

eDTLR – Dicționarul Tezaur al Limbii Române în format electronic

Numar contract/Data: eDTLR 910013/18.09.2007

Program: PNCDI II, programul 4 *Parteneriate în domeniile prioritare*

Tip proiect: PC - Proiect Complex

Durata contractului: 18 septembrie 2007 - 17 septembrie 2010

Finanțator: Ministerul Educației, Cercetării și Tineretului ([MECT](#)), prin Centrul Național de Management Programe ([CNMP](#))

Coordonator de proiect: Universitatea "Alexandru Ioan Cuza" Iași ([UAIC](#)), Facultatea de Informatică ([FII](#))

Director de proiect: prof. dr. [Dan Cristea](#)

Rolul IIT: partener

Responsabil proiect din partea IIT: Prof.Dr. Horia-Nicolai Teodorescu

Colectiv:

Nume	Funcția în cadrul proiectului
Horia-Nicolai Teodorescu	Responsabil proiect
Neculai Curteanu	Responsabil științific
Vasile Apopei	Colaborator intern
Cecilia Bolea	Colaborator intern
Cristina Niță	Colaborator intern
Diana Maria Trandabăț	Colaborator intern
Alex Moruz	Colaborator intern
Iustin Dornescu	Colaborator intern
Gabriela Pavel	Colaborator extern
Cristina Popovici	Colaborator extern
Cristina Mihaela Vereștiuc	Colaborator extern

Obiective proiect:

eDTLR va include patru tipuri de activități definitorii:

1. elaborarea de tehnici informatice de digitizare a Dicționarului, de corectare a conținutului său și de indexare a citatelor în sursele bibliografice – ce vor fi elaborate de partenerii informaticieni;

2. activități de digitizare a Dicționarului propriu-zis și a surselor bibliografice – efectuate de asemenea de parteneri informaticieni;

3. activități lexicografice de corectură a conținutului și formatului electronic al Dicționarului și stabilirea listei surselor bibliografice de scanat – efectuate de lexicografi;

4. activități de exploatare și diseminare a metodelor create și a produsului obținut – efectuate de toți partenerii.

Descriere proiect:

În contextul actual și de perspectivă al societății informaționale, se impune stringent existența formatului digital al lucrărilor fundamentale ale unei culturi naționale.

Proiectul pornește de la forma tipărită a Dicționarului tezaur al limbii române (conținând seriile vechi și nouă ale Dicționarului limbii române), ale cărui volume au început să fie tipărite începând cu 1913 sub egida Academiei Române, și care, conform planului Academiei, urmează a fi terminat până la sfârșitul acestui an. Proiectul urmărește, ca prim obiectiv, transpunerea în format electronic a celor cca 13.000 de pagini ale Dicționarului (doar pentru ultimele volume tipărite existând o formă culeasă pe calculator), lucru ce va permite consultări complexe ale acestuia, precum și continuarea activităților de editare și actualizare (nu numai că volumele seriei vechi, tipărite înainte de război, trebuie reconsiderate pentru actualizarea lexicului și a definițiilor, dar Dicționarul trebuie să permită actualizări continue, sincrone evoluției limbii). Pentru ca formatul electronic al Dicționarului să devină unul dintre principalele instrumente de lucru în cercetarea lingvistică academică românească, al doilea obiectiv al proiectului este constituirea unei colecții de texte, în formă scanată și text, cuprinse în lista surselor bibliografice ale celor două serii ale Dicționarului. Aceste texte ar urma să fie indexate, în format imagine și text, în vederea accesării și fișării automate.

Proiectul pune bazele cooperării interinstituționale în domeniul cercetării fundamentale în domeniul lexicografiei computaționale dintre Academia Romană și instituțiile de învățământ superior. Colaborarea interdisciplinară dintre partenerii proiectului, recunoscuți în societatea academică românească și internațională, certifică concretizarea obiectivelor propuse la un standard de calitate competitiv la nivel mondial în domeniu. eDTLR reunește un consorțiu format din informaticieni, specialiști în lingvistică computațională și prelucrarea limbajului natural, și lingviști lexicografi, ce aparțin mediilor academice și universitare din București, Iași și Cluj-Napoca.

Activitățile inovatoare ale proiectului sunt preponderent de natură informatică, ceea ce explică coordonarea de către o instituție cu profil informatic. Pe de altă parte, rezultatele proiectului sunt esențialmente de natură socio-umanistă, cu o extrem de înaltă audiență la public, ceea ce explică încadrarea lui în domeniul socio-uman. Marele Dicționar al limbii române, cu semnificație de tezaur al patrimoniului național, reprezintă nu numai o înregistrare a evoluției limbii române de la origini până în momentul actual, dar, datorită formatului lui electronic ce va fi elaborat în proiect, este potențat astfel încât să

urmărească prin mijloacele tehnologiilor informatice și ale cunoașterii, specifice secolului XXI, evoluția limbii române în contextul globalizării și a performanței comunicative.

Rezultatele proiectului vor fi făcute publice pe Internet și diseminate atât în conferințe naționale și internaționale, cât și în școli de vară. Toate drepturile intelectuale și de autor asupra monumentalei lucrări aparțin Academiei Române.

Lucrari publicate

Pana in prezent, colectivul IIT a publicat urmatoarele lucrari aferente temei grantului:

1. Curteanu, N., Moruz, A., **Trandabăț, D.** (2008) *Extracting Sense Trees from the Romanian Thesaurus by Sense Segmentation & Dependency Parsing*, Proceedings of CogAlex Cognitive Aspects of the Lexicon: Enhancing the Structure, Indexes and Entry Points of Electronic Dictionaries, COLING 2008 Workshop endorsed by SIGLEX, pp. 55 - 63, ISBN 978-1-905593-56-9.
2. Curteanu, N., **Trandabăț, D.**, Moruz, A., Pavel, G., Verestiuc, C. (2008) *Dictionary Sense Segmentation & Dependency - A Flexible Strategy for Thesauri Parsing*, Proceedings of 5th European Conference on Intelligent Systems and Technologies ECIT 2008, CD edition.

Mai multe informatii pot fi gasite pe site-ul proiectului

https://consilr.info.uaic.ro/edtlr/wiki/index.php?title=Despre_proiect