

Vizualizacija podatkov

Uvod

- Kako izgledajo poti milijonov spletnih obiskovalcev skozi svetovni splet?
- Kako lahko bilijone človeških genotipov A, C, G in T primerjamo z drugo vrsto (opice, miši)?
- Med vsemi datotekami na disku, katere zasedejo največ prostora in kako pogosto so uporabljene?

Uvod

- Odgovore na vsa vprašanja nam lahko ponudi vizualizacija podatkov
- Seveda v kombinaciji z drugimi tehnikami
 - Računalniških znanosti
 - Statistike
 - Rudarjenja po podatkih
 - Grafičnega oblikovanja
- Odgovor mora imeti pomen in biti mora razumljiv za široko množico

Uvod – možni problemi

- Veličina podatkov
 - Težko je ustvariti jasno sliko
- Spremenljiva narava podatkov
 - Novi podatki
 - Stari podatki se neprestano spreminjajo

= to kliče po novih orodjih za vizualizacijo

Analiza podatkov - cilj

- Poudariti attribute po njeni pomembnosti
- Odkriti vzorec
- Istočasno prikazati attribute skozi več dimenzij

Povzetek

- Uporaba podatkov na drugačen način
- Osnove vizualizacije
 - Kako izbrati pravi prikaz
 - Kako dodati interaktivnost, ki naredi podatke vedno znova zanimive
 - Uporaba orodja Processing za vizualizacijo
 - Način razmišljanja enak tudi pri drugih orodjih

Planiranje

- Preveč informacij
- Zbiranje podatkov
- Razmišljanje o podatkih
- Podatki niso nikoli enaki
- Kakšno je vprašanje?
- Kombinacija več disciplin

Planiranje - Preveč informacij

- Informacijska preobremenitev
 - Dnevni časopis (beri New York Times, sobotno Delo) nosi več informacij kot jih je renesančni človek sploh imel na voljo v svojem celem življenju
- Računska moč današnjih računalnikov
 - hramba informacij
 - dostop
 - obdelava
- Grafična moč (igrice)
 - 2D in 3D grafika
 - Omogočajo gladko, interaktivno animacijo

Planiranje - Zbiranje podatkov

- Vedno boljši smo pri kopičenju podatkov
 - ne znamo pa jih uporabiti
- Veliko podatkov je brezplačno dostopnih na internetu
 - Ne izkoristi pa se njihov informacijski potencial ker niso pravilno prikazani oziroma sploh niso.
- Tudi če imamo zbranih veliko podatkov nam ne odgovorijo na naša vprašanja
 - **Velik izziv vizualizacije**

Planiranje - Razmišljanje o podatkih

- Premalo razmišljamo o podatkih
 - Primer: AOL objavil podatke o iskalnih pojmih milijon uporabnikov
 - Identifikacija uporabnikov
 - Zadrega ob razkritju njihovih navad
- Kaj sploh pomenijo zbrani podatki
 - Navade vseh ljudi ali samo peščice
- Kaj pomeni, če izgubimo milijone osebnih informacij ljudi ali če nelegalno pridejo v javnost?

Planiranje - Podatki niso nikoli enaki

- Navajeni smo, da gledamo na podatke kot na fiksne vrednosti
 - Podatki se nenehno spreminjajo
- Kako predstaviti podatke, ki se spreminjajo vsako sekundo, minuto?
 - To je v resničnem svetu velikokrat res
 - Vreme se spreminja
 - Vlaku ima zamudo
 - Predstavitel novega izdelka spremlja promet na internetu
 - Kaj se zgodi ko se stvari premikajo?
 - Kako rokujejo z “živimi” podatki?
 - Omogočimo sled, ki kaže spremembe
 - Omogočimo potovanje skozi čas

Planiranje - Kakšno je vprašanje?

- Primer: Zemljevid podzemne železnice
 - Prikazuje kompleksno strukturo mesta
 - Se skoncentrira na cilj uporabnika
 - Kako priti iz točke A v točko B

Harry Beck, 1930



Planiranje - **Kakšno je vprašanje?**

- Pogledati moramo na celoto zbranih podatkov
- Preveč se koncentriramo na veličino podatkov (koliko GB imamo podatkov, kako težak bo problem predstavitve)

in premalo na vprašanja

- Zakaj smo podatke zbrali?
- Kaj je zanimivega v podatkih?
- Kakšno zgodbo lahko povemo?

Planiranje - Kakšno je vprašanje?

- Najpomembnejši del pri razumevanju podatkov – **Na katera vprašanje bi radi s svojo predstavitevjo odgovorili!**
- Enostavno je reči:
 - Kaj se skriva v podatkih?
 - Kaj podatki predstavljajo?**ampak**
 - Za kaj so podatki pomembni?**
 - Kakšna je njihova vrednost?**

Planiranje - **Kakšno je vprašanje?**

- Bolj specifično kot je vaše vprašanje
 - **lažja, jasnejša bo vaša predstavitev**
- Iz celotne množice podatkov mora biti sporočilo jasno vidno
- Tudi vprašanje mora biti dobro
 - Prinese interes na druge
 - Poveča radovednost

Planiranje - **Kombinacija več disciplin**

- Statistika
 - Podatkovno Rudarjenje
 - Grafično oblikovanje
 - Vizualizacija podatkov
-
- V procesu predstavitve podatkov moramo vse discipline med seboj povezati

7 stopenj oblikovanja informacij

- Pridobivanje (**Acquire**)
- Razčlenjevanje (**Parse**)
- Filtriranje (**Filter**)
- Rudarjenje (**Mine**)
- Predstavitev (**Represent**)
- Izboljšava (**Refine**)
- Interakcija (**Interact**)

Prodobivanje (**Acquire**)

- Podatki so lahko:
 - v datoteki
 - podatkovni bazi
 - na internetu

Razčlenjevanje (**Parse**)

- Določimo strukturo podatkov po pomenu
- Uredimo v kategorije

Filtriranje (**Filter**)

- Odstranimo nepomembne podatke
- Ohranimo samo podatke, ki so za dano vizualizacijo interesantni
- Umestimo podatke v matematični kontekst

Rударjenje (**Mine**)

- Uporabimo matematične in statistične metode
- Metode rudarjenja po podatkih
- Spoznamo vzorec v podatkih

Predstavitev (**Represent**)

- Za začetek izberemo osnovno predstavitev
 - stolpčni diagram
 - seznam
 - drevo

Izboljšava (**Refine**)

- Izboljšamo osnovno predstavitev
 - jasnejše
 - vizualno privlačnejše

Interakcija (**Interact**)

- Dodamo metode za manipulacijo s podatki
- Z interakcijo spreminjamo pogled na podatke
 - Izbiramo kateri atributi bodo vidni