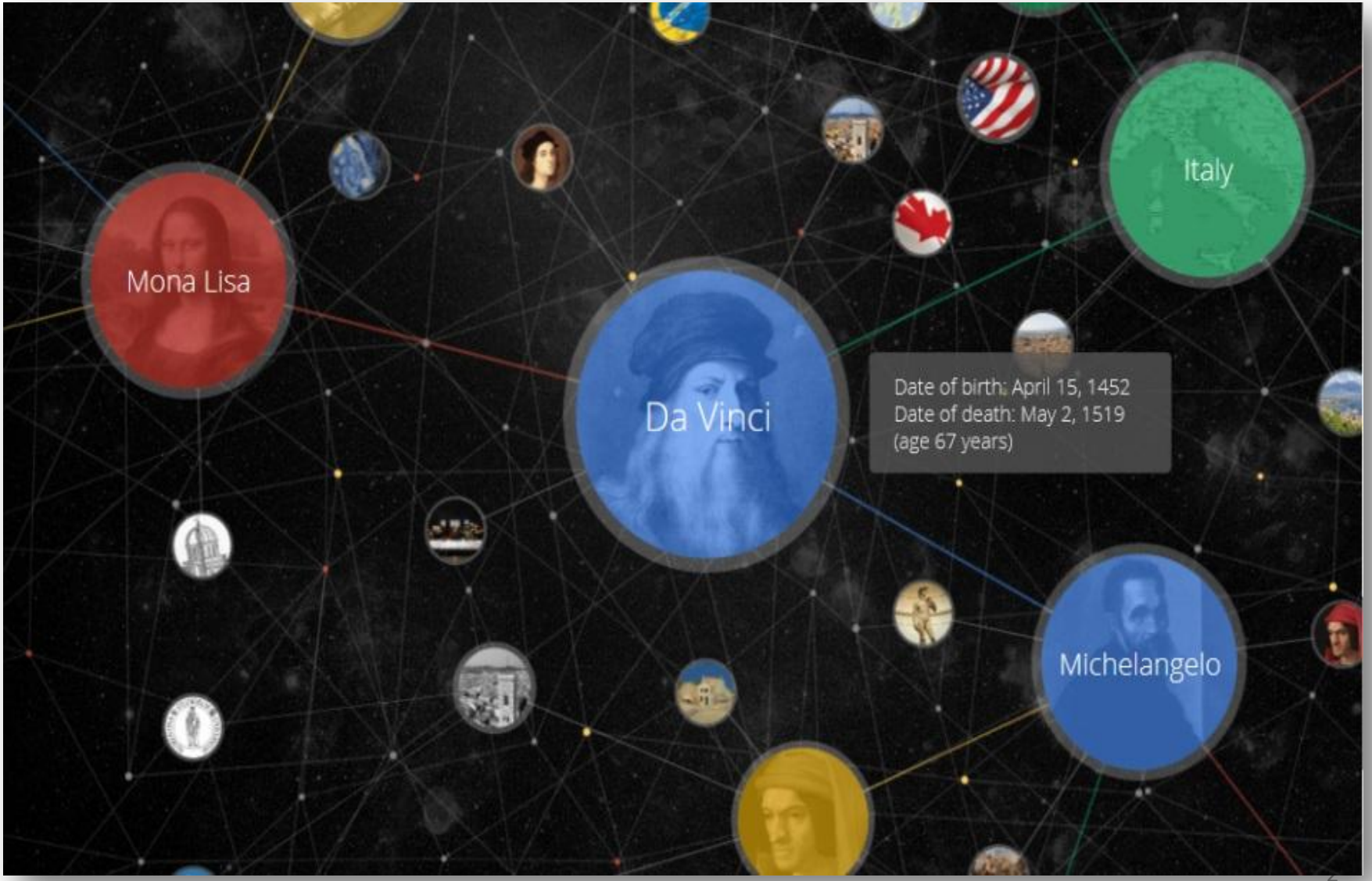
Language Understanding and Knowledge Base Construction: Walk to Arrive Together

Maosong SUN
Dept. of CS, Tsinghua University

Nanjing, Google Workshop

knowledge graph is great



Infobox is far from enough for knowledge discovery

主条目：南京历史

古代 [编辑]

南京一带在100万年至120万年前就有古人类活动。约7000年前，出现了以北阴阳营文化为代表的新石器时代原始村落。3000年前，相当于中原的商周之际，秦淮河流域出现了密集的原始聚落，被称为湖熟文化。春秋战国时，在这些聚落的基础上形成了南京地区最早的城邑。前571年（周灵王元年），楚国在今六合区设棠邑，是南京政区建置的开始。据传说，吴王夫差前495年曾在南京筑冶城[注 3]。前472年（周元王四年），越国大夫范蠡在今中华门外秦淮河南岸筑越城[注 4]，是现南京城区建城史的开始。前333年（周显王三十六年），楚威王在石头山筑金陵邑[注 5]，是现南京城区内设治所之始，南京的别称"金陵"由此而来[13]。前210年（秦始皇三十七年），金陵邑改为秣陵县[注 6]。直至汉末，现南京地区只设县级政区。[14]

东汉末年，割据江东的孙权于211年将治所移到秣陵，在金陵邑旧地筑石头城要塞，次年改秣陵为建业。229年，孙权称帝建立东吴，将都城从武昌迁至有"钟山龙盘，石头虎踞"之称的建业，开启了南京的都城史。[注 7]西晋灭吴后，282年（太康三年）改建业为建邺，313年（建兴元年）又改建康。西晋灭吴后仅三十年，就亡于永嘉之乱，317年晋宗室司马睿在建康建立东晋，北方人口纷纷南迁。在此后约三百年的南北大分裂时期，建康成为华夏的正朔所在。420年东晋灭亡后，宋、齐、梁、陈四朝

人口	
总人口(2012)	816.1万人
- 市区常住人口(2012)	816.1万人
- 市区城镇常住人口(2012)	654.8万人
人口密度	1238.95人/平方千米
官方语言	普通话
方言	江淮官话洪巢片南京小片南京话
经济	
GDP(2012)	7,201.57 亿元（本币）
人均GDP	88,908元
市区GDP(2012)	7,201.57 亿元（本币）
其他	
时区	UTC+8（东八区）
市树	雪松
市花	梅花
邮政编码	210000-213000
电话区号	+86 (0)25
车牌号码	苏A

KB is far from enough

- Low coverage
- Lack of event description

  related to verb

- NLP, Open text understanding
- Parsing the Web (Slav Petrov and Ryan McDonald 2012)

The best accuracies are in the 80-84% range for F1 and LAS; even part-of speech accuracies were just above 90%.

Parsing Chinese is extremely difficult

约7000年前，出现了以北阴阳营文化为代表的新石器时代原始村落。

Parse

```
(ROOT
  (IP
    (VP
      (ADVP (AD 约))
      (LCP
        (IP
          (VP
            (LCP
              (NP (NT 7000年))
              (LC 前))
            (PU ，)
            (VP (VV 出现) (AS 了))))
        (LC 以北))
      (ADVP (AD 阴阳))
      (VP (VV 营)
        (NP
          (CP
            (IP
              (NP (NN 文化))
              (VP (VV 为)
                (NP (NN 代表))))
            (DEC 的))
          (NP
            (NP
              (ADJP (JJ 新))
              (NP (NN 石器)))
            (NP (NN 时代)))
          (ADJP (JJ 原始))
          (NP (NN 村落))))))
  (PU 。)))
```

Parsing in a big data thinking

约7000年前，出现了以北阴阳营文化为代表的新石器时代原始村落。

出现了原始村落。

Parse

```
(ROOT
  (IP
    (VP (VV 出现) (AS 了)
      (NP
        (ADJP (JJ 原始))
        (NP (NN 村落))))
    (PU 。)))
```

Parsing in a big data thinking

八大古都 - 搜搜百科
6000年前南京就出现了原始村落,在1950年代发掘的北阴阳营遗址年代约为为公元前4000年至前3000年,为长江下游地区新石器时代文化类型——北阴阳营文化,南京地区早期...
baike.soso.com/v1342...htm 2013-01-14 - 百度快照

土木工程发展简史-正文 - 【人人分享-人人网】
[图文] 人类最初居无定所,利用天然掩蔽物作为居处,农业出现以后需要定居,出现了原始村落,土木工程开始了它的萌芽时期。随着古代文明的发展和社会进步,古代土木工程经历了它...
blog.renren.com/share/233359176/1069... 2013-10-24 - 百度快照

6000年前南京就出现了原始村落,

出现了原始村落。

Parse

```
(ROOT
  (IP
    (VP
      (LCP
        (NP (NT 6000年))
        (LC 前))
      (NP (NR 南京))
      (ADVP (AD 就))
      (VP (VV 出现) (AS 了)
        (NP
          (NP
            (ADJP (JJ 原始))
            (NP (NN 村落)))
          (NP (NN ,)))))))
```

Philosophy of paring in the big data thinking

- 林子大了什么鸟都有

  Any kind of birds could be possible as the forest is so huge!

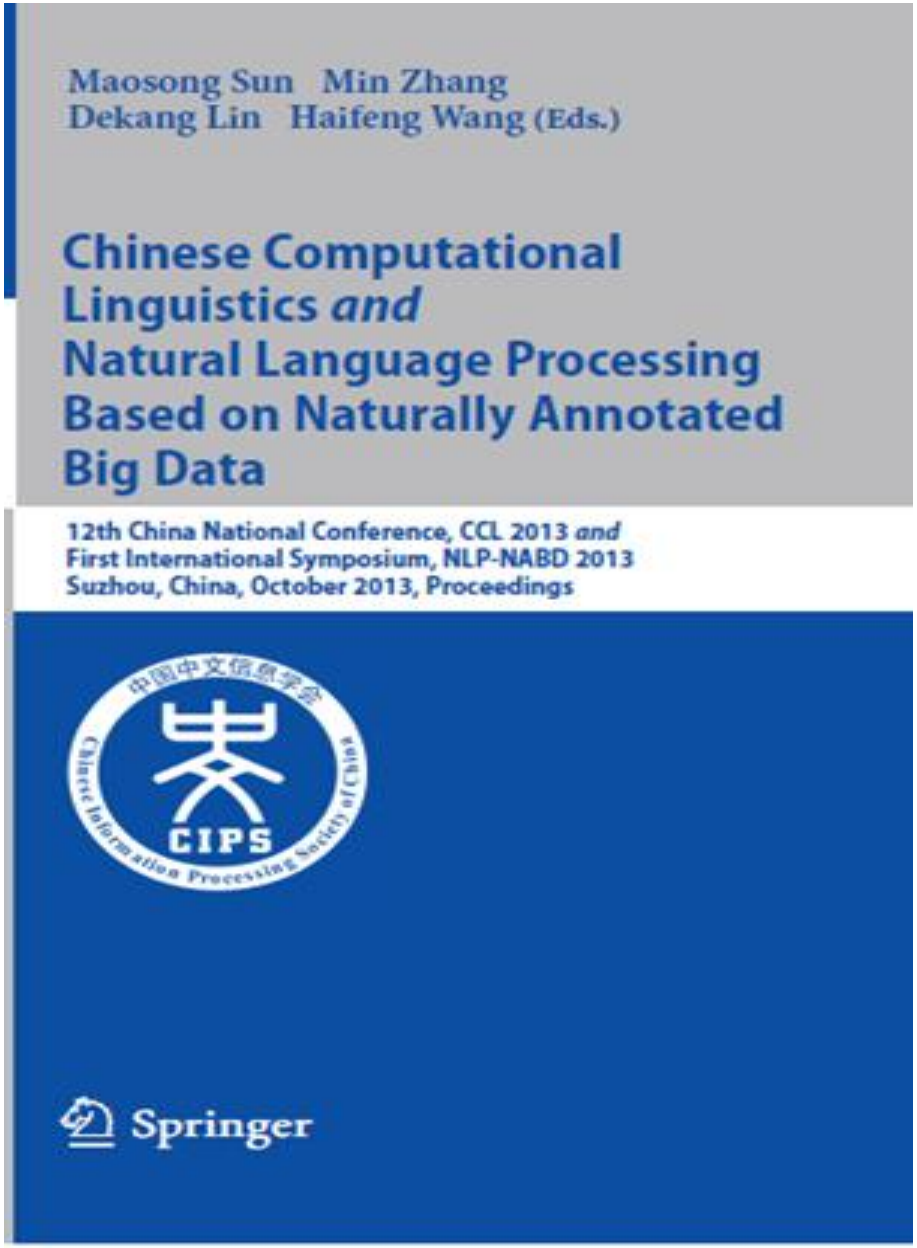
- 东方不亮西方亮

  (The country's territory is so vast that) when it is dark in the east, it is light in the west.

- **Global view parsing:**

  Easy sentences first, then difficult one.

Philosophy of NLP in the big data thinking



Thanks !