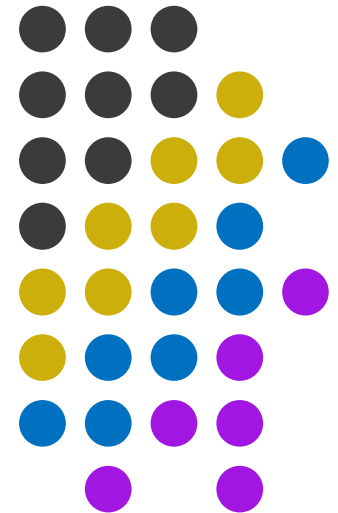


# Named Entity Recognition in Biomedical Domain with an Optimized Tag-set

---

Han-Cheol Cho  
Tsujii Lab., The University of Tokyo



# Background

- What is Named Entity Recognition (NER) in biomedical domain?
  - To **1) recognize** and **2) classify** the target names such as *Protein, DNA, RNA* and so on
- How can it be solved?
  - Researchers love **a sequential labeling method**

|                |             |
|----------------|-------------|
| ...            |             |
| Number         | O           |
| of             | O           |
| glucocorticoid | B-protein   |
| receptors      | I-protein   |
| in             | O           |
| lymphocytes    | B-cell_type |
| and            | O           |
| their          | O           |
| sensitivity    | O           |
| to             | O           |
| hormone        | O           |
| action         | O           |
| .              | O           |
| ...            |             |

# Tag-sets for NER

- Various tag-sets in NER
  - IO, IOB1, IOB2, IOE1, IOE2, BILOU, ...
  - **IOB2 tag-set** is the most popular one
- A problem of IOB2 tag-set
  - IOB2 tag-set is **too coarse-grained**

|                |                  |
|----------------|------------------|
| ...            |                  |
| of             | O                |
| glucocorticoid | <b>B-protein</b> |
| receptors      | <b>I-protein</b> |
| (              | O                |
| GR             | <b>B-protein</b> |
| )              | O                |
| ...            |                  |

|                  |                  |
|------------------|------------------|
| ...              |                  |
| of               | O                |
| thyroid          | <b>B-protein</b> |
| <b>hormone</b>   | <b>I-protein</b> |
| <b>receptors</b> | <b>I-protein</b> |
| in               | O                |
| ...              |                  |

# Contradictory Result

- (Ratinov et al, 2009) claimed that **BILOU tag-set outperforms IOB2 tag-set** in general domain
- However, experiment results of NER in bio-domain shows that **IOB2 tag-set is slightly better** than BILOU tag-set