

# O ABORDARE BAZATĂ PE ARBORI PENTRU MODELAREA DATELOR DIN CERCETAREA MEDICALĂ. COMPARAȚIE CU ABORDĂRI EXISTENTE ÎN APLICAȚII COMERCIALE

PÉTER OLÁH<sup>1</sup>, MARIUS MĂRUȘTERI<sup>2</sup>, CARMEN BIRIȘ<sup>3</sup>, MARIA MIHAELA OPRIS<sup>4</sup>, DANIELA DOBRU<sup>5</sup>

<sup>1</sup>Doctorand Universitatea de Medicină și Farmacie Tîrgu-Mureș, <sup>2,3,4,5</sup>Universitatea de Medicină și Farmacie Tîrgu-Mureș

**Cuvinte cheie:** baze de date medicale, sisteme informatice în domeniul medical, registru electronic de laborator

**Rezumat:** În lucrările noastre anterioare am dezvoltat o tehnică de structurare a datelor bazată pe arbori și multiarbori pentru a modela date din cercetarea clinică medicală. În acest articol încercăm să evaluăm în ce măsură acest model este utilizabil pentru a dezvolta un registru electronic pentru experimente de laborator, comparând punctele tari și punctele slabe ale acestuia cu produsele câtorva dintre ofertanții de pe piața de software. Această comparație se concentrează pe cinci aspecte critice: adaptabilitate, utilizarea structurilor de tip arbore, folosirea de șabloane, interogarea datelor și funcții de colaborare în grup. În ultimele două secțiuni ale articolului vom prezenta câteva discuții referitoare la rezultatele acestei comparații și vom trage câteva concluzii care vor avea un rol în orientarea eforturilor noastre viitoare privind dezvoltarea sistemului.

**Keywords:** medical databases, healthcare information systems, user interface, electronic laboratory notebook

**Abstract:** In our past work we have developed a data structuring technique based on trees and multitrees to model clinical research data. In this paper we try to assess the suitability of this model for keeping an electronic laboratory notebook by comparing its strengths and weaknesses with the products of some of the leading software providers on the market. The comparison focuses on five critical aspects: customizability, the use of tree-like structures, the use of patterns or templates, the techniques used to retrieve the data and functions that support collaboration. In the last two sections of the paper we discuss our findings and draw some conclusions that will serve as guidelines for our future work.

## INTRODUCERE

Într-o lucrare anterioară (1) am prezentat dezvoltarea unei tehnici de structurare a datelor pentru a modela date din cercetarea clinică în domeniul gastroenterologiei folosind arbori și multiarbori.(2)

În cadrul eforturilor noastre de a testa portabilitatea acestui model la alte domenii din cercetarea medicală, am încercat să adaptăm aplicația pentru cercetare de laborator în domeniul Bioingineriei și al Medicinii Regenerative. Aici am întâlnit conceptul deja consacrat de Registru Electronic de Laborator (REL), care urmărește să înlocuiască registrele de hârtie pe care majoritatea cercetătorilor le folosesc.

Trebuie menționat totuși că un simplu REL nu acoperă întreaga gamă de necesități informaționale ale unui proiect complex de cercetare. Proiectul în cadrul căruia testăm modelul nostru include de asemenea cercetare teoretică și cercetare experimentală pe un model animal.

## OBIECTIVE

În acest articol încercăm să evaluăm în ce măsură acest model este utilizabil pentru a dezvolta un registru electronic pentru experimente de laborator, comparând punctele tari și punctele slabe ale acestuia cu produsele câtorva dintre ofertanții de pe piața de software.

Comparația noastră va urmări în primul rând structurile și funcțiile de bază care permit cercetătorului să-și modeleze propriile date și va pune un accent mai mic pe aspecte legate de comunicare și securitate.

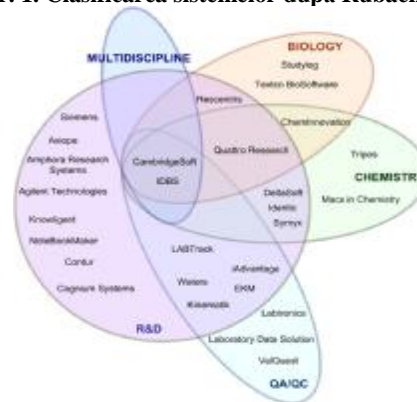
Analiza va fi una calitativă, care va urmări două obiective majore:

- Validarea conceptelor pe care le-am utilizat în abordarea noastră
- Identificarea de noi cerințe cu scopul de a îmbunătăți sistemul nostru astfel încât acesta să se ridice la nivelul standardelor din industrie.

## MATERIAL ȘI METODĂ DE LUCRU

Ca și punct de pornire pentru selecția sistemelor pe care le-am inclus în analiza noastră am utilizat clasificarea realizată de Rubacha et al. (3) (figura nr. 1).

Figura nr. 1. Clasificarea sistemelor după Rubacha



<sup>1</sup>Autor corespondent: Péter Oláh, Str. Ghe. Marinescu, Nr. 50, Cod 540136, Tîrgu-Mureș, România, E-mail: olah\_peter@yahoo.com, Tel: +40265 212111

Articol intrat în redacție în 09.11.2012 și acceptat spre publicare în 17.12.2012  
ACTA MEDICA TRANSILVANICA Martie 2013;2(1):10-12

Lista finală a fost construită folosind următoarele criterii:

- Au fost incluse doar sistemele care au fost dezvoltate în mod special pentru domeniul cercetării - dezvoltării, conform autorilor listei originale;(3)
- Soluțiile care nu au avut un site web funcțional la momentul analizei au fost eliminate;
- Lista a fost filtrată suplimentar prin selecția acelor aplicații care oferă capturi de ecran sau versiuni demonstrative direct în site-ul web de prezentare.

În final, sistemele pe care le-am evaluat pentru comparația noastră sunt:

- OpenLAB ELN de la Agilent Technologies;
- e-CAT produs de Axiope;
- iPad ELN de la Cognium Systems.

Caracteristicile pe care le-am investigat sunt:

- Adaptabilitate – poate utilizatorul să definească noi structuri fără intervenția dezvoltatorului de software?
- Folosirea arborilor – sunt utilizate structuri de tip arbore în modelarea datelor?
- Șabloane – există șabloane predefinite și/sau poate utilizatorul să definească noi tipare?
- Interogarea datelor – se folosesc interogări structurate sau motoare de căutare de uz general?
- Colaborare – există funcții implementate în cadrul sistemului care să faciliteze colaborarea între mai mulți utilizatori?

De asemenea, am analizat în detaliu modul de structurare a datelor din două puncte de vedere:

- Date administrative – ce structuri sunt utilizate pentru a organiza experimentele pe care cercetătorii le derulează într-o perioadă de timp?
- Date științifice – cum sunt structurate datele în cadrul fiecărui experiment?

O analiză similară a fost efectuată și pentru sistemul nostru pentru a-i compara funcționalitatea cu cea a sistemelor de pe piață.

### REZULTATE

Utilizând criteriile menționate mai sus am identificat unele similarități, dar și câteva diferențe între aplicațiile analizate. În această secțiune vom prezenta rezultatele obținute.

#### Agilent Technologies - OpenLAB ELN

Conceptele centrale în arhitectura sistemului sunt **tiparele și formularele**.

*Adaptabilitatea* sistemului este obținută prin folosirea de „tipare dinamice” și „formulare dinamice” pentru a oferi modalități de structurare personalizată a tiparelor și a datelor științifice stocate.

*Arborii* sunt folosiți pentru a structura datele administrative ale sistemului, cum ar fi ferestrele deschise, mulțimea tiparelor sau a formularelor existente. Datele științifice sunt definite în interiorul formularelor, neavând așadar o structură arborescentă.

*Șabloanele* sunt reprezentate de tiparele utilizate pentru a modela fluxurile de lucru ale cercetătorilor. Ele pot fi configurate prin asocierea de formulare la acestea și reutilizate în funcție de nevoi.

*Interogarea datelor* este realizată cu ajutorul unui motor de căutare, îmbunătățit cu un instrument de căutare structurat pentru a regăsi experimente în diverse faze de realizare sau supervizare.

*Colaborarea* este obținută folosind Internetul prin configurarea unor echipe virtuale în funcție de nevoi.

#### Axiope - e-CAT

Filozofia sistemului se învârtă în jurul conceptului de **înregistrare**.

Sistemul este *adaptabil* prin utilizarea unei serii de șabloane pentru a structura atât datele administrative, cât și pe cele științifice. Lista de șabloane este aceeași pentru ambele categorii, acest fapt făcând ca distincția între ele să fie uneori dificilă.

Partea administrativă a aplicației este structurată folosind *arbori*. O structură denumită Arborele Înregistrărilor este utilizată pentru a organiza Proiecte, Utilizatori și Alte Date (de exemplu date legate de aprovizionare). Pentru a structura datele științifice au fost preferate listele și câmpurile.

Sistemul folosește *șabloane*, acesta fiind preîncărcat cu peste 20 de tipare, cu posibilitatea de a defini unele noi.

*Interogarea datelor* este nestructurată, dar sistemul dispune de un motor de căutare puternic cu posibilitatea de a rafina rezultatele în funcție de proiect, respectiv în funcție de tipul sau numele câmpurilor.

*Colaborarea* este implementată prin utilizarea Internetului, partajarea datelor putând fi astfel configurată încât Investigatorul Principal să poată accesa toate înregistrările non-private ale membrilor unui grup de lucru.

#### Cognium Systems - iPad

Această aplicație reprezintă un Sistem de Gestionare de Conținut, care permite crearea și gestionarea de documente XML, adaptat domeniului cercetării experimentale.

Fiind un sistem bazat pe XML, acesta utilizează în mod intensiv **tag**-urile.

Construit în jurul conceptelor de „document” și „tag”, sistemul poate categoriza practic orice prin aplicarea de tag-uri asupra datelor elementare, fiind astfel foarte *adaptabil*.

Aplicația utilizează *arbori* pentru a organiza datele administrative într-un Browser de Documente, dar și pentru a naviga în structura experimentelor. Tag-urile utilizate formează o ierarhie în cadrul căreia utilizatorul poate naviga cu ajutorul arborilor. Există și posibilitatea de a crea legături directe între tag-uri, ceea ce sugerează construcția unei structuri de tip graf.

În încercarea de a găsi tipare, am identificat trei *șabloane* predefinite: notițe de cercetare, notițe de proiect și un document generic. Șablonul documentului generic permite utilizatorului definirea de noi tipare în funcție de necesități.

*Interogarea datelor* este susținută de un motor de căutare complex, care convertește criteriile de căutare în tag-urile corespunzătoare, pentru a le căuta apoi în baza de date.

*Colaborarea* între utilizatorii sistemului este realizată prin accesarea de depozite de date externe.

#### Abordarea bazată pe arbori / multiarbori

Conceptele centrale în abordarea noastră sunt **arborii** și **multiarborii**.

Adaptabilitatea acestei abordări derivă din modul de proiectare al bazei de date. Structura datelor științifice nu este modelată intensional, ci extensional, fiind astfel modificabilă prin simpla adăugare sau actualizarea unor înregistrări în zona de metadata a bazei de date.

*Arborii* și *multiarborii* sunt folosiți pentru a modela structura datelor științifice și nu pe cea a datelor administrative. Fiecare tip de experiment poate fi structurat ca un arbore. Diferite variante ale aceluiași experiment vor deveni subarborii unui multiarbore. În caz de nevoie, noduri ce aparțin de arbori diferiți pot fi legați în mod direct, rezultând astfel o structură mai complexă, cunoscută sub numele de Graf Aciclic Direcționat (GAD).

*Tiparele* sunt parte integrantă a proiectării sistemului. De fiecare dată când un utilizator definește un nou tip de experiment, se crează un șablon reutilizabil. Tiparele pot fi de asemenea transferate de la o structură la alta prin copiere/alipire.

E o modalitate specială de copiere și alipire care realizează replicarea unui subarboare pornind dintr-un anumit nod.

*Interogarea datelor* este implementată cu prioritate prin interogări structurate. Acestea sunt interogări complexe ale bazei de date ce se află la baza sistemului, care se folosesc de structurile de tip arbore definite de către utilizator.

*Colaborarea* este limitată pe moment la utilizatorii aceluiși server, sistemul nefiind adaptat utilizării prin Internet. În această formă colaborarea se poate realiza prin acordarea de drepturi de acces specifice utilizatorilor sau grupurilor de utilizatori.

### DISCUȚII

În această secțiune ne vom concentra discuțiile asupra modului în care modelul bazat pe arbori este potrivit pentru implementarea unui REL pentru cercetare academică. Ar trebui menționat faptul că cerințele utilizatorilor academici diferă de cerințele celor care lucrează în industria chimică și farmaceutică.(4)

În modelul nostru *flexibilitatea* este considerată esențială, rezultând un sistem extrem de personalizabil de către utilizatorul final. Analiza celorlalte trei aplicații de mai sus susține ideea că personalizarea este un factor cheie pentru un REL de succes.

Structurile de tip *arbore* sunt din ce în ce mai prezente în interfețele grafice utilizator, într-o varietate de sisteme. În aplicațiile REL analizate această structură a fost utilizată în principal pentru a organiza diferite tipuri de înregistrări (experimente, modele, tipuri de date etc). Niciuna dintre aceste abordări nu a utilizat arborii pentru a modela structura datelor științifice.

Interfețele ce oferă funcționalitatea de introducere a datelor de către cercetători utilizează în general agregare de obiecte. Aceste obiecte pot fi privite ca fiind elemente constructive extrem de specializate, care pot fi puse împreună pentru a reprezenta structura sau fluxul unui experiment. Aceste elemente constructive pot fi combinate în variante nelimitate, oferind un grad ridicat de flexibilitate. Totuși, datele colectate în acest mod nu par să fie structurate în conformitate cu regulile unei baze de date relaționale. Prin comparație, modelul bazat pe arbori oferă o flexibilitate similară, dar păstrează datele extrem de structurate, susținute de o bază de date relațională normalizată în partea de back-end a sistemului.

Cercetarea presupune de multe ori repetarea aceluiași pași, astfel încât nevoia de *șabloane* este inerentă. Toate sistemele descrise mai sus oferă într-un fel sau altul funcționalitate pentru definirea șabloanelor sau tiparelor. În timp ce această funcție este opțională în aplicațiile comerciale analizate, în abordarea noastră ea este încorporată în proiectarea sistemului. Astfel, un cercetător nu poate înregistra un experiment înainte de a defini un șablon pentru acesta. Acest lucru poate constitui un dezavantaj în anumite situații, dar pe termen lung se va asigura structura corespunzătoare a tuturor datelor introduse în sistem.

În ceea ce privește *interogarea datelor*, folosirea motorului de căutare pare a fi norma generală în aplicațiile comerciale. Sistemul nostru nu oferă încă această facilități, dar ea poate fi dezvoltată prin implementarea funcției de căutare de tip „full text” în baza de date.

Unele dintre motoarele de căutare analizate sunt foarte complexe, folosind mai multe nivele de rafinare a rezultatelor căutării. Cu toate acestea, complexitatea algoritmilor de căutare nu poate substitui exactitatea interogărilor executate asupra unor date structurate. Acest tip de interogare a datelor oferă utilizatorului posibilitatea de a viza elemente specifice ale

informațiilor și să aștepte rezultate precise. Acest tip de interogare este, în opinia noastră, punctul forte al abordării bazate pe structurile arborescente.

Analiza sistemelor REL comerciale a arătat, de asemenea, că *colaborarea* este un factor cheie în realizarea de experimente științifice de succes. În plus, colaborarea trebuie să funcționeze indiferent de distanța dintre membrii echipelor, sistemele web, stabilind în mod clar tendința în acest sens.

### CONCLUZII

Analiza a trei sisteme comerciale disponibile de tip REL a arătat că personalizarea, utilizarea de șabloane / tipare și colaborarea sunt cele mai urmărite funcționalități în acest tip de aplicație.

Utilizarea structurilor arborescente pentru a modela datele științifice îndeplinește aceste cerințe și, în opinia noastră, prezintă unele avantaje clare legate de interogarea datelor. Cu toate acestea, această abordare nu a fost pe deplin validată încă.

#### Notă:

Această lucrare este parțial susținută de Programul Operațional Sectorial Dezvoltarea Resurselor Umane (POS DRU), finanțat din Fondul Social European și de către Guvernul României, în cadrul proiectelor cu numerele de contract POSDRU 64331 și POSDRU 60782.

### REFERINȚE

1. Olah P, et al. A database design pattern for structuring hierarchical medical data. Acta Medica Marisensis, 2013. to be published in vol. 58 no. 6.
2. Furnas GW, Zacks J. Multitrees: Enriching and Reusing Hierarchical Structure, in ACM CHI 94 Human Factors in Computing Systems; 1994. p. 330-336.
3. Rubacha M, Rattan AK, Hosselet SC. A review of electronic laboratory notebooks available in the market today. J Lab Autom; 2011;16(1):90-8.
4. Rudolphi F, Goossen LJ. Electronic laboratory notebook: the academic point of view. J Chem Inf Model. 2012;52(2):293-301.