Linking System at TAC KBP2014

Ludovic Jean-Louis - *Netmail Inc.* **Eric Charton** - *Yellow Pages Group*

Université Concordia University

Correction Processes & Mutual Disambiguation

- Use the semantic knowledge of annotations in a document to locally improve the precision of each specific annotation
- The Mutual Disambiguation module has 3 main steps:

1. Build a set of ranked candidate annotations
2. Apply correction processes (NE Label and Co-reference)
3. Apply Mutual Disambiguation Process

Document:

IBM has 12 research laboratories worldwide. In 1952, Thomas J. Watson, Jr., became president of the company.

Annotation object:

NE Mentions	Candidates annotations
IBM	International Brotherhood of Magicians
	International Business Machines
Thomas J. Watson	Thomas Watson, Jr.

Direct Semantic relations:

International Business Machines ⟷ Thomas Watson, Jr (10)
International Brotherhood of Magicians ⟷ Thomas Watson, Jr (0)

Common Semantic relations:

International Business Machines ⟷ Thomas Watson, Jr {IBM 7070, Software, History of IBM, ...}
International Brotherhood of Magicians ⟷ Thomas Watson, Jr {∅}

Query Merger

- 11 heuristics to combine the output of the two annotation engines

The merging algorithms consider the following information when evaluating each entity:

- NE Label (PER, ORG, GPE)
- Link (KB Node or NIL)
- Document type (DF, NW, WB)

Merging Heuristics

Best five strategies:

1. All entities from Wikimeta: **Wikimeta**$_{ALL}$
2. [h1]: **Wikimeta**$_{ALL}$ + (**AIDA**$_{ALL}$ - *redundancies*)
3. [h2]: **Wikimeta**$_{ALL}$ + (**AIDA**$_{ORG}$)
4. [h3]: **Wikimeta**$_{ALL}$ + (**AIDA**$_{ORG}$ - **AIDA**$_{NIL}$)
5. [h4]: (**Wikimeta**$_{ALL}$ - **Wikimeta**$_{ORG}$) + (**AIDA**$_{ORG}$)

SemLinker results on KBP2014 training set

SemLinker run	WikiF1	CEAFmP	CEAFmR	CEAFmF1	DiscP	DiscR	DiscF1	LinkP	LinkR	LinkF1
Wikimeta only	0.475	0.396	0.431	0.413	0.546	0.595	0.570	0.503	0.548	0.525
AIDA only	0.422	**0.457**	0.335	0.387	**0.647**	0.471	0.545	0.508	0.370	0.428
heuristic 1	**0.480**	0.380	**0.456**	0.414	0.534	**0.638**	**0.581**	0.487	**0.583**	**0.531**
heuristic 2	0.476	0.393	0.436	0.414	0.543	0.604	0.572	0.500	0.556	0.527
heuristic 3	0.478	0.389	0.447	**0.416**	0.539	0.620	0.577	0.494	0.569	0.529
heuristic 4	0.419	0.454	0.312	0.370	0.607	0.418	0.495	**0.562**	0.387	0.459

SemLinker results on KBP2014 test set

SemLinker run	WikiF1	CEAFmP	CEAFmR	CEAFmF1	DiscP	DiscR	DiscF1	LinkP	LinkR	LinkF1
CSFG1 (Wikimeta)	0.466	0.516	0.608	**0.558**	0.525	0.617	0.567	0.474	0.557	0.512
CSFG2 (h1)	**0.509**	0.359	**0.692**	0.473	0.516	**0.717**	**0.600**	0.434	**0.646**	0.519
CSFG3 (h2)	0.467	0.492	0.618	0.548	0.523	0.628	0.571	0.468	0.566	0.512
CSFG4 (h3)	0.505	0.447	0.652	0.530	0.524	0.671	0.588	0.460	0.605	**0.523**
CSFG5 (h4)	0.381	**0.582**	0.432	0.496	**0.597**	0.439	0.506	**0.537**	0.395	0.455

Related Publications

Eric Charton, Marie-Jean Meurs, Ludovic Jean-Louis, Michel Gagnon, **Mutual Disambiguation for Entity Linking**. ACL, Maryland USA, June 2014

Eric Charton, Marie-Jean Meurs, Ludovic Jean-Louis, Michel Gagnon, **SemLinker system for KBP2013: A disambiguation algorithm based on mutual relations of semantic annotations inside a document.**. TAC KBP, NIST 2013

Eric Charton, Juan-Manuel Torres-Moreno, **NLGbAse: a free linguistic resource for Natural Language Processing systems.** LREC2010, Malta, May 2010