
Contents

Introduction	ix
Chapter 1. Named Entities for Accessing Information	1
1.1. Research program history	2
1.1.1. Understanding documents: an ambitious task	2
1.1.2. Detecting basic elements: named entities	3
1.1.3. Trend: a return to slot filling	7
1.2. Task using named entities as a basic representation	9
1.3. Conclusion	10
Chapter 2. Named Entities, Referential Units	11
2.1. Issues with the named entity concept	12
2.1.1. A heterogeneous set	12
2.1.2. Existing defining formulas	17
2.1.3. An NLP object	21
2.2. The notions of meaning and reference	22
2.2.1. What is the reference?	22
2.2.2. What is meaning?	24
2.3. Proper names	27
2.3.1. The traditional criteria for defining a proper name	28
2.3.2. Meaning and referential function of proper names	30
2.3.3. The “referential load” of proper names	34
2.4. Definite descriptions	35
2.4.1. What is a definite description?	35
2.4.2. The meaning of definite descriptions	38

2.4.3. Complete and incomplete definite descriptions	39
2.5. The meaning and referential functioning of named entities	41
2.5.1. Reference to a particular	42
2.5.2. Referential autonomy	44
2.5.3. A “natural” heterogeneity	45
2.6. Conclusion	46
Chapter 3. Resources Associated with Named Entities	47
3.1. Typologies: general and specialist domains	48
3.1.1. The notion of category	48
3.1.2. Typology development	49
3.1.3. Typologies beyond evaluation campaigns	53
3.1.4. Other uses of typologies	54
3.1.5. Illustrated comparison	57
3.1.6. Issues to consider regarding entities	57
3.2. Corpora	59
3.2.1. Introduction	59
3.2.2. Corpora and named entities	60
3.2.3. Conclusion	65
3.3. Lexicons and knowledge databases	65
3.3.1. Lexical databases	66
3.3.2. Knowledge databases	72
3.4. Conclusion	75
Chapter 4. Recognizing Named Entities	77
4.1. Detection and classification of named entities	78
4.2. Indicators for named entity recognition	79
4.2.1. Describing word morphology	79
4.2.2. Using lexical databases	81
4.2.3. Contextual clues	83
4.2.4. Conclusion	85
4.3. Rule-based techniques	85
4.4. Data-driven and machine-learning systems	88
4.4.1. Majority class models	91
4.4.2. Contextual models (HMM)	92
4.4.3. Multiple feature models (Softmax and MaxEnt)	93

4.4.4. Conditional Random Fields (CRFs)	95
4.5. Unsupervised enrichment of supervised methods	95
4.6. Conclusion	96
Chapter 5. Linking Named Entities to References	99
5.1. Knowledge bases	100
5.2. Formalizing polysemy in named entity mentions	102
5.3. Stages in the named entity linking process	103
5.3.1. Detecting mentions of named entities	103
5.3.2. Selecting candidates for each mention	103
5.3.3. Entity disambiguation	104
5.3.4. Entity linking	106
5.4. System performance	106
5.4.1. Practical application: DBpedia Spotlight	107
5.4.2. Future prospects	108
Chapter 6. Evaluating Named Entity Recognition	111
6.1. Classic measurements: precision, recall and F-measures	112
6.2. Measures using error counts	115
6.3. Evaluating associated tasks	120
6.3.1. Detecting entities and mentions	121
6.3.2. Entity detection and linking	122
6.4. Evaluating preprocessing technologies	126
6.5. Conclusion	128
Conclusion	131
Appendices	137
Appendix 1. Glossary	139
Appendix 2. Named Entities: Research Programs	141
Appendix 3. Summary of Available Corpora	147
Appendix 4. Annotation Formats	151
Appendix 5. Named Entities: Current Definitions	153
Bibliography	157
Index	169