NNE: A Dataset for Nested Named Entity Recognition in English Newswire

Nicky Ringland, Xiang Dai, Ben Hachey, Sarvnaz Karimi, Cecile Paris, James R. Curran
University of Sydney and CSIRO Data61

Why recognize nested named entities?

- Most NER tools capture **only flat mention structure**, reflecting the available annotated datasets
- **Ignores important information** useful for downstream tasks, e.g.:

Entity-entity relationships

... the Ontario Supreme Court said it will postpone ...
STATE
GOVERNMENT

Entity attribute values

Former U.N. Ambassador Jeane Kirkpatrick ...
ORG:OTHER ROLE FIRST NAME
ROLE PER
ROLE
PER

Part-whole relationships

... this wealthy Southern California beach community ...
STATE
REGION

Annotation Schema and Process

- Use the flat **BBN** annotation as a *starting point*
- Augment with nested structure and fine-grained entity types
- Annotate **all named mentions** including *time*, *date* and *numerical* entities
- Annotate **all structural elements** including *nested mentions*
- Add **consistent substructure** to *avoid spurious ambiguity* e.g., University of Toronto
- **4 annotators**, background in linguistics and/or computational linguistics
- **270 hours** total annotation time
- **2x annotation** of Sections 00 and 23; **4x annotation** of Section 02
- **17 hours** additional time for adjudicating multiple Sections 00, 02, 23
- **0.907 Fleiss' kappa** over token-level tag stacks on Section 02

References

[1] Bailin Wang and Wei Lu. Neural Segmental Hypergraphs for Overlapping Mention Recognition. In EMNLP. 2018.

[2] Bailin Wang et al. A Neural Transition-based Model for Nested Mention Recognition. In EMNLP. 2018.

Why use NNE?

Item	NNE	GENIA	ACE05
Documents	2,312	2,000	464
Sentences	49,208	18,546	12,548
Sentences w. nesting	32,387	9,533	4,266
Tokens	1.1M	0.5M	0.3M
Mentions	279,795	92,681	30,966
Entity types	114	36	7
Mentions per sentence	5.69	4.99	2.46
Top-level mentions	118,525	76,582	23,464
Maximum depth	6	4	6

- **Large**, **nested**, **fine-grained** named entity recognition dataset
- **279,795 mentions** of **114 entity types** with up to **6 layers** nesting
- Built on the **Penn Treebank**, providing opportunity for *joint exploration with other NLP tasks*

Benchmark Results

	P	R	F_1
BiLSTM-CRF-OUTER	89.9	38.0	53.5
BiLSTM-CRF-INNER	93.8	62.0	74.7
BiLSTM-CRF-UNION	92.2	85.8	88.9
Hypergraph [1]	91.8	91.0	91.4
Transition [2]	77.4	70.1	73.6

- Flat NER models can achieve high precision but suffer from **low recall**
- Hypergraph-based model performs best on our dataset, but with substantially **low decoding speed**

Using NNE

- NNE comprises **new standoff annotation** over the Penn Treebank
- Also includes **code for knitting, evaluation and analysis**
- **Freely available** under permissive licences