內容串流中實體特性偵測之研究
Detection of Entity Properties in Content Stream

Qing-Cheng Li

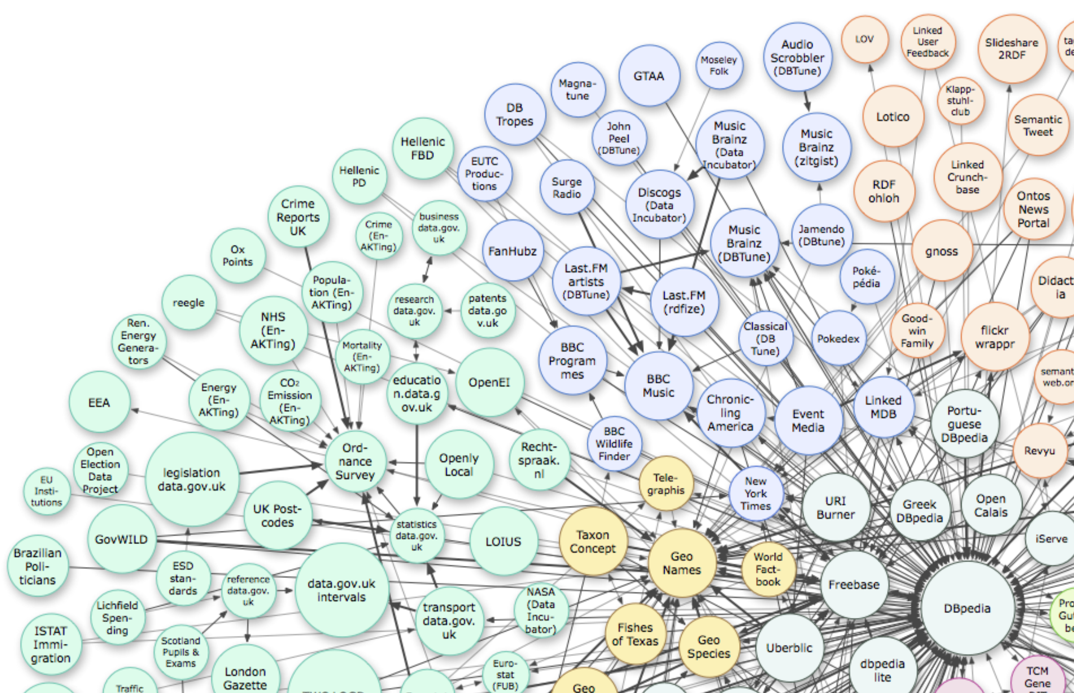
Advisor: Hsin-Hsi Chen, Ph.D.

Outline

1. Introduction
2. Related Works
3. Methods
4. Experiments
5. Conclusion

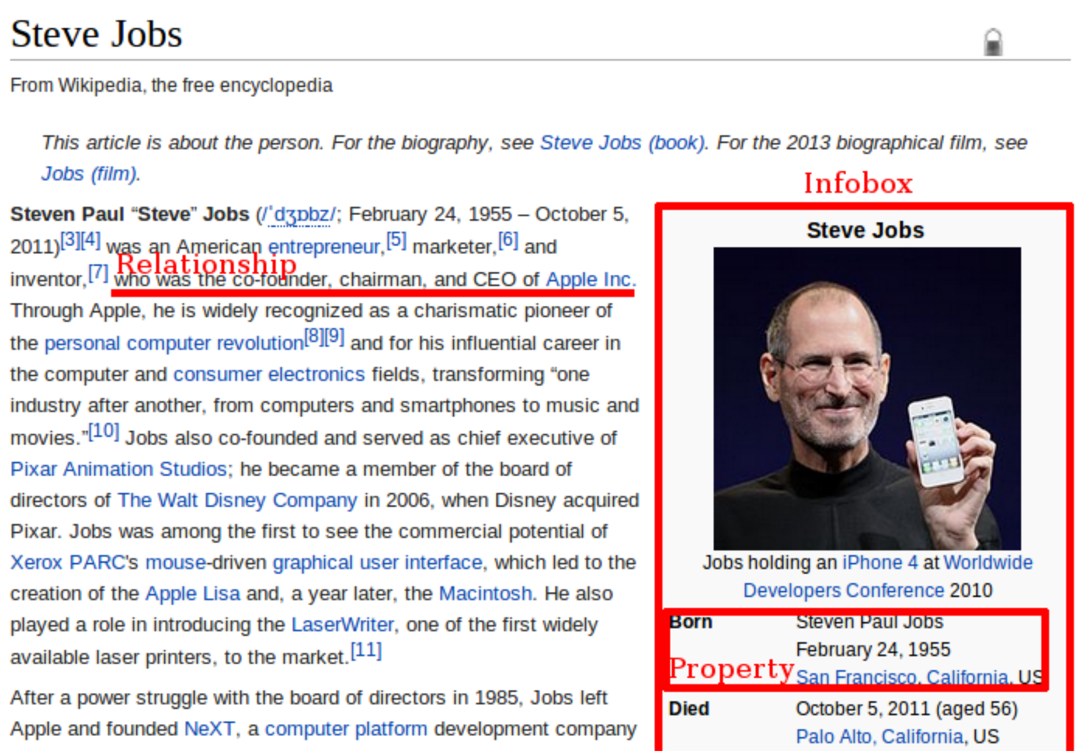
Introduction

- Many resources store human knowledge on the web.
- Those resources are directly or indirectly edited by human editors.



Knowledge Base

- Knowledge Base (KB) is a DB of content about entities and their properties and relationships.
- e.g. Wikipedia



Steve Jobs

From Wikipedia, the free encyclopedia

This article is about the person. For the biography, see Steve Jobs (book). For the 2013 biographical film, see Jobs (film).

Steven Paul "Steve" Jobs (/ˈdʒɒbz/; February 24, 1955 – October 5, 2011)[3][4] was an American entrepreneur,[5] marketer,[6] and inventor,[7] who was the co-founder, chairman, and CEO of Apple Inc. Through Apple, he is widely recognized as a charismatic pioneer of the personal computer revolution[8][9] and for his influential career in the computer and consumer electronics fields, transforming "one industry after another, from computers and smartphones to music and movies."[10] Jobs also co-founded and served as chief executive of Pixar Animation Studios; he became a member of the board of directors of The Walt Disney Company in 2006, when Disney acquired Pixar. Jobs was among the first to see the commercial potential of Xerox PARC's mouse-driven graphical user interface, which led to the creation of the Apple Lisa and, a year later, the Macintosh. He also played a role in introducing the LaserWriter, one of the first widely available laser printers, to the market.[11]

After a power struggle with the board of directors in 1985, Jobs left Apple and founded NeXT, a computer platform development company

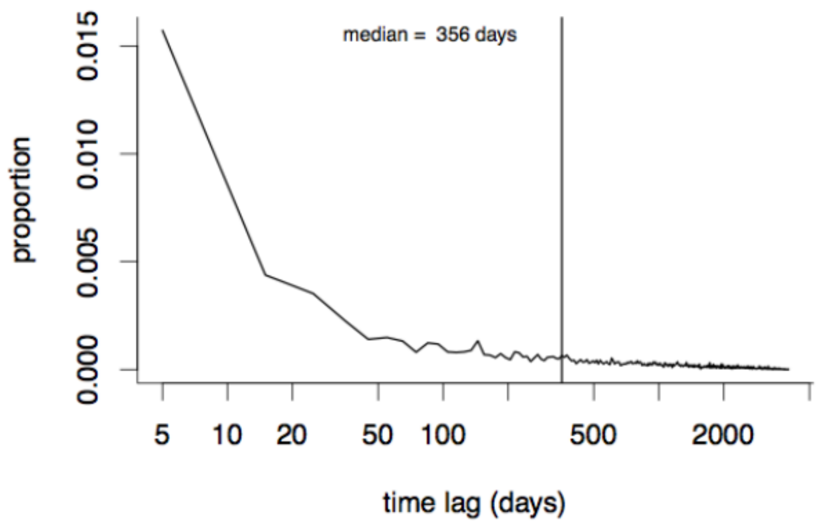
Steve Jobs

Jobs holding an iPhone 4 at Worldwide Developers Conference 2010

Born	Steven Paul Jobs February 24, 1955 San Francisco, California, US
Died	October 5, 2011 (aged 56) Palo Alto, California, US

Knowledge Base Acceleration

- World knowledge varies with time.
- Waits for human editors update the information in KB.
- How to acquire knowledge to reflect the changes of real world is very important.
- TREC KBA
    - Filtering content stream
    - Find target entity
    - Recommend edits to KB



Enitiy Property

- Documents in content stream contain entities and their properties.
- e.g. “Jobs was bron in San Francisco”
    - Entities
        - Jobs
        - San Francisco
    - Entity property
        - Jobs’ bron place (was bron in)

Goal

- Use patterns to detect entity properties in content stream efficiently and effectively.
- Efficiently
  - Content stream is big (~100,000 docs/hour)
- Effectively
  - Pattern’s quality
  - Pattern’s coverage
  - Pattern’s reliability
  - Pattern’s ambiguity

Related Works

- Structural KB
  - DBpedia
    - Extract information from wiki's infobox
  - YAGO
    - Extract from wiki and wordnet
  - Freebase
    - Scalable tuple
    - Contributed by human editors

Related Works

- KBA
    -

	2012	2013
Corups	7 months (4,973 hrs) > 400M docs	17 months (11,948 hrs) > 1B doc
Queries	27 people 2 organizations	98 people 19 organizations 24 facilities

    - Kjersten et al. (2012)
        - SVM, Topic classification
    - Wang et al. (2013)
        - Query expansion, Random Forest, Learning to rank.

Related Works

- Compare our work with KBA
    - Both want to accelerate the update of KB
    - KBA focus on entity
    - Our work focus on properties

Related Works

- Application of KB
  - QA System
    - Adolphs et al. (2011), Yao et al. (2014), Berant et al. (2013)
  - Named Entity Disambiguation
    - DBpedia Spotlight (Mendes et al, 2011)
- Relation between Pattern and Property
  - Extract relation
    - Fader et al. (2011), Moro et al. (2012)
  - PATTY - a texonomy of relational patterns with semantic types
    - Nakashole et al. (2012)

PATTY’s patterns

- <*Domain*> Pattern <*Range*>
- Pattern notation
  - [POS] - POS tag
  - word - word
  - * - wildcard

- e.g. “<*person*>’s [adj] voice * <*song*>”
  - “ Amy Winehouse’s soft voice in `Rehab’ ”

PATTY's Relation Parapharsing

- Given a relation from a KB, identify patterns that can be used to express that relation.

PATTY Relation Mining

MPI-INF|Databases

Thesaurus | Relations | Taxonomy

▶ DBPedia Relations

▼ YAGO Relations

actedIn
created
dealsWith
diedIn
directed
graduatedFror
happenedIn
hasAcademic/
hasCapital
hasChild
hasWonPrize

Relation: yago:wasBornIn

1-60 of 214

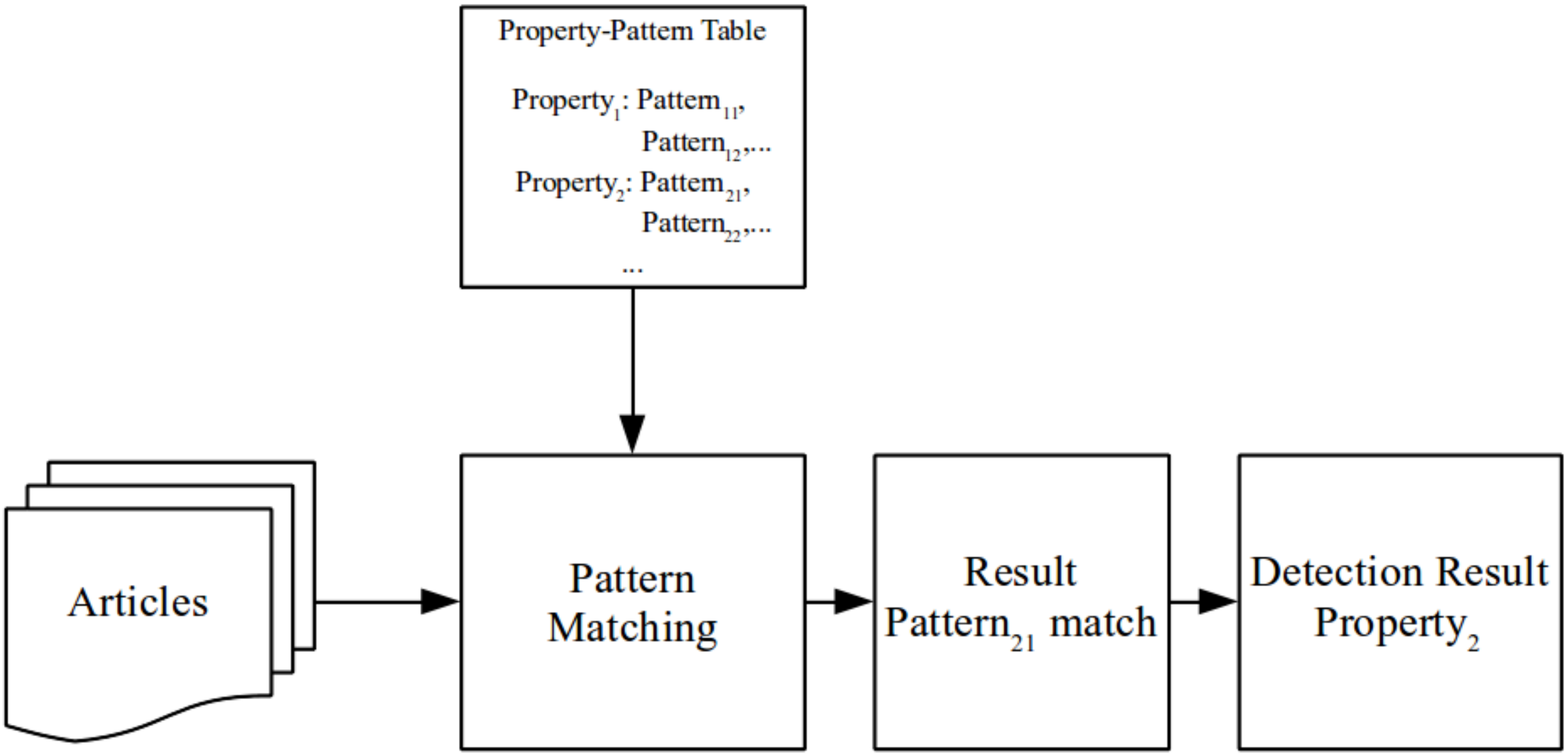
Pattern	Domain	Range	Confidence
died while;	person	city	0.963
[[con]] grew up;	person	city	0.795
buried in;	person	city	0.9
was born [[con]] was raised in;	person	city	0.927
[[con]] graduated in;	person	city	0.948
had worked in;	person	city	0.78

Methods

- Detect entity property by patterns
- “Jobs was born in San Francisco”
  - Property: *wasBornIn*
  - Pattern: was born in
- Property-Pattern Table

Property	Patterns
wasBornIn	was born in [[adj]] childhood in lived in
LivesIn	lived in been working in

Methods

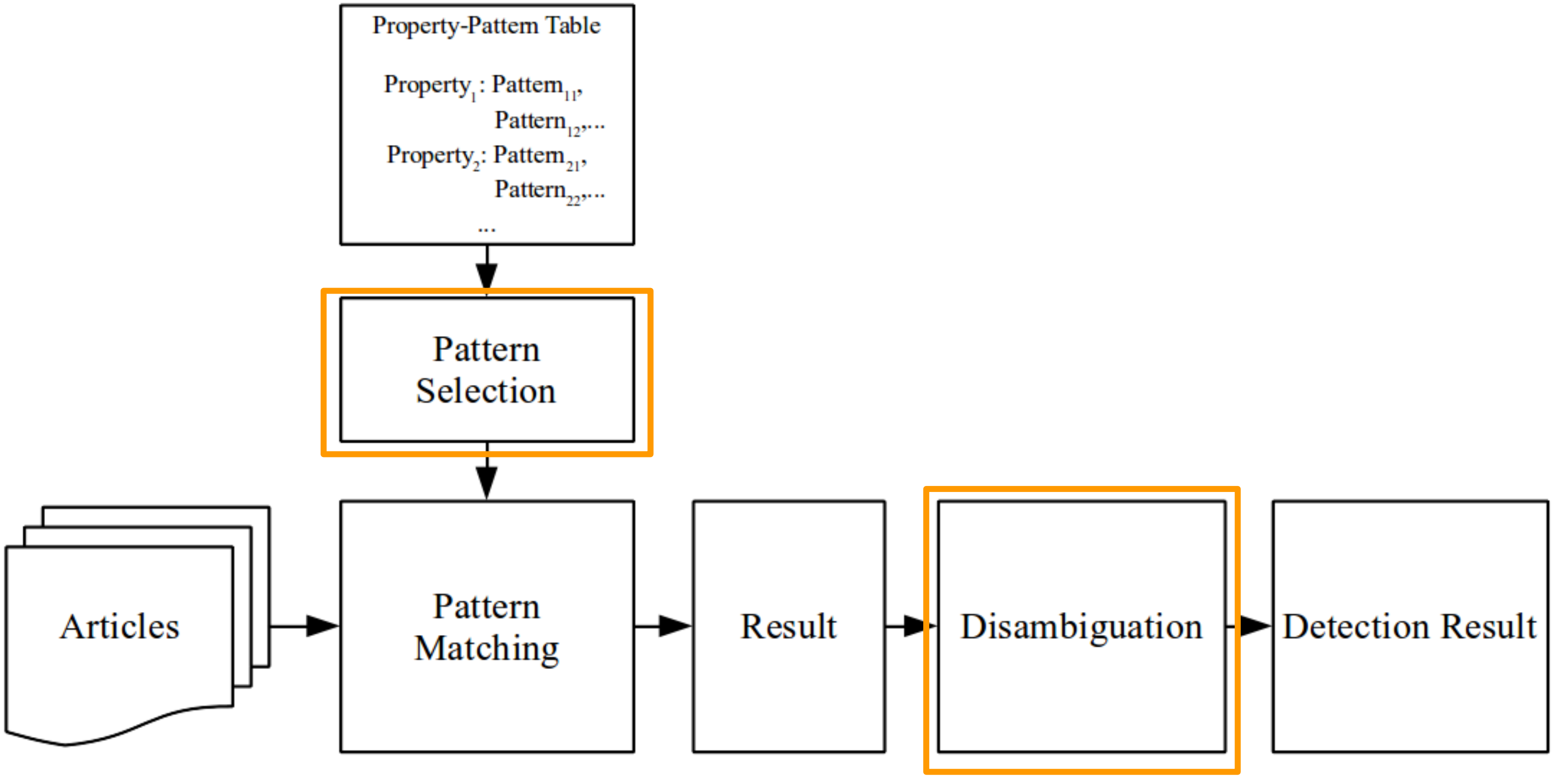


“Jobs **was born in** San Francisco” **was born in** **wasBornIn**

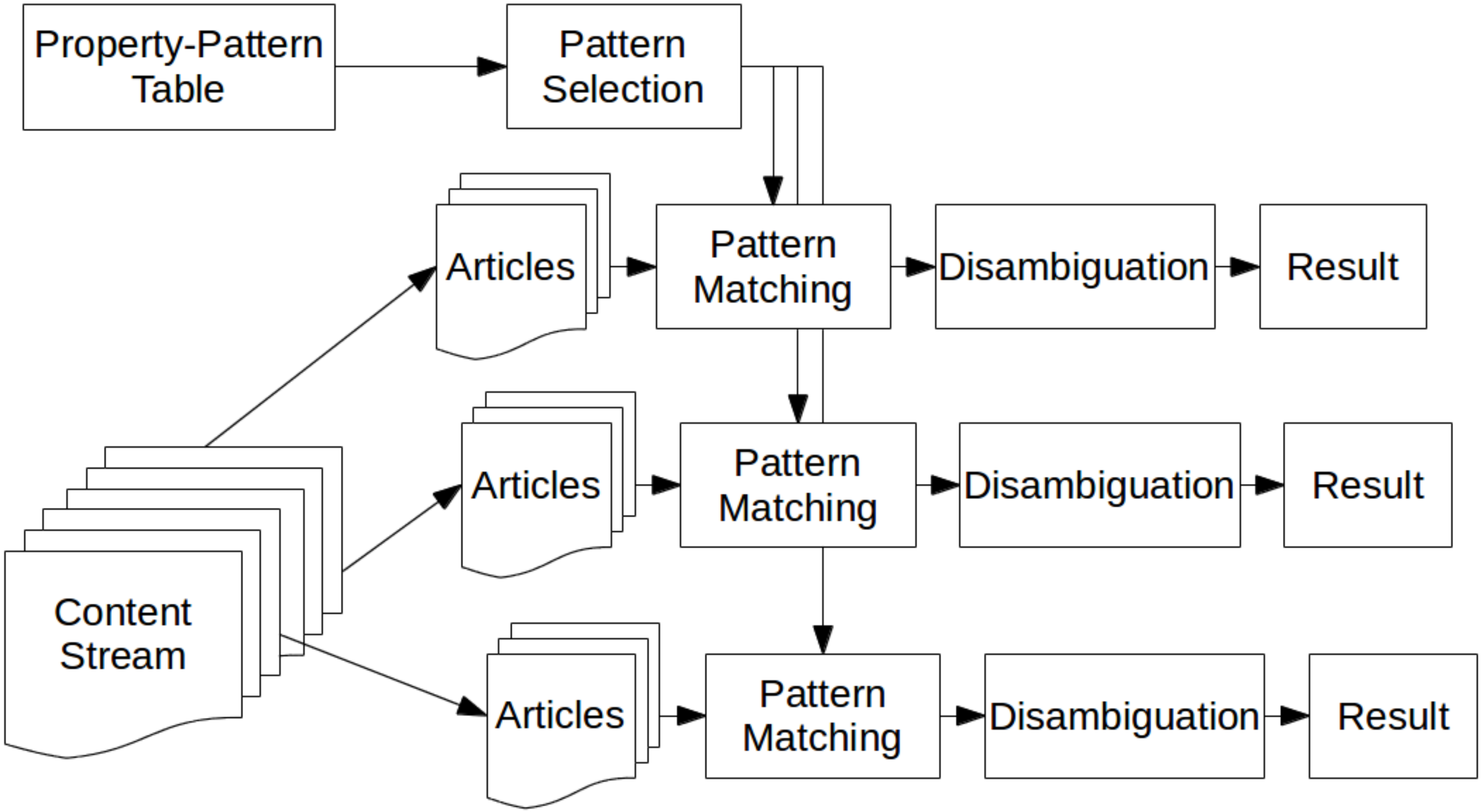
Issues

- Coverage
  - Property-Pattern table may not cover all patterns.
- Quality
  - A pattern is too general.
- Reliability
  - A pattern cannot express the property in table.
- Ambiguity
  - A pattern can express more than one property.

Detection Process

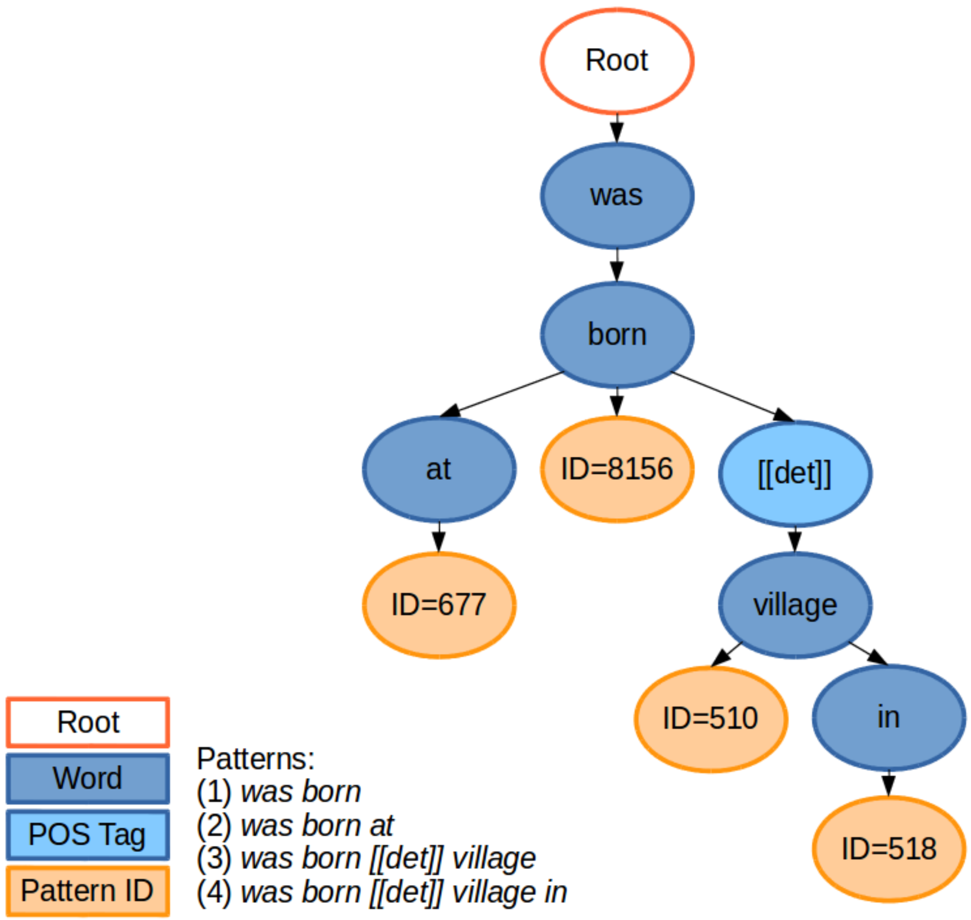


Detection System



Pattern Matching

- Pattern Prefix Tree



Pattern Matching

演算法 1 樣式比對演算法

輸入： A: 文章; T: 樣式前綴樹;
輸出： P: 文章 A 中出現的樣式;

```
P ←[]                                                          ▷ 初始化 P
for all Sentence S in A do
    對 S 進行詞性標記
    tmpP ←[]                                                   ▷ 初始化暫存陣列
    for i from 0 to length of S do
        (word,POStag) ← S[i]
        for all possiblePattern in tmpP do
            if word or POStag in T's possiblePattern.depthInTree+1 level nodes then
                continue
            else
                if T's possiblePattern.depthInTree+1 level node is PatternID then
                    add (PatternID,startPoint,i as endPoint) into P
                end if
                remove possiblePattern from tmpP
            end if
        end for
        if word or POStag in T's first level nodes then
            add (depthInTree,i as startPoint) into tmpP
        end if
    end for
end for
return P
```

Pattern Selection

- Select pattern by PATTY's confidence value
  - Confidence value
    - support set
    - {support set size} / {untyped variant}
  - (1) YAGO:actedIn - "links" : 0.143
  - (2) YAGO:actedIn - "starred in [[det]] film": 0.937
  - Select patterns which have high confidence values.

Pattern Selection

- Select pattern by its reliability
- Reliability
  - # there is a property / # appear in docs
  - e.g.
    - “was born” appears in 22,543 docs, and 12,038 docs have property “wasBornIn”
    - The reliability of “was born” for “wasBornIn” is 0.534

Pattern Selection

- Select pattern by its ambiguity degree
- Ambiguity degree
  - How many properties a pattern can express
  - e.g.
    - “fisrt met with” ‘s ambiguity degree is 5
    - It can express “hasAcademicAdvisor”, “isKnownFor”, “isMarriedTo”, “influences”, “hasChild”

Confidence, Ambiguity and Patterns

歧義度	信心值 > 0 樣式數量	信心值 > 0.7 樣式數量	信心值 > 0.8 樣式數量	信心值 > 0.9 樣式數量
1	11381	8913	5951	1879
2	4778	4175	2328	316
3	1255	1048	683	187
4	666	600	473	67
5	635	605	278	18
6	77	68	50	10
7	132	131	52	12
8	49	49	13	2
9	26	26	12	8
10	13	13	0	0
11	12	12	9	5
12	2	2	0	0
13	0	0	0	0
14	2	2	0	0
15	0	0	0	0
16	0	0	0	0
17	3	3	0	0
總計	19031	15647	9849	2504

Property Disambiguation

- Use entity type information
- e.g.
  - “appeared in” in wiki article “Jahsha Bluntt”
  - “appeared in”
    - pattern that express “actedIn”
    - Domain is <actor>
  - “Jahsha Bluntt”
    - basketball player
    - doesn’t have property “actedIn”

Pattern Overlap between Properties

Property	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) actedIn	1.00	0.33		0.01	0.31	0.01		0.02		0.09		
(2) created	0.21	1.00		0.01	0.17	0.11		0.01		0.10	0.02	0.03
(3) dealsWith			1.00						0.01	0.01		
(4) diedIn	0.01	0.03		1.00	0.01	0.04		0.01		0.03		0.01
(5) directed	0.51	0.44		0.01	1.00	0.07		0.03		0.06		
(6) graduatedFrom	0.01	0.16		0.02	0.04	1.00		0.01		0.02		0.04
(7) happenedIn			0.02				1.00					
(8) hasAcademicAdvisor	0.06	0.07		0.02	0.06	0.05		1.00		0.18		
(9) hasCapital			0.21	0.08					1.00			
(10) hasChild	0.05	0.09		0.01	0.02	0.01		0.03		1.00		0.13
(11) hasWonPrize	0.08	0.69			0.01			0.04		0.21	1.00	0.21
(12) holdsPoliticalPosition	0.01	0.08		0.01		0.07				0.40	0.01	1.00
(13) influences	0.07	0.13		0.01	0.02	0.03		0.13		0.35		0.13
(14) isCitizenOf	0.02	0.07		0.27	0.02	0.10		0.01		0.03		0.03
(15) isKnownFor	0.09	0.48		0.04	0.15	0.24		0.10		0.18	0.03	0.05
(16) isLeaderOf	0.01	0.58		0.18	0.01	0.41		0.01		0.04	0.11	0.11
(17) isLocatedIn		0.01	0.02	0.23		0.03		0.01		0.02		
(18) isMarriedTo	0.06	0.10		0.01	0.02	0.02		0.09		0.43		0.06
(19) isPoliticianOf	0.01	0.07		0.50	0.02	0.11		0.01		0.04		0.02
(20) livesIn	0.01	0.10		0.54	0.02	0.09		0.02		0.03		0.02
(21) participatedIn	0.02	0.01		0.02	0.01					0.07		0.01
(22) playsFor		0.04		0.02		0.13				0.02		
(23) wasBornIn	0.01	0.03		0.55	0.01	0.06		0.02		0.02		
(24) worksAt	0.02	0.35		0.04	0.06	0.63		0.02		0.03	0.03	0.06

(i,j) = | Property i's patterns ∩ Property j's patterns | / | Property i's patterns |

Pattern Overlap between Properties

Property	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
(1) actedIn	0.08	0.01	0.07			0.14					0.01	0.02
(2) created	0.10	0.02	0.23	0.08		0.14	0.01	0.02		0.02	0.01	0.17
(3) dealsWith	0.01				0.07	0.01					0.01	
(4) diedIn	0.02	0.16	0.05	0.06	0.22	0.04	0.17	0.29		0.02	0.50	0.05
(5) directed	0.05	0.01	0.19			0.08	0.01	0.01		0.01	0.01	0.07
(6) graduatedFrom	0.03	0.04	0.18	0.09	0.02	0.05	0.02	0.03		0.09	0.03	0.47
(7) happenedIn												
(8) hasAcademicAdvisor	0.51	0.01	0.25		0.02	0.59	0.01	0.02			0.04	0.05
(9) hasCapital												
(10) hasChild	0.24	0.01	0.08	0.01	0.01	0.51	0.01	0.01		0.01	0.01	0.01
(11) hasWonPrize		0.04	0.68	0.64				0.04			0.03	0.64
(12) holdsPoliticalPosition	0.27	0.02	0.07	0.04		0.23	0.01	0.01				0.09
(13) influences	1.00	0.01	0.09		0.01	0.51		0.01			0.01	0.03
(14) isCitizenOf	0.04	1.00	0.11	0.22	0.20	0.07	0.36	0.39		0.06	0.33	0.13
(15) isKnownFor	0.14	0.06	1.00	0.18	0.03	0.23	0.04	0.06		0.05	0.05	0.35
(16) isLeaderOf	0.02	0.38	0.62	1.00	0.22	0.11	0.20	0.26		0.04	0.26	0.64
(17) isLocatedIn	0.01	0.12	0.04	0.08	1.00	0.03	0.11	0.17		0.01	0.28	0.03
(18) isMarriedTo	0.30	0.01	0.09	0.01	0.01	1.00	0.01	0.01		0.01	0.02	0.03
(19) isPoliticianOf	0.02	0.63	0.14	0.20	0.31	0.09	1.00	0.61		0.06	0.44	0.15
(20) livesIn	0.04	0.44	0.13	0.17	0.30	0.06	0.40	1.00		0.03	0.56	0.10
(21) participatedIn						0.07			1.00		0.02	0.02
(22) playsFor		0.03	0.05	0.01	0.01	0.01	0.02	0.02		1.00	0.02	0.11
(23) wasBornIn	0.03	0.22	0.07	0.10	0.30	0.06	0.17	0.33		0.03	1.00	0.07
(24) worksAt	0.04	0.06	0.34	0.19	0.03	0.07	0.04	0.05		0.11	0.05	1.00

(i,j) = | Property i's patterns ∩ Property j's patterns | / | Property i's patterns |

Property Disambiguation

- Property selection strategy
  - Select one
    - Sort the candidate propertiy by its reliability
    - e.g.
      - “have lived in”: “isLocatedIn”: 52/95, “wasBornIn”: 21/95, “isCitizenOf”: 8/95, “livesIn”: 6/95, “isLeaderOf”: 2/95
      - Select first property “isLocatedIn”
  - Select by a threshold
    - Normalize reliability
    - e.g.
      - “have lived in”: “isLocatedIn”: 1, “wasBornIn”: 0.404, “isCitizenOf”: 0.154, “livesIn”: 0.115, “isLeaderOf”: 0.038
      - If threshold = 0.2, select “isLocatedIn” and “wasBornIn”
  - Select all

Property Disambiguation

- Sometime pattern can express property is related to the context.
  - dealsWith → contry, location...
  - holdsPoliticalPosition → Primier, Mayor, Minister...
- Naive Bayes Classifier
  - For each property, train a binary NBC
  - Feature: words in sentences
  - TrainMultinomialNB($\mathbb{C}$, $\mathbb{D}$)

```
TrainMultinomialNB(C, D)
V ← ExtractVocabulary(D)
N ← CountDocs(D)
for each c ∈ C
do N_c ← CountDocsInClass(D, c)
    prior[c] ← N_c / N
    text_c ← ConcatenateTextOfAllDocsInClass(D, c)
    for each t ∈ V
    do T_ct ← CountTokensOfTerm(text_c, t)
    for each t ∈ V
    do condprob[t][c] ← (T_ct + 1) / Σ_t'(T_ct' + 1)
return V, prior, condprob
```

```
ApplyMultinomialNB(C, V, prior, condprob, d)
W ← ExtractTokensFromDoc(V, d)
for each c ∈ C
do score[c] ← log prior[c]
    for each t ∈ W
    do score[c] += log condprob[t][c]
return arg max_{c∈C} score[c]
```

Experiment

- Wikipedia (2013/03) article as entity
- Properties in YAGO Facts as answer for each wiki article
- PATTY’s patterns

- 334,469 wiki articles
- 24 YAGO relations as properties
- 19,031 patterns

Dataset

- Docs number of each property

Property	Number of documents
actedIn	3971
created	8665
dealsWith	111
diedIn	11729
directed	2036
graduatedFrom	10474
happenedIn	1792
hasAcademicAdvisor	628
hasCapital	363
hasChild	2113
hasWonPrize	8916
holdsPoliticalPosition	1190
influences	869
isCitizenOf	10974
isKnownFor	73
isLeaderOf	1397
isLocatedIn	216772
isMarriedTo	3555
isPoliticianOf	170
livesIn	7661
participatedIn	346
playsFor	42751
wasBornIn	47141
worksAt	1981

Pattern Coverage

歧義度	信心值 > 0				信心值 > 0.7			
	樣式數	出現	涵蓋文章	比例	樣式數	出現	涵蓋文章	比例
1	11381	7267	250737	74.97	8913	5854	244888	73.22
2	4778	2976	275736	82.44	4175	2697	272395	81.44
3	1255	886	278820	83.36	1048	745	274265	82.00
4	666	503	281256	84.09	600	458	276782	82.75
5	635	465	283830	84.86	605	442	279603	83.60
6	77	64	283858	84.87	68	57	279648	83.61
7	132	100	284128	84.95	131	100	280096	83.74
8	49	43	284275	84.99	49	43	280342	83.82
9	26	25	284295	85.00	26	25	280385	83.83
10	13	11	284297	85.00	13	11	280387	83.83
11	12	9	284299	85.00	12	9	280391	83.83
12	2	2	284299	85.00	2	2	280391	83.83
13	0	0	284299	85.00	0	0	280391	83.83
14	2	2	284299	85.00	2	2	280391	83.83
15	0	0	284299	85.00	0	0	280391	83.83
16	0	0	284299	85.00	0	0	280391	83.83
17	3	3	284299	85.00	3	3	280391	83.83
總計	19031	12356	–	–	15647	10448	–	–

歧義度	信心值 > 0.8				信心值 > 0.9			
	樣式數	出現	涵蓋文章	比例	樣式數	出現	涵蓋文章	比例
1	5951	4029	222371	66.48	1879	1121	155402	46.46
2	2328	1697	251552	75.21	316	206	163265	48.81
3	683	487	258715	77.35	187	110	170820	51.07
4	473	355	261633	78.22	67	47	171978	51.42
5	278	243	262321	78.43	18	16	173397	51.84
6	50	43	262364	78.44	10	8	173457	51.86
7	52	41	262452	78.47	12	8	174396	52.14
8	13	15	263419	78.76	2	2	174476	52.17
9	12	12	263514	78.79	8	8	175287	52.41
總計	9840	6922	–	–	2499	1526	–	–

Evaluation

- Efficiency
  - How many docs/min per core
- Performance
  - Precision
  - Recall
  - F1
  - Macro Average
  - Micro Average

- Evaluate on document level because we only have document level's answer.

Efficiency

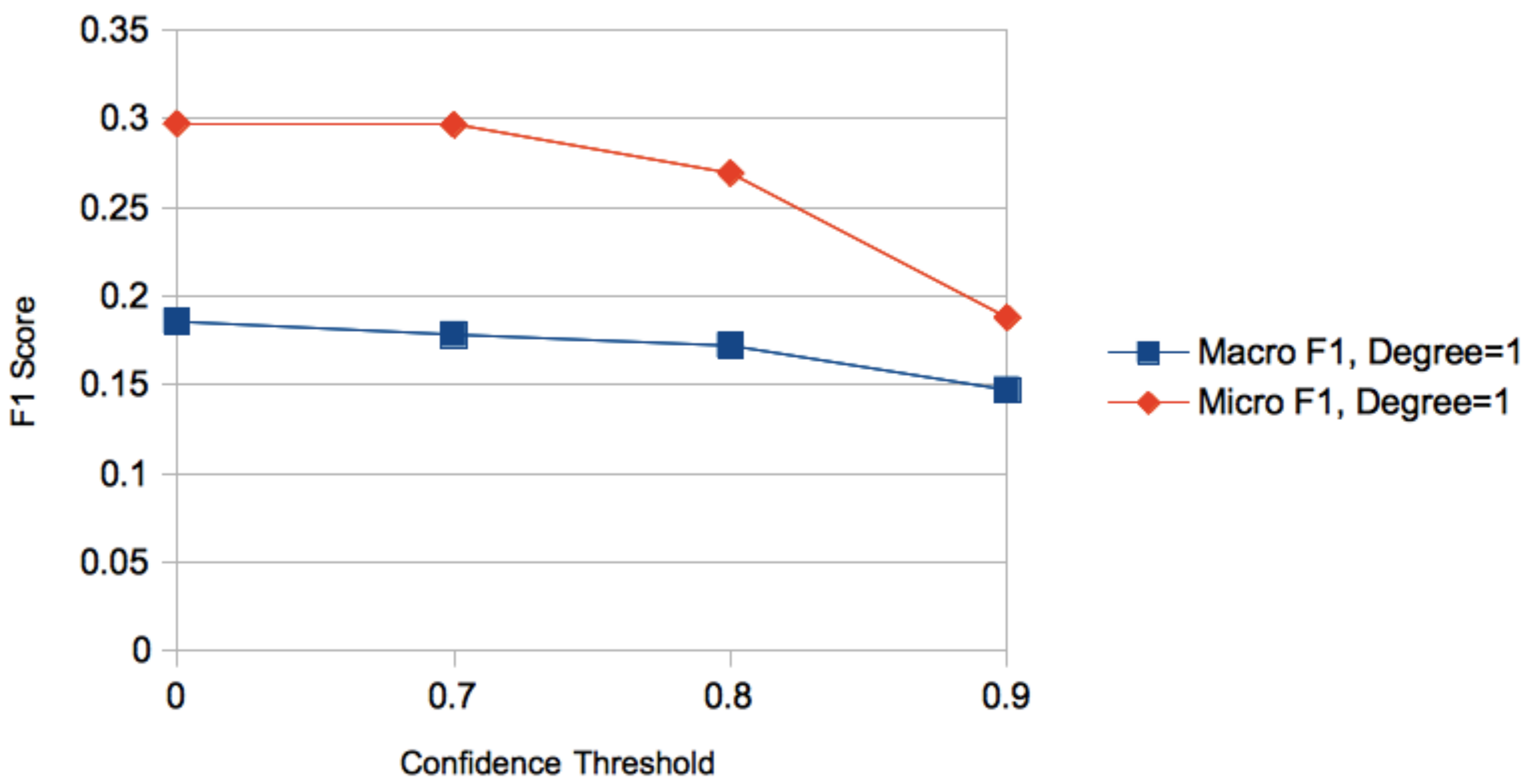
- Pattern Matching
  - 529 docs/min per core
- Pattern Disambiguation
  - 2750 docs/min per core

- Overall
  - 3 core for 100,000 docs in a hour
  - 334,469 docs in 30 mins on a 24 core workstation

Baseline

Property	無歧義			歧義度 ≤ 5		
	Precision	Recall	F_1 Score	Precision	Recall	F_1 Score
actedIn	0.0435	0.5163	0.0802	0.0281	0.7976	0.0543
created	0.0823	0.8171	0.1495	0.0558	0.9500	0.1055
dealsWith	0.0034	1.0000	0.0067	0.0022	1.0000	0.0043
diedIn	0.0622	0.3688	0.1065	0.0551	0.6468	0.1015
directed	0.0393	0.2867	0.0691	0.0290	0.9403	0.0562
graduatedFrom	0.1472	0.4752	0.2248	0.1130	0.7857	0.1976
happenedIn	0.1323	0.6676	0.2209	0.1315	0.6731	0.2201
hasAcademicAdvisor	0.0281	0.4329	0.0528	0.0129	0.6892	0.0254
hasCapital	0.0530	0.3430	0.0917	0.0059	0.4582	0.0116
hasChild	0.0284	0.6302	0.0544	0.0182	0.9724	0.0358
hasWonPrize	0.1212	0.0110	0.0202	0.0969	0.0738	0.0838
holdsPoliticalPosition	0.0302	0.7598	0.0580	0.0151	0.8945	0.0298
influences	0.0153	0.8271	0.0301	0.0082	0.9559	0.0162
isCitizenOf	0.1199	0.1295	0.1245	0.0879	0.3901	0.1434
isKnownFor	0.0021	0.4464	0.0042	0.0008	0.9380	0.0016
isLeaderOf	0.0468	0.0233	0.0311	0.0193	0.4696	0.0371
isLocatedIn	0.8452	0.5208	0.6445	0.7709	0.5618	0.6499
isMarriedTo	0.0383	0.6623	0.0724	0.0274	0.9665	0.0534
isPoliticianOf	0.0163	0.2355	0.0305	0.0037	0.7025	0.0073
livesIn	0.0747	0.0553	0.0635	0.0673	0.4438	0.1169
participatedIn	0.0223	0.6549	0.0431	0.0202	0.7151	0.0393
playsFor	0.5239	0.5181	0.5210	0.3829	0.6096	0.4703
wasBornIn	0.2862	0.0825	0.1280	0.2523	0.3962	0.3083
worksAt	0.0439	0.3563	0.0781	0.0210	0.8650	0.0409
Macro Average	0.1169	0.4509	0.1857	0.0927	0.7040	0.1639
Micro Average	0.2257	0.4356	0.2973	0.1255	0.5603	0.2051

Confidence

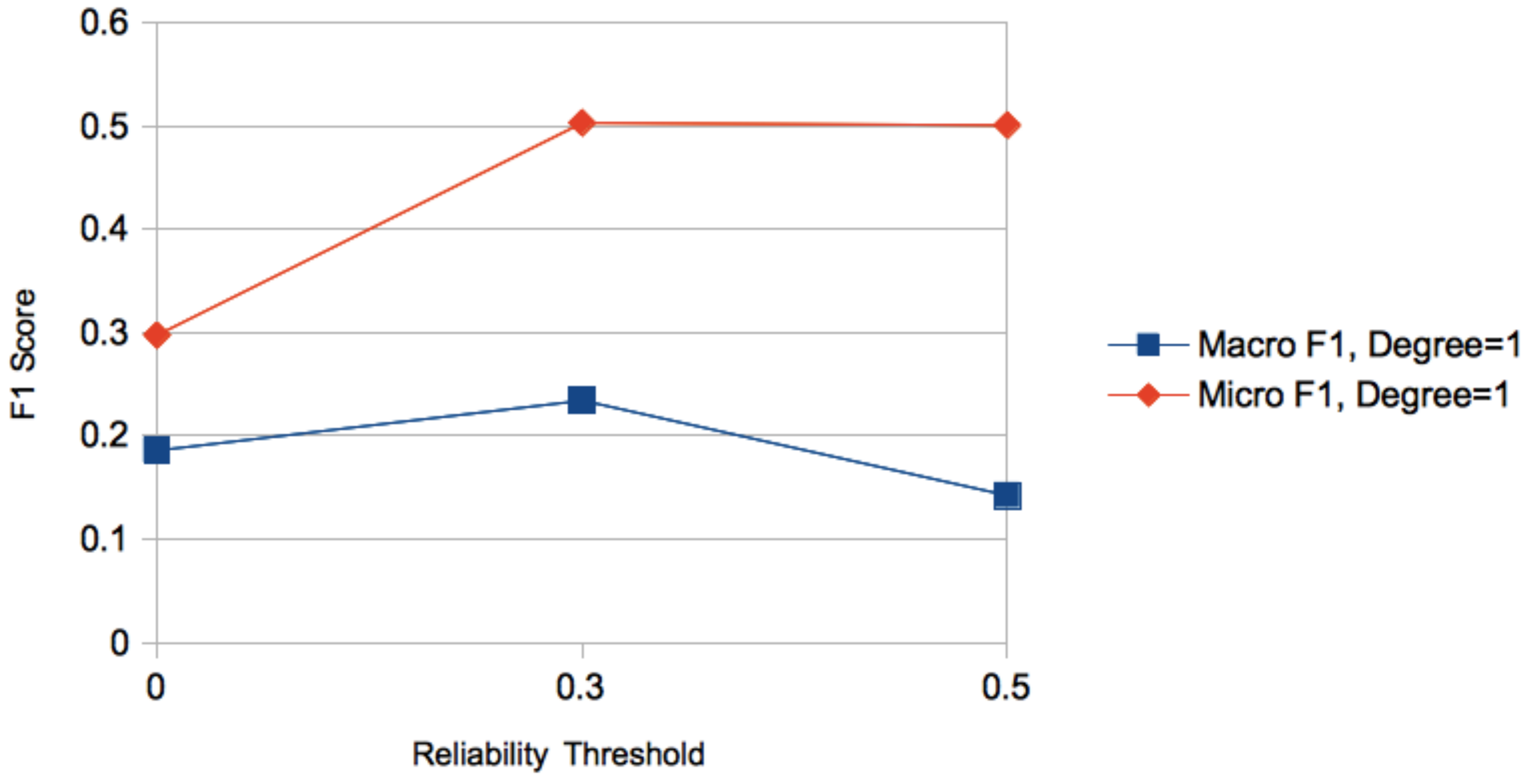


0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
F1 Score
0
0.7
0.8
0.9
Confidence Threshold
Macro F1, Degree=1
Micro F1, Degree=1

Confidence

Property	Confidence > 0.8			Confidence > 0.9		
	Precision	Recall	F_1 Score	Precision	Recall	F_1 Score
actedIn	0.0490	0.3196	0.0850	0.0833	0.1032	0.0922
created	0.1005	0.6682	0.1747	0.1564	0.2208	0.1831
dealsWith	0.0128	0.8935	0.0252	0.0114	0.5805	0.0224
diedIn	0.0603	0.3479	0.1028	0.1191	0.2179	0.1540
directed	0.0258	0.1403	0.0436	0.1693	0.0145	0.0267
graduatedFrom	0.1022	0.2326	0.1420	0.1129	0.0736	0.0891
happenedIn	0.0638	0.1374	0.0872	0.3010	0.0139	0.0266
hasAcademicAdvisor	0.0288	0.3254	0.0530	0.0468	0.2426	0.0785
hasCapital	0.1117	0.0113	0.0206	0.0000	0.0000	–
hasChild	0.0297	0.5289	0.0563	0.0337	0.1872	0.0572
hasWonPrize	0.1212	0.0110	0.0202	0.1193	0.0008	0.0015
holdsPoliticalPosition	0.0350	0.7283	0.0667	0.0437	0.4819	0.0801
influences	0.0195	0.7098	0.0379	0.0616	0.3715	0.1057
isCitizenOf	0.1131	0.0790	0.0930	0.1368	0.0293	0.0483
isKnownFor	0.0024	0.4068	0.0047	0.0000	0.0000	–
isLeaderOf	0.0419	0.0120	0.0187	0.0264	0.0056	0.0092
isLocatedIn	0.8806	0.4001	0.5502	0.9066	0.1853	0.3078
isMarriedTo	0.0405	0.5531	0.0755	0.0935	0.2915	0.1416
isPoliticianOf	0.0142	0.1527	0.0260	0.0118	0.0720	0.0202
livesIn	0.0709	0.0466	0.0563	0.0643	0.0080	0.0143
participatedIn	0.0220	0.5917	0.0423	0.0266	0.4552	0.0503
playsFor	0.4966	0.2762	0.3550	0.5432	0.0121	0.0236
wasBornIn	0.3194	0.0736	0.1196	0.3073	0.0313	0.0569
worksAt	0.0435	0.1779	0.0699	0.0404	0.0524	0.0456
Macro Average	0.1169	0.3260	0.1721	0.1423	0.1521	0.1471
Micro Average	0.2331	0.3190	0.2694	0.3213	0.1328	0.1880

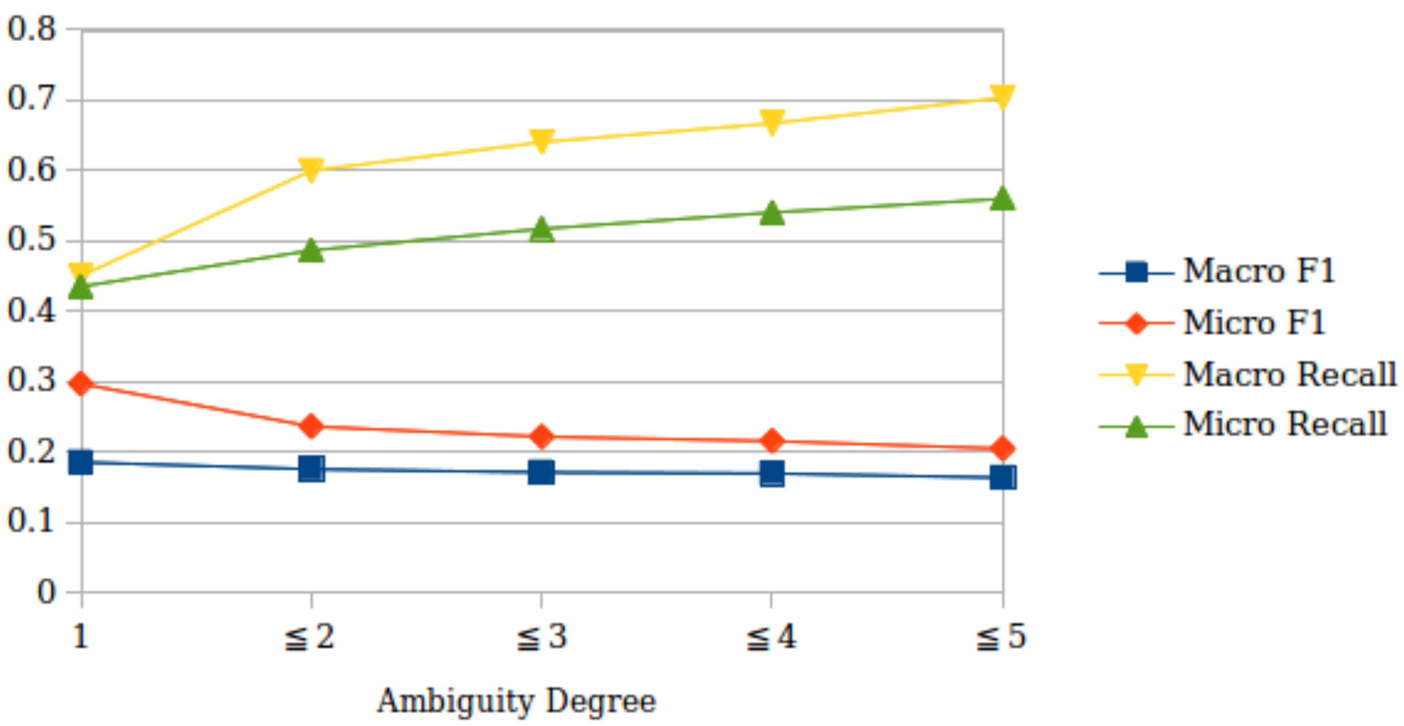
Reliability



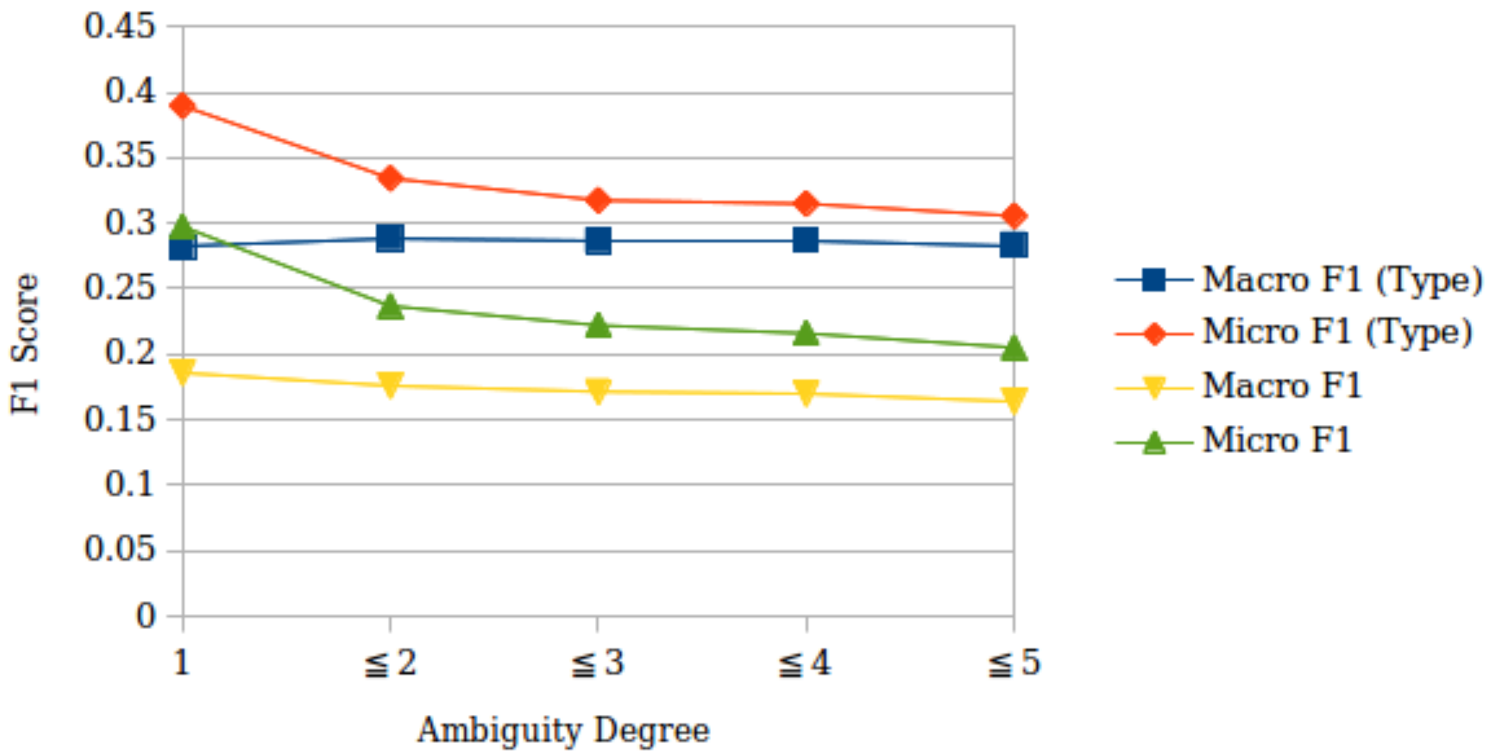
Reliability

Property	可信賴度 > 0			可信賴度 > 0.3			可信賴度 > 0.5		
	Precision	Recall	F_1 Score	Precision	Recall	F_1 Score	Precision	Recall	F_1 Score
actedIn	0.0435	0.5163	0.0802	0.4038	0.3380	0.3680	0.5109	0.1773	0.2632
created	0.0823	0.8171	0.1495	0.3122	0.5858	0.4073	0.4790	0.4485	0.4632
dealsWith	0.0034	1.0000	0.0067	0.2499	0.7309	0.3725	0.0250	0.0037	0.0065
diedIn	0.0622	0.3688	0.1065	0.3737	0.1130	0.1736	0.4004	0.0145	0.0280
directed	0.0393	0.2867	0.0691	0.5506	0.1305	0.2110	0.6140	0.1133	0.1913
graduatedFrom	0.1472	0.4752	0.2248	0.3869	0.2307	0.2891	0.4965	0.0521	0.0943
happenedIn	0.1323	0.6676	0.2209	0.5055	0.3098	0.3842	0.7744	0.1833	0.2964
hasAcademicAdvisor	0.0281	0.4329	0.0528	0.3272	0.0513	0.0887	0.0333	0.0014	0.0027
hasCapital	0.0530	0.3430	0.0917	0.0000	0.0000	–	0.0000	0.0000	–
hasChild	0.0284	0.6302	0.0544	0.2886	0.2432	0.2640	0.3607	0.0267	0.0497
hasWonPrize	0.1212	0.0110	0.0202	0.3697	0.0027	0.0054	0.5167	0.0008	0.0015
holdsPoliticalPosition	0.0302	0.7598	0.0580	0.0000	0.0000	–	0.0000	0.0000	–
influences	0.0153	0.8271	0.0301	0.2137	0.2900	0.2461	0.1952	0.0103	0.0195
isCitizenOf	0.1199	0.1295	0.1245	0.3640	0.0118	0.0229	0.5043	0.0039	0.0077
isKnownFor	0.0021	0.4464	0.0042	0.0000	0.0000	–	0.0000	0.0000	–
isLeaderOf	0.0468	0.0233	0.0311	0.0000	0.0000	–	0.0000	0.0000	–
isLocatedIn	0.8452	0.5208	0.6445	0.8527	0.5206	0.6465	0.8665	0.5188	0.6490
isMarriedTo	0.0383	0.6623	0.0724	0.3307	0.2113	0.2579	0.5033	0.0523	0.0947
isPoliticianOf	0.0163	0.2355	0.0305	0.1082	0.0191	0.0324	0.0000	0.0000	–
livesIn	0.0747	0.0553	0.0635	0.3983	0.0110	0.0214	0.4500	0.0013	0.0026
participatedIn	0.0223	0.6549	0.0431	0.1000	0.0114	0.0204	0.0000	0.0000	–
playsFor	0.5239	0.5181	0.5210	0.6013	0.5157	0.5552	0.6683	0.4971	0.5701
wasBornIn	0.2862	0.0825	0.1280	0.3951	0.0588	0.1024	0.4841	0.0083	0.0162
worksAt	0.0439	0.3563	0.0781	0.3516	0.1095	0.1670	0.4951	0.0271	0.0514
Macro Average	0.1169	0.4509	0.1857	0.3118	0.1873	0.2340	0.3491	0.0892	0.1421
Micro Average	0.2257	0.4356	0.2973	0.7046	0.3908	0.5027	0.8014	0.3638	0.5004

Ambiguity Degree



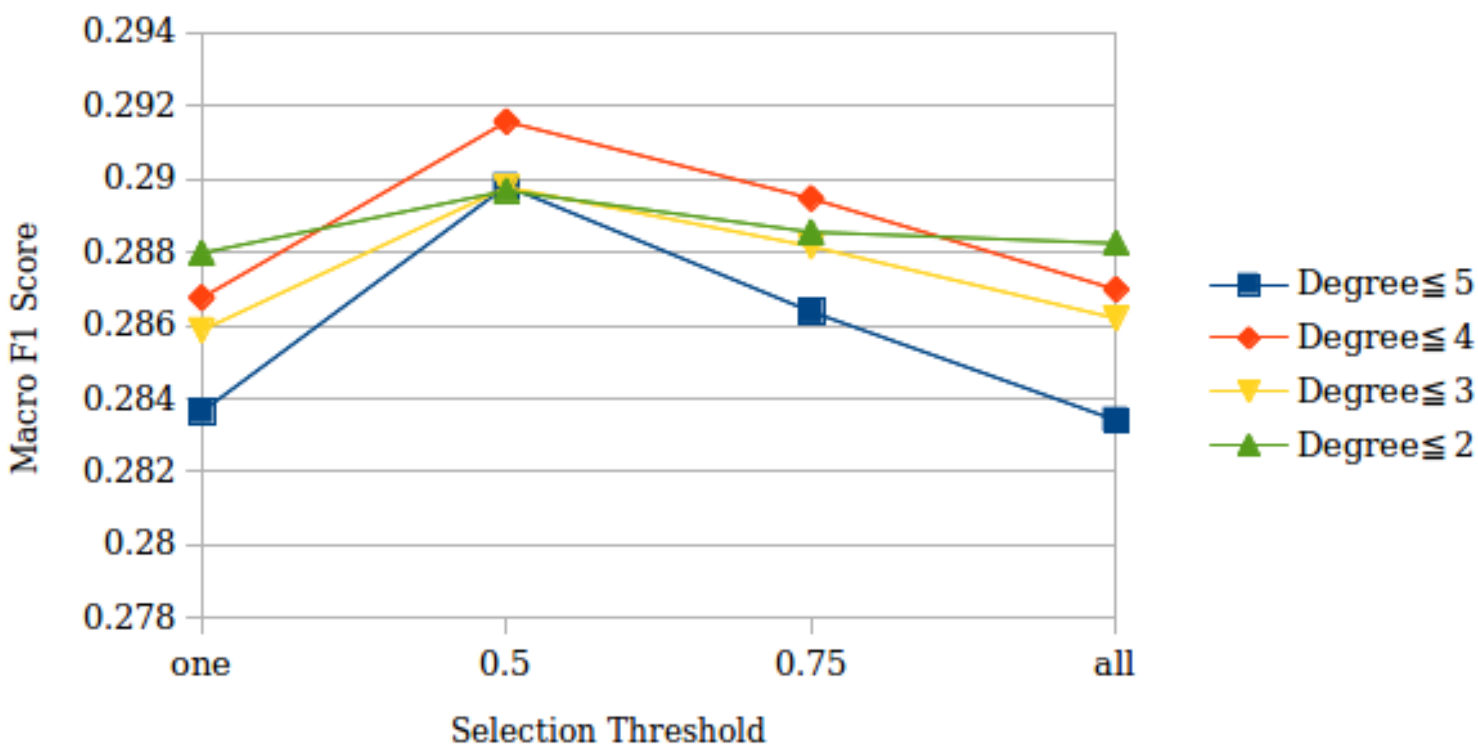
Entity’s Type Information



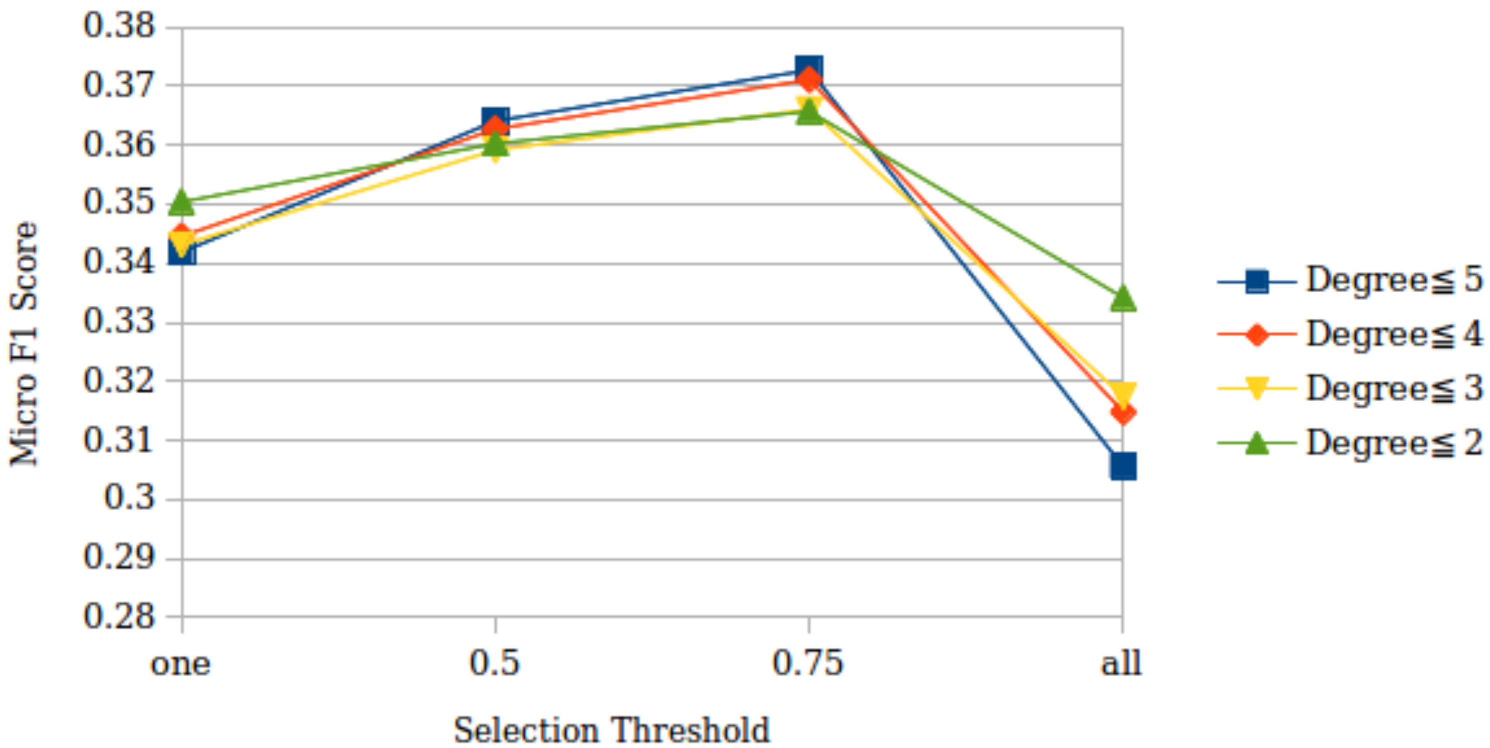
Entity's Type Information

Property	無實體類型資訊			加入實體類型資訊		
	Precision	Recall	F_1 Score	Precision	Recall	F_1 Score
actedIn	0.0373	0.6691	0.0707	0.5300	0.6691	0.5915
created	0.0729	0.8853	0.1348	0.1072	0.8853	0.1912
dealsWith	0.0022	1.0000	0.0043	0.0049	1.0000	0.0098
diedIn	0.0490	0.4720	0.0888	0.0995	0.4720	0.1643
directed	0.0482	0.9067	0.0915	0.0826	0.9067	0.1514
graduatedFrom	0.1336	0.6335	0.2206	0.1881	0.6335	0.2900
happenedIn	0.1315	0.6731	0.2201	0.7641	0.6731	0.7157
hasAcademicAdvisor	0.0158	0.5892	0.0308	0.0258	0.5892	0.0494
hasCapital	0.0059	0.4582	0.0116	0.0064	0.4582	0.0126
hasChild	0.0219	0.9453	0.0429	0.0357	0.9453	0.0689
hasWonPrize	0.1160	0.0111	0.0203	0.1782	0.0111	0.0209
holdsPoliticalPosition	0.0180	0.8733	0.0352	0.0284	0.8733	0.0549
influences	0.0108	0.9184	0.0214	0.0195	0.9184	0.0383
isCitizenOf	0.0871	0.2361	0.1273	0.1667	0.2361	0.1954
isKnownFor	0.0013	0.7799	0.0027	0.0016	0.7799	0.0033
isLeaderOf	0.0252	0.0900	0.0393	0.0386	0.0900	0.0541
isLocatedIn	0.8287	0.5401	0.6540	0.9816	0.5401	0.6968
isMarriedTo	0.0349	0.9561	0.0674	0.0558	0.9561	0.1055
isPoliticianOf	0.0052	0.4960	0.0102	0.0125	0.4960	0.0243
livesIn	0.0729	0.1141	0.0890	0.1196	0.1141	0.1168
participatedIn	0.0203	0.7012	0.0395	0.0233	0.7012	0.0451
playsFor	0.4526	0.5311	0.4887	0.5489	0.5311	0.5398
wasBornIn	0.2459	0.1916	0.2154	0.4806	0.1916	0.2740
worksAt	0.0355	0.7368	0.0678	0.0520	0.7368	0.0971
Macro Average	0.1030	0.6003	0.1759	0.1896	0.6003	0.2882
Micro Average	0.1564	0.4870	0.2367	0.2544	0.4870	0.3342

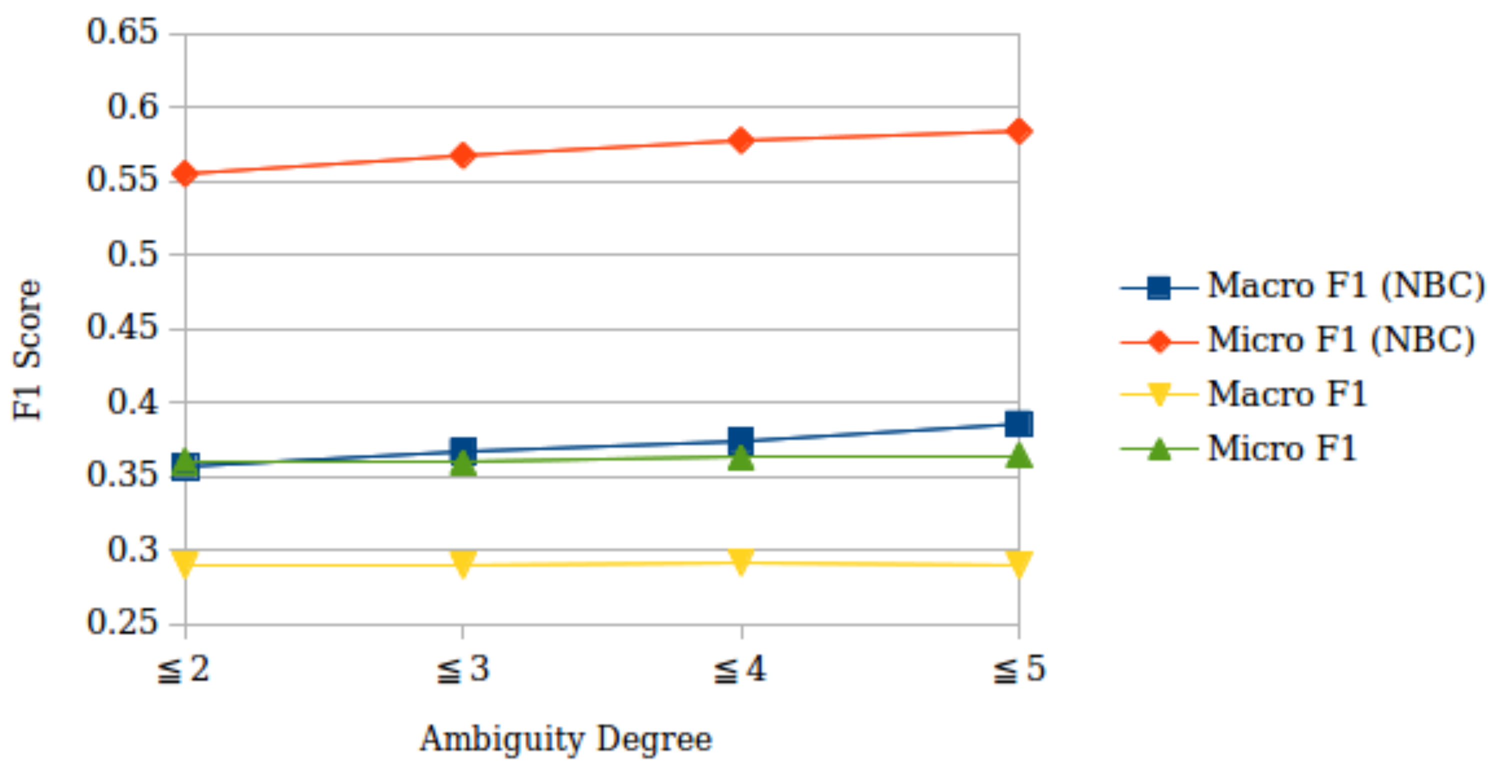
Property Selection Strategy



Property Selection Strategy



Naive Bayes Classifier



Naive Bayes Classifier

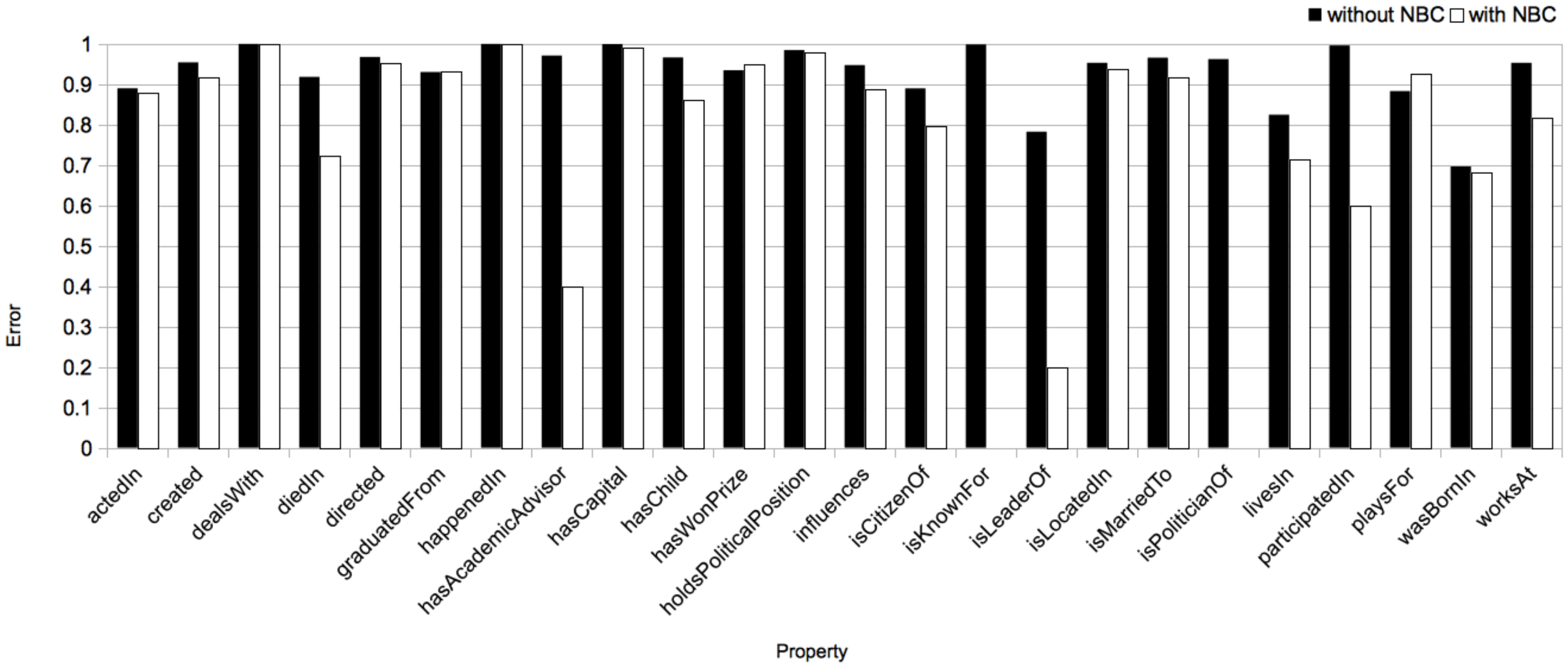
Property	不使用分類器			使用分類器		
	Precision	Recall	F_1 Score	Precision	Recall	F_1 Score
actedIn	0.5332	0.7420	0.6205	0.5985	0.5495	0.5730
created	0.0904	0.9429	0.1651	0.3036	0.7617	0.4342
dealsWith	0.0107	1.0000	0.0211	0.3724	0.6352	0.4696
diedIn	0.1145	0.5688	0.1906	0.3868	0.3807	0.3837
directed	0.0784	0.9186	0.1445	0.4229	0.7072	0.5293
graduatedFrom	0.1706	0.7783	0.2798	0.4446	0.6204	0.5180
happenedIn	0.7641	0.6731	0.7157	0.9916	0.5238	0.6855
hasAcademicAdvisor	0.0393	0.5086	0.0730	0.4667	0.0130	0.0253
hasCapital	0.0064	0.4582	0.0126	0.2336	0.3891	0.2919
hasChild	0.0351	0.9603	0.0678	0.3155	0.6858	0.4322
hasWonPrize	0.1165	0.0727	0.0895	0.7733	0.0065	0.0129
holdsPoliticalPosition	0.0348	0.8548	0.0669	0.4324	0.0271	0.0511
influences	0.0199	0.9224	0.0390	0.1878	0.6360	0.2900
isCitizenOf	0.1631	0.2843	0.2073	0.5582	0.0241	0.0461
isKnownFor	0.0027	0.4934	0.0054	0.0000	0.0000	–
isLeaderOf	0.1055	0.1680	0.1296	0.2000	0.0008	0.0015
isLocatedIn	0.9804	0.5614	0.7139	0.9917	0.5499	0.7075
isMarriedTo	0.0529	0.9626	0.1003	0.2380	0.5994	0.3407
isPoliticianOf	0.0281	0.4393	0.0528	0.0000	0.0000	–
livesIn	0.1453	0.1740	0.1584	0.6474	0.0097	0.0191
participatedIn	0.0232	0.6970	0.0450	0.3000	0.0074	0.0144
playsFor	0.5055	0.6091	0.5525	0.8742	0.6043	0.7146
wasBornIn	0.4782	0.3779	0.4221	0.5231	0.2856	0.3695
worksAt	0.0538	0.5618	0.0981	0.4770	0.1137	0.1836
Macro Average	0.1897	0.6137	0.2898	0.4475	0.3388	0.3856
Micro Average	0.2740	0.5426	0.3641	0.7428	0.4813	0.5841

Error Analysis

- Error
  - False Postive
    - Ambiguity
    - Other
      - Pattern appear but no property
  - False Negative
    - Coverage

Error Analysis

- Other error in False Postive



Conclusion

- We use pattern to detecet entity properties in content stream efficiently and effectively.
- Entity’s type information and NBC play an important role in detection process.

Future work

- Solve ambiguity error, other error and coverage problem
- Pattern extraction
- Sentence level evaluation
- Short text (mircoblog, social message)
- Entity linking + property detection

Thanks!