

INFORMATIZAREA DICȚIONARULUI LIMBII ROMÂNE – UN VIS PE CALE DE REALIZARE

THE CYBERNATION OF THE DICTIONARY OF ROMANIAN LANGUAGE – A DREAM TURNING TO REALITY

(Abstract)

The article presents the general framework of the project of cybernation of the *Dictionary of Romanian Language*. The decision of the Board of the Romanian Academy to resume work on the first part of the Dictionary (DA) according to the norms of the new series (DLR) constituted a fitting opportunity to start work on writing the academic *Dictionary* in an XML format. The digitalization of the dictionary is accomplished with the help of the Oxygen program put forward by a computer specialist who achieved the work interface. This will offer the possibility of the dictionary's being posted on the internet.

Cuvinte-cheie: dicționar, cibernetică, programul oxygen, interfață, câmpuri de redactare.

Key-words: dictionary, cybernation, oxygen program, interface, section of the writing formula.

Obiective

La Institutul de Lingvistică al Academiei Române „Iorgu Iordan–Al. Rosetti” urmează să se redacteze, în urma hotărârii Prezidiului Academiei Române, prima parte a *Dicționarului limbii române* după normele celei de a doua părți.

Dicționarul limbii române a constituit, alături de gramatică și ortografie, unul dintre cele trei deziderate istorice ale înființării Societății Academice Române, devenită ulterior Academia Română. El este cunoscut sub numele de *Dicționarul Academiei*, întrucât este elaborat sub egida acestei prestigioase instituții.

Dicționarul limbii române este un dicționar istoric, explicativ, etimologic și normativ și este denumit și dicționarul *tezaur*, grație mării sale bogății lexicale. Redactarea primei ediții a început în anul 1906 și a fost finalizată în anul 2010. Prima parte, elaborată între 1906 și 1949, cuprinde literele A–L, iar a doua parte, numită Serie nouă, literele M–Z, precum și literele neredactate sau nefinalizate (E, D și L). *Dicționarul limbii române* a fost publicat, între 1913 și 2010, într-un număr de 37 de volume, însumând peste 175.000 de cuvinte și variante în cca 18.000 de pagini.

Cu sprijinul Băncii Naționale a României, în anul 2010, a fost retipărită o ediție anastatică a întregului dicționar, în 19 tomuri, în condiții grafice excelente.

O tehnologie nouă. O structură nouă

În noul context al societății informaționale, se impune cu necesitate ca noua ediție a *Dicționarului limbii române* să fie redactată în format XML, pentru a putea oferi posibilitatea consultării lui pe internet și a intra astfel în circuitul valorilor mondiale.

Digitalizarea noii ediții a *Dicționarului* se va face în cadrul unui proiect de anvergură cu ajutorul programului Oxygen, produs în România.

Institutul a luat legătura cu compania producătoare a programului Oxygen, un program de editare XML, bine cunoscut și foarte apreciat, cu scopul de a îndeplini obiectivul convertirii formatului nestructurat al dicționarului într-un format structurat.

În cursul acestei operațiuni, institutul a hotărât adoptarea unor standarde tehnice larg răspândite și acceptate, bazate pe XML, respectiv TEI (Text Encoding Initiative = Inițiativa de codare a textelor), adecvate pentru redactarea dicționarului.

Instrumente noi

Programul Oxygen permite redactorilor să utilizeze abordări accesibile pentru utilizator în ceea ce privește editarea documentelor în format XML, respectiv folosirea de interfețe similare formularelor web.

Această modalitate de editare este extrem de utilă și de facilă, întrucât redactorii pot astfel să adauge conținut folosind un formular și nu un arbore XML, a cărui structură ar trebui să o cunoască, așa cum este cazul pentru majoritatea editoarelor XML existente.

Perioada de învățare a modului în care se realizează redactarea dicționarului este foarte scurtă, interfața este ușor de utilizat și folosirea sa a întrunit consensul tuturor redactorilor care îl vor utiliza.

Această abordare ne-a permis să satisfacem chiar și cele mai sofisticate doleanțe, prin crearea unui cadru specializat pentru conținutul dorit al *Dicționarului*, un cadru care conține module de redactare, fișiere de adaptare a interfeței utilizatorului și clase Java pentru funcții specializate. Dintre aceste

doleanțe enumerăm: necesitatea de căutare și consultare a serverului dicționarului pentru interogări complexe privind conținutul în curs de redactare, adăugarea de fragmente de conținut și de șabloane de conținut în funcție de valorile selectate de utilizator din liste derulante, componente noi pentru formularul de redactare, altele decât cele oferite în mod obișnuit de programul Oxygen, redarea datelor în formatele dorite etc.

Redactare versus redare

Elaborarea formularului de redactare s-a făcut, pe cât a fost posibil, cu respectarea ordinii secțiunilor unei intrări în dicționar.

Cu toate că redactarea folosind acest formular nu este identică cu redactarea anterioară, care folosește editoare de text obișnuite, noi considerăm că formularul este un progres și un avantaj, din mai multe considerente, dintre care enumerăm următoarele:

1. conținutul rezultat este structurat din punct de vedere semantic, nu doar stilistic, așa cum este cazul când conținutul este redactat cu editoare de text;
2. conținutul structurat poate fi redat în formate multiple,: HTML (pagină web), PDF, EPUB (format de carte electronică) etc.;
3. structura conținutului poate fi oricând modificată, după necesități;
4. se pot face căutări complexe în conținut, lucru care nu este posibil în cazul conținutului nestructurat obținut cu editoare de text;
5. se pot înființa servicii care să furnizeze conținut în orice format dorit, pentru terți;
6. redactarea este ușurată prin folosirea de liste derulante care conțin valori predefinite, standardizate;
7. în cursul introducerii de conținut sau în cursul etapei de validare a conținutului, există posibilitatea verificării existenței unor corelații complexe, atât între diferitele elemente ale intrării curente, cât și între elemente ale intrării curente și elemente ale altor intrări din dicționar.

Prezentarea secțiunilor formularului de redactare

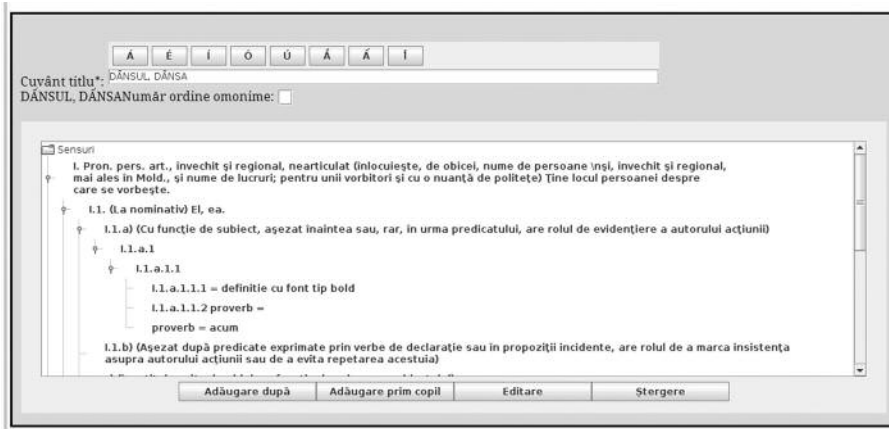
Secțiunea cu metadate ale intrării



Validare	Redare
Redactor*: <input type="text" value="Clim Marius"/>	
Revizor*: <input type="text"/>	
Creat: 2013-09-23T14:05:41+03:00	
Jurnal modificări	
Modificat: 2013-11-27T11:05:20+01:00	
Modificat: 2013-11-27T19:25:20+01:00	

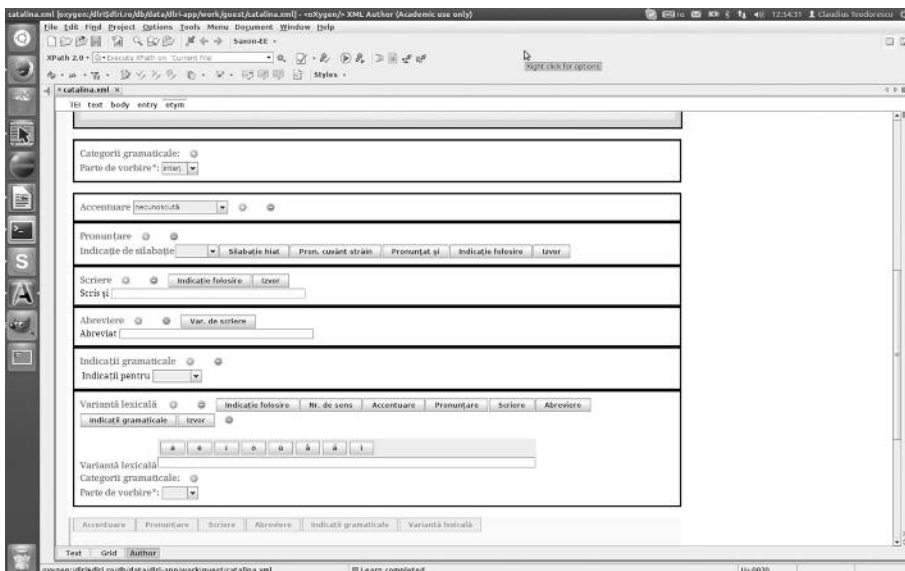
Această secțiune permite introducerea numelor redactorilor și a revizorilor care au contribuit la redactarea intrării și afișează data creării intrării, precum și un jurnal cu data și ora modificărilor intrării, validate prin apăsarea butonului Validare.

Secțiunea cu cuvântul-titlu și cu arborele de sensuri



Această secțiune permite introducerea cuvântului-titlu și a numărului de omonim, precum și redarea în formă sintetică a arborelui de sensuri ale intrării, redare care permite o viziune de ansamblu asupra sensurilor, precum și diferite operații asupra lor, cum ar fi adăugarea, ștergerea sau editarea de sensuri.

Secțiunea cu informații despre cuvântul-titlu



Această secțiune permite selectarea categoriei gramaticale a cuvântului-titlu și, după caz, introducerea de informații privind formele de accentuare, de pronunțare, de scriere, de abreviere, indicațiile gramaticale și variantele lexicale ale cuvântului-titlu.

Secțiunea etimologie

The screenshot shows a web interface for the 'Etimologie' (Etymology) section. At the top, there is a title 'Etimologie' with a small icon. Below it, there are several input fields and dropdown menus:

- 'Etimologie pentru' with a dropdown menu set to 'cuvântul titlu'.
- 'element extern' with a dropdown menu.
- 'Tip element extern' with a dropdown menu set to 'împrumut'.
- 'Probabilitate etimologie' with a dropdown menu.
- 'Tip împrumut' with a dropdown menu set to 'etimon sigur'.
- Four buttons labeled 'a', 'â', 'y', and 'ô'.
- 'Limba' with a dropdown menu set to 'gr.' and an 'etimon sigur' input field.

 The interface is clean and functional, designed for data entry and selection.

Această secțiune permite clasificare etimologiei cuvântului-titlu, precum și introducerea de alte informații etimologice.

Structura de clasificare etimologică este de departe cea mai complexă dintre toate secțiunile unei intrări, dar credem că s-a reușit obținerea unei interfețe „prietenoase”, prin folosirea de liste derulante cu tipuri de clasificări și prin introducerea de conținut specializat o dată cu selectarea unei intrări dintr-o asemenea listă.

Concluzii

Considerăm ca fiind excelentă inițiativa colectivului redacțional de a continua redactarea acestui dicționar într-un format structurat, folosind standarde consacrate în domeniu, și suntem siguri că în viitor multiple reușite vor fi posibile, printr-o cercetare inteligentă a datelor astfel structurate și prin corelarea acestor date cu datele structurate conținute în alte dicționare, fie românești, fie străine.

Monica BUSUIOC
 Institutul de Lingvistică al Academiei Române
 „Iorgu Iordan – Al. Rosetti”, București
 Claudiu TEODORESCU