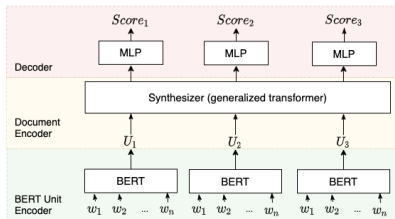


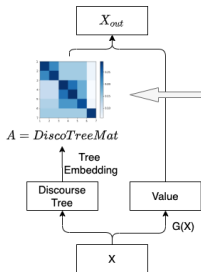
Do We Really Need That Many Parameters In Transformer For Extractive Summarization? Discourse Can Help !

Wen Xiao, Patrick Huber and Giuseppe Carenini | University of British Columbia

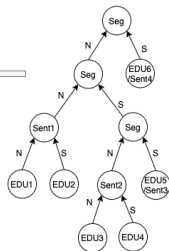
- ▶ **Goal:** Reduce the number of learnable parameters while keeping a competitive performance in extractive summarization
- ▶ **Idea:** Incorporate the discourse information to the summarizer's attention module
- ▶ **How:**
 - 1 Embed the discourse tree (Dep/Const) into an attention matrix (three variants)
 - 2 Replace the original dot-product self-attention with the fixed tree attention



Extractive Summarization Model



Discourse Tree Attention



Discourse Tree