

Matches encoding

We will use **BIOES** coding (Begin, Inside, Outside, End, Single)

- B – if token matches the beginning of some entity
- B, I – if two tokens match as prefix
- I, E – if two tokens match as postfix
- S – if matched single token entity
- ...

Example for 4 lexicon dictionaries:

Text	Hayao	Tada	,	commander	of	the	Japanese	North	China	Area	Army
LOC	–	–	–	–	–	B	I	–	S	–	–
MISC	–	–	–	S	B	B	I	S	S	S	S
ORG	–	–	–	–	–	B	I	B	I	I	E
PERS	B	E	–	–	–	–	–	–	S	–	–

B, I, O, E, S are later encoded as one-hot vectors