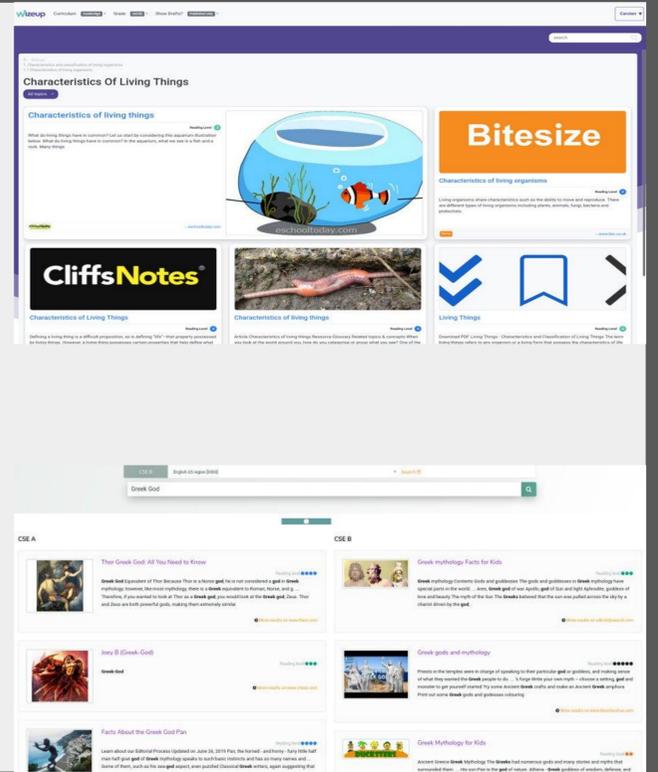


Neural Information Retrieval for Educational Resources

ECIR 2022 – Industry Day

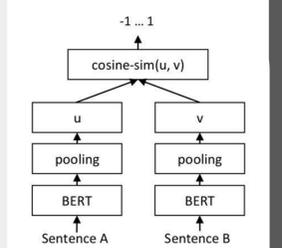
Background and Motivation

- **Wizenoze - The Company**
 - “Wizenoze brings structure in a world of information overload”
 - EdTech provider since 2013
- **Product**
 - API providing educational content
 - Content selection: human curators plus AI / NLP
 - Tailored for custom-defined curricula
 - Backed by search engine
 - **Better search results -> less human curation effort**



Train a Sentence BERT Model for Educational Resources

- Fine-tuned Sentence BERT on historic curation data
 - ~33k in-domain query-document-judgement triplets
- Model selection on held-out validation set
- Vanilla S-BERT vs fine-tuned S-BERT on test sub-set to evaluate training impact



	P@5	P@10	MAP@5	MAP@10
MiniLM-L6-v2	0.7926	0.7888	0.8505	0.8504
mpnet-base-v2	0.7851	0.7926	0.8619	0.8654
Fine-tuned	0.8889	0.8667	0.9370	0.9260

Evaluation and Validation

- Evaluate on separate test set:
 - 39 hand-picked test queries
- Sanity checks: manually validate 50 most different results sets (RBO score)
- A/B testing (WIP):
 - which model is preferred by curators?

	P@5	P@10	MAP@5	MAP@10
BM25+	0.6963	0.6407	0.8402	0.8039
Fine-tuned	0.7926	0.7519	0.8727	0.8632

Summary and Conclusions

- Neural Information Retrieval: significant improvements on education domain search
- Learning in-domain expert preferences
- Affordable with serverless implementation, gradual transition

