

Romlig datamanipulering

Gunnar Tenge, 18.04.08

Romlige manipuleringsmetoder brukes i GIS-analyser. I denne artikkelen forklares alle manipuleringsmetoder som man kan forvente å finne i et GIS-program.

Et GIS-program du har installert på din PC kalles ofte en tykk klient, mens en tynn klient er et kartinnsynsverktøy i web-browsersen din. De tynne klientene får mer og mer funksjonalitet, men det vil ta mange år før de overtar for de tykke klientene når det gjelder å kunne manipulere geodata. Manipuleringsmetoder kalles også operasjoner eller kommandoer eller verktøy (tools).

Innhold:

1. Noen basisoperasjoner.....	1
2. Utvalg.....	1
2.1. Utvalg på egenskaper:.....	2
2.2. Geografiske utvalg:.....	2
3. Forandre, forenkle og øke informasjonen til egenskapene.....	2
4. Forandre, forenkle og øke informasjonen til geometrien.....	3
4.1. Overlay-operasjoner:.....	4
5. Oppsummering.....	6

1. Noen basisoperasjoner

I alle GIS-verktøy bør man kunne utføre beskrivende statistikk på alle felter i databasen. Som å telle opp, finne max-verdier, finne min-verdier, gjennomsnittsverdier, standardavvik, etc. Man bør gjerne også få opp grafer som viser fordelingen av verdiene. Disse operasjonene er ikke særegent for GIS-verktøy, men er likevel meget nyttige. Det er lett å glemme etter ett par tiår med datamaskiner på kontorpulten at en datamaskin er utrolig mye raskere og nøyaktigere enn mennesker til å telle opp, finne max-verdien, osv. til en lang rekke (mange tusen) verdier.

To basisfunksjoner som er særegne for GIS-verktøy, er å beregne lengde og areal. Alle GIS-verktøy skal kunne gjøre dette. Tidligere brukte man linjal, hyssing og evt. en mekanisk "rullemåler" (<http://en.wikipedia.org/wiki/Opisometer>) for å måle lengder. For arealer kunne man bruke prikketelling (http://www.geog.ubc.ca/courses/geob373/labs/dot_grid_instructions.html) eller planimeter (<http://en.wikipedia.org/wiki/Planimeter>). Nå gjøres dette med tilgjengelige lengde- og arealberegningsverktøy i GIS-programvaren. Det er også lett å glemme hvor utrolig mye raskere og nøyaktigere en datamaskin er til beregne lengder og arealer (http://en.wikipedia.org/wiki/Polygon#Area_and_centroid) enn mennesker. Det å kunne beregne arealene til mange (mange tusen) polygoner fort, ble sett på som noe av det viktigste til noen av GIS-programmene som kom på 70-tallet.

2. Utvalg

Skal man gjøre en analysejobb må man få tak i data. Ofte får man tak i data med flere objekter enn man trenger eller data som dekker et større område enn man trenger. Da må det gjøres et utvalg.

Å gjøre et utvalg fra et stort datasett er gjerne der man begynner når man skal gjøre en analysejobb.

- a. Gjøre et utvalg
- b. Analysere dette utvalget

2.1. Utvalg på egenskaper:

Eks. velg alle mennesker som er 6 år. Da må man typisk skrive denne select-setningen, avhengig av det GIS-programmet man har: `Select age = 6`.

Det er meget streng syntaks (~ grammatiske regler for hvordan spørringen skal skrives) i select-setninger som GIS-programmene vil kunne tolke. Det finnes en standard som heter Standard Query Language (SQL). Den har denne syntaksen: `select .. from .. where`, men de forskjellige GIS-programmene har gjerne sin egen syntaks. Her må det leses i help'en til det enkelte program.

Det settes opp en utvetydig påstand og det som er sant blir valgt.

Påstandene inneholder en logisk operasjon som "=", ">", "<", osv. som kan kombineres med "and", "or", osv. for å lage kombinert utvalg.

Digresjon: En morsomhet fra logikkens verden. Hvis vi på norsk (og for så vidt på engelsk) skal si muntlig at du fra en gruppe med barn skal velge alle som er 6 år og 7 år, så bruker vi og som operator. Men i logikkens verden blir dette feil. Påstanden: `Select age = 6 and 7`, returnerer ingen barn fordi ingen barn er 6 og 7 år samtidig! Man må bruke eller. `Select age = 6 or 7`. Så hvis du en dag får beskjed om å velge alle barn som er 6 år og 7 år, så kan du være vanskelig å si at det er dessverre ingen!

2.2. Geografiske utvalg:

Det å kunne gjøre geografiske utvalg er særegent for GIS-verktøy. Dataene har en romlig dimensjon som gjør at vi kan velge ut fra beliggenhet. Typiske å velge alle innefor, utenfor, som overlapper, osv.

Å klippe ut noe av en geometri er noe mer enn et geografisk utvalg. Klipper linjer og polygoner med omrisset av det polygonet man klipper med. En form for overlay (overlay er forklart senere i teksten).

Husk at geometrien og egenskapene i et geodatasett henger sammen. Gjøres det et utvalg på egenskapene, så velges også de geometriske formene til objektet, og gjøres det et geografisk utvalg så velges også egenskapene til de geometriske formene som er valgt.

3. Forandre, forenkle og øke informasjonen til egenskapene

De følgende manipulerings-teknikkene har vært kjent i flere tiår i relasjonsdatabaser (tabelldatabaser) og er ikke særegne for GIS-verktøy.

Forandre - kalles også å reklassifisere eller å omkode.

Har for eksempel en kolonne med kvadratmeterverdier og forandrer den til dekar ved å dele på 1000. Innholdet er kun forandret, ingen informasjon har blitt tilført eller tatt bort.

Forenkling – kalles også å gruppere eller å generalisere.

Har for eksempel en kolonne med alder i år og forenkler denne til aldergruppene ung, middelaldrene og gammel. Innholdet er nå forenklet. Informasjon har gått tapt.

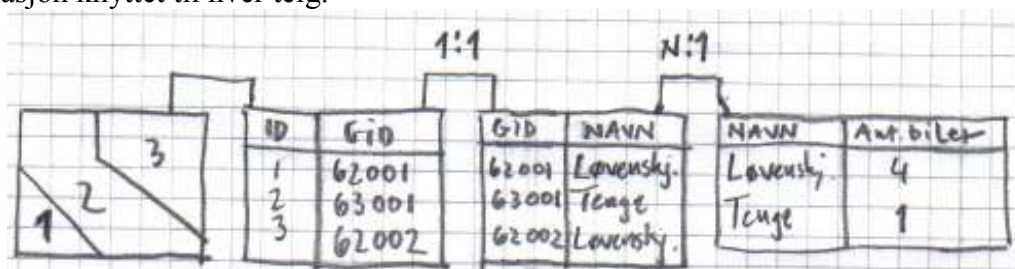
Øke informasjonen – kalles også å koble.

Det går egentlig ikke an å øke informasjonen til en tabell, men man kan koble på en ny tabell med felles koblingsnøkkel (for eksempel fødselsnummer). Dette er den viktigste operasjonen i et informasjonssystem – man kan koble flere tabeller som til sammen fremstår som mer informasjon. Dette er også hele ideen med relasjonsdatabaser.

De enkleste er 1:1- og N:1-koblinger, men det er også fullt mulig å gjøre en N:N- og 1:N-koblinger.

Eksempel:

Egenskapene til tre eiendomsteiger kobles med to andre tabeller slik at man får mer informasjon knyttet til hver teig.



4. Forandre, forenkling og øke informasjonen til geometrien

De følgende manipulerings-teknikkene er særegne for GIS-verktøy.

Forandre - en type forandring av geometrien er **buffer**. Med et bufferverktøy kan man lage polygon fra punkt, linje eller polygon.

Nærhet er et meget viktig romlig konsept. Ofte ønsker vi å finne ut hva som ligger i nærheten av noe annet, da kan man lage en buffer som strekker seg for eksempel 200 meter ut fra objektet. Bufferpolygoner brukes ofte til å klippe med eller "overlay" (se under) med et annet geodatasett. Men å si at man har gjort en GIS-analyse når man har laget en buffer er vel å ta litt hardt i.

Forenkling - innholdet blir forenklet. Informasjon går tapt.

- **dissolve** - slå sammen polygoner eller linjer med en eller flere like egenskaper.
- **generaliser linjeforløp og polygoner** – finnes gjerne noen kommandoer for dette som "glatter" linjer. Være varsom med dette. Dette er vanskeligere å gjøre på en ordentlig måte enn det ser ut til. Generalisering er en gren av kartografien og mye er skrevet om dette.
- **gjøre et utvalg** og lagre dette utvalget

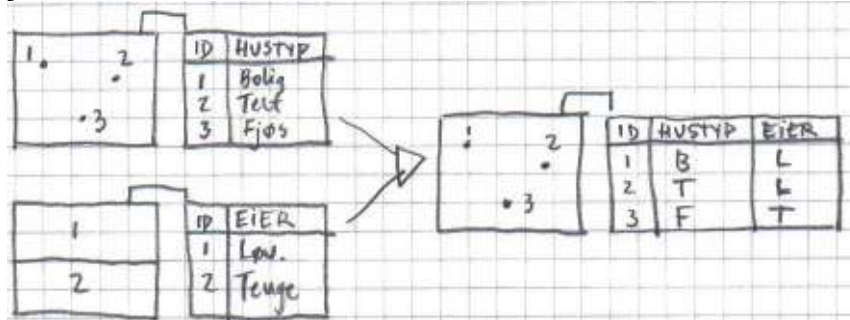
Øke informasjonen – kalles også romlige koblinger, spatial join eller **overlay**.

Det går egentlig ikke an å øke informasjonen til en geometri, men man kan koble den med en annen geometri vha. koblingsnøkkelen sted. Data som ligger samme sted blir koblet sammen.

Dette er den viktigste operasjonen i et Geografisk InformasjonsSystem – man kan koble flere romlige datasett som til sammen fremstår som mer informasjon.

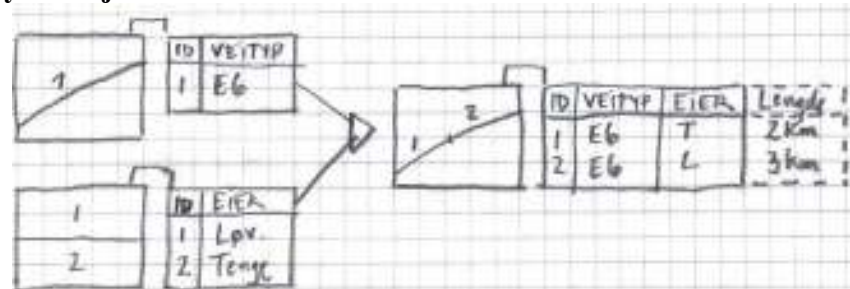
4.1. Overlay-operasjoner:

Punkt1 + Poly = Punkt2



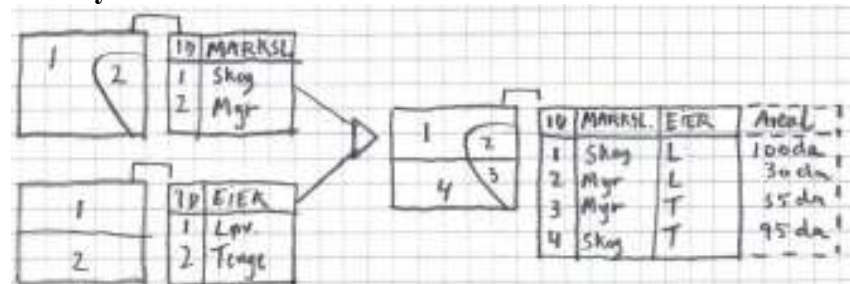
Punkt2 er geometrisk identisk med Punkt1, men inneholder informasjon om hvilke polygoner de havnet innenfor.

Linje1 + Poly = Linje2



Linje2 er geometrisk identisk med Linje1, men inneholder informasjon om hvilke polygoner de går gjennom. Linjen blir delt av polygonomrisset og linjetopologien oppdateres.

Poly1 + Poly2 = Poly3



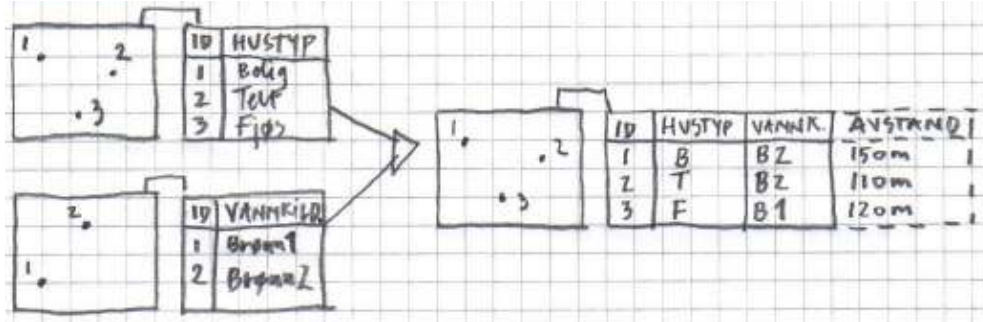
Poly3 er en kombinasjon av geometriene til Poly1 og Poly2 og inneholder informasjon fra både Poly1 og Poly2. Nye polygoner dannes og polygontopologien oppdateres.

Det splittes linjer, dannes knutepunkter, regnes ut arealer og omkretser og tildeles nye id-er.

Dette er en maskinkrevende operasjon som medfører oppdatering av en masse pekere og tabeller. De nye polygonene kalles CMU'er (Comparable Map Units) eller MMU'er

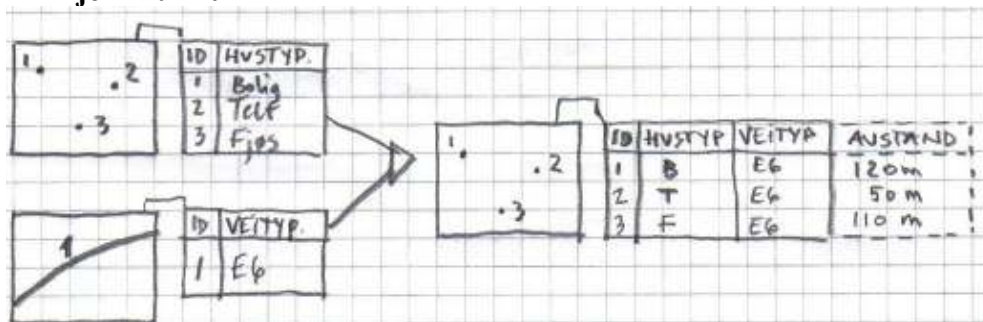
(Minimum Map Units). Det er også vanlig at det dannes en hel masse små-polygoner, slivers.

Punkt1 + Punkt2 = Punkt3



Punkt3 er geometrisk identisk med Punkt1, men inneholder informasjon om hvilke punkter fra Punkt2 som ligger nærmest og avstanden til punktene. Ofte kalles dette å finne den nærmeste naboen (nearest neighbour) i et annet punktdatasett.

Punkt1 + Linje = Punkt2



Punkt2 er geometrisk identisk med Punkt1, men inneholder informasjon om hvilke linjer som ligger nærmest og avstanden til dem.

I tillegg til disse overlay-operasjonene som skal finnes i alle GIS-programmer, så kunne man tenke seg et tre andre som jeg ikke har sett implementert noe sted:

Poly1 + Punkt = Poly2

Poly2 er geometrisk identisk med Poly1, men inneholder informasjon fra punktene som ligger innenfor polygonene. Kan løses med:

- Punkt + Poly = Punkt, gjør så en summering på de punktene som havner innefor hvert polygon, og så kobles dette med en vanlig tabellkobling til polygonene igjen.
- Punktene gjøre om til thiessenpolygoner som polygonoverlayes med Poly1.

Linje1 + Punkt = Linje2

Linje2 er geometrisk identisk med Linje1, men inneholder informasjon fra punktene som ligger nærmest linjene. Kan løses med:

Punkt + Linje = Punkt, gjør så en summering på de punktene som er nærmeste hver linje, og så kobles dette med en vanlig tabellkobling til linjene igjen.

Poly1 + Linje = Poly2

Poly2 er geometrisk identisk med Poly1, men inneholder informasjon fra linjebitene som ligger innenfor polygonene. Kan løses med:

- Linje1 + Poly = Linje2, gjør så en summering på de linjebitene som havner innefor hvert polygon, og så kobles dette med en vanlig tabellkobling til polygonene igjen.

5. Oppsummering

Alle disse manipulerings-teknikkene brukes i GIS-analyser. De aller viktigste er overlay-operasjonene som gjør at vi kan koble romlige datasett for å få ut mer informasjon.

Det er også verdt å nevne at vi trenger gode data, dvs. godt dokumentert og standardisert, før vi går i gang med å manipulere dem. Dårligere data blir ikke noe bedre om de generaliseres eller kobles med andre data. Her som ellers gjelder ”shit in = shit out”.