

## Descrierea folosind diagrame a prelucrării distribuite a datelor spațiale

Vasile Dănuț Mihon, Angela Minculescu, Vlad Colceriu, Dorian Gorgan

Universitatea Tehnică din Cluj-Napoca

Str. Memorandumului, Nr. 28, 400114, Cluj-Napoca

E-mail: [vasile.mihon, vlad.colceriu, dorian.gorgan}@cs.utcluj.ro](mailto:vasile.mihon, vlad.colceriu, dorian.gorgan}@cs.utcluj.ro),  
[angela.minculescu@gmail.com](mailto:angela.minculescu@gmail.com)

**Rezumat.** De cele mai multe ori modelarea și simularea fenomenelor din lumea reală implică un număr mare de operații complexe, supuse anumitor constrângeri și reguli de compatibilitate. Soluția propusă în cadrul acestei lucrări optimizează întregul proces de descriere a fenomenelor prin intermediul grafurilor de execuție, punând la dispoziția utilizatorului tehnici interactive de specificare a nodurilor și a relațiilor dintre acestea, validarea sintactică automată a fenomenului descris, organizarea ierarhică a grafurilor, tehnici de dezvoltare colaborativă, reutilizarea resurselor, personalizarea componentelor grafice. Soluția descrisă permite folosirea practică a conceptelor chiar și de către utilizatorii care nu dețin cunoștințe despre infrastructura de calcul. Lucrarea de față prezintă mecanismele folosite în cadrul aplicației WorkflowEditor, pentru definirea interactivă a grafurilor de prelucrare distribuită a datelor spațiale.

**Cuvinte cheie:** graf de prelucrare, hipergraf, prelucrare distribuită, date spațiale, caz de utilizare, reguli sintactice, mediu colaborativ.

### 1. Introducere

Descrierea și prelucrarea fenomenelor naturale, a experimentelor reale sau simulate din diferite domenii de activitate reprezintă un proces complex, care implică o bună cunoaștere a problemei, colectarea datelor experimentale, descrierea sintactică și semantică a soluției propuse și, analiza și interpretarea corectă a rezultatelor obținute. Un fenomen (experiment, scenariu) este modelat ca o succesiune de operatori (algoritmi, funcții) interconectați sub forma unui graf de execuție, care pe baza unui set de intrări, produce într-un timp finit un rezultat valid.

Scopul reprezentării fenomenelor (experimentelor sau scenariilor) sub formă de graf este acela de a oferi o modalitate optimă de organizare a modelului de date. Folosind această metodă grafică, reprezentarea fenomenelor este mai ușor de vizualizat și interpretat de către diferite categorii de utilizatori. Modelul poate fi folosit ulterior în execuția scenariului descris prin intermediul grafului asociat. Fiecare nod de intrare primește o resursă-sistem, iar ordinea prelucrării este conform fluxul descris de conectarea nodurilor prin arce orientate.

Majoritatea aplicațiilor actuale nu permit descrierea vizuală a fenomenului în sine, ceea ce face metoda vulnerabilă din punctul de vedere al erorilor sintactice care pot fi introduse de către persoanele care realizează descrierea acestuia. Mai mult, utilizatorii trebuie să dețină cunoștințe din domeniul din care provine problema, cunoștințe de programare, precum și o experiență vastă în ceea ce privește utilizarea acestor tipuri de aplicații.

Soluția propusă în lucrare încearcă să elimine aceste neajunsuri și să ofere asistență utilizatorului pe întregul parcurs al procesului de descriere a scenariului. Astfel, sistemul se adresează atât utilizatorilor care dețin cunoștințe minimale despre domeniul problemei, cât și utilizatorilor experți. Fără a necesita un bagaj bogat de cunoștințe din partea utilizatorilor, sistemul le oferă suport, acționând transparent pentru evitarea eventualelor erori.

În starea inițială, aplicația WorkflowEditor oferă un set de funcții (operatori) de bază din domeniul prelucrării datelor geospațiale, disponibile pentru toți utilizatorii platformei. Folosind acest set de bază se pot construi noi scenarii de utilizare, prin interconectarea modulelor existente. Se oferă astfel posibilitatea de reutilizare a resurselor pentru generarea unor funcții noi.

Lucrarea de față descrie conceptele teoretice și practice care au fost implementate în cadrul aplicației WorkflowEditor, pentru a oferi suport în crearea și gestionarea interactivă a scenariilor. Execuția ulterioară a acestor grafuri nu este realizată în cadrul aplicației WorkflowEditor. Pentru aceasta se pot folosi platforme de prelucrare externe, cum ar fi GreenLand (Mihon et al, 2012), care oferă suport pentru execuția, monitorizarea și interpretarea acestor scenarii.

În continuare, secțiunea 2 prezintă alte realizări asemănătoare din domeniul lucrării. Secțiunea 3 evidențiază conceptele teoretice care au stat la baza

dezvoltării sistemului, urmând ca în ultima parte, secțiunea 4, să se descrie modul în care aceste concepte au fost dezvoltate ca și acțiuni-utilizator.

## 2. Alte realizări în domeniu

Există în literatura de specialitate un număr relativ mic de lucrări și sisteme care se referă la posibilitatea de a descrie în mod interactiv scenarii complexe pentru a fi executate în medii distribuite: Pegasus (Deelman et al, 2004), Taverna (Oinn et al, 2004) sau GridFlow (Cao et al, 2003). Toate aceste soluții folosesc o abordare similară cu cea propusă în aplicația WorkflowEditor. Abordarea se referă la descrierea scenariilor sub formă de grafuri, în care fiecare nod încapsulează un algoritm (funcție matematică) prin care se reprezintă un pas al fluxului de execuție din cadrul scenariului complex.

Fiecare dintre sistemele enumerate anterior se bazează pe aceleași principii, dar modul de implementare diferă de la caz la caz. Astfel, unele dintre ele permit generarea de grafuri cu noduri izolate, în timp ce altele oferă posibilitatea de a conecta simultan un nod de ieșire la mai multe noduri de intrare, sau permit aranjarea automată a nodurilor pe bază de algoritmi de afișare.

Soluția propusă în cadrul acestei lucrări se diferențiază de sistemele prezentate anterior prin:

Organizarea grafurilor pe nivele ierarhice. Astfel, un nod poate conține un alt graf imbricat, denumit sub-graf. Mecanismul acesta poate fi aplicat recursiv pe mai multe nivele de adâncime, fără a exista o limită impusă;

Descrierea colaborativă a grafurilor: fiecare utilizator își poate aduce contribuția la dezvoltarea scenariilor complexe, prin adăugarea de noi algoritmi, sau prin actualizarea celor existenți;

Reutilizarea resurselor pentru extinderea scenariilor existente sau pentru generarea de noi scenarii. Pentru aceasta se pot folosi resursele deja existente, la care se adaugă funcțiile dorite;

Extinderea setului de bază: majoritatea sistemelor descrise anterior sunt destinate unui domeniu specific de activitate. Momentan scenariile existente în baza de date WorkflowEditor sunt din domeniul prelucrării informațiilor

geospațiale. Există însă posibilitatea de a include și fenomene studiate în cadrul altor discipline, precum: arheologie, astrologie, chimie, fizică, matematică;

Personalizarea grafurilor pentru o mai bună identificare a funcțiilor oferite. Există posibilitatea de a aplica palete de culori, de a stabili stiluri de desenare a liniilor, dimensiunea nodurilor etc.

Există și concepte similare între soluția propusă și editoarele de grafuri prezentate anterior. Dintre acestea cele mai importante sunt:

Validarea operațiilor-utilizator pe baza unui set de reguli, aplicate pentru trasarea grafică a elementelor;

Poziționarea și organizarea automată a grafurilor complexe, cu zeci de noduri și conexiuni multiple. Se pot folosi în acest sens diferiți algoritmi de ierarhizare structurală: circulari, propuși de Baur&Brandes (2004), ortogonali (Eiglsperger et al, 2001), bazați pe forțe (Didimo et al, 2011) sau algoritmi ierarhici (Di Battista et al, 1999);

Structura modelului de date folosită pentru salvarea scenariilor se poate realiza în diferite formate: Extensible Markup Language (XML, 2012), Guideline XML (GXML, 2012), JavaScript Object Notation (JSON) conform recomandărilor lui Eriksson&Hallberg (2011), sau alte formate proprii. Datorită faptului că modelarea scenariilor se realizează sub formă de grafuri, s-a ales folosirea structurii de date XML.

### 3. Concepte teoretice

Un graf poate fi definit ca o pereche ordonată  $G=(X, U)$ , unde  $X$  este o mulțime finită și nevidă de noduri, iar  $U$  este o mulțime nevidă de perechi de elemente din  $X$ , care poartă denumirea de arce uni-direcționale. Trebuie menționat faptul că toate aceste grafuri trebuie să aibă proprietatea de aciclitare, în caz contrar există posibilitatea de execuție infinită a iterațiilor din cadrul buclelor definite de utilizator.

Pentru modelarea și simularea scenariilor, soluția propusă în cadrul lucrării este dezvoltată pe baza noțiunilor din teoria grafurilor. Elementele mulțimii  $X$  se identifică ca și algoritmi, funcții și modele matematice descrise prin intermediul limbajelor de programare Java și C. Odată încărcăți în sistem

(sub diferite forme: fișiere executabile, fișiere jar, fișiere bash) acești algoritmi (operatori) au un identificator unic, putând fi folosiți ca și noduri în graful de descriere a scenariilor.

Fiecare nod are o mulțime finită și nevidă de intrări și ieșiri. Fiecare intrare/ieșire corespunde unui anumit tip de date: valori numerice, valori șir de caractere, fișiere în diferite formate (.txt, .sh, .dbx) și imagini satelitare în formatul Tagged Image File Format (TIFF).

Descrierea scenariilor sub această formă este benefică din două puncte de vedere. Primul se referă la reprezentarea vizuală a nodurilor și a conexiunilor dintre ele. În acest caz, tipurile de date ale intrărilor și ieșirilor au o semnificație abstractă, ele fiind folosite doar pentru validarea setului de reguli sintactice (ex.: se pot realiza conexiuni numai între noduri de același tip). Crearea modelului de date specific acestor scenarii reprezintă al doilea avantaj. În acest caz tipurile de date sunt instanțiate ca resurse-sistem, care apoi sunt folosite în prelucrarea fizică a scenariilor.

Elementele mulțimii  $U$ , din definiția anterioară a grafului, reprezintă arce uni-direcționale care descriu fluxul de execuție a scenariilor modelate prin grafuri. Pe baza setului de reguli sintactice, arcul poate lega doar ieșiri-intrări de același tip. În cadrul aplicației WorkflowEditor nu există posibilitatea de a defini arce frânte, acestea fiind reprezentate printr-un singur segment de dreaptă.

În cazul în care cel puțin un nod conține un graf imbricat, structura poartă denumirea de hipergraf. Matematic, el poate fi definit astfel: o pereche ordonată  $G_H=(X_H, U_H)$ , în care cel puțin un nod din  $X_H$  conține o altă pereche ordonată  $G=(X, U)$ . Celelalte elemente constitutive se păstrează, în sensul că fiecare arc conectează două noduri independente, compatibile din punct de vedere al tipului de date.

Figura 1 reprezintă structura conceptuală a unui hipergraf care conține trei operatori de bază (OP1, OP2 și OP3), un graf imbricat (SW1), 5 intrări (corespunzătoare operatorilor OP1 și OP2) și două ieșiri. Nodul SW1 conține la rândul său alte elemente grafice, interconectate prin arce uni-direcționale. Trebuie remarcat faptul că cele două intrări ale nodurilor OP4 și SW2 corespund ieșirilor

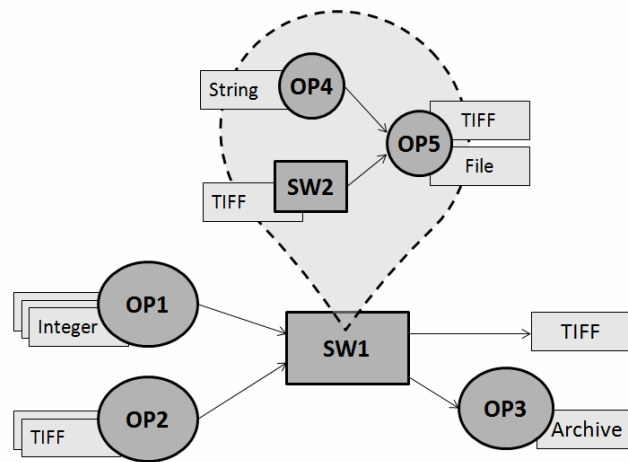


Figura 1. Structura conceptuală a unui hipergraf

lui OP1 și OP2. Ieșirea de tipul “File” a grafului imbricat SW1, devine intrare pentru operatorul OP3.

### 3.1 Estetici de vizualizare

Modelarea și simularea fenomenelor complexe din lumea reală implică un număr mare de operații, necesitând folosirea unor grafuri de dimensiuni mari (zeci de noduri, cu multiple legături interne). Pentru o mai bună vizualizare a acestora, a fost implementat un algoritm de afișare și organizare automată a grafurilor. Esteticile care au stat la baza acestui algoritm sunt descrise în cele ce urmează:

*Minimizarea suprafeței de desfășurare a grafului:* se dorește folosirea unui spațiu cât mai redus pentru afișarea corectă și inteligibilă a grafului. Soluția propusă se bazează pe determinarea poziției și dimensiunii celui mai mic dreptunghi care permite acoperirea integrală a nodurilor și arcelor din graf. În cazul redimensionării (scalării) maxime a grafului, dreptunghiul este identic cu suprafața de afișare. Datorită acestui aspect, s-a convenit

eliminarea vizuală a elementelor care nu pot fi cuprinse în cadrul acestei regiuni;

*Uniformizarea lungimii arcelor și a rezoluției unghiulare:* este în strânsă legătură cu prima estetică, prin care se dorește reducerea spațiului ocupat de reprezentarea vizuală a grafului. Uniformizarea rezoluției unghiulare se folosește în afișarea simetrică a nodurilor grafului;

*Păstrarea hărții mentale a grafului,* propusă de Purchase&Hoggan (2007): se referă la determinarea unei modalități de păstrare a poziției, dimensiunii și a ordinii nodurilor, indiferent de starea inițială a lor. Prin satisfacerea acestei cerințe, se dorește minimizarea efortului utilizatorilor de a memora structura de date. Prin identificarea coordonatelor relative și a dimensiunii primitivelor grafice folosite în reprezentarea nodurilor, sistemul este capabil de a satisface această cerință, indiferent de acțiunile anterioare (ex.: re poziționarea nodurilor, modificarea lungimii și a direcției arcelor);

*Simetrie:* se dorește afișarea simetrică a nodurilor în funcție de un punct pivot, care de regulă reprezintă centrul grafului. În acest sens pot fi folosiți diferiți algoritmi de afișare grafică, care vor fi prezentați în secțiunea următoare. Soluția propusă se bazează pe modelul algoritmilor ierarhici, care permit afișarea nodurile frunză pe primul nivel, urmate apoi de părinții acestora. Procedul continuă în mod recursiv până la parcurgerea integrală a grafului;

*Minimizarea numărului de intersecții dintre arce:* folosirea arcelor frânte reprezintă o posibilă soluție pentru evitarea excesivă a intersecțiilor. Datorită dificultății implementării unui astfel de sistem, momentan există disponibilă doar varianta de reprezentare a arcelor prin segmente de dreaptă. Planarizarea grafurilor constituie a doua soluție relativ la această problemă. Combinarea celor două soluții funcționează în majoritatea cazurilor, dar nu și pentru grafuri cu un număr ridicat de arce. În concluzie, cerința de minimizare a numărului de intersecții dintre arce este o problemă NP-completă, așa cum a fost descrisă de Garey&Johnson (1983), care nu poate fi satisfăcută în totalitate;

Există situații în care esteticile prezentate anterior se exclud reciproc, fiind necesar un compromis pentru afișarea și organizarea grafurilor. Soluția propusă în cadrul acestei lucrări asociază fiecărei cerințe o prioritate pe baza

căreia se realizează afișarea nodurilor și arcelor. Ordinea descrisă anterior reflectă prioritatea atribuită, prima estetică fiind de prioritate maximă.

### 3.2 Algoritmi de organizare spațială a grafurilor

În funcție de categoria din care face parte graful, se poate determina un algoritm optim de afișare a acestuia, care să satisfacă cât mai multe dintre esteticile definite în secțiunea anterioară. Determinarea categoriei de apartenență presupune analiza grafului din mai multe perspective: dacă graful este ciclic sau aciclic, dacă este orientat sau neorientat, dacă are o structură arborescentă sau, dacă un nod poate avea mai mulți părinți etc.

De regulă, convențiile de desenare se referă la modalitatea de reprezentare a arcelor din graf (Figura 2). Pentru aceasta se pot folosi segmente de dreaptă, linii frânte, împărțirea matricială a spațiului și poziționarea nodurilor în dreptul liniilor și coloanelor din matricea virtuală etc. Avantajul folosirii liniilor frânte este dat de reducerea numărului de intersecții dintre arce. Poziționarea nodurilor în zonele de intersecție dintre liniile și coloanele matricei, minimizează numărul arcelor oblice, satisfăcând în același timp estetica de uniformizare a lungimii arcelor și a rezoluției unghiulare.

Există patru clase de algoritmi care se pot aplica pentru ordonarea automată a nodurilor și arcelor unui graf: algoritmi circulari, ortogonali, bazați pe forțe și, algoritmi ierarhici. O scurtă descriere a fiecărei clase este redată în cele ce urmează. Algoritmii circulari, propuși de Baur&Brandes (2004), oferă posibilitatea ordonării nodurilor sub forma unei elipse. Se recomandă folosirea acestora pentru rețele mici de noduri. În cazul grafurilor complexe, nu oferă o viziune clară asupra structurii rețelei.

Afișările ortogonale (Eiglsperger et al, 2001) se bazează pe re poziționarea nodurilor în centrul matricei virtuale, astfel încât majoritatea arcelor devin paralele cu axele de coordonate. Acest mod de reprezentare reduce numărul de intersecții dintre arce.

Algoritmii bazați pe forțe (Didimo et al, 2011) satisfac esteticile de simetrie, nesuprapunere a nodurilor și, de uniformitate a lungimii arcelor și a rezoluției unghiulare. Metoda constă în mutarea repetată a nodurilor pe baza forțelor dintre ele. În fiecare pas se calculează o funcție de optim local, pe baza euristicilor definite anterior. Procesul continuă până la determinarea



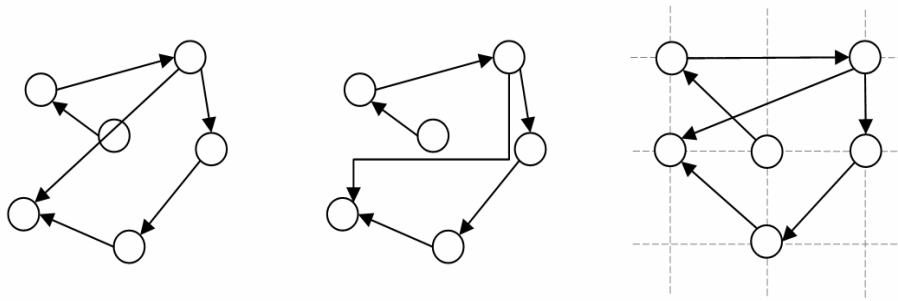


Figura 2. Convenții de desenare a grafurilor prin: (a) segmente de dreaptă, (b) linii frânte, (c) poziționarea nodurilor în matricea virtuală

valorii de optim global, care corespunde soluției care respectă cele mai multe dintre euristicile de vizualizare.

Soluția propusă în cadrul acestei lucrări folosește algoritmi ierarhici (Di Battista et al, 1999) pentru organizarea grafurilor. Procesul de afișare pornește de la identificarea nodurilor frunză (de pe cele mai de jos nivele ierarhice). În continuare se încearcă poziționarea pe nivele consecutive a nodurilor conectate prin muchii directe. Procesul continuă până la parcurgerea integrală a elementelor grafului.

### 3.3 Constrângeri și reguli de validare

Se aplică nodurilor și arcelor care intervin în structura unui graf, având ca și efect minimizarea numărului de erori-utilizator. Validările aplicate la nivelul reprezentării grafice, asigură și corectitudinea modelului de date XML. Acest model este folosit ulterior în prelucrarea scenariului pe infrastructuri de calcul paralele și distribuite. Fiecare intrare primește o resursă-sistem, iar ordinea prelucrării se bazează pe fluxul descris prin intermediul arcelor.

Principalele validări care sunt implementate în cadrul aplicației WorkflowEditor sunt descrise în cele ce urmează:

*Suprapunerea nodurilor:* nu este permisă acțiunea de poziționare a unui nod deasupra altuia. Există o distanță ( $d$ ) limită între nodurile adiacente care trebuie satisfăcută. În cazul în care se produc asemenea situații, sistemul acționează automat prin poziționarea nodului respectiv în regiunea corectă;

*Unicitatea nodurilor:* fiecare element nou adăugat este unic identificat în cadrul grafului. Astfel, dacă se folosește aceeași denumire pentru două elemente, utilizatorul este avertizat prin mesaje sugestive;

*Conectarea ieșirilor:* fiecare nod are asociat un număr finit de intrări și ieșiri. Constrângerea impusă de sistem este aceea de a conecta prin arce toate ieșirile nodurilor cu intrări din cadrul altor noduri. Astfel se evită dezvoltarea grafurilor cu elemente izolate;

*Specificarea arcelor:* un arc reprezintă o legătură uni-direcțională între ieșirea unui nod și intrarea altuia. Acțiunea de conectare a ieșirii unui nod la una dintre intrările aceluiași nod este considerată incorectă;

*Potrivirea tipurilor de date:* fiecare arc trebuie să conecteze noduri de același tip (ex.: numeric, șir de caractere, imagini satelitare etc.);

*Aciclicitatea grafului:* aplicația WorkflowEditor nu oferă suport pentru modelarea scenariilor prin grafuri ciclice. Aceste tipuri de grafuri necesită mecanisme avansate de gestionare și control, existând posibilitatea introducerii execuțiilor infinite. Fără să existe o condiție specifică de oprire, după terminarea unei iterații se continuă cu următoarea și tot așa. În acest caz procesul de prelucrare continuă la infinit;

*Graf nenul:* existența a cel puțin unui nod în cadrul grafului dezvoltat.

### 3.4 Sistemul de coordonate

Pentru dezvoltarea și afișarea grafurilor, aplicația WorkflowEditor folosește spațiul bi-dimensional XOY. Există două modalități de lucru: folosirea coordonatelor normalizate și a coordonatelor ecran. În ambele cazuri colțul din stânga sus a suprafeței active de lucru are coordonatele (0, 0). Restul punctelor se determină în funcție de acesta, valoarea lor fiind dată de diferența de pixeli, pe axele X și Y, dintre punctul (0, 0) și noua poziție.

Folosirea coordonatelor normalizate presupune aducerea lor în intervalul [0, 1]. În acest caz precizia de calcul trebuie să fie mare, datorită faptului că există foarte multe puncte de valori apropiate (ex.: 0,03; 0,032). Dezavantajul acestei metode îl reprezintă spațiul ridicat de stocare a coordonatelor numere-reale pentru nodurile grafului, în raport cu folosirea coordonatelor ecran de tipul numere-naturale. Ambele metode asigură portabilitatea coordonatelor pe monitoare de rezoluții diferite. Ca urmare,

descrierea de scenarii complexe prin intermediul grafurilor folosește coordonatele ecran pentru poziționarea elementelor grafice.

### **3.5 Tehnici colaborative de dezvoltare a grafurilor**

De cele mai multe ori descrierea scenariilor complexe implică un grup de specialiști. Se dorește existența unui mediu colaborativ care să permită accesul partajat la resursele-sistem. Colaborarea dintre specialiști se referă în special la implementarea și conectarea algoritmilor (funcțiilor) care intervin în cadrul fiecărui nod din graf.

Sistemul descris în această lucrare oferă posibilitatea de acces partajat la resursele existente și de dezvoltare colaborativă a scenariilor. Pentru o mai bună delimitare a contribuțiilor aduse de fiecare persoană, este nevoie de o identificare a acestora. În acest scop se folosesc palete de culori pentru marcarea nodurilor specifice fiecărui utilizator. De exemplu, elementele în curs de dezvoltare sunt marcate cu culoarea roșie, iar cele deja finalizate și validate de comunitatea științifică sunt afișate cu verde.

### **3.6 Reutilizarea resurselor**

Scenariile dezvoltate într-un anumit domeniu de activitate, pot fi reutilizate în același domeniu, prin extinderea sau prin îmbunătățirea lor, sau în alte domenii științifice, prin adaptarea acestora la noile cerințe. Procesul de reutilizare este dependent de complexitatea informațiilor prezentate, descrierea utilizării funcțiilor implementate, gradul de pregătire al utilizatorilor în domeniile pentru care au fost dezvoltate aceste scenarii etc.

Astfel, se pot identifica patru modalități de realizare a unor noi resurse: modificări de conținut, adaptarea resurselor pentru alte domenii de activitate, adaptarea resurselor în funcție de cunoștințele utilizatorilor și, prezentarea în diferite forme a conținutului.

Modificările aduse conținutului presupun actualizarea acestuia prin adăugarea și eliminarea de elemente grafice. Acestea sunt frecvente mai ales în faza incipientă de dezvoltare a scenariilor.

Reutilizarea grafurilor în alte domenii decât cel pentru care au fost proiectate inițial, presupune existența unei documentații complete în ceea ce privește semnificația nodurilor, a intrărilor și ieșirilor. Algoritmii, metodele

și funcțiile descrise prin limbaje de programare, încapsulate ca și resurse abstracte în cadrul nodurilor, necesită de asemenea o descriere completă.

În funcție de cunoștințele utilizatorilor într-un anumit domeniu, există posibilitatea de adaptare a resurselor existente la nevoile proprii. Ca și în cazul precedent este foarte importantă disponibilitatea unei documentații complete a întregului mecanism de funcționare.

Sistemul permite personalizarea elementelor grafice, în funcție de preferințele utilizatorului. Se pot defini palete de culori, stiluri de linii, algoritmi de ordonare automată a elementelor, nivele de acuratețe în ceea ce privește procesul de specificare al arcelor etc. Toate aceste elemente grafice oferă posibilitatea de prezentare a aceleiași resurse în diferite forme. Un studiu de caz interesant îl constituie folosirea tehnicilor de realitate îmbogățită peste rezultatele obținute în urma procesării datelor spațiale. Conceptele descrise în (Dârdală et al, 2011) constituie punctul de plecare în abordarea acestui subiect.

#### **4. Aplicația WorkflowEditor**

Prin intermediul sistemului descris în cadrul acestei lucrări, utilizatorul are posibilitatea de a dezvolta structuri complexe de grafuri, denumite hipergrafuri, prin specificarea interactivă a nodurilor și a relațiilor dintre acestea.

Setul de reguli de validare a acțiunilor-utilizator reduce semnificativ numărul erorilor sintactice care pot interveni în cadrul dezvoltării scenariilor complexe. Aceste reguli produc un efect vizual ușor de interpretat de către utilizator, iar ajustările necesare sunt aplicate automat de către sistem.

Din graful construit se poate genera o structură de date de tip XML, care este apoi folosită ca și suport pentru prelucrarea fizică a scenariului. Editarea manuală a modelului XML nu îi asigură corectitudinea din punct de vedere sintactic. Cu toate că fișierul este corect structurat, pot să intervină o serie de inconsistențe cum ar fi suprapunerea nodurilor, apariția nodurilor izolate, conectarea eronată a arcelor, incompatibilitatea tipurilor de date între extremitățile arcelor (ex.: nodul de start este de tipul șir de caractere, iar cel final de tipul întreg) etc.

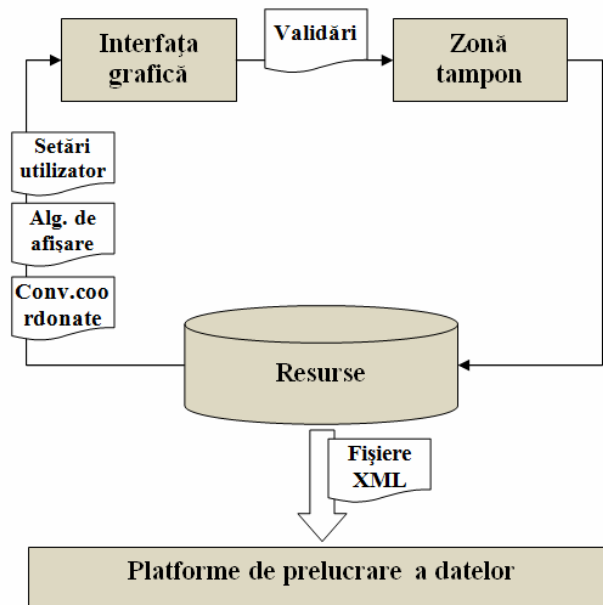


Figura 3. Arhitectura aplicației WorkflowEditor

#### 4.1 Arhitectura aplicației

Arhitectura aplicației este prezentată în Figura 3. Transferul de date către interfața grafică se bazează pe servicii Web, implementate prin paradigma REST (Pautasso et al, 2008). Afișarea grafurilor în cadrul suprafeței active se realizează într-o succesiune de trei pași. În primul rând are loc ajustarea coordonatelor inițiale în funcție de noua rezoluție a dispozitivului pe care este vizualizată aplicația. Al doilea pas constă în poziționarea și organizarea automată a grafului pe suprafața activă de lucru, folosind algoritmi ierarhici de vizualizare. În cazul în care autorul acestui scenariu a folosit elemente grafice personalizate (palette de culori, stiluri de desenare a liniilor etc.) sistemul oferă posibilitatea de a le reproduce, indiferent de utilizatorul și de sistemul care le folosește.

Sistemul permite dezvoltarea unor noi scenarii, prin reutilizarea resurselor existente. Fiecare acțiune-utilizator este validată prin intermediul unui set de

reguli și constrângeri. Datele intermediare sunt memorate într-o zonă tampon, iar în momentul salvării grafului acestea se transmit bazei de date atașate aplicației.

Expunerea resurselor către platformele externe de prelucrare a datelor se realizează prin intermediul fișierelor XML. Aplicația WorkflowEditor se folosește doar pentru descrierea scenariilor, prin intermediul grafurilor. Prelucrarea și execuția acestor scenarii trebuie însă realizată prin intermediul altor platforme, cum ar fi GreenLand. Procesul de execuție distribuită a datelor spațiale folosind platforma GreenLand este descris în Colceriu&Gorgan (2011) și (Mihon et al, 2012).

## 4.2 Crearea primitivelor grafice

Există trei tipuri de primitive grafice utilizate în cadrul aplicației: cercuri, dreptunghiuri și linii uni-direcționale. Primele două se folosesc pentru reprezentarea nodurilor, iar primitiva de linie pentru reprezentarea arcelor din graf. În cazul în care nodul este de tipul hipergraf (conține alte grafuri imbricate) se recurge la modalitatea de afișare prin dreptunghi. În caz contrar, când nodul este format doar din algoritmul (funcția matematică de bază) pentru care a fost creat inițial, se folosesc primitivele de tip cerc. Nodurile de acest tip sunt denumite operatori.

Fiecare nod al grafului conține o mulțime finită de intrări și ieșiri. Convenția de desenare presupune afișarea intrărilor în partea stângă a nodului, iar ieșirile în partea din dreapta.

Atunci când numărul de intrări/ieșiri este mare, se realizează o redimensionare automată a nodurilor (cercuri sau dreptunghiuri) pentru a permite afișarea integrală a lor. Un scenariu complex este redat în Figura 4. Acesta presupune clasificarea regiunilor de vegetație, a regiunilor urbane și a celor de apă din regiunea Istanbul (Bektaş et al, 2012).

Scenariul prezentat este din domeniul prelucrării datelor satelitare și folosește un set de cinci algoritmi, reprezentați vizual sub forma a cinci noduri (Atmospheric correction, Mosaic, EVI, Density slicing și Accuracy assessment). Descrierea detaliată a acestora este prezentată în (Bektaş et al, 2012). Parcurgerea se face din partea stângă înspre dreapta. Figura 4 reprezintă descrierea interactivă a scenariului. Fluxul de prelucrare a datelor, care se realizează prin intermediul platformelor externe (Figura 3), este descris în cele ce urmează.

Operatorul Atmospheric correction este reprezentat sub formă de cerc, ceea ce înseamnă că acesta conține doar algoritmul de bază prin care se realizează metoda de corecție a imaginilor satelitare. Primește ca și intrare o imagine satelitară și generează ca și rezultat imaginea corectată.

Nodul Mosaic conține un graf imbricat, motiv pentru care este reprezentat prin intermediul unui dreptunghi. Pe baza a două imagini satelitare, realizează alipirea acestora într-o singură imagine. Nodul Enhanced

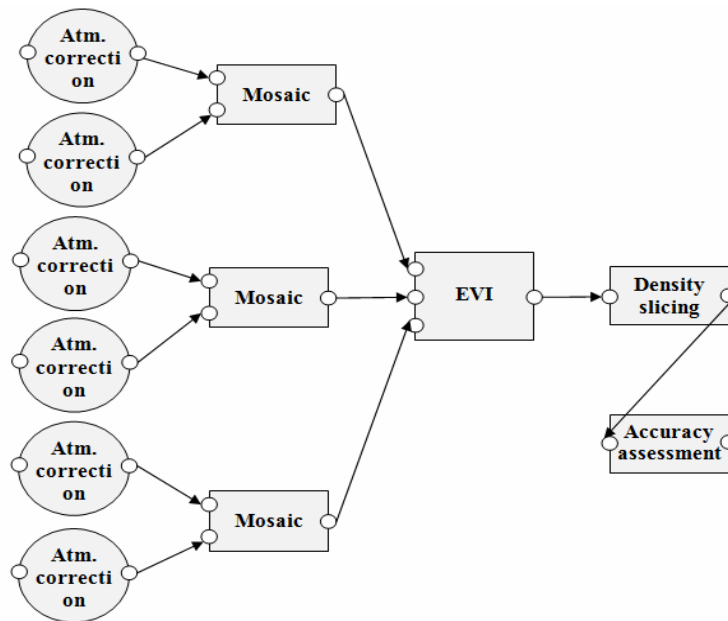


Figura 4. Dezvoltarea scenariilor complexe folosind aplicația WorkflowEditor

Vegetation Index (EVI) se folosește pentru a calcula indicele de vegetație prin intermediul unei formule matematice, care implică prelucrarea datelor din trei imagini satelitare, primite ca și intrare. Rezultatul produs este o imagine alb-negru.

Pentru o mai bună vizualizare a regiunilor clasificate, se aplică algoritmul de pseudo-colorare (Density slicing). Implicit se folosește culoarea verde pentru zonele de vegetație, culoarea albastră pentru regiunile de apă și cu roșu se reprezintă zonele urbane. Nodul Accuracy assessment produce ca și rezultat diferite date statistice referitoare la performanța algoritmului de clasificare. Pe baza acestor date, utilizatorul-specialist poate să valideze rezultatul, sau să aplice o nouă prelucrare a scenariului.

Rezultatele obținute în urma prelucrării algoritmilor sunt prezentate în

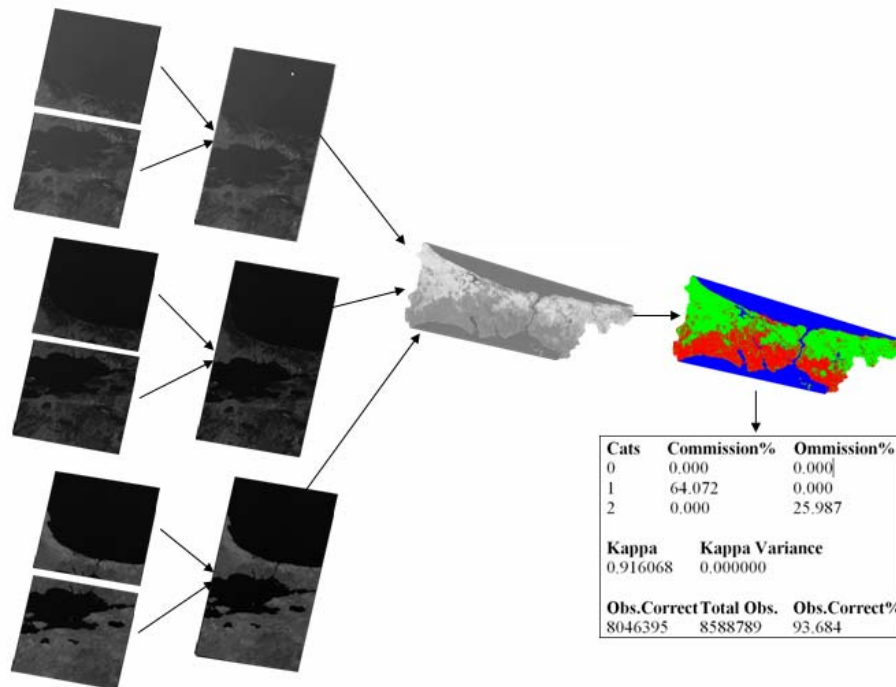


Figura 5. Rezultatele prelucrării algoritmilor din cadrul studiului de caz din regiunea Istanbul



Figura 5. Imaginile satelitare conțin informații despre anumite zone geografice. În funcție de unghiul de vizualizare a senzorului (dispozitivului) de măsurare și a proiecției regiunii geografice, imaginea satelitară prezintă anumite deformări spațiale (cum este cazul celor din Figura 5).

### 4.3 Funcționalitățile aplicației

Interfața grafică a aplicației WorkflowEditor este prezentată în Figura 6. Momentan, există un set de operatori de bază din domeniul de prelucrare a imaginilor satelitare, care prin intermediul editorului de grafuri poate fi extins și la alte domenii.

Suprafața activă reprezintă zona centrală a interfeței (Figura 6) și conține nodurile și arcele scenariului definit prin intermediul grafurilor. Intrările sunt marcate prin elemente grafice de tipul cerc și sunt vizibile în partea din stânga a nodului. Ieșirile sunt poziționate în partea dreaptă, fiind reprezentate tot sub formă de cerc. Conform conceptelor teoretice descrise

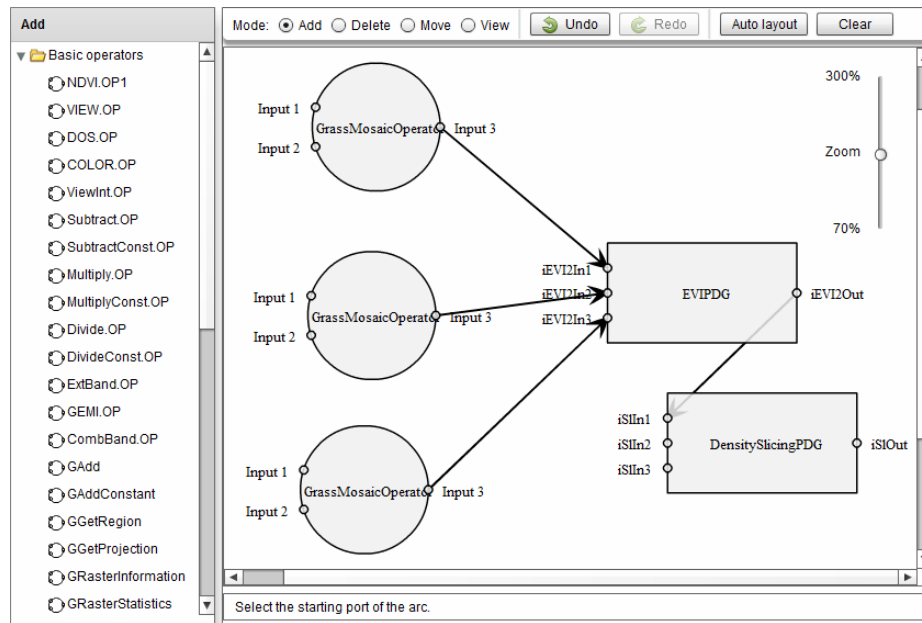


Figura 6. Interfața grafică a aplicației WorkflowEditor

anterior, ieșirile nodurilor sunt conectate cu intrările altor noduri prin arce uni-direcționale, fără a exista posibilitatea de a defini grafuri ciclice.

Aplicația oferă utilizatorilor patru moduri de lucru: adăugare de noi elemente grafice pe suprafața activă, ștergerea elementelor grafice, modificarea poziției nodurilor și arcelor din graf, vizualizarea informațiilor.

### **Adăugarea elementelor grafice**

Prin intermediul listelor de operatori de bază și de grafuri complexe, utilizatorul are posibilitatea de a defini noi scenarii (hipergrafuri). Adăugarea nodurilor în graf se face în doi pași. În primul pas, utilizatorul parcurge lista de algoritmi (operatori) și alege funcția de care are nevoie, prin acționarea butonului stânga al dispozitivului Maus. Urmează apoi specificarea locației de poziționare pe suprafața activă. La fiecare acțiune de adăugare de noi elemente, sistemul verifică regula de suprapunere a nodurilor. În cazul în care această regulă nu este îndeplinită are loc re poziționarea automată a elementului grafic în vecinătatea zonei specificate de utilizator.

### **Specificarea conexiunilor**

După selectarea punctului de început al unui arc, sistemul activează un algoritm de determinare și evidențiere a intrărilor compatibile. Pe tot parcursul procesului, linia urmărește deplasarea cursorului mouse. În cazul în care se ajunge în vecinătatea unui nod, se realizează o afișare a intrărilor compatibile, de același tip, atât grafică cât și sub formă de text.

La diferite nivele de redimensionare (scalare) a grafului, selectarea extremităților arcelor devine dificilă. Ca urmare, a fost definită o zonă de selecție extinsă pentru acestea. Chiar dacă cursorul dispozitivului Maus nu a atins intrarea/ieșirea nodului, sistemul este capabil de a deduce intenția utilizatorului și de a realiza conexiunea dorită.

### **Ștergerea elementelor grafice**

În cazul în care utilizatorul dorește să modifice structura hipergrafului, trebuie să comute pe modul de lucru *Delete* (Figura 6). Fiecare nod sau arc al structurii poate fi șters prin acționarea butonului stânga al dispozitivului Maus. Opțiunile de revenire înapoi (undo) și revenire înainte (redo) pot fi de asemenea folosite pentru a modifica structura hipergrafului.

### **Poziționarea manuală a elementelor**

Resursele sistemului sunt disponibile la nivelul interfeței sub formă de liste. Utilizatorul are posibilitatea de a selecta o resursă și de a o plasa în suprafața de lucru activă, la poziția (x, y), prin acționarea butonului stânga al dispozitivului maus. Modificarea poziției inițiale și operațiile de revenire înapoi și revenire înainte (undo, redo) sunt acțiuni valide în cadrul aplicației WorkflowEditor.

### **Poziționarea automată a elementelor**

Folosind algoritmi de afișare descriși în capitolul 3, Concepte teoretice, sistemul oferă posibilitatea de ordonare automată a elementelor grafului. În acest mod, se asigură satisfacerea unui număr cât mai mare de estetici de vizualizare definite anterior. Este important de menționat faptul că performanța acestor algoritmi este în strânsă legătură cu complexitatea structurii de date. În urma experimentelor s-a observat o relație invers proporțională între numărul de elemente grafice și gradul de satisfacere a esteticilor de minimizare a numărului de intersecții dintre arce și, uniformizare a lungimii arcelor și a rezoluției unghiulare.

### **Vizualizarea nodurilor**

În cazul în care nodul conține grafuri imbricate, există posibilitatea de a vizualiza conținutul nodurilor într-o altă suprafață de lucru. Sistemul poate gestiona simultan mai multe astfel de zone active, iar utilizatorul poate naviga între ele prin intermediul tastelor direcționale și a dispozitivului maus.

Pentru nodurile de tip operator (care conțin abstractizarea unui algoritm sau a unei funcții matematice), vizualizarea informațiilor presupune descrierea în detaliu a structurii algoritmului, a tipurilor de intrări și ieșiri, precum și a legăturilor dintre ele. În cazul grafurilor de dimensiuni mari este necesară optimizarea procesului de afișare a informațiilor, similară cu cea propusă de Grigoriu&Buraga (2011).

## 5. Mulțumiri

Această lucrare a beneficiat de suport financiar prin proiectul ”Creșterea calității studiilor doctorale în științe inginerești pentru sprijinirea dezvoltării societății bazate pe cunoaștere”, contract: POSDRU/107/1.5/S/78534, proiect cofinanțat din Fondul Social European prin Programul Operațional Sectorial Dezvoltarea Resurselor Umane 2007-2013, precum și prin proiectul FP7 enviroGRIDS, cofinanțat de Comisia Europeană, prin contractul 226740.

## 6. Concluzii

Lucrarea de față prezintă conceptele teoretice și practice care au stat la baza dezvoltării aplicației WorkflowEditor și o analiză a soluțiilor experimentale pentru realizarea unei utilizabilități ridicate a tehnicilor de interacțiune cu utilizatorul din interfața grafică.

Aplicația WorkflowEditor facilitează modelarea interactivă a scenariilor complexe, provenite din fenomene naturale, sau din simulări ale acestora. Dificultățile care intervin în realizarea acestor scenarii au fost subliniate în capitolul 3, Concepte teoretice, oferind diferite soluții de rezolvare.

De asemenea, au fost evidențiate tehnicile de interacțiune cu utilizatorul în contextul descrierii funcționalităților complexe prin intermediul metaforelor vizuale, care permit folosirea sistemului atât de către utilizatorii neinițiați cât și de către experți în domenii precum Știința Pământului, matematică, infrastructuri de prelucrare distribuită și paralelă a datelor spațiale.

Descrierea diagramatică a scenariilor implică diferite aspecte legate de dezvoltarea, editarea, gestionarea și, vizualizarea hipergrafurilor realizate din operatori de bază, grafuri și arce. Punerea la dispoziție a unui mediu colaborativ de lucru, posibilitatea de reutilizare a resurselor, asistența oferită de sistem în diferite situații de utilizare, personalizarea elementelor grafice, reprezintă câteva dintre facilitățile oferite utilizatorilor, care au ca și scop dezvoltarea scenariilor complexe de prelucrare a datelor spațiale.

## Referințe

Baur, M., Brandes, U. (2004) Crossing Reduction in Circular Layouts. Proc. Graph-

- Theoretic Concepts in Computer-Science, 332-343.
- Bektaş, B.F., Goksel, C., Sozen, S., Allenbach, K., Gvilava, M., Rahman, K., Gorgan, D., Mihon, D. (2012) Remote Sensing Services – ESIP Platform and Hot Spot Inventory Case Studies. enviroGRIDS Deliverable D2.11, [http://envirogrids.net/index.php?option=com\\_jdownloads&Itemid=13&view=finish&cid=139&catid=11](http://envirogrids.net/index.php?option=com_jdownloads&Itemid=13&view=finish&cid=139&catid=11)
- Cao, J., Jarvis, S.A., Saini, S., Nudd, G.R. (2003) GridFlow: Workflow Management for Grid Computing. In 3rd International Symposium on Cluster Computing and the Grid (CCGrid), IEEE CS Press, Tokyo, Japan, Los Alamitos, 198-206
- Colceriu, V., Gorgan, D. (2011) Tehnici de interacțiune utilizator în prelucrarea pe Grid a imaginilor satelitare. *Revista Română de Interacțiune Om-Calculator*, 4(număr special), 95-98
- Dârdală, M., Reveiu, A., Felix, F.T. (2011) Model de realitate virtuală integrat în sisteme informatice geografice. *Revista Română de Interacțiune Om-Calculator*, 4(număr special), 115-118
- Deelman, E., Blythe, J., Gil, Z., Kesselman, C., Mehta, G., Patil, S., Su, M. H., Vahi, K., Livny, M. (2004) Pegasus: Mapping Scientific Workflow onto the Grid. Across Grids Conference, Nicosia, Cyprus, 11-20
- Di Battista, G. Eades, P., Tamassia, R., Tollis, I. G. (1999) Graph Drawing: algorithms for the visualization of graphs. Prentice-Hall.
- Didimo, W., Liotta, G., Romeo, S. A. (2011) Topology driven force-directed algorithms. *International Symposium on Graph Drawing of Lecture Notes in Computer Science, Springer*, 6502, 165–176.
- Eiglsperger, M., Fekete, S., Klau, G. (2001), Orthogonal graph drawing. *Drawing Graphs, Lecture Notes in Computer Science*, 121–171.
- Eriksson, M., Hallberg, V. (2011) Comparison between JSON and YAML for data serialization. *Bachelor's thesis in Computer Science*, Royal Institute of Technology.
- Garey, M. R., Johnson, D. S. (1983) Crossing number is np-complete. *Siam Journal On Algebraic And Discrete Methods*, 4(3), 312-316.
- Grigoriu, A., Buraga, S. C. (2011) Studii asupra vizualizării datelor și infograficelor. *Revista Română de Interacțiune Om-Calculator*, 4(număr special), 25-28
- GXML, Descriere și elemente de bază. (2012) <http://xml.coverpages.org/gxml.html>
- Mihon, D., Colceriu, V., Minculescu, A., Bacu, V., Gorgan, D. (2012) Descrierea diagramatică a prelucrării imaginilor satelitare în aplicația GreenLand. *Revista Română de Interacțiune Om-Calculator*, 5(2), 129-134.
- Oinn, T., Addis, M., Ferris, J., Marvin, D., Senger, M., Greenwood, M., Carver, T., Glover, K., Pocock, M.R., Wipat, A., Li, P. (2004) Taverna: a Tool for the Composition and Enactment of Bioinformatics Workflows. *Bioinformatics*, Oxford University Press, 20, 3045-3054.
- Pautasso, C., Zimmermann, O., Leymann, F. (2008) RESTful Web Services vs. Big Web

Services: Making the Right Architectural Decision. *17<sup>th</sup> International World Wide Web Conference (WWW2008)*, 804-814.

Purchase, H., Hoggan, E. (2007) How important is the mental map? An empirical investigation of a dynamic graph layout algorithm. *Graph Drawing, ser. Lecture Notes in Computer Science*, 4372, 184-195.

XML, Descriere și elemente de bază. (2012)

<http://www.w3.org/TR/REC-xml/>