FastText: Subword Embeddings via Character N-grams

Words

Character N-grams (3-6 chars)

playing

<pl	pla	lay	ayi	yin	ing	ng>

unhappiness

<un	unh	nha	hap	app	ppi	...

COVID

<CO	COV	OVI	VID	ID>
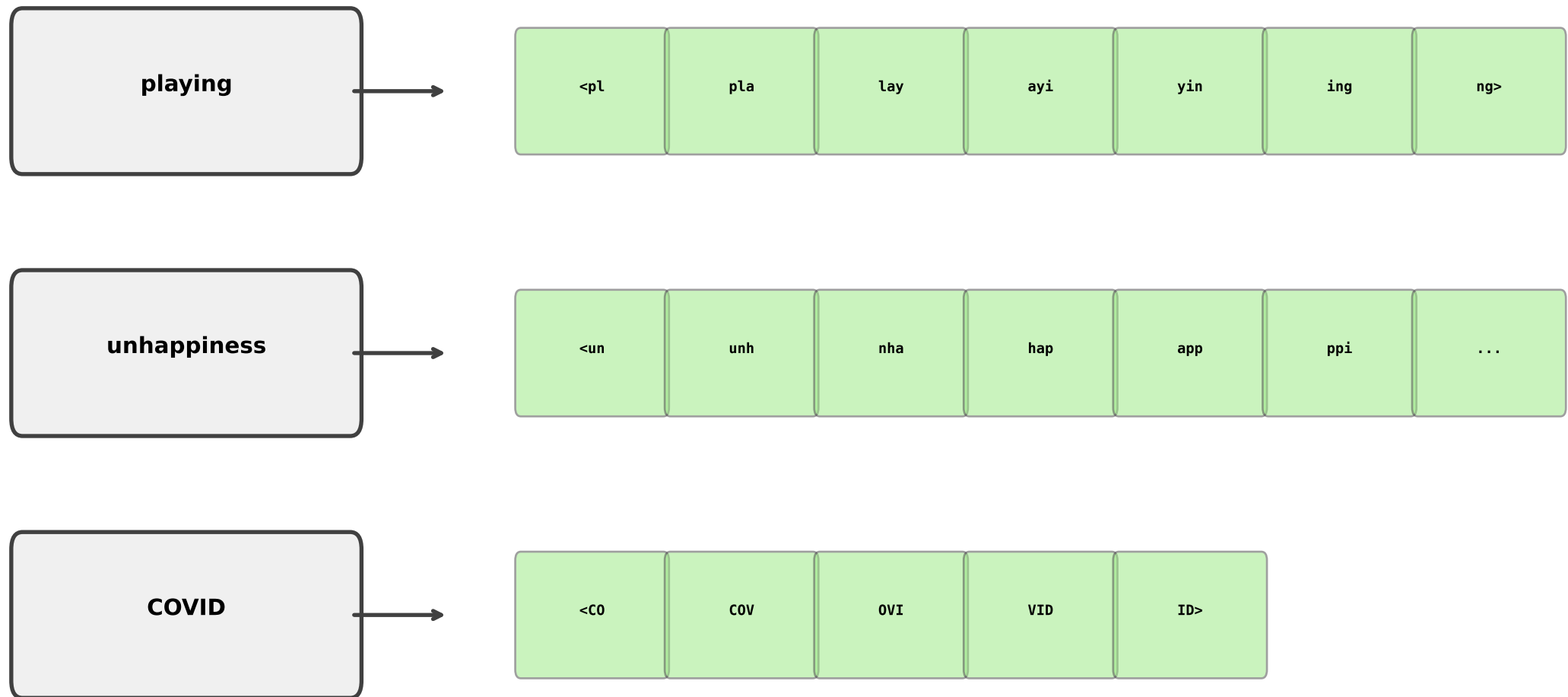
$$FastText: v_{word} = \sum_{g \in ngrams} v_g$$