

Studii asupra vizualizării datelor și infograficelor

Alecsandru Grigoriu

Facultatea de Informatică,
Universitatea „Al. I. Cuza” Iași, România
alecsandru.grigoriu@info.uaic.ro

Sabin Corneliu Buraga

Facultatea de Informatică,
Universitatea „Al. I. Cuza” Iași, România
busaco@info.uaic.ro

REZUMAT

În acest articol se descriu o serie de contribuții în domeniul vizualizării datelor și elaborarea infograficelor în prisma a două studii de caz. Primul studiu reprezintă o abordare originală a vizualizării datelor rețelelor sociale și livrarea analizelor în format vectorial. Al doilea studiu constă în proiectarea unei soluții de tip infografic pentru provocarea lansată de Mozilla Labs – *Open Data Visualization Competition* – menită să ofere o interpretare statistică și vizuală a datelor oferite de Fundația Mozilla cu privire la categoriile de utilizatori și modul lor de interacțiune cu browser-ul Web Firefox – date obținute prin inițiativa *Mozilla Test Pilot*.

Cuvinte cheie

Studiu, vizualizarea datelor, infografic, rețea socială, interacțiune Web, grafică vectorială

Clasificare ACM

H5.2. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCERE

Odată cu “explozia” Internetului, utilizatorii pot accesa zilnic un volum imens de informații (de la știri, filme, muzică, la notificări frecvente). O problemă ar fi gestionarea acestora, dar mai degrabă cum am putea reprezenta informațiile astfel încât să putem înțelege – la nivel de utilizator – mesajul dorit a fi transmis. Actualmente, există colecții de date având mari dimensiuni, aspectele de interes fiind modelarea, procesarea, agregarea, interpretarea și vizualizarea acestora. După tip sau după context, datele diferă între ele și nu pot fi reprezentate în același mod. Prin urmare, provocarea constă, mai întâi, în înțelegerea a ceea ce transmit datele și, în al doilea rând, maniera cum sunt expuse utilizatorului.

Conform Ben Fry [1], pentru a înțelege datele trebuie să începi cu o mulțime de numere și o întrebare. Pentru a ajunge la un răspuns trebuie parcurși următorii pași:

- *acquire*: obținem datele și le salvăm;
- *parse*: prelucrăm datele respective pentru o mai bună înțelegere a lor;
- *filter*: eliminăm datele de care nu avem nevoie – păstrăm doar ceea ce este de interes pentru problema de rezolvat;
- *mine*: aplicăm diverse metode pentru a genera statistici care să ne aproprie cât de cât de răspunsul dorit;

- *represent*: reprezentăm datele; vizualizările pot diferi în funcție de cerințe și de datele în sine. De exemplu, pentru a reprezenta evoluția în timp a anumitor date se poate apela la un grafic.
- *refine*: aducem ajustări la adresa celor reprezentate pentru a clarifica, îmbunătăți și chiar oferi un aspect estetic conform vizualizării curente;
- *interact*: interacționăm – o simplă reprezentare nu este de ajuns. Pentru a manipula cu adevărat datele încât să extragem informațiile dorite, proiectul obținut trebuie să ofere interacțiune cu utilizatorul, astfel obținem informații noi și valide pentru întărirea puternică a certitudinii răspunsului obținut.

Ultima etapă (*interact*) privește domeniul interacțiunii om-calculator. Interacțiunea se realizează la nivel de interfață cu utilizatorul, cuprinzând atât partea software, cât și cea hardware. De exemplu, datele noastre sunt afișate pe ecran și interacționăm/realizăm operațiile specifice (depinde de fiecare aplicație în parte) prin intermediul mouse-ului și tastaturii (fie că este vorba să tastăm date de intrare, fie să executăm o comandă) sau, mai nou, prin gesturi.

Reprezentările vizuale au evoluat în timp și s-au dezvoltat noi abordări. Vizualizarea datelor se poate realiza și la nivel de infografic [2].

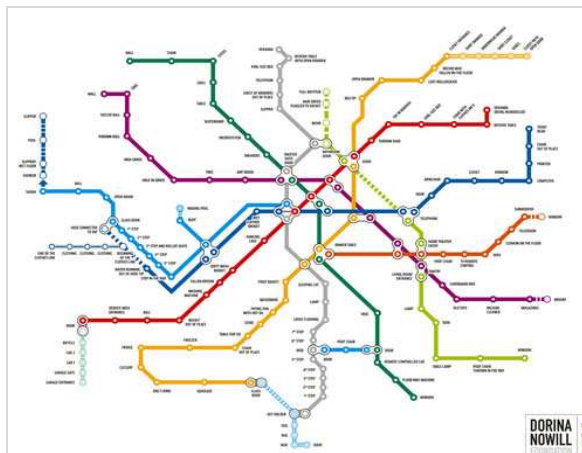


Figura 1. Foundation for the Blind: Metro Map (Dorina Nowill)

Diferența majoră între cele două stă în accentul pus pe arhitectura și structurarea informațiilor [3] la nivel de infografic, unde partea de interacțiune lipsește. Un infografic se focalizează asupra livrării de conținut drept sursă de informare pentru utilizatorul de rând (consumatorul). Figura 1 ilustrează ideea de informare publică, prin oferirea unei hărți de rute de metrou cetățenilor, inclusiv celor cu dezabilități.

Având o imagine de ansamblu asupra vizualizării datelor și infograficelor, articolul de față oferă o contribuție originală cu privire la aceste tehnici, din perspectiva a două studii de caz. Prima parte va aborda utilizarea graficii vectoriale pentru vizualizarea datelor din cadrul rețelelor sociale. Al doilea studiu va aborda contextul reprezentării datelor utilizatorilor navigatorului Mozilla Firefox sub formă de infografic.

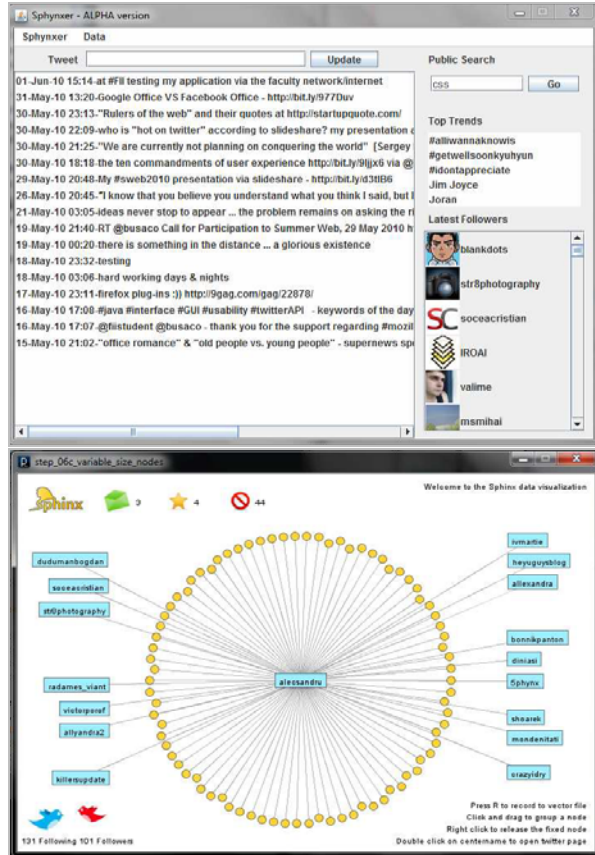


Figura 2: Interfața principală a soluției propuse

STUDIUL EXPERIENȚEI GRAFICII VECTORIALE

Scopul proiectului este realizarea unui studiu comparativ privind facilitățile oferite de tehnologiile actuale vizând grafica vectorială – ca exemple tipice, se pot enumera JavaFX, SVG (*Scalable Vector Graphics*), Adobe Flex/AIR, WPF (*Windows Presentation Foundation*) și X3D (*Extensible 3D Graphics*).

S-a proiectat o interfață de programare a aplicațiilor (API – *Application Programming Interface*) de vizualizare a datelor care să uniformizeze maniera de acces la soluțiile menționate anterior. Am urmărit reprezentările alternative ale informațiilor privitoare la aplicațiile Web sociale, în special rețele sociale (e.g., Facebook, Jaiku, MySpace, Twitter etc.) – pe baza API-urilor oferite de acestea.

În esență, provocarea pornește de la rețelele sociale, crearea unei vizualizări asociate acestora și se termină în a determina acele formate vectoriale care satisfac așteptările utilizatorului. Astfel, beneficiarii devin utilizatorii (aceștia variind în funcție de tip și mediu) care sunt legați în mod direct sau indirect de studiul realizat. Pe baza datelor culese din cadrul rețelei sociale prestabilite, se proiectează

o vizualizare a acestor date, în final aceasta este exportată în diverse formate vectoriale. Tot acest lanț de procese poate fi cuprins de o serie de aplicații menite să evidențieze avantajele/dezavantajele graficii vectoriale.

Primii beneficiari direcți sunt utilizatorii siturilor rețea socială (totul pornește de la datele aferente fiecărui profil de utilizator). Pentru o mai bună particularizare, ne-am rezumat la facilitățile oferite de una dintre cele mai populare rețele sociale (astfel numărul de beneficiari fiind satisfăcător de numeros): Twitter.

Soluția propusă

Am urmat o serie de pași pentru a ajunge la implementarea dorită. Am stocat datele corespunzătoare profilului, pentru a le putea manipula. Am procesat datele obținute în funcție de interesul nostru (nu avem nevoie de toate datele – le putem grupa pe categorii – astfel depistăm categorii asociate „tweets”, „timeline”, „following”, „followers”).

După ce am filtrat datele obținute, am aplicat tehnici specifice domeniului *data mining*. Mai exact, pentru fiecare tip de dată obținem un șablon (*pattern*) – pentru a plasa datele într-un context statistic. De exemplu, în prima fază ne-ar interesa să aflăm minimul și maximum pentru un anumit tip de date – putem determina numărul de mesaje (*tweet-uri*) și a-l compara la nivel de abonați de tip *followers/following*.

Având determinate datele în forma lor „particularizată” se poate trece la etapa de reprezentare a datelor într-o formă vizuală corespunzătoare. Reprezentarea implementată sub formă de graf (Figura 2) poate fi extinsă din punct de vedere al interacțiunii. Astfel, pe baza metodei de vizualizare stabilite putem obține formatele vectoriale dorite (din *data visualization* exportăm spre *vector graphics*). Formatele vectoriale obținute au fost testate, după caz, după integritatea și funcționalitatea lor. Astfel, formate precum PDF au fost verificate pentru portabilitate, iar documentele SVG pentru integrarea lor la nivel de browser.

Instrumente folosite

În dezvoltarea soluției de vizualizare am apelat la diverse tehnologii disponibile gratuit, oferind suportul necesar proiectului nostru.

Am recurs la limbajul de programare disponibil *open source* Processing [4] pentru a putea realiza vizualizarea datelor obținute. Pentru a facilita integrarea între module am ales Java – JDK 6. La nivel de integrare cu Twitter, am apelat la un API disponibil liber: twitter4j [5] implementat în Java. Conectarea la rețeaua socială se realizează prin intermediul protocolului de autentificare OAuth [6].

Pentru export grafică vectorială, am apelat la metode native Processing/Java sau biblioteci *open source* (de exemplu, pentru export în SVG am avut nevoie de biblioteca proSVG – dezvoltată de comunitatea utilizatorilor Processing).

Propunerea finală [7] a constat din două module principale: un modul client realizat cu Java SWING pentru accesul la datele aferente contului de Twitter și modulul dedicat vizualizării datelor stocate prin intermediul clientului descris anterior.

Proiectarea interfeței cu utilizatorul

Soluția dezvoltată este compusă din două module. Primul reprezintă o interfață-client menită să ofere acces la funcționalitățile Twitter. Provocarea a constat în alegerea acelor proprietăți necesare utilizatorului: cu cât mai multe funcționalități sunt suportate într-o singură aplicație, cu atât mai complexă devine interfața cu utilizatorul. Prin urmare, modulul client a fost proiectat pentru actualizarea stărilor, căutarea publică a preferințelor și vizualizarea datelor (tendențele actuale, ultimii abonați, precum și numărul de mesaje primite, de actualizări favorite și utilizatori blocați).

La nivel de șabloane de proiectare [8], aplicația este construită sub forma unei singure ferestre principale (*one-window-drilldown*), unde pot fi accesate toate funcționalitățile descrise anterior (*center stage*). Pentru a facilita parcurgerea conținutului, elementele textuale sunt aliniate la stânga (conform șablonului de prezentare *left-right alignment*). Fiecare funcționalitate este încadrată în propriul panou (*task pane*), aceste mini-aplicații putând fi accesate direct din fereastra principală (*hub and spoke*) – de exemplu, căutarea publică.

Interfața principală aduce în prim plan utilizatorului acțiunile de bază corespunzătoare Twitter.

Al doilea modul este dedicat extinderii vizualizării datelor. Ultimii abonați sunt „conectați” la utilizatorul curent printr-un sistem de grafuri. Scopul aceste reprezentări este de arăta gradul de apropiere dintre utilizatori (unii pot să fie îndepărtați față de alții) și gradul de expunere la nivel de rețea socială (utilizatorul este înconjurat de abonați și orice mișcare este strict legată de aceștia). În vizualizarea extinsă, utilizatorii pot analiza comparații privind numărul de utilizatori urmăriți (*following*) versus numărul de abonați (*followers*) prin contrastul unor pictograme metaforice („păsări”) de mărimi diferite în funcție de valorile comparate.

Pentru vizualizarea numărului de mesaje, de actualizări favorite și utilizatori blocați s-au folosit metafore reprezentative (plic, stea, respectiv acces interzis). La nivel de recurgere la șabloane de proiectare, s-a apelat tot la o scenă principală (*center stage*) unde au fost expuse toate datele necesare. În plus, interacțiunea cu datele (de exemplu, nodurile grafului) se realizează prin manipulare directă (*direct manipulation*): utilizatorii pot „prinde”, „trage” sau „muta” nodurile, astfel datele pot fi grupate pe diverse categorii. Nodurile revin în poziția lor inițială (cu ajutorul algoritmului de relaxare descris de Ben Fry în lucrarea [1]).

Acționarea butoanelor de mouse (click, respectiv dublu-click) ajută la accesarea etichetei fiecărui nod (care conține numele utilizatorului) și la accesarea paginii de profil corespunzătoare utilizatorului etichetat. Pentru a nu

crea confuzie utilizatorului, vizualizarea oferă sugestii scurte pentru o navigare ușoară.

STUDIUL ASUPRA VIZUALIZĂRII DATELOR ÎN INFOGRAFIC ȘI COMPORTAMENTULUI UTILIZATORILOR DE BROWSER WEB

Competiția *Open Data Competition* [9] organizată de Mozilla Labs se bazează pe datele oferite de programul Mozilla Test Pilot [10]. Aceste date au urmărit modul cum utilizatorii folosesc navigatorul Web Firefox și maniera în care aceștia interacționează cu interfața versiunii 4 beta.

Datele culese în cadrul Test Pilot au fost oferite publicului pentru analiză și modelare. Provocarea finală a constat în conceperea unei soluții de vizualizare a datelor sau expunere infografică care să ofere un răspuns la întrebarea „Cum folosesc oamenii Firefox?”.

Livrarea conceptului final [6] a urmat o serie de pași similar studiului de caz prezentat anterior.

Întâi, au fost descărcate datele de test, disponibile în format CSV (*Comma Separated Values*), dar în număr limitat de înregistrări. Pe baza acestor fișiere, s-au filtrat și sortat datele, astfel încât am putut efectua primele analize.

De exemplu, numărul de utilizatori care intrau în *private mode* era legat de numărul de descărcări de fișiere sau procentajul celor care nu folosesc Firefox era dominat de utilizatorii browser-ului Google Chrome.

Având obținute astfel de analize, infograficul a fost împărțit în două secțiuni: descrierea profilului unui utilizator de Firefox (numită *FX User Profile*) și descrierea de activități și acțiuni ale respectivului profil de utilizator (*FX Modus Operandi*).

Statisticile oferite în infografic nu reprezintă statistici oficiale ale fundației Mozilla, ci sunt concluzii proprii, pe baza tehnicilor de analiză realizate. Ca instrumente de lucru am recurs la Microsoft Office Excel 2007 și Adobe Illustrator.

FX User Profile

Profilul utilizatorului (vezi Figura 3) este dedicat oferirii de informații generale. Pe baza datelor puse la dispoziție am putut deduce următoarele:

- 84% din utilizatorii înregistrați folosesc versiunea Firefox 4.0.6b. 89% au folosit versiunea Test Pilot 1.0.3, iar 185 este numărul maxim de extensii instalate.
- 35% dintre utilizatorii înregistrați folosesc Firefox de 3—5 ani, 32% au între 26—35 ani, 26% utilizează Internetul între 2—4 ore pe zi și 91% sunt subiecți de sex masculin.
- 69% dintre utilizatori folosesc și alt browser Web alături de Firefox, dintre care 91% sunt utilizatori de Google Chrome.
- 93% din utilizatorii înregistrați recurg la sistemul de operare Windows (49% pentru Windows XP, iar 41% utilizează Windows 7).
- 40% dintre utilizatorii nu folosesc dispozitive de tip *smartphone*; totuși, 14% interacționează cu produse

Apple (iPhone, iPod Touch, iPad), 11% Android, 10% Symbian.

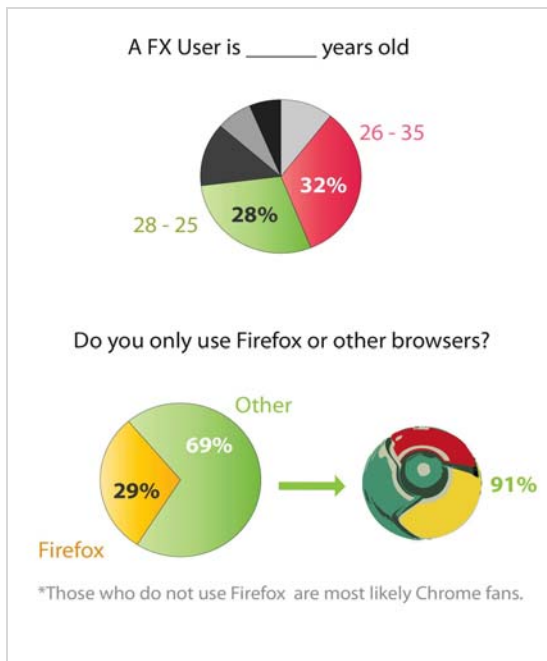


Figura 3. Firefox User Profile: diferite statistici

FX Modus Operandi

Prin intermediul datelor puse la dispoziție am putut trage următoarele concluzii cu privire la comportamentul utilizatorilor (Figura 4):

- 67% din utilizatorii înregistrați preferă să apese tasta ENTER decât să apese butonul Go atunci când introduc o adresă Web în bara de adrese. 18% preferă să aibă bara de meniu ascunsă, iar 37% preferă să mute *tab*-urile deasupra. Majoritatea preferă lucrul cu 1—3 *tab*-uri.
- 98% dintre utilizatorii înregistrați preferă să deschidă un *tab* nou cu ajutorul butonului dedicat.
- Majoritatea utilizatorilor accesează butonul de *feedback*, fără să introducă vreun mesaj. Cea mai puțin folosită parte a interfeței este *Subscribe*.
- De cele mai multe ori, utilizatorii folosesc împreună instrumentele/extensiile Web Developer, Inspect și Web Console atunci când vizualizează codul sursă al unei pagini Web.
- În meniu, cel mai utilizat buton este cel pentru Firefox. Cea mai folosită scurtătură de taste este cea pentru *tab* nou (CTRL + T), urmată de cea pentru căutare (CTRL + F) și revenirea la pagina anterioară (tasta *backspace*).

CONCLUZII

Lucrarea de față reprezintă o contribuție aplicativă în domeniul interacțiunii om-calculator, fiind focalizată asupra vizualizării datelor și realizarea de infografice datorită.

Studiile de caz originale realizate au urmărit vizualizarea în format vectorial a datelor rețelelor sociale, inclusiv cu

marcarea relațiilor dintre utilizatori, reprezentarea infografică a categoriilor de utilizatori de browser Web și a comportamentului lor.

Drept direcții de viitor, studiile de caz vor fi extinse. Ele pot aborda contextul vizualizării datelor pentru platforme mobile (*smartphone*, tabletă etc.), dar și îmbunătățirea interacțiunii la nivel de interfață tactilă (*multi-touch*).

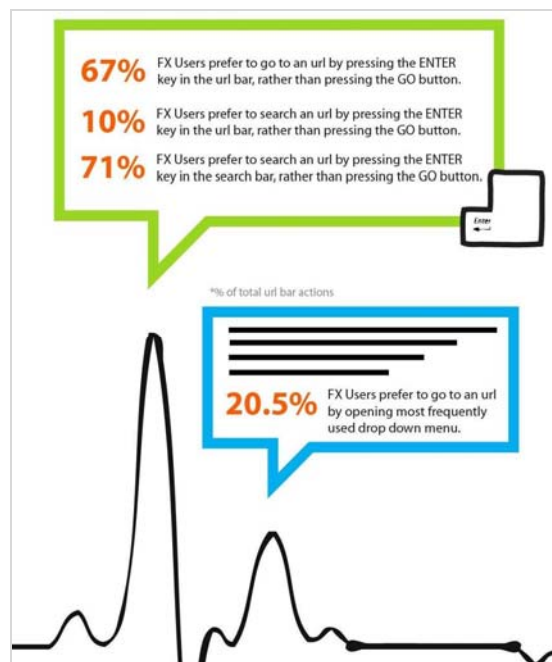


Figura 4. Firefox Modus Operandi: diverse comportamente ale utilizatorilor

REFERINȚE

1. Ben Fry, *Visualizing Data*, O'Reilly (2008)
2. Ecaterina Valică, "Comunicarea vizuală prin intermediul infograficelor", în S. Buraga și I. Juvină (eds.), *Proceedings of the 5th National Conference on Human-Computer Interaction — RoCHI 2008*, MATRIX ROM, București (2008)
3. Peter Morville, Luis Rosenfeld, *Information Architecture for the World Wide Web*, O'Reilly (2006)
4. * * *, Situl Processing: <http://processing.org/>
5. * * *, Situl bibliotecii Twitter4j: <http://twitter4j.org/>
6. * * *, Situl protocolului OAuth: <http://oauth.net/>
7. * * *, Film demonstrativ – vizualizarea datelor dintr-o rețea socială: <http://students.info.uaic.ro/~alecsandru.grigoriu/moviedemo.swf>
8. J. Tidwell, *Designing Interfaces*, O'Reilly (2005)
9. * * *, Situl Mozilla Open Data Competition: <http://design-challenge.mozillalabs.com/open-data/OpenDataCompetition/>
10. * * *, FX User Profile and Modus Operandi: <https://testpilot.mozillalabs.com/testcases/fx-user-profile-and-modus-operandi.html>