

start sequence



Embedding



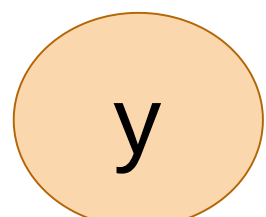
concat



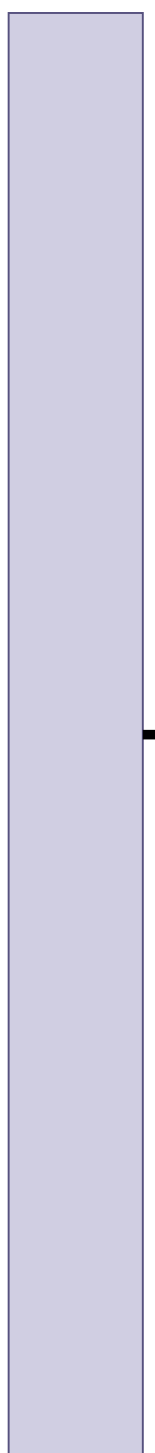
MLP



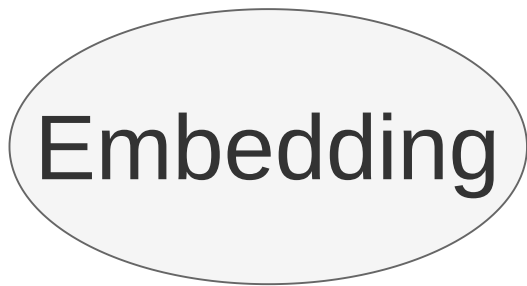
y



end sequence



Embedding



Lengths

