



Source material



Dialogue extraction, character clustering,  
context extraction



Large language  
models



Evaluation

*A Song of Ice and Fire*

5 books - 1778216  
words

*Harry Potter*

7 books - 1087549  
words

*BookNLP*

"Harry - yer a  
wizard" said by  
**Hagrid**

*Character  
clustering*

**Harry** is the same  
as **Harry Potter** is  
the same the **Chosen  
one**

*Context  
extraction using  
SpaCY*

Extract a number of  
sentences before  
the dialogue

*In-Context  
Learning*

Retrieval augmented  
generation

*Summarization*

*Sorting hat test*

*Trivia questions*

*Manual evaluation*