

Applikationsutveckling för överföring av adressinformation till GIS-databas

Erik Bengtsson

Fastighetsvetenskap
Lunds Tekniska Högskola
Lunds Universitet

Real Estate Science
Lund University, Sweden

ISRN LUTVDG/TVLM 00/5048, Lund 2000



Applikationsutveckling för överföring av adressinformation till GIS-databas

Development of an application for transfer of address
information to GIS database

Examensarbete omfattande 20 poäng utfört av:

Erik Bengtsson

Fastighetsvetenskap

Lunds Tekniska Högskola

Lunds Universitet

Handledare:

Bengt Rystedt, Avdelningen för Fastighetsvetenskap, Lunds Tekniska Högskola

Stefan Kallur, T-Kartor Sweden AB, Kristianstad

Examinator:

Åsa Knutsson, Avdelningen för Fastighetsvetenskap, Lunds Tekniska Högskola

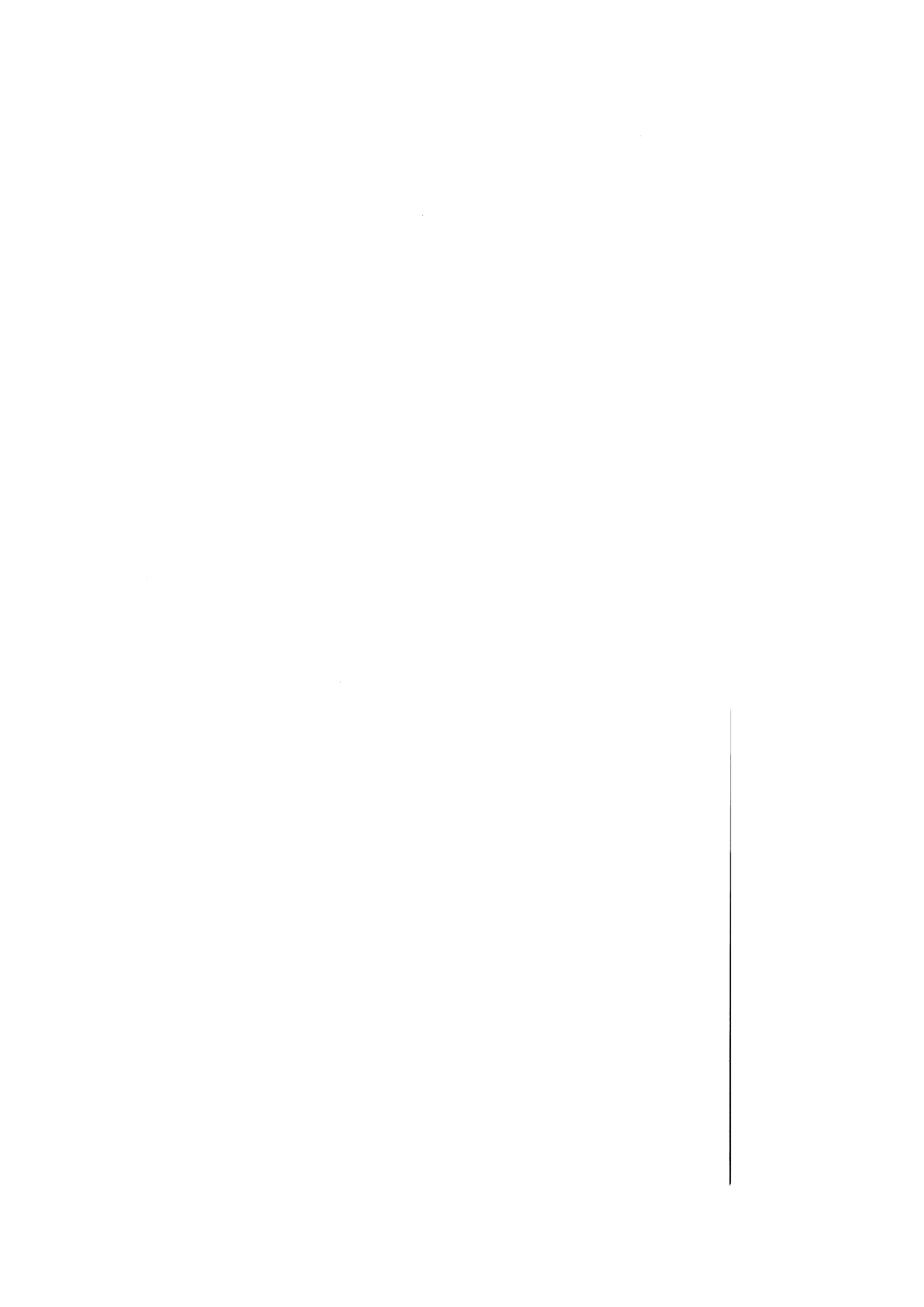
Juni 2000

ISRN LUTVDG/TVLM 00/5048

Sökord: GIS, applikationsutveckling, ArcView, geokodning

Key words: GIS, application development, ArcView, geocoding

Language: Swedish



INNEHÅLLSFÖRTECKNING

Sammanfattning	5
Summary	5
1 INLEDNING	7
1.1 Allmänt	7
1.2 Syfte	7
1.3 T-Kartor	7
1.3.1 Historia och verksamhet	7
1.3.2 Kartproduktion	7
1.3.3 Databasdriven kartografi, CPS	7
1.3.4 Data och konsultverksamhet	8
1.3.5 Internet	8
1.4 Rapportbeskrivning	8
2 GEOKODNING MED HJÄLP AV ADRESSMATCHNING	10
2.1 Belägenhetsadress	10
2.2 Adressmatchning	11
2.3 Användning av geokodning	11
2.4 Geokodning i ArcView	12
2.4.1 Adressmatchning i ArcView	12
2.4.2 Hur ArcView bestämmer adressens koordinater	13
2.5 Geokodning med Avenue	13
2.6 Sammanfattning	15
3 GIS OCH ARCVIEW	16
3.1 Vad är GIS?	16
3.2 Geografiska data	17
3.2.1 Raster	17
3.2.2 Vektor	18
3.2.3 Topologi	18
3.3 Datakvalitet	18
3.3.1 Positionsnoggrannhet	18
3.3.2 Noggrannhet i attributdata	19
3.3.3 Konsistens av länkning mellan geometriska data och attributdata	19
3.3.4 Konsistens av topologi	19
3.3.5 Geografisk täckning och upplösning	19
3.3.6 Aktualitet	19
3.3.7 Felkällor	19
3.4 ArcView	20
3.4.1 Arbeta med ArcView	20
3.4.2 Geometriska data i ArcView	21
3.4.3 Avenue	21
3.5 Objektorientering	21
3.5.1 Introduktion	21
3.5.2 Objektorientering i Avenue	22

3.5.3	Skapa applikationer i ArcView m.h.a. Avenue	23
3.5.4	Vad är ett script?	23
3.5.5	Anpassning	23
4	GIS-DATABASEN GEOTÄTORT	24
4.1	Databasbeskrivning	24
4.1.1	Geografisk täckning	24
4.1.2	Struktur	25
4.1.3	Innehåll	25
4.1.4	Databasstruktur	26
4.1.5	Kvalitet	27
4.1.6	Metadata	30
4.2	Tillämpningar och användningar av GeoTätort	30
4.2.1	Kartografi	30
4.2.2	Marknadsanalys	31
4.2.3	Transport	31
5	PROBLEMLÖSNING	34
5.1	Syfte	34
	Problembeskrivning	34
5.1.2	Krav	35
5.1.3	Användning	35
5.2	Förutsättningar	35
5.2.1	Programmiljö	35
5.2.2	Indata	35
5.3	Överföring av adressinformation från ett adressregister till GeoTätort i ArcView	37
5.3.1	Koppling av adressnummer till vägsegment	37
5.3.2	Geokodning	38
5.3.3	Ändring av felaktiga adresser	39
5.3.4	Repetering	40
5.4	Testning	41
5.4.1	Testdata	41
5.4.2	Testresultat	41
5.4.3	Tidsåtgång	41
5.5	Slutsatser	42
Litteraturförteckning		45
	Referenser på Internet	46

Sammanfattning

I ett geografiskt informationssystem (GIS) kan uppgifterna i ett adressregister, en databas som innehåller information kopplat till adresser, visualiseras grafiskt. För att göra denna visualisering görs en s.k. geokodning av adressregistret mot en databas där adresserna finns geografiskt lokaliserade. Geokodning är en process för att koppla ihop en geokod (t.ex. adress) med koordinater. Vid geokodningen påförs koordinaterna i den geografiska databasen till uppgifterna i adressregistret.

Genom denna geokodningsprocess kan uppgifterna i adressregistret få en utökad användning för presentation, marknadsanalys, planering o.s.v.

För att geokodningen skall kunna genomföras krävs alltså att det finns en databas med geografisk information. T-Kartor Sweden AB använder databasen GeoTätort som geografisk referens i en stor del av sin verksamhet. För att öka databasens kvalitet vill man kontrollera och tillföra adressinformation till vägnätet i GeoTätort genom att använda externa register med koordinatsatta adresser. I vissa fall vill man använda adressregistret för att överföra adressinformationen i det externa adressregistret till den egna databasen. I andra fall vill man främst använda uppgifterna i adressregistret till att jämföra med uppgifterna i GeoTätort och på så sätt kunna identifiera och justera eventuella felaktigheter.

Genom detta examensarbete vill T-Kartor få en praktisk lösning för hur man med hjälp av ett adressregister skall kunna tilldela vägsegmenten i GeoTätort attribut för vilka adressnummer som finns på vänster respektive höger sida. Den främsta fördelen med en sådan lösning är att operationen med att överföra adressinformation automatiseras och därmed sparar arbetstid. Kvaliteten på resultatet av informationsöverföringen är emellertid svår att bestämma. Resultatet beror till stor del på lägesnoggrannheten på informationen i adressregistret.

Examensarbetet är utfört med största tonvikt på den praktiska lösningen av uppgiften, som utnyttjat i ett program för GIS-programvaran ArcView.

Rapporten redogör dels för en lösning för att tilldela vägsegmenten dess attribut för adressnummer, dels för att identifiera och rätta eventuella felaktigheter som kan finnas i GeoTätort. Rapporten beskriver förutom den praktiska lösningen även databasen GeoTätort, GIS-programvaran ArcView och principerna för geokodning med hjälp av adressmatchning.

Summary

The information in an address database can be graphically visualised using geocoding. The address database is geocoded with a database containing addresses with geographic coordinates. In the geocoding process the coordinates are transferred to the address database by comparing the addresses.

By using geocoding the information in an address database can be of more use in presentations, market analyzing, planning and so on.

T-Kartor Sweden AB is using the geographic database GeoTätort in many areas of the business. By using address databases with geographically located addresses T-Kartor wish to check, correct and maybe add information to GeoTätort and thereby increase the quality of the database. In some cases T-Kartor wish to copy the address information in the address database to GeoTätort. In other cases the objective is to check the information for spelling mistakes, incorrect coordinates or other defective information.

The objective of this Master of Science thesis is to achieve an implementation of how to assign the address number attributes of the street segments in GeoTätort to represent the first and last number on each side of every segment by using an address database with geographically located addresses. The result of the implementation is a program application for the GIS software ArcView. This application will mean an automation of the process of assigning the address attributes for the street segments in GeoTätort.

This paper describes the implementation of the ArcView application. The program application assigns the address attributes for the street segments and also identifies and corrects errors that might occur in GeoTätort. The paper describes the GeoTätort database, the GIS software ArcView and the principles of geocoding.

1 Inledning

1.1 Allmänt

Föreliggande examensarbete har utförts som en del av utbildningen vid civilingenjörsutbildningen i lantmäteri, 180 poäng, vid Lunds Tekniska Högskola. Uppdragsgivare för examensarbetet har varit T-Kartor Sweden AB, Kristianstad. Handledare har på LTH varit prof. Bengt Rystedt, på T-Kartor Stefan Kallur.

Jag riktar ett tack till de anställda på T-Kartor som varit till hjälp vid arbetet, speciellt till min handledare Stefan Kallur för hjälp med information och litteratur och till Eric Ericsson för råd om implementering.

1.2 Syfte

Uppgiften är att utveckla en metod att använda ett adressregister för att koppla adressinformation till T-Kartors databas GeoTätort. Med hjälp av ett adressregister med koordinatsatta adresser önskar man tillföra information om vilka adressnummer som finns på varje vägsegment i den befintliga databasen. T-Kartor söker också en metod för att automatiskt identifiera, granska och åtgärda de fel som upptäcks vid jämförelse mellan ett adressregister och adressuppgifterna i T-Kartors databas GeoTätort.

Examensarbetets syfte är till största delen en praktisk lösning av ovanstående problem; problemlösning och implementering av ett program i GIS-programvaran ArcView för informationsöverföring från adressregistret till GeoTätort. Vägsegmenten i GeoTätort ska förses med uppgift om första och sista adressnummer som finns på varje vägsegments vänstra respektive högra sida.

1.3 T-Kartor

1.3.1 Historia och verksamhet

T-Kartor Sweden AB grundades 1985 av Sten Ravhed och var då ett rent kartproduktionsbolag. Tanken var att utnyttja den digitala tekniken för framställning av kartor. Sedan 1989 då T-Kartor inledde satsningen på databasdriven kartografi har företaget utvecklats i hög takt. Idag har företaget fyra arbetsområden: kartproduktion, mjukvaruutveckling, data och konsultverksamhet.

T-Kartor ägs idag av VD Sten Ravhed och Jack Dangermond, grundare av och ägare till Environmental Systems Research Institute (ESRI), Inc.

1.3.2 Kartproduktion

T-Kartor är Sveriges största privata producent av kartmaterial. Bland företagets kunder finns Telia InfoMedia, Vägverket och flera kommuner, turistorganisationer och företag inom reklam och marknadsföring.

Företagets största produkt inom kartografiområdet är sedan flera år kartorna som finns i Telias telefonkataloger. De s.k. Röda Sidorna är Sveriges mest använda kartor och utnyttjas dagligen för att leta upp adresser i tätorter (T-Kartor Sweden AB, 1997a).

1.3.3 Databasdriven kartografi, CPS

T-Kartors databasdrivna kartografi gör att en rad produkter kan skapas utifrån samma databasunderlag (Produktion Data CPS Konsulting, 1996).

1989 bestämde sig företaget för att satsa på databasdriven kartografi. ARC/INFO, en av världens ledande GIS-programvaror, saknade dock en applikation för kartografi. Med bistånd från ESRI, Inc., skapare av ARC/INFO, utvecklade T-Kartor en egen applikation. Den första versionen kallades Cartographics vilken senare utvecklades vidare till Cartographics Production System, CPS.

Från början fanns inga intentioner om att sälja CPS, det var ett program endast avsett för intern användning. Men när andra kartografiföretag visade intresse för vilka möjligheter som CPS kunde erbjuda tog T-Kartor steget ut på mjukvarumarknaden.

Idag använder drygt 160 olika företag och organisationer i 30 olika länder CPS. CPS används såväl för att tillverka atlas och cityguider, som för att göra militära, topografiska och geologiska kartor (www.t-kartor.se, 1998).

1.3.4 Data och konsultverksamhet

T-Kartor tillhandahåller digitala kartdatabaser och information lagrad geografiskt (t.ex. från Statistiska Centralbyrån, SCB). Dessa data kan användas som grunddata i kundens geografiska informationssystem eller för kartframställning.

Konsultverksamheten inom T-Kartor är fokuserad på tre huvudområden: kartografi, transport/navigation och marknadsanalys. Verksamheten bygger på företagets fleråriga erfarenhet inom GIS-branschen. T-Kartor kan också erbjuda implementering och utveckling av kundspecifika applikationer för geografiska informationssystem (Produktion Data CPS Konsulting, 1996).

T-Kartor har också varit med i utvecklingen av Hydrographic Information System, HIS, för Sjöfartsverken i Sverige och Finland. HIS är ett system för insamling, hantering och kontroll av hydrografiska data. HIS erbjuder också möjligheter för att tillverka elektroniska sjökort (T-Kartor Sweden AB, 1997b).

1.3.5 Internet

Kartor på Internet/Intranet är ett av T-Kartors senaste arbetsområden. Bland annat har företaget varit med och utvecklat applikationen för Gula Sidorna på Internet (www.gulasidorna.se), där man kan söka efter ett företag och sedan få dess lokaler markerade på en digital karta på sin bildskärm. Kartan är dessutom zoom- och panoreringsbar vilket underlättar för att kunden ska kunna navigera rätt (T-Kartor Sweden AB, 1997b).

1.4 Rapportbeskrivning

Rapporten inleds med en sammanfattning av examensarbetet. I första kapitlet beskrivs vilket syfte och vilka förutsättningar som funnits för arbetets genomförande.

Geokodning är ämnet för kapitel 2. Dels beskrivs vilka möjligheter som geokodningen erbjuder, dels förklaras geokodningsproceduren i ArcView och dess programmeringsspråk Avenue.

I kapitel 3 görs dels en kortfattad förklaring av begreppet geografiska informationssystem, GIS, och vad som karakteriserar ett sådant, dels en summarisk genomgång av den programvara som använts för examensarbetet, ArcView.

I det fjärde kapitlet görs en omfattande beskrivning av GeoTätort, den databas som T-Kartor främst använder i sin verksamhet och som använts vid utförandet av examensarbetet.

Kapitel 5 beskriver hur lösningen av uppgiften som ligger till grund för arbetet i praktiken ser ut och vilka slutsatser som har dragits.

2 Geokodning med hjälp av adressmatchning

De flesta organisationer och företag har register eller databaser innehållande adresser som t.ex. kundregister, elevregister, brottsregister. Med s.k. geokodning kan man lokalisera sina tabelldata innehållande adresser genom att varje element representeras som en punkt på en karta med hjälp av en process som kallas adressmatchning. På så sätt kan man göra t.ex. marknadsanalyser där ett adressregister knyts till en geografisk databas för grafisk presentation. Med en koppling till t.ex. tidigare SCB:s socioekonomiska data kan analysen kompletteras med information om ett område som t.ex. befolkningens sammansättning, ålder, inkomster mm.

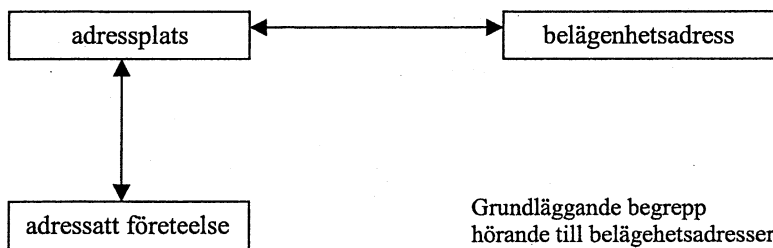
En adress representerar ett läge på samma sätt som en geografisk koordinat gör. Men eftersom en adress är en enkel textsträng innehållande information om gatunamn, husnummer och kanske ett postnummer och postort, krävs en operation för att ta fram den koordinat som ska representera adressen. För att göra detta krävs att adressregistret förbinds med ett geografiskt bestämt referensdata över motsvarande område, vanligen ett gatunät. Om referensmaterialet innehåller adresser kan detta användas för att bestämma och tilldela objekten i adressregistret koordinater. Denna operation som används för att lokalisera adresser på en karta kallas geokodning.

Genom användning av ArcView-applikationen som beskrivits i föregående kapitel har vägsegmenten i GeoTätort försetts med adressuppgifter. GeoTätort-databasen kan sedan användas som referensdata vid geokodning av register som innehåller information kopplad till adress.

2.1 Belägenhetsadress

Belägenhetsadresser är ett sätt att ange en plats utan att använda koordinater. En belägenhetsadress anger var en plats finns, till skillnad mot postala adresser som anger vart post skall skickas för att hamna hos en adressat. Sveriges kommuner sätter belägenhetsadresser och användarna är många inom både offentlig och enskild sektor (SIS, 1998, s. 2). Standardiseringen i Sverige, SIS, har upprättat en standard för hur belägenhetsadresser skall utformas. En belägenhetsadress sätts på en plats för att där finns något man vill peka ut. Plats kallas i SIS standarden för adressplats och det som pekas ut för adressatt företeelse. Via adressplatsen kan man koppla koordinater till belägenhetsadressen. Rekommendationerna i standarden är till för att öka kvalitén på Sveriges belägenhetsadresser och därmed öka dess användbarhet (SIS, 1998, s. 3).

För att en belägenhetsadress entydigt ska peka ut en adressplats byggs adressen upp i fem



Grundläggande begrepp
hörande till belägenhetsadresser
Fig. 2.1

nivåer.

1. Kommun
2. Geografisk kommunedel
3. Adressområde (gata, väg, torg, gård eller liknande)
4. Gårdsadressområde (ej obligatoriskt)
5. Adressplats (entré, port eller liknande)

(SIS, 1998, s. 5-6)

I denna rapport används begreppet adress och består i regel av adressnivå tre och fem. Vidare beskrivning av adressbegreppen kan hittas i SIS rapport för standardisering *Belägenhetsadresser, utgåva 1 1998, Svensk Standard SS 63 70 03*.

2.2 Adressmatchning

Adressmatchning är en process som jämför två adresser för att bestämma om de är lika. Vanligtvis jämförs adresserna i ett adressregister med adresserna i ett koordinatsatt referensdata med hjälp av ett GIS-program.

Adressens olika komponenter (se ovan, avsnitt 2.1) i adressregistret jämförs med motsvarande element i referensdatafilen. Adressplatsen (adressnumret) lokaliseras i rätt nummerintervall för vägsegmenten. En adress har matchat om adressens komponenter är desamma i adressregister och referensdata och adressnumret kan lokaliseras i något av vägsegmentens adressnummerintervall.

Exempel: KAPTENSGATAN 11 29 12 24, är ett vägsegment i referensdatafilen som på vänster sida har udda nummer mellan 11 och 29 och på höger sida jämna nummer mellan 12 och 24. KAPTENSGATAN 23, ett element i adress Tabellen, matchar mot vägsegmentet KAPTENSGATAN 11 29 12 24 eftersom dess adressnummer är inom intervallet för vägsegmentets udda nummer.

Adressmatchning är emellertid sällan en helt automatiserad process. Man bör känna till de problem som kan uppstå vid processen och hur resultatet kan påverkas. En adress matchar inte p.g.a. att den inte är kompatibel med de digitala kartdata som används som referens. Det kan vara felaktigheter i de adressdata som ska matchas eller i de digitala kartdata (The Department of Geography, University of Iowa, 1999). Det kan vara varierande stavning av gatunamn, olika förkortningar, dubbla adresser i korsningar eller postadresser. I de digitala kartdata med gatunätet kan det vara felaktiga eller saknade adressintervall eller kanske saknade vägsegment. (Department of Geography, Central Connecticut State University, 1999). Adressmatchningsgraden, d.v.s. graden av korrekt matchade adresser, kan öka om man försöker öka kvalitén på referensdata och adressfilen som ska matchas, t.ex. genom att expandera de vanligaste förkortningarna på gatu- och vägnamn.

2.3 Användning av geokodning

Via ett geokodat adressregister kan data baserat på adress presenteras på en karta eller en bildskärm. Denna visualisering förbättrar betraktarens möjligheter att uppfatta och förstå stora datamängder. Den grafiska presentationen ökar betraktarens perceptionsförmåga.

Exempel på vad man kan göra m.h.a. geokodning:

- Man kan skapa "nålkartor" för att visa placering av olika objekt som har adress.

- Man kan förfråga och lokalisera objekt m.h.a. adress.
- Man kan göra "point-in-polygon" analys för att t.ex. hitta alla kunder inom ett distributionsområde.
- Ett företag kan göra marknadsanalyser för sina produkter eller tjänster.
- Räddningstjänsten använder geokodning när de genom att ange en adress bestämmer vilka fordon och personal som ska göra utryckning.
- En skola kan använda geokodning för att matcha elevernas bostadsadresser mot ett gatunät för att t.ex. bestämma skolskjutsar.
- Geokodning används för att t.ex. identifiera riskområden kring miljöfarliga industrier genom att matcha data mot befolkningsstatistik.

Precision Marketing

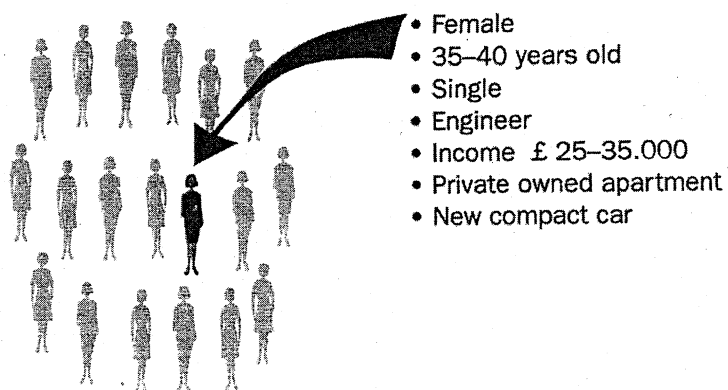


Fig. 2.2
Precision marketing.
M.h.a geokodning kan
man inrikta sin
marknadsföring på
speciella kundtyper
(Burbach, 1994).

2.4 Geokodning i ArcView

Geokodning av adresser i ESRI's GIS-program ArcView (se vidare kapitel 3) är en process som skapar ett tema baserat på de geokodade data som läggs till vyn. Under geokodningsprocessen läser ArcView adresserna och lokaliserar dem i ett referenstema som representerar gatunätet för det område där adresserna finns. Adresserna "ärver" koordinater från referenstemat. ArcView skapar därefter ett geokodat tema baserat på en ArcView shapefil för att lagra alla element och dess attribut från adresstabellen samt koordinat för de adresser som har matchat mot gatunätet. ArcView lägger till temat till vyn och de matchade adresserna representeras av en punktsymbol.

2.4.1 Adressmatchning i ArcView

Vid adressmatchningen tittar ArcView på adresskomponenterna i både adresstabellen och shapefilen med vägnätet. ArcView söker sedan efter adresstandarder för att dela upp adresstabellens textsträng i gatunamn och adressnummer och jämför sedan om adresserna stämmer överens.

2.4.1.1 När matchar adresserna?

I referensdata, shapefilen, har varje vägsegment två attribut som representerar intervallet av udda nummer på segmentet, d.v.s. de nummer som finns på vägens vänstra sida, och två attribut som representerar intervallet av jämna nummer.

Om ett element i adresstabellen har samma gatunamn som ett vägsegment i referensdatafilen och dess nummer är inom något av samma vägsegments nummerintervall, har adressen matchat.

Felaktig, ofullständig eller tvetydig adressinformation reducerar matchningens resultat. Därför kan det krävas manuell inverkan som att justera matchningsparametrarna eller interaktivt editera elementens adresser om de är felaktiga.

2.4.2 Hur ArcView bestämmer adressens koordinater

Om ett element i adresstabellen matchar mot ett vägsegment, placerar ArcView adressen på rätt sida av vägsegmentet beroende på huruvida elementets adressnummer är udda eller jämnt. En punkts placering längs gatan interpoleras fram genom att jämföra adressnumret med gatulänkens adressnummerintervall. Adresserna positioneras med jämna avstånd längs vägsegmentet.

ArcView placerar de udda adressnummerna ett standardavstånd till vänster om vägsegmentet och de jämna nummerna till höger.

Exempel: KAPTENSGATAN 23 matchar mot segmentet KAPTENSGATAN 11 29 12 24 och placeras ungefär två tredjedelar av segmentets längd in på gatans vänstra sida. Uträkningen av punktens koordinater baseras på var det matchade elementet är lokaliserat.

ArcView lägger en procentuell "klämfaktor" till den interpolerade placeringen genom att flytta in alla adresser en procentuell sträcka från ändarna på vägsegmentet. På så sätt förhindras att adresserna vid vägsegmentens ändrar placeras i en vägkorsning.

Offsetavstånd anger hur långt från den faktiska positionen som den matchade adressen ska placeras. Om ett adressnummer matchat mot ett vägsegment kan det vara önskvärt att placera adressen bredvid vägsegmentet för att förenkla visualiseringen av resultatet.

2.5 Geokodning med Avenue

ArcView's möjligheter för geokodning kan även utnyttjas vid programmering av egna script med programspråket Avenue.

Vid geokodningsprocessen matchas en eller flera adresser i en tabell med ett referenstema, ett tema som innehåller adressattribut, t.ex. gatunamn och adressnummer. Elementen i referenstemat är lokaliserade i vyn via koordinater.

För att matcha adresser krävs att referenstemat är matchningsbart. Ett tema är matchningsbart i ArcView om det finns ett MatchSource-objekt och ett geokodningsindex knutet till det. MatchSource skapas genom att man anger vilken typ av adresser, AddressStyle, som används vid geokodningen, t.ex. gatunamn med adressnummer eller postnummer och ort. MatchSource innehåller information om vilka attribut i referenstemat som används som behövs för att lokalisera en adress i referenstemat. Geokodningsindex innehåller information om i vilken ordning som elementen i referenstemat ska sökas vid geokodningsproceduren. Man kan säga att indexet sorterar referenstemat beroende på hur AddressStyle ser ut, för att proceduren ska gå så fort som möjligt.

För varje adress som matchas mot referenstemat delar ArcView upp adressen i separata komponenter, MatchKeys, t.ex. gatunamn och adressnummer. Dessa komponenter jämförs med referenstemat m.h.a. MatchSource. ArcView söker genom referenstemat och använder MatchKey för att hitta matchande kandidater. Matchningskandidaterna sparas i en temporär lista och poängsätts efter hur väl de matchar den uppdelade adressen i MatchKey. Om kandidaten har högre poäng än det förutbestämda gränsvärdet i MatchSource har ArcView funnit ett vägsegment som adressen matchar mot. Kandidaterna i listan rangordnas och den kandidat som har högst poäng, som är större än gränsvärdet, är det vägsegment som adressen placeras vid i vyn med en punktsymbol.

1. Inläsning av adressen som ska matchas.
2. Sätter MatchKey genom att dela upp adressen.
3. Söker genom referenstemat med MatchKey.
4. Poängsätter kandidater, MatchCand, som placeras i en lista, MatchCase.
5. Plockar fram bästa MatchCand från MatchCase.
6. Lokaliserar adressen m.h.a. MatchCand's position.

Fig. 2.3
Geokodningsproceduren
i ArcView.

Exempel: Referenstemat har ett MatchSource objekt knutet till sig, där AddressStyle är angiven som gatunamn_adressnummer. MatchSource innehåller också information om att attributen `l_add_fr`, `l_add_to` (lägsta respektive högsta nummer på vänster sida) och `r_add_fr`, `r_add_to` (lägsta respektive högsta nummer på höger sida) representerar vägsegmentens nummerintervall. Adressen KAPTENSGATAN 23 läses in från adresstabellen. Adressen delas upp i komponenterna KAPTENSGATAN och 23. Dessa komponenter är adressens MatchKey.

ArcView söker genom referenstemat och poängsätter de vägsegment som liknar komponenterna i MatchKey. Dessa kandidater poängsätts och placeras i MatchCase. När alla objekt i referenstemat är genomsökta plockas det vägsegment som har högst poäng i MatchCase fram. Gränsvärdet i MatchSource för att adressen skall matcha är normalt 80. Alla objekt i referenstemat som heter KAPTENSGATAN har fått hög poäng, men det enda objekt med poängen 100 är KAPTENSGATAN 11 29 12 24. Detta vägsegment är det enda objekt som helt matchar komponenterna i MatchKey. Adressen KAPTENSGATAN 23 placeras två tredjedelar av vägsegmentets längd in på gatans vänstra sida.

När geokodning genomförs med Avenue kontrollerar användaren hela processen med adressmatchning. Man kan ange hur känsligt ArcView ska vara med stavningen av adressnamnen, ett lägre värde ger fler matchningskandidater, även om det medför att det ökar risken för felaktig matchning. Fler kandidater gör också att hela geokodningsprocessen blir långsammare eftersom varje kandidat ska poängsättas beroende på hur väl den matchar adressen. Om indata är av hög kvalitet är det att rekommendera att starta geokodningen med ett högre tal för stavningskänsligheten.

Efter att adresserna geokodats och det nya temat med de matchade adresserna lagts till vyn kan en repetering av matchningen göras. En förenklad geokodningsprocess kan då genomföra ommatchning av de adresser som inte matchat. Stavningskänsligheten kan ändras, liksom minstavärdet för att en adress ska vara kandidat eller minstavärdet för att matcha med en kandidat. Förändringar av hur ArcView ska hantera kandidater innebär att

man kan få ett nytt resultat av geokodningen och öka antalet matchade adresser, naturligtvis på bekostnad av matchningens tillförlitlighet.

2.6 Sammanfattning

Geokodning är en GIS-operation som möjliggör sammanlänkning av olika datakällor. Ett register eller databas som baseras på en position (adress, fastighet, postnummer och -ort, region, land) kan i ett GIS-program kopplas samman med geografiska data och på så sätt lokaliseras på "kartan". Detta öppnar i sin tur möjligheter för att genomföra analyser av sitt material. Ett företags eller en organisations databaser kan, när de är kopplade till en geografisk databas i ett GIS, vara till stor hjälp vid planering, marknadsanalyser och presentation.

3 GIS och ArcView

Detta kapitel ger inledningsvis en presentation av geografiska informationssystem, GIS. Det GIS-program som T-Kartor använder i sin verksamhet är ESRI's ArcView, som beskrivs senare i detta kapitel.

3.1 Vad är GIS?

Ett geografiskt informationssystem, GIS, är ett datorbaserat informationssystem med funktioner för insamling, bearbetning, lagring, analys och presentation av geografiska data (Lantmäteriverket, 1993a). Denna definition är vanligt förekommande som beskrivning av geografiska informationssystem.

GIS har många olika användningsområden:

- Dataprocessering: kartproduktion, 3D-visualisering
- Dataanalyssystem: konfliktanalyser, transportoptimering
- Presentationsverktyg
- Förvaltningsverktyg
- Planeringsverktyg
- Navigationssystem

(Bernhardsen, 1994)

Vid en jämförelse med vanliga kartor har GIS en stor fördel i att lagring och presentation av data är separerade. Data kan därför presenteras och visualiseras på olika sätt, t.ex. med en kartbild eller med tabeller med samma data som underlag. En ytterligare fördel är att GIS kan länka samman data från olika databaser med geometriska data. Det som krävs är att data på något sätt är kopplade till en position t.ex. via adress, fastighet, postnummerområde. På så sätt kan olika data jämföras och visualiseras på en kartbild. I ett GIS beskrivs verkligheten med hjälp av digital kartdata, geografiska data som beskriver var objekt är lokaliserade i rummet och attributdata som vanligt består av tabeller med karaktäristika, attribut, vilka kan konverteras till geografiska symboler för att visualiseras tillsammans med andra geometriska data i en kartbild.

Databasen är en mycket viktig beståndsdel i ett GIS. Den lagrar de data som GIS-programmet använder vid modellering, analys och presentation. I databasen lagras data på ett strukturerat och likartat sätt och är lätt att uppdatera med nya eller ändrade data.

3.2 Geografiska data

Geografiska data är dels geometriska data dels attributdata och topologi (se vidare avsnitt 3.2.3). Geometriska data beskriver en företeelse geometriskt, dess läge i förhållande till ett referenssystem. Geometriska data representeras av punkter, linjer och ytor. Attributdata är de icke-geometriska data som ytterligare beskriver ett objekt (t.ex. namn, färg, material, längd, adress) och kan delas in i kvalitativa data och kvantitativa data. Kvalitativa data beskriver objektets typ medan kvantitativa data ger objektet ett värde. Attributdata är nödvändiga för att kunna göra analyser. Ett GIS som saknar databas med attributdata kan bara presentera en kartprodukt med geometriska data. Geometriska data är vanligtvis lagrade separerade från attributdata i olika tabeller i en databas. Unika identifierare används för länkning och dataåtkomst mellan de olika tabellerna i databasen (Bernhardsen, 1994).

Nod-ID	x	y			
1564	10115,00	14119,12	Geometriska data		
1565	10248,80	13997,65			
Länk-ID	nod-ID	nod-ID	vägnamn	Beläggning	hastighet
101	1564	1565	Ågatan	Asfalt	50
102	1565	1841	Ågatan	Asfalt	70
Geometriska data			Attributdata (kursivt)		

Fig. 3.1 Ågatan beskrivs i ett GIS av geometriska data och attributdata. Gatan består av länkar som i sin tur beskrivs av dess koordinatsatta ändpunkter (noder).

Traditionellt delas geometriska data in i två typer, raster och vektor.

3.2.1 Raster

Vid användande av rasterdatamodellen generaliseras verkligheten till likformade celler, vanligtvis i form av kvadrater eller rektanglar men även triangulära eller hexagonala celler förekommer (Bernhardsen, 1994). Den geometriska upplösningen beror på cellernas storlek i rutnätet. De geografiska objekt som ska presenteras i en bild representeras av numeriska värden i rutnätet. En cell (bildpunkt, pixel) är det minsta elementet som kan visualiseras i en bild och verkligheten som faller inom en cell måste därför förenklas till att representeras av endast ett värde. En punkt kan representeras av en cell, en linje av en sekvens av celler (se fig. 3.2) och en area av en samling celler som gränsar till varandra (Worboys, 1995). Varje cell adresseras med dess rad- och kolumnnummer i rutnätet, vanligtvis räknat från bildens övre vänstra hörn.

Eftersom en cell endast kan innehålla ett numeriskt värde lagras oftast olika objekttyper i var sitt rasterlager. I en cell kan det t.ex. finnas en väg och ett vattenskyddsområde. Det finns därför ett rasterlager där cellen har ett värde som anger att det finns en väg av någon typ. Det finns också ett rasterlager som beskriver vattenskyddsområden. Det är vanligt att det finns olika tematiska lager för t.ex. topografi, markanvändning etc.

Rasterdata är främst lämpligt vid datainsamling, utmatning och logisk analys av ytor och höjdmodeller.

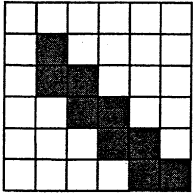


Fig. 3.2
En linje beskriven
med rasterdata.

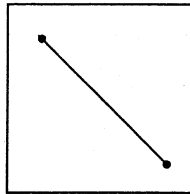


Fig. 3.3
En linje beskriven
med vektordata.

3.2.2 Vektor

Spatiella data lagrade som vektordata baseras på punkter och dess koordinater. En vektor är en rät linje mellan två koordinater på en digital karta. Ett punktobjekt representeras av en punkt i kartan. En linje beskrivs av en vektor mellan två punkter (se fig. 3.3) och en yta beskrivs av dess avgränsningar som representeras av ett antal vektorer.

Lagring av vektordata är inte så utrymmeskrävande som lagring av rasterdata eftersom den enda information som behöver lagras är de punkter som beskriver ett objekts avgränsningar (Worboys, 1995). Det är vanligt att även vektordata lagras i olika tematiska lager för t.ex. infrastruktur, fastighetsindelning osv.

Vektordata är lämpligt då man kräver hög precision och detaljeringsgrad. En ytterligare fördel är att det är enkelt att koppla attributdata till vektordata där objekten kan ha attribut som t.ex. ett namn eller en färg som är kopplad till dess identitet.

3.2.3 Topologi

Topologi utgörs av vissa typer av samband mellan två geometriskt beskrivna objekt. Topologin anger ett visst objekts läge i förhållande till ett annat objekt (t.ex. gränsar till, ligger inom) (Lantmäteriverket, 1993a).

Om en databas innehåller topologi är det möjligt att besvara frågor som:

- Vilka punktobjekt finns inom en yta?
- Vilka vektorer hänger samman?
- Vilka ytor gränsar till varandra?

Topologin är i många fall nödvändig för att kunna utföra analyser, t.ex. vid s.k. overlay-operationer och nätverksanalyser. Nätverksanalyser används t.ex. i navigations-tillämpningar, där det krävs att GIS-programmet har information om vägnätet, vilka vägar som korsar varandra o.s.v. för att kunna beräkna hur ett fordon skall vägledas från en destination till en annan.

3.3 Datakvalitet

3.3.1 Positionsnoggrannhet

På en vanlig karta är positionsnoggrannheten omvänt proportionell till kartskalan, en karta i skala 1:1000 har större positionsnoggrannhet än en i skala 1:100000. Datainfångsten till en digital kartdatabas har vanligtvis gjorts genom att digitalisera en vanlig papperskarta. Därför finns det ett flertal källor för felaktigheter i databasen: mätning, fotogrammetrisk kartframställning, kartritning och digitalisering. Digitala kartdata som skapats direkt från flygfoton eller fältmätning har genomgått färre steg från ursprunglig datainsamling och har

därför vanligtvis större positionsnoggrannhet än data som kommer från digitaliserade kartor (Bernhardsen, 1994).

Konverteringar av data mellan raster och vektor, och vice versa, innebär också förlust av noggrannhet av geometrisk data.

3.3.2 Noggrannhet i attributdata

Attributdata som är dåligt definierade kan innebära att objekt klassificeras felaktigt när skillnaden mellan objekttyperna är oklara.

För rasterdata kan det finnas osäkerheter speciellt vid gränser mellan två ytor av olika typ eftersom de är skapade av en subjektiv bedömning av originaldata av en operatör.

3.3.3 Konsistens av länkning mellan geometriska data och attributdata

Länkningen mellan geometriska data och attribut data är beroende av användningen av identifierare. Uppdateringar av geometriska data och attributdata bör göras samtidigt för att inkonsistens av länkningar inte skall uppstå (Bernhardsen, 1994).

3.3.4 Konsistens av topologi

För att kunna utföra analyser krävs att konsistensen av topologin upprätthålls; ytor som gränisar skall ha gemensamma gränslinjer, ett vägnät skall hänga samman, polygoner skall vara slutna o.s.v. Dålig kvalitet på topologin medför felaktigheter i t.ex. överlagringsoperationer och nätverksanalyser.

3.3.5 Geografisk täckning och upplösning

Den geografiska täckningen kan verifieras med genom att jämföra hur många objekt som finns i databasen jämfört med hur många som finns i verkligheten. Upplösningen kan översättas med hur många observationer som gjorts av verkligheten; cellstorleken i ett raster-GIS, minsta storleken för en polygon i ett vektor-GIS eller antalet mätpunkter som använts.

3.3.6 Aktualitet

Datum för när datainsamling gjorts är också ett mått på kvalitet. Men tiden ökar risken för att nya objekt tillkommit eller att objekts geometri och attribut har förändrats.

3.3.7 Felkällor

Felaktigheter eller kvalitetsbrister på innehållet i en digital kartdatabas kan uppstå i all fasen av databehandling. De felkällor som kan förekomma: (Bernhardsen 1994)

- Ordinära felkällor
 - datainsamling: mätinstrument, flygfotokameror, GPS, satellitsensorer o.s.v.
 - kartframställning, dataeditering: beräkningar, kartritningsinstrument
 - förändringar i verkligheten: tillkomst eller förändringar av objekt
 - begränsad täckning eller upplösning
- Felkällor vid databehandling
 - datainsamling: digitalisering, insamling av attributdata
 - lagring av data: bristande numerisk precision, fel i lagringsmedia

- manipulering av data: raster till vektor, vektor till raster, generalisering, interpolering

- Metodfel

- felaktig metod for datainsamling
- bristande definition av objektclasser för klassificering
- bristande erfarenhet av datainsamlare
- osäkerhet i begränsningslinjer mellan områden

3.4 ArcView

Det GIS-program som T-Kartor använder mycket i sin verksamhet är ArcView GIS. ArcView fungerar som ett komplement till ESRI's produktpaket ARC/INFO. ARC/INFO är skapat för att användas i DOS- eller UNIX-miljö, medan ArcView är skapat för de användare som arbetar i MS-Windows.

För närvarande finns ArcView i version 3. Programmet är främst anpassat för analys och presentation av data.

All aktivitet i ArcView sker i ett projekt. Projektet innehåller information om de dokument och s.k. scripts (se nedan) du arbetar med. Dokumenten är de medel som används för att analysera, bearbeta och presentera data; vyer, tabeller och diagram (Environmental Systems Research Institute, 1996). Scripts är speciella program som skrivs i programmeringspråket Avenue för att utföra speciella operationer i ArcView (se vidare avsnitt 3.4.3).

I projektet finns också information om vilket gränssnitt som används för att interagera med dokumenten i projektet. Det kan vara information om vilka kontroller, knappar och menyer som används i det aktuella projektet.

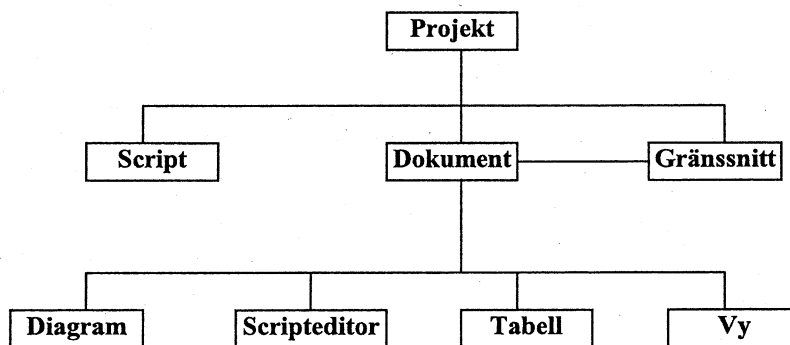


Fig. 3.4 Uppbyggnaden av ett projekt i ArcView.

3.4.1 Arbeta med ArcView

ArcView erbjuder alla sedvanliga möjligheter att arbeta med digitala kartdata, t.ex. zoomning, etikettering av kartobjekt, skapa speciella utskrifter, ändra färger och symboler,

mäta avstånd, söka objekt m.m. Man kan genom att klicka på ett objekt få upp all information som finns om detta objekt i attributdatatabeller.

Geometriska data visualiseras som teman i en vy (kartbild). Varje tema i en vy är baserat på informationen i en tabell. Det är därför möjligt att specificera vilka teman som ska visas i vyn såvida de data som finns är uppdelade i olika tabeller för olika objekttyper.

Man kan göra avancerade urval av objekt i en tabell och få de objekt som uppfyller kriterierna markerade i vyn.

3.4.2 Geometriska data i ArcView

ArcView använder ett eget format för geometriska data, benämnt shapeformat. Till varje shapefil kopplas en databastabell som innehåller attributdata. Programmet kan också använda ARC/INFO's format för lagring av geometriska data.

Om användaren vill använda en databas av annat format innehållande objekt med lägesinformation kan den importeras i ArcView som en tabell. ArcView kan importera tabeller som är lagrade i dBASE-format (dbf) eller som kommaseparerade textfiler. Dessa tabeller kan sedan användas som de är eller konverteras till shapeformat för ökade användningsmöjligheter som t.ex. nätverksanalyser (Environmental Systems Research Institute, 1996).

3.4.3 Avenue

ArcView's objektorienterade programmeringspråk och utvecklingsverktyg heter Avenue. Avenue är fullt integrerat med ArcView och plattformsoberoende. Avenue's användningsområden är många. Man kan anpassa hur man vill använda ArcView, ändra dess utseende, utveckla speciella operationer eller applikationer som använder och integreras med ArcView's grafiska gränssnitt (Environmental Systems Research Institute, 1994b).

3.5 Objektorientering

3.5.1 Introduktion

En objektorienterad modellering bygger på abstraktion av verkligheten där de karaktäristika som skiljer objekt av en objekttyp från objekt av alla andra objekttyper definieras och utgör beskrivningen av objekten i denna objekttyp (Booch, 1991, s.39). Modellen (dataprogrammet) är på så sätt en formaliserad beskrivning av en aspekt av verkligheten. Modellen består av beskrivningar av de objekt och förhållanden mellan objekt som förekommer (Holm, 1992). Ett objekt har en unik identitet i modellen, har attribut, kan utträta operationer och befinna sig i vissa tillstånd (fig. 3.5).

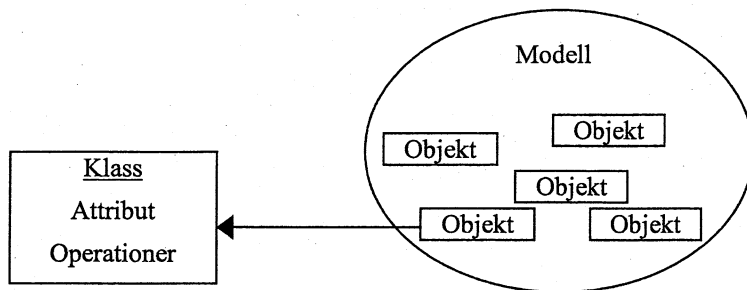


Fig. 3.5 Attribut och operationer definierar objektens egenskaper.

Attributen och operationerna som utgör objektens egenskaper beskrivs i en klass. De operationer som ett objekt kan utföra beskrivs av rutiner i klassen för den objekttypen. Dessa rutiner kan sedan anropas för att utföra en operation på ett objekt. Endast objekt av den objekttyp som en klass representerar kan använda sig av de operationer som finns beskrivna för klassen.

Exempel:

Cell_1.Rita är ett anrop av rutinen *Rita* som utför en operation, t.ex. att rita en kvadrat, med objektet *Cell_1*. *Cell_1* är ett objekt av klassen *Kvadrat*. I *Kvadrat* finns alla operationer (rutiner) angivna som objekten av den klassen kan utföra. Objektet *Cell_1* har attributen *X*, *Y* och *Sida* som rutinen *Rita* använder för att rita *Cell_1*. *Cell_2* är ett annat objekt av klassen *Kvadrat* som har egna värden för attributen *X*, *Y* och *Sida*. Ett objekts attribut kan förändras genom att objektets operationer används och kan på så sätt befinna sig i olika tillstånd men objektets identitet förändras aldrig (Booch, 1991, s.85).

3.5.2 Objektorientering i Avenue

Ett objekt i ArcView kan vara ett element som bygger upp programmets gränssnitt som t.ex. en knapp eller ett fönster. Det kan vara ett element i ett projekt som t.ex. en vy, en tabell eller en utskriftslayout. Ett objekt kan också vara ett grundelement i Avenue som ett tal, datum eller en textsträng eller också vara ett grafiskt element som en punkt, linje eller yta (Environmental Systems Research Institute, 1994b).

Varje klass i ArcView's klasshierarki definierar de gemensamma egenskaperna (attribut och operationer) för varje objekt som tillhör den klassen.

Med Avenue kan användaren skapa, ändra eller hämta information om objekt genom att anropa operationer som finns i objektklassen.

Exempel:

Man kan flytta ett grafiskt objekt genom att anropa objektets operation *Flytta* med parametrar för hur stor förflyttningen skall vara. Objektet flyttar då sig själv ett avstånd i x- och y-led.

Man kan få produkten av två tal genom att anropa en operation för multiplikation.

Vid programmering i Avenue liksom i andra objektorienterade programspråk används s.k. punktnotation. Ett uttryck i Avenue har formen
variabel = objekt.egenskap

Egenskapen för objektet som anropas kan vara ett attribut eller en operation. När det sänds ett anrop till ett objekt i Avenue resulterar anropet i att ett nytt objekt returneras som i detta fall tilldelas variabeln.

Exempel:

Programraden `vyNamn = minVy.GetName` returnerar ett strängobjekt med namnet på objektet `minVy` som tilldelas variabeln `vyNamn`.

Uttrycket `mittTal = 2 + 5` returnerar ett talobjekt som tilldelas variabeln `mittTal`.

3.5.3 Skapa applikationer i ArcView m.h.a. Avenue

ArcView erbjuder stora möjligheter att utveckla egna applikationer genom att använda Avenue. Med speciella Avenue-scripts kan användaren utföra särskilda operationer för att använda ArcView's möjligheter på bästa sätt. Genom att utnyttja andra ArcView-applikationer kan användarens Avenue-script bli en del av ett större system.

3.5.4 Vad är ett script?

Ett script är en komponent i ett ArcView-projekt som innehåller Avenue-kod. Precis som makron, procedurer eller scripts i andra programmeringsspråk, är syftet med Avenue-scripts att automatisera operationer och utöka ArcView's möjligheter. Många delar av ArcView är egentligen uppbyggda av en samling interna scripts, t.ex. är alla knappar associerade med ett script som utför en uppgift (Environmental Systems Research Institute, 1994b).

Man skapar ett script i ArcView's Script Editor-fönster eller i vilken texteditor som helst. Innan scriptet kan köras måste ArcView kompilera koden. Under kompileringen kontrollerar ArcView att syntaxen, ordföljden i en programrad, är riktig. Kompileringen kontrollerar också att koden på varje programrad kan förstås av ArcView.

3.5.5 Anpassning

Med Avenue kan användaren manipulera alla objekt som finns i ArcView. Scripts kan skrivas som utför exakt de operationer som krävs av användarens applikation.

Användargränssnittet kan modifieras för att anpassas speciellt för användarens behov. Projektets utseendet kan förändras, kontroller kan läggas till eller tas bort för att passa slutanvändarens önskemål.

Applikationer kan skapas som programtillägg vilka sedan kan kopplas samman med vilket projekt som helst. Ett programtillägg är oberoende av projektfilen vilket gör att applikationens funktionalitet och utökade möjligheter kan utnyttjas av flera olika projekt (Environmental Systems Research Institute, 1994b).

Applikationens alla komponenter som t.ex. scripts och dokument lagras i en speciell ODB-fil (object data base).

4 GIS-databasen GeoTätort

4.1 Databasbeskrivning

GeoTätort är en geografisk databas för användning inom GIS och kartografi. Detaljnivån är främst anpassad för att användas i skalområden mellan 1:5 000 - 1:50 000 (Bjöersdorff, 1994). Det ursprungliga syftet med GeoTätort var att utifrån denna databas producera kartor till telefonkatalogen. Detta är fortfarande T-Kartors främsta användning av GeoTätort.

1997 påbörjades en revidering av GeoTätort, vilken beräknas pågå fram till sekelskiftet

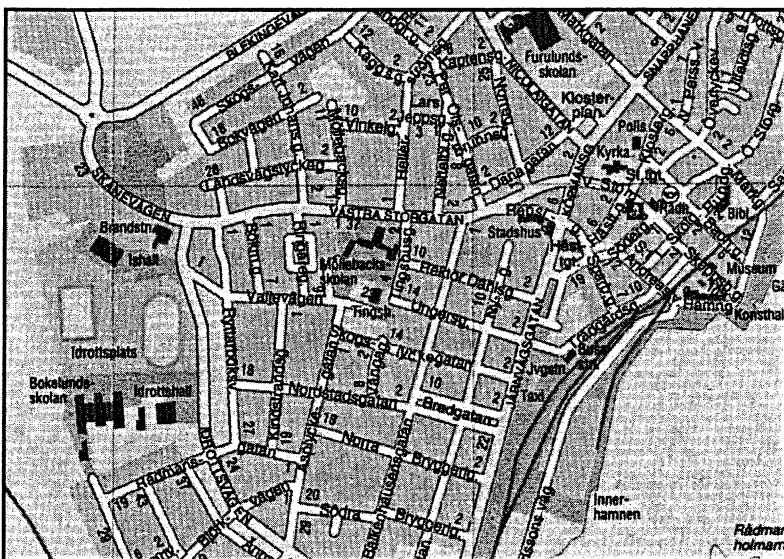


Fig. 4.1
Karta över
Sölvesborg ur
Röda Sidorna.

(T-Kartor Sweden AB, 1997b). Revideringen utförs för att få förbättrad och jämnare kvalitet på databasen. M.h.a. flygbilder justeras de geografiska positionerna för vägar och byggnader. Information samlas också in "på fältet" för att kontrollera vägnamn och tillföra databasen nyttillkomna attribut. Även uppgifter om trafikrestriktioner samlas in för att utöka basens användning för transporttillämpningar.

4.1.1 Geografisk täckning

GeoTätort finns för mer än 300 svenska orter. Med några få undantag innefattar detta alla orter med fler än 5 000 invånare, samt de orter som är huvudort i någon kommun. Denna målsättning är idag i det närmaste helt uppfyllt (Agnarsson, Vingå, Kallur, Johnsson, 1997).

Detta innebär att GeoTätort täcker ca 65 % av landets invånare och 90 % av Sveriges industri. GeoTätort används bl.a. för kartframställningen till landets 28 telefonkatalogsdelar, Röda Sidorna. Den geografiska utbredningen för GeoTätort är i huvudsak den som redovisas på kartorna i telefonkatalogen. Hur mycket av respektive ort som täcks av dagens databas varierar något (se vidare 4.1.5.2).

4.1.2 Struktur

Databasen innehåller linjer, ytor samt för kartografi inom speciella skalintervall även grafiska texter. Till varje objekttyp finns dessutom beskrivande data som exempelvis klassindelning m.m. (se vidare 4.1.3.2).

Vägnätet i GeoTätort är uppbyggt enligt "mittlinjesprincipen". För kartografitillämpningar såväl som snyggare presentation, finns även generaliserade vägkantslinjer att tillgå.

4.1.3 Innehåll

Innehållet i dagens GeoTätort är starkt präglad av det ursprungliga syftet med databasen; att producera telefonkatalogskartor. Detta gäller såväl geografisk utbredning som innehåll och grafik. Den största utvecklingen på senare år har skett på databasens attribut innehåll, där en del nya attribut tillkommit för att möjliggöra fler GIS-tillämpningar.

4.1.3.1 Geografiska objekt

Detaljnivån för vilka objekt som ingår i GeoTätort härrör sig från de kartografiska behoven vid produktionen av kartor i skalområden mellan 1:10 000 och 1:25 000.

Strukturen i databasen skiljer sig dock från den tryckta grafiska. De mest centrala objekten i databasen är de linjer som representerar vägar och gator. Databasen är uppbyggd efter "vägmittlinjesprincipen", där varje väg eller gata representeras av en linje som utgör vägens mitt. Övriga linjer i databasen är övriga transportleder (järnvägar etc) samt konturlinjer för olika ytklasser (bebyggelse, skog, vatten etc.).

Även objektinnehållet har sina historiska arv. Exempelvis är de enda byggnader som finns markerade offentliga byggnader.

4.1.3.2 Typkodning

Dagens typkodning är till största delen baserad på samma information som fanns i det ursprungliga kartografiska utseendet. De olika typer av objekt som fanns beskrivna i telefonkatalogen har legat till grund för databasens typkodning.

Det finns dock ett kontinuerligt arbete med att göra om typkodningen för en del företeelser.

Adressattributen är den grupp av attribut som utvecklats mest under den senaste tiden. Dessa attribut möjliggör sökning av adress med hjälp av GIS (s.k. geokodning). Varje enskilt vägmittlinjesegment förses med uppgifter om gatans namn, samt första och sista adressnummer på vänster resp. höger sida. Ett inmatningsarbete som hittills utförts manuellt med vanliga "papperskartor" som underlag.

4.1.3.3 Linjeattribut

Till varje enskilt linjesegment finns ett antal beskrivande data kopplat.

Namn Namn på linjesegment, t.ex. gatunamn.

Typer Olika klassificeringar av vägar, rälstrafikleder, konturlinjer för bebyggelse och vattenområden o.s.v.

Adressattribut som i GIS-system möjliggör geokodning på adressnivå. Adressattributen används också för olika transporttillämpningar av GeoTätort.

L add fr Vägsegmentets första adressnummer vänster sida

L add to	Vägsegmentets sista adressnummer vänster sida
R add fr	Vägsegmentets första adressnummer höger sida
R add to	Vägsegmentets sista adressnummer höger sida
Vis	Anger synlig resp osynlig linje för en grafisk standard kartografi. 1 = synlig, 0 = osynlig
Längd	Längd i meter.

4.1.3.4 Ytattribut

Area	Area i kvadratmeter.
Perimeter	Omkrets i meter.
Typer	Bebyggelse, vattenområden, vegetationsområden.
Namn	Namn på offentlig byggnad.

4.1.3.5 Grafik

Till produkten GeoTätort hör dessutom ett antal grafiska "företeelser" som inte är några objekt i egentlig mening. Detta gäller framförallt texter, symboler och kantlinjer på vägar. Det som skiljer dessa "företeelser" från annat innehåll som t.ex. vägmittlinjer, är att de är skalberoende och därmed begränsade i sin GIS-användning.

Dessa grafiska "företeelser" är i hög grad knutna till specifika produktuttag ur databasen, exempelvis "Röda sidor-kartan", och kan knappast ses som en del av en generell geografisk databas.

Idag är enbart gatunamn och offentlig byggnad sökbara i GeoTätort. Denna sökning görs dock via attribut till dessa objekt. Även "textobjekten" bör lagras som attribut till de geografiska objekt som de refererar till för att vara sökbara i en GIS-programvara. En relation mellan grafisk text och motsvarande objekt bör skapas. Detta medför dessutom enklare revidering av texterna och säkerställer att stavningen är riktig.

Stavningen av alla texter ska kontrolleras mot databaser som antas ha objektens korrekta stavning.

4.1.4 Databasstruktur

Tidigare fanns GeoTätort lagrat i ArcInfo-format för uppdatering och underhåll. Varje ort representerades m.a.o. av en egen isolerad databas.

För varje ort fanns en egen databas, där uppdateringar och revideringar gjordes. Dessutom fanns ett antal produktspecifika (tematiska) databaser sparade för varje ort. Exempelvis för "Röda sidornas" och "Din Dels" olika kartutseende.

Denna databasstruktur har i takt med GeoTätorts ökade användning för annat än telefonkatalogskartor blivit ohållbar.

Dagens GeoTätort är uppbyggt enligt en datamodell med master-/produkt-databaser. Samtliga tätorter finns lagrade i samma referenssystem i en masterdatabas, där objekten kan lagras och uppdateras på ett kontrollerat sätt. Produktspecifika data lagras i produkt-databaser. Kraven på datamodellen är höga när databasens innehåll ska användas både för kartografisk produktion och geografisk analys. Kartografisk produktion kräver en databas som är mer eller mindre anpassad efter innehållet och skalan på kartprodukten. Data

i en kartografiskt produktdatabas kan ha genomgått olika steg av generalisering för att förbättra den visuella presentationen, t.ex. förenkling, förflyttning, aggregering eller kanske utelämnning av kartografiska objekt. Geografisk analys kräver å sin sida data av hög geometrisk kvalitet (Oxenstierna, 1997). Masterdatabasen är den geografiskt korrekta databasen som används för geografisk analys och där görs alla revideringar av databasens innehåll. I produktdatabaserna utförs all generalisering och kartografisk bearbetning.

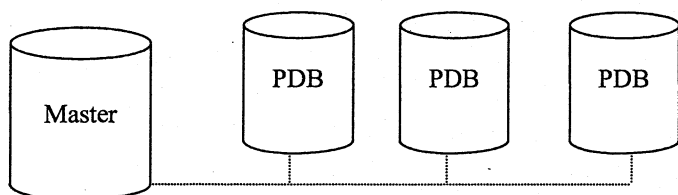


Fig 4.2
I produktdatabaser lagras produktspecifika kopior av objekt i masterdatabasen.

Denna lösning är också ekonomiskt fördelaktig eftersom ändringar i masterdatabasen reflekteras i produktdatabaserna vilket gör att problemen med uppdatering av representationer av samma data i olika databaser har reducerats. Med master-/produktdatabasmodellen är det möjligt att hålla geometriskt riktiga basdata och kartografiska data separerade men ändå relaterade till varandra (Carlsson, Johnsson, 1997). Alla generaliserade data är länkade till dess ursprungsdata i masterdatabasen via unika identifierare vilka gör det möjligt att uppdatera en produktdata när objekt har ändrats i masterdatabasen.

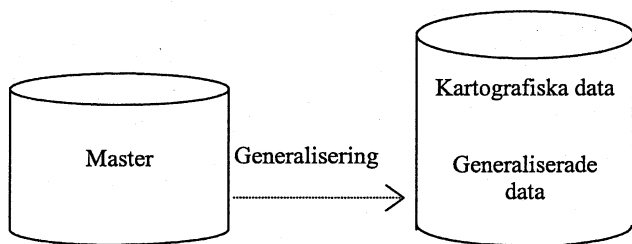


Fig 4.3
Förutom generaliserade data, lagras även kartografiska data som är speciella för

Genom att GeoTätort lagras i en storskalig masterdatabas kan vägnätet mellan orterna byggas ut, vilket är av stor betydelse för t.ex. navigationstillämpningar (se avsnitt 4.2.3 nedan).

4.1.5 Kvalitet

Eftersom innehållet i GeoTätort är hämtat från så många datakällor är det vanskligt att ange ett mått på databasens kvalitet. Beaktat denna problematik och vissa antaganden (se avsnitt 4.1.5.1), är den enda säkra utsagan att det förekommer lokala geometriska fel som är betydligt större än de som gäller GeoTätort generellt. Dessa lokala fel härstammar från uppdateringar gjorda från kommunala underlag. De av kommunerna insända underlagen varierar mycket i kvalitet, allt från "frihandskomplettering av katalogsida", till väl markerade hänvisningar på storskalig karta.

4.1.5.1 Lägesnoggrannhet

Den geometriska kvaliteten i GeoTätort varierar kraftigt. Det enda befintliga kvalitetsmättet på lägesnoggrannheten för varje ords databas som finns att tillgå är resultatet från koordinattransformationerna. Resultatet anger hur stor avvikelser var i meter på den punkt som stämde sämst. Detta säger m.a.o. inget om lokala avvikelser inom respektive databas.

Koordinattransformationer till Rikets Nät, RT90, har gjorts med samtliga tätorters databaser. Av medelfelen vid koordinattransformationerna kan man utläsa att de geometriska felen är måttliga och relativt jämnt fördelade i små och medelstora orter. De större orterna däremot, speciellt Stockholm, Göteborg och Malmö kan lokalt ha stora geometriska fel.

Det är svårt att ange ett generellt noggrannhetsmått på GeoTätort av idag, då de lokala skillnaderna är stora. Man kan dock säga att små orter håller relativt sett jämnare kvalitet än de större städerna, och att medelfelet vid transformationen till Rikets Nät där ligger på ca 5-10 m, samt att grova avvikelser är relativt sparsamt förekommande. Detta gör det rimligt att anta detta som en grund för ett generellt kvalitetsmått med reservation för de lokala fel som kan ha uppkommit i samband med uppdateringar i databaserna. De uppdateringar som gjorts är generellt sett av sämre kvalitet än originaldata, framför allt beroende på brister i underlagsmaterialet. I större orter, framför allt Stockholm, Göteborg och Malmö, varierar medelfelet mellan 5 och 30 m, beroende på stadsdel. Grova fel är avsevärt vanligare i databasen över dessa städer, och kan uppgå till 100-150 m (Agnarsson, Vingå, Kallur, Johnsson, 1997).

Observera att detta är ett kvalitetsmått relaterat till hur väl basen överensstämmer med Lantmäteriverkets "Ekonomiska karta", vilken naturligtvis också innehar olika mått av kvalitetsbrister. Transformationsfelet i meter finns dokumenterat för respektive ort.

4.1.5.2 Geografisk täckning

En vanlig reaktion från användare av GeoTätort vare sig det gäller den digitala produkten eller de tryckta kartorna i olika publikationer handlar om *var* kartkanten går, t.ex:

- Vi bor/vårt företag ligger precis utanför kartan.
- Varför täcker ni inte in hela orten?
- Varför finns det ingen GeoTätort över vårt område?
- Sidindelningen av katalogen borde göras om, med tanke på alla nya bostadsområden.

Det förefaller orimligt att täcka in hela Sverige med den detaljnivå som finns i GeoTätort. Frågan är var gränsen ska gå, vilka områden skall täckas in?

Ett tydligt arv från kartproduktionssyftet är GeoTätorts fyrkantiga utbredning. De största orterna täcks oftast in ganska väl tack vare att man från början ändå varit tvungna att ha många sidor för att kunna redovisa ortens utbredning på telefonkatalogens kartsidor. Även de allra minsta orterna täcks in bra då de med lätthet har ritats på en eller en halv sida. På de mellanstora orterna är däremot täckningsområdet för GeoTätort mer ojämnt. Vissa orter täcks in väl medan andra kan vara väldigt tufft beskurna, där endast de centrala delarna finns med. Detta varierar dessutom en del beroende på landsände, men det är svårt att hitta något generellt mönster (Agnarsson, Vingå, Kallur, Johnsson, 1997).

Den något ojämna täckningen av orter kan betraktas som en kvalitetsbrist i nuvarande GeoTätort. Alternativ till begränsning som diskuterats för en eventuell utbyggnad av GeoTätort, är SCB's tätortsavgränsningar, en definition som konkurrenten "Tätort 2000" (produceras av Metria) tillämpar. Tätortsavgränsningarna är definierade ur

befolkningsstatistiskt syfte och bygger på avstånd mellan bebodda hus, (med hänsyn tagen till regionala avvikelser). Inte heller denna begränsning överensstämmer alltid med den folkliga uppfattningen om var gränsen för en tätort går. Men den sammanfaller trots allt oftare med en mer naturlig gräns, än den fyrkantiga sidan gör. GeoTätorts utbredning jämfört med SCB's tätortsavgränsningar varierar, ibland är den större och ibland mindre (Agnarsson, Vingå, Kallur, Johnsson, 1997).

4.1.5.3 Topologikonsistens

Ett omfattande arbete har gjorts för att säkerställa att kvalitén är god på detta område. Databasen kan sägas vara helt utan brister i topologi efter de databearbetningar som genomförts.

4.1.5.4 Kvalitet på attributdata

Attributdata samlas kontinuerligt in för att säkerställa att attributdata finns tillgänglig för de områden som motsvarar den geometriska täckningen. Kopplingen mellan attributdata och geometriska uppdateras samtidigt.

4.1.5.5 Koordinatsystem

Dagens uppdaterings- och lagringsarbete gjordes tidigare i ett för varje tätort lokalt koordinatsystem, medan transformation till koordinater i Rikets Nät, RT90, gjordes vid uttagstillfället. Detta förfarande tillämpades för att T-Kartor ville behålla sidindelningen av telefonkatalogskartorna. Vid övergången till en masterdatabas med alla tätorter krävdes att alla koordinater lagrades i samma koordinatsystem. Samtliga lokala koordinater har därför transformerats till koordinater i RT90 (Agnarsson, Vingå, Kallur, Johnsson, 1997).

4.1.5.6 Metoder för transformation

Helmert-transformation (fig. 4.4) i planet innebär att en punktmängd i ett koordinatsystem genom skalförändring, parallellförflyttning och vridning passas in ett annat koordinatsystem. Om koordinatsystemen inte ligger i samma projektionssystem (RT90 är ett projektionssystem med medelmeridianen 2,5 gon väst) måste först s.k. överräkning genomföras. Koordinaterna i ett projektionssystem överräknas till ett annat projektionssystem via referensellipsoidens latitud- och longitudvärden.

Skalförändringen är vid Helmert-transformation lika i båda riktningarna vilket gör att punktmängden behåller sin form vid transformationen (Galvenius, 1985).

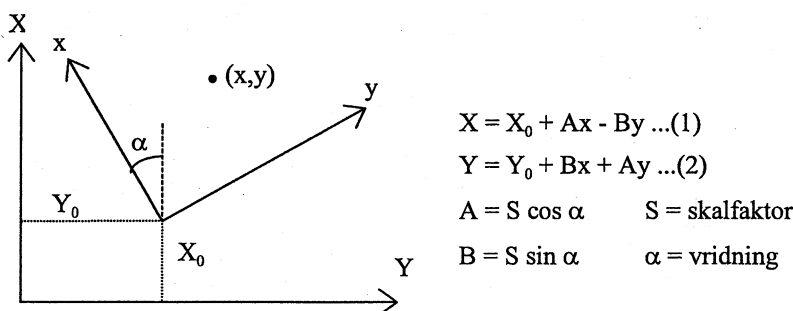


Fig. 4.4 Helmerttransformation

Sambanden (1) och (2) i figuren ovan innehåller fyra obekanta och kräver för lösbarhet minst två punkter med koordinater i båda systemen. Minst fyra punkter rekommenderas dock att användas i praktiken.

De transformationspunkter som använts vid transformation av GeoTätort till Rikets Nät, har mätts in manuellt från "Ekonomiska kartan". Det vanligaste antalet referenspunkter vid transformationen är fyra men i storstäderna har fler punkter använts.

Transformationspunkterna placeras så långt ut i "hörnena" som möjligt, för att undvika extrapolering (Agnarsson, Vingå, Kallur, Johnsson, 1997).

4.1.5.7 Aktualitet

Uppdatering av databasen utgör en viktig aspekt på kvaliteten. GeoTätort uppdateras med ganska god frekvens, varje år i storstadsområdena, vartannat år i övriga städer och vart tredje år för övriga mindre orter. Däremot är aktualiteten bara angiven på denna ganska höga nivå och inte på objektnivå, vilket får ses som en begränsning (se vidare 4.1.6). Så länge Röda Sidorna är den dominerande användaren av GeoTätort ter sig en mer frekvent uppdatering än telefonkatalogens utgåvor ganska orimligt. En förbättring kan vara att alla orter uppdateras varje år (Agnarsson, Vingå, Kallur, Johnsson, 1997).

4.1.6 Metadata

Eftersom GeoTätort ska säljas till externa kunder är det av stor vikt att kvaliteten på databasen kan mätas. Någon metadatabas för GeoTätort finns inte idag men kommer att byggas upp som antingen en del av databasen eller som ett intilliggande system. Man vill kunna redovisa när varje objekt lagts till databasen, från vilket underlagsmaterial och av vem samt vilka uppdateringar av objekten som gjorts (Agnarsson, Vingå, Kallur, Johnsson, 1997).

4.2 Tillämpningar och användningar av GeoTätort

Under denna rubrik beskrivs hur efterfrågan för GeoTätort ser ut och vilka krav olika typer av användare ställer på GeoTätort. Avsnittet avser att peka på de olika användningsområdenas skillnader i krav på databasen.

4.2.1 Kartografi

Kartografianvändningen representerar den största användningen av GeoTätort, (t.ex. kartframställning till telefonkatalogen). Efter att GeoTätort lyftes in i GIS-miljön har den moderniserade kartframställningsprocessen möjliggjort produktion av olika typer av kartografiska produkter ur en och samma databas. Exempelvis för olika skalor och olika tematiskt innehåll.

Ett övergripande krav på GeoTätort ur kartografisk synvinkel, är att databasen på ett så flexibelt sätt som möjligt skall kunna användas till att producera olika typer av kartor. Detta omfattar även digital kartografi. Att presentera kartor på skärm jämfört på papper innebär nämligen både nya krav och möjligheter ur kartografisk synvinkel (Agnarsson, Vingå, Kallur, Johnsson, 1997).

Exempel på produkter kan förutom de redan nämnda telefonkatalogskartorna i "Röda Sidorna" och "Din Del" vara kartor för Internetanvändning eller CD-atlaser. Dessa är båda exempel på produkter från det område som benämns digital kartografi (T-Kartor Sweden AB, 1997a).

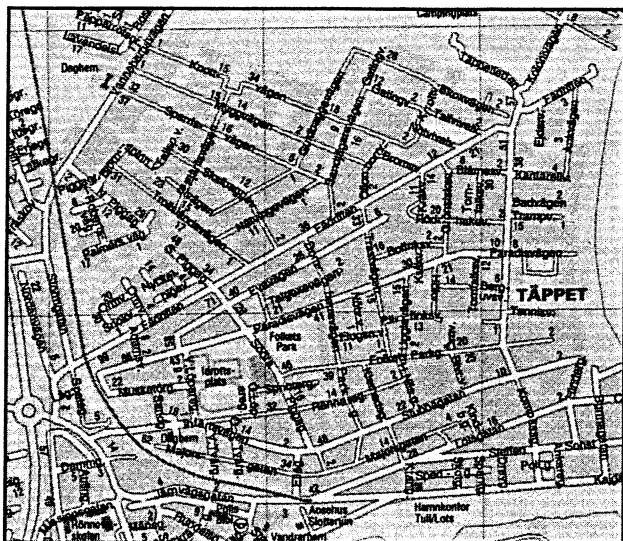


Fig. 4.5
Karta över Åhus
ur Din Del.

4.2.2 Marknadsanalys

Redan i dag finns användare av GeoTätort som kan hänföras till användartypen för marknadsanalys. De två centrala behoven hos denna typ av användare är att kunna visualisera statistik på en bakgrundskarta, men kanske framför allt att utnyttja GeoTätorts geokodningsmöjligheter (se kapitel 2 om geokodning). Eftersom GeoTätort är en databas, kan den samköras med andra databaser, t.ex adressbaserna till Telekatalogen, SCB:s befolkningsdatabaser, eller egna register. Användaren har kanske ett kund- eller medlemsregister innehållande adresser som denne med geokodning kan knyta till GeoTätort och få utplacerade på kartan. Resultatet kan sedan grupperas, analyseras eller visualiseras mot ytterligare fler databaser, så länge de har en geografisk referens (Camarata, 1994). Resultatet kan användas som underlag för analyser av typen:

- Var är det bäst att etablera nya butiker?
- Hur optimeras säljresurserna?

4.2.3 Transport

Även inom transportområdet finns idag potentiella användare av GeoTätort för olika digitala tillämpningar. På grund av de starkt varierande behovskraven är detta avsnitt för transporttillämpningarna uppdelat i ett antal undertillämpningar som positionering, ruttplanering och navigation. GeoTätort saknar idag trafikinformation som krävs för ruttplanering och navigation. En utbyggnad av databasens attribut planeras dock för att möjliggöra dessa tillämpningar. Framkomlighetsbegränsningar som enkelriktning, avspärningar, planskilda korsningar och juridiska körrestriktioner skall redovisas (Agnarsson, Vingå, Kallur, Johnsson, 1997).

4.2.3.1 Positionering

Positionering innebär att användaren, vanligen med hjälp av Global Positioning System, GPS, lokaliserar var någon eller något befinner sig i realtid (ofta ett fordon eller hel fordonssflotta). Positionen ritas upp på kartbild som tolkas visuellt av trafikledningscentral eller av fordonsförare. En koppling till trafikinformation kan dessutom ge besked om bästa och snabbaste vägen med hänsyn till vägarbeten, trafikstockningar mm. Databasen

GeoTätort tillsammans med GPS kan för taxi, budbilsföretag, vårdtransportföretag, vaktbolag, färdtjänst m.fl. innebära nya möjligheter till effektivare transporter och ökad säkerhet.

Exempel på användare idag är SOS alarmering som med hjälp av denna positioneringsteknik på ledningscentralen har kontroll på var utryckningsfordon för tillfället är lokaliserade, om de är lediga osv. På så sätt kan alltid den bil som är närmast en olycksplats larmas.

Denna transporttillämpning används också av enklare typer av navigeringssystem. M.a.o de navigeringssystem som inte aktivt guidar föraren med körinstruktioner, utan bara anger *var du är* och eventuellt *vart du ska*.

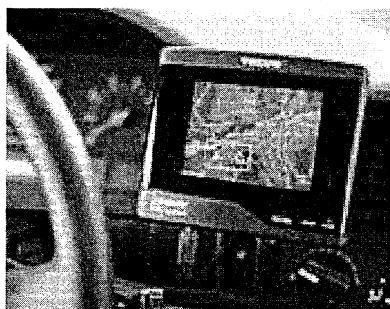


Fig 4.6
Volvos positionerings-
system Dynaguide

4.2.3.2 Ruttplanering

Under denna kategori finns tillämpningar för att svara på frågor av typen; *Hur tar jag mig lämpligast från A till B?* Den här typen av användare finns inte idag eftersom GeoTätort ännu inte innehåller nödvändiga trafikrestriktioner som enkelriktning och svängningsförbud etc. Intresset för detta framtida användningsområde är dock stort och möjligheterna för ruttplanering, eller s.k. nätverksanalys, finns redan idag i GIS-program på marknaden (Agnarsson, Vingå, Kallur, Johnsson, 1997).

4.2.3.3 Navigation

Med navigation avses de tillämpningar som används av avancerade system i fordonet som kontinuerligt ger föraren instruktioner av typen; *sväng höger i nästa korsning, fortsätt rakt fram, etc.*

Redan idag används avancerade navigationssystem i fordonstrafiken. Såväl biltillverkare som elektronikföretag är aktiva på denna nya marknad. Nationella vägbatabaser är under uppbyggnad i ett flertal länder för användning i navigationssystem.

Kraven som ställs för denna typen av tillämpningar är summan av de krav som ställs för både positionering och ruttplanering plus lite till. Kvalitetskraven kan dock i många fall vara systemspecifika.

Det är framför allt intressant att hålla en hög geometrisk noggrannhet på väg- och gatunätet, då det är detta som är av intresse för navigering och positionering. Den siffra som nämnts vid diskussioner om noggrannhet vid fordonsnavigering är att den genomsnittliga lägesnoggrannheten för vägar i tätorter ska vara 5 meter. Maximal avvikelse får vara 10 m. Man kan ifrågasätta om detta noggrannhetskrav är rimligt. För navigeringsapplikationer måste det väsentliga vara att man svänger på rätt gata. Krävs det verkligen 5 m noggrannhet för detta (Agnarsson, Vingå, Kallur, Johnsson, 1997)?

En annan väsentlig fråga är att fordonet grafiskt ligger på vägen vid kombinerad

presentation av fordonsposition och vägdata. Vid grafisk presentation lägger man ofta på en kartografisk vägbuffer, som gör att vägen på kartan blir mycket bredare, upp till ca 100 m, än i verkligheten. 5 m lägesnoggrannhet är alltså ett onödigt högt krav för integrerad presentation (Agnarsson, Vingå, Kallur, Johnsson, 1997).

5 Problemlösning

5.1 Syfte

För att T-Kartor fullt ut ska kunna använda GeoTätort i de huvudsakliga tillämpningsområdena kartografi, marknadsanalys och transport krävs att databasen innehåller så riktiga uppgifter som möjligt.

Vägsegmentens adressattribut möjliggör sökning av adress med hjälp av GIS-programmens möjligheter till adressmatchning. För att kunna utföra adressmatchningen krävs att varje enskilt vägmittlinjesegment är försett med uppgifter om gatans namn, samt första och sista adressnummer på vänster respektive höger sida. Inmatningen av dessa uppgifter har hittills utförts manuellt med vanliga kartor som underlag.

För att utnyttja de adressregister som redan finns vill T-Kartor per automatik kunna överföra information från adressregistret till GeoTätort. För att förbättra informationen i GeoTätort önskar man:

1. att från ett adressregister med punktdata föra över information om adressnummer till de attribut som representerar vägsegmentens nummerintervall i GeoTätort.
2. att finna en metod för att automatiskt identifiera och åtgärda de felaktigheter som kan uppdagas vid jämförelse med adressregistret, som t.ex. felstavningar av gatunamn.

Uppgiften är att lösa problemen med informationsöverföring och databashantering och att implementera lösningen i en ArcView-applikation i programspråket Avenue.

5.1.1 Problembeskrivning

T-Kartor använder sin databas GeoTätort bl.a. för geokodning av databaser där informationen är kopplad till gatuadresser. Geokodning är en process för att koppla ihop en geokod (t.ex. adress) med koordinater. Eftersom objekten i GeoTätort är försedda med adressnummeruppgifter kan en adress geokodas till ett vägsegment med hjälp av adressmatchning. Geokodning är närmare beskrivet i kapitel 2. Vägsegmentens adressattribut möjliggör sökning av adress med hjälp av GIS-programmens möjligheter till geokodning med hjälp av adressmatchning. För att kunna utföra adressmatchningen krävs

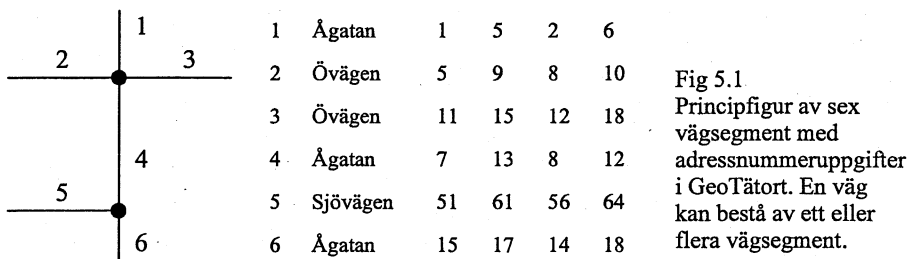


Fig 5.1
Principfigur av sex vägsegment med adressnummeruppgifter i GeoTätort. En väg kan bestå av ett eller flera vägsegment.

att varje enskilt vägmittlinjesegment är försett med uppgifter om gatans namn, samt första och sista adressnummer på vänster respektive höger sida. Inmatningen av dessa uppgifter till GeoTätort har hittills varit ett tidskrävande arbete som utförts manuellt med vanliga kartor som underlag. En automatisering av denna informationsöverföring skulle innebära stora fördelar i T-Kartors databasproduktion främst med avseende på tids- och arbetsbesparing.

5.1.2 Krav

Ett krav som T-Kartor ställer på ArcView-applikationen är att den ska kunna användas för en GeoTätort-databas som helt saknar adressattribut såväl som en databas som redan har adressattribut.

Vidare krävs att minst 95 procent av adresserna ska matcha vid en geokodning efter en körning av applikationen på en GeoTätort-databas över storstäder som Stockholm, Göteborg och Malmö.

Vid förändringar av uppgifter i GeoTätort-databasen skall användaren ha kontroll över vilka ändringar som utförs av applikationen eftersom det inte kan förutsättas att uppgifterna adressregistret alltid är riktiga. De ändringar som utförs skall dokumenteras.

5.1.3 Användning

T-Kartors användning av denna ArcView-applikation är framför allt när det ska skapas GeoTätort-databaser över nya samhällen, då ett adressregister kan användas för att sätta adressattributen till de geografiska objekten i GeoTätort. Att manuellt läsa av en adresskarta och sedan lägga till information om adressnummer till varje element i GeoTätort är ett mycket tidskrävande arbete. En automatiserad process skulle därför innebära stora arbetsbesparingar, vilket i sin tur förbättrar konkurrenskraften eftersom produktionskostnaden för en ny databas kan sänkas.

Vidare kan applikationen användas vid kontroll och förbättring av befintliga delar av GeoTätort-databasen.

5.2 Förutsättningar

Ingångsdata för att lösa uppgiften är dels ett adressregister och dels T-Kartors egna databas GeoTätort över motsvarande område.

5.2.1 Programmiljö

Det GIS-verktyg som innehar en central roll i T-Kartors verksamhet är Esri's ArcView, för närvarande i version 3. ArcView erbjuder de möjligheter för adressmatchning som behövs vid geokodningen och som är en förutsättning för att kunna lösa uppgiften (se avsnitt 3.2.3).

5.2.1.1 Programmeringspråk

För att kunna utnyttja ArcView's verktyg för adressmatchning och underlätta presentation av resultat skrivs programkoden i ArcView's egna programmeringsspråk Avenue (se avsnitt 3.4.3 och 3.5).

5.2.2 Indata

De enda indata som antas förekomma är den del av T-Kartors egna databas för GeoTätort som ska förbättras samt ett adressregister över motsvarande område. Adressregister förutsätts innehålla data från endast en kommun, då det annars föreligger stor risk för att samma gatunamn förekommer i flera kommuner.

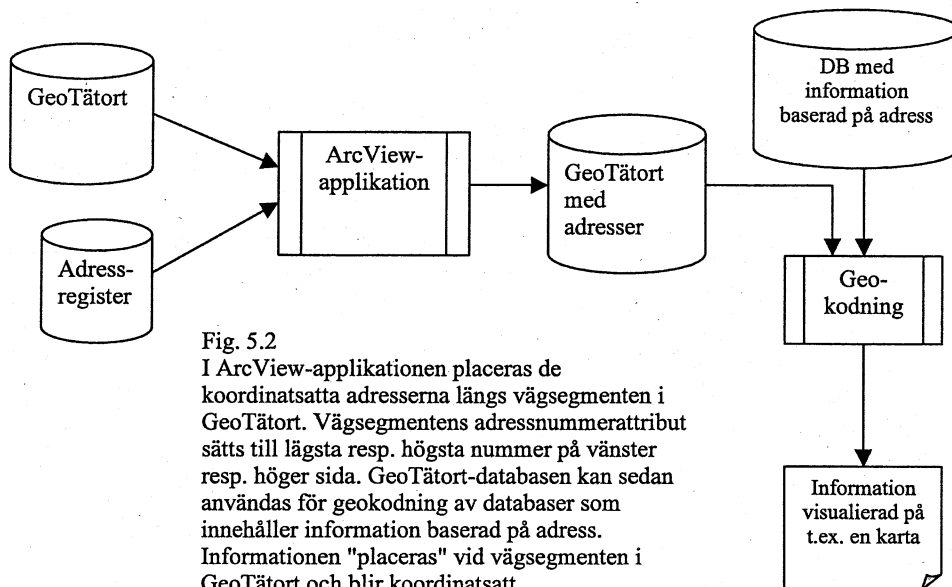


Fig. 5.2
 I ArcView-applikationen placeras de koordinatsatta adresserna längs vägsegmenten i GeoTätort. Vägsegmentens adressnummerattribut sätts till lägsta resp. högsta nummer på vänster resp. höger sida. GeoTätort-databasen kan sedan användas för geokodning av databaser som innehåller information baserad på adress. Informationen "placeras" vid vägsegmenten i GeoTätort och blir koordinatsatt.

5.2.2.1 GeoTätort

GeoTätort-databasen används i originalskick (se kapitel 4 om GeoTätort).

De attribut i databasen som är nödvändiga för att genomföra processen är vägsegmentens ID, namn och första respektive sista adressnummer på höger och vänster sida (r_add_fr, r_add_to, l_add_fr, l_add_to). Information finns också om vilken riktning varje vägsegment har. P.s.s. kan ett segments högra och vänstra sida identifieras.

Exempel:

ID	Namn	Startnod	Slutnod	l_add_fr	l_add_to	r_add_fr	r_add_to
2071	KAPTENSGATAN	25687	25901	5	9	2	10
2072	KAPTENSGATAN	21451	16899	11	29	12	24
2073	STRANDVÄGEN	11425	9854	1	21	2	26

5.2.2.2 Adressregister

Adressregistret förutsätts vara i databasformatet dBase (dbf). T-Kartor hanterar ofta adressregister som tillhandahållits i vanligt textformat men sådana register formateras enkelt om till dBase-format i ett databashanteringsprogram.

Alla förekomster i adressregistret antas innehålla uppgifter om gatunamn och adressnummer. Vidare antas att den geografiska positionen för varje adress i registret representeras av x- och y-koordinater.

Varje adress är ett element i tabellen. Adressen är angiven som gatunamn_adressnummer. Ex. KAPTENSGATAN_14

Om adresstabellen är sorterad i bokstavsordning efter adressernas gatunamn är det till

avsevärd fördel vid processen. Tidsåtgången kan minskas drastiskt eftersom ett flertal av de operationer i programmet som behandlar ett element åt gången inte behöver upprepas om en adress har samma gatunamn som föregående adress.

Exempel:

Adress	x	y
KAPTENSGATAN 22	12112	37554
KAPTENSGATAN 23	12101	37520
KATODVÄGEN 1	11245	43552

Adressregistret läggs till den vy i ArcView-projektet som innehåller GeoTätort-temat. Adressregistret finns då i shapeformat vilket möjliggör de operationer som ska genomföras. Vid konverteringen till shapeformat placeras elementen (adresserna) på kartan m.h.a. av dess koordinater.

5.3 Överföring av adressinformation från ett adressregister till GeoTätort i ArcView

Adressregistret och GeoTätort finns nu som teman, shapefiler, i en vy i ArcView. Dessa teman baseras på tabeller innehållande den information som finns för varje element i temat.

Programmet, eller applikationen, består av fyra Avenue-script. Ett huvudscript sköter den övergripande hanteringen av de olika teman som finns och som temporärt skapas. Detta script anropar tre olika script som genomför de olika operationerna som ska genomföras med dessa teman.

Programkoden i scripten beskrivs nedan med s.k. pseudoprogramkod.

5.3.1 Koppling av adressnummer till vägsegment

Den första operationen är att för varje adress i adresstabellen identifiera vilken gatulänk, med samma namn, som ligger närmast. Därefter identifieras vilka adressnummer som finns på gatulänken. Vid programstart är alla adressnummer noll men allt eftersom programkörningen fortgår kommer länkarna att tilldelas adressattribut. Adressnumret jämförs sedan med gatulänkens lägsta resp. högsta adressnummerattribut beroende på om numret är udda eller jämnt. Om numret är lägre än det befintliga lägsta numret eller högre än det högsta numret ändras gatulänkens felaktiga adressnummerattribut.

Tillvägagångssätt

Pseudoprogramkod:

För varje adress i adresstabellen

adressnamn = adress - adressnummer

Om adressnamn <> föregående adressnamn

Markera alla element i GeoTätort-tabellen med gatunamn = adressnamn

(För varje gatunamn som förekommer i adresstabellen markeras de element i GeoTätort-tabellen som har samma namn.)

För varje vägsegment i den markerade GeoTätort-tabellen

Beräkna avståndet till adressen

Om avståndet < tidigare minsta avstånd

minsta avstånd = avstånd

närmaste vägsegment = vägsegmentets ID

(För varje adressedelement beräknas vilket av de markerade vägsegmenten som ligger närmast.)

För varje vägsegment i den markerade GeoTätort-tabellen

Om vägsegmentets ID = närmaste vägsegment

Om adressnummer är udda

Om adressnummer < vägsegmentets l_add_fr

vägsegmentets l_add_fr = adressnummer

Om adressnummer > vägsegmentets l_add_to

vägsegmentets l_add_to = adressnummer

Om adressnummer är jämnt

Om adressnummer < vägsegmentets r_add_fr

vägsegmentets r_add_fr = adressnummer

Om adressnummer > vägsegmentets r_add_to

vägsegmentets r_add_to = adressnummer

(Kontroll om adressnumret är udda eller jämnt, d.v.s. om det skall finnas på vägsegmentets vänstra eller högra sida. Om adressnumret är lägre än det nummer som representerar segmentets lägsta adressnummer, eller om det är högre än det nummer som representerar segmentets högsta adressnummer, ändras detta attribut till adressnumret.)

Nästa adress i adresstabellen

Slut pseudoprogramkod

5.3.2 Geokodning

Geokodningsproceduren (se kapitel 2) matchar adresserna mot GeoTätort-temat och skapar ett tema som visar vilka adresser som har matchat. Denna procedur genomgås främst för att på ett enkelt sätt använda ArcViews inbyggda funktioner för att upptäcka vilka adresser som kan misstänkas ha en annorlunda stavning än vägsegmenten i GeoTätort. Geokodningen ger även ett mått på GeoTätorts kvalitet gentemot adressregistrets, hur många adresser som kan hittas vid en sökning på adress i GeoTätort. Geokodningsoperationen är en något modifierad version av den geokodningsprocedur som ArcView använder.

Observera att det tema som skapas efter operationen redovisar adressernas position baserat på deras adress och inte på koordinaterna som finns i adresstabellen.

Geokodningsproceduren är endast avsedd som en kvalitetskontroll av GeoTätort. Resultatet av proceduren används sedan som underlag för nästa procedur där eventuella felaktigheter ska indentifieras.

I scriptet anges att den AddressStyle som ska användas vid geokodningen är gatunamn_adressnummer samt vilka villkor som måste uppfyllas för att adresserna ska matcha.

Geokodningen genomförs som beskrivs i avsnitt 2.4.

5.3.3 Ändring av felaktiga adresser

Det krävs en betydligt mer omfattande och mer invecklad procedur för att indentifiera de gatunamn i GeoTätort som inte stämmer med adressnamnen i adressregistret och sedan rätta till eventuella felaktigheter. Eftersom det är förenat med en del risker att ändra gatunamnen i GeoTätort-databasen krävs det att denna procedur är interaktiv. Användaren måste kunna styra om de förändringar som föreslås av applikationen ska göras. Genomförda förändringar redovisas i en dbf-fil som skapas vid proceduren.

5.3.3.1 Tillvägagångssätt

Den modifierade geokodningsproceduren har skapat ett tema med elementen i adresstabellen där ett attribut anger om adressen har matchat mot GeoTätort eller inte. Huvudscriptet skapar ett tema som består av de adresser som inte matchat. Detta tema (tabell) kallas nedan för omatchtabell. Proceduren använder de x- och y-koordinater som de omatchade adresserna "ärvt" från den ursprungliga adresstabellen.

För varje omatchad adress skapar applikationen en lista som innehåller de gator som ligger i adressens närhet. Dessa gatunamn presenteras i en dialogruta (fig. 5.3) varifrån användaren väljer vilket gatunamn som skall ändras till överensstämmelse med den omatchade adressens namn.

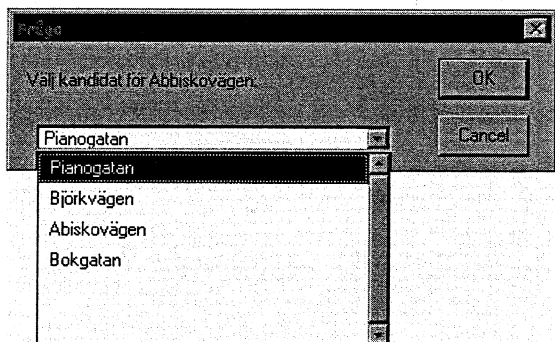


Fig. 5.3
Användaren väljer vilket gatunamn i GeoTätort som ska ändras.

Pseudoprogramkod:

Skapa en lista för bokföring av genomförda ändringar

För varje adress i omatchtabellen

adressnamn = adress - adressnummer

Om adressnamn <> föregående adressnamn

Skapa kandidatlista

(För varje gatunamn som förekommer i omatchtabellen skapas en lista som ska innehålla de vägsegment som finns i närheten av adresser med samma adressnamn.)

För varje vägsegment i GeoTätort-tabellen

Om vägsegmentets attribut l_add_fr, l_add_to, r_add_fr, r_add_to = 0

Om avståndet mellan adressen och vägsegmentet < 200

Markera vägsegmentet i GeoTätort-tabellen

Lägg till vägsegmentets namn i kandidatlistan

(För varje adress i omatchtabellen markeras de element i GeoTätort-tabellen som ligger inom 200 längdenheter (här meter) radie från adressen. Dessa vägsegments namn läggs till listan med kandidater.)

Om antalet kandidater i kandidatlistan > 0

Användaren får välja vilken av kandidaterna som ska ändra namn = vald

(Via en dialogruta med s.k. rullningsmeny frågar applikationen användaren om och i så fall vilket av gatunamnen i kandidatlistan som ska ändra namn till samma adress som den omatchade adressen (se fig. 5.3).)

För varje vägsegment i GeoTätort-tabellen

Om vägsegmentets namn = vald

*Om vägsegmentets attribut l_add_fr, l_add_to, r_add_fr,
r_add_to <> 0*

Ändring inte tillåten

(GeoTätort-tabellen söks än en gång igenom. Om det existerar något vägsegment i tabellen som har samma namn som det valda gatunamnet och har något adressnummer i något av de attribut som representerar vilka adressnummer som finns på segmentet, kommer någon ändring av namnet inte att göras.)

Om ändring tillåten

Användaren får en kontrollfråga om namnändring ska genomföras

För varje vägsegment i GeoTätort-tabellen

Om vägsegmentets namn = vald

Sätt vägsegmentets namn = adressnamn

*Lägg vägsegmentets gamla och nya namn till
listan med genomförda förändringar*

Annars

Användaren får besked att ändring inte kan genomföras

(Om inget av vägsegmenten i GeoTätort-tabellen med det valda namnet har några adressnummer kopplade till sig frågar applikationen än en gång användaren om adressändringen ska göras. Om så är fallet ändras alla segmentens namn till samma som den omatchade adressen. Den genomförda ändringen läggs till en lista för att hålla reda på vilka förändringar som gjorts i GeoTätort-tabellen.)

Skapa dbf-fil innehållande elementen i listan med genomförda ändringar

(Listan med utförda ändringar i GeoTätort sparas i dbf-format.)

Slut pseudoprogramkod.

5.3.4 Repetering

Om det har gjorts några korrigeringar i GeoTätort frågar applikationen om hela proceduren skall repeteras eftersom adresser nu kan kopplas till de element i GeoTätort-tabellen som

haft felaktiga gatunamn.

5.4 Testning

5.4.1 Testdata

Vid de upprepade tester som gjorts har ett adressregister över Stockholm använts. Adressregistret innehåller drygt 95000 adresser med sitt ursprung från Stockholms kommun.

GeoTätort över motsvarande område innehåller drygt 29000 vägsegment att överföra information om adressnummer till.

5.4.2 Testresultat

Det är svårt att fullständigt testa hur bra kopplingen av adresser fungerar. Geokodningen ger dock ett mått på hur många av adresserna som har matchat mot något vägsegment i GeoTätort. Det krav som ställt av T-Kartor anger att minst 95 procent av adresserna ska matcha vid en geokodning efter en körning av applikationen på en GeoTätort-databas över storstäder som Stockholm, Göteborg och Malmö. Vid första geokodningen visar siffrorna på att knappt fyra procent av adresserna inte har matchat. Redan det är ett bra resultat som ligger under under den gräns på fem procent som är målet vid adressmatchning på GeoTätort i storstäder som Stockholm, Göteborg och Malmö (Kravspecifikation GeoTätort, 1997). Resultatet efter geokodning då hela processen har körts för andra gången, d.v.s. då användaren har ändrat felaktiga adressnamn, visar att mer än hälften av dessa fel har eliminerats. Som jämförelse kan nämnas att den koppling av adresser för Stockholm som tidigare utförts manuellt endast kunde matcha knappt 92 procent av adresserna.

Operation	Omatchade adresser
Geokodning 1	3,9 %
Geokodning 2	1,5 %

De fel som applikationen upptäcker i GeoTätort är till största delen felstavningar i form av utelämnade eller omkastade bokstäver eller att t.ex. ändelsen -gatan använts i stället för -vägen. Andra typer av fel går inte att åtgärda eftersom det förekommer adresser vars gatunamn inte finns i GeoTätort. I testerna med Stockholm har sådana fel ofta visat sig vara enstaka adresser som -torg, -plats eller -plan. I flera av dessa fall saknas adressen p.g.a. att det finns en annan adress på samma punkt, byggnaden har s.k. dubbel adress.

En ändring av ett felstavat gatunamn har stor betydelse för det förbättrade resultatet eftersom alla vägsegment med det felaktiga namnet ändras och alla adresser på den gatan kommer att matcha då processen upprepas.

5.4.3 Tidsåtgång

Processens första del, kopplingen av adressnummer till vägsegment, tar mycket datorkapacitet. För varje nytt gatunamn söks först alla vägsegment med samma namn upp och sedan beräknas vilket av dessa som ligger närmast adressen.

Den interaktiva delen av processen med att ändra de felaktiga adresserna är ganska tidskrävande för användaren av så här stora datamängder, naturligtvis beroende av hur bra resultatet är vid geokodningen.

Hela processen är ganska tidskrävande och tar många datortimmar i anspråk. Skillnaden är dock stor mot tidigare då en liknande manuell adressättning kunde ta flera arbetsdagar. Den

operation som kräver några timmars interaktiv medverkan av användaren är proceduren för ändring av adressnamn.

Operation	Tidsåtgång
Koppling av adressnummer	11 timmar
Geokodning	2 timmar
Ändra felaktiga namn	2-3 timmar (kräver operatör)

Tidsåtgången är proportionell mot databasens och adressregistrets storlek och till viss del också hur många fel som påträffas. De ovan angivna tiderna är för ett adressregister med 95000 adresser och en del av GeoTätort innehållande 29000 vägsegment.

5.5 Slutsatser

Proceduren för att koppla adresserna till vägsegmenten är en ganska enkel. Adresserna placeras vid det vägsegment med samma namn som ligger närmast adressen. Avgörande här är vilka koordinater som adressen har tilldelats. Är det mittpunktskoordinater för fastigheten där adressen finns? Är det koordinater för byggnad på adressen eller representerar de var ingången till adressen finns?

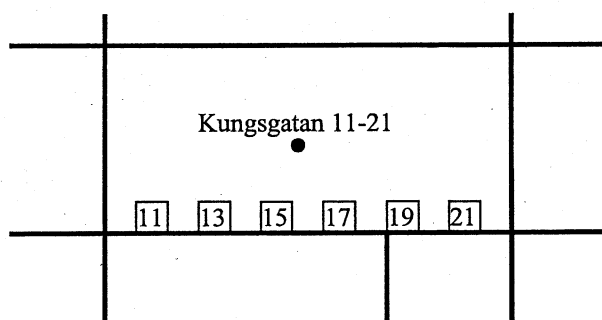


Fig. 5.4
Adresserna har samma koordinat; fastighetens centralpunkt. I verkligheten är de jämnt fördelade längs gatan.

Om adressregistret som används inte har speciellt hög lägesnoggrannhet kan det uppkomma ett fel som ej funnit någon lösning på automatiskt sätt. Om adressregistret kommer från t.ex. FDS är adresserna lokaliserade till en fastighet (se fig. 5.4).

Vid koppling av adresser kommer alla adresserna i fastigheten att kopplas till samma länk;

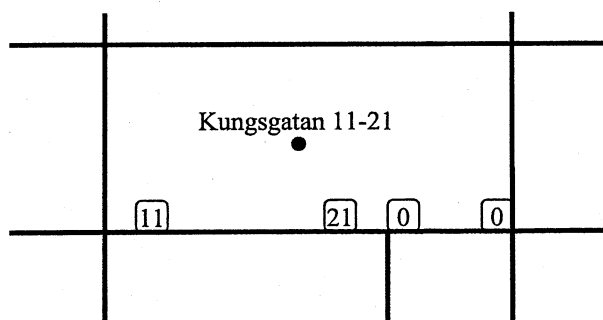


Fig. 5.5
Efter proceduren har samtliga adresser kopplats till samma vägsegment; det som ligger närmast fastighetens centralpunkt.

den som är närmast fastighetens centralpunkt. Det kortare vägsegmentet kommer att inte att få några adressnummer på dess vänstra sida (se fig. 5.5). Det är svårt att se hur en helautomatisk lösning av detta fel skulle se ut. Tänkbart är att lägga till en kontrollfunktion som söker upp var sådana fel kan misstänkas, t.ex. där en länk med adressattribut som gränsar till en kortare länk med samma namn som saknar adressattribut. Användaren får sedan manuellt rätta till eventuella felaktigheter.

En annan lösning skulle kunna vara att bilda polygoner på vägnätet. På så sätt skulle vägsegmenten i exemplet ovan kunna identifieras till att utgöra ett sammanhängande element. Adresserna skulle då kunna placeras med jämna mellanrum längs de båda vägsegmenten. Denna lösning skulle kräva en hel del ytterligare beräkningar och jämförelser, dels för att identifiera de fall då lösningen bör användas, dels för att utföra fördelningen av adressnummer på flera vägsegment. Ett problem med denna lösning finns också i de fall då en återvändsgata "sticker in" i en polygon.

Det är mycket svårt att ange ett mått för hur väl kopplingen av adressinformation fungerar. Resultatet av kopplingen beror mycket på kvaliteten, främst noggrannheten, i adressregistret. En viktig faktor för hur resultatet blir är hur ortens stadsstruktur ser ut. I ett villaområde fungerar kopplingen i princip hundra procentigt, medan det i en mer komplicerad struktur som flerbostadsbebyggelse uppkommer fler fel. Resultatet av kopplingsproceduren kan bara kontrolleras ordentligt genom att jämföras manuellt mot en detaljerad adresskarta. Geokodningsproceduren ger ett kvalitetsmått på hur många adresser som har matchat mot ett vägsegment. Vid de tester som gjorts har andelen omatchade adresser varit under de fem procent som angivits som krav för storstäder.

De förändringar som applikationen föreslår vid ändring av gatunamn är endast baserade på de uppgifter som finns i adressregistret. Det finns egentligen inget som talar för att dessa uppgifter skulle vara mer riktiga än de som finns i GeoTätort. Det är därför av stor vikt att känna till av vilken kvalitet det använda adressregistret har. Vid förändringar av gatunamn bör adressnamnet kontrolleras mot ytterligare någon datakälla. Vidare bör ett attribut tillföras vägsegmenten i GeoTätort som anger vilken källa och dess kvalitet som ligger till grund för vägsegmentens adressattribut. De genomförda ändringarna vid adressändring sparas i en speciell fil för att en eventuellt felaktig förändring av GeoTätort skall kunna återställas.

Denna automatiserade adressättning är ganska tidskrävande. Skillnaden mot tidigare förhållanden är dock att proceduren utförs under ett antal "datortimmar" i stället för ett flertal manuella arbetsdagar vilket naturligtvis kan innebära avsevärda kostnadsbesparingar. Störst nytta har man av applikationen om den används på en databas som helt saknar adressattribut eller om adressinformationen anses vara behäftad med så mycket fel att en ombearbetning bör göras.

Applikationen får anses ha löst problemet. De krav som ställts på applikationen har vid tester kunnat verifierats som uppfyllda.

Litteraturförteckning

- Agnarsson, M., Vingå, G., Kallur, S., Johnsson, K. - *Behovsanalys GeoTätort*
Telia InfoMedia / T-Kartor Sweden AB, Kristianstad, 1997
- Bernhardsen, T. - *Geographic Information Systems*
Viak IT, Arendal, Norge, 1992
- Bjöersdorff, T. - *Digital city maps plus telephone directory databases: a new marketing opportunity*
GIS in Business 94 Europe, Proceedings, s. 55-59
Longman GeoInformation, Cambridge, Storbritannien, 1994
- Booch, G. - *Object Oriented Design With Applications*
The Benjamin/Cummings Publishing Company Inc., 1991
- Burbach, M. - *How to Effectively Use Customer Data*
GIS in Business 1994, Conference Proceedings, s. 205-206
GIS World, Inc., Fort Collins, USA, 1994
- Camarata, S.J. - *Geography for Business: Who needs it?*,
GIS in Business 1994, Conference Proceedings, s. 19-22
GIS World, Inc., Fort Collins, USA, 1994
- Carlsson, Å., Johnsson, K. - *The Master-Product Model For Geographic And Cartographic Data Management*
T-Kartor Sweden AB, Kristianstad, 1997
- Environmental Systems Research Institute, Inc. - *Customizing ArcView with Avenue*
Environmental Systems Research Institute, Inc., Redlands, USA, 1994a
- Environmental Systems Research Institute, Inc. - *Introducing Avenue*
Environmental Systems Research Institute, Inc., Redlands, USA, 1994b
- Environmental Systems Research Institute, Inc. - *Using ArcView GIS*
Environmental Systems Research Institute, Inc., Redlands, USA, 1996
- Galvenius, G. - *Geodetisk plan- och nivåmätning*
Stockholm, 1985
- Holm, P. - *Objektorienterad programmering och Simula*
KF-Sigma, Lund, 1992
- Lantmäteriverket - *Handbok till mätningsskuggörelsen - Databaser*
Lantmäteriverket, Gävle, 1993a
- Lantmäteriverket - *Handbok till mätningsskuggörelsen - Geodesi, GPS*
Lantmäteriverket, Gävle, 1993b
- Marx, R.W. - *The Census Bureau's TIGER System*
Cartography and Geographic Information Systems, Volume 17, No. 1, 1990

American Congress on Surveying and Mapping, Bethesda, USA, 1990

Oxenstierna, A. - *Generalisation Rules For Database-Driven Cartography*
T-Kartor Sweden AB, Kristianstad, 1997

Standardiseringen i Sverige, SIS - *Belägenhetsadresser, utgåva 1 1998, Svensk Standard SS 63 70 03*
SIS Förlag, Stockholm, 1998

Telia InfoMedia - *Kravspecifikation GeoTätort*
Telia InfoMedia, Stockholm, 1997

T-Kartor Sweden AB - *Produktion Data CPS Konsulting*
T-Kartor Sweden AB, Kristianstad, 1996

T-Kartor Sweden AB - *T-Kartor's informationsblad nr 2*
T-Kartor Sweden AB, Kristianstad, 1997a

T-Kartor Sweden AB - *T-Kartor's informationsblad nr 3*
T-Kartor Sweden AB, Kristianstad, 1997b

Worboys, M.F. - *GIS: A Computing Perspective*
Taylor & Francis, Keele, Storbritannien, 1995

Referenser på Internet (November 1999)

www.gulasidorna.se - Telia InfoMedia

www.t-kartor.se - T-Kartor Sweden AB

www.uiowa.edu/~geog/health/address/address.html - The Department of Geography, University of Iowa, USA

http://www.spea.indiana.edu/jgant/Addr_match/index.htm - The School of Public and Environmental Affairs, Indiana University, Bloomington, USA

<http://www.geography.ccsu.edu/harmonj/g378/html/busapps/index.htm> - Department of Geography, Central Connecticut State University, USA

