

Andmebaaside projekteerimine

Sissejuhatus. Kursuse korraldus.
Andmebaas ja andmebaasihalduse süsteem.
Põhimõisteid.
Andmemudel.

Kursuse üldandmed

Kood: IFI6013.DT

Maht: 3 EAP ($14 \times 2 = 28$ kontakttundi, $3 \times 26 - 28 = 50$ tundi iseseisvat tööd)

Kursuse veebileht:

http://www.cs.tlu.ee/~inga/AB_proj/

Korraldus

Igal nädalal 2 tundi arvutiklassis.

Osa sellest ajast toimub loengu vormis, osa praktikumina.

Praktiliste ülesannete lõpetamine kodus on üks osa iseseisvast tööst.

NB! Iseseisev töö ei ole ainult selleks, et seda õppejõule ette näidata, vaid ennekõike selleks, et oma teadmisi kinnistada ja seeläbi targ[em]aks saada.

Materjalid ja kirjandus

Täpsem info on kursuse veebilehel. Mõned kirjanduse viited:

- R. Elmasri, S. Navathe. *Fundamentals of Database Systems*, 7th Ed., Pearson 2016
- Priit Rospel. *Andmebaasisüsteemide alused*, loengumaterjal
<https://enos.itcollege.ee/~priit/1.%20Andmebaasid/1.%20Loengumaterjalid/>
- Erki Eessaar. *Andmebaaside projekteerimine*, TTÜ Kirjastus 2008
- Jaagup Kippar. *Andmebaaside projekteerimine 2017* (SQL näited),
<http://minitorn.tlu.ee/~jaagup/kool/java/kursused/17/ab/>
- A. Silberschatz, H. F. Korth, S. Sudarshan. *Database System Concepts*, 7th Ed., McGrawHill Education 2019

Teemad ja praktilised tööd

Teoorias:

- Üldised mõisted, olem-seos mudel, relatsioonmudel, võtmed, mudeli kvaliteet ja selle hindamine, normaalkujud, formaalsed soovitusel modelleerimiseks, aegpüsivad andmemudelid, SQL, ...

Praktikas:

- Relatsioonandmebaaside kavandamine, andmete modelleerimine – olem-seos skeem (ERD, Chen, varesejala tähistus).
- Relatsioonandmebaaside loomine, andmete lisamine, muutmine, päringud, – struktuurpäringukeel (SQL – *Structured Query Language*).

Ei tegele:

- igasugused muu arhitektuuriga andmebaasid ja AB haldussüsteemid, samuti relatsiooniliste baaside spetsiifilisemad teemad (kursus Andmebaasid tarkvaraarenduse moodulis);
- terviklikud rakendused ja infosüsteemid;
- kaunid kasutajaliidesed.

Tarkvara

Praktilisteks ülesanneteks on mõistlik kasutada tarkvara (ehkki modelleerida saab kuhugi maani ka paberi-pliiatsiga).

- Modelleerimiseks (ERD joonistamine):
 - QSEE SuperLite <https://www.leedsbeckett.ac.uk/qsee/>
 - ERDPlus (<https://erdplus.com>)
 - Vertabelo (<https://www.vertabelo.com/>)
- Andmebaasi loomiseks ja päringute katsetamiseks (SQL):
 - greeny.cs.tlu.ee-s relatsiooniline andmebaasihaldussüsteem MariaDB (MySQL-i autorite poolt loodud) ja tema käsurea kasutajaliides.
 - Äärmise vajaduse korral phpMyAdmin aadressil: <http://greeny.cs.tlu.ee/phpMyAdmin/>

Hindamine

- Iseseisev töö moodustab lõpphinde 40%.
 - Ülesannete lahendamine. Ülesanded sisaldavad olem-seos mudeli (kontseptuaalse) ja relatsioonmudeli (andmemudeli) loomist, andmebaasi ehitamist, andmebaasis päringute tegemist. Konkreetsed ülesanded antakse kursuse käigus. Ülesanded lahendatakse valdavalt väikestes meeskondades.
 - Iseseisvatest töödest arvestatakse lõpphinde peamiselt suuremat kodutööd, kus valitakse valdkond, modelleeritakse seda, koostatakse andmebaas ja mõned olulisemad päringud. Kogu tegevus dokumenteeritakse.
- Eksam moodustab lõpphinde 60%.
 - Eksamitööks on teooriatest ja mõned praktilised ülesanded.

Üldisematest mõistetest

Mis on andmebaas?

Kus on andmebaas?

Millised on Sinu kogemused andmebaasidega
ja kas need kogemused on head või halvad?

Mis on andmed?
Mis on informatsioon?

Andmed ja informatsioon

Andmed (kuidas on ainsus?)

Andmed (*data*) - informatsiooni taastõlgendatav esitus formaliseeritud kujul, mis sobib edastuseks, tõlgenduseks või töötluks.

Informatsioon (*information*) - teadmused, mis puudutab objekte: näiteks fakte, sündmusi, asju, protsesse või ideid, sealhulgas mõisteid, ja millel on teatavas kontekstis eritähendus.

Andmed – mingil viisil kogutud ja salvestatud toormaterjal.

Informatsioon – omavahel seotud andmed töödeldud ja organiseeritud kujul, tal on tähendus.

Teadmised (*knowledge*) – kogutud andmete põhjal tehtud järeldused ja üldistused.

Andmed on informatsiooni kandja (nagu paber on andmete kandja).

Andmebaas

Andmebaas (*database*) on kogum omavahel seostatud andmeid.

Andmebaasi (AB) omadused:

- esindab mingit osa tegelikust maailmast;
- on loogiliselt sidus ja tähendusega andmekogum;
- on loodud kindla eesmärgiga ja on vajalik kasutajatele nende eesmärkide saavutamiseks;
- koos andmetega hoitakse andmete struktuuri ja koosluse kirjeldust.

Andmebaaside tekkest ja ajaloost loe P. Raspeli e-kursuse 1. peatükis.

Andmebaasihalduse süsteem

Andmebaasihalduse süsteem (*database management system - DBMS*) on enamasti üsna keerukas tarkvara. Ta vahendab ligipääsu andmetele, lubades andmeid kirjeldada, hallata ja taasesitada. Lisaks tagab ta mitmete omaduste täidetust nii andmete kui ka nendele ligipääsu osas.

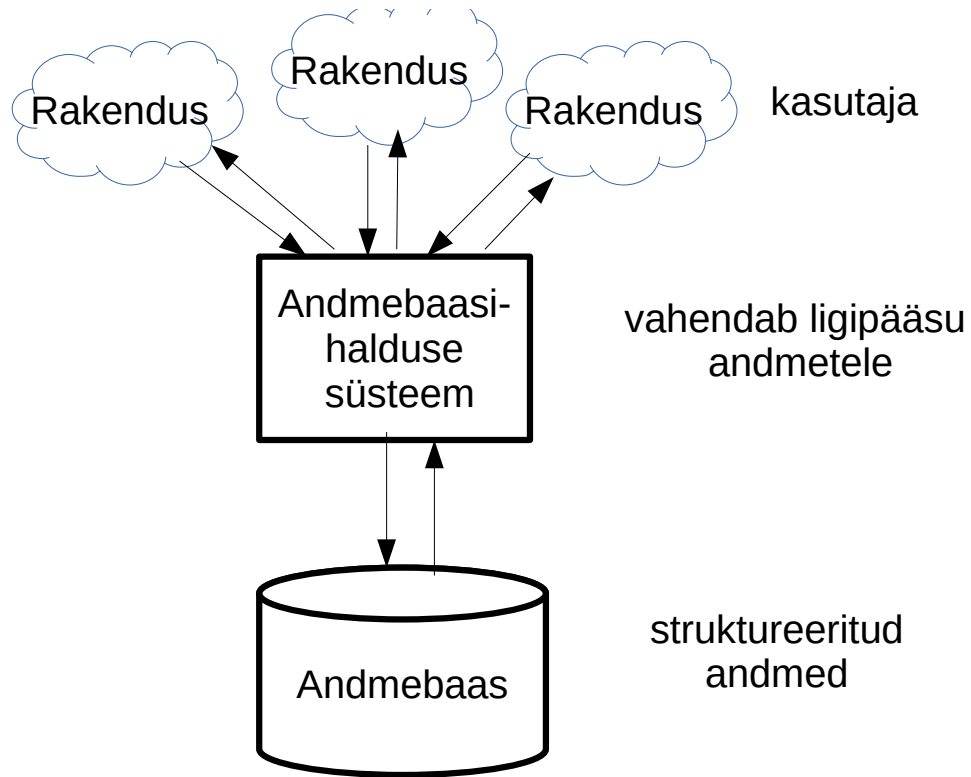
ABHS täidab järgmisi ülesandeid:

- On võimeline salvestama andmeid.
- Tagab andmete kestvuse / püsivuse.
- Annab vahendid soovitud andmete kiireks väljaotsimiseks.
- Tagab turvalisuse.
- Võib ka aidata leida andmetes mustreid ja arengusuundi.

Andmebaasihalduse süsteemid

Andmebaasihalduse süsteemide populaarsuse edetabel:

<https://db-engines.com/en/ranking>



ABHS funktsioonid

ABHS vahendab olulisi tegevusi andmebaasiga.

AB loomine, kustutamine, haldamine

- AB struktuuri loomine ja muutmine, kasutajate loomine ja õiguste andmine, andmete varundamine ja taastamine

AB kasutamine

- Andmete lisamine, muutmine, kustutamine, päringute tegemine, protseduuride käivitamine

Kust (miks) tulevad andmebaasid?

Andmebaase ei tehta reeglina andmebaasi enda pärast.

Andmebaasid aitavad lahendada ülesandeid, kus on tarvis tegelda suuremate andmehulkadega.

Nad on osaks paljudes veebirakendustes, sotsiaalmeedias, ettevõtete infosüsteemides, raamatupidamisprogrammides, personaliarvestuses jne.

Andmebaasid luuakse koos neid kasutava ja inimestele vahendava tarkvaraga ning nii on nad tihedalt seotud tarkvaraarendusega.

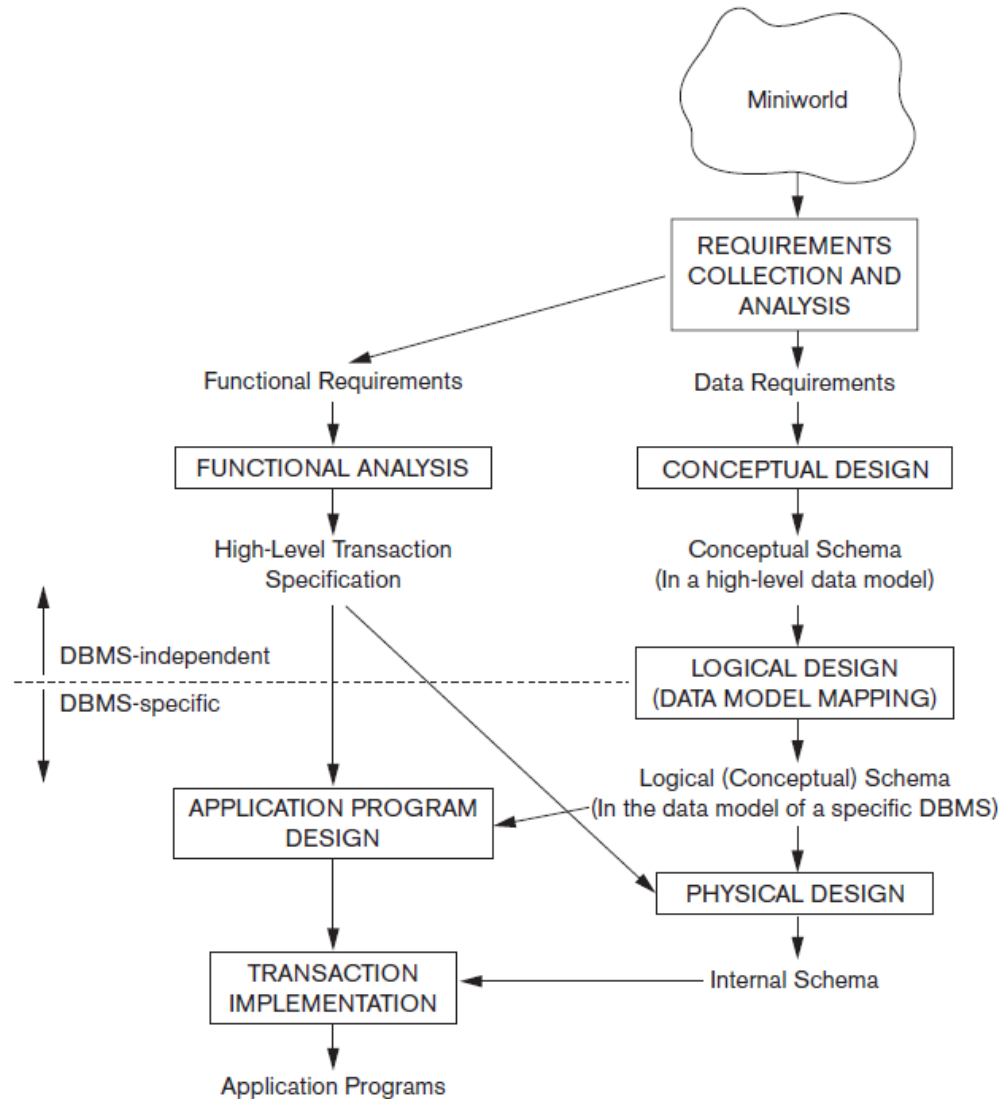
Tarkvaraarendus

Tarkvaraarendust “tükeldatakse” etappideks / tegevusteks mitmel erineval viisil. Üks võimalusi on järgmine:

- Nõuete analüüs ja kirjeldamine
- Arhitektuuri kavandamine
- Realisatsioon ja moodulite testimine
- Integratsioon ja süsteemi testimine
- Kasutamine ja hooldus

Kuhugi siia peab enamasti mahtuma ka andmebaasi loomine.

Järgmisel slaidil on joonis Elmasri & Navathe raamatust, mis kirjeldab tarkvaraarendust rõhuga andmebaasi loomisele.



Kavandamine

Jooniselt tulevad välja kaks olulist tegevust andmebaasi loomisel:

- **kontseptuaalne disain:** tulemuseks üldine kontseptuaalne (mõiste)mudel
- **loogiline disain:** tulemuseks loogiline/tehniline andmemudel, realisatsioonimudel (andmebaasitüübi jaoks spetsiifiline)

Disain on siin kontekstis **kavandamine** ehk **projekteerimine** (mitte kaunistamine / kujundamine).

Kavandamise tulemuseks on mudelid, mida ingl keeles siin kontekstis tavatsetakse *schema*'deks kutsuda.

Andmebaasi kavandamine on tegeliku elu mõistete-andmete ja nende vaheliste suhete teisendamine parimal võimalikul viisil andmebaasi struktuuriks.

Mudel

Mis on mudel? Näiteid?

Mudel on tegelikkuse lihtsustatud kirjeldus, mis toob välja olulise ja jätab kõrvale ebaolulise.

Oluline ja ebaoluline on suhteline ja sõltub kontekstist.

Mudel peab olema täpne ning esitatud arusaadavalt sellele “seltskonnale”, kelle jaoks seda vaja on.

Mudel võib olla kirjeldatud tekstina, kuid tihti on abiks joonis (pilt vs 1000 sõna).

Modelleerimine on mudeli koostamine.

Andmemudel

Andmemudel (*data model*) on mõistete ja põhimõtete süsteem, mille abil saab kirjeldada andmestiku ülesehitust, andmeelementide omavahelisi seoseid.

Meie kontekstis on andmemudeleid vaja andmete struktuuri ja seoste kirjeldamiseks ning teisalt konkreetse andmebaasi struktuuri kirjeldamiseks (andmetüübid, seosed, kitsendused).

Andmemudelisse võivad kuuluda ka põhioperatsioonid andmete manipuleerimiseks ning võimalus täiendavaid operatsioone kirjeldada.

Andmemudelite kategooriad

Jaotus vastavalt andmete kirjelduse tasemele:

- **Kontseptuaalne andmemudel** (*conceptual data model*) – kõrgema taseme mudel, mis annab vahendid andmete kujutamiseks nii, nagu kasutajad neid tajuvad.
- **Füüsiline andmemudel** (*physical data model*) – madala taseme mudel, annab vahendid, et kirjeldada andmete salvestamist salvestusseadmel, näiteks kõvakettal.
- **Tehniline andmemudel** (*representational data model*) – jääb eelnevate vahele ja annab andmete kirjeldamiseks vahendid, mis on arusaadavad kasutajale, kuid teisalt ei ole liiga kaugel andmete salvestamisest.

Tehniline andmemudel

Andmemudelid, mida on kasutatud ja kasutatakse andmebaasihalduse süsteemides:

- Pärandandmemudelid (*legacy data models*):
 - **Hierarhiline andmemudel** (1960ndad, IBM) – vanim AB struktuur, andmete vahel on puukujulised seosed – üks vanem, mitu last. Inkarnatsioon XML-s.
 - **Võrkmodel** (1960ndate lõpp) – püüdis parandada eelmise jäikust, kuid kaotas populaarsuses relatsioonilisele mudelile. Tänapäevane graafimudel on mingis mõttes järglane.
- **Relatsioonandmemudel** (1970ndad, Edgar F. Codd)
- Isekirjeldavad andmemudelid, dokumendiandmebaasid, NoSQL süsteemid.

Andmemudelite alusel on loodud andmebaasihalduse süsteemid, mis toetavad vastavat tüüpi andmebaase. Näiteks relatsioonandmebaasisüsteemid töötavad relatsioon-andmemudelitega.

Andmemudel (uesti)

Praktikas nimetatakse andmemudeliks tihti konkreetse andmebaasi kirjeldust (struktuuri).

Andmemudel (*data model*) on andmebaasi ülesehituse kirjeldus. Kasutatakse mõistet **andmebaasi struktuur** või **skeem** (*schema*).

Andmebaasi struktuuri kirjeldamisel lähtutakse andmebaasi tüübist ning vastavast andmemudelist (meie kasutame **relatsioonmudelit**).

Andmebaasi struktuur ja andmebaas on erinevad mõisted.

Andmemudel

Kontseptuaalne andmemudel ja kontseptuaalne kavandamine – kirjeldab tegelikkuses eksisteerivaid andmeid, mõisteid, seoseid

Loogiline andmemudel ja loogiline kavandamine – kontseptuaalne mudel vastendatakse loogiliseks ehk realisatsioonmudeliks. Viimase järgi luuakse tegelik andmebaas.