Appendix A

Anonymous ACL-IJCNLP submission
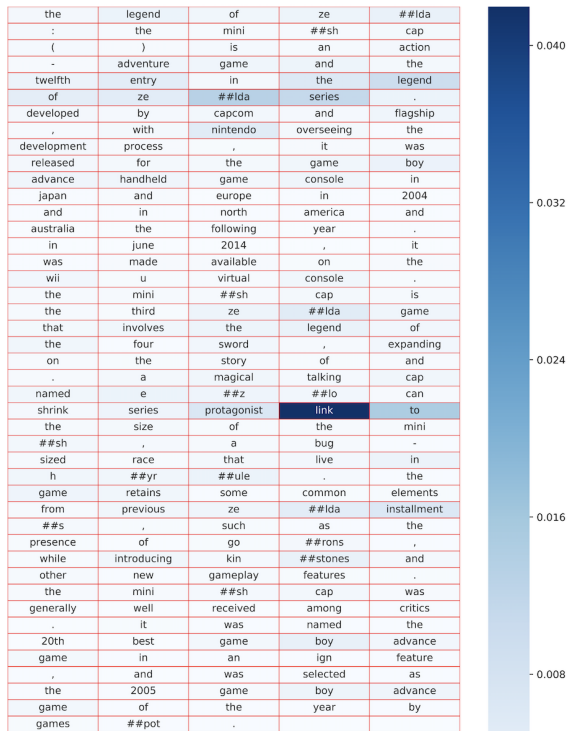

Figure 1: Attention heatmap of BERT$_{\text{BASE}}$ over the tokenized document of the DocRED example

Figure 2: Attention heatmap of the proposed *E2GRE* over the tokenized document of the DocRED example

We show attention heatmaps for an example document in the DocRED development set, one for the BERT$_{\text{BASE}}$ and one for our *E2GRE* model, as a case study to see how our model is working.

We show that the the attention values for the BERT$_{\text{BASE}}$ model are highly concentrated on the single token "Link", and that all other tokens in the document share very low attention values. The differences in attention between evidence and non-evidence sentences are very small. These attention values are much smaller in scale that the attention values of our *E2GRE* model.

On the heatmap for our *E2GRE* model, we notice that the attention values of our pretrained language model is more focused on the correct evidence sentences. On average, the tokens within evidence sentences 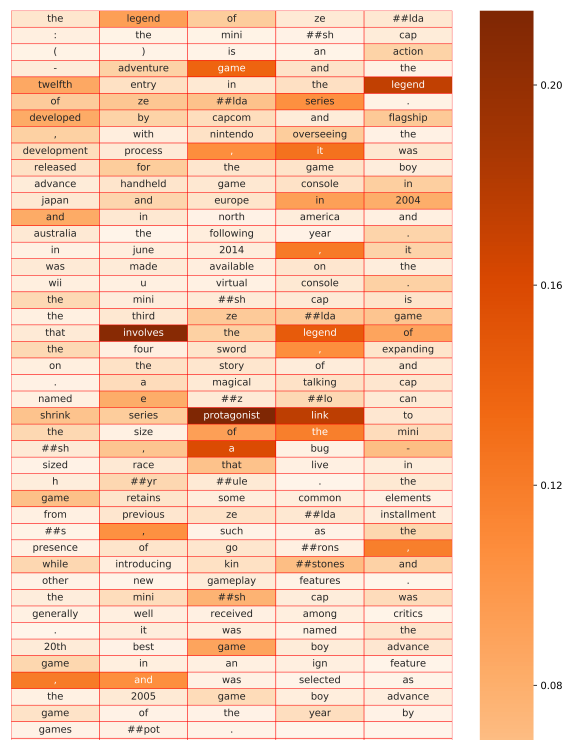are higher than the tokens outside of the evidence sentences.