

Универзитет у Новом Саду, Факултет техничких наука

ОАС Софтверско инжењерство и информационе технологије

Организација података

Увод – подаци

Садржај

- **Подаци**
- Структуре података
- Складиштење података
- Ресурси

Подаци

- Подаци
 - подаци представљају предмет изучавања и интензивне употребе у многим областима у вези с рачунарством и информатиком
 - информациони системи
 - системи за управљање базама података
 - инжењеринг података
 - наука о подацима
 - вештачка интелигенција и машинско учење
 - ...

Подаци

- Подаци – језички аспект
 - податак
 - енгл. *datum*
 - лат. *datum*
 - партицип перфекта пасива од глагола *do*, *dare*, *dedi*, *datum* (дати), номинатив једнине средњег рода
 - подаци
 - енгл. *data*

Подаци

Давимо се у информацијама и гладни смо знања.

Радерфорд Д. Роџер (Hastie et al., 2009)

Подаци

- Хијерархија мудрости (Rowley, 2007; Ackoff, 1989)
 - хијерархија подаци–информације–знање–мудрост (ПИЗМ)
 - енгл. *data–information–knowledge–wisdom hierarchy, DIKW*
 - хијерархија знања, пирамида знања...
 - модел у информационим наукама
 - постоје разне варијације у структури модела
 - постоје разна схватања основних појмова



Подаци

- Хијерархија мудрости (Rowley, 2007; Ackoff, 1989)
 - компоненте модела
 - подаци, информације, знање, мудрост
 - везе између компоненти модела
 - формирање информација на основу података
 - формирање знања на основу информација
 - формирање мудрости на основу знања



Подаци

- Хијерархија мудрости
 - примери дефиниција основних концепата
 - подаци
 - дискретне, објективне чињенице или опажања, који су неорганизовани и необрађени, и не преносе никакво посебно значење (Rowley, 2007)
 - информације
 - подаци који су организовани тако да имају значење и вредност за примаоца (Rowley, 2007)



Подаци

- Хијерархија мудрости
 - примери дефиниција основних концепата
 - знање
 - разумевање информација засновано на њиховој перципиранијо важности или значају за проблемско подручје (Rowley, 2007)
 - мудрост
 - акумулирано знање, које ти омогућава да разумеш како да примениш концепте из једног домена на нове ситуације или проблеме (Rowley, 2007)



Подаци

- Хијерархија мудрости
 - пример схватања повезаности између основних концепата
 - подаци самостално немају значај, али након што су интерпретирани и погодно доведени у узајамну везу, они пружају информације које нам омогућавају да побољшамо наше знање о свету(Atzeni et al., 1999)



Подаци

- Хијерархија мудрости
 - примери за податке (могу подразумевати различита схватања самог концепта)
 - 20
 - Европа
 - Уторак је.
 - Тихо је.
 - За Нови Сад температура ваздуха у степенима Целзијуса је 15 (у овом тренутку).



Садржај

- Подаци
- **Структуре података**
- Складиштење података
- Ресурси

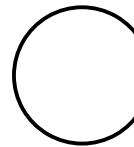
Структуре података

- Структура података је оријентисани граф с функцијама које његовим чворовима и ивицама придржују семантику (Mogin, 2008)
 - граф $G = (S, R)$
 - S – коначан скуп, представља чворове
 - R – релација у скупу S , представља ивице
 - придрживање семантике
 - чворови семантички одређени над скупом V
 - $f : S \rightarrow V$
 - ивице семантички одређене над скупом V'
 - $g : R \rightarrow V'$
 - може постојати и за чворове и за ивице више функција за придрживање семантике над разним скуповима

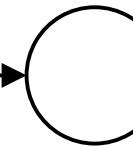
Структуре података

- Структура података – пример

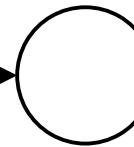
$s_1: (E27, 8)$



$s_2: (E14, 6)$



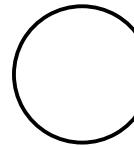
$s_3: (E27, 4)$



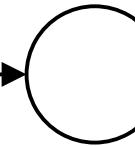
Структуре података

- Структура података – пример

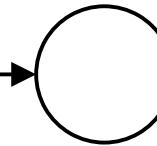
$s_1: (E27, 8)$



$s_2: (E14, 6)$



$s_3: (E27, 4)$

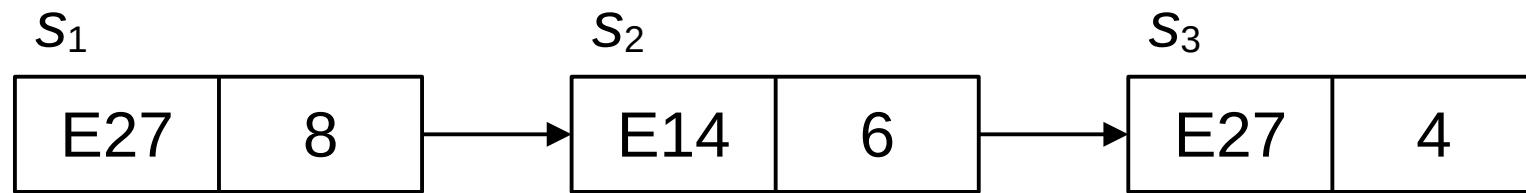


подаци о сијалицима

(придружене семантика: тип грла и снага у ватима)

Структуре података

- Структура података – пример (другачији начин приказа)



Структуре података

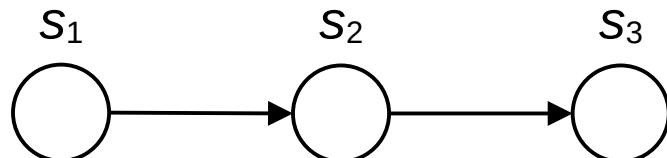
- Врсте структура података (Mogin, 2008)
 - према дозвољеном броју непосредних претходника и следбеника једног чвора (семантика се не узима у обзир)
 - линеарне структуре
 - структуре стабла
 - мрежне структуре

Структуре података

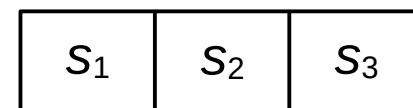
- Врсте структура података – линеарне структуре (Mogin, 2008)
 - сваки чврт може имати највише једног директног претходника и највише једног директног следбеника
 - врсте линеарних структура података
 - ацикличне
 - познате и као ланци, отворене или просте листе и низови
 - цикличне
 - познате и као затворене листе и прстенови

Структуре података

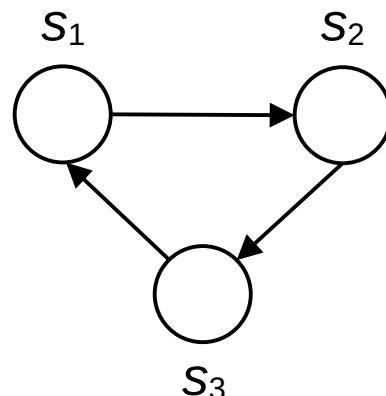
- Врсте структура података – линеарне структуре



*пример ацикличне
линеарне структуре*



*пример ацикличне
линеарне структуре
(другачији приказ)*



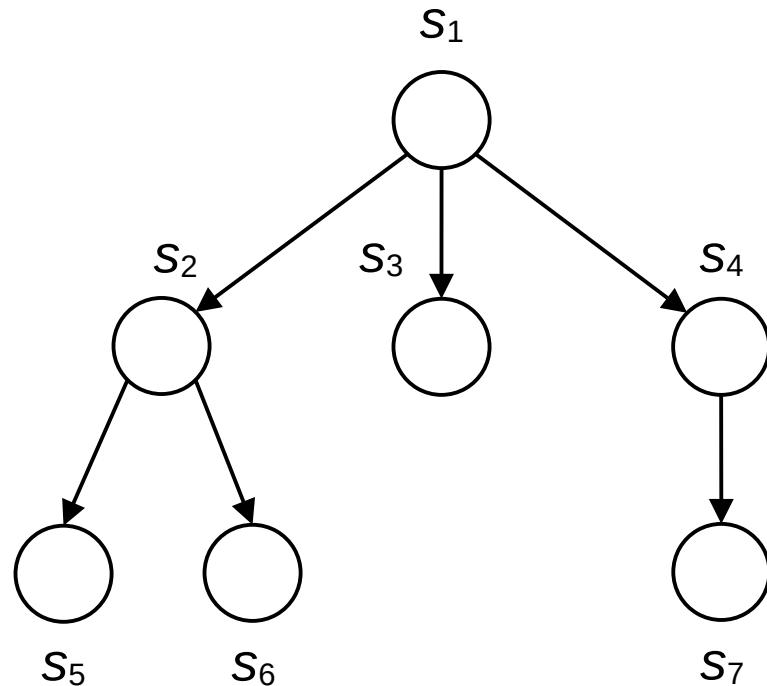
*пример цикличне
линеарне структуре*

Структуре података

- Врсте структура података – структуре стабла (Mogin, 2008)
 - сваки чврт може имати највише једног директног претходника и до n директних следбеника
 - $0 \leq n \leq N - 1$
 - N представља кардинални број скупа који се снабдева структуром
 - структуре стабла су ацикличног карактера

Структуре података

- Врсте структура података – структуре стабла



пример структуре стабла

Структуре података

- Врсте структура података – структуре стабла (Mogin, 2008)
 - корен је чвр без иједног претходника
 - лист је чвр без иједног следбеника
 - чворови стабла су организовани по нивоима хијерархије
 - корен стабла се налази на 1. нивоу
 - чвр се налази на k -том нивоу ако је на крају пута који је дужине $k - 1$ и који креће од корена
 - висина стабла h одговара броју нивоа хијерархије у стаблу
 - стабло је реда n (n -арно) ако сваком чврлу који није лист одговара максимално n директно подређених чворова
 - $1 < n < N$

Структуре података

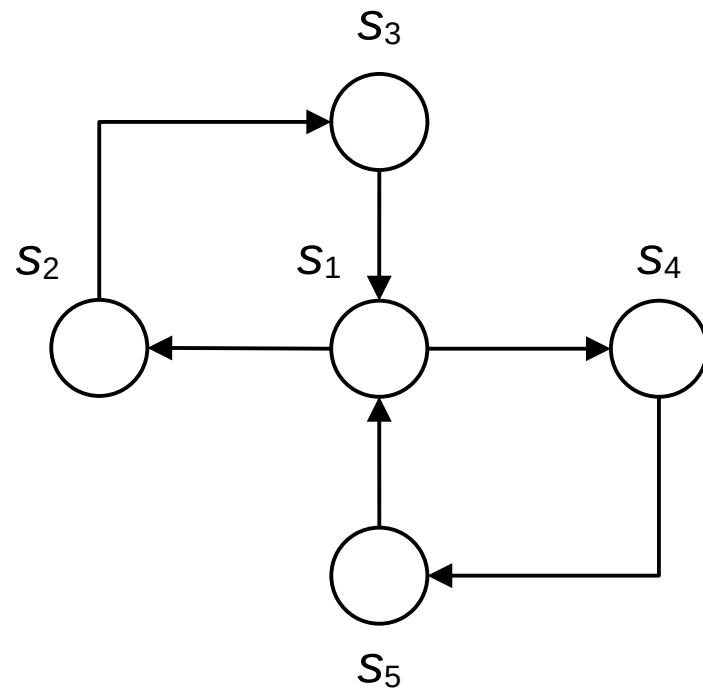
- Врсте структура података – структуре стабла (Mogin, 2008)
 - стабло је празно ако је без корена
 - стабло је пуно ако су сви листови на истом растојању од корена
 - стабло реда n је комплетно ако сви чворови који нису листови имају n директно подређених чворова
 - стабло је балансирано ако за сваки чвор важи да се број чворова у сваком његовом подстаблу не разликује за више од један
 - стабло реда n је оптимално балансирано ако су сви чворови на нивоима од 1 до $h - 2$ комплетни
 - чвор је комплетан ако има свих n директно подређених чворова

Структуре података

- Врсте структура података – мрежне структуре (Mogin, 2008)
 - сваки чврт може имати до N директних претходника и до N директних следбеника
 - N представља кардинални број скупа који се снабдева структуром
 - структура се сматра мрежном ако постоји чврт који има више од једног директног претходника
 - мрежне структуре су обично цикличног карактера

Структуре података

- Врсте структура података – мрежне структуре



пример мрежне структуре

Структуре података

- Врсте структура података (Mogin, 2008)
 - према апстрактности скупа над којим се одређује семантика
 - структуре над скупом обележја (структуре обележја)
 - структуре над скупом података (структуре података)
 - логичке структуре података
 - физичке структуре података

Садржај

- Подаци
- Структуре података
- **Складиштење података**
- Ресурси

Складиштење података

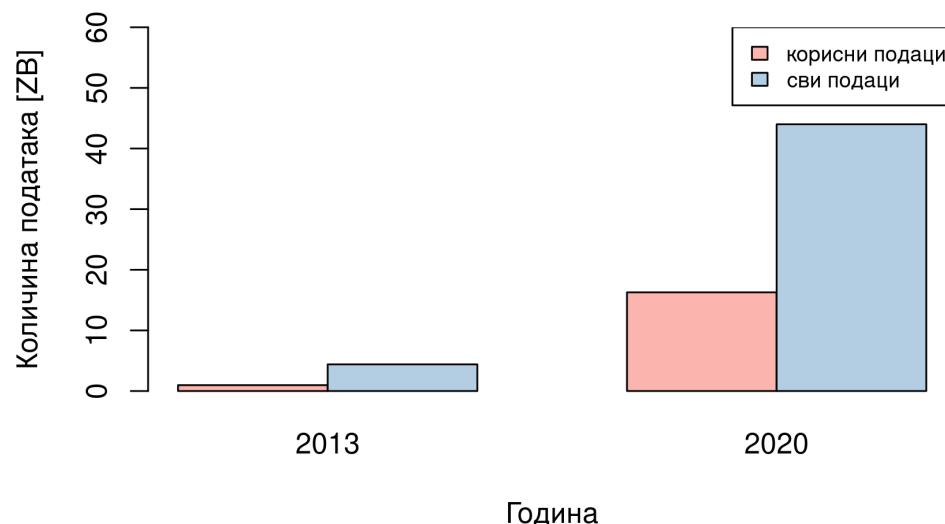
- Количина података у свету

- истраживање компаније *IDC* из 2014. године

(Dell Technologies, 2014; Cyclone Interactive, 2014)

- студија *The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things*
 - процене за дигитални универзум

Процене компаније IDC (2014)



Складиштење података

- Количина података у свету
 - истраживање компаније *IDC* из 2014. године (Dell Technologies, 2014; Cyclone Interactive, 2014)
 - студија *The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things*
 - процене за дигитални универзум
 - 2013. година
 - 4,4 ZB је величина дигиталног универзума
 - 22% је удео корисних података (података погодних за анализу)
 - већина података у вези са зрелим тржиштима (САД, Западна Европа, Јапан, Канада, Аустралија, Нови Зеланд)

Складиштење података

- Количина података у свету

- истраживање компаније *IDC* из 2014. године
 - студија *The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things*
 - процене за дигитални универзум
 - 2013. година
 - 4,4 ZB је величина дигиталног универзума
 - 22% је удео корисних података (података погодних за анализу)
 - већина података у вези са зрелим тржиштима (САД, Западна Европа, Јапан, Канада, Аустралија, Нови Зеланд)
 - 2020. година
 - 44 ZB је величина дигиталног универзума
 - преко 35% је удео корисних података
 - већина података у вези са растућим тржиштима (Кина, Бразил, Индија, Русија, Мексико)

Складиштење података

- Количина података у свету
 - истраживање компаније *IDC* из 2018. године (Business Wire, 2018)
 - студија *The Digitization of the World – From Edge to Core*
 - процене за глобалну датасферу
 - квантификација количине података који су креирани, прикупљени и репликовани широм света
 - 2025. година
 - 175 ZB је потенцијална величина глобалне датасфере
 - 49% ускладиштених података могло би бити у јавним клауд окружењима

Складиштење података

- **Дата центри**

- *Дата центар је физичка просторија, зграда или постројење у којем је смештена ИТ инфраструктура за изградњу, извршавање и испоручивање апликација и сервиса. Такође складишти податке повезане с тим апликацијама и сервисима и управља тим подацима.* (IBM, 2024)

Складиштење података

- **Дата центри** (IBM, 2024; Amazon Web Services, n.d.)
 - инфраструктура у дата центрима
 - рачунска инфраструктура
 - складишна инфраструктура
 - мрежна инфраструктура
 - помоћна инфраструктура
 - често се примењује виртуализација физичких ресурса (централних процесорских јединица, складишних уређаја...)

Складиштење података

- Дата центри
 - класификација дата центара према организацији *Uptime Institute*
(Uptime Institute, n.d.; Stansberry, 2021; Amazon Web Services, n.d.)
 - дата центри 1. реда (основног капацитета)
 - основни ниво могућности
 - заступљени непрекидно напајање за случај аномалија, простор за ИТ системе, посебна расхладна опрема и генератор за случај нестанка струје
 - потпуно заустављање рада неопходно ради одржавања и поправки
 - неочекивано заустављање рада има утицаја
 - годишњи застој у раду од ≈ 29 сати
 - ...

Складиштење података

- Дата центри
 - класификација дата центара према организацији *Uptime Institute*
(Uptime Institute, n.d.; Stansberry, 2021; Amazon Web Services, n.d.)
 - дата центри 2. реда (редудантног капацитета)
 - додатне могућности за напајање и хлађење
 - заступљени генератори, јединице за хлађење, модули за непрекидно напајање, пумпе, резервоари за гориво...
 - неочекивано заустављање рада има утицаја
 - годишњи застој у раду од ≈ 22 сата
 - ...

Складиштење података

- Дата центри

- класификација дата центара према организацији *Uptime Institute*
(Uptime Institute, n.d.; Stansberry, 2021; Amazon Web Services, n.d.)

- дата центри 3. реда (конкурентно одрживи)
 - могућност конкурентног одржавања
 - у случају одржавања или замене опреме, није потребно заустављање рада
 - било који део може бити искључен без утицаја на ИТ делатност
 - годишњи застој у раду од $\approx 1,6$ сати
 - ...

Складиштење података

- Дата центри
 - класификација дата центара према организацији *Uptime Institute*
(Uptime Institute, n.d.; Stansberry, 2021; Amazon Web Services, n.d.)
 - дата центри 4. реда (тolerантни на испаде)
 - постојање неколико независних и физичких изолованих система
 - одвојеност система како неки догађај не би угрозио више система
 - планирани и непланирани догађаји немају утицаја
 - у случају заустављања делова ради одржавања, може бити повећан ризик од прекида ако наступи проблем
 - годишњи застој у раду од ≈ 26 минута

Складиштење података

- **Дата центри**^(IBM, 2024)
 - врсте сервера
 - рек сервери
 - блејд сервери
 - мејнфрејмови
 - врсте система за складиштење података
 - системи за блоковско складиштење
 - системи за складиштење датотека
 - системи за складиштење објекта

Садржај

- Подаци
- Структуре података
- Складиштење података
- **Ресурси**

Ресурси

- Извори и литература
 - Pavle Mogin. Strukture podataka i organizacija datoteka. 3. izdanje. Računarski fakultet (Beograd, Srbija), CET (Beograd, Srbija). 2008.
 - Glava 1. Klasifikacija struktura podataka
 - Paolo Atzeni, Stefano Ceri, Stefano Paraboschi, Riccardo Torlone. Database Systems: Concepts, Languages and Architectures. McGraw-Hill (Maidenhead, England, UK). 1999.
 - Trevor Hastie, Robert Tibshirani, Jerome Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd edition. Springer (New York, NY, USA). 2009.

Ресурси

- Извори и литература
 - Jennifer Rowley. The Wisdom Hierarchy: Representations of the DIKW hierarchy. *Journal of Information Science*. 2007; 33(2); 163–180.
 - Russel L. Ackoff. From Data to Wisdom. *Journal of Applied Systems Analysis*. 1989; 16; 3–9.
 - Dell Technologies. Digital Universe Invaded by Sensors. [Internet]. 2014. URL:
<https://www.dell.com/en-us/dt/corporate/newsroom/announcements/2014/04/20140409-01.htm>

Ресурси

- Извори и литература
 - Cyclone Interactive. Executive Summary: Data Growth, Business Opportunities, and the IT Imperatives | The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things. [Internet]. 2014. URL: <http://idccclients.cycloneinteractive.net/emc-digital-universe-iview-2014/executive-summary.htm>
 - Business Wire. Seagate Launches New Data-Readiness Index Revealing Impact Across Four Global Industries as 30 Percent of Data Forecasted to be Real-Time by 2025. [Internet]. 2018. URL: <https://www.businesswire.com/news/home/20181126005585/en/Seagate-Launches-New-Data-Readiness-Index-Revealing-Impact-Across-Four-Global-Industries-as-30-Percent-of-Data-Forecasted-to-Be-Real-Time-by-2025>

Ресурси

- Извори и литература
 - IBM. What is a Data Center?. [Internet]. 2024. URL: <https://www.ibm.com/topics/data-centers>
 - Amazon Web Services. What is a Data Center?. [Internet]. (n.d.). URL: <https://aws.amazon.com/what-is/data-center/>
 - Uptime Institute. Tier Classification System. [Internet]. (n.d.). URL: <https://uptimeinstitute.com/tiers>
 - Matt Stansberry. Explaining the Uptime Institute's Tier Classification System (April 2021 Update). [Internet]. 2021. URL: <https://journal.uptimeinstitute.com/explaining-upptime-institutes-tier-classification-system/>

Ресурси

- Додатни ресурси
 - Google Workspace. Inside a Google Data Center. [Internet]. 2014. URL: <https://www.youtube.com/watch?v=XZmGGAbHqa0>
 - Google Cloud Tech. Google Data Center Security: 6 Layers Deep. [Internet]. 2020. URL: <https://www.youtube.com/watch?v=kd33UVZhnAA>
 - Amazon. What It's Like Inside an AWS Data Center. [Internet]. 2023. URL: <https://www.aboutamazon.com/news/aws/aws-data-center-inside>

Ресурси

- Додатни ресурси
 - Microsoft. Virtual Datacenter Tour | Azure Global Infrastructure Experience. [Internet]. (n.d.). URL: <https://datacenters.microsoft.com/globe/explore/datacenter/>
 - Hetzner. 360° Tour. [Internet]. (n.d.). URL: <https://www.hetzner.com/unternehmen/360-tour/>
 - imgix. The Newest imgix Datacenter by imgix - Exposure. [Internet]. 2014. URL: <https://photos.imgix.com/building-a-graphics-card-for-the-internet>
 - IBM. A Tour Inside the IBM z16. [Internet]. 2022. URL: <https://developer.ibm.com/blogs/a-tour-inside-the-ibm-z16/>