

Gagnasafnsfræði

Páll Melsted

26. ágúst

Yfirlit

- Inngangur
- Af hverju gagnagrunnar
- Praktísk atriði
- Kostir og gallar venslagagnagrunna
- sqlite

Yfirlit

- Hefðbundin notkun
 - Geymsla talna, texta
- Margmiðlunargagnagrunnar
 - Myndir, hljóð, vídó
- Landfræðigrunnar - GIS
 - Geyma og greina landakort, gervitunglamyndir og upplýsingar tengt staðsetningu

Yfirlit

- Vöruhús gagna og greiningakerfi (OLAP)
 - Sækja og greina gögn s.s. viðskipta upplýsingar úr stórum gagnagrunnum
 - Styðja rauntíma ákvarðanatöku
 - "Venjulegir" notendur
- Rauntímagagnagrunnar
 - Stýra iðnaðar- og framleiðsluferlum.

Inngangur

- Gagnagrunnur
 - Safn skyldra gagna
 - Þekktar staðreyndir sem hafa merkingu og unnt er að skrá
 - Gögn sem hægt er að túlka í samhengi
 - Byggður í tilteknum tilgangi
- Dæmi um stóran gagnagrunn með viðskiptalegan tilgang
 - Amazon.com
 - midi.is
 - Valitor

Inngangur

- Gagnagrunnskerfi (DBMS)
 - Safn forrita
 - Gerir notendum kleift að smíða gagnagrunn og viðhalda honum
- Skilgreining gagnagrunns
 - Tilgreina gagnatög, gagnamót og skorður á gögn
- Gögn um gögn (Meta-data)
 - Gagnagrunnskilgreining
 - Geymd í gagnagrunninum sem efnisyfirlit

Inngangur

- Aðgerðir á gagnagrunn
 - Fyrirspurnir og uppfærslur á gögnum sem skilgreina sýndarheiminn
 - Framleiðsla á skýrslum
- Samnýting gagnagrunns
 - Leyfir mörgum notendum samtímis aðgang að gagnagrunninum
- Notendaforrit
 - Senda fyrirspurnir og uppfærslur á gagnagrunninn gegnum gagnagrunnskerfið (DBMS)

Inngangur

- Fyrirspurn
 - Sækir hluta gagnanna úr gagnagrunninum
- Færsla (Transaction)
 - Veldur lestri og skrift í gagnagrunninn
- Meðal öryggisþátta eru:
 - Kerfisvernd (System protection)
 - Gagnavernd (Security protection)
- Viðhald gagnagrunns
 - Þróa þarf kerfið eftir því sem kröfur breytast

Af hverju þurfum við að læra um gagnagrunna?

- Fræðilegt svar
 - Blandar saman áhugaverðum reikniritum, rökfræði og tölvunotkun
- Svar fyrir forritara
 - Við þurfum að kreista upplýsingar úr gagnagrunnum í vinnunni
- Svar fyrir gagnanörda
 - Við þurfum að vinna með mikið magn gagna sem breytist sífelld

Markmið í námskeiðinu

Að öðlast sterkan bakgrunn í gagnagrunnum. Eftir námskeiðið eigið þið að geta útfært miðlungsstóran venslagagnarunn, þ.m.t.

- Hannað í E/R líkaninu og fært yfir í venslagangagrunn
- Forritað flóknar fyrirspurnir í SQL
- Notað fræðileg tæki, t.d. staðalform og venslaalgebru, til að bæta hönnun gagnagrunna
- Búið til gagnagrunn með skorðum t.d. vísunarheilleika
- Bætt hraða og afköst gagnagrunna með vísun

Praktísk atriði

- Heimasíða
- Vikublað 1
- Piazza
- Gradescope

Hvernig eigum við að geyma gögnin?

- Í textaskrá, TSV?
- Í excel skjölum?
- Í gagnagrindum, t.d. tré og tengdir listar?
- Á gataspjöldum?

Kostir gagnagrunna

- Leyfa flóknari fyrirspurnir en excel og textaskrár
- Gerðir fyrir samskiða vinnslu
- Lifa af náttúruhamfarir
- Allt geymt á disk

Hvernig eru gagnagrunnar útfærðir?

Vitlaus spurning!

Undir lok námskeiðsins munum við aðeins íhuga hvernig gagnagrunnar eru útfærðir. Að öðru leyti **viljum** við ekki þurfa að hugsa um það.

Við þurfum líkan af gögnum sem samsvarar venjulegri notkun en bindur okkur ekki við einhverja útfærslu.

"Any problem in computer science can be solved by another level of indirection" -David Wheeler (eða Butler Lampson)

Venslalíkanið

Langflestir gagnagrunnar nota venslalíkanið, við munum fylgja því eftir þar til í lok námskeiðsins. Öll gögn eru geymd í töflum þar sem hver dálkur hefur **nafn**. Töflurnar eru kallaðar **vensl**.

Movie: title	year	length
-----	-----	-----
Pretty Woman	1990	119
The Man Who Wasn't There	2001	116
Logan's run	1976	
Star Wars	1977	124
Empire Strikes Back	1980	111
Star Trek	1979	132

Venslalíkanið

Fyrst lagt fram af Codd árið 1969

(The relational model) provides a basis for a high level retrieval language which will yield maximal independence between programs on the one hand, and machine representation and organization of data on the other.

Venslalíkanið

- Öll gögn eru geymd í venstum (relation).
- Öll vensl hafa eigindi (attribute) með nöfnum
- Vensl er mengi af n-dum (tuples) fyrir hvert eigindi
- Hvert eigindi hefur óðal (domain) af löglegum gildum.

Venslalíkanið

- gagnagrunnur == safn af tölum
- vensl == tafla
- eigindi == dálkur í töflunni
- n-d == röð í töflunni (borið fram "ennd")

Venslalíkanið og SQL

SQL er mál sem leyfir okkur að skilgreina vensl og framkvæma fyrirspurnir

- Allir venslagagnagrunnar nota SQL
- Hægt að nota beint á grunn eða í gegnum forrit
- Myndar millilag óháð forritunarmáli eða gagnagrunni

Hönnun gagnagrunna

Ekki alltaf ljóst hvernig er best að hanna gagnagrunn.

Hönnun þýðir hvernig venlin eru skilgreind

Stór hluti námskeiðsins mun fara í hönnun gagnagrunna þegar við höfum náð tökum á SQL fyrirspurnum.

Hvernig er best að geyma eftirfarandi gögn

- Nemendaskráning í HÍ
- Pantanir hjá Dominos
- Venjulegur dagur hjá Hagkaupum

Veljið eitt af verkefnum og ræðið saman 2-4 í hóp (10 mínútur).

Hversu flókin eru gögnin? Hversu mikið? Hversu hratt breytast þau?

Fyrirspurnir í gagnagrunnum

Movie gagnagrunnurinn

title	year	length
-----	-----	-----
Pretty Woman	1990	119
The Man Who Wasn't There	2001	116
Logan's run	1976	
...		

SQL fyrirspurn til að ná í gögnin

```
SELECT * FROM Movie;
```

Meira SQL

```
SELECT title, year  
FROM Movie  
WHERE length > 120;
```

```
SELECT *  
FROM Movie  
WHERE year > 15*length;
```

SELECT þekking

Hve margir hafa séð

```
SELECT title, year  
FROM Movie  
WHERE length > 120;
```

eða svipað áður?

En hvað með

$$\pi_{\text{title,year}} (\sigma_{\text{length} > 120} (\text{Movie}))$$

Uppbygging SELECT

```
SELECT name1, name2, ...  
FROM relation1, relation2, ...  
WHERE <some condition>;
```

Við notum * sem styttingu á að fá allar n-dir

Tölum um í næstu viku hvernig við vinnum með mörg
vensl í SELECT.

Málfræði og merking SQL

SQL er tiltölulega læsilegt, yfirleitt er hægt að skilja einfaldar SELECT fyrirspurnir.

Eftir því sem við lærum meira um SQL þurfum við að hafa í huga

- Málfræði SQL. Hvernig **má** skrifa SQL setningar
- Merkingarfræði SQL. Hvað **þýðir** SQL setningin

Málfræðin er einföld. Til að tala um merkingarfræðina þurfum við að hafa líkan.

Venslalíkanið

- SQL er byggt á venslalíkaninu og venslaalgebru.
- Venslalíkanið gefur okkur nákvæma merkingu á SQL setningum.
- Venslaalgebra leyfir okkur að tala um jafngildar aðgerðir og endurskrifa fyrirspurnir.

Önnur hlutverk SQL

SQL er meira en bara fyrirspurnir. Við notum SQL til að

- Skilgreina ný vensl
- Breyta gögnum
- Setja upp skorður og kveiki
- Halda utan um notendur og öryggi
- Stýra hreyfingum á gagnagrunni fyrir marga notendur

Fyrir næstu viku

- Dæmi á vikublaði. Skil á mánudag.
- Setja upp sqlite.
- Skrá sig á [gradience.com](https://www.gradience.com) og gradescoppe
- Lesa kafla 1, 2.1, 2.2 og 6.1, 6.2