

Alexandru Anghelina, Claudiu Teodorescu

DIGITALIZAREA *DICȚIONARULUI LIMBII ROMÂNE* PROBLEME. SOLUȚII. PERSPECTIVE

Digitalizarea *Dicționarului limbii române* (DLR) nu presupune doar dezvoltarea interfeței de redactare (Busuioc, Teodorescu 2015: 443), ci și crearea de instrumente auxiliare pentru redactori. În acest context, în cadrul Institutului de Lingvistică al Academiei Române „Iorgu Iordan – Alexandru Rosetti” a fost dezvoltată platforma **soliorom.ro** pentru a veni în sprijinul redactorilor dicționarului academic.

Platforma dispune de mai multe instrumente digitale, cele mai importante în opinia noastră fiind **corpusul de citate**, **siglarul digital** și **interfața de editare a bibliografiei**.

O asemenea abordare are avantaje multiple:

- informația poate fi diseminată cu rapiditate către toți redactorii;
- orice modificare adusă anumitor componente se reflectă în interfața de redactare (avem aici în vedere modificări făcute în cadrul bibliografiei);
- se pot valorifica resursele comune (corpusul de citate).

În momentul de față, sistemul DLRI se bazează pe mai multe tehnologii și instrumente software:

- Oxygen XML Author – pentru redactarea intrărilor;
- standardul TEI – pentru codarea informației lexicografice;
- existDB – pentru stocarea datelor;
- diverse componente web refolosibile – pentru platforma soliorom.ro.

Primul aspect pe care dorim să îl prezentăm are în vedere **atestările** cuvintelor din DLR, elemente esențiale în redactarea dicționarului. Pentru valorificarea tuturor resurselor disponibile – excerptarea de citate fiind făcută în toate cele trei institute de lingvistică (din București, Cluj și Iași) responsabile de redactarea dicționarului –, soliorom.ro dispune de o secțiune dedicată, unde se pot adăuga, edita sau căuta contextele în care este atestat un anumit cuvânt.

Segmentarea fiecărei atestări și adăugarea ei în baza de date poate să se facă uneori imperfect. De aceea, a fost dezvoltată componenta de editare a atestărilor care permite marcarea strictă a părților componente: cuvânt-titlu, categorie gramaticală, citat, sursă, localizare în sursă sau diverse note ori comentarii.

LR, LXIX, nr. 3–4, p. 491–496, București, 2020



ACADEMIA ROMÂNĂ

Secția de filologie și literatură

Ontologie
Corpus de dicționare
Corpus de texte
Corpus de citate
Bibliografia filologică românească
BSL
Terminologii
Colțul cercetătorului

Corpus de citate



Citate

Căutarea se face cu diacritice.

Dacă se apasă pe butonul din stânga citatului, acesta va fi copiat automat și va putea fi inserat în documentul word în care lucrați.

Deocamdată, baza de date cuprinde doar citate pentru cuvintele-titlu de la litera A. Pe măsură ce vor fi disponibile citate pentru alte cuvinte-titlu, vor fi integrate și acelea.

Nr. crt.	Cuvânt-titlu	Citat	Sursă
1	AICI	<input type="checkbox"/> var. aice	Cu cinste alăturz aice tîlmăcire acestui izvodu(a. 1812). SAVA, DOC. II, 138.
2	AICI Var. AICEA Eu sănt aicea, mă pot afla Fără pre altul a întreba. ZILLOT, CRON. 62	<input type="checkbox"/>	
3	AICI	<input type="checkbox"/> Ei au căutat în multe rînduri ajutoariți de la Spama, dar aici nu este vorba decât de vestitul conte D Egnon.	POGOR. HENR. XI/19
4	AICI adv.	<input type="checkbox"/> ACI	Sunt de vreo două săptămăni, întors de pe front unde căt p-aci era să vin să te văd.VIANU, C. în Ő. XIV, 29[1918]
5	AICI	<input type="checkbox"/> Aici stă administratorul – căpitan Ghețu. Pensionarul numai locuiește aici.	CIPRIAN, A. 54 [1958]
6	AICI adv: Eu aci, mândra la munte, Două inimioare frănte. FOLC. DOBR. 475.ver.	<input type="checkbox"/>	
7	AICI Var. AICEA	<input type="checkbox"/> Zice Domul cătră Sellum, fiul lui Iosue, împăratul Iudei, cel ce au împărțit pentru Iosue, fratele său, cel ce au ieșit din locul acesta: „Nu să va mai întoarce aicea, Ci în locul la care l-am mutat”.	BIBLIA VULGATA IV, 188 [a. 1760-1761]
8	AICI	<input type="checkbox"/> Prea înălțate Doamne, ca acest regulament să se publice sub formă de ordonanță domnească, am onoarea a vă supune cu respect la semnătura Măriei Voastre anexatul aci proiect de decret.	BOLINTINEANU, C. 55 [1863]
9	AICI	<input type="checkbox"/> var. ACI (pop.)	Căt pentru modul predării și aci este, ca pretutindea, unul și același totuleana [1855] D. PAVEL, E.A.R. VII/10
10	AICI var. AIS	<input type="checkbox"/> Nu ci ceme, cin nu m-a mincat pri mine pri pădure, nu ce manincă	iei pră cine ais-în sat SANDU-TIMOC, POV. 364[1970]
11	AICI adv. Noi aici numai de vărutea cea creștinească vom grăi. MICU, T. I, 70/21.	<input type="checkbox"/>	

În condițiile în care informațiile din fișele fizice pot să difere cantitativ, a fost adoptată o soluție de multiplicare a anumitor câmpuri (atestare și note). În acest fel, se poate reține toată informația din fișe.

Al doilea aspect la care dorim să facem referire este **siglarul digital**, un instrument necesar pentru corelarea siglelor folosite în diverse perioade de redactare a dicționarului.

În ceea ce privește siglele (nume convenționale date lucrărilor din bibliografia dicționarului) se pot distinge trei faze:

1. siglele utilizate în DA, elaborat sub conducerea lui Sextil Pușcariu;
2. siglele utilizate în DLR, serie nouă;
3. siglele utilizate în DLRi, varianta actuală a dicționarului.

Astfel, platforma permite redactorului identificarea și transpunerea rapidă a siglelor folosite fie în DA (siglă veche), fie în DLR (siglă nouă) în forma actuală (siglă DLRi). În același timp, siglarul oferă informații referitoare la data publicării sau redactării textului din care este excerptată o atestare (o inovație prezentă în ultima formă a dicționarului), la ediția de lucru și la localizarea textului din cadrul ediției respective.

Siglar digital

Clătarea se face cu diacritice.

Se pot face căutări după știri de litere localizate oriunde în interiorul unei sigle, indiferent că este veche, din DLR sau din DLRI.

Rezultatele vor fi afișate în două tabele: primul cuprinde corespondențele între siglele vechi și cele din DLR, iar al doilea cuprinde corespondențele între siglele DLR și sigle DLRI, respectiv, care au la început șirul de litere cu care s-a început căutarea.

Dacă se apasă pe butonul din stânga atestării în format DLRI, aceasta va fi copiată automat și va putea fi inserată în documentul word în care lucrați, cu tot cu formatare (îngroșat, drept, înclinat, după caz).

Asachi, F

Sigle vechi - sigle DLR

Nr. crt.	Sigla veche	Sigla DLR
1	Asachii, F.	ASACHI, FABULE 99

Sigle DLR - sigle DLRI

Nr. crt.	Sigla DLR	Sigla DLRI	Ediție de lucru + editor	An ediție lucru	Număr de pagini	Identificator
1	ASACHI, F. ASACHI, FABULE ASACHI, O. I ASACHI, S. L. I	■ 1819-1830; 1836; 1844; 1862 ASACHI, Fab., I	Opere, I	1973	p. 247-300	■ #unid-2156eb39-756a-7ac1-53a6-b2ab0bd38e95
2	ASACHI, F. ASACHI, FABULE ASACHI, O. II ASACHI, S. L. II	■ 1839-1862; 1844; 1862 ASACHI, Fab., II	Opere, I	1973	p. 301-361	■ #unid-4ad8f0d1-401e-40be-38a3-cba4be9a032e

Al treilea și ultimul aspect pe care dorim să îl prezentăm are în vedere modulul dedicat administrării **bibliografiei**.

În condițiile în care orice modificare la nivelul resursei bibliografice are reflexe în interfața de redactare a dicționarului – modificarea sau adăugarea unei sigle trebuie să fie reflectată și în formularul de redactare din Oxygen XML Author–interfața de editare permite actualizarea rapidă a tuturor informațiilor necesare.

The screenshot shows a web-based bibliographic interface. On the left, a search results list is visible, with the top entry being 'Lestajul Țării Moldovei de la domnia întâi și până la a patru domnie a lui Constantin Mavrocordat (1713-1774)'. The main area on the right is a form for editing or viewing a record. It contains several sections: 'Sigla DLRI' with a dropdown menu set to 'Cond. Mavrocordat, II'; 'Titlu' with a text input field containing 'Condiția lui Constantin Mavrocordat'; 'editat' with a dropdown menu set to 'Bibun'; 'prefațat' with a dropdown menu set to 'Bibun'; 'ilustrat de glosar' with a dropdown menu set to 'Bibun'; 'Cronă lucrării' with a dropdown menu set to 'text juridic'; 'Volume' with a text input field; 'Ediție de lucru' with a text input field; and 'Nr. de pagini' with a text input field containing '777 p.'.

The screenshot shows a form for citation and source information. At the top, there is a dropdown menu for 'Atestare:' set to 'Autor'. Below it, the 'Citat:' field contains the text 'Să rămâne moția de istov a abăgerilor, să o vândă, să-și scoată banii.' The 'Izvor:' field is set to 'Cond. Mavrocordat, II' and 'mavroc'. The 'Căutare' field is set to 'Data: 1742'. The 'Localizare în sursă:' field is set to 'T²' and '195'.

În același timp, siglarul și bibliografia sunt interconectate, evitându-se astfel editarea manuală a datelor în mai multe fișiere. Orice editare realizată în cadrul bibliografiei se va reflecta automat și în siglar (cu condiția ca aceasta să fie relevantă pentru ambele documente).

Dezvoltarea resurselor complementare interfeței de redactare poate eficientiza, în opinia noastră, efortul de redactare prin partajarea tuturor resurselor necesare. Deși procesul este încă în desfășurare, ne propunem realizarea unui sistem de trei servere de date XML sincronizate, care să deservească cele trei centre unde se redactează *Dicționarul limbii române* informatizat.

BIBLIOGRAFIE

Busuioc, Monica, Claudiu Teodorescu, 2015, *Informatizarea „Dicționarului limbii române” – un vis pe cale de realizare*, în Sala, Marius, Maria Stanciu Istrate, Nicoleta Petuhov (ed.), 2015, p. 441–445.

- DA = *Dicționarul limbii române*. [DA] [Sub conducerea lui Sextil Pușcariu]. Tomul I. Partea I: A–B. Academia Română. Întocmit și publicat după îndemnul și cu cheltuiala Maiestetii Sale Regelui Carol I. București, Librăriile Socec & Comp. și C. Sfetea, 1913, LXX + 716 p.; Tomul I. Partea II: C. Academia Română. Întocmit și publicat după îndemnul și cu cheltuiala Maiestetii Sale Regelui Carol I. București, Tipografia Ziarului „Universul”, 1940, [2] + 1 066 p.; Tomul I. Partea III. Fascicula I: D–De. Academia Republicii Populare Române. București, Universul Intreprindere Industrială a Statului, 1949, 80 p.; [Fascicula II: De–Desțina; șpalt, 1948]; Tomul II. Partea I: F–I. [de fapt: I/Î]. Academia Română. Întocmit și publicat după îndemnul Maiestetii Sale Regelui Carol I. București, Monitorul Oficial și Imprimeriile Statului. Imprimeria Națională, 1934, [2] + 955 p.; Tomul II. Partea II. Fascicula I: J–Lacustru. Academia Română. Întocmit și publicat după îndemnul Maiestetii Sale Regelui Carol I... București, Tipografia Ziarului „Universul” S. A., 1937, 66 p.; Tomul II. Partea II. Fascicula II: Ladă–Lepăda. Academia Română. Întocmit și publicat după îndemnul Regelui Carol I. București, Tipografia Ziarului „Universul” S. A., 1940, 160 p.; Tomul II. Partea II. Fascicula III: Lepăda–Lojniță. Academia Republicii Populare Române. București, Tipografia Ziarului „Universul” S. A., 1948, 50 p.
- DLR = *Dicționarul limbii române*. (DLR). Serie nouă. (I₃₋₈, III–V, VI–XIV). Academia Republicii Populare Române (VI₁₋₂), Academia Republicii Socialiste România (VI₃–X₂, XI₁₋₃), Academia Română (X₃–X₅, XII–XIV, I₃₋₈, III–V). [Redactori responsabili: acad. Iorgu Iordan, acad. Alexandru Graur și acad. Ion Coteanu. Din anul 2000, redactori responsabili: acad. Marius Sala și acad. Gheorghe Mihăilă]. București, Editura Academiei Republicii Populare Române (VI₁₋₂); Editura Academiei Republicii Socialiste România (VI₃–XI); Editura Academiei Române (XII–XIV, I₃₋₈, III–V). Tomul VI. Litera M: 1965–1968, CXXXVI + 1 076 p.; Tomul VII. Partea 1. Litera N: 1971, 584 p.; Tomul VII. Partea a 2-a. Litera O: 1969, 400 p.; Tomul VIII. Litera P: Partea 1: P – Păzui, 1972, VIII + 357 p.; Partea a 2-a: Pe–Pînar, 1974, p. 359–693; Partea a 3-a: Pînă–Pogribanie, 1977, p. 695–948; Partea a 4-a: Pogrijenie–Presimțire, 1980, p. 949–1 341; Partea a 5-a: Presin–Puzzolană, 1978, p. 1 343–1 866. Tomul IX. Litera R: 1975, 625 p.; Tomul X. Litera S: 1986–1994. Partea 1. S–Sclabuc: 1986, 388 p.; Partea a 2-a. Scladă–Semînțarie: 1987, p. 389–690; Partea a 3-a. Semn–Sîveică: 1990, p. 691–1 040; Partea a 4-a. Slab–Spongios, 1992, p. 1 039–1 412; Partea a 5-a. Spongiar–Swing: 1994, p. 1 413–2 147; Tomul XI. Partea 1. Litera Ș: 1978, 271 p.; Tomul XI. Partea a 2-a: Litera T. T–Tocăliță: 1982, 376 p. Partea a 3-a. Litera T: Tocăna–Twist, 1983, p. 377–763; Tomul XII. Partea 1. Litera Ț: 1994, 240 p.; Tomul XII. Partea a 2-a. Litera U: 2002, 468 p.; Tomul XIII. Partea 1. Litera V. V–Veni: 1997, 325 p.; Partea a 2-a: Litera V. Venial–Vizurină: 2002, p. 327–751; Partea a 3-a. Litera V. Vîclă–Vuzum. Literele W, X, Y, 2005, p. 753–1 340; Tomul XIV. Litera Z: 2000, 409 p.; Tomul I. Partea a 3-a. Litera D. D – Deînmulțit: 2006, CXLVI + 342 p.; Tomul I. Partea a 4-a. Litera D. Deja – Deținere: 2006, p. 343–780; Tomul I. Partea a 5-a. Litera D. Deținut–Discopotiriu: 2007, p. 781 – p. 1 167; Tomul I. Partea a 6-a. Litera D. Discord–Dyke: 2009, p. 1 169 – 1 624; Tomul IV. Litera L. L–Lherzolită: 2008, 778 p.; Tomul V. Litera L. Li–Luzulă: 2008, 815 p.; Tomul I. Partea a 7-a. Litera E. E–Erzaț: 2009, 388 p.; Tomul I. Partea a 8-a. Litera E. Es–Ezredeș: 2010, 361 p.; Tomul III. Literele J, K, Q: 2010, [3] + 98 p.

Sala, Marius, Maria Stanciu Istrate, Nicoleta Petuhov (ed.), 2015, *Lucrările celui de-al cincilea Simpozion Internațional de Lingvistică, București, 27–28 septembrie 2013*, București, Editura Univers Enciclopedic Gold.

*

Text Encoding Initiative (<http://tei-c.org>)

THE DIGITIZATION OF THE ROMANIAN LANGUAGE DICTIONARY
CHALLENGES. SOLUTIONS. PERSPECTIVES

(Abstract)

The paper aims to describe some of the auxiliary instruments developed for the drafting of the digitized version of the *Romanian Language Dictionary* (DLRi) and to provide some insights regarding the decisions made.

Keywords: soliom.ro, digital dictionary, Oxygen XML Author, TEI

Alexandru Dan ANGHELINA
Institutul de Lingvistică al Academiei Române
„Iorgu Iordan – Alexandru Rosetti”
București, Calea 13 Septembrie, nr. 13

Claudius TEODORESCU
Institutul de Filologie Română
al Academiei Române
„Alexandru Philippide”
Iași, str. Th. Codrescu, nr. 9