

# VIZUALIZAREA DATELOR MASIVE ȘI VISUAL ANALYTICS. ABORDĂRI ȘI TENDINȚE

Mihnea Horia VREJOIU

Mădălina Cornelia ZAMFIR

Vladimir FLORIAN

mihnea@dossv1.ici.ro

madalina@ici.ro

vladimir@ici.ro

Institutul Național de Cercetare-Dezvoltare în Informatică - ICI București

**Rezumat:** În zilele noastre, în tot mai multe domenii, sunt colectate și stocate volume uriașe de date care cresc continuu. Posibilitatea utilizării efective a acestora, prin extragerea informației utile conținute, devine o provocare din ce în ce mai mare. Domeniul *visual analytics* (VA) caută să pună la dispoziția oamenilor mijloace mai potrivite și mai eficiente pentru analiza și înțelegerea acestor seturi mari de date, permițându-le astfel să decidă și să acționeze în mod corespunzător în timp real. Lucrarea de față prezintă o imagine generală asupra a ceea ce numim astăzi *visual analytics*, cu abordările și tendințele specifice. Sunt acoperite aspecte privind integrarea tehnicilor de vizualizare a datelor masive, a tehnologiilor de management și metodelor automate de explorare și analiză avansată a acestora, cu implicarea interactivă decisivă a utilizatorului uman în lanțul de acțiuni și procesări specifice *visual analytics* pentru descoperirea și extragerea de informații și cunoștințe din date, utilizabile în elaborarea și formularea de evaluări, suport pentru decizie și acțiune imediată. Sunt redată sintetic ideile principale referitoare la definiția, metoda / procesul și componentele *visual analytics* rezultate din proiectul VisMaster, acțiune de coordonare din cadrul programului TVE al PC7 pentru C-D al UE. Sunt trecute în revistă câteva din domeniile majore de aplicație. De asemenea, este prezentată situația actuală, prin câteva produse open source și comerciale și este rezumată o perspectivă asupra abordărilor și tendințelor în VA, cu probleme, provocări, oportunități și potențiale soluții și dezvoltări așteptate. În final, sunt formulate câteva concluzii.

**Cuvinte cheie:** vizualizare date, date masive, analiză avansată vizuală, explorare vizuală date, gestiune date.

**Abstract:** In our days, in more and more areas, huge volumes of data – that are continuously growing – are collected and stored. The possibility of effectively using them, by extracting the useful contained information, becomes an increasing challenge. The *visual analytics* (VA) field tries to provide better and more effective means for humans to analyze and understand such huge data sets, thus enabling them to decide and act accordingly in real time. This paper presents an overview on what we are calling today *visual analytics*, with specific approaches and trends. Aspects regarding the integration of Big Data visualization techniques, data management technologies, automated methods for data mining and analytics, with human user interaction for steering the chain of processings and actions specific to *visual analytics* to find and extract useful information and knowledge for evaluations that may support immediate decisions and actions, are covered. The main ideas regarding the definition, method / process, and components of the *visual analytics*, resulted from the VisMaster Project, a Coordination Action in the FET program of EU's FP7 for R&D, are briefly reproduced. Some of the major application areas are reviewed. Also, the current status is presented, through some open source and commercial products, and an overview of the approaches and trends in VA, with problems, challenges, opportunities and potential solutions, is summarized. Some conclusions are finally presented.

**Keywords:** Data Visualization, Big Data, Visual Analytics, Visual Data Mining, Data Management.

## 1. Introducere

În prezent, blocajele și piedicile în calea progresului științific nu mai apar în zona achiziției de date, ci în posibilitățile de analiză a acestora. Datorită furnizării continue de date diverse de la tot mai multe rețele de senzori, telescoape, experimente științifice și instrumente cu rate înalte de transfer, seturile de date disponibile pentru cercetare cresc exponențial. Tehnicile și tehnologiile clasice de stocare și analiză nu mai pot opera eficient în acest nou context. Provocările se mută din zona captării de date și a activităților de gestionare și întreținere a lor, spre analiza și vizualizarea acestora. Existența seturilor de date masive (*Big Data*) reprezintă atât o oportunitate, deoarece sunt posibile perspective mai profunde când volumul datelor disponibile este

mai mare, cât și o provocare, deoarece de multe ori metodologiile actuale nu sunt adecvate gestionării seturilor extrem de mari de date, fiind deci necesară găsirea de noi soluții.

Domeniul *visual analytics* (VA) caută să pună la dispoziția oamenilor mijloace mai bune și mai eficiente pentru analiza și înțelegerea seturilor mari de date pe baza tehnicilor de vizualizare și explorare pentru extragerea informației și a cunoștințelor din acestea, permițându-le să ia decizii și să acționeze imediat în baza celor descoperite, în timp real. *Visual analytics* integrează capacitățile analitice ale calculatorului cu abilitățile analistului uman, oferindu-i acestuia posibilitatea să preia controlul procesului analitic și facilitând descoperiri și/sau perspective noi, surprinzătoare.

Lucrarea de față prezintă o imagine generală asupra a ceea ce numim astăzi *visual analytics*, cu abordările și tendințele specifice și detalierea raportului între vizualizarea datelor masive (Big Data) și *visual analytics*. Sunt acoperite aspecte privind împletirea tehnicilor de vizualizare a datelor masive cu tehnologiile de management și metodele automate de explorare și analiză avansată a acestora, cu implicarea decisivă a utilizatorului uman în lanțul de acțiuni și procesări specifice *visual analytics* pentru descoperirea și extragerea de informații și cunoștințe „ascunse” în date, utilizabile în elaborarea și formularea de evaluări și concluzii suport pentru decizie și acțiune imediată.

În continuarea lucrării, în secțiunea 2 sunt redate sintetic câteva dintre ideile principale referitoare la definiția, metoda / procesul și componentele *visual analytics*, rezultate din proiectul VisMaster – acțiune de coordonare din cadrul programului Future and Emerging Technologies (FET), al Programului Cadru 7 al Uniunii Europene pentru Cercetare și Dezvoltare Tehnologică [1]. În secțiunea 3 sunt trecute în revistă câteva domenii de aplicație pentru *visual analytics*. Secțiunea 4 prezintă o imagine asupra situației actuale prin câteva produse *open source* și comerciale de VA. Secțiunea 5 rezumă o perspectivă asupra abordărilor și tendințelor în VA, probleme, provocări, oportunități și potențiale soluții și dezvoltări așteptate, conform [2]. Lucrarea se încheie cu o secțiune de concluzii.

## 2. Vizualizare și *Visual Analytics*

De cele mai multe ori, datele brute nu au o valoare în sine, pentru majoritatea aplicațiilor trebuind extrasă cumva informația conținută în acestea. Astăzi, achiziția datelor brute nu mai reprezintă o problemă, dimpotrivă, ajungându-se adesea la ceea ce se numește problema supraîncărcării cu informație (*information overload*). Apare riscul ca aceasta să fie în mare parte irelevantă pentru o anumită sarcină, sau să fie procesată sau prezentată nepotrivit, sau să rămână nevalorificată deși ar fi utilă.

Datele brute trebuie să fie transformate, prin aplicarea unor metode și modele adecvate, în informații și cunoștințe semnificative, inteligibile, utilizabile. Pentru aceasta este necesar să se stabilească:

- cine / ce definește relevanța informației pentru o anumită sarcină;
- cum pot fi identificate procedurile adecvate / neadecvate într-un proces complex de luare a deciziilor;
- cum poate fi prezentată informația rezultată într-un mod orientat pe decizie sau orientat pe sarcină;
- ce tipuri de interacțiuni pot facilita rezolvarea problemei și luarea deciziilor.

Țelul *visual analytics* este de a transforma supraîncărcarea cu informație într-o oportunitate, utilizând vizualizarea pentru a face modul de procesare a datelor și informațiilor transparent pentru un discurs analitic. Prin vizualizarea acestor procese se oferă posibilitatea examinării nu numai a rezultatelor, ci și a proceselor efective, furnizând și mijloace de comunicare despre acestea în loc de a rămâne în final doar cu rezultatele. *Visual analytics* facilitează astfel evaluarea constructivă, corecția și îmbunătățirea rapidă a proceselor și modelelor și, în final, extragerea cunoștințelor din date și luarea deciziilor bazate pe acestea.

*Visual analytics* furnizează tehnologii care permit împletirea capacităților analitice umane cu cele computaționale în procesarea datelor. Vizualizarea capătă noi valențe, devenind mediul unui proces analitic semiautomat, interactiv, în care oameni și calculatoare cooperează fiecare cu capacitățile lor distincte pentru obținerea celor mai bune rezultate. Utilizatorul rămâne autoritatea decisivă în direcționarea analizei. Sistemul trebuie să ofere mijloacele efective de interacțiune pentru focalizarea sarcinilor specifice. În multe aplicații, în decursul lanțului de procesări de la date la decizie pot participa mai multe persoane. O reprezentare vizuală va schița acest lanț și va furniza o referință pentru colaborarea acestora de-a lungul diferitelor sarcini și la diferite niveluri de detaliere. Cercetarea *visual analytics* este în esență interdisciplinară, combinând diferite domenii de cercetare relaționate, cum sunt: vizualizarea, explorarea datelor (*data mining*), managementul datelor (*data management*), fuzionarea datelor (*data fusion*), statistica și știința cognitivă (*cognitive science*), interacțiunea om-calculator (*human-computer interaction* - HCI), între altele. Deoarece *visual analytics* este o disciplină integratoare, arile

de cercetare specifice aplicațiilor pot contribui la procedurile și modelele existente. Comunitățile de cercetare provenind din zone foarte orientate pe aplicație, lucrează la soluții specifice utilizând instrumente și standarde specifice domeniului respectiv. Cerințele *visual analytics* introduc noi dependențe între aceste domenii. Integrarea disciplinelor amintite mai sus în *visual analytics* va genera un set de concepte și teorii bine stabilite și general acceptate, permițând oricărei evoluții dintr-o disciplină să aibă impact asupra întregului domeniu al *visual analytics*.

## 2.1 Definiția *Visual Analytics*

Tehnicile de analiză automată, cum sunt statistica și explorarea datelor (*data mining*), s-au dezvoltat independent de tehnicile de vizualizare și interacțiune. Astăzi, în cercetarea *visual analytics*, s-a trecut de la analiza confirmatorie a datelor, utilizând vizualizări exclusiv pentru prezentarea rezultatelor ipotezelor / modelelor generate, la analiza exploratorie a datelor, implicând interacțiunea utilizatorului cu datele și rezultatele.

Odată cu îmbunătățirile interfețelor grafice utilizator și a dispozitivelor de interacțiune, au fost dedicate eforturi importante vizualizării informației. A fost facilitată astfel integrarea utilizatorului în procesele de descoperire a cunoștințelor și de explorare a datelor (*data mining*) prin tehnici de vizualizare eficiente și eficiente, capabilități de interacțiune și transfer de cunoștințe. Aceasta a condus la ceea ce se numește explorarea vizuală a datelor (*visual data mining*).

Termenul *visual analytics* este utilizat astăzi într-un context mai larg, descriind un nou domeniu multidisciplinar, care combină diferite arii de cercetare, incluzând: vizualizarea, interacțiunea om-calculator, analiza datelor, managementul datelor, procesarea datelor geospațiale și temporale, suportul pentru decizia spațială și statistica.

O definiție inițială pentru *visual analytics* a fost: „știința raționamentelor analitice facilitate de interfețe vizuale interactive” [3]. O definiție mai specifică poate fi: „*visual analytics* combină tehnici de analiză automată cu vizualizări interactive pentru înțelegerea efectivă, elaborarea de raționamente și luarea deciziilor pe baza seturilor de date masive și complexe” [4].

Scopul *visual analytics* este acela de a permite utilizatorilor, prin instrumente și tehnici specifice, să:

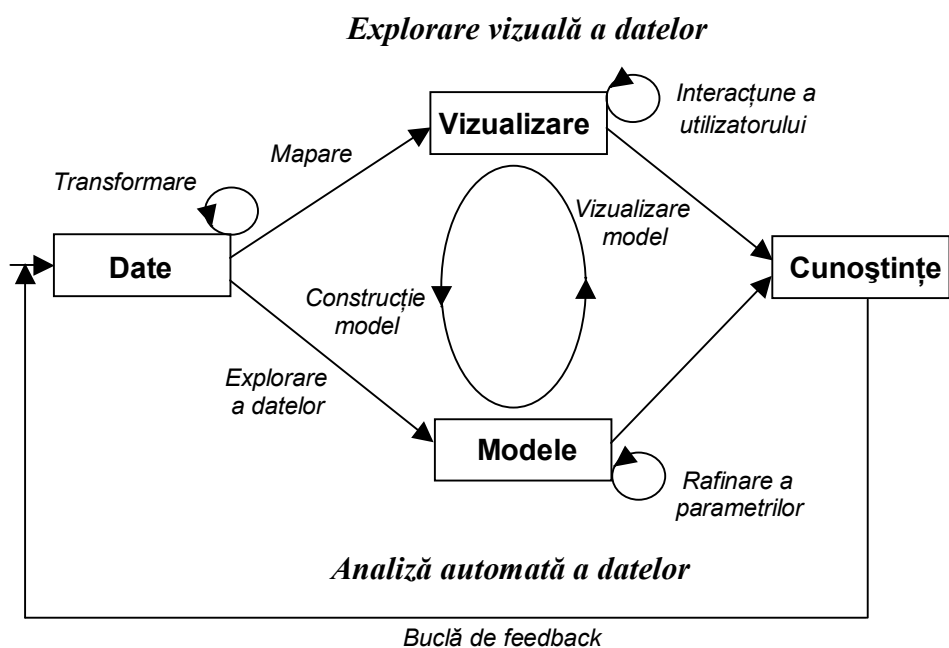
- sintetizeze informația și să deducă perspective din date masive, dinamice, ambigue și adesea conflictuale (modelare);
- detecteze ce era de așteptat (validare) și să descopere lucruri (modele) surprinzătoare;
- furnizeze evaluări (pe baza ipotezelor / modelelor generate) în timp util, sustenabile și inteligibile;
- comunice efectiv aceste evaluări.

## 2.2 Procesul *Visual Analytics*

Procesul *visual analytics* combină metode de analiză automată și vizuală, înlănțuite prin interacțiunea umană, pentru extragerea cunoștințelor din date masive. Figura 1 prezintă diferitele etape ale unui proces *visual analytics* și tranzițiile între acestea.

În multe scenarii de aplicație, sunt utilizate surse de date eterogene care trebuie integrate înainte de a putea fi aplicate metode vizuale sau automate. Prin urmare, primul pas este adesea reprezentat de preprocesarea și transformarea datelor pentru obținerea de diferite reprezentări pentru explorări ulterioare. Alte sarcini de preprocesare includ curățirea datelor, normalizare, grupare (*clusterization*), sau integrarea surselor de date eterogene.

După transformare, analistul poate alege între aplicarea metodelor vizuale sau, respectiv, automate de analiză. Dacă este utilizată mai întâi o metodă automată, sunt aplicate metode de explorare a datelor (*data mining*) pentru generarea modelelor datelor originale. Odată creat un model, analistul trebuie să-l evalueze și să-l rafineze, ceea ce se poate realiza cel mai bine prin interacțiune cu datele. Vizualizările îi permit să interacționeze apoi cu metodele automate prin modificarea parametrilor sau prin selectarea altor algoritmi. Vizualizarea poate fi apoi utilizată din nou pentru evaluarea constatărilor modelelor generate. Alternarea între metode vizuale și automate este caracteristică pentru procesul *visual analytics* și conduce la o rafinare continuă și la verificarea rezultatelor preliminarilor. Pot fi astfel descoperite într-o etapă incipientă eventuale rezultate înșelătoare, în pașii intermediari.



**Figura 1. Procesul visual analytics este caracterizat prin interacțiuni între date, vizualizări,**

Dacă este efectuată mai întâi explorarea vizuală, utilizatorul trebuie să confirme apoi printr-o analiză automată ipotezele astfel generate. Interacțiunea utilizatorului cu vizualizarea este necesară pentru a releva informația importantă, de exemplu prin mărire (*zoom*) a diferitelor zone de date, sau prin considerarea unor observări vizuale diferite asupra datelor. Constatările din vizualizări pot fi utilizate pentru a ghida construirea modelului în analiza automată.

În rezumat, în procesul *visual analytics*, cunoștințele pot fi obținute din vizualizare, din analiză automată, precum și din interacțiuni precedente între vizualizări, modele și analistul uman.

Procesul *visual analytics* urmărește împletirea strânsă a metodelor de analiză automată cu reprezentări vizuale interactive. Sintetic, un proces *visual analytics* constă în secvența: 1) – mai întâi analiză; 2) – vizualizare a lucrurilor importante; 3) – *zoom/filtrare*; 4) – analiză mai departe; 5) – detalii la cerere.

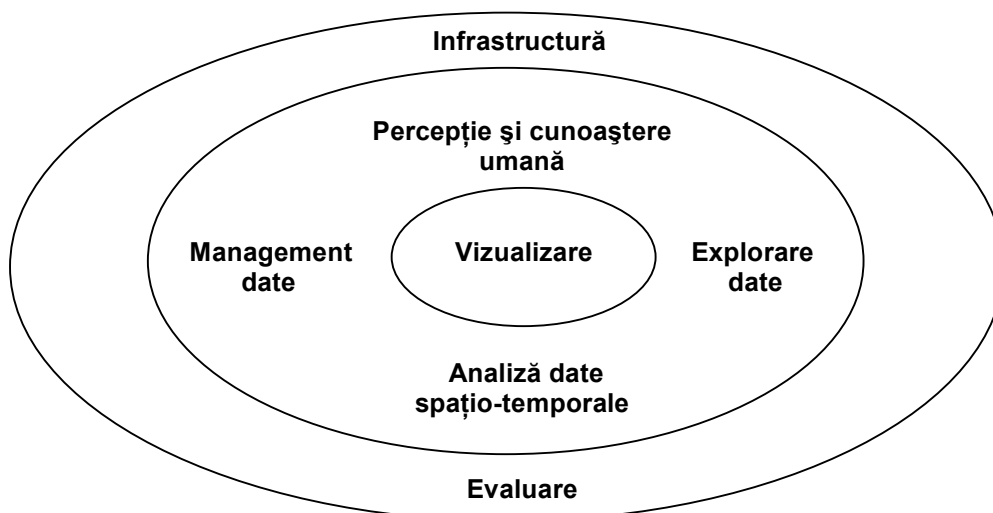
### 2.3 Componentele *Visual Analytics*

În cadrul *visual analytics* sunt integrate știința și tehnologia din mai multe discipline, așa cum este reprezentat în Figura 2.

### Vizualizarea (*visualization*)

Poate fi împărțită grosier în vizualizare științifică și vizualizare a informațiilor. Vizualizarea științifică se ocupă în primul rând de vizualizarea 3-D a datelor din lumea ingineriei, din biologie, meteorologie, astronomie etc., cu scopul reprezentării acestora, care adesea conțin și o dimensiune temporală, ca entități fizice sugestive cum ar fi: suprafețe, volume și fluxuri. Vizualizarea informațiilor a dezvoltat metode pentru vizualizarea datelor abstracte atunci când nu există referințe spațiale explicite. Dintre exemplele tipice enumerăm: datele de business, demografice, științifice și din rețelele sociale. Nu numai că vorbim în astfel de cazuri de volume uriașe de date, dar, de cele mai multe ori și de sute de dimensiuni. De asemenea, pe lângă tipuri de date numerice și de tip text standard, aceste dimensiuni pot fi reprezentate de tipuri de date complexe cum ar fi grafică, video, sunet, precum și alte tipuri sofisticate de date definite pentru web semantic.

Capacitatea de interacțiune cu datele este foarte importantă. Au fost dezvoltate noi vizualizări cum ar fi: coordonate paralele, hărți arborescente, reprezentări vizuale ale datelor pictografice sau bazate pe pixeli etc., împreună cu o varietate de tehnici pentru reducerea densității datelor la afișare.



**Figura 2. Visual analytics integrează vizualizarea cu discipline adiacente și depinde de disponibilitatea unei infrastructuri adecvate și a unor facilități de evaluare (adaptare după [1])**

Există de asemenea tehnici speciale pentru vizualizarea datelor structurate, cum ar fi abordările bazate pe grafuri pentru rețele și pentru vizualizarea dimensiunilor spațiale și temporale în cazul geovizualizării. Mai mult, unele vizualizări utilizează tehnici de analiză automată a datelor (de ex. clusterizare, reducere dimensională) ca preprocesare anterioară vizualizării.

### **Managementul datelor (data management)**

Managementul eficient al datelor de tipuri și calități diferite reprezintă o componentă cheie a *visual analytics*, deoarece furnizează în mod tipic datele de intrare care trebuie analizate. În general, o condiție necesară pentru efectuarea oricărui tip de analiză de date este aceea de a dispune de o bază de date integrată și consistentă. Cercetările în domeniul bazelor de date s-au concentrat până de curând mai cu seamă pe aspecte legate de eficiența și scalabilitatea unor căutări exacte pe date structurate uniforme. Odată cu dezvoltarea Internetului, care a facilitat accesul la diferite surse de date de tipuri variate, accentul în cercetarea din domeniul bazelor de date s-a mutat către integrarea acestor date eterogene.

Găsirea unor reprezentări eficiente pentru diferite tipuri de date cum ar fi date numerice, grafuri, text, semnale audio și video, date semistructurate, reprezentări semantice etc., reprezintă o problemă cheie pentru tehnologiile

moderne de baze de date. Pe de altă parte, existența datelor eterogene nu necesită numai integrarea mai multor tipuri de date și formate, ci necesită, de asemenea, curățirea datelor deoarece apar situații cu valori de date lipsă sau imprecise. Aplicațiile moderne necesită ca astfel de fuzionări inteligente de date să poată fi realizate aproape în timp real și pe cât posibil în mod automat. De asemenea, noi forme de surse de informații cum ar fi sursele de fluxuri de date, rețelele de senzori sau extragerea automată a informației din colecții mari de documente (de ex. text, HTML) conduc la o problemă dificilă de analiză a datelor. Aceasta constituie obiectul focalizării actuale a cercetărilor în domeniul bazelor de date. Tehnicile de management al datelor utilizează din ce în ce mai mult și tehnici de analiză a datelor și de vizualizare pentru optimizarea proceselor și informarea utilizatorului.

### **Explorarea datelor (data mining)**

Disciplina explorării datelor dezvoltă metode computaționale pentru extragerea automată a informației relevante din datele brute prin intermediul algoritmilor de analiză automată. În timp ce analiza datelor s-a dezvoltat inițial pentru date structurate, cercetările recente tind spre analiza unor date semistructurate și complexe, ca documentele web, sau datele multimedia. În aproape toți algoritmi de analiză a datelor trebuie specificați o serie de parametri, ceea ce adeseori nu este tocmai trivial și necesită

supervizarea de către un expert uman. Vizualizarea interactivă poate facilita aceasta și, de asemenea, poate fi utilizată pentru prezentarea rezultatelor analizei automate – așa numita explorare vizuală a datelor.

### **Analiza datelor spațio-temporale (*spatio-temporal data analysis*)**

Datele spațiale sunt date cu legături cu lumea reală, cum ar fi măsurători geografice, date de poziționare GPS și date de la aplicații cu senzori la distanță, care pot fi reprezentate pe o hartă sau diagramă. Găsirea relațiilor spațiale și a tiparelor în aceste date prezintă un interes special și necesită dezvoltarea unor funcții de management, reprezentare și analiză. Datele temporale pe de altă parte, sunt o funcție de timp, valorile variabilelor din date putându-se modifica de-a lungul timpului. Sarcinile importante de analiză în acest caz includ identificarea tiparelor, tendințelor și corelațiilor elementelor de date în timp.

### **Percepția și cunoașterea (*perception and cognition*)**

Percepția și cunoașterea reprezintă latura (mai) umană a *visual analytics*. Percepția vizuală este mijlocul prin care oamenii interpretează lumea înconjurătoare și în acest context, imaginile de pe ecranul unui calculator. Cunoașterea este abilitatea de a înțelege această informație vizuală, de a face deducții / inferențe logice, bazate în mare măsură pe învățări anterioare. Cunoștințele despre cum „gândim vizual” sunt importante în proiectarea interfețelor utilizator și, împreună cu experiența practică din domeniul interacțiunii om-calculator, vor fi de ajutor în crearea de metode și instrumente pentru proiectarea unor tehnici de interacțiune multimodale bazate pe percepție pentru vizualizarea și explorarea spațiilor de informații mari, precum și pentru evaluarea utilizabilității acestui tip de sisteme.

## **3. Domenii de aplicație**

*Visual analytics* (VA) este esențial în domeniile de aplicație unde trebuie procesate și analizate spații vaste de date pentru extragerea de informații și cunoștințe din acestea.

Dintre principalele domenii de aplicație, astrofizica oferă multe oportunități tehnicilor

*visual analytics*: volume masive de date nestructurate, provenind din diferite direcții ale spațiului și acoperind întreg spectrul de frecvențe, de la fluxuri continue de terabytes de date care pot fi înregistrate și analizate. Prin tehnici obișnuite de analiză a datelor, astronomii pot separa datele relevante de zgomot, pot analiza similarități sau modele / șabloane repetitive complexe și obține cunoștințe utile despre univers. Abordarea *visual analytics* facilitează procesul de identificare a fenomenelor neașteptate în fluxurile masive și dinamice de date, care altfel nu ar fi putut fi găsite prin intermediul algoritmilor standard.

Monitorizarea mediului și a climei reprezintă de asemenea domenii care implică volume uriașe de date colectate de senzori de pe întreg globul și din sateliți, continuu, la intervale scurte de timp. O abordare vizuală poate ajuta la interpretarea acestor volume masive de date și la căpătarea unei înțelegeri asupra dependenței factorilor de climă și a scenariilor de modificare a climei, care altfel nu ar fi putut fi ușor identificate. În afară de prognoze ale vremii, aplicațiile actuale vizează încălzirea globală, topirea gheții la poli, subțierea stratului de ozon din stratosferă, avertizările de uragane și tsunamiuri.

În domeniul managementului situațiilor de urgență, *visual analytics* poate fi de ajutor în determinarea evoluțiilor și identificarea măsurilor ce trebuie luate pentru limitarea pagubelor (de ex. necesitatea unor construcții fizice sau a evacuării populației). Astfel de scenarii pot include catastrofe naturale sau meteorologice cum ar fi inundații sau valuri mari, activitate vulcanică, furtuni, incendii sau epidemii, dar și catastrofe tehnologice generate de om, cum ar fi accidente industriale, de transport sau poluare. *Visual analytics* pentru securitate și geografie reprezintă de asemenea domenii importante de cercetare. Aplicațiile pot acoperi soluții informatice antiterorism, supravegherea și protecția granițelor, detectarea traseelor (*path detection*), sau securitatea rețelelor informatice.

*Visual analytics* permite investigarea și detectarea similarităților și/sau a anomaliilor în seturi foarte mari de date. De exemplu, conform [1], la nivel european, Europe Media Monitor (<https://ec.europa.eu/jrc/en/scientific-tool/europe-media-monitor-newsexplorer>) colectează documente din câteva mii de surse

de știri: portaluri media, site-uri web guvernamentale și agenții de știri și procesează sute de mii de articole pe zi, în 43 de limbi. În biologie și medicină, tomografia computerizată și imagistica cu ultrasunete pentru reconstrucția și vizualizarea digitală 3-D produce *gigabytes* de date medicale. Domeniul de aplicații bioinformatică utilizează tehnici *visual analytics* pentru analiza unor mari cantități de date biologice (de ex. cazul proiectului genomului uman, cu 3 miliarde de perechi de bază per individ).

#### 4. Situația actuală

Sistemele pentru cercetare și comerciale în domeniul explorării și vizualizării datelor disponibile astăzi pot fi clasificate astfel:

- instrumente statistice și matematice;
- instrumente algoritmice specifice;
- biblioteci pentru visual analytics;
- instrumente pentru explorare vizuală a datelor;
- instrumente și pachete web;
- instrumente pentru vizualizare științifică;
- metode combinate;
- proiectare orientată pe informații computaționale.

Există mai multe tehnici și seturi de instrumente (*toolkits*) pentru *visual analytics* (VA) de tip *open source* care au fost dezvoltate în mediul academic [5]. Pe partea comercială [6], de asemenea, un larg spectru de companii, de la cele specializate pe descoperirea datelor, ca: Tableau, QlikTech și TIBCO, la corporații internaționale ca: IBM, Microsoft, Oracle și SAP, a dedicat eforturi importante pentru cumpărarea și/sau dezvoltarea unor produse proprii de analiză dedicate Big Data. De ex.: IBM a cumpărat Cognos, Oracle a achiziționat Siebel și Hyperion, SAP a cumpărat Business Objects, iar TIBCO, Spotfire, în timp ce SAS a dezvoltat JMP, iar Microsoft, Sharepoint, Teradata și PowerPivot, dar și SQL Server BI, care au devenit produse cheie pe piață. Dintre seturile de instrumente *open source*, fiecare acoperind un set specific de funcționalități de vizualizare, analiză și interacțiune, enumerăm: InfoVis Toolkit [7], Prefuse [8], Improvise [9] și JUNG [10]. Aceste seturi de instrumente pot

fi utilizate în dezvoltarea de aplicații VA noi. O alternativă poate fi utilizarea de sisteme software comerciale, de sine stătătoare, sau integrate în infrastructuri informatice existente, cum sunt, de ex.: Tableau [11], Spotfire [12], QlikView [13], JMP (SAS) [14], ADVIZOR [15], Centrifuge [16] și Visual Analytics [17].

În ceea ce privește managementul datelor, funcționalitățile sistemelor comerciale VA pot fi împărțite în următoarele categorii: 1) de management al datelor (încărcare, integrare, opțiuni de export sau prezentare); și 2) de manipulare a datelor (preprocesare, curățare, transformare, aspecte de utilizabilitate). Din punct de vedere al funcționalităților de management al datelor, majoritatea sistemelor de VA permit conectarea la sisteme de baze de date relaționale (SQL, PostgreSQL, Oracle). Puține însă permit acces la sisteme de stocare scalabile vertical (Hadoop, Vertica – orientate coloană, Mongo-DB – orientat document), sau la sisteme de baze de date la cerere, bazate pe web (Amazon S3 și Salesforce Database System – Non-SQL și orientate obiect). Din punct de vedere al datelor de intrare structurate și/sau nestructurate suportate, majoritatea sistemelor și instrumentelor de VA suportă formate ca: text brut (CSV) și Microsoft Excel, dintre cele mai utilizate. Totodată, majoritatea sistemelor permite accesul simultan la mai multe surse de date și oferă posibilități de procesare pentru unificarea datelor și consolidarea surselor respective. Pe de altă parte, scrierea rezultatelor înapoi în baza de date este un mecanism rar implementat (de ex. de Tableau, JMP și Visual Analytics). Prezentarea rezultatelor se face de regulă prin panouri de control (*dashboards*), interactive, fie găzduite local (pe servere securizate), fie pe serverele furnizorului soluției VA, prin accesare *web site* HTML sau Adobe Flash. Mobilitatea este de asemenea un aspect actual foarte important pentru sistemele de VA comerciale. De ex., Tableau, Spotfire, QlikView și JMP oferă posibilitatea accesării panourilor de comandă interactive de prezentare pe tablete sau telefoane inteligente, prin protocol HTML5. După încărcare, adesea este necesar un pas de curățare și transformare a datelor (de ex. pentru gestionarea situațiilor valorilor lipsă sau nule, normalizarea datelor după una sau mai multe dimensiuni etc.). Majoritatea sistemelor comerciale permit opțiuni pentru manipularea datelor cu ajutorul unor limbaje de expresii proprietare (de ex. VizQL în cazul Tableau).

Dintre principalele operațiuni de manipulare a datelor, comune sistemelor VA, menționăm: calcularea de coloane, din linii ale bazelor de date, combinarea liniilor sau coloanelor într-una singură, cuplări (*joins*) sau cuplări pe tabele filtrate. Majoritatea sistemelor de VA comerciale au dificultăți în cuplarea tabelelor filtrate după opțiunile utilizatorului. Acesta trebuie să exporte tabelele filtrate și să le reîncarce din fișiere, realizând astfel operațiunea de cuplare într-un pas intermediar.

Tehnicile și metodele de analiză automată utilizate de sistemele VA pot fi grupate în patru categorii principale: 1) statistice; 2) modelarea datelor; 3) reducerea dimensionalității; 4) explorare vizuală. Prima categorie include funcții statistice pentru analiza datelor uni-, bi-, sau multi-dimensionale, de ex.: calcul medie, minim, maxim, deviație standard; sau corelații și/sau coeficienți de corelare de diverse tipuri; sau, respectiv, analiză discriminantă sau a varianței. Ultimul tip este mai rar suportat (de ex. de Spotfire, JMP și ADVIZOR). Metodele din a doua categorie permit modelarea datelor și identificarea de tipare prin utilizarea de algoritmi de explorare a datelor (*data mining*), de ex. clusterizare, clasificare, tehnici de modelare de rețea (grafuri), tehnici de modelare predictivă. A treia categorie, cuprinde tehnici de reducere dimensională pentru reducerea datelor multi-dimensionale la spații cu foarte puține dimensiuni, înainte de analiză sau vizualizare, păstrând însă esența datelor intactă. Rezultatul este utilizat adesea pentru generarea de proiecții 2D sau 3D ale datelor. Tehnicile cele mai utilizate pentru reducerea dimensiunilor sunt: analiza componentelor principale, scalarea multi-dimensională și hărțile cu auto-organizare. Modelarea de rețea este suportată de Visual Analytics și Centrifuge, care suportă de asemenea și analiză de clusterizare pe rețea. JMP și Spotfire acoperă toată gama de funcționalități de modelare și implementează și tehnici de reducere a dimensiunii pentru manipularea datelor cu multe dimensiuni. Căutarea de tipare (*pattern search*) este o altă caracteristică importantă a analizei automate a datelor. Unele sisteme permit definirea tiparului țintă de căutat prin intermediul interfeței grafice utilizator. Aceasta intră în a patra categorie, care furnizează mijloace rapide și intuitive pentru analiza de tipare.

Tehnicile de vizualizare pot fi împărțite în:

1) reprezentări grafice ale datelor și 2) tehnici de interacțiune. Cea dintâi se referă la forma vizuală statică în care sunt afișate grafic datele sau modelele și este cea referită în general ca „vizualizare”. Tehnicile de interacțiune se referă la posibilitățile de interacțiune ale utilizatorului cu modelele grafice statice prin: mărire, panoramare (*zoom, pan*) etc. Toate produsele de VA implementează tehnici de vizualizare standard pentru analiza datelor cu foarte puține dimensiuni, de tip: grafice cu linii, cu bare, de tip „plăcintă” și histogramme. Pentru dimensiuni mai mari sunt utilizate diagrame de dispersie (*scatterplot*), matrici de astfel de diagrame, sau hărți calorice (*heatmap*). Puține sisteme implementează însă vizualizări cu coordonate paralele, care sunt considerate eficiente pentru reprezentarea datelor cu multe dimensiuni, sau funcționalități pentru vizualizarea datelor de tip text. Unele sisteme oferă funcționalități pentru vizualizarea datelor de tip rețea (de ex. QlikView, Spotfire, JMP, Visual Analytics și Centrifuge). Totodată, funcționalități pentru vizualizarea datelor geografice sunt destul de limitate în multe sisteme, deși cele mai multe permit proiecții pe o hartă statică. Tableau și JMP oferă facilități de recomandare, care sugerează ce tipuri de vizualizări se potrivesc datelor de intrare. Pentru majoritatea sarcinilor de analiză vizuală este esențială interacțiunea cu modelele de date și vizualizarea. De exemplu, pentru filtrarea datelor, pentru explorarea pe un subset de dimensiuni sau de elemente de date, pentru mărire și/sau panoramare a „ferestrei” din date vizualizate pentru a observa modelul la diferite niveluri de detaliu, pentru a schimba interactiv focalizarea vizualizării pe o zonă din date sau alta fără a pierde întreaga imagine (tehnici de distorsionare a focalizării și contextului) și de a lega și peria (*link and brush*) diferite vizualizări pentru a vedea datele din diferite perspective. Cele mai multe instrumente suportă astăzi filtrarea și *zoom*-ul interactiv, precum și distorsionarea vizualizării (de ex., scara logaritmică). Posibilitatea de a avea mai multe vederi simultan, conectate prin *link-and-brushing* este o abordare eficientă și un avantaj al instrumentelor respective.

## 5. Abordări și tendințe

Dificultățile inerente asociate cu algoritmi de explorare a datelor (*data mining*) și nevoia de justificare a rezultatelor acestora au stimulat



dezvoltarea de medii integrate în care sunt utilizate vizualizări adecvate ca tehnică complementară pentru asistarea explorării datelor. Combinația între vizualizare și explorarea datelor este cunoscută ca explorare vizuală a datelor (*visual data mining*).

Instrumentele care oferă facilități de căutare, navigare și vizualizare au fost îmbunătățite substanțial în ultima decadă. Au apărut sisteme care permit accesul unui public larg la consultarea vizuală interactivă a unor informații / date utile, cu volume mari și mai multe dimensiuni (de ex. GapMinder – [www.gapminder.org](http://www.gapminder.org)). Asistăm astăzi la o explozie a datelor care promite oportunități nebănuite în progresul științific și tehnologic. Acest potențial se bazează nu numai pe posibilitățile de a colecta și accesa acest volum extraordinar de date, ci și pe capacitatea de a-l înțelege. Din acest punct de vedere, capacitatea noastră de a extrage cunoștințe din datele brute este departe de a-și fi atins limitele. Sunt produse cu mult mai multe date decât suntem în măsură astăzi să explorăm, ceea ce conduce la volume masive de date achiziționate care sunt încă neatinse, așteptând descoperiri și dezvoltări viitoare pentru a putea fi utilizate efectiv. Există două abordări majore pentru descoperirea de cunoștințe în date: fie sunt utilizate metode matematice (de ex. învățarea automată – *machine learning*), fie este utilizată judecata umană prin observarea directă a datelor (de ex. *visual analytics*). Oricum, direct sau indirect, în ambele situații elementul uman este implicat în descoperirea cunoștințelor. Mai mult, în final, acesta este și beneficiarul acestor cunoștințe, cel care le va înțelege și utiliza [2].

Procesul de înțelegere presupune în mod normal crearea unui model mental, pe baza căruia lucrurile capătă un sens. În general, un astfel de proces se realizează de cele mai multe ori intuitiv, prin tehnici convenționale de vizualizare. Dimensiunile uriașe ale datelor multidimensionale forțează granițele capacităților actuale de vizualizare a acestora. Este practic imposibil să fie vizualizate seturi de date care au sute sau mii de dimensiuni. În acest context, se fac astăzi eforturi pentru a se identifica linii directoare de proiectare și instrumente analitice care să permită vizualizarea, explorarea și analizarea seturilor de date mari și complexe, în scopul descoperirii de cunoștințe din datele respective mai repede și mai ușor.

## Descoperirea cunoștințelor, vizualizarea și *Big Data*

Procesul de descoperire al cunoștințelor este un aspect fundamental al științei în general. Un model bogat pentru descrierea acestui proces este descris în [18]. Autorii susțin că cercetătorii navighează într-un spațiu 4-dimensional pentru a extrage sens / înțeles / semnificație din observațiile lor. Prima dimensiune din această paradigmă o constituie reprezentarea datelor. Aici se formează o reprezentare abstractă dintr-un set de caracteristici. A doua dimensiune este spațiul ipotezelor. Aici, cercetătorul generează noi presupuneri privind posibilele relații cauzale. Apoi se trece la a treia dimensiune, a spațiului experimental, pentru testarea acestor ipoteze. Trebuie menționat faptul că experimentele însele se regăsesc într-un cadru experimental care definește limitele experiențelor posibile și ale rezultatelor așteptate. Prin urmare, a patra dimensiune este spațiul paradigmei experimentale, unde cercetătorul poate alege o clasă complet diferită de experimente pentru sarcina sa.

Cu instrumente de *visual analytics*, un proces de descoperire de cunoștințe poate fi modelat pe baza primelor trei dimensiuni. O vizualizare specifică este un exemplu al spațiului reprezentării datelor. Interacțiunea cu datele are loc în spațiul experimental. În fine, opțiunile de vizualizare / explorare formează o serie de decizii în spațiul ipotezelor.

Prin utilizarea spațiului vizual pentru reprezentarea datelor, se obține o conexiune tangibilă și mai directă cu datele efective. Prin formarea unei întrebări vizuale, se formează de fapt o ipoteză mentală și când se face o căutare vizuală se experimentează cu datele în scopul confirmării sau rejectării ipotezei. Acest proces a fost explicat în literatură în mai multe moduri. În [19] este prezentat acest proces în șapte pași. Mai întâi trebuie achiziționate datele. Apoi, acestea trebuie separate (*parsed*) și făcute citibile de către computer (*machine readable*). Urmează o filtrare prin care este selectat un subset al datelor care este relevant pentru ceea ce se dorește. Apoi, se caută în acesta informație folositoare, ceea ce implică în mod normal un tip de transformare matematică. Rezultatele sunt apoi reprezentate într-o vizualizare inițială. Urmează rafinarea și pașii finali de interacțiune în care vizualizarea

este explorată și îmbunătățită prin reluarea pașilor anteriori până când sunt extrase sau descoperite cunoștințele dorite.

În domeniul *visual analytics* este avută în vedere o perspectivă mai generală asupra descoperirii cunoștințelor [1][4]. Scopul aici este să se identifice modul în care oamenii înțeleg datele și apoi să se genereze o disciplină algoritmică care să beneficieze atât de puterea tehnicilor de procesare automată, cât și de capacitățile umane în discernerea și analiza modelelor / tiparelor vizuale.

Pe lângă problemele legate de scalare, vizualizarea și înțelegerea seturilor complexe de date cu sute de caracteristici ridică provocări deosebite. Astfel, pe măsură ce dimensiunea datelor crește, latența oricărei căutări începe să pună probleme serioase. O idee poate fi aceea ca, dacă un sistem nu întoarce un răspuns la o întrebare particulară în câteva secunde, atunci analistul să renunțe la acea întrebare și să nu beneficieze de răspunsul la ea. Ar avea loc astfel o deplasare de la instrumente convenționale, precise dar lente, spre o paradigmă de tip „imprecise dar rapide”, ceea ce ar putea depăși problemele de latență din cazul seturilor de date mari și complexe [20]. O altă problemă a sistemelor analitice actuale este reprezentată de lipsa de *feedback* și de ghidare. Interactivitatea este un alt aspect important al sistemelor de vizualizare. Ar trebui să ne așteptăm ca noile instrumente de *analytics* să ofere mecanisme de explorare multimodală responsive în sprijinul extragerii de semnificație / sens, să ofere noi abilități de ghidare pentru explorarea datelor mari și multidimensionale, să se concentreze pe latențe mici de căutare chiar și cu costul unor răspunsuri mai puțin precise, sau imprecise și să asigure posibilitatea interacțiunii prin care utilizatorul să poată modifica parametrii căutării în timpul procesării acesteia.

### **Tratarea dimensiunii: Instrumente orientate ecran (*screen-aware*)**

În timp ce dimensiunile datelor cresc continuu fără o limită previzibilă, abilitățile cognitive umane sunt destul de limitate. Pot fi percepute probabil numai câteva milioane de caracteristici, sau chiar mai puține [19]. Faptul că limitarea în înțelegerea vizualizării datelor mari apare ca fiind datorată capacității umane, a condus la dezvoltarea unei noi clase de

soluții, care se focalizează pe datele de ieșire în loc de cele de intrare. Astfel de instrumente orientate ecran utilizează diferite niveluri de abstractizare a datelor pentru a reduce dimensiunea informației prezentate și pentru a evita analiza datelor care ies din suprafața ecranului. Acestea utilizează apoi mecanisme interactive și exploratorii pentru a veni în sprijinul utilizatorului spre a naviga prin vizualizări și a înțelege datele mai bine. Aceste instrumente se bazează pe ideea că în vizualizarea Big Data, detaliile pot fi adesea neglijate. Astfel, de exemplu, la vizualizarea a milioane de puncte, de cele mai multe ori relevantă este forma generală a vizualizării și nu poziția exactă a fiecărui punct. Eventual, mai poate fi de interes posibilitatea de interacțiune cu vizualizarea în sensul scalării / panoramării (*zoom / pan*) și/sau glisării ferestrei de date observate de o manieră cursivă, continuă, pentru formarea unui model mental mai bun al caracteristicilor globale ale vizualizării. Există abordări care utilizează o tehnică denumită *zoom semantic* [21]. Spre deosebire de *zoom*-ul geometric care redesenează toți pixelii funcție de scală, *zoom*-ul semantic oferă mai multe detalii la mărire și ascunde unele dintre acestea la micșorare. Aceasta duce la reduceri importante ale încărcării de procesare și comunicație. *Zoom*-ul semantic este utilizat în conjuncție cu structuri de date multirezoluție, respectiv preprocesarea datelor vizibile pentru fiecare nivel de rezoluție dorit. Alte abordări merg chiar mai departe, propunându-și vizualizarea doar a lucrurilor importante. Adesea, chiar algoritmul este cel care decide ce este considerat important, fie în urma unor operațiuni de clusterizare, separare, agregare, filtrare etc., fie prin alegerea aleatoare a unui eșantion redus de date pentru vizualizare.

### **Tratarea complexității: Navigare condusă de utilizator**

O altă abordare permite intervenția utilizatorului pentru a furniza un *feedback* despre ce este important și ce trebuie vizualizat, sau a selecta pentru ce regiune anume și vecinătăți ale acesteia dorește mai multe detalii. Este astfel adăugat un nivel suplimentar de sofisticare a sistemelor de vizualizare și sunt extinse capacitățile acestora de generare de proiecții cu semnificație din datele complexe. Utilizatorul

ghidează sistemul spre subspațiul dorit din setul de date original, iar algoritmul de proiecție este apoi focalizat pe acesta, evitând procesări inutile pe restul setului de date. Fiind direct implicat în procesul de reducere a complexității, utilizatorul își poate forma un model mental mai bun al datelor. În unele situații, operatorul poate ajusta parametrii algoritmilor de reducere dimensională pentru a ajunge la o reprezentare acceptabilă.

## 6. Concluzii

Abordările prezentate mai sus remodelează paradigma vizualizării convenționale. Se acordă prioritate vitezei și responsivității chiar dacă aceasta este în detrimentul preciziei, prezentându-se un subset al datelor sau un rezumat și o versiune comprimată a acestora. Aceste abordări sunt adesea orientate ecran, complexitatea lor computațională depinzând mai degrabă de dimensiunea / rezoluția ecranului decât de dimensiunea datelor.

Sistemele analitice vizuale viitoare trebuie de asemenea să ofere posibilitatea interacțiunii permanente cu datele, utilizatorul putând ajusta parametrii căutărilor în mod constant în timp ce sistemul demonstrează instantaneu noi vizualizări. Astfel de interacțiuni pot fi utile deoarece în procesul de descoperire a datelor este necesar să fie puse mai multe întrebări și să fie parcurse mai multe iterații pe ipoteze înainte de a putea fi formulate întrebările adecvate / corecte. Rareori un analist de date pune o singură întrebare. Memoria umană este foarte limitată, mai ales când sunt implicate cantități vaste de informație vizuală. O latență redusă la căutare și mecanisme de navigare intuitive pot fi utile în parcurgerea înainte și înapoi a mai multor vizualizări și observarea datelor din perspective multiple, mărinș șansa identificării / găsirii unor modele / șabloane semnificative în acestea.

Managementul datelor și *visual analytics* sunt două discipline care, împreună, sunt capabile să exploateze oportunitățile aduse de era informațională. Cum managementul datelor este practic intrinsec *visual analytics*, rezolvarea unora dintre problemele deschise ale managementului de date va duce (și) la îmbunătățirea aplicațiilor de *visual analytics*. Pe de altă parte, probleme specifice *visual analytics* vor ridica noi provocări comunității implicate în managementul datelor. Prin

urmare, în cadrul acestei dependențe simbiotice, este important să se reflecteze asupra diferitelor perspective și relațiilor inerente atunci când sunt analizate rolurile mutuale jucate de managementul datelor și *visual analytics*. De exemplu, managementul datelor oferă tehnici și algoritmi (de interpolare, statistici) în cazul datelor lipsă, dar o abordare complet automată poate pierde fapte importante, cum ar fi date lăsate lipsă intenționat sau disfuncționalități ale senzorilor. Vizualizarea oferă de asemenea metode pentru tratarea automată a situațiilor datelor lipsă, într-o manieră interactivă. Trebuie deci stabilit când / dacă responsabilitatea gestionării unor astfel de situații va fi în zona managementului datelor sau în cea a vizualizării.

Integrarea datelor eterogene este o activitate centrală a managementului datelor care devine din ce în ce mai importantă și mai utilizată. În vreme ce partea de stocare și regăsire a datelor din seturi foarte mari de date este bine înțeleasă și acoperită, suportul pentru vizualizările eficace și eficiente ale datelor cu (de exemplu) miliarde de elemente / articole și/sau sute de dimensiuni reprezintă încă o activitate provocatoare de cercetare. În particular, utilizatorul trebuie să primească *feedback* rapid în timp ce explorează datele.

Pentru a se face progrese în domeniul *visual analytics* este necesar să fie abordate câteva probleme critice cum sunt: problemele de incertitudine, integrarea datelor semantice și managementul semanticilor, proveniența datelor, fluxurile de date, vizualizarea interactivă a seturilor masive de date, soluționarea activităților intensive la procesare și proiectarea de sisteme de *visual analytics* destinate publicului larg. Rezolvarea acestor probleme reprezintă o provocare pentru ambele comunități – de management al datelor și de *visual analytics* – pentru a se putea profita de oportunitățile oferite de volumul crescând de informații disponibile astăzi.

Dezvoltatorii de aplicații *visual analytics* sunt confrunțați cu mai multe provocări fundamentale în încercarea de a dezvolta metodologii iterative integrate care implică colectarea informației, preprocesarea datelor, reprezentarea cunoștințelor, interacțiune și luarea deciziei. Între oportunitățile majore, amintim: dezvoltarea de instrumente și metode generice / generale, reglementarea și controlul calității, vizualizarea modelelor. Probabil cea

mai importantă oportunitate este aceea de a se oferi liniile directoare și cadrul pentru ca aceste componente să fie dezvoltate [1].

## Mențiuni

Prezentul articol are la bază parte din lucrările Etapei I a Proiectului PN1609-0201, derulat în cadrul Programului Național Nucleu COGNOTIC, la ICI București, 2016.

## BIBLIOGRAFIE

1. **KEIM, D.; KOHLHAMMER, J.; ELLIS, G.; MANSMANN, F. (EDS.):** Mastering The Information Age. Solving Problems with Visual Analytics. Eurographics Association, Germany, ISBN 978-3-905673-77-7, electronic version at Eurographics Digital Library: <http://diglib.eg.org>, 2010.
2. **AMRAIL, S. A.; LEWIS, M.; SARGENT, R.; NOURBAKHS, I.:** Explorable Visual Analytics. Knowledge Discovery in Large and High-Dimensional Data. Proceedings of the Association for Computing Machinery Special Interest Group Knowledge Discovery and Data Mining (ACM SIGKDD) Workshop on Interactive Data Exploration and Analytics, New York City, USA, 2014, pp. 26-34.
3. **THOMAS, J. J.; COOK, K. A.:** Illuminating the Path: The Research and Development Agenda for Visual Analytics. IEEE Computer Society Press, Los Alamitos, 2005.
4. **KEIM, D.; ANDRIENKO, G.; FEKETE, J.-D.; GORG, C.; KOHLHAMMER, J.; MELANCON, G.:** Visual analytics: Definition, process, and challenges. In Information Visualization, LNCS 4950, Kerren, A. et al. (Eds.), Springer-Verlag Berlin Heidelberg, 2008, pp. 154-175.
5. **HARGER, J. R.; CROSSNO, P. J.:** Comparison of open-source visual analytics toolkits. In Proceedings of the SPIE Conference on Visualization and Data Analysis, 2012.
6. **ZHANG, L.; STOFFEL, A.; BEHRISCH, M.; MITTELSTÄDT, S.; SCHRECK, T.; POMPL, R.; WEBER, S. H.; LAST, H.; KEIM, D.:** Visual Analytics for the Big Data Era – A Comparative Review of State-of-the-Art Commercial Systems. In Proceedings of IEEE Conference on Visual Analytics Science and Technology 2012, Oct. 14-19, Seattle, WA, USA, pp. 173-182.
7. **FEKETE, J.-D.:** The infovis toolkit. In INFOVIS, 2004, pp. 167-174.
8. **HEER, J. ; CARD, S. K.; LANDAY, J. A.:** Prefuse: a toolkit for interactive information visualization. In Proceedings of the SIGCHI conference on Human factors in computing systems, CHI '05, New York, NY, USA, (ACM), 2005, pp. 421-430.
9. **WEAVER, C.:** Building highly-coordinated visualizations in Improvise. In INFOVIS, 2004, pp. 159-166.
10. **\*\*\*:** Java Universal Network/Graph Framework, <http://jung-sourceforge.net/>.
11. **\*\*\*:** <http://www.tableausoftware.com/>.
12. **\*\*\*:** <http://spotfire.tibco.com/>.
13. **\*\*\*:** <http://www.qlikview.com/>.
14. **\*\*\*:** <http://www.jmp.com/>.
15. **\*\*\*:** <http://www.advizorsolutions.com/>.
16. **\*\*\*:** <http://www.centrifugesystems.com/>.
17. **\*\*\*:** <http://www.visualanalytics.com/>.
18. **SCHUNN, C. D.; KLAHR, D.:** A 4-space model of scientific discovery. In Proceedings of the seventeenth annual conference of the Cognitive Science Society, 1995, pp. 106-111.
19. **FRY, B. J.:** Computational information design. Ph. D. Thesis, Massachusetts Institute of Technology, School of Architecture and Planning, Program in Media Arts and Sciences, 2004.
20. **FEKETE, J.-D.:** Visual analytics infrastructures: From data management to exploration. Computer, 46(7), July 2013, pp. 22-29.
21. **HERMAN, I.; MELANCON, G.; MARSHALL, M.:** Graph visualization and navigation in information visualization: A survey. IEEE Transactions on Visualization and Computer Graphics, 6(1), January 2000, pp. 24-43.