



Workshop 8:

Inovare la intersecția dintre Prelucrarea Limbajului Natural și Digital
Humanities: aplicabilitate pentru studii de limbă și literatură română

Coordonator UVT:

CS II Dr. habil. Mădălina Chitez

Coordonator din diaspora:

Dr. Vionica Pătrăucean
Prof. Univ. dr. Ioana Galleron

Coordonator național:

Conf. dr. Traian Rebedea
CS I Dr. Roxana Patraș

Concluzii Workshop 8 (1)

A. Concluzii privind domeniile de cercetare abordate: NLP (*Natural Language Processing*) și DH (*Digital Humanities*)

✓ Limba română este catalogată drept *lesser used language* sau chiar *low-resource language* atunci când este vorba de studii de Inteligență Artificială. De aceea, **valoarea patrimonială a resurselor digitale de limbă, literatură sau cultură românească** este incontestabilă. La workshop s-au prezentat exemple de astfel de resurse: corpusuri reprezentative (CoRoLa), corpusuri specializate (ROGER, EXPRES, LEMI, HAIRO, ELTEC, POPLITE, RoDia), arhive digitale de literatură (MDRR), corpus de ediții digitale (re-editări ale operei lui Mihai Eminescu), instrumente digitale bazate pe corpus (CoToHiLi – lingvistică istorică; ReaderBench – readability; RELATE – platformă de procesare date de corpus în limba română; ComparaQ - platformă de procesare corpusuri multilingve). A fost exprimată dorința de a se identifica **finanțare pentru evenimente dedicate acestor tematici** pentru a stimula cooperarea națională și internațională a specialiștilor din **domeniile IT și AI, Filologie, Psihologie, Educație, Legislație**, interesați să creeze, acceseze, analizeze și să dezvolte aplicații pe baza acestor resurse, precum și noi resurse.

✓ Reflecțiile participanților au avut în vedere, ca teme centrale sau conexe, foarte **recenta evoluție a interfeței Chat GPT3** și posibilitățile sale de utilizare pentru cercetare.

✓ A fost remarcat nivelul scăzut al resurselor de limbă română disponibile cercetării comparativ cu alte limbi (de exemplu, seturi de date NER 2, NLI). Soluțiile identificate împreună cu colegii din diaspora au fost: **(a) crearea de repozitorii digitale de date lingvistice, cu generare de DOI / ISBN** (beneficii de copyright), găzduite și securizate pe servere bugetate de Ministerul Cercetării, Inovării și Digitalizării; **(b)** asigurarea de **finanțări dedicate colectării de date lingvistice** și **consultanței juridice** privind colectarea și folosirea acestora pentru cercetare-dezvoltare; **(c)** asigurarea de **linii de finanțare dedicate direcțiilor emergente din științele umaniste** (de ex. Științele Umaniste Digitale), care sunt, din păcate, sub-finanțate în comparație cu alte discipline.

✓ S-a observat lipsa datelor reprezentative pentru limba română din infrastructuri europene de date lingvistice, cum ar fi **CLARIN**. Un motiv ar fi **ne-plata cotizației de participare din partea României**. De exemplu, din acest motiv, corpusul recent **ROGER**, a fost «găzduit» de repozitoriul sloven.

Workshop 8:

Inovare la intersecția dintre Prelucrarea Limbajului Natural și Digital
Humanities: aplicabilitate pentru studii de limbă și literatură română

Coordonator UVT:

CS II Dr. habil. Mădălina Chitez

Coordonator din diaspora:

Dr. Vionica Pătrăucean
Prof. Univ. dr. Ioana Galleron

Coordonator național:

Conf. dr. Traian Rebedea
CS I Dr. Roxana Patraș

Concluzii Workshop 8 (2)

B. Concluzii privind stimularea cercetării din România, în general, și stimularea colaborării cu colegii din diaspora, în special:

- ✓ Crearea unui cadru legislativ (simplu și ușor de implementat) care să **faciliteze mobilitatea cercetătorilor din și spre mediile academice din diaspora**. Cel existent, care vizează mobilități pe termen scurt ale experților din diaspora în România, ar putea fi extins și adaptat și pe direcția de mobilitate din România spre colegi români din diaspora.
- ✓ Crearea unui cadru legislativ (simplu și ușor de implementat) privind **Achizițiile din domeniul cercetării** (de exemplu, licențe software).
- ✓ Creșterea **atractivității mediului de cercetare din România** prin deschiderea de poziții *non entry-level* pe perioadă nedeterminată pentru specialiști din diaspora, sau poziții pe perioadă nedeterminată pentru cercetători tineri sau cu experiență, din țară, prin măsuri ca: (a) **burse de (re)integrare urmate de angajare**; (b) adaptarea cadrului legislativ prin care **scutirea pe impozitul pe venit de 10%** din domeniul cercetării să nu mai fie acordată exclusiv contractelor part-time din proiectele de cercetare ci și normei de bază a angajaților cu funcții de cercetare (este absurd ca același cercetător, cu aceeași fișă a postului și aceleași sarcini de cercetare, care are scutire de la plata impozitului pe venit pentru salariul din proiect să nu beneficieze de această scutire pentru salariul de bază; acest lucru este valid pentru cei din universități, Institute de Cercetare și Academia Română); (c) **creșterea salariilor pentru pozițiile de cercetare din mediul academic** (evident, corelate cu atingerea de ținte academice).