

Genereren van een 6-positie postcodebestand op basis van de kadastrale registratie

Kwalitatief goede postcodebestanden spelen een belangrijke rol in veel GIS-toepassingen, aangezien de postcode gebruikt kan worden als koppeling tussen het administratieve adres van gebouwen en de bijbehorende geografische locaties. Deze koppeling is in beide richtingen mogelijk: adres-matching wordt gebruikt om een bepaald adres te vinden bij een gegeven coördinaat en geocoderen wordt gebruikt om aan een adres een coördinaat toe te kennen.

Bestaande adres-matching en geocoderingstechnieken kennen echter een aantal beperkingen. Zo maken de meeste adres-matching-algoritmes gebruik van lineaire interpolatie tussen twee knooppunten in een wegennetwerk. Hierbij wordt een reeks van huisnummers gekoppeld aan een wegsegment. In combinatie met de veronderstelling dat de adressen regelmatig verspreid langs de straat liggen, kan de locatie van een bepaald adres op het wegsegment worden berekend. Dit geeft acceptabele resultaten zolang de adressen inderdaad min of meer regelmatig verspreid liggen, maar complexere situaties vragen om een slimmere aanpak.

De geografische locatie van een postadres, en daarmee dus ook van de postcode, is niet eenduidig te definiëren. Het middelpunt van een huis kan gezien worden als de geografische locatie van de bijbehorende combinatie postcode-huisnummer, maar de van deze punten afgeleide geografische locatie van een postcode, bestaande uit gemiddeld zo'n 16 adressen, blijft een vrij willekeurig punt. De voor de berekening gebruikte coördinaten blijken namelijk in de praktijk vaak willekeurig toegekend of berekend (bijv. de centroïde) binnen de grenzen van een gebouw. En hoe kan een 6-positie postcodevlakkenbestand geconstrueerd worden op basis van



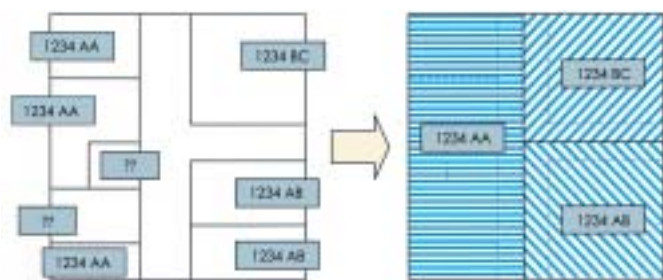
*Friso Penninga,
en co-auteurs
Edward Verbree,
Wilko Quak en
Peter van
Oosterom,
TU Delft-OTB,
sectie GIS-t.
f.penninga@
otb.tudelft.nl*

deze postcodepunten? Het berekenen van Thiessen-polygonen (Voronoi diagram) rond deze postcodepunten kan leiden tot dubieuze resultaten. Zo is in fig. 1 te zien dat grenzen van postcodegebieden bebouwde percelen doorsnijden en de meeste grenzen niet samenvallen met fysieke grenzen, zoals wegen en rivieren.

Om deze problemen op te lossen is een aanpak, die uitgaat van de kadastrale registratie, ontwikkeld en geïmplementeerd met behulp van de Computational Geometry Algorithms Library (CGAL). De 6-positie postcode functioneert nu als koppeling tussen een adres en de bijbehorende locatie. Bij een specifiek coördinaat kan een adres-matching-algoritme nu op basis van een point-in-polygon-operatie eerst het bijbehorende postcodegebied vinden en vervolgens het bijbehorende perceel binnen het postcodegebied. Door de kadastrale registratie is bij dit perceel de complete adresbeschrijving van plaats, straat en huisnummer te geven. In de omgekeerde richting kan een geocoderingsalgoritme nu het perceel gebruiken als uitgangspunt om zo de postcode en het bijbehorende polygoon te vinden.

*Fig. 1.
Thiessen-polygonen
representeren
postcodegebieden
(bron: Bridgis GIS
starterkit).*





Deze methode werkt probleemloos voor bebouwde percelen. De gebouwen zijn niet alleen beschreven met een plaats, straat en huisnummer, maar ook met een door de TPG toegekende postcode. Aan onbebouwde percelen is echter geen postcode toegekend. Om tot een volledig dekkend postcodebestand te komen, zal voor deze onbebouwde percelen de meest waarschijnlijke postcode worden bepaald. Aangezien we willen dat de postcodegebieden zoveel mogelijk door infrastructuur worden begrensd, splitst het algoritme eerst de infrastructuurpercelen op de hartlijn om vervolgens de afzonderlijke delen aan de meest waarschijnlijke postcodegebieden toe te voegen. Dit alles resulteert uiteindelijk in een vlakverdeling van Nederland in niet-overlappende aangrenzende polygonen, waarbij elke polygoon een postcode representeert en de infrastructuure objecten op de grenzen liggen van deze postcode-polygonen. Dit concept is geïllustreerd in fig. 2.

Creëren postcodegebieden

Voor het creëren van de postcodegebieden zijn gegevens nodig uit zowel AKR als uit LKI. Als een perceel bebouwd is en er eigendomsrechten op rusten, dan is er een AKR-objectadres beschikbaar en daarmee een postcode. Indien er geen AKR-objectadres aan een perceel gekoppeld is, is er daarmee ook geen postcode bekend. Ook is het mo-

Fig. 2.
Concept: van
kadastrale situatie
(links) naar
postcodegebieden
(rechts).

gelijk dat aan een perceel meerdere AKR-objectadressen gekoppeld zijn, wat kan resulteren in meerdere postcodes. Deze laatste situatie, die bijvoorbeeld bij een groot appartementencomplex of een groot perceel met huurwoningen kan voorkomen, wordt in het algoritme niet meegenomen, maar er worden wel suggesties voor gedaan in de aanbevelingen. In AKR wordt een indicatie gegeven van het soort gebruik van het perceel in de bebouwings- en de cultuurcode. Op basis hiervan kan worden bepaald of een perceel bebouwd is of dat het gebruikt wordt als infrastructuur (wegen, spoorbanen of waterlopen).

Als de postcode die via AKR-objectadres gevonden is, gekoppeld wordt aan de grenzen uit LKI, is de postcode fysiek gerepresenteerd door de eigendomsgrenzen. Vaak zijn deze eigendomsgrenzen zichtbaar in het terrein en daarmee dus ook de postcode. Door nu de percelen met dezelfde postcode samen te voegen, ontstaan er postcodegebieden waarvan de grenzen herkenbaar zijn in het terrein. Aangezien voor adres-matching een vlakdekkend postcodebestand vereist is, is het noodzakelijk om ook postcodes toe te kennen aan percelen waaraan in AKR geen postcode gekoppeld is. Dit kan eenvoudig gebeuren middels een iteratief proces waarbij de meest waarschijnlijke postcode wordt bepaald aan de hand van de postcodes van de aangrenzende percelen. Als dit proces ook toegepast wordt op infrastructuurpercelen treden er echter ongewenste situaties op. De praktijk leert dat huizen (en daarmee percelen) aan de ene zijde van de weg vaak een andere postcode hebben dan de huizen aan de andere zijde. Als in het iteratieve proces nu één van deze twee postcodes aan de weg wordt toegekend, gaat de informatie dat aan deze weg ook huizen met de andere postcode liggen verloren. Adres-matching met punten die op deze weg liggen, wordt hierdoor moeilijk uitvoerbaar. Ook wordt het kaartbeeld minder leesbaar, omdat sommige postcodegebieden onevenredig hard groeien. Gevoelsmatig ligt de grens tussen de twee postcodes echter op de hartlijn van de weg. Daarom zou de ene helft van de weg aan de ene postcode toegekend moeten worden en de andere helft aan de andere postcode. Het splitsen van infrastructuurpercelen door skelettering om de afzonderlijke delen aan verschillende postcodes te kunnen toekennen, is één van de belangrijkste kenmerken van het nieuw ontwikkelde algoritme.

Skeletteren van infrastructuurpercelen

Alle percelen worden ingelezen in de Planar map-datastructuur in CGAL. Dit kan eenvoudig gebeuren door alle perceelsgrenzen uit het LKI toe te voegen. Zoals eerder aangegeven moeten de percelen zonder adres die op basis van hun cultuurcode als infrastructuur zijn aangemerkt, worden geskeletterd om vervolgens op dit skelet te worden gesplitst. Eerst wordt een constrained triangulatie van de perceelsgrenzen opgebouwd, waarbij de oorspronkelijke edges van het infrastructuurperceel in de triangulatie moeten terugkomen (groen in fig. 3). De overige edges (zwart in fig. 3) worden bij de triangulatie berekend. Op basis van deze triangulatie wordt het skelet (rood in fig. 3) berekend door gebruik

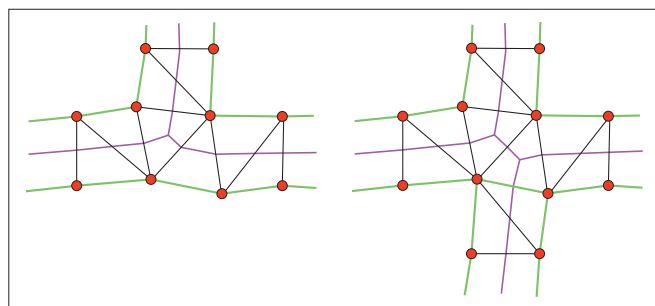


Fig. 3.
DeLucia en Black
skeletten van een
T-splitsing en een
kruising.

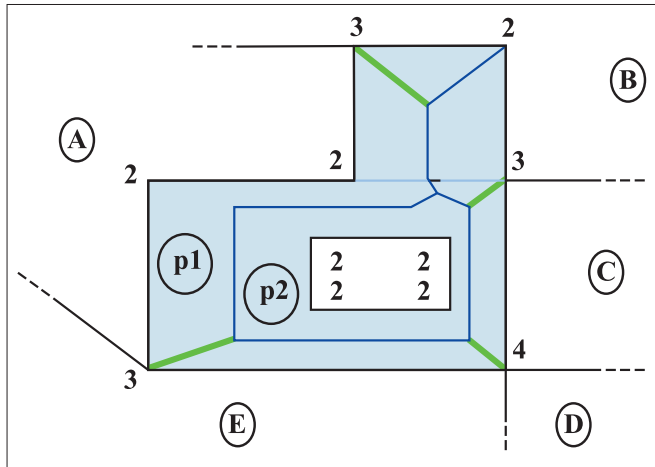
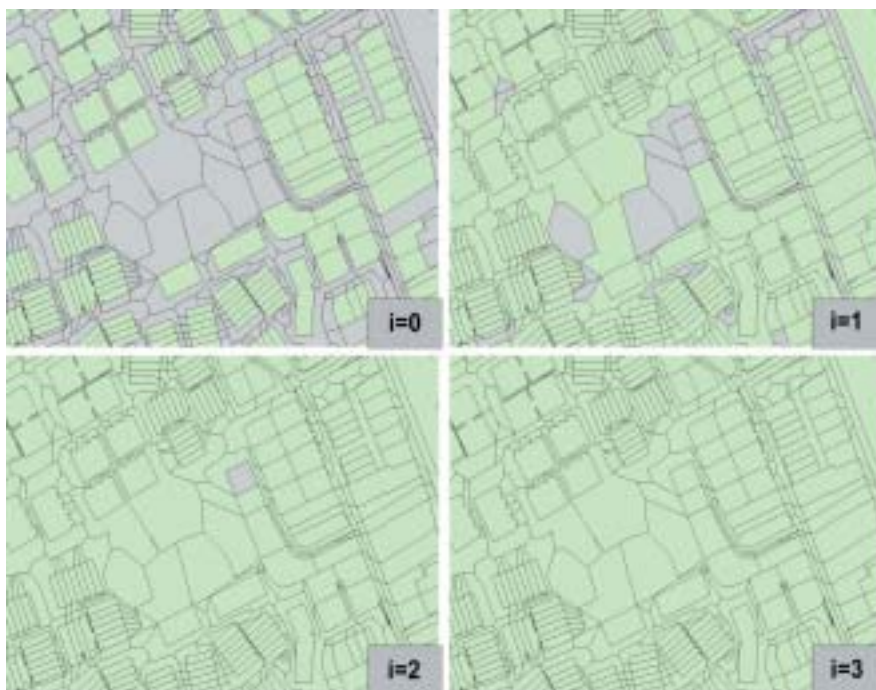


Fig. 4.
Perceel met skelet
(blauw) en extra
zijtakken (groen) om
uiteindelijk in zes
delen te splitsen.

te maken van de methode van DeLucia en Black [3]. Soortgelijke technieken worden ook regelmatig toegepast bij kaart-generalisatie [1][2]. Het aantal constrained edges in een driehoek is bepalend voor de wijze waarop het skeletsegment in deze driehoek wordt berekend. Fig. 3 toont een paar voorbeelden van de methode. Inwendige driehoeken zonder constrained edges (0-driehoeken) worden door het skelet in drieën gesplitst (bij t-splitsingen en kruisingen). Driehoeken met één constrained edge worden in twee delen gesplitst (doorgaande weg). De driehoeken met twee constrained edges worden ook in tweeën gesplitst, waarbij het skelet aansluit op een aangrenzend perceel (komt niet voor in de figuur).

Na deze berekeningen worden enkele extra zijtakken (groen in fig. 4) van het skelet berekend om het skelet met de buitengrens te verbinden en zo het perceel in meer afzonderlijke delen te kunnen splitsen. Dit gebeurt daar waar in de originele kaart het aantal verbonden lijnen op de punten (van het infrastructuurperceel) groter is dan twee. Met dit uitgebreide skelet wordt het originele perceel gesplitst, zie fig. 4. Deze gesplitste percelen worden vervolgens toegevoegd aan de Planar map-datastructuur in CGAL.

Fig. 5.
Iteratief postcodes
toekennen. Percelen
met postcode in
groen, zonder
postcode in grijs.



Toekennen postcodes aan (delen van) percelen zonder postcode

De infrastructuurpercelen zijn nu geskeletteerd en gesplitst in afzonderlijke delen. Aan deze delen moet nu de meest waarschijnlijke postcode toegekend worden, net als aan de percelen zonder AKR-objectadres. Om een postcode toe te kennen berekent het algoritme (de som van) de lengte van de grenzen die het perceel deelt met percelen mét een postcode. Vervolgens wordt gekozen voor de postcode met de langst gezamenlijke grens. Dit proces wordt herhaald totdat alle percelen een postcode hebben. Deze iteratieslagen zijn weergegeven in fig. 5. Een nadeel van deze aanpak is dat er soms twee of meer afzonderlijke, niet verbonden, gebieden kunnen ontstaan waaraan dezelfde postcode is gekoppeld. Dit is zichtbaar in fig. 6, waar de postcode 2761 JE is gekoppeld aan vier afzonderlijke gebieden.

Het samenvoegen van percelen met dezelfde 6-posities postcode gebeurt in de Planar map door het verwijderen van alle grenssegmenten met dezelfde postcode aan de linker en rechterkant.

Toekennen betrouwbaarheidsindicator aan postcodegebieden

Hoe meer iteratiestappen er nodig zijn om een postcode aan een perceel toe te kennen hoe onzekerder het wordt of de toegekende postcode 'juist' is. Deze betrouwbaarheid wordt weergegeven met de relatieve betrouwbaarheidsindicator Q . Deze indicator is afhankelijk van het aantal iteratiestappen dat nodig was om een postcode toe te kennen aan een perceel. De relatieve betrouwbaarheidsindicator wordt genormaliseerd tussen 0 en 1 door de som van de inverse van het aantal iteratiestappen te delen door het totaal aantal percelen in het postcodegebied. Een postcodegebied dat voornamelijk bestaat uit infrastructuur of onbebouwde percelen krijgt hierdoor een waarde dichtbij 0; een postcodegebied dat met name bestaat uit bebouwde percelen een waarde dichtbij 1. Men kan zich voorstellen dat in bebouwd gebied de relatieve betrouwbaarheid hoger ligt dan in landelijk gebied. Dit is ook duidelijk zichtbaar in fig. 7.



Fig. 6. Toekennen postcodes kan leiden tot afzonderlijke gebieden met dezelfde postcode.



Fig. 7. Berekende relatieve betrouwbaarheidsindicator Q .

Gebruik van CGAL binnen het project

De complete procedure van het skeletteren tot de uitvoer van de polygonen is geprogrammeerd in C++, waarbij gebruik is gemaakt van de Computational Geometry Algorithms Library, versie 2.4 (CGAL, 2002). CGAL is een gezamenlijk project van meerdere computationele geometrie-instituten in Europa en Israël. Het doel van het CGAL-project is het middels een C++-bibliotheek toegankelijk maken van de belangrijkste algoritmes uit de Computational Geometry voor gebruik in het bedrijfsleven en de wetenschappelijke wereld. Deze algoritmes dienen eenvoudig maar robuust te zijn. Het beschreven algoritme maakt met name gebruik van de Planar Map en Constrained Delaunay triangulatie-classes. Binnen de Planar Map is de Trapezoidal Map gebruikt voor de point-locate queries in de Planar Map.

De grootste voordelen van CGAL zijn de flexibiliteit en de efficiency. Een gebruiker kan de CGAL-classes zelf uitbreiden met de benodigde functionaliteit, terwijl de standaard efficiënte bewerkingen op de classes mogelijk blijven. Daarnaast zijn de datatypen geparametriseerd door het gebruik van C++-templates, waardoor de gebruiker de mogelijkheid heeft om een efficiënter algoritme te kiezen of het gebruik van geheugen te minimaliseren.

Het gevolg van het streven naar flexibiliteit is de toegenomen complexiteit van CGAL. Het zal de gemiddelde GIS-gebruiker vrij veel tijd kosten om goed te leren werken met C++ en de CGAL bibliotheek. Daarnaast is er een aantal valkuilen bij het kiezen van de juiste kernel. De kernel definieert welk number type gebruikt wordt in de representatie van de coördinaten van de ruimtelijke objecten. In de software van het ontwikkelde algoritme is gebruik gemaakt van exact reals in een homogeen coördinatenstelsel.

Resultaten

Het algoritme is getest met een bestand van de kadastrale gemeente Zevenhuizen (ZH). Tijdens deze test kwam een aantal beperkingen van de methode van DeLucia en Black aan de orde. Met name het skeletteren van kruisingen geeft vaak problemen. Een kruising bestaat uit twee aangrenzende 0-driehoeken, waardoor er een niet zinnige edge ontstaat tussen de twee hoeken (zie nogmaals fig. 3). Andere beperkingen zijn de aanwezigheid van valse 2-driehoeken en verschoven 0-driehoeken, zie fig. 8 en 9. Hierdoor ontstaan onterechte zijtakken respectievelijk een aantal pieken. Dit zijn allemaal bekende problemen bij het skeletteren en oplossingen hiervoor zijn beschreven in [4] en [6]. Deze en andere verbeteringen kunnen een startpunt vormen voor toekomstig onderzoek (zie kader).

Fig. 8.
Een valse 2-driehoek ontstaat door een niet-rechte wegkant met een onterechte zijtak als gevolg.

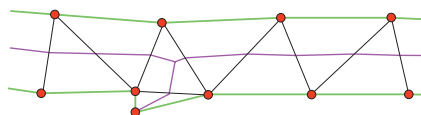
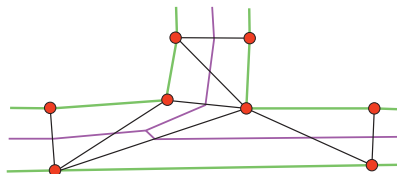


Fig. 9.
Een verschoven 0-driehoek leidt tot een piek in het skelet.



Relevant toekomstig onderzoek

1. Het verbeteren van de resultaten van skeletteren door
 - a. verwijderen van pieken;
 - b. verbeterde technieken toepassen op kruisingen, zoals voorgesteld in [4][6];
2. Onderzoeken van de relatie tussen de kwaliteit van de administratieve gegevens (cultuur- en bebouwingscode) en de resultaten van het algoritme
3. Afleiden van grotere postcodegebieden naast 6-positiënniveau: 5 ppc, 4ppc (en misschien zelfs 3ppc, 2ppc of 1ppc)
4. Leveren van kwantitatief bewijs voor het verbeterde resultaat van het ontwikkelde algoritme, ook voor grotere bestanden en andere type gemeenten
5. Gebruik van ACN-coördinaten voor het opsplitsen van grote 'huurpercelen' (via Thiesen polygonen)
6. Bij appartementen complexen, waarbij verschillende postcodes eigenlijk boven elkaar liggen, kan dit in de 2D-kaart worden gevisualiseerd door het betreffende perceel in een aantal symbolische 'taartpunten' op te splitsen (evenveel als er postcodes zijn).



Om de mogelijkheden van de nieuwe methode te onderstrepen is in fig. 10 een deel van de postcodegebieden in Zevenhuizen gecombineerd met de bebouwingslaag uit de Top10 vector. Duidelijk zichtbaar is dat met het nieuwe algoritme bebouwing vrijwel nooit meer doorsneden wordt door willekeurige postcodegrenzen. Een enkele uitzondering is weergegeven in fig. 11. De twee gebouwen die doorsneden worden zijn respectievelijk het gemeentehuis en een openbare school. Beide staan op één zeer groot perceel in eigendom van de gemeente, waarop naast deze openbare gebouwen ook negen wegen en een marktplein liggen. Vanwege deze verschillende functies is voor dit perceel geen cultuurcode ingevoerd en staat de bebouwingscode op 'onbebouwd'. Dit perceel, dat met name infrastructuur bevat, is geskeletteerd waarbij de grenzen de bebouwing doorsneden.



Fig. 10.
Resultaten van
het testgebied
Zevenhuizen,
gecombineerd
met de bebouwing
uit de Top10vector.

Alternatieven en verbeteringen

Er is een aantal mogelijke verbeteringen op het gepresenteerde algoritme te suggereren. Zo kan het algoritme eenvoudig veranderd worden op de volgende wijze:

1. ken postcodes toe (aan alle niet-infrastructurele percelen zonder adres) en voeg de percelen samen tot postcodegebieden zoals ook beschreven in [5];
2. bereken het skelet inclusief zijtakken van de infrastructuurpercelen en voeg de afzonderlijke delen toe aan de omliggende postcodegebieden;
3. ken postcodes toe aan overige percelen zonder postcode (percelen die geheel ingesloten zijn door infrastructuur).

Het voordeel van deze alternatieve werkwijze is dat gewaarborgd wordt dat postcodes de straat niet oversteken, wat overeenkomt met de werkelijke situatie. Een tweede voordeel is dat er bij de berekening van het skelet minder extra zijtakken (groen in fig. 4) nodig zijn, zodat de rekentijd beperkt wordt. Het is echter nog onduidelijk of deze alternatieve methode het probleem van de afzonderlijke gebieden met dezelfde postcode zal oplossen.

Fig. 11.
Twee gebouwen
worden doorsneden
als gevolg van
incorrecte
bebouwingscode.

Een behoorlijk andere implementatie zou kunnen zijn het alleen maar gebruik maken van een triangulatie voor alle percelen: voor de percelen met een postcode (alle driehoeken krijgen die postcode), voor de infrastructuurpercelen (driehoeken hebben geen postcode, maar worden gebruikt voor de skelettering, welke ook wordt toegevoegd na berekening) en voor de percelen zonder adres (driehoeken hebben geen postcode, maar krijgen die iteratief per driehoek toegekend). Percelen zonder adres kunnen gesplitst worden waardoor mogelijk voorkomen kan worden dat er meerdere niet-samenhangende gebieden bij één postcode ontstaan. Het is bij een dergelijke implementatie aan te bevelen om extra punten toe te voegen aan de te trianguleren objecten, met name aan grote niet-infrastructuurpercelen zonder postcode, om zo een regelmatigere triangulatie te waarborgen. Nader onderzoek moet uitwijzen of dit alternatief een realistische optie is.

Een laatste alternatieve oplossing voor het probleem van de afzonderlijke gebieden met dezelfde postcode is het toekennen van meer dan één waarschijnlijke postcode aan de percelen zonder postcode. Als alle percelen één of meer mogelijke postcodes hebben, kan geprobeerd worden om postcodegebieden zo te berekenen dat percelen met dezelfde postcode met elkaar verbonden zijn. Hierbij kan dan de meest praktische postcode gekozen worden voor individuele percelen om zo tot verbonden postcodegebieden te komen in plaats van de meest waarschijnlijke postcode.

Conclusies

In dit artikel is een nieuwe methode gepresenteerd voor het afleiden van een planaire partitie van 6-positie postcodegebieden door van zowel geometrische als van administratieve gegevens uit de kadastrale registratie gebruik te maken. Infrastructuurpercelen worden gesplitst op de hartlijnen, waarna de delen afzonderlijk worden toegevoegd aan de meest waarschijnlijke postcodegebieden.

De methode is getest met behulp van een bestand van één kadastrale gemeente. De resultaten zijn veelbelovend: een beter kaartbeeld in vergelijking met de vaak willekeurige grenzen van de op Thiessen-polygonen gebaseerde methoden, met name doordat de gebruiker de ligging van de infrastructuur duidelijker herkent. Uit visuele controle blijkt bovendien dat slechts in uitzonderingsgevallen gebouwen nog doorsneden worden door de grenzen van postcodegebieden. De kwaliteitsindicator is toegevoegd om de relatieve betrouwbaarheid van de toegekende postcode weer te geven. Een duidelijk nadeel van de procedure van toekennen van postcodes is dat er afzonderlijke gebieden met dezelfde postcode kunnen ontstaan. Dit zou best kunnen overeenstemmen met de werkelijkheid, maar zal voor een aantal toepassingen toch zoveel mogelijk vermeden moeten worden.

Ten aanzien van het gekozen implementatieplatform kan geconcludeerd worden dat wanneer een GIS-pakket voor een ruimtelijk probleem niet volstaat, bijvoorbeeld wegens prestatie of ontbrekende functionaliteit,

CGAL een goed alternatief kan zijn. Hoewel CGAL veel mogelijkheden biedt, is bij de gebruiker in ieder geval ervaring met C++ vereist. ■

Literatuur

- [1] Ai, T. and P.J.M. van Oosterom, *GAP-Tree Extensions Based on Skeletons*. In: D. Richardson en P.J.M. van Oosterom (eds), *Advances in Spatial Data Handling*, 10th International Symposium on Spatial Data handling. Berlin: Springer-Verlag, 2002, p. 501-513.
- [2] Chithambaram, R. et al, *Skeletonizing polygons for map generalization*. In: Technical papers, ACSM-ASPRS Convention, Cartography and GIS/LIS, Volume 2. Bethesda: ACSM, 1991, p. 44-54.
- [3] DeLucia, A. and T. Black, *A comprehensive approach to automatic feature generalization*. In: Proceedings of 13th International Cartographic Conference. International Cartographic Association, 1987, p. 169-191.
- [4] Jones, C.B. et al, *Map Generalization with a Triangulated Data Structure*. *Cartography and Geographic Information Systems*, 22-4 (1995), p. 317-331.
- [5] Schut, T.G., en B. Maessen, *Haalbaarheid 6PPC-vlakkenkaart (eindrapportage)*. Apeldoorn: Kadaster, 2000.
- [6] Uitermark, H. et al, *Semantic and Geometric Aspects of Integrating Road networks*. In: A. Vckovski et al (eds), *Interoperating geographic information system: second international conference, INTEROP '99*. Berlin: Springer-Verlag, 1999, p. 177-188.

Samenvatting

Dit artikel beschrijft de procedure om op basis van de kadastrale registratie te komen tot een betrouwbare en goed leesbare planaire partitie van 6-positie postcodegebieden. Problemen van bestaande postcodebestanden, zoals het doorsnijden van bebouwing en vaak willekeurige ligging van de grenzen, zijn vermeden. Om tot een volledige planaire partitie te komen, zal aan alle percelen waarvan geen postcode bekend is, de meest waarschijnlijke postcode worden toegekend. Hierbij is speciale aandacht voor infrastructuurpercelen. Deze percelen worden op de hartlijn gesplitst en de twee helften worden ieder afzonderlijk aan het aangrenzende postcodegebied gekoppeld. Deze nieuwe aanpak resulteert in nauwkeurige postcodebestanden die bovendien visueel aantrekkelijk zijn.

TREFWOORDEN

GIS-technologie, gegevensbestanden, systeembeschrijving

KEYWORDS

GIS-technology, databases, system description