IOB/BILOU

The term "IOB/BILOU" in the context of Named Entity Recognition (NER) refers to two common tagging schemes used for annotating entities in text data. Here is a brief explanation of each tagging scheme:

1. **IOB (Inside, Outside, Beginning)**:
   * In the IOB tagging scheme, each token in a sequence is tagged with one of three labels: "I" for tokens inside an entity, "O" for tokens outside any entity, and "B" for the beginning of an entity.
   * For example, in a sentence "Apple is a company," the IOB tags for "Apple" (a company name) would be "B-ORG" (beginning of an organization) and "O" (outside any entity) for the other tokens.
2. **BILOU (Beginning, Inside, Last, Outside, Unit)**:
   * The BILOU tagging scheme is an extension of the IOB scheme that includes additional tags to mark the last token of an entity and single-token entities.
   * The tags used in BILOU are "B" for the beginning of an entity, "I" for inside an entity, "L" for the last token of an entity, "U" for single-token entities, and "O" for tokens outside any entity.

These tagging schemes are commonly used to annotate text data for NER tasks to indicate the boundaries and types of entities present in the text. By using these standardized tagging schemes, NER models can learn to recognize and extract entities accurately from text data.If you have annotated data using either the IOB or BILOU tagging scheme, you can use this annotated data to train NER models such as spaCy's NER model or BERT by converting the annotations into a format compatible with the model's training requirements.