

Inside–outside–beginning (tagging)

The **IOB format** (short for inside, outside, beginning) is a common tagging format for tagging tokens in a chunking task in computational linguistics, (ex.named-entity recognition).^[1] It was presented by Ramshaw and Marcus in their paper "Text Chunking using Transformation-Based Learning", 1995 ^[2] The B- prefix before a tag indicates that the tag is the beginning of a chunk, and an I- prefix before a tag indicates that the tag is inside a chunk. The B- tag is used only when a tag is followed by a tag of the same type without O tokens between them. An O tag indicates that a token belongs to no chunk.

Another similar format which is widely used is IOB2 format, which is the same as the IOB format with the difference that the B- tag is used in the beginning of every chunk (i.e. all chunks start with the B- tag).

A readable introduction to entity tagging is given in Bob Carpenter's blog post, "Coding Chunkers as Taggers" ^[3]. 'BIO' is plausibly a synonym for 'IOB'.

References

1. "Entity Recognition" (<http://www.evalita.it/2009/tasks/entity>)
2. Ramshaw and Marcus (1995). "Text Chunking using Transformation-Based Learning" (<https://arxiv.org/abs/cmp-lg/9505040>).
3. Bob Carpenter (1995). "Coding Chunkers as Taggers: IO, BIO, BMEWO, and BMEWO+" (<https://lingpipe-blog.com/2009/10/14/coding-chunkers-as-taggers-io-bio-bmewo-and-bmewo/>)

Retrieved from '[https://en.wikipedia.org/w/index.php?title=Inside–outside–beginning_\(tagging\)&oldid=804180597](https://en.wikipedia.org/w/index.php?title=Inside–outside–beginning_(tagging)&oldid=804180597)

This page was last edited on 7 October 2017, at 08:35.

Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.