

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

November 18, 2022



- 1 Quality of question generation
- 2 Data preparation
- 3 Question generation
- 4 Baseline System
- 5 Next steps

- 1 Quality of question generation
- 2 Data preparation
- 3 Question generation
- 4 Baseline System
- 5 Next steps

sumy library space for automatic text summarization

This is a space for [sumy](#), an automatic text summarization library. The summary can be extracted either from an HTML page or plain text. You can find a list of available summarization methods [here](#).

method

edmundson ▾

language

french ▾

sentence count

1

input type

☐ URL
 ☒ text

input

les tribunaux judiciaires ont competence pour se prononcer tant sur les dommages - interets a allouer aux tiers leses par le voisinage d' une installation classee pour la protection de l' environnement que sur les mesures propres a faire cesser le prejudice que cette installation pourrait causer dans l' avenir , a condition que ces mesures ne contrarient pas les prescriptions edictees par l' administration en vertu des pouvoirs de police speciale qu' elle detient . le

Nettoyer

Soumettre

output

les tribunaux judiciaires ont competence pour se prononcer tant sur les dommages - interets a allouer aux tiers leses par le voisinage d' une installation classee pour la protection de l' environnement que sur les mesures propres a faire cesser le prejudice que cette installation pourrait causer dans l' avenir , a condition que ces mesures ne contrarient pas les prescriptions edictees par l' administration en vertu des pouvoirs de police speciale qu' elle detient .

Activer Windows

Figure: Summary of a case's decision made by the Court de cassation

Text simplification models

- MILES: Multilingual Lexical Simplifier
- ACCESS: AudienCe-CEntric Sentence Simplification
- Lexical simplification with BERT

Evaluation : EASSE (Easier Automatic Sentence Simplification Evaluation)

MILES: Multilingual Lexical Simplifier

```
BERT model loaded!

Attempting to load embeddings...

No embeddings for fr found. Using without...

Starting MILES Flask server...

Loaded as HTTP Server on port 80, running forever:
Traceback (most recent call last):
  File "/mnt/c/Users/didij/Documents/Université/M2/S1/CBR/Miles-master/app.py", line 64, in <
    http_server.serve_forever()
  File "/mnt/c/Users/didij/Documents/Université/M2/S1/CBR/MILES-master/env/lib/python3.10/sit
    self.start()
  File "/mnt/c/Users/didij/Documents/Université/M2/S1/CBR/MILES-master/env/lib/python3.10/sit
    self.init_socket()
  File "/mnt/c/Users/didij/Documents/Université/M2/S1/CBR/MILES-master/env/lib/python3.10/sit
    StreamServer.init_socket(self)
  File "/mnt/c/Users/didij/Documents/Université/M2/S1/CBR/MILES-master/env/lib/python3.10/sit
    self.socket = self.get_listener(self.address, self.backlog, self.family)
  File "/mnt/c/Users/didij/Documents/Université/M2/S1/CBR/MILES-master/env/lib/python3.10/sit
    return _tcp_listener(address, backlog=backlog, reuse_addr=cls.reuse_addr, family=family)
  File "/mnt/c/Users/didij/Documents/Université/M2/S1/CBR/MILES-master/env/lib/python3.10/sit
    sock.bind(address)
  File "/mnt/c/Users/didij/Documents/Université/M2/S1/CBR/MILES-master/env/lib/python3.10/sit
    return self._sock.bind(address)
PermissionError: [Errno 13] Permission denied: ('0.0.0.0', 80)
```

Figure: Permission issue

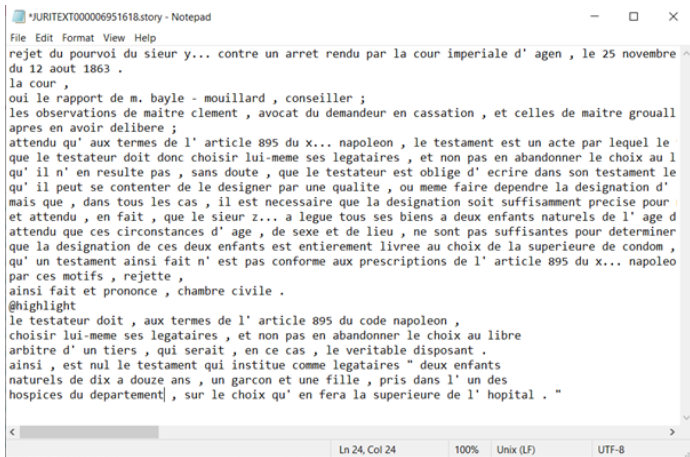
- 1 Quality of question generation
- 2 Data preparation**
- 3 Question generation
- 4 Baseline System
- 5 Next steps

Preprocessing steps

- preprocessing script¹ => generate the .story files
 - ▶ 53220 civile cases files
 - ▶ 17314 cases after year 2000
- some inconsistencies => delete unwanted characters

¹<https://github.com/euranova/CASS-dataset>

Story files



```
*JURITEXT000006951618.story - Notepad
File Edit Format View Help
rejet du pourvoi du sieur y... contre un arret rendu par la cour imperiale d' agen , le 25 novembre
du 12 aout 1863 .
la cour ,
oui le rapport de m. bayle - mouillard , conseiller ;
les observations de maitre clement , avocat du demandeur en cassation , et celles de maitre grouall
apres en avoir delibere ;
attendu qu' aux termes de l' article 895 du x... napoleon , le testament est un acte par lequel le
que le testateur doit donc choisir lui-meme ses legataires , et non pas en abandonner le choix au l
qu' il n' en resulte pas , sans doute , que le testateur est oblige d' ecrire dans son testament le
qu' il peut se contenter de le designer par une qualite , ou meme faire dependre la designation d'
mais que , dans tous les cas , il est necessaire que la designation soit suffisamment precise pour
et attendu , en fait , que le sieur z... a legue tous ses biens a deux enfants naturels de l' age d
attendu que ces circonstances d' age , de sexe et de lieu , ne sont pas suffisantes pour determiner
que la designation de ces deux enfants est entierement livree au choix de la superieure de condom ,
qu' un testament ainsi fait n' est pas conforme aux prescriptions de l' article 895 du x... napoleo
par ces motifs , rejette ,
ainsi fait et prononce , chambre civile .
@highlight
le testateur doit , aux termes de l' article 895 du code napoleon ,
choisir lui-meme ses legataires , et non pas en abandonner le choix au libre
arbitre d' un tiers , qui serait , en ce cas , le veritable disposant .
ainsi , est nul le testament qui institue comme legataires " deux enfants
naturels de dix a douze ans , un garcon et une fille , pris dans l' un des
hospices du departement| , sur le choix qu' en fera la superieure de l' hopital . "
```

Ln 24, Col 24 100% Unix (LF) UTF-8

Figure: Generated .story file

- 1 Quality of question generation
- 2 Data preparation
- 3 Question generation**
- 4 Baseline System
- 5 Next steps

- based on .story files
- context - the summarization made by the lawyer
- split the summarization by point
- one or more questions per context

Json file

```
{
  "documentID": "JURITEXT000007040210.story",
  "items": [
    {
      "context": "l'article 1. 21 - 10 du code des assurances subordonne , en cas d'alienatio",
      "questions": [
        {
          "question": "Quel article du code des assurances subordonne la transmission de"
        }
      ]
    },
    {
      "context": "des lors la cour d'appel , ayant constate que le dommage , susceptible d'ou",
      "questions": [
        {
          "question": "Qu'est-ce que la cour d'appel a considéré comme susceptible d'ouvr"
        }
      ]
    }
  ]
},
{
```

Figure: Example json file

- 1 Quality of question generation
- 2 Data preparation
- 3 Question generation
- 4 Baseline System**
- 5 Next steps

Baseline model

- Performing information retrieval
 - ▶ Using pre-trained model
 - all-mpnet-base-v2 [Song et al., 2020]
 - all-distilroberta-v1 [Sanh et al., 2019]
 - ▶ cosine similarity – basic method AND
 - ▶ FAISS - (Facebook AI Similarity Search)
 - an algorithm that searches for embeddings of documents that are similar to each other

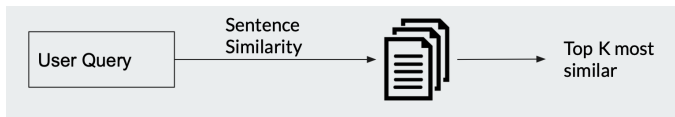


Figure: Baseline system

Example from baseline

Q: Ai-je droit à l'aide juridique ?

"la securite juridique , invoquee sur le fondement du droit a un proces equitable pour contester l'application immediate d'une solution nouvelle resultant d'une evolution de la jurisprudence , ne saurait consacrer un droit acquis a une jurisprudence figee , des lors que la partie qui s'en prevaut n'est pas privee du droit a l'acces au juge (arret n 1 , pourvoi n 07 - 14.932 et arret n 2 , pourvoi n 08 - 16.914)

"lorsque des decisions etrangeres sont annulees , l'arret qui les avait declarees executoires se trouve prive de fondement juridique "

Time Taken: 129.82898473739624|

Figure: Q1: result

Example from baseline

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

"la participation de la personne protegee au financement de la mesure de protection exercee par un mandataire judiciaire a la protection des majeurs est fonction de ses ressources et ce n'est que lorsque le juge des tutelles est saisi d'une demande d'indemnité exceptionnelle que des diligences particulierement longues ou complexes peuvent etre prises en consideration pour allouer cette indemnité au mandataire , en complement des sommes percuees ",

"lorsqu'un conjoint survivant , marie sous le regime legal et designé beneficiaire en premier lieu de contrats d'assurance - vie , est decede sans en avoir accepte le benefice , l'enfant designé beneficiaire en dernier lieu et acceptant est , en application de l'article l. 132 - 12 du code des assurances , repute avoir droit aux sommes stipulees aux contrats a partir du jour de leur souscription .il en resulte que les dispositions de l'article l. 132 - 16 du code des assurances ne sont pas applicables et qu'en vertu de l'article 1437 du code civil , la succession du souscripteur des contrats est redevable envers la communaute des deniers communs ayant servi a acquitter une charge contractee dans l'interet personnel de celui-ci ",

"l'inobservation , par le fonds de garantie des assurances obligatoires de dommages , du delai de six mois qui lui est imparti par l'article r. 421 - 68 , alinea 2 , du code des assurances , ne le prive pas du droit d'agir en justice pour contester l'exception de non-garantie soulevee par l'assureur "

Time Taken: 122.8533890247345]

Figure: Q2: result

Dense retrieval model

- As we saw, baseline system does not perform good enough
- Create better retrieval system
 - ▶ Model to retrieve relevant info (DPR)
 - ▶ Model to re-rant retrieved info (BERT, ELECTRA)

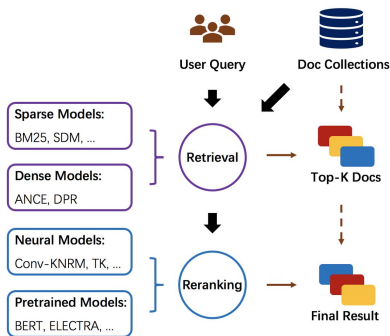


Figure: Architecture for dense model²

²<https://github.com/thunlp/OpenMatch>

- 1 Quality of question generation
- 2 Data preparation
- 3 Question generation
- 4 Baseline System
- 5 Next steps**

Timeline

- Nov 21st - Dec 30:
 - ▶ Find and test text simplification model
 - important for QA generation → works better with short sentences
 - ▶ Dense information retrieval model
- January:
 - ▶ Front end development
 - ▶ Final tests and report writing

References



Sanh, V., Debut, L., Chaumond, J., and Wolf, T. (2019).
Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter.



Song, K., Tan, X., Qin, T., Lu, J., and Liu, T.-Y. (2020).
Mpnnet: Masked and permuted pre-training for language understanding.