

Enrich4All – proiect european pentru digitalizarea administrației publice

Radu ION

radu@racai.ro

Institutul de Cercetări pentru Inteligență Artificială „Mihai Drăgănescu”

Enrich4All (<https://www.enrich4all.eu/>)

- Proiect european desfășurat în perioada 1 iunie 2021 – 31 mai 2023
- Obiective:
 - Integrarea platformei de traducere automată eTranslation în chatbot-ul inteligent creat de SupWiz Aps
 - Instalarea chatbot-ului multi-lingual pe site-uri ale unor instituții publice din România, Luxemburg și Danemarca
- Parteneri
 - [ICIA](#)
 - [Luxembourg Institute of Science and Technology](#)
 - [BEIA Consult International](#)
 - [SupWiz Aps](#)

Resurse dezvoltate în proiect

- Un model BERT pentru luxemburgheză
 - Această limbă nu avea niciun model pe Hugging Face
- Un detector de limbă pentru limbile din proiect care include și luxemburgheza
 - Disponibil într-un container Docker
- Corpusuri specializate pentru COVID-19 și autorizații de construcție
- Seturi de date cu întrebări și răspunsuri în cele două domenii
- Un corpus de documente ale Școlii Doctorale a Academiei Române (SCOSAAR)
- Un set de întrebări și răspunsuri pentru cei interesați de desfășurarea doctoratului la Academia Română

QA cu BING și BERT

- R. Ion, A. Avram, V. Păiș, M. Mitrofan, V. Barbu Mititelu, E. Irimia și V. Badea. „An Open Domain QA System for eGovernance”. CLIB 2022.
- Este un sistem de întrebare-răspuns care:
 - Trimite întrebarea utilizatorului la motorul de căutare BING pentru a obține fragmente de text care pot conține răspunsul
 - Utilizează o rețea neuronală BERT românească pentru a căuta răspunsul în fragmentele de text găsite de BING

Query gen.	MRR	Exact %	F1 %
Baseline query gen.	0.4056	33.85	65.07
CW query gen.	0.5337	50.77	76.69

Modele BERT pentru limba română

Model	Exact %	F1 %
BERT-base-ro-uncased	71.33	77.25
BERT-base-ro-cased	73.33	76.75
RoBERT-small	58.00	61.64
RoBERT-medium	59.33	63.06
RoBERT-large	61.00	64.12
Distil-BERT-base-ro	51.33	70.39
Distil-RoBERT-base	55.33	61.72
DistilMulti-BERT-base-ro	51.33	70.39

Chatbot SCOSAAR

- ICIA a adaptat chatbot-ul SupWiz pentru a răspunde la întrebări în română (și engleză, prin traducere automată) despre desfășurarea doctoratului în Academia Română
- Chatbot-ul este disponibil la <https://acad.ro/institutia/scosaar.html>
- Pentru a conduce un dialog cu utilizatorul, chatbot-ul SupWiz clasifică automat întrebările/răspunsurile utilizatorului pentru a identifica nodul din arborele de dialog care se potrivește cel mai bine cu intenția exprimată

Întrebări și răspunsuri pentru chatbot-ul SCOSAAR

- Clasificatoarele din chatbot-ul SupWiz au nevoie de date de antrenare reprezentate de întrebări asociate cu etichete
- Echipa ICIA a dezvoltat manual un set de întrebări și răspunsuri, clasificate în domenii și subdomenii:
 - admitere: 800
 - bursă: 468903
 - coordonator: 425
 - durata studiilor: 112
 - studii cu frecvență sau fără: 10
 - metodologie: 121908
 - obligații de publicare: 204
 - formatarea tezei de doctorat: 726
 - regulations: 251
 - susținerea tezei de doctorat: 67
 - transfer: 619

Formatul unei perechi întrebare-răspuns

- L: durata
- Q: Care este durata [studiilor universitare de doctorat/doctoratului]?
- Q: [Cât/Câți ani] [durează/ține] doctoratul?
- Q: [Cât/Câți ani] [durează/țin] studiile universitare de doctorat?
- Q: Cât îmi ia să [fac/urmez/termin] [studiile universitare de doctorat/doctoratul]?
- A: Conform Regulamentul de organizare a studiilor universitare de doctorat în cadrul Școlii de Studii Avansate a Academiei Române, capitolul I, articolul 6, alineatul (1), **[durata studiilor universitare de doctorat este de trei ani, cu posibilitatea întreruperii și/sau prelungirii instituționale]**. Mai multe informații aici <https://acad.ro/scosaar/doc2013/doc2013-0910Regulament.pdf>

Dezvoltarea chatbot-ului în BotStudio

The screenshot displays the BotStudio interface for a chatbot named SCOSAAR. The interface is divided into three main sections:

- Left Sidebar (Navigation):**
 - BUILD:** Conversation Flow (selected), Responses, Configuration, Advanced.
 - ANALYZE & DEBUG:** Conversation Logs, Statistics, Insights, Training.
- Main Workspace:** A flowchart titled SCOSAAR. The flow starts with a 'Welcome' node, followed by 'Issue too short', 'Issue sufficiently long', and 'admitere'. The 'admitere' node is currently selected and highlighted in blue. A tooltip is visible over this node, containing the text: "Conform regulamentului de organizare a studiilor uni...". The flow continues to 'admitere : 15', 'Ask if problem was solved', 'Problem solved', 'Problem not solved', and 'Transfer chat to support'. The flow ends with 'admitere : 16' and 'admitere : 17'.
- Right Sidebar:** 'Graph view' and 'Search' buttons.

Demo chatbot SCOSAAR

- <https://acad.ro/institutia/scosaar.html>

Articole publicate

- Ion, Radu and Avram, Andrei-Marius and Păiș, Vasile and Mitrofan, Maria and Mititelu, Verginica Barbu and Irimia, Elena and Badea, Valentin. **An Open-Domain QA System for e-Governance**. In *Proceedings of the 5th International Conference on Computational Linguistics in Bulgaria (CLIB 2022)*. Department of Computational Linguistics, IBL -- BAS, Sofia, Bulgaria, pp. 105--112, September 2022
- Dimitra Anastasiou, Anders Ruge, Ion, Radu, Svetlana Segărceanu, George Suciu, Olivier Pedretti, Patrick Gratz, Hoorieh Afkari. **A Machine Translation-Powered Chatbot for Public Administration**. In *Proceedings of the 23rd Annual Conference of the European Association for Machine Translation*, pp. 327-328
- Dimitra Anastasiou, Ion, Radu, Valentin Badea, Olivier Pedretti, Patrick Gratz, Hoorieh Afkari, Valerie Maquil, Anders Ruge. **ENRICH4ALL: A first Luxembourgish BERT Model for a Multilingual Chatbot**. In *Proceedings of SIGUL 2022@LREC2022, pages 207–212, Marseille, 24-25 June 2022*

Concluzii

- Proiectul Enrich4All s-a încheiat cu succes, prin instalarea chatbot-ului SupWiz pe site-ul SCOSAAR
- Instituțiile publice nu sunt dispuse să investească în digitalizarea interacțiunii lor cu cetățenii
 - Din zeci de cereri de colaborare trimise (în România și Luxemburg) doar două instituții au acceptat colaborarea, cu condiția ca ele să nu fie implicate deloc în procesul de dezvoltare
 - Există temeri legate de bugetul alocat, pierderea unor slujbe, dificultatea menținerii sistemelor în funcțiune
- Avansul modelelor de limbă gigant (LLMs), de ex. GPT-4 și ChatGPT, va face ca abordarea proiectului Enrich4All să fie depășită