

**CREAREA, GESTIUNEA ȘI PARTAJAREA DATELOR
LINGVISTICE: PRACTICI EXISTENTE ȘI PROVOCĂRI ÎN
ROMÂNIA
VALOAREA DATELOR PENTRU DEZVOLTAREA DE
TEHNOLOGII LINGVISTICE DE TOP**
Panel



IMPORTANȚA DATELOR

- Datele reprezintă nutrientul vital pentru dezvoltarea economică:
 - Bază pentru multe produse și servicii noi
 - Generează productivitate și eficientizează resursele în toate sectoarele economiei
 - Permit crearea de produse și servicii personalizate
 - Susțin politici mai bune și îmbunătățirea serviciilor guvernamentale
- Viziunea Europei:
 - Creează un mediu atractiv pentru politici astfel încât, până în 2030, contribuția UE la economia bazată pe date – date stocate, prelucrate și valorificate superior în Europe – să fie cel puțin la nivelul ponderii sale economice.
 - Creează un spațiu integrat de date european

(Source: https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf)

DATE PENTRU TEHNOLOGIA LIMBAJULUI

- TL (și IA în general) necesită date care implică codifică ceea ce vrem să obținem
- Datele lingvistice sunt un tip special de date: limbajul este generat tot timpul, dar pentru a antrena sisteme de IA cu date lingvistice avem nevoie de date **digitale și procesabile**
 - în cantități mare – cu cât mai multe, cu atât mai bine
 - În diferite limbi, domenii, genuri ..., dependent de problema pe care dorim să o rezolvăm
- Acest lucru s-a dovedit și se dovedește în continuare o activitate anevoioasă, cu numeroase obstacole...

Panelul va încerca să răspundă la aceste întrebări cu ajutorul specialiștilor care au dezvoltat resurse textuale și orale pentru limba română, servicii și instrumente.

MODERATOR CORNELIU BURILEANU



Profesor,
Președinte al
Institutului de
Cercetare
CAMPUS, UPB

LIVIU PETRIȘOR DINU



Profesor, Șeful
colectivului NLP de
cercetare la
Universitatea din
București

DAN CRISTEA



Profesor, Șeful
colectivului NLP de
cercetare la
Institutul de
Informatică
Teoretică, Iași

MIRCEA GIURGIU



Profesor, Șeful
colectivului de
cercetare
tehnologia vorbirii,
Universitatea
Tehnică, Cluj-
Napoca

VERGINICA BARBU MITITELU



Cercetător
principal, Șefa
colectivului
Resurse lingvistice,
ICIA

ÎNTREBĂRI LA CARE ÎNCERCĂM SĂ RĂSPUNDEM ÎN PANEL

- Care este cadrul pentru distribuția de resurse lingvistice în România?
- Sunt datele lingvistice create, gestionate și partajate într-un mod organizat?
- Care sunt, în opinia fiecăruia dintre dv., cele mai mari provocări în partajarea datelor lingvistice în România ?
- Care sunt obstacolele? Politici pentru tehnologia limbajului? Lipsa finanțării? Lipsa infrastructurii? Lipsa expertizei/personalului calificat?
- Din punctul dv. De vedere, ce ar fi nevoie să fie făcut pentru a îmbunătăți situația și a facilita partajarea datelor lingvistice în România ?

CONCLUZII: DATELE...

- ... implicit codifică ceea ce dorim să realizăm
- ... trebuie să fie alese conform cu necesitățile problemei de rezolvat
- ... sunt necesare în cantități foarte mari – cu cât mai multe, cu atât mai bine
- ... câteodată sunt greu de găsit și obținut



MUȚUMIM PENTRU ATENȚIE!

Website: www.lr-coordination.eu

Twitter: @LR_Coordination

Facebook: www.facebook.com/EuropeanLanguageResourceCoordination

LinkedIn: www.linkedin.com/in/lrcoordination

Email: info@lr-coordination.eu

