	R	F1	P	R	F1	P	R	F1
10K	0.947	0.032	0.062	<b>0.960</b>	0.043	0.082	0.933	0.075	0.139
50K	0.894	0.075	0.138	0.922	0.105	0.189	0.907	0.139	0.241
100K	0.897	0.093	0.169	0.924	0.130	0.228	0.909	0.160	0.272
200K	0.901	0.114	0.202	0.926	0.157	0.268	0.892	0.191	0.315
500K	0.891	0.146	0.251	0.909	0.196	0.322	0.860	0.230	0.363
1M	0.860	0.164	0.275	0.885	0.219	0.351	0.842	0.248	0.383
2M	0.797	0.182	0.296	0.819	0.250	0.383	0.788	0.278	0.411
3M	0.784	0.187	0.302	0.802	0.253	0.385	0.778	0.282	0.414
4M	0.739	0.178	0.287	0.801	0.258	0.390	0.778	0.287	0.419
5M	0.779	0.189	0.304	0.798	0.260	0.392	<b>0.768</b>	<b>0.289</b>	<b>0.420</b>

Table 5: Accuracies on different numbers Wiki sentences.

5

## Conclusion

- A parsing-based System
- A system for Chinese open information extraction
- Joint WSD and relation extraction
- Both the system and data are freely available at: <https://sourceforge.net/projects/zore>

6