Teorija baza podataka
Uvod

Izv. prof. dr. sc. Markus Schatten

Fakultet organizacije i informatike,
Sveučilište u Zagrebu
Pavlinska 2, 42000 Varaždin
markus.schatten@foi.hr

Jeste li znali?

- Cijena jednog GiB?
  - 1956. ~ $ 2,000,000

Jeste li znali?

- Cijena jednog GiB?
  - 1956. ~ $ 2,000,000
  - 2021. < $ 0.02

Koliko bajta ...?

- 1,024 MiB = 1 GiB = kamion natovaren papirima

Koliko bajta ...?

- 1,024 MiB = 1 GiB = kamion natovaren papirima
- 1,024 GiB = 1 TiB = 50,000 stabala

Koliko bajta ...?

- 1,024 MiB = 1 GiB = kamion natovaren papirima
- 1,024 GiB = 1 TiB = 50,000 stabala
- 1,024 TiB = 1 PiB = 250 milijardi stranica teksta

Koliko bajta ...?

- 1,024 MiB = 1 GiB = kamion natovaren papirima
- 1,024 GiB = 1 TiB = 50,000 stabala
- 1,024 TiB = 1 PiB = 250 milijardi stranica teksta
- 1,024 PiB = 1 EiB = 1 bilijun knjiga!

Digitalni podaci

- 2002. 5 EiB podataka pohranjeno u računalima = 37 × Library of Congress = ... ?

Digitalni podaci

- 2002. 5 EiB podataka pohranjeno u računalima = 37 × Library of Congress = ... ?
- Sve riječi koje je izgovorila ljudska vrsta!

Digitalni podaci

- 2002. 5 EiB podataka pohranjeno u računalima = 37 × Library of Congress = ... ?
- Sve riječi koje je izgovorila ljudska vrsta!
- Ikada!!

Digitalni podaci

- 2002. 5 EiB podataka pohranjeno u računalima = 37 × Library of Congress = ... ?
- Sve riječi koje je izgovorila ljudska vrsta!
- Ikada!!
- 2006. 161 EiB = 12 kula knjiga od zemlje do mjeseca!

Digitalni podaci

- 2002. 5 EiB podataka pohranjeno u računalima = 37 × Library of Congress = ... ?
- Sve riječi koje je izgovorila ljudska vrsta!
- Ikada!!
- 2006. 161 EiB = 12 kula knjiga od zemlje do mjeseca!
- 2007. 295 EiB ...

Digitalni podaci

- 2002. 5 EiB podataka pohranjeno u računalima = 37 × Library of Congress = ... ?
- Sve riječi koje je izgovorila ljudska vrsta!
- Ikada!!
- 2006. 161 EiB = 12 kula knjiga od zemlje do mjeseca!
- 2007. 295 EiB ...
- 2010. 988 EiB = knjige od Sunca do Plutona i nazad!

Digitalni podaci

- 2002. 5 EiB podataka pohranjeno u računalima = 37 × Library of Congress = ... ?
- Sve riječi koje je izgovorila ljudska vrsta!
- Ikada!!
- 2006. 161 EiB = 12 kula knjiga od zemlje do mjeseca!
- 2007. 295 EiB ...
- 2010. 988 EiB = knjige od Sunca do Plutona i nazad!
- 2013. WWW sadrži 4 ZiB (4096 EiB) = 2 × projicirani broj zvijezda u svemiru

Digitalni podaci

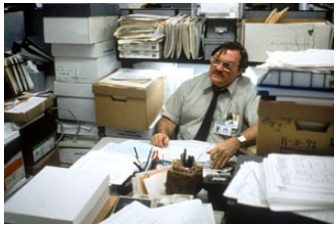
- Većina tih podataka potencijalno sadrži relevantne informacije!
- 98 % svih informacija su kreirane elektronskim putem
- Preko 80 % naših dokumenata nikada nisu ispisani

E-mail

- Godina prve poslane elektroničke poruke:
  - 1971.
- Broj poslanih e-mailova Clintonove administracije (2001)
  - 32 milijuna
- Prva parnica "Bijele kuće" zbog e-maila:
  - 1989.
- Procjena 2017.
  - 1 milijarda

E-mail in business

- 2/3 radne snage u SAD-u koristi e-mail kao dio dnevne rutine
- Radnici u znanjem intenzivnim djelatnostima dnevno 2 – 3 sata ih odgovaraju i pišu



Mobiteli

- 2019. broj korisnika mobilnih telefona dosegao je 5 milijardi!
- 2017. 22 milijarde tekstualnih poruka (SMS) se šalje svaki dan (ne uključujući app-to-app messaging)
- WhatsApp i Facebook Messenger šalju zajedno više od 60 milijardi poruka dnevno

Mobilni uređaji

- Oko 81 % poslovnih ljudi koristi mobilne uređaje
- Prosječni tinejđer u SAD pošalje 80 poruka dnevno

Eksplozija društvenih medija

- Enciklopedija Brittanica sadrži 65,000 članaka
- Wikipedija u engleskoj inačici sadrži preko 6.3 milijuna članaka (2021.)

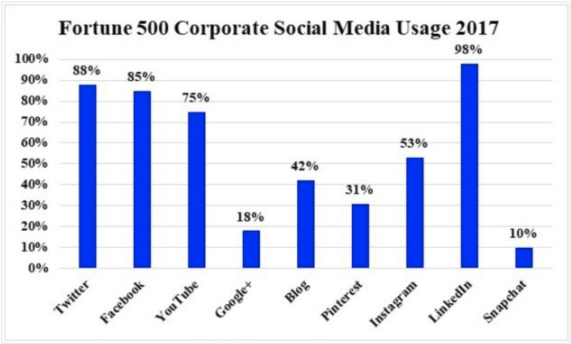
Eksplozija društvenih medija

- Dnevno se pošalje oko 58 milijuna tweetova
- Broj tweetova u sekundi je oko 9100
- Ukupno više od 10 milijardi tweetova

Eksplozija društvenih medija

- 2018. Facebook ima prekao 2.20 milijarde mjesečno aktivnih korisnika!
- Svakih 60 sekundi na Facebook-u: poslano 510,000 komentara, ažurirano 293,000 statusa, postavljeno 136,000 fotografija

Društveni mediji i poslovni svijet



Source: http://www.umassd.edu/cmr/socialmediaresearch/2017fortune500/#d.en.963986

Koliko je informacija na Internetu?

- Vrijeme potrebno da bi jedna osoba bez spavanja pregledala svaku web stranicu na jednu minutu:
  - 95 000 godina

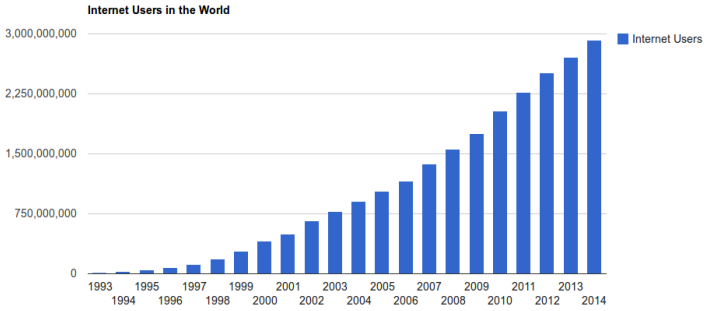
Koliko je informacija na Internetu?

- Vrijeme potrebno da bi jedna osoba bez spavanja pregledala svaku web stranicu na jednu minutu:
  - 95 000 godina
- Vrijeme potrebno da sve pročita:
  - Oko 8.7 milijuna godina

Koliko je informacija na Internetu?

- Vrijeme potrebno da bi jedna osoba bez spavanja pregledala svaku web stranicu na jednu minutu:
  - 95 000 godina
- Vrijeme potrebno da sve pročita:
  - Oko 8.7 milijuna godina
- Prema nekim procjenama na Internetu je pohranjeno više od jednog YiB (8,000,000,000,000,000,000,000,000 bitova)

Broj Internet korisnika



Google

- Oko 100 milijardi Google pretraga svaki mjesec
- Indeksira oko 45 milijardi stranica

Internet of Things

- Cisco procjenjuje da je krajem 2019. IoT generirao preko 500 zetabajta podataka godišnje
- U godinama pred nama predviđa se da će taj broj rasti eksponencijalno, ne linearno!

Što će se dogoditi kada ...

- ... svaka osoba na svijetu posjeduje tisuće računala?

Što će se dogoditi kada ...

- ... svaka osoba na svijetu posjeduje tisuće računala?
- ... računala budu milijardu puta brža od današnjih?

Što će se dogoditi kada ...

- ... svaka osoba na svijetu posjeduje tisuće računala?
- ... računala budu milijardu puta brža od današnjih?
- ... će biti pohranjeno više digitalnih podataka od molekula u svemiru?

Što će se dogoditi kada ...

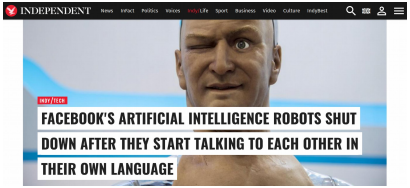
- ... svaka osoba na svijetu posjeduje tisuće računala?
- ... računala budu milijardu puta brža od današnjih?
- ... će biti pohranjeno više digitalnih podataka od molekula u svemiru?
- ... umjetna inteligencija nadmaši našu vlastitu?

Što će se dogoditi kada ...

- ... svaka osoba na svijetu posjeduje tisuće računala?
- ... računala budu milijardu puta brža od današnjih?
- ... će biti pohranjeno više digitalnih podataka od molekula u svemiru?
- ... umjetna inteligencija nadmaši našu vlastitu?
- ... pametna računala krenu graditi računala koje mi više ne razumijemo?

Što će se dogoditi kada ...

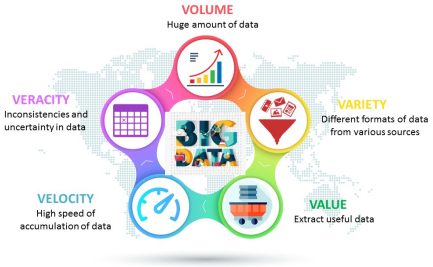
- ... svaka osoba na svijetu posjeduje tisuće računala?
- ... računala budu milijardu puta brža od današnjih?
- ... će biti pohranjeno više digitalnih podataka od molekula u svemiru?
- ... umjetna inteligencija nadmaši našu vlastitu?
- ... pametna računala krenu graditi računala koje mi više ne razumijemo?



Metodologija

5Vs of BigData

- Volume – količina
- Variety – različitost
- Veracity – povjerenje
- Velocity – brzina
- Value – upotrebljivost



Potrebne su nam nove metode ...

- Polustrukturirani i nestrukturirani podaci
- Strujanje podataka (engl. streaming data)
- Distribuirana pohrana, procesiranje, postavljanje upita, analiza ...

Napredni podatkovni sustavi

- Napredne relacijske baze podataka
  - Parcijalne baze podataka
  - Temporalne baze podataka
  - Aktivne baze podataka
  - Poopćene baze podataka
  - Objektno-relacijske baze podataka
  - ...
- Ne (nužno) relacijske baze podataka
  - Deduktivne baze podataka
  - Polustrukturirane baze podataka (podatkovni grafovi)
  - Objektno-orijentirane baze podataka
  - Sustavi strujanja podataka
  - ...

Podatkovna znanost

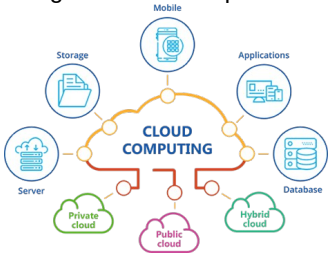
- Podatkovna znanost (engl. data science) je interdisciplinarno polje koje koristi znanstvene metode, procese, algoritme i sustave kako bi izlučila znanje i spoznaje iz mnogih strukturiranih i nestrukturiranih izvora podataka.
- Podatkovna znanost usko je vezana uz područja rudarenja podataka, umjetne inteligencije (posebice strojno učenje), podatkovno inženjerstvo i velike podatke (engl. Big Data)

Umjetna inteligencija

- Umjetna inteligencija (engl. Artificial intelligence – AI) može se smatrati inteligecijom koju pokazuju računala za razliku od prirodne intelligencije koju pokazuju ljudi i druga živa bića koja uključuje svijest i emocionalnost.

Računalsvo u oblaku

- Računalstvo u oblaku (engl. Cloud computing) je dostupnost računalnih resursa (posebice pohrane podataka i procesorske snage) prema potrebama korisnika ali bez izravne administracije od strane korisnika.
- Podatkovni centri dostupni su mnogim korisnicima putem Interneta.



Želite postati stručnjak za podatke?

- Steknite diplomu prvostupnika u području informatike, računarstva, matematike, fizike ili povezanog područja;
- Steknite diplomu magistra u području baza podataka ili povezanog područja;
- Steknite iskustvo u području kojim se želite baviti (npr. zdravstvo, poslovanje, fizika, …);
- Iskoristite svaku mogućnost učenja i usavršavanja vaših znanja i vještina.

Pitanja?

Izvori

1. Unknown Author: E-discovery presentation*
2. Oard, D. W., Baron, J. R., Hedin, B., Lewis, D. D., & Tomlinson, S. (2010). Evaluation of information retrieval for E-discovery. Artificial Intelligence and Law, 18(4), 347-386.
3. Oard, D. W., & Webber, W. (2013). Information retrieval for e-discovery. Foundations and Trends® in Information Retrieval, 7(2–3), 99-237.
4. Baron, J. R. (2011). Law in the age of exabytes: Some further thoughts on 'information inflation'and current issues in e-discovery search. Richmond Journal of Law & Technology, 17(3), 9.
5. Allman, T. Y. (2006). Managing Preservation Obligations After the 2006 Federal E-Discovery Amendments. Rich. JL & Tech., 13, 1.
6. Conrad, J. G. (2010). E-Discovery revisited: the need for artificial intelligence beyond information retrieval. Artificial Intelligence and Law, 18(4), 321-345.
7. Lee, T., Kim, H., Rhee, K. H., & Shin, U. S. (2013). Design and Implementation of E-Discovery as a Service based on Cloud Computing. Computer Science and Information Systems, 10(2), 703-724.
8. McNee, S. M., & Arnette, B. (2008, April). Productivity as a metric for visual analytics: reflections on e-discovery. In Proceedings of the 2008 Workshop on BEyond time and errors: novel evaLuation methods for Information Visualization (p. 1). ACM.
9. Lee, T., Kim, H., Rhee, K. H., & Shin, S. U. (2013). Implementation and performance of distributed text processing system using hadoop for e-discovery cloud service. Journal of Internet Services and Information Security (JISIS), 4(1), 12-24.

*This presentation is partially based on another presentation which wasn't signed or had any indication about authorship. I have done my best to find the original author but without success. In case you are the original author, or know the original author, please contact me, I will be glad to add you the sources to the acknowledgements.