



Učinkovita stvarnovremenska obrada brzih geoprostornih podataka (RETROFIT)

Javna uvodna prezentacija projekta

Ovaj rad je sufinancirala
Hrvatska zaklada za znanost
projektom UIP-2017-05-9066

Osnovne informacije o projektu

- Naziv projekta: Učinkovita stvarnovremenska obrada brzih geoprostornih podataka
- Izvor financiranja: Hrvatska zaklada za znanost (HRZZ)
- Vrsta natječaja: Uspostavni istraživački projekti (UIP-2017-05)
- Trajanje:
 - od 01. ožujka 2018.
 - do 28. veljače 2023.
- Vrijednost financiranja: 1.263.693,24 HRK
- Znanstvena područja: Tehničke i prirodne znanosti
 - Interdisciplinarni projekt
- Istraživanje je vezano uz aktivnosti Znanstvenog centra izvrsnosti za znanost o podacima i kooperativne sustave
 - Istraživačka jedinica za znanost o podacima
 - Istraživačko područje (IP) 3: Heterogeno računarstvo i napredne usluge u oblaku
 - IP 3.5 Raspodijeljena obrada tokova podataka u stvarnom vremenu

Projektni tim

□ Voditelj

- izv. prof. dr. sc. Krešimir Pripužić (FER)



□ Suradnici

- doc. dr. sc. Mladen Maradin (PMF)
- dr. sc. Martina Marjanović (FER)
- Aleksandar Antonić, mag. ing. inf. et. (FER)
- mr. sc. Ivan Livaja (VUS)
- dr. sc. Mirjana Pripužić (VERN)

- novozaposleni doktorand(i)

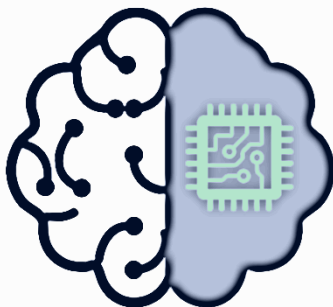


Glavni znanstveni ciljevi projekta

- Razviti **algoritme za stvarnovremensku usporedbu (filtriranje) brzih geoprostornih podataka** (O4)
- **Primijeniti tehnike strojnog učenja na brze geoprostorne podatke** (O5)
- Razviti raspodijeljeno rješenje za **stvarnovremensku obradu brzih geoprostornih podataka** u računalnom grozdu (O6)



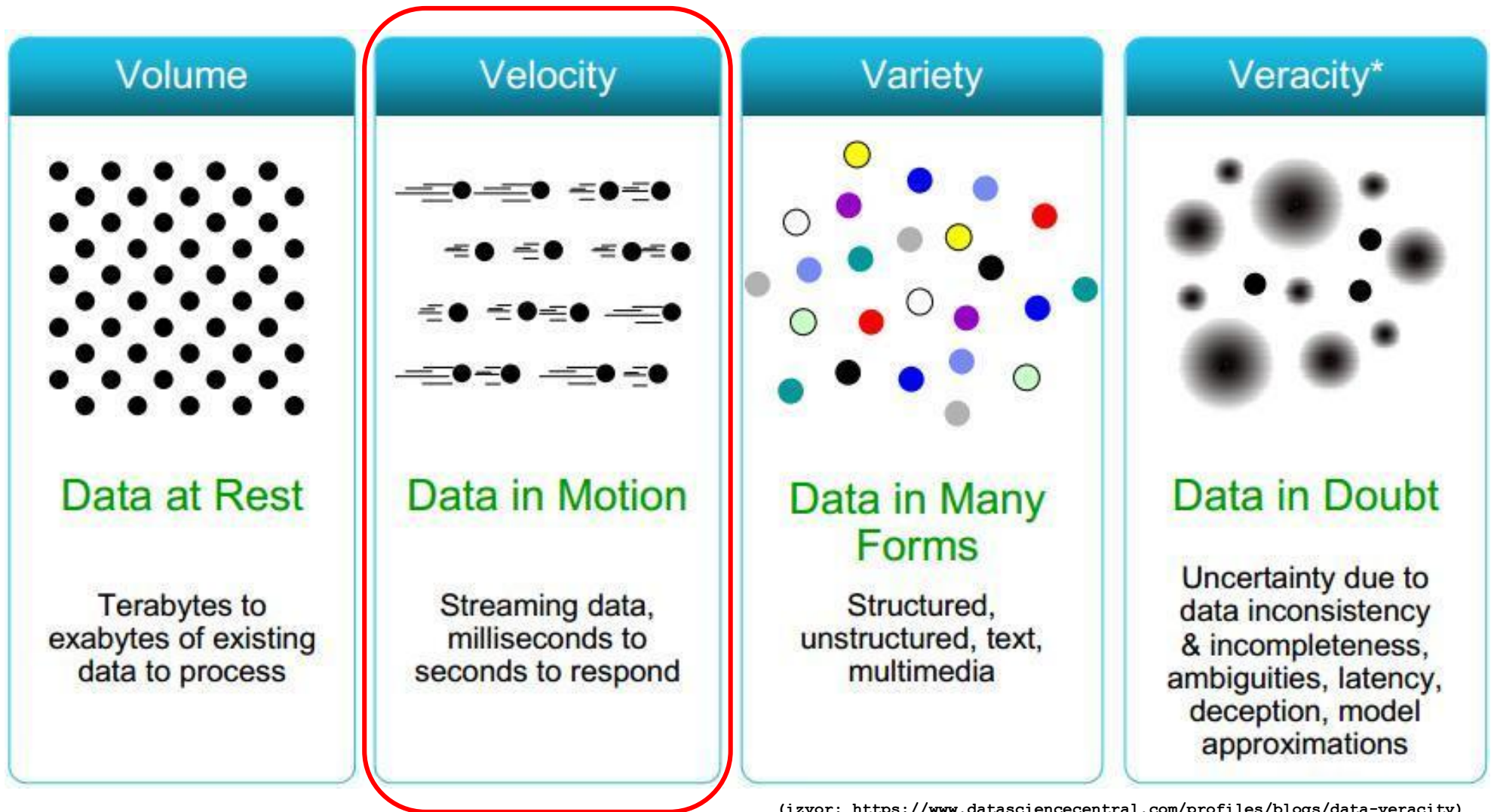
+



=



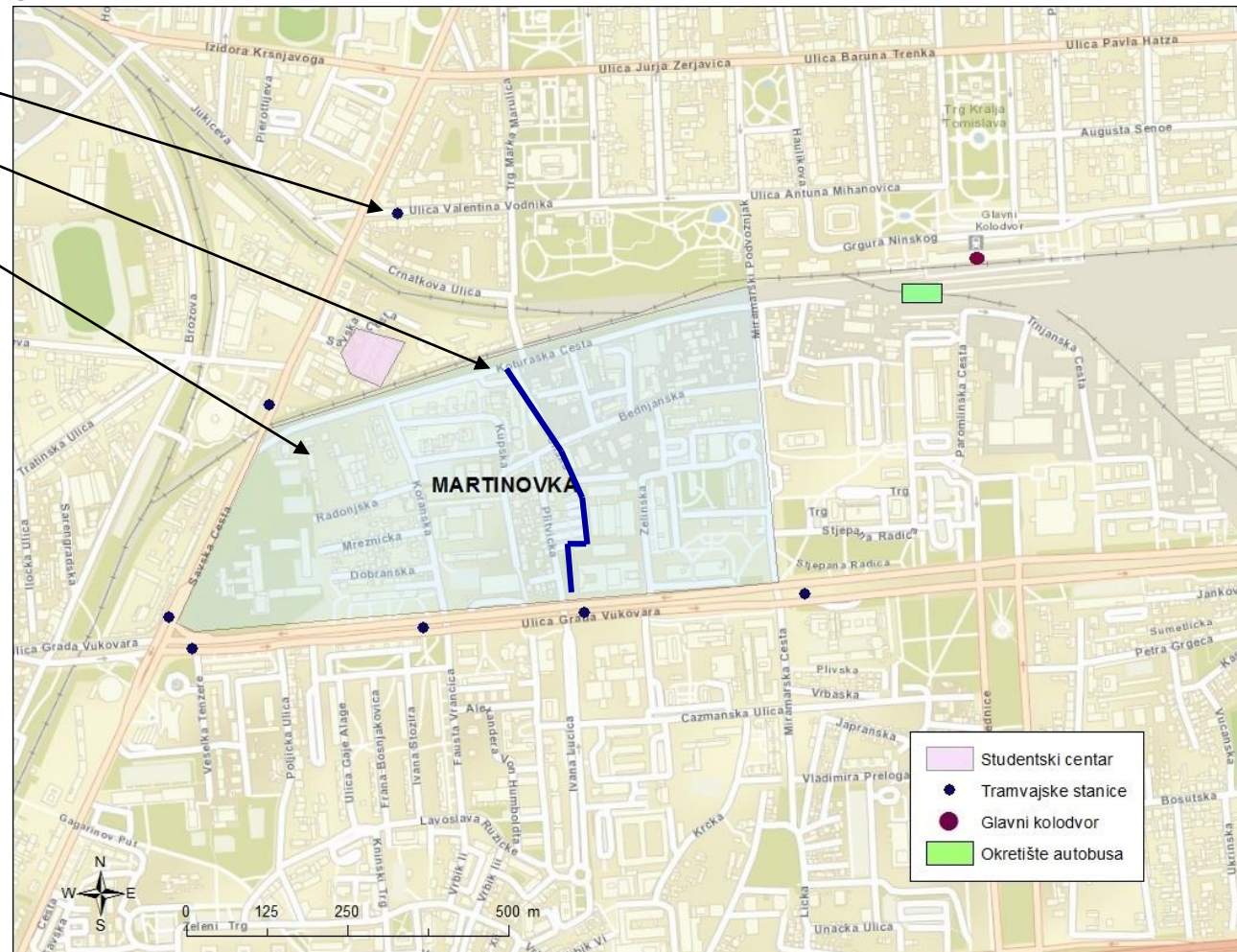
Brzi podaci (*Fast Data*)



Geoprostorni podaci

□ Podaci s geografskom komponentom

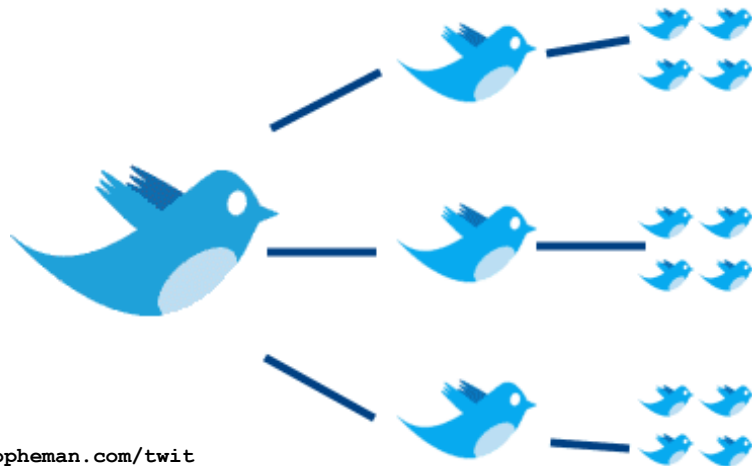
- Point
- Polyline
- Polygon



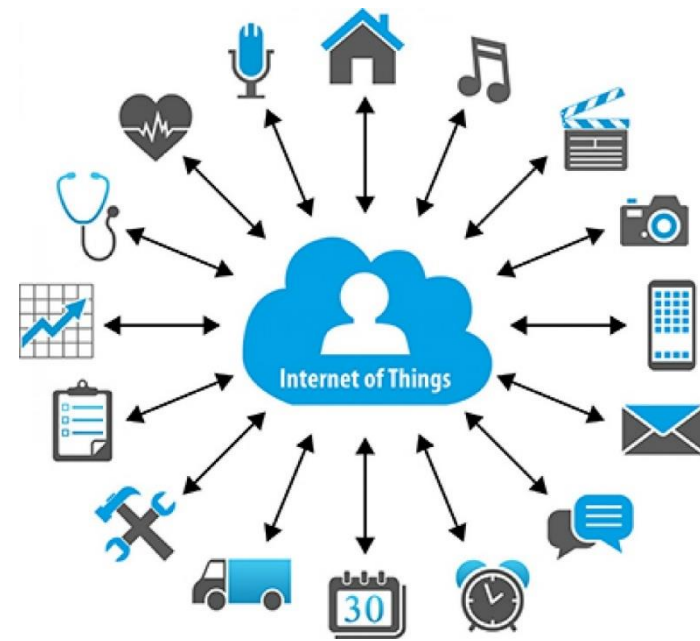
(izvor: <http://atlas.geog.pmf.unizg.hr/gkp/martinovka/index.html>)

Brzi geoprostorni podaci

- Tok podataka velikog intenziteta koji sadrži podatkovne objekte s geoprostornom komponentom
 - Strukturirani podaci
 - senzorski podaci s Interneta stvari
 - Nestrukturirani podaci (tekst)
 - objave na društvenoj mreži Twitter



(izvor:
<http://labs.topheman.com/twitter-stream-channels/>)



(izvor:
<https://semielectronics.com/sensors-lifblood-internet-things/>)

Filtriranje brzih geoprostornih podataka

- Konzumente brzih geoprostornih podataka zanima samo podskup objava (objavljenih brzih geoprostornih podataka)
- Konzumenti interes za objavama trebaju izraziti pretplatom
 - **Geoprostorni objekt** koji predstavlja geografsko područje interesa
 - **Dodatni uvjeti na sadržaj** objava kojima se precizno definira interes
- Glavni problem je **učinkovito u stvarnom vremenu pronaći pretplate koje zanima pojedina objava** (usporedba pretplata s objavama)
- Rješenje se svodi na sustav objavi-pretplati s više razina usporedbe
 1. Usporedba geoprostornih objekata
 2. Usporedba sadržaja

(izvor:
<https://prateekvishshi.com/2014/01/11/what-is-fuzzy-matching/>)



Praktična izvedba filtriranja brzih geoprostornih podataka

□ Centralizirano rješenje

- Koristi (jednostavni) centralizirani prostorni indeks (grid, quadtree, STR tree) za pretplate
- Postoje gotove implementacije indeksa koje su otvorenog koda
- Nije skalabilno
- Jedinствена točka ispada

□ Raspodijeljeno rješenje

- Koristi (komplicirani) raspodijeljeni prostorni indeks za pretplate
- Ne postoje gotove implementacije raspodijeljenih indeksa
- Skalabilno
- Ne postoji jedinствена točka ispada
- Potrebno je upravljati raspodijeljenim radnim procesima

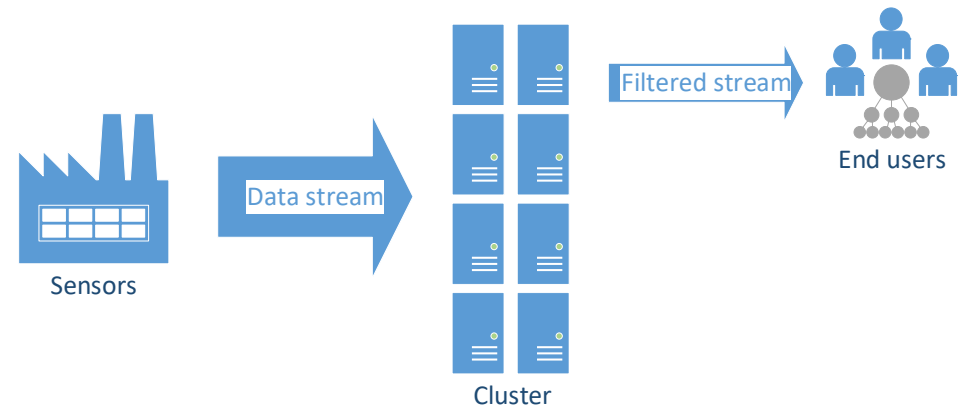
- Izravno upravljanje (npr. A. Antonić, M. Marjanović, K. Pripužić, I. Podnar Žarko: „A Mobile Crowd Sensing Ecosystem Enabled by CUPUS: Cloud-based Publish/Subscribe Middleware for the Internet of Things”, Future Generation Computer Systems, 56 (2016), 607-622)
- Korištenjem odgovarajućih programskih okvira za računalni grozd (**Apache Hadoop YARN** ili Apache Mesos)

(izvor:
<https://mindmajix.com/apache-hadoop-yarn>)

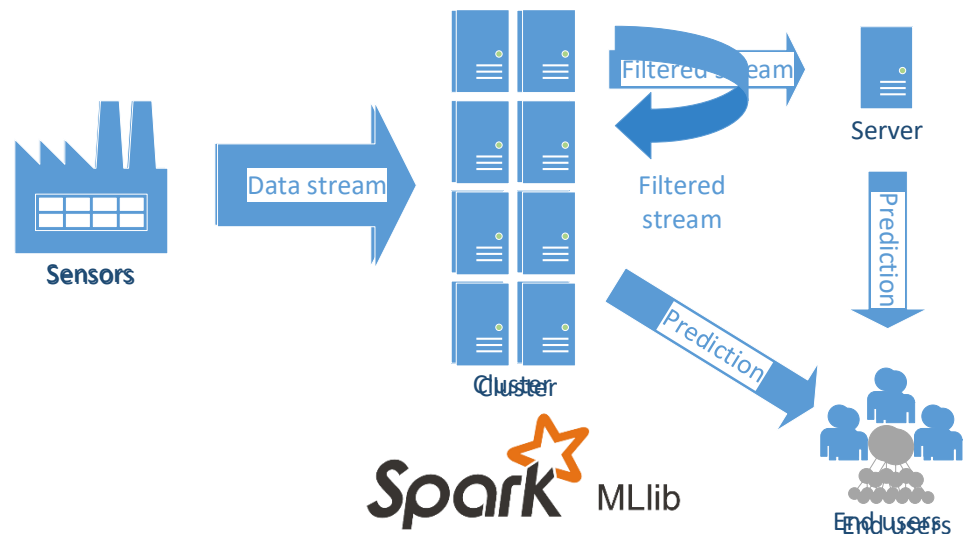


Konzumenti brzih geoprostornih podataka

- Klasični pristup (računalni grozd samo za filtriranje)



- Naš pristup (računalni grozd za sve)
 - Grozd može pohraniti veliku količinu podataka
 - Grozd može brzo obraditi veliku količinu podataka
 - Grozd se već koristi za filtriranje brzih geoprostornih podataka (konzumenti objava će biti procesi koji vrše predikciju)
 - Grozd podržava gotov API za strojno učenje (npr. **Apache Spark MLlib**)

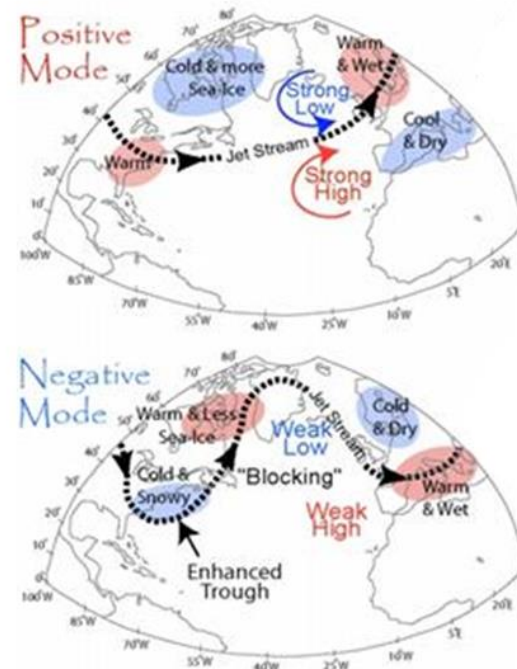


Potencijalni studijski slučajevi brzih geoprostornih podataka

- Prognoze pojave ekstremnih protoka s obzirom na količinu padalina u porječju
- Telekonekcije (uzročno-posljedične veze između vremenskih prilika u međusobno udaljenim dijelovima Zemlje)
 - utjecaj klimatskih ciklusa na vremenske obrasce u Hrvatskoj
 - North Atlantic oscillation, Artic oscillation...
- Ostali studijski slučajevi



North Atlantic Oscillation



(izvor: <https://muchadoaboutclimate.wordpress.com/2013/09/06/the-north-atlantic-oscillation/>)

Ostali ciljevi projekta

- Uspostavljanje istraživačke grupe i **Laboratorija za Brze podatke**
- Financiranje doktoranda/doktoranada do obrane doktorske disertacije (i nakon nje)
- Postavljanje i održavanje laboratorijske opreme
 - **Računalni grozd od 16 računala** s neophodnom mrežnom opremom
 - Studentski laboratorij (monitori, tipkovnice, miševi)
- Uključivanje studenata u istraživanja
- Širenje informacija o postignućima na projektu
 - **Sudjelovanje na barem 12 konferencija**
 - Objava barem 4 članka u uglednim časopisima
- **Pohađanje barem 5 ljetnih škola**
- Nabavljanje najnovije znanstvene literature
- Prijavljivanje na ostale izvore financiranja