Факултет организационих наука
Универзитета у Београду
Лабораторија за електронско пословање

Информационе технологије
- од усменог предања до семантичког веба -

Проф. др Божидар Раденковић

boza@elab.rs

Садржај предавања

- Увод – преглед појмова
- Моделирање и модели података
- Генерисање, пренос и чување података
- Претраживање и коришћење информација
- Семантички веб
- Откривање знања у подацима
- Cloud computing базе података
- Правци будућег развоја
- Закључак

УВОД

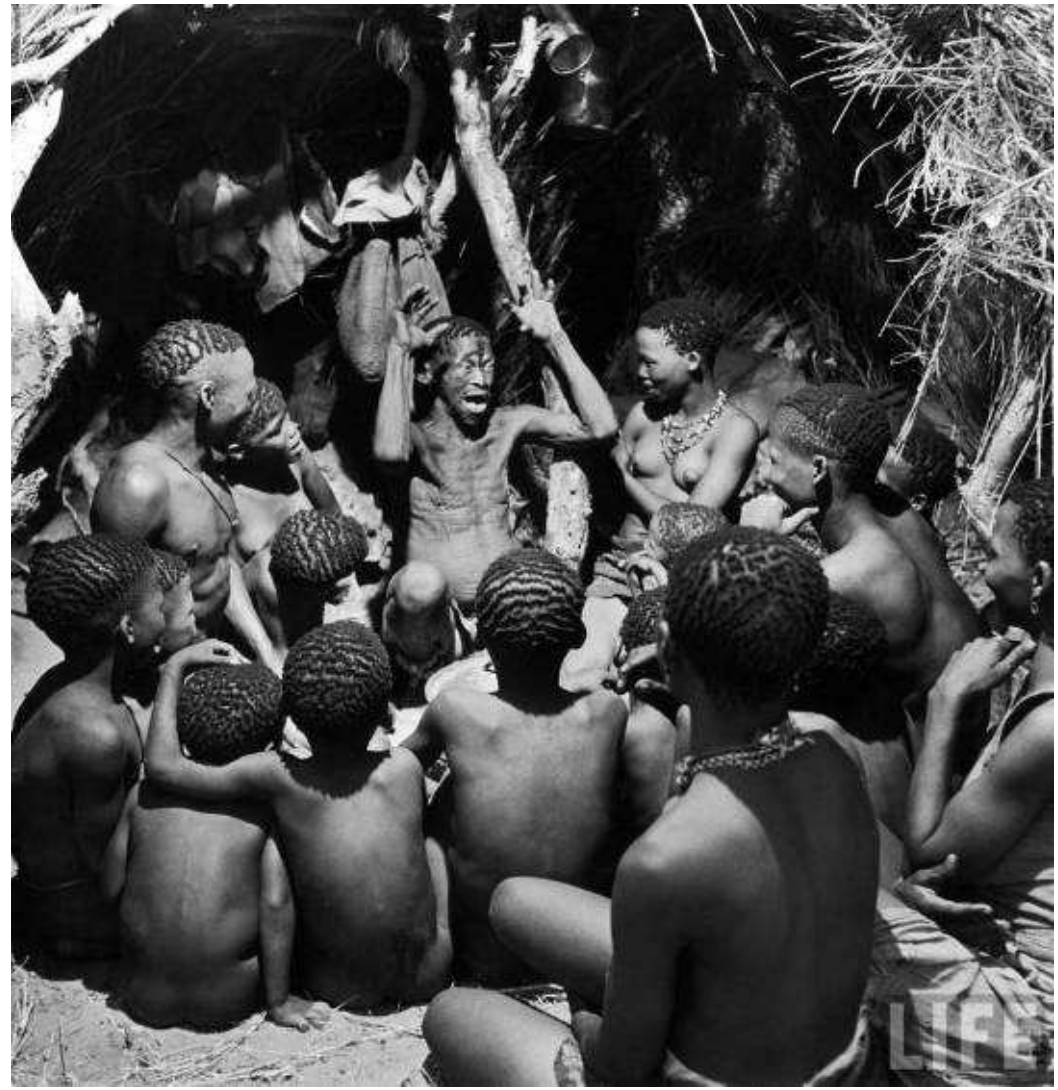
Преглед појмова

- Податак је симболички и формализован приказ чињеница, појмова и инструкција, погодан за комуникацију и интерпретацију.
- Информација је перцепиран податак који представља инкремент знања.
- Комуникација је пренос информација од извора до одредишта у простору и времену.
- Семантика је наука о значењу појмова.
- Знање је скуп факата о неком ентитету.
- Подаци + Семантика = Знање

Усмена предања

- У периоду од пре 160.000 година до пре 6.000 година, усмено предање је било једини начин преноса информација.

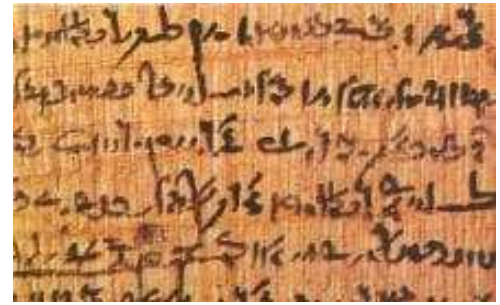
- Недостаци:
  - Дисторзија информација
  - Губљење информација



N.R. Farbman LIFE, "Bushman story-telling", South Africa, 1947.

Писана реч

- Технологије за чување информација:
  - Камене плоче
  - Глинене табле
  - Воштане табле
  - Папирус
  - Пергамент
  - Папир
  - ...



- Недостаци:
  - Секвенцијално читање
  - Спор пренос кроз простор
  - Губитак писаног материјала
  - Потребан већи простор за складиштење

Рачунарске технологије

- Задатак рачунарских технологија је да омогуће ефикасно чување, претраживање и упоређивање података.

- За чување података, неопходна је одговарајућа организација података на медијуму:
  - Релационе базе података
  - Апстракције релационих база података (NO-SQL)
  - Нерелационе базе података
  - XML

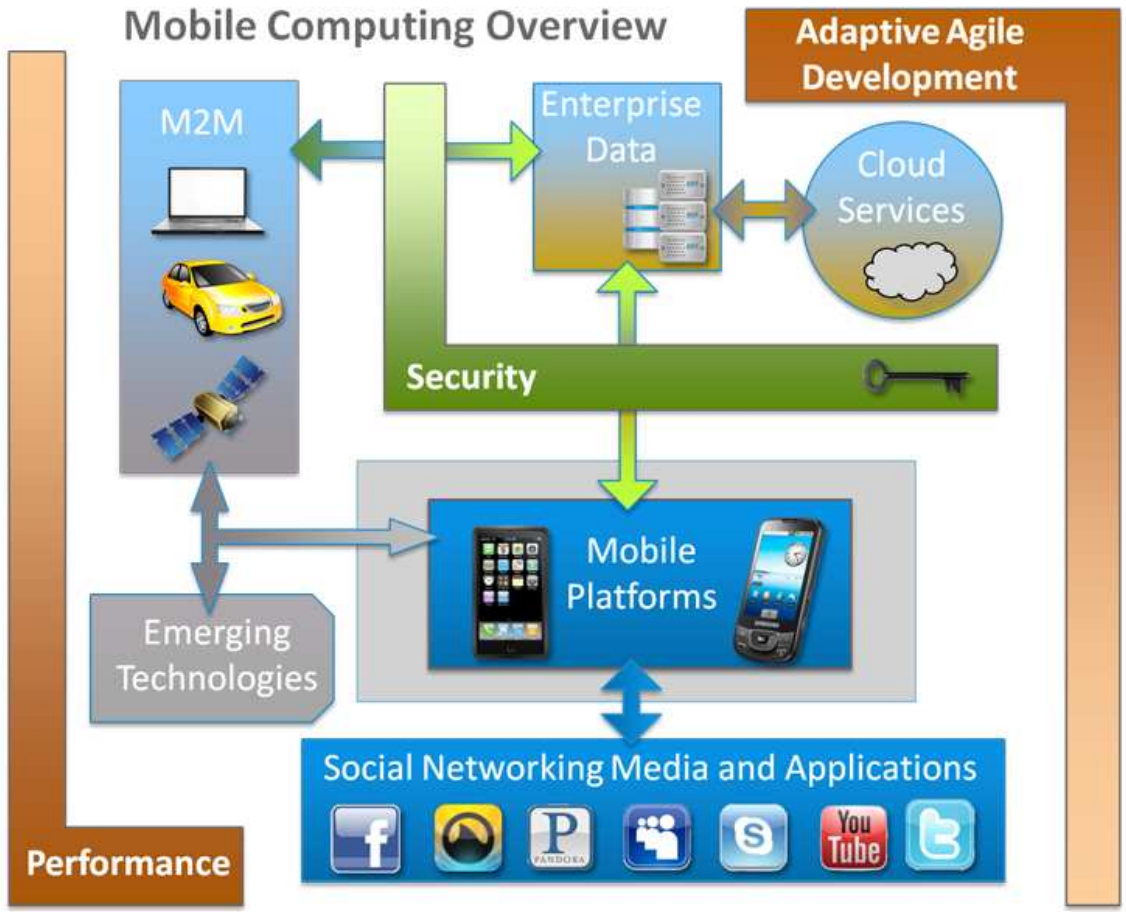
- За брзо претраживање и упоређивање података, користе се:
  - Индекси
  - Хеш табеле

- За ефикасно коришћење података, неопходно је чувати и семантику података:
  - Асоцијације помоћу спољних кључева у релационом моделу
  - Онтологије

- Технологије за подршку доношења одлука:
  - Упити (релациона алгебра)
  - OLAP (On Line Analytical Processing)
  - Истраживање података (data mining)
  - Big data

Мобилно рачунарство

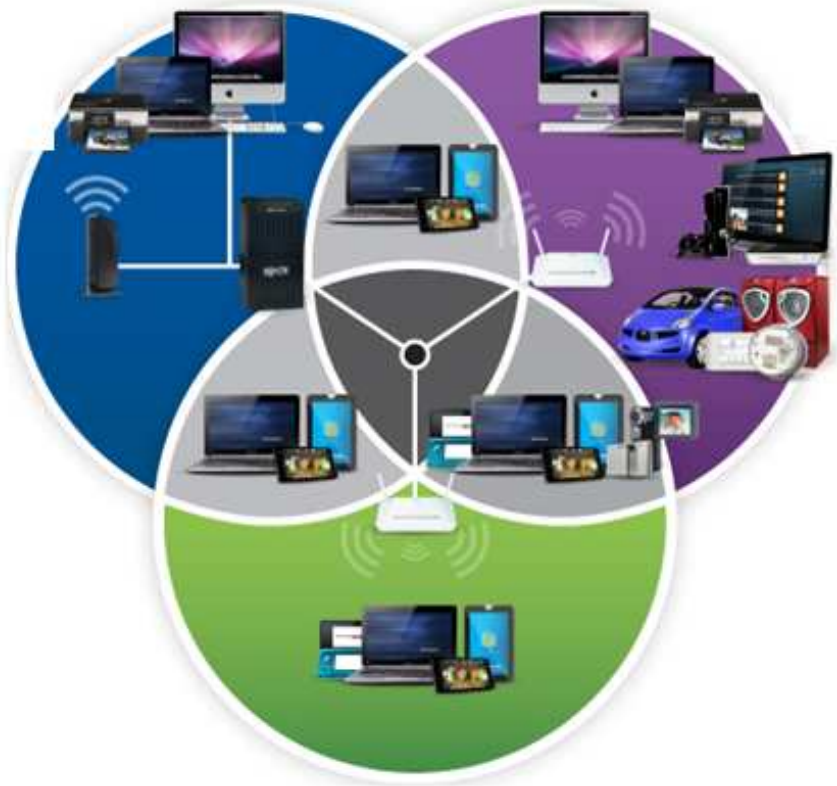
- Подразумева интеракцију између човека и рачунара који се може користити у покрету помоћу мобилне комуникације, хардвера (мобилни уређаји) и софтвера (мобилне апликације).
  Извор: George Forman and John Zahorjan



Интернет интелигентних уређаја

- Појам први пут представљен 1999. године од стране Kevin Ashton-а, извршног директора MIT Auto-ID центра.

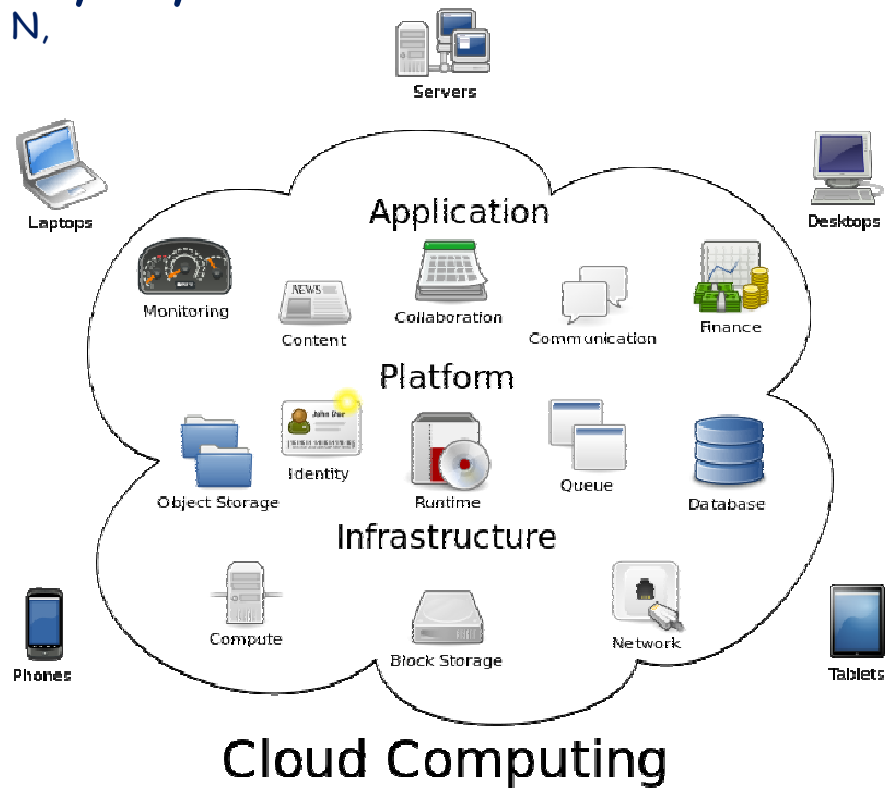
- Сап: "Свет у коме су физички објекти интегрисани у информациону мрежу и постају активни учесници у пословним процесима."



Рачунарство у облаку
Cloud Computing

- Појам се користи да опише низ различитих рачунарских концепата који укључују велики број рачунара који су повезани у реалном времену путем Интернета.

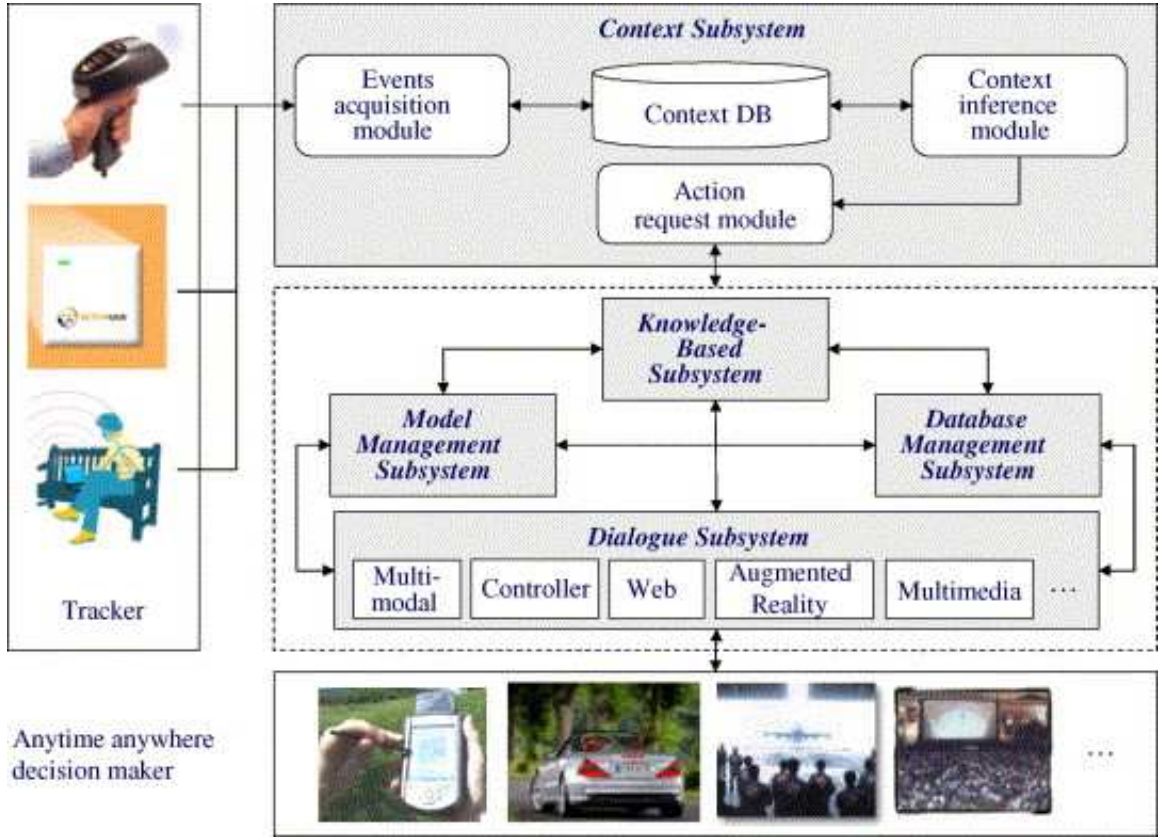
- Cloud computing представља нову парадигму за испоруку рачунарских услуга.
  Извор: Cloud computing for education: A new dawn?, Sultan N, International Journal of Information Management, 2010



Context-aware computing

- Подразумева системе који препознају карактеристике окружења и прилагођавају се контекстима у којима се корисник налази.

Извор: The potential roles of context-aware computing technology in optimization-based intelligent decision-making, Ohbyung K, Expert Systems with Applications, 2006



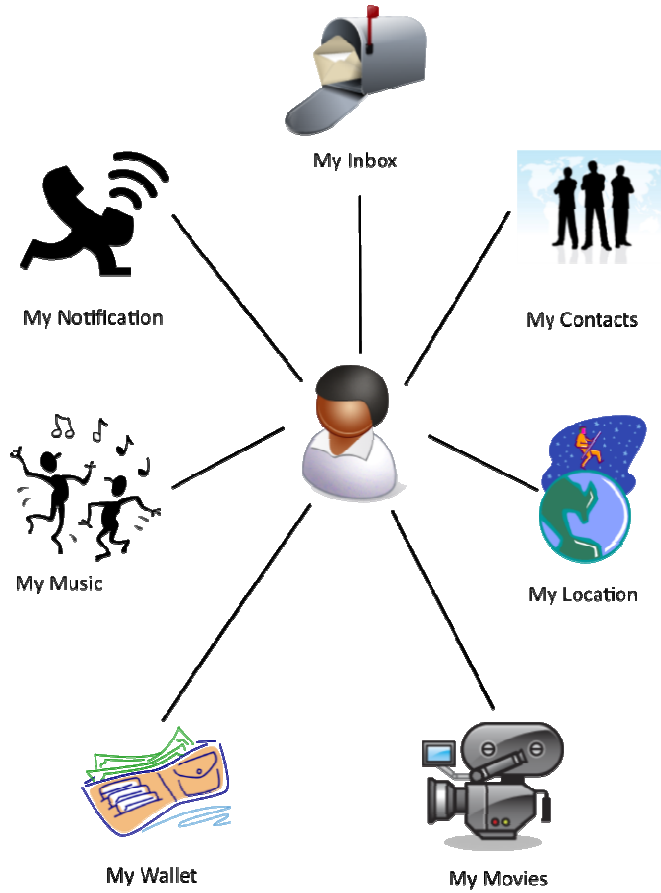
Свеприсутно рачунарство
ubiquitous computing

- Појам је први пут дефинисао Mark Weiser 1988. године.

- Већи број мањих уређаја који имају могућност да реализују рачунарске операције и комуникацију, као што су: смарт телефони, смарт картице, сензорске мреже, RFID.
  - Уређаји се користе у свакодневном животу.
  - Обезбеђују комуникацију између уређаја и окружења

Pervasive Computing

- Стварање паметаног окружења које се прилагођава контексту и потребама корисника. David Garlan, Daniel Siewiorek, Asim Smailagic, и Peter Steenkiste



Pervasive Computing in Home and Community

Друштвени медији
Social computing

- Друштвени медији се односе на онлајн технологије и праксе које се користе за дељење садржаја, мишљења, искустава, перспективе.

извор: Users of the world, unite! The challenges and opportunities of Social Media, Kaplan A, Haenlein M, Business Horizons, 2010.

Виртуелна реалност

Рачунарски модел реалног света доступан нашим чулума

МОДЕЛИРАЊЕ И МОДЕЛИ ПОДАТАКА

Моделирање

- Као резултат процеса који називамо интелигентно људско понашање, моделирање представља свакодневну активност и велики део онога што нас чини људским (интелигентним) бићима.

- Резултат моделирања је модел који представља апстракцију реалности, јер не може да обухвати све њене аспекте.

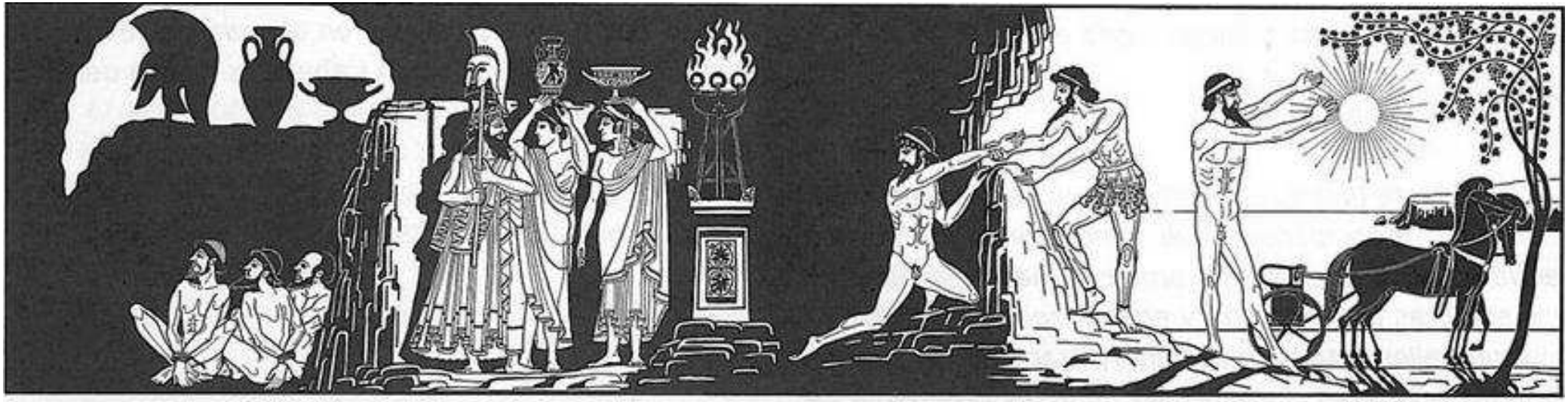
Модел

- Модел је упрошћена и идеализована слика реалности.
- Модели су увек апстракције реалног света, због тога задржавају само оне карактеристике које су битне за сврху његовог изучавања.
- Ниво апстракције у процесу моделирања утиче на валидност модела, односно на успешност представљања реалног система моделом.
- Имплементације модела се мењају у зависности од развоја технологије.

Апстракције и експериментални услови

- Модел је апстракција за одређене експерименталне услове.

- Анегдота о Платоновој пећини
  Шта се види из пећине, а шта је стварно напољу?



Апстракције и информациони токови у виртуелној реланости

- Информациони модел реалног или виртуелног света је апстрактна ментална слика креатора. Чува се на медијумима. Медијум има ограничења, па долази до аномалија:
  - Губљење информација
  - Дисторзија информација

- Прималац преузима записану информацију са медијума, при чему поново долази до губитка информација, због:
  - Ограничења медијума
  - Особина примаоца и његових менталних апстракција

- Пренос информација:
  - Кроз време - меморисање
  - Кроз простор – телекомуникације

Апстракције и информациони токови у виртуелној реалности

Апстракције података

- У моделима података користе се следеће врсте апстракција:
  - Асоцијација - апстракција која пресликавања врши од вишег ка нижем нивоу у хијерархији спецификације система.
  - Агрегација - агрегирање више типова у један апстрактнији сложенији тип.
  - Генерализација – типови података се апстрахују у генерички апстрактнији тип уочавањем заједничких особина и операција које се наслеђују од надтипа.
  - Специјализација – додавање детаља у циљу креирања валидног модела
  - Редукција – избацивање непотребних података

Претраживање и везе између података

- Асоцијација се користи за уређивање података у форму погодну за претраживање, приказ и упоређивање.
- Класификација је методолошки поступак груписања података како би се над њима могла једноставно и директно применити апстракција асоцијације.
- Резултати претраживања података и њихово упоређивање се користе у процесу доношења одлука.
- Процес доношења одлука се у великој мери заснива на математичкој теорији скупова и бинарним релацијама.

Скупови као математички апарат за моделирање података

- Савремена теорија скупова настаје крајем 19. века када Кантор даје описну математичку теорију која се још назива и интуитивна или наивна теорија скупова.

- Представља покушај да се обједини до тада позната математика.

- Скуп је апстрактан појам који обједињује извесне елементе у јединствену целину.

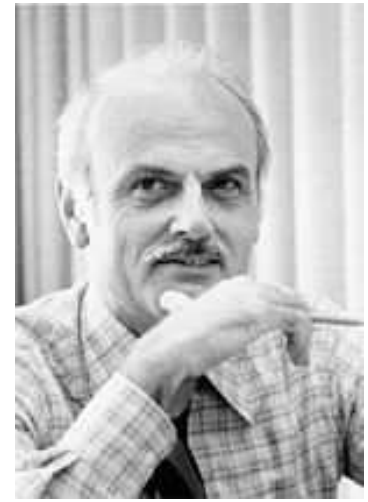
- Декартов производ је основна операција над два или више скупова. Издвајањем подскупова декартовог производа по неком критеријуму се добијају релације.

- Релације су најчешће коришћен концепт за моделирање података у информационим технологијама .

Релациони модел

- Информациони систем је модел реалног света, тј. апстрактна слика стварности, па су за моделовање погодни скупови и релације као апстрактни математички концепти.

- Релациони модел базе података уводи Едгар Ф. Код 1969.



- E.F.Codd (1969) Derivability, redundancy and consistency of relations stored in large data sets, http://www.liberidu.com/blog/images/rj599.pdf

- E.F.Codd (1970) A relational model of Data for Large Shared Data Banks, http://www.seas.upenn.edu/~zives/03f/cis550/codd.pdf

- Релациони модел је дуго година најшире коришћени модел у реализацији информационих система предузећа.

Организација података у релационом моделу

- Релација је скуп n-торки, може се представити табелом.
  РЕЛАЦИЈА = ТАБЕЛА

- Релациони модел базе података је апстрактан. Табела има само редове и колоне.

- Пример релације:

23	Пера	2000	BMW 730
24	Жика	3000	Audi S8
25	Лаза	2500	Mercedes CLS 63 AMG

Семантика података у релационом моделу

- Семантика података се у релационом моделу своди на увођење класификације за врсте колона :
  - Кандидати за кључеве
  - Кључеви – по кључу се сортирају редови табеле
  - Описни атрибути
  - Спољни кључеви

Шифра	Име	Плата	Аутомобил
23	Пера	2000	BMW 730
24	Жика	3000	Audi S8
25	Лаза	2500	Mercedes CLS 63 AMG

- Ограничења над атрибутима су део релационог модела.

- За претраживање се користи релационе алгебра, преко упита у SQL језику (Structured Query Language).

Трансакције у релационом моделу

- ACID (Atomicity, Consistency, Isolation, Durability)
  Скуп послова у при извршењу важног пословног задатка, који чини једну логичку операцију, назива се трансакцијом. Трансакција над базом података мора да буде атомична, конзистентна , изолована и трајна.

- Трансакције и скуп технологија за њихово праћење је по први пут дефинисао Jim Gray 1970.

- SQL је инхерентно трансакцион. Нова трансакција почиње када се претходна оконча. Трансакција се може и експлицитно дефинисати.

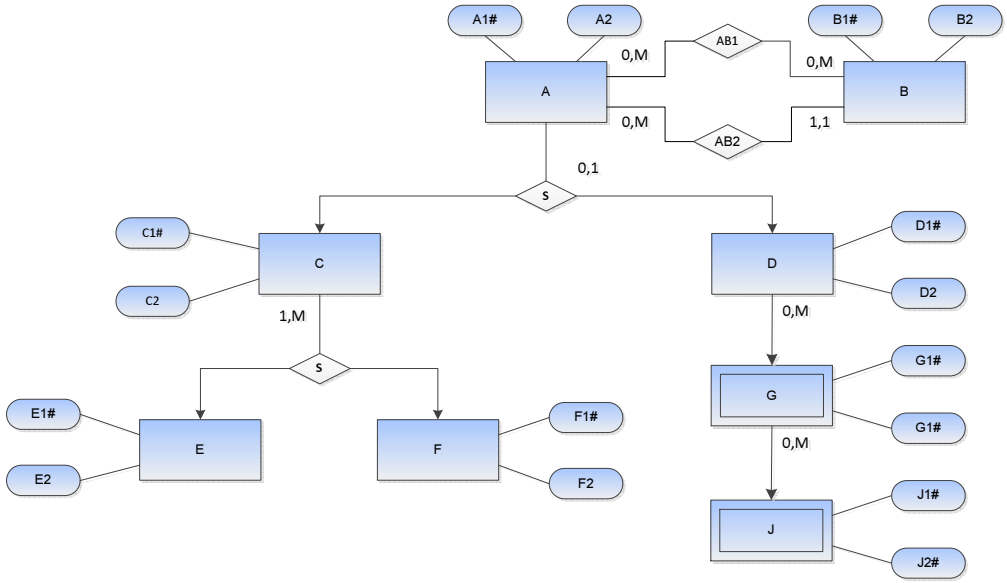
- Подршка трансакцијама је повећала поузданост обраде у релационом моделу, али је доста успорила брзину обраде.

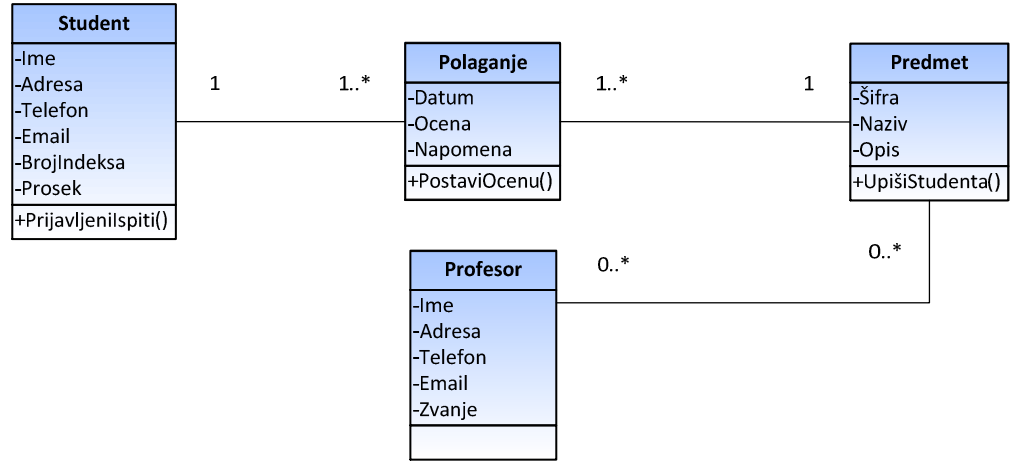
- ACID трансакције се тешко реализују у дистрибуираним базама.

Детаљно моделирање семантике података релационог модела

- Семантика података релационог модела се може имплицитно специфицирати кроз структуру релационог модела. За ово се користе технике као што су ПМОВ и UML.

- У ПМОВ-у или UML-у семантика података се држи одвојено од табела релационог модела, на посебним дијаграмима и табелама.





Моделирање семантике припада пројектовању информационих система

Предности и недостаци релационог модела

- Предности релационог модела:
  - Висок ниво апстракције, широка примена.
  - Коришћење релационе алгебре у извештавању.
  - Погодан за класификацију података и комплексне извештаје.
  - Подржава ACID трансакције.

- Недостаци релационог модела:
  - Подаци се чувају у табелама, али не и њихова семантика.
  - Концепт релације није дистрибуиран нити је погодан за паралелизацију и скалирање перформанси.

Хијерархијске базе података - XML

- У трослојној клијент-сервер архитектури релациони модел је погодан за чување података на слоју података. Слој пословне логике те податке обрађује. Излаз из пословне логике је углавном хијерархијски структуриран.

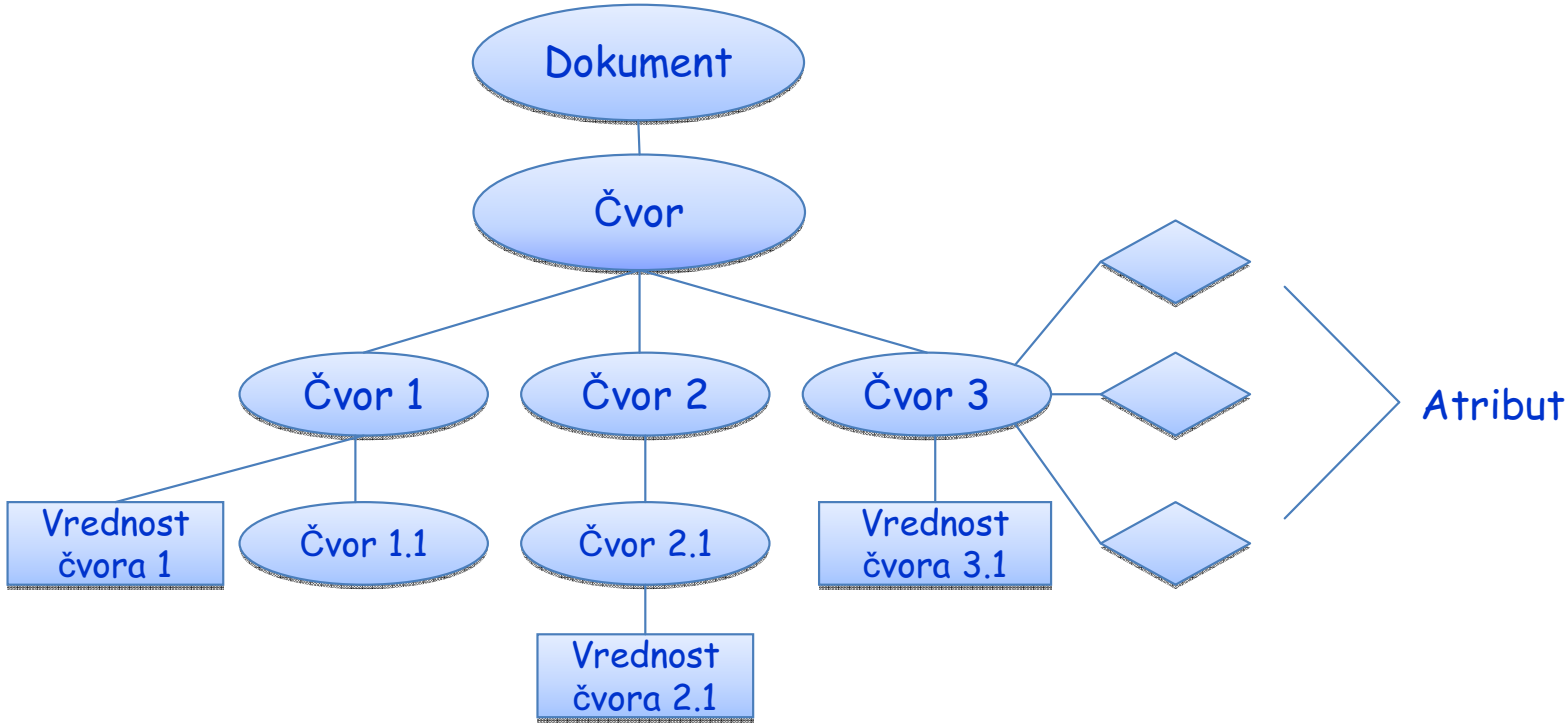
- XML је стандард који подржава меморисање и обраду података у форми стабла, заједно са њиховом семантиком.

- Технологије које подржавају XML су:
  - за пренос података (XML, XML shema),
  - меморисање (DOM,SAX),
  - претраживање (XQUERY) и
  - трансформације стабла (XSLT).

XML

- XML и скуп технологија које га прате се користе највише у трослојној клијент-сервер архитектури за пренос и обраду података из слоја пословне логике у слој презентације.

- У новије време се све више самостално користи као технологија за опис, смештање и пренос објеката са хијерархијском структуром.



XML – Примена у будућности

- Релационе базе података су инкорпоририрале XML, тако да је XML постао њихов нераздвојни део:
  Oracle, DB2, MS-SQLj, MySQL, PostgreSQL...

- XML је основна форма за меморисање и манипулисање објектима у семантичком вебу, тако да ће у доброј мери његова примена у будућности зависити од степена прихватања семантичког веба.

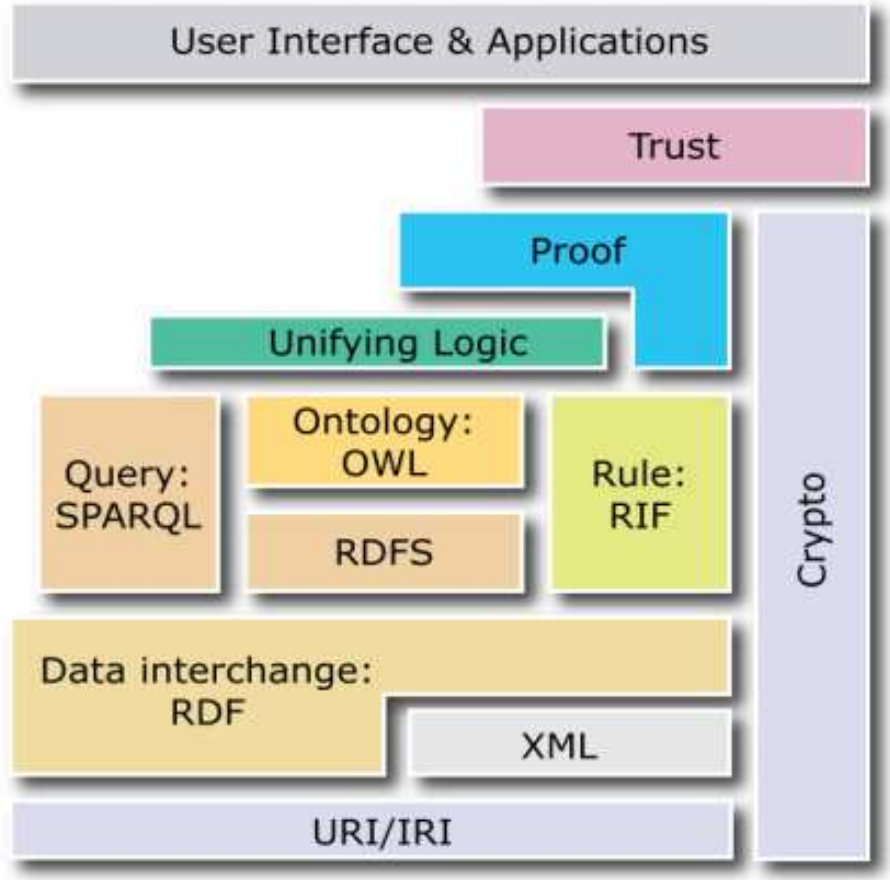
- Већ сада се појављују алтернативе које су у неким областима примене флексибилније и једноставније за коришћење
  - нпр: JSON – JavaScript Object Notation

СЕМАНТИЧКИ ВЕБ

Семантички веб

- Семантички веб је мрежа повезаних докумената, који описују релације између објеката реалног или виртуелног света, путем података и придружених семантичких информација које се користе за аутоматизовано претраживање и извештавање.

- Стандарде за семантички веб прописује W3C конзорцијум.



Експлицитно моделирање семантике података

- Значење података на семантичком вебу се унапред моделира дефинисаним класификацијама тагова метаподатака – онтологијама.

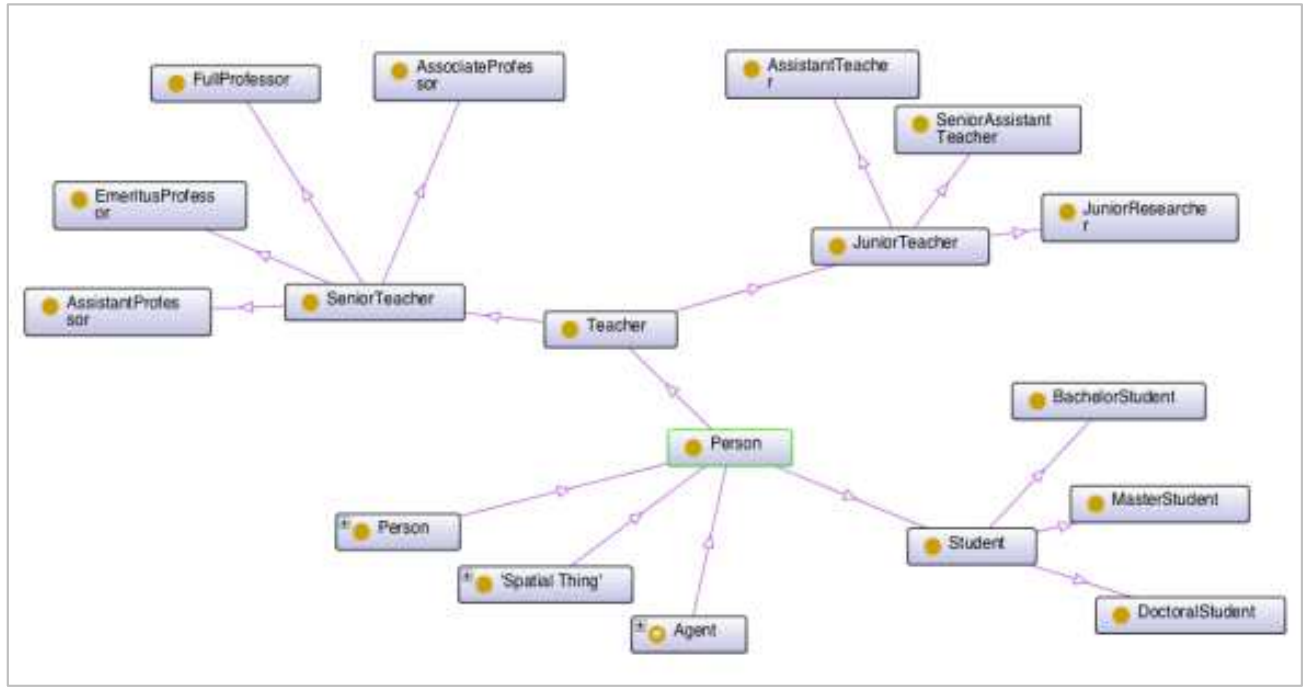
- Не постоји свеобухвана онтологија, већ се бирају локалне онтологије које су валидне за одређене експерименталне услове.

- За експлицитно моделирање семантике података користе се језици као што су:
  - OWL,
  - RDF,
  - XBRL и др.

Онтологије и семантички веб

- У контексту рачунарских наука, онтологије:
  - Представљају скупове концепата који описују одређене области.
  - Описују концепте и везе између тих концепата.
  - Ограничене су јер се дефинишу статички, па им придружујемо метаподатке.

- Треба да су једноставне, свеобухватне и динамичке. Ово су све противречни захтеви.



Ограничења у примени онтологија и семантичког веб-а

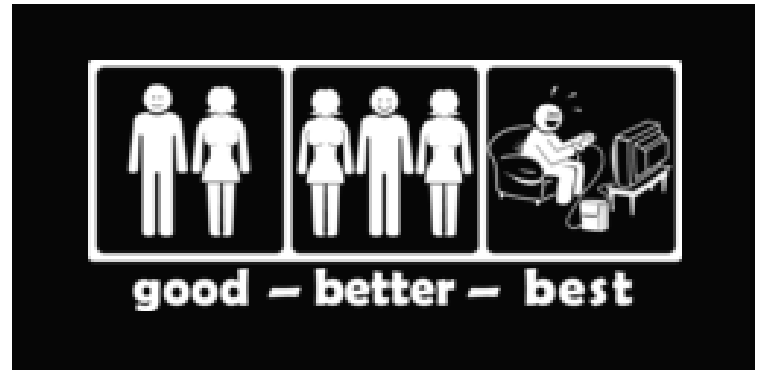
- Недостаци :
  - Репрезентација објеката је ограничена на XML и његову хијерархијску структуру.
  - Моделирање онтологија се врши ван семантичког веба . Ово треба да буде обједињено. Онтологије за описивање веба будућности треба да се динамички креирају, мењају и машински уче међусобна пресликавања.
  - У односу на стандардан веб, захтевају много више рачунарских и људских ресурса за имплементацију.
  - Повећана претња приватности, због могућности злоупотреба семантичких података.

ПРЕТРАЖИВАЊЕ И КОРИШЋЕЊЕ ИНФОРМАЦИЈА

Претраживање података и њихова употеба у пословном одлучивању

- Многи проблеми у инжењерској пракси и у свакодневном животу се своде на проналажење и упоређивање података задатих у енумеративној форми или форми реалних бројева.

- Сортирање таквих података је операција која омогућује брзо претраживање и упоређивање података, поготово када их има много.



Бинарне релације

- Као математички концепт настале су крајем 19. века.
- Порекло појма датира од де Моргана, 1860. године "On the Syllogism, No. IV. And on the Logic of Relations"
- Бинарна релација је подскуп Декартовог производа два скупа.

$$\rho \subseteq AxB$$

- Бинарна релација се у информационим системима представља табелом са две колоне.

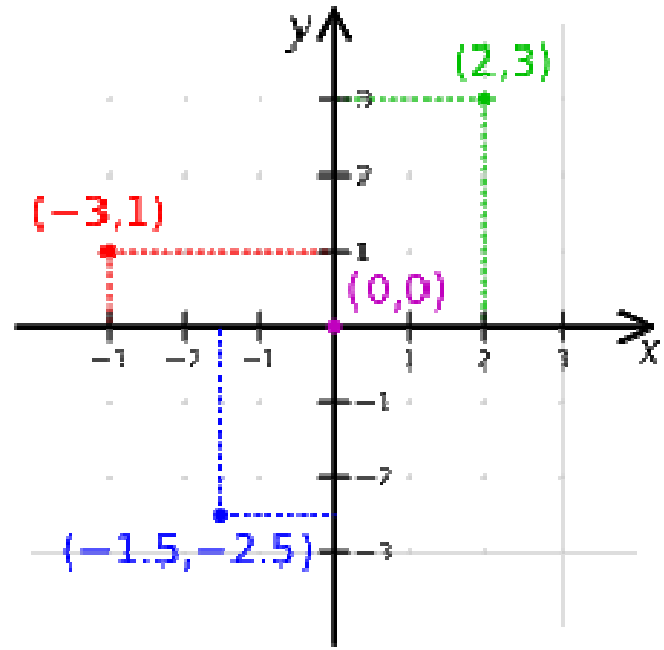
Декартов координатни систем

- Рене Декарт (1596-1650)

- Свака тачка у координатном систему је одређена уређеним паром координата. Било који скуп тачака у кординатном систему је бинарна релација .

- Декартов координатни систем је погодан за упоређивање појава из реалног света јер се бинарне релације реалних бројева увек могу сортирати и упоређивати.

- Осим бинарних релација, рационални људски ум не поседује други механизам за упоређивање појава у реалном свету.



Коцкање као начин одбране од диктатуре упрошћено уређених бинарних релација ☺

Сортирање података и бинарне релације

- Релација која податке уређује (сортира) у поредак погодан за упоређивање је ≤ (мање или једнако).
- Релација мање или једнако је релација поретка за скупове података који узимају вредности из скупа реалних бројева. [деф.1.3.15 http://www.mi.sanu.ac.rs/~gvm/Teze/LinearAlgebra.pdf]
- Релација ≤ није релација поретка над тринарним и н-арним релацијама.
- Да би се појаве које се описују реалним бројевима упоређивале и сортирале, потребно је да се н-арне релације неком апстракцијом (трансформацијом) сведу на бинаране, уз минимални губитак информација.
- Ово је најчешће коришћен метод у друштвеним и природним наукама.

Пословно одлучивање : свођење н-арне на бинарну релацију

Проблем Топаловића: Човек се апстрахује пројекцијом од три на само једну димензију - дужину. Резултат мерења не задовољава критеријуме пословања, (Покојник, 228цм).

Претраживање релационих база података

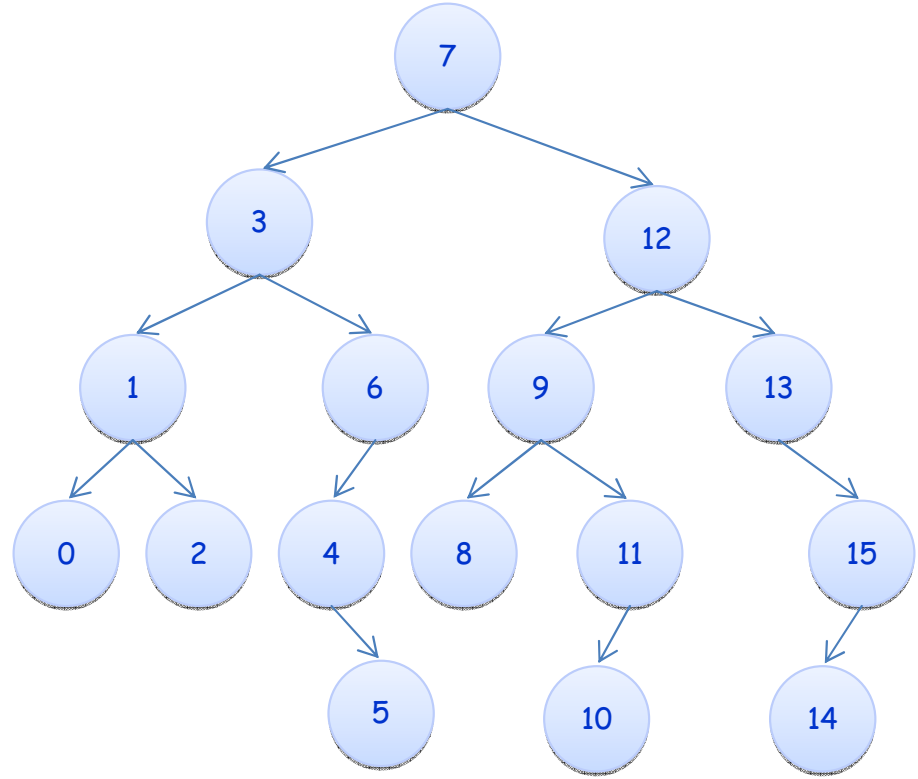
- У релационом моделу база података је скуп табела које су уређене по кључевима.

- За претраживање, тј. налажење података, табеле се сортирају по кључу, а затим се сортирани кључеви неком методом брзо претражују (нпр: метода половљења...).

- За брже налажење података на основу кључа користе се и хеш табеле које се добијају трансформацијом вредности кључева неким алгоритмом (нпр: MD5) у адресе на које су подаци смештени на меморијском медијуму. На тај начин се може директно приступити подацима на основу кључа, без претраживања.

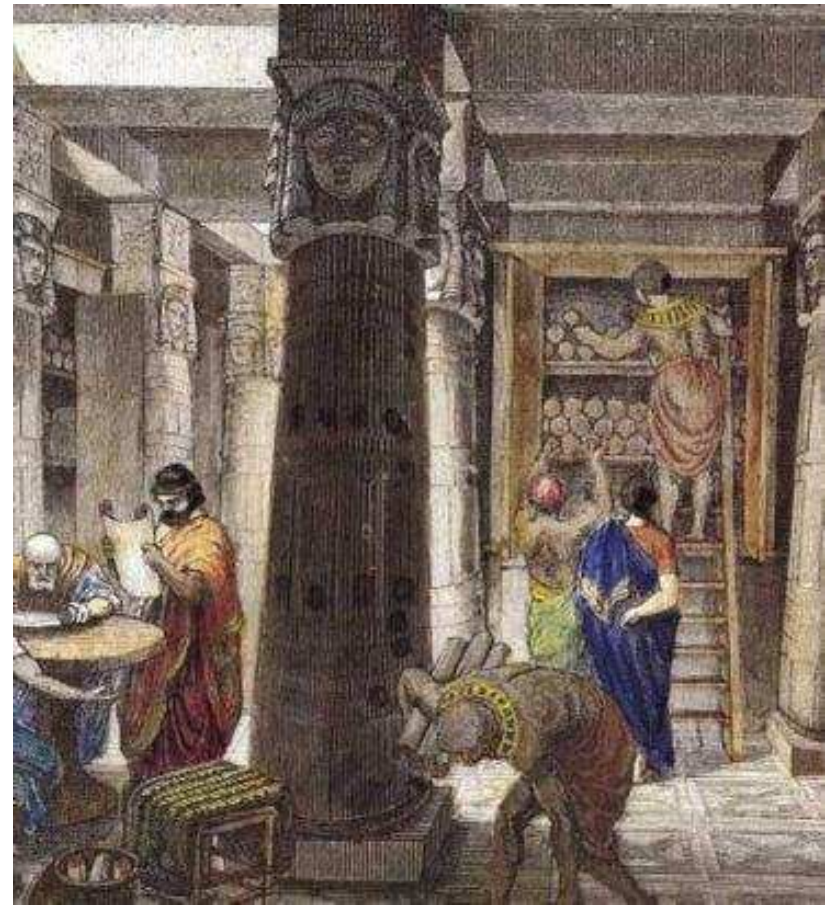
Технологије за брзо претраживање и индексирање (конструисање бинарних релација)

- За брзо претраживање релационих база података по описним атрибутима, користе се индекси.
- Индекси су додатне сортиране бинарне релације (колона табеле, кључ релације)
- Индекси се због брзине претраживања имплементирају коришћењем :
  - Бинарног стабла
  - Б-стабла
  - Балансираног б-стабла
- Технологије имплементације индекса омогућавају да се пролазом кроз стабло добију сортиране бинарне релације.



Бинарне релације и индекси у античком времену

- Пример примене: Александријска библиотека
- Александријска библиотека је складиштила књиге из књижевности и филозофије целог света.
- У Александријској библиотеци чувало се 700.000 свитака уређено по насловима и ауторима.
- Први каталог Александријске библиотеке саставио је Калимах. Назван је Пинакес.
- Уређен је по колонама и користио се као модел за библиотечке каталоге до 19. века.



Анегдота: бинарне релације и савремени научно-истраживачки рад

- Методолошки поступак који се често примењује у научној анализи :

1. Избор предмета истраживања - откривање топле воде:
   Многе природне појаве су по својој природи већ описане у форми бинарних релација. Такве појаве се често узимају за предмет научне анализе у научним радовима и дисертацијама.

2. Увођење оператора компликатора $Д : P \rightarrow N$
   који познато своди на непознато:
   Представљање појава из тачке 1. ради детаљније анализе са више аспеката, у форми n-арних релација, које не могу више да се упоређују или сортирају.

3. Тражење инверзног оператора – откривање рупе на саксији :
   Тражи се инверзни оператор који обично важи само за партикуларне случајеве, који n-арну релацију из тачке 2. своди на бинарну релацију из тачке 1., да би се појава могла анализирати и модалитети појаве упоређивати.

4. Објављивање радова у SCI часописима и одбрана дисертације.

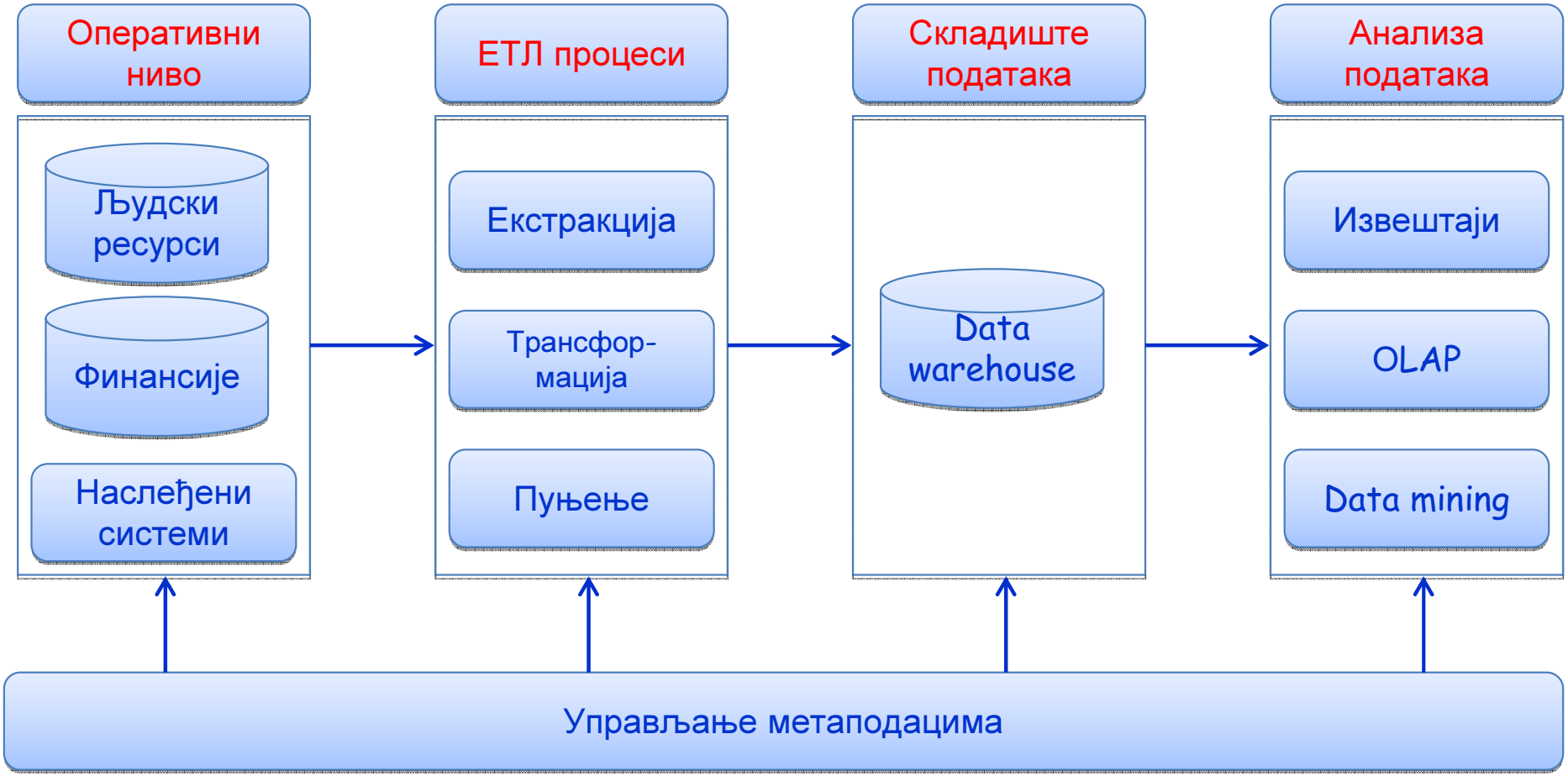
ОТКРИВАЊЕ ЗНАЊА У ПОДАЦИМА

Некласификовано знање у подацима

- У релационом моделу подаци су смештени у табелама.
- Релације и спољни кључеви су погодни за класификацију података, али у базама података постоје и информације ван ове класификације, нарочито у динамици података.
- За потребе откривања знања ван класификације унапред задате у релационом моделу подаци се преносе у складишта (стоваришта) података.
- Складиште података или Data Warehouse је репозиторујум података над којим корисници могу спроводити ad-hoc анализе података и креирати извештаје.
  Извор: Building the Data Warehouse, Inmon B, Wiley and sons, 1992
- Основни циљ складиштења података је прикупљање, дистрибуција и доступност информација кроз простор и време ради њихове касније анализе, извештавања и откривања новог знања.

Трансформација из OLTP у OLAP

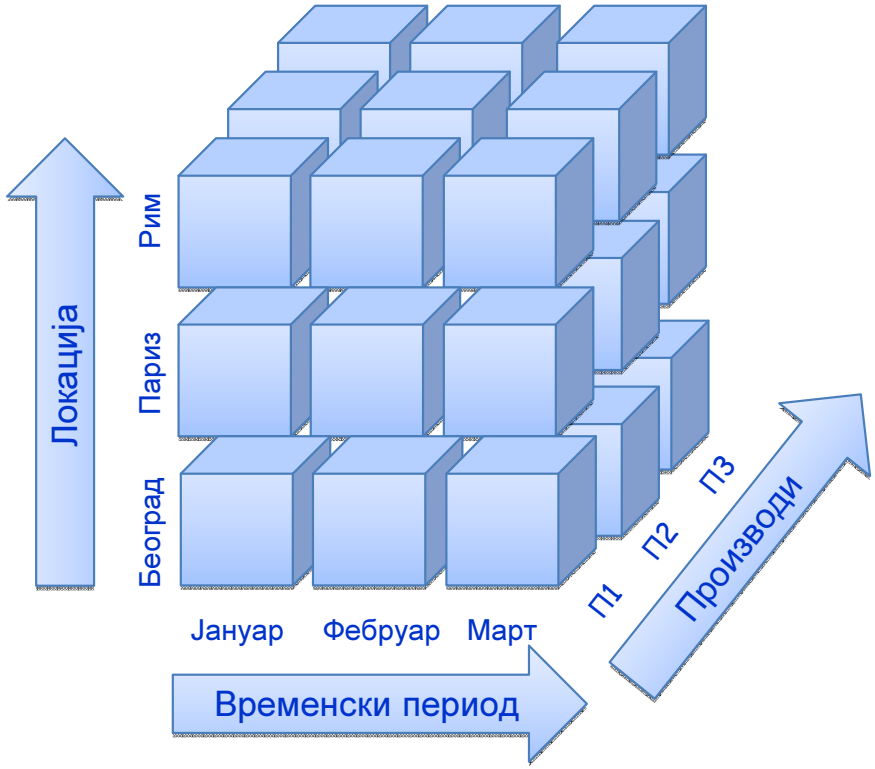
- Подаци се у складиште података смештају екстракцијом, трансформацијом и учитавањем података из трансакционих релационих база података.

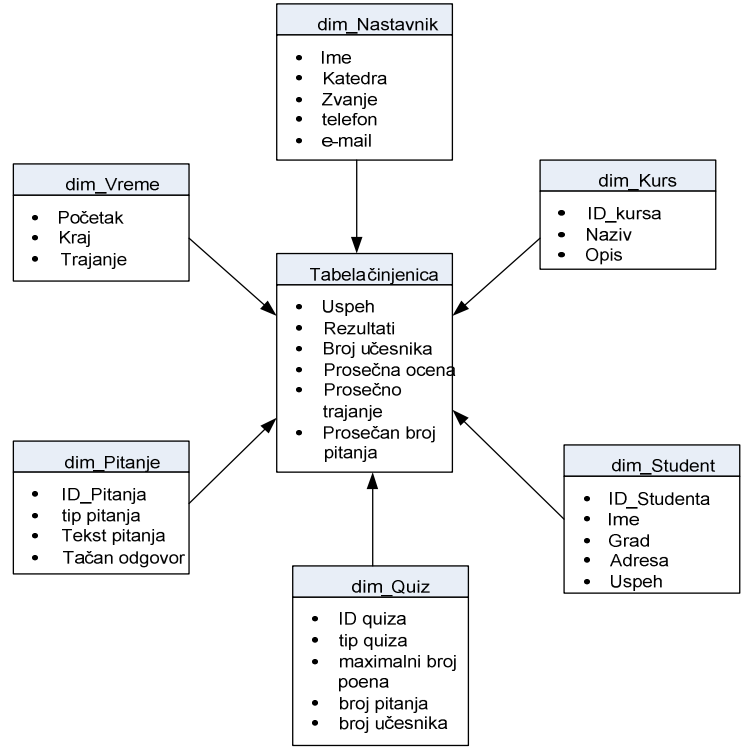


OLAP и пословна интелигенција

- OLAP хипер коцка је начин организовања података из складишта података у мултидимензионе табеле.

- Пословна интелигенција је процес прикупљања расположивих интерних и релевантних екстерних података, и њихова обрада у корисне информације.





OLAP у пословном одлучивању

- Елементи хипер коцке су нумеричке вредности повезане са појавом која се анализира. Називају се "мере" (measures). Распоређене су по координатама димензија хипер коцке.

- За потребе одлучивања, "мере" појаве која се анализира, одабраним OLAP трансформацијама прилагођеним проблему одлучивања, се трансформишу у једну колону бинарне релације. Друга колона бинарне релације је изведена из димензија хиперкоцке.

- Тако добијена бинарна релација се сортира, евентуално се графички приказује и на основу критеријума одлучивања се доносе пословне одлуке.

Технике анализе података

- Стабло одлучивања – за разврставање, предвиђање, процену вредности, груписање, описивање података и визуелизацију
  Пример примене: Системи препоруке у електронским продавницама
  Извор: Data mining with decision trees and decision rules, Apte C, Weiss S, Future generation computer systems, 1997

- Класификација - сврставање ентитета у једну од неколико претходно дефинисаних група.
  Пример примене: анализа Интернет саобраћаја
  Извор: Toward integrating feature selection algorithms for classification and clustering, Liu H, Yu L, IEEE Transactions on knowledge and data engineering, 2005

- Кластеровање – груписање објеката сличних карактеристика у групе које се међусобно разликују.
  Пример примене: анализа типова понашања корисника веб сајта.
  Извор: Survey of clustering algorithms, Xu R., Wunsch D., IEEE Transactions on neural networks, 2005

Технике анализе података

- Правила повезивања – за проналажење корелације између ставки у огромним сетовима података.
  Пример примене: анализа потрошачке корпе, тј. производа који се често купују заједно.
  Извор: Mining generalized association rules, Srikant R, Agrawal R, Future generation computer systems, 1997

- Неуронске мреже – за моделовање комплексних односа између улазних и излазних података и проналажења шаблона у подацима.
  Пример примене: финансијско предвиђање
  Извор: Effective data mining using neural networks, Lu HJ, Setiono R, Liu H, IEEE Transactions on knowledge and data engineering, 1996

- Генетски алгоритми – технике засноване на еволуцији живих организама (наслеђивање, селекција, и сл.) примењују се за решавање различитих оптимизационих проблема.
  Пример примене: анализа природног језика
  Извор: Genetic algorithms: Concepts and applications, Man KF, Tang KS, Kwong S, IEEE Transactions on industrial electronics, 1996

Правци будућег развоја пословне интелигенције

- Интеграција са Big data базама података
- Интензивно коришћење MapReduce алгоритма у процесу откривања знања
- Откривање некласификованог знања у реалном времену.
- Генерисање информација коришћењем OLAP метода представља искорак ка опонашању начина рада људског мозга. У том правцу ће се пословна интелигенција и развијати.

CLOUD COMPUTING
BAZE PODATAKA

Big data

- Због великих количина података (Big data) које се данас генеришу у информационим системима веб 2.0 предузећа, јављају се тешкоће у коришћењу релационих база података.

- Димензије big data:

  - Разноврсност – подаци нису структурирани и у различитим су форматима: текст, аудио, видео, clickstream, лог фајлови, подаци са сензора...

  - Брзина – подаци су често временски осетљиви, потребно је брзо их доставити и анализирати.

  - Обим – ред величине је терабајти и већи.

Релациони модел и Big data

- Захтеви пројектовања и имплементације Web 2.0 апликација:
  - Поузданост - Способност ИС да обавља одређене функције, под одређеним условима, у току времена.
  - Дистрибуираност - Могућност да се различити делови ИС складиште на различитим местима.
  - Скалабилност - Способност ИС да лако подносе повећану количину обраде или да је лако могућа надоградња уколико дође до повећане количине обраде података.
- Да би наведени захтеви били испуњени неопходно је упростити - апстраховати релациони модел база података.

Big Data
NoSQL и нерелационе базе података

- Апстракција над хеш табелама и над релационим моделом.
- Погодне су за глобалну примену (једна база података за читав свет).
- Обезбеђују поузданост, дистрибуираност и скалабилност.
- Омогућују брзо индексирање и претраживање.
- Добра интеграција и API са cloud платформама као сервис.
- Повољност за cloud провајдере јер једну табелу продају свим својим корисницима.
- Нема добре класификације као у релационом моделу. Критеријуми класификације нису јасно дефинисани.
- Недостатак за пројектанте ИС и програмере – пресликавање из ПМОВ-а у нерелациону базу није једнозначно. Потребан је механизам за семантичко пресликавање модела података у нерелациону базу.

Map Reduce - паралелно и дистрибуирано претраживање и извештавање

- Google је 2010. патентирао овај алгоритам који претражује податке уређене по паровима: кључ, податак (k,v) [US Patent 7,650,331]

1. Map(k1,v1) → list(k2,v2)
   је паралелно и дистрибуирано процесирање сваког уређеног пара из првог домена: Map (k1,v1) и одабирање само оних који се пресликавају у парове из другог домена: list(k2,v2)

2. Reduce(k2, list (v2)) → list(v3)
   Паралелна и дистрибуирана обрада парова: (k2, list (v2)) која на основу критеријума за извештавање ове парове пресликава у резултат претраживања: list(v3)

- Представљен је по први пут у раду:
  Jeffrey Dean and Sanjay Ghemawat,
  MapReduce: Simplified Data Processing on Large Clusters, Google inc 2004.

- Овај алгоритам се користи као основни механизам за претраживање и извештавање у већини Big data база података.
  Hadoop – open source имплементација MapReduce.

Not Only SQL

- За реализацију дистрибуираности – потребно је мапирање између података и рачунара на којима се подаци чувају и обрађују. Ово се постиже хијерархијским хеширањем кључева .

- За обезбеђење скалабилности – табеле из релационих база података се апстрахују једном глобалном табелом. Овој табели приступа се коришћењем кључева или хеш табела.

- Релациона алгебра се у великој мери напушта као механизам за генерисање извештаја. Користи се MapReduce алгоритам.

"Brewer's CAP Theorem"
Consistency, *Availability* and *Partition Tolerance*

- Basically Available, Soft state, Eventual consistency
  Потпуна супротност од ACID у релационим базама.

Not Only SQL
(кључ, податак) модел

- Кључ-податак модел (key,value) – Amazon Dynamo
  - Табела са две колоне - бинарна релација - која може бити сортирана или несортирана.
  - Примењује се хијерархијски хешинг. Хеш кључ се састоји од хијерерхијски уређене путање до податка.
  - Корисници имају погледе (view) над глобалном табелом.
  - Провајдер одржава само једну инфраструктуру.
  - Скалабилност се постиже увођењем нових рачунара.
  - Проблем ажурирања се решава увођењем time-stamp-а за праћење историје измене податка. Не гарантује се увек конзистентност података.

Више детаља се може наћи у:
Dynamo: Amazon's highly available key-value store,
Giuseppe DeCandia et. al., SOSP 2007.
MySQL од верзије 5.6 подржава NoSQL memcached API

Not Only SQL
Big Table

- BigTable модел - Google
  - Релациони модел података се апстрахује једном великом глобалном табелом
  - Сваки елемент табеле има идентификатор реда, идентификатор колоне и time-stamp
  - Индексирање може да буде глобално, и може се вршити и над редовима и над колонама
  - Додатне апстракције: делови табела – таблети
  - Модел је једноставан и омогућава дистрибуираност.

Bigtable: A Distributed Storage System for Structured Data, Fay Chang et. al., OSDI 2006.

Cassandra – open source Big Table

Нерелационе базе података засноване на документима

- Документ модел
  - Дистрибуиране, хијерархијске и мрежне базе података
  - Користе се стабла и графови
  - Семантика је у чворовима
  - Проблем је представљање докумената и сематике једноставним концептима (XML,JSON)

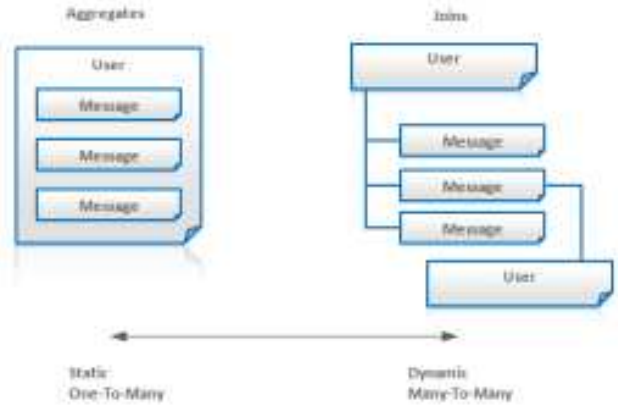
  - MongoDB open source имплементација.

Концепти моделовања података у Not only SQL базама података

- Денормализација - копирање истих података у више докумената или табела у циљу поједностављења/оптимизације упита или корисничких података у одређени модел података.

- Агрегација ентитета:
  - Формирају се флексибилне класе ентитета које омогућавају угњеждавање ентитета са комплексном и различитом интерном структуром.

- Моделовање M:M релација – спајања се врше у време моделирања, а класични "join"-и се не користе.



Релационе и NoSQL базе података - упоредне карактеристике -

- NoSQL базе података решавају проблеме дистрибуираности, поузданости и скалабилности. Побољшавају се перформансе.

- Жртвује се конзистентност података и нормализација. Јављају се проблеми у ажурирању.

SQL базе	NoSQL базе
Висок ниво апстракције	Нижи ниво апстракције
Одвојеност података од семантике	Већа повезаност података и семантике
Споро претраживање	Брже претраживање
Непогодне за дистрибуирана окружења	Погодне за дистрибуирана окружења
Нема редудансе, подаци су ажурни	Проблем ажурности због редудантности
Нормализовани подаци	Денормализовани подаци

Врсте и примери NoSQL база података

Граф базе	BigTable модел	Документ модел	Перзистентни Кључ-податак модел	Меморијски Кључ-податак модел
neo4j	BigTable (Google)	MongoDB (BigTable)	Dynamo (Amazon)	memcached
FlockDB (Twitter)	HBase (BigTable)	CouchDB	Voldemort (Dynamo)	Hazelcast
InfiniteGraph	Cassandra (Dynamo + BigTable)	Riak (Dynamo)	Redis	
	Hypertable (BigTable)		Membase (memcached)	
	SimpleDB (AmazonAWS)		TokyoCabinet	

NoSQL
Правци будућег развоја

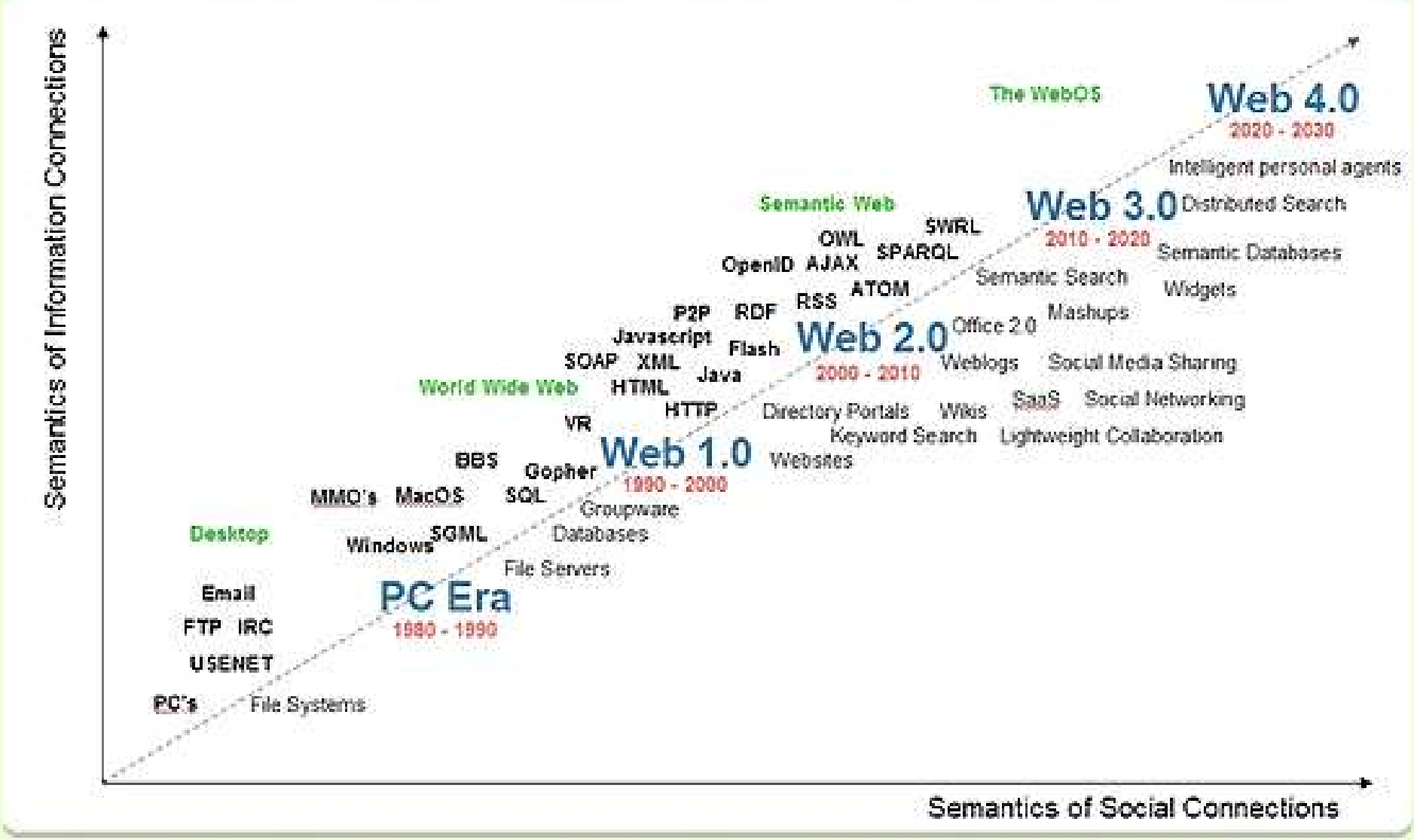
- Кључ-податак модели
  - Већ су имплементирани механизми за брзо претраживање. Очекује се развој функционалности као што су управљање транскацијама, упитни језици, интеграција са методама за проналажење знања, и др.

- BigTable модели
  - Модели засновани на BigTable развијаће се у правцу правог и потпуног глобалног дистрибуираног и паралелизованог релационог модела који подржава ACID трансакције.

- Документ модели
  - Базе података засноване на документима (објектима) омогућиће рачунарима аутономно моделовање семантике објеката на вебу. Резултат би могао да буде семантички веб у коме машине могу самостално да обрађују и стварају информације.

ЗАКЉУЧАК

Будући развој информационих технологија



Source: Radar Networks & Nova Spivack, 2007 – www.radarnetworks.com

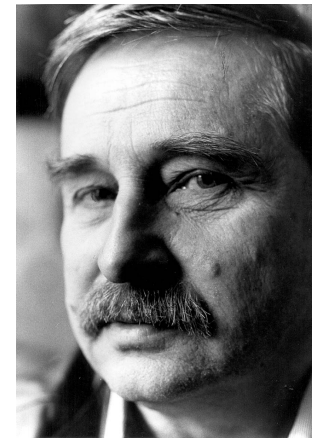
Закључак

- Савремени информациони системи треба да омогуће поуздано, дистрибуирано и скалабилно окружење за генерисање, пренос, чување и претраживање знања, без губитка семантике.

- Тежи се откривању онога што је трансцендентно – непојмиво.

- Будући развој информационих технологија треба да омогући везу између реалног и виртуелног света и обраду информација на начин ближи људима:

"Људски снови и мисли нису линерани, они се роје и бокоре на све стране, имају симултаност која захвата у живот и из живота знатно више и шире него било какав језички исказ.
То сви знамо из искуства."



Милорад Павић

Факултет организационих наука
Универзитета у Београду
Лабораторија за електронско пословање

Информационе технологије
- од усменог предања до семантичког веба -

Проф. др Божидар Раденковић

boza@elab.rs