

GesTE

UN INSTRUMENT DE GESTIONARE A TERMINOLOGIEI

Dorina CHIȘ

Universitatea „Tibiscus” din Timișoara

Valeriu Dorin CHIȘ

Universitatea „Politehnica” Timișoara

1. Apelul Uniunii Latine privind finanțarea creării unui software de gestionare a terminologiei

Uniunea Latină a lansat, în 24 iunie 2006, un apel al cărui obiect îl constituia crearea unui software „open source”, gratuit, de gestionare a terminologiei care să țină cont de standardele actuale în materie, ergonomic și eficient în manipularea informației. El urma să fie destinat în primul rând specialiștilor de limbă română, cu posibilitatea ca, pe termen lung, să fie tradus și în alte limbi. Pentru aceasta era necesar să se prevadă toate funcționalitățile multilingve.

Această inițiativă a Uniunii Latine a fost determinată de aprecierea că, pentru limba română, nu sunt disponibile la ora actuală instrumente fiabile, standardizate, care să permită gestionarea și armonizarea datelor și să fie accesibile tuturor celor implicați în munca terminologică. Pentru că nu există întotdeauna posibilități materiale de achiziționare a unor programe eficiente, dar costisitoare, numeroase terminologii se realizează utilizând aplicații ad-hoc care nu sunt compatibile cu cele folosite de alte grupuri de lucru.

Am răspuns acestui apel (la a doua lansare a lui) și ne-am angajat să realizăm gestionarul terminologic cerut.

2. Premise și abordare

În vederea elaborării structurii bazei de date terminologice am cercetat bibliografia accesibilă nouă în biblioteci și pe Internet precum și baze de date ale unor mari universități, centre și laboratoare de cercetare, instituții și organisme naționale și internaționale, dintre care menționăm câteva :

- Banque de données terminologique - Service de la langue française, de la Communauté française de Belgique, <http://www.cfwb.be/franca/bd/bd.htm>,
- Base de terminologie du CILF (Conseil international de la langue française) <http://www.cilf.org/bt.fr.html>,
- CRITER (Corpus du Réseau Interministériel de Terminologie), La délégation générale à la langue française et aux langues de France <http://www.culture.gouv.fr/culture/dgltf/terminologie/base-donnees.html>,
- EURODICAUTOM – actualmente IATE (Inter Active Terminology for Europe) - Bază de date a Comisiei europene <http://iate.europa.eu/iatediff/SearchByQueryLoad.do?method=load>,
- ILOTERM – Bază de date a Organizației internaționale a muncii, <http://www.ilo.org/iloterm/>
- IMF Terminology, Bază de date terminologice a fondului monetar internațional, http://www.imf.org/external/np/term/index.asp?index=fra&index_langid=2,
- Le Grand dictionnaire terminologique de l'Office québécois de la langue française, http://www.granddictionnaire.com/btml/fra/r_motclef/index1024_1.asp,
- TERMROM - Baza Centrului Național de Terminologie, <http://www.cfwb.be/franca/bd/bd.htm>,
- UNESCOTERM, Bază de date terminologice a UNESCO, <http://termweb.unesco.org/>
- UNTERM, United Nations Multilingual Terminology Database, <http://157.150.197.21/dgaacs/unterm.nsf...>

Am putut astfel să luăm cunoștință de modul în care sunt prezentate datele oferite și posibilitățile de informare puse la dispoziția utilizatorilor.

Pe de altă parte, la sediul din București al Uniunii Latine au avut loc mai multe întâlniri, mese rotunde, consacrate întocmirii fișelor terminologice și bazelor de date.

Am putut constata cu aceste prilejuri că, în funcție de instrumentele de lucru utilizate, de metodele abordate în realizarea terminologiilor și în funcție de

publicul destinat, structurile propuse erau diferite, chiar dacă mare parte din elemente se regăseau în mai toate variantele.

Această constatare ne-a determinat să pornim de la premisa că gestionarul nostru trebuie să fie un instrument flexibil, care să pună la dispoziție utilizatorului o structură la standarde internaționale, dar, în același timp, să permită fiecăruia, pentru fiecare caz în parte, să creeze o structură proprie, aceea pe care o consideră cea mai potrivită scopurilor propuse.

În ceea ce privește latura informatică, trebuiau alese instrumente care să asigure caracterul gratuit al gestionarului.

3. Caracteristicile gestionarului de terminologie GesTE

3.1. Elemente de programare

Gestionarul este compatibil cu sistemele de exploatare acceptate pe piața românească. Operațiile pe partea de server sunt efectuate de un script PHP care interoghează o bază de date MySQL. Bazată pe un software *open source*, aplicația poate fi utilizată *on line*, în sistem Intranet sau pe calculatorul propriu.

Atât datele referitoare la structura bazei terminologice (felul în care sunt organizate câmpurile și grupurile, listele și intrările lor) cât și datele terminologice sunt reținute într-o bază de date MySQL, care mai cuprinde și informațiile despre utilizatorii cu acces (restrâns sau total) la aplicație. Tabelul datelor terminologice nu are o structură fixă, fiind modificat de program în funcție de structurile care sunt create pe parcurs. Totuși, fiecare câmp al oricărei structuri trebuie asociat cu o coloană a acestui tabel.

Interogările MySQL se fac pe partea serverului de către un script PHP. Clientului îi sunt apoi trimise informațiile în format XHTML. Totuși, codul XHTML a fost separat de codul PHP, ceea ce îi asigură o mare flexibilitate și posibilitatea de a face modificări de la nivelul interfeței utilizator. Formatarea datelor se realizează cu ajutorul unor fișiere *.tpl* care pot fi editate pentru afișarea datelor în orice format fără a modifica algoritmiul programului.

Navigarea prin paginile programului se face prin parametrii GET trimiși fișierului index.php. Astfel, nu sunt dezvoltate căi către alte fișiere ale programului, totul realizându-se intern. De asemenea, parametrii folosiți pentru comenzi nu pot dezvălui în vreun fel structura bazei de date și nu permit atacuri de genul *MySQL Injection*.

La crearea unui utilizator nou, datele acestuia sunt înregistrate într-un tabel împreună cu un identificator. Referirea ulterioară la acel utilizator se face prin intermediul identicatorului creat la înregistrare. Parola folosită la autentificare nu este salvată în baza de date (nu poate fi recuperată). În schimb, programul înregistrează identicatorul (hash-ul md5) parolei și îl compară cu cele ale parolelor introduse la autentificările ulterioare. Astfel, un acces la baza de date MySQL care nu folosește interfața aplicației nu va dezvălui datele de autentificare ale utilizatorilor.

Aplicația permite exportul și importul datelor terminologice. Exportul presupune generarea unui fișier în format propriu (.gte) care conține rezultatul unei căutări. Pentru import e necesar fișierul .gte exportat anterior precum și o structură compatibilă cu datele importate. Atât importul cât și exportul permit selectarea câmpurilor intrărilor terminologice asupra cărora se va opera. Astfel, se poate exporta doar o listă de termeni, sau doar termenii și definițiile lor etc.

3.2. Standarde utilizate

Standardele folosite în elaborarea bazei de date GesTE au fost :

- ISO 1087 – concepte utilizate în terminologie,
- ISO 12618 – principii de creare a bazelor de date terminologice,
- ISO 12620 – categorii de date în terminologie,
- ISO 12200 sau MARTIF – *Machine-Readable Terminology Interchange Format* – principii de conversie și programare XML a băncilor de date terminologice.

3.3. Limbi de lucru

Baza de date oferă interfețe în limbile: catalană, franceză, italiană, portugheză, română și spaniolă. Traducerile textelor au fost asigurate de biroul din Paris al Direcției de terminologie și inginerie lingvistică (DTIL) al Uniunii latine.

Utilizatorul poate alege o limbă de lucru pe care o stăpânește mai bine, și va primi în această limbă toate informațiile legate de un termen (definiție, context, explicații...), chiar dacă ea diferă de cea căreia îi aparține termenul în cauză.

3.4. Structura bazei de date

Structura implicită a bazei de date include categorii de informații cuprinse în standardul ISO 12620. A fost introdus, în plus, câmpul *limba căreia îi aparține termenul*, tocmai pentru că există posibilitatea, evocată mai sus, de a alege

pentru interfață o altă limbă decât cea căreia îi aparține termenul în cauză.
Principalele categorii de date (ISO) sunt grupate după cum urmează :

- Grup 1 : Categoriile de date despre termen și informații legate de termen
A Termenul
 - Subgrup1 : Termeni
 - Subgrup 2 : Categoriile de date legate de termen
 - Subgrup 3 : Echivalență
 - Subgrup 4 : Domeniu și sub-domeniu
 - Grup 2 : Categoriile de date legate de descrierea noțiunii
 - Subgrup 5 : Descrierea noțiunii
 - Subgrup 6 : Relații ale noțiunii
 - Subgrup 7 : Poziția noțiunii
 - Subgrup 8 : Nota
 - Grup 3 : Categoriile de date administrative
 - Subgrup 9 : Limbaj documentar
 - Subgrup 10 : Informații administrative
- B : Categoriile de date bibliografice

Atât grupurile, cât și câmpurile create, pot fi re poziționate, modificate sau șterse, după cum se poate vedea mai jos:

The screenshot displays the UNITERM web interface. At the top, there are two main sections: 'A.10.22.3 origine-bază de date (Cuvânt) [T] [F]' and 'B categorii de date bibliografice (vizibil) [T]'. The first section is associated with 'origine_base_donnees' and describes it as a database considered as a document in bibliographic documentation. The second section describes elements used for processing and indexing manuscripts. Below these sections, there are buttons for 'Ascunde', 'Arată', 'Șterge', and 'Mută în: Conform ISO 12620:1999'. There is also a language dropdown set to 'română'. The 'Adaugă:' section has a 'grup' dropdown. Below that, there are input fields for 'Nume:' and 'Descriere:', and a 'Părinte:' dropdown set to 'Conform ISO 12620:1999'. A 'Creează' button is at the bottom right. On the right side, there is a search bar with 'Căutare' and 'Caută' buttons, and a link for 'Importă intrări'. The footer contains the text 'Creat de Dorina Chiș, Valeriu Dorin Chiș | Ghid de utilizare'.

Există și posibilitatea de a edita fiecare câmp, pentru a efectua modificări în ceea ce privește descrierea, tipul, asocierea și utilizarea lui în generarea fișei precum și în căutare:

Numele și descrierile câmpurilor, fiecare cuvânt/ fragment de text pot fi, de asemenea, editate și apoi modificate prin adăugare sau eliminare a unor elemente, reformulare. La fel și traducerile. Fiecare câmp dispune de o descriere clară, astfel încât confuziile să fie excluse și completarea fișei terminologice să nu pună probleme. Aceasta este generată în conformitate cu structura activă (În cazul în care se lucrează cu mai multe structuri în paralel, structura activă trebuie selectată și marcată ca atare).

Pentru a stabili lista domeniilor în care s-ar putea încadra termenii ce vor fi cuprinși în bazele terminologice, am folosit un model de ontologie deja utilizat în Uniunea Latină (banca terminologică a Italiei), propus și dezvoltat de Ingetraut Dahlberg.¹

3.5. Funcționalitate

Operațiile care se pot efectua cu ajutorul GesTE sunt : înregistrarea datelor terminologie și prelucrarea lor, căutare simplă/ avansată, adăugarea/ ștergerea sau modificarea unor câmpuri, crearea/ consultarea/ validarea/ suprimarea unor fișe, importul și exportul datelor, administrarea bazei de date.

Căutarea simplă se poate, de exemplu, realiza:

- folosind doar steluța * (se afișează toate intrările introduse) ;
- după un cuvânt, o sintagmă ;
- după primele litere ale unui cuvânt urmate de steluță * (se afișează toate cuvintele care conțin grupul respectiv de litere) ;
- după un grup de litere plasate între steluțe (omțând în felul acesta prima și ultima parte a cuvântului) ;
- după autor (dacă îl cunoaștem) ;
- după echivalentul unui termen, de exemplu echivalentul lui *contrat* în germană **Vertrag** el trebuie pus însă între steluțe, având alte date înainte și după el.

Căutarea avansată (folosind butonul *Opțiuni*), permite căutarea unui termen sau a unei expresii doar în unele câmpuri, pe care le selectează utilizatorul.

Pentru **căutarea cu filtre**, fiecare utilizator trebuie să-și aleagă filtrele din lista *Adaugă pentru câmpul*. Se alege apoi între butoanele **conține/ exclude** din lista care derulează și se consemnează ce anume trebuie să conțină/ excludă câmpul respectiv. Se adăugă apoi alt câmp, după același procedeu. Dacă se alege, de exemplu, câmpurile *termen* și *limba*, iar în acesta din urmă se înscrie

¹ așa cum este el prezentat în articolul Gilioliei Negrini, *Quels systèmes conceptuels peut-on préconiser pour les banques de données terminologiques en ligne?*, publicat în revista *Terminologies nouvelles*, nr.18, juin 1998, *Terminotique et documentation*, pp. 41-53.

franceza, la apăsarea butonului *Caută*, se vor afișa toți termenii în limba franceză existenți în baza de date.

GesTE este însoțit de un **Ghid**, tradus în toate limbile de lucru, care oferă utilizatorilor toate informațiile necesare pentru o exploatare a lui în condiții optime.

4. Concluzii

Gestionarul de terminologie GesTE pune la dispoziția celor interesați un instrument flexibil și eficient. El permite colectarea, prelucrarea, modificarea, stocarea, exportul și importul informațiilor. Fiecare utilizator are posibilitatea, de la nivelul interfeței de lucru, să adapteze acest instrument de lucru nevoilor proprii.

Realizarea lui a fost posibilă datorită inițiativei și suportului Uniunii Latine.

Utilizarea acestui software poate contribui la armonizarea activității de gestionare a terminologiei și la facilitarea schimburilor de informații între diversele grupuri de lucru.

În prezent el este utilizat cu succes de la predarea lui, în mai 2007 de colectivul biroului din Paris al Direcției pentru Terminologie și Inginerie Lingvistică (DTIL) al Uniunii Latine și prezentat în varianta *demo* la adresa <http://dtil.unilat.org/ro/index.htm>.