Dense retrieval system for general court laws

Adriana NICOARA, Dalila LADLI, Silviya SILWAL

Institut des Sciences du Digital
Pôle Herbert Simon
13 rue Michel Ney
54000 Nancy, France

January 23rd, 2023

Overview

- Baseline system - simple retrieval of k most similar contexts
- Dense retrieval model - retrieval + re-ranking of the outputs based on the synthetic nli dataset generated from our questions and contexts
- Creation of a first draft of the front-end
- Automatic and Human Evaluation

Front-end

- Used streamlit to create front-end
- Users can select any one of the 3 models and get related contexts
- Users are able to select the number of contexts they want to retrieve
- Online version hosted on HuggingFace spaces, running on CPU

Front-end

Civil Legal Query

Quelles protections la Loi sur la protection du consommateur accorde-t-elle aux individus?

Choose Number of Result:

10

Choose Model

- (•) Model 1
- ( ) Model 2
- ( ) Model 3

Run

Context: la convocation a une assemblee generale de coproprietaires concernant des epoux proprietaires d'un lot dependant de leur communaute de biens doit etre libellee au nom des deux

Context: le juge n'est pas tenu , en cas de desaccord des parents , d'ordonner la residence en alternance a titre provisoire

Context: l'indemnite d'eviction due au locataire doit etre evaluee en excluant la taxe sur la valeur ajoutee du chiffre d'affaires qui a servi de base a son calcul si tels sont les usages de la profession exercee par le preneur

Context: pour le calcul de la surface maximale prevue par l'article l. 12 - 6 du code rural et de la peche

Human Evaluation

- Human evaluation done using likert scale (0-5)
- Same interface as front-end but with added rating feature
- Users rate the retrieved information of each model, but are unaware of model type
- User rating data is stored in a json file
- Storing query, retrieved contexts, model type and rating
- No personal information is stored

Automatic Evaluation

- Using embedding based scoring system, similar to BERTScore[2], DepthScore[1]
- Compute similarity of query and contexts as a average of cosine similarity between their embeddings
- Use different model sentence transformer model[1] unseen by our system to get embeddings
- Compute average score from a list of queries to get overall score
- Compare results among the three models
- Preliminary result based on 14 questions

Models	Mean Similarity Score	Std.
Civile-law-IR	0.88	0.018
STSB	0.83	0.022
DR-Baseline	0.86	0.019

Table: Overall Mean similarity score and Standard deviation for 3 dense retrieval models

[1] https://huggingface.co/nreimers/MiniLM-L6-H384-uncased

Automatic Evaluation results

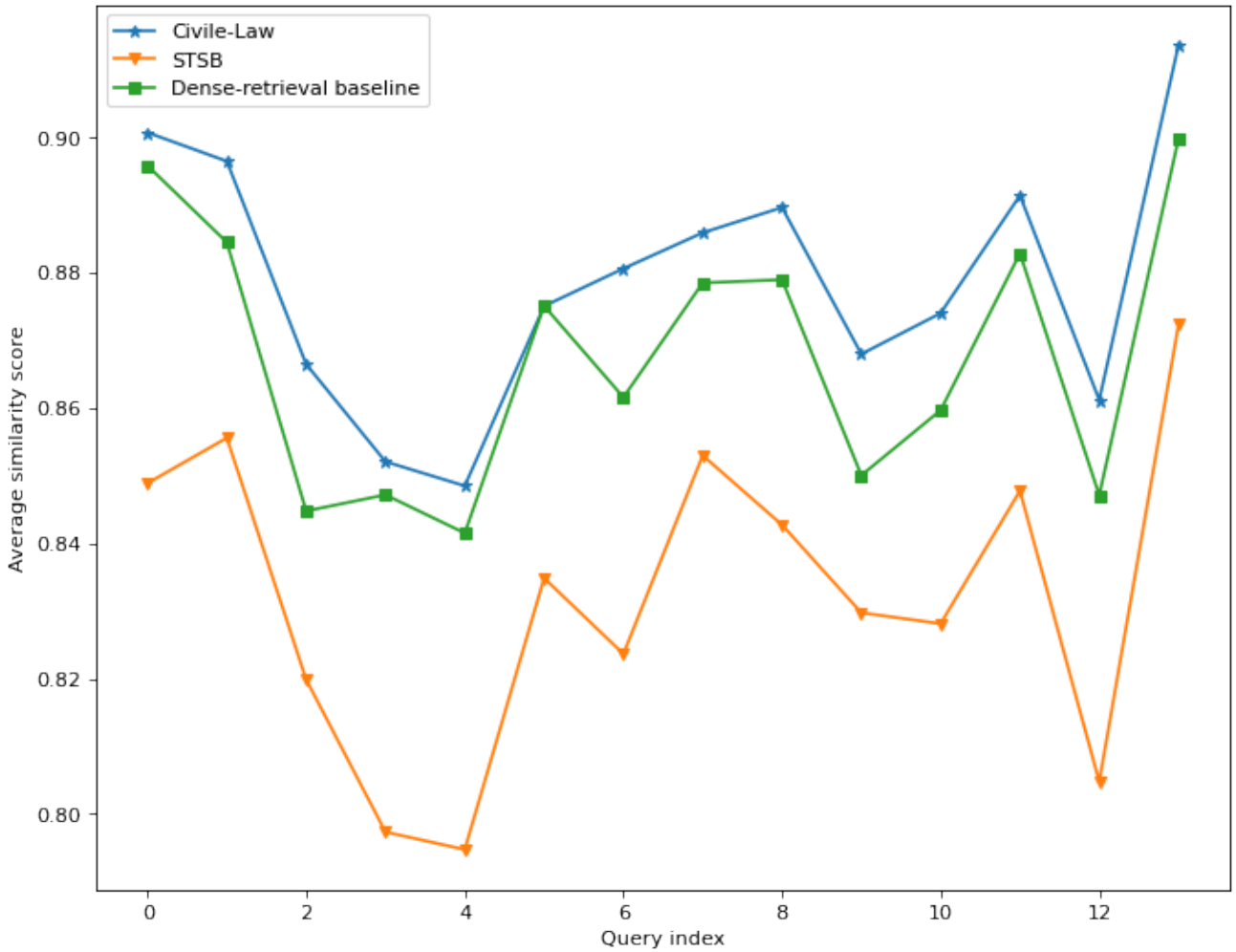

Figure: Comparison based on mean similarity score for the 3 dense retrieval models

Timeline

- Jan 23 - Jan 27:
  - Complete human evaluation
  - Prepare the defense
  - Final report corrections

References


G. Staerman, P. Mozharovskyi, P. Colombo, S. Clémençon, and F. d'Alché Buc.
A pseudo-metric between probability distributions based on depth-trimmed regions, 2021.


T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi.
Bertscore: Evaluating text generation with BERT.
CoRR, abs/1904.09675, 2019.

Thank you for your attention !