

Shapefile-formaatti

3D-Win / Tiedosto / Formaattit / Vektoritiedosto / GIS

Avainsanat: Shapefile, Shp, GIS-ohjaintiedosto

Artikkelin toiminnot 3D-Win 6.5.1 alkaen

Julkaistu 18.01.2019, Mikko Syrjä

Shapefile-formaatti on Esrin kehittämä tiedostomuoto GIS-datan siirtämiseen ja on hyvin laajalti tuettu paikkatieto- ja mittaussovelluksissa. Monasti se on ainoa tapa siirtää tietoa joidenkin järjestelmien välillä. Formaatti on avoin ja dokumentoitu, mutta binäärisenä hieman hankala käsitellä. Hyvä yleiskäyttö rakenteesta löytyy englanninkielisestä [wikipediasta](#). Seuraavassa keskitytään formaatin käyttöön 3D-Winin kanssa.

Formaatti koostuu useista tiedostoista. Tiedosto päätteellä .shp sisältää kohteiden geometrian ja ominaisuudet ovat päätteellä .dbf varustetussa tietokantataulussa. Molemmista tiedostoista ovat kohteet samassa järjestyksessä, mutta muuta linkkiä niiden välillä ei ole. Toisesta tiedostosta puuttuvasta kohteesta seuraa siis väärät ominaisuustiedot kaikille seuraaville kohteille. 3D-Win antaa luettaessa virheilmoituksen, jos tiedostoissa on eri määrä kohteita. Jos dbf-tiedosto puuttuu, luetaan vain geometria ilman ominaisuuksia. Kirjoitettaessa ominaisuustiedosto kirjoitetaan aina.

Lisäksi voi löytyä esimerkiksi koordinaattijärjestelmän määrittely päätteellä .prj ja indeksitiedosto päätteellä .shx. Indeksitiedosto tarvitaan lukemiseen, mutta se luodaan tarvittaessa automaattisesti samaan hakemistoon muiden tiedostojen kanssa.

Geometria

Shapefile tukee kolmea geometristä perustyyppiä: point (piste), polyline (taiteviiva) and polygon (sulkeutuva alue). Näistä kustakin on vielä erikseen 2D- ja 3D-versiot. Lisäksi on joitain muita tyyppisiä (esim. multipoint), jotka luetaan muuntamalle ne joksikin edellä mainituista tyypeistä. Kirjoitus tapahtuu aina jollain kolmesta perustyyppistä.

Yhdessä shp-tiedostossa on aina vain yhtä geometristä tyyppiä eli pisteet ja viivat on kirjoitettava eri tiedostoihin. Pisteiden tai viivojen kirjoitus valitaan kirjoitusmuuntimen Muut-kohdan asetuksista. Jos kirjoitettavaksi on valittu viivat, muunnin tarkistaa tiedoston kaikki taiteviivat. Jos kaikki viivat ovat sulkeutuvia, kirjoitetaan tiedosto polygon-kohteina. Muuten kaikki viivat kirjoitetaan polyline-tyyppisinä.

2D-tiedostoissa on korkeus joskus mukana ominaisuustietona. Se voidaan luettaessa siirtää Z-koordinaatiksi muuntimen kenttäasetusten kohdassa Z. Kirjoitettaessa 2D / 3D valitaan muuntimen XYZ-kohdassa olevan Kirjoita Z -asetuksen mukaan. Jos asetus on päällä, kirjoitetaan 3D-muotoinen tiedosto. Muussa tapauksessa kirjoitetaan 2D-muotoinen tiedosto eli vain koordinaatit X ja Y.

Ominaisuudet

Koska kutakin shp-tiedostoa vastaa aina yksi tietokantatyyppinen dbf-taulu, on kaikilla tiedostojen kohteilla aina samat ominaisuudet, joiden arvot luetaan taulun sarakkeista. Kirjoitettaessa on käytettävä [GIS-ohjaintiedostoa](#), joka määrittää taulun sarakkeiden nimet ja tyytit. Kuten muutkin GIS-formaattit, Shapefilen luku luo käyttäjän asetushakemistoon tiedostot auto_attr.dat ja auto_attr.txt, joista ensimmäinen on dbf-tilin ominaisuuksien perusteella automaattisesti rakennettu ohjaintiedosto.

Taulun sarakkeiden nimien pituus on rajoitettu kymmeneen merkkiin. Kirjoitettaessa ominaisuuksien nimet katkaistaan automaattisesti tähän rajaan. Jos katkaisun seurauksena kahdella sarakkeella on sama nimi, annetaan virheilmoitus. Tässä tapauksessa on käytettävä ominaisuuksien nimien muunnostiedostoa muuntimen kenttäasetuksissa. Tiedostossa annetaan alkuperäinen ominaisuuden nimi vasemmassa sarakkeessa ja lyhennetty nimi oikeassa sarakkeessa.

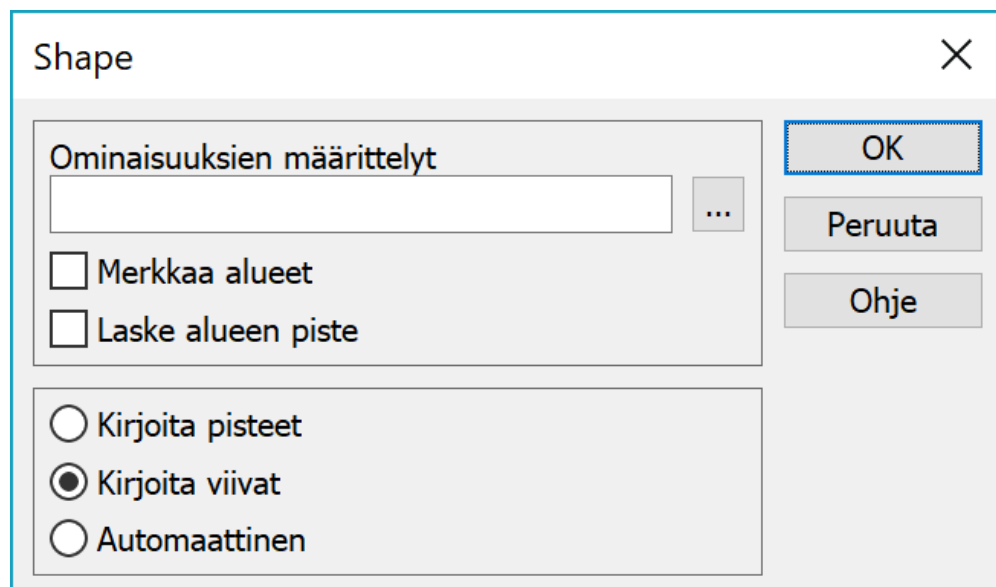
Teknisesti ominaisuudet sisältävä dbf-tiedosto on vanhan dBase-tietokantaohjelman kannan tauluille käyttämä muoto, jota tukevat monet nykyisetkin ohjelmat. Esimerkiksi Excel ymmärtää sen, kun tiedostomuodoksi valitaan avatessa dBase files (*.dbf). Tämä on usein näppärä tapa tarkistaa mitä ominaisuuksia tiedostossa on.

Koordinaatisto

Jos kirjoitettavan aineiston koordinaatisto on tunnettu, ohjelma kirjoittaa myös koordinaatiston määrittelytiedoston päätteellä .prj. Esimerkiksi GK27-koordinaatiston määrittely:

Muunnin

Luku- ja kirjoitusmuunnin löytyy 3D-Winin formaatinmuuntainen funktiolistasta nimellä Shape. Muunnin noudattaa samoja sääntöjä kuin muutkin GIS-muuntimet. Koska geometrialtaan ja ominaisuuksiltaan erilaiset kohteet ovat eri tiedostoissa, joudutaan yleensä avaamaan samalla kertaa useita shp-tiedostoja. Nämä voidaan sitten tarvittaessa yhdistää elementtistassa.



Ominaisuuksien määrittelyt -asetuksessa määrätään GIS-ohjaintiedosto, jota tarvitaan kirjoitettaessa määrittelemään dbf-aulun sarakkeet. Tämän käytöstä on tarkempi selvitys alempana.

Merkkaa alueet -asetus tekee kaikista alunperin polygon-tyyppisistä taiteviivoista alueita. Muita sulkeutuvia taiteviivoja ei koskaan merkata alueiksi. Laske alueen piste -asetus laskee alueille automaattisesti referenssipisteen alueen sisäpuolelle.

Alaosan asetuksilla valitaan kirjoitetaanko pisteet vai viivat. Automaattiasetus tarkistaa kirjoitettavan tiedoston sisällön ja kirjoittaa viivat, jos tiedostossa on yksikin taiteviiva. Muuten kirjoitetaan hajapisteet.

Ohjaintiedosto

Shapefilen GIS-ohjaintiedosto tukee useita osioita. Kussakin osiossa määritellään kirjoitettavat ominaisuudet hakasuluissa annetulle koodille. Yksittäisen koodin sijasta hakasuluissa voidaan antaa myös mikä tahansa ominaisuuhauun mukainen hakuehto (esim. 4?? tai #TYPE=POINT). Näin samaa osiota voidaan käyttää usealle samantyyppiselle koodille. Esimerkiksi kaikilla kaivoilla tai putkilla voi olla samat ominaisuudet. Jos koodiin täsmäävää osiota ei löydy, käytetään oletuksena ensimmäistä osiota.

Kirjoitettaessa käytettävä osio määräytyy tiedoston ensimmäisen pisteen tai viivan perusteella. Moniosainen ohjaintiedosto ei siis automaattisesti jaa tiedostoa erityyppisiin koodeihin, vaan ainoastaan mahdollistaa saman muuntimen ja ohjaintiedoston käytön eri tyyppisten tiedostojen kirjoitukseen. Jos tiedosto sisältää eri tyyppisiä kohteita, ne on ensin eroteltava esimerkiksi toiminnolla Tiedosto / Elementit / Jaa osiin.

Seuraavassa esimerkissä on erilliset määrittelyt kaivoille (pistemäiset kolminumeroiset neljällä alkavat koodit: #TYPE=POINT;4??) ja putkille (viivamaiset kolminumeroiset neljällä alkavat koodit: #TYPE=LINE;4??). Lisäksi alussa on oletusmäärittely kaikille muille kohteille.

```
[ ] T3 string 12
[ #TYPE=POINT;4?? ]
T3 string 12
DIAMETER real 10 1
MATERIAL string 32
COVER string 32

[ #TYPE=LINE;4?? ]
T3 string 12
DIAMETER real 10 1
MATERIAL string 32
```

Kaivoille kirjoitetaan ominaisuuksina koodi, halkaisija, materiaali ja kannen tyyppi. Vastaavasti putkille kirjoitetaan koodi, halkaisija ja materiaali. Kaikille muille kohteille kirjoitetaan vain koodi. Halkaisijat kirjoitetaan reaalityyppinä ja muut ominaisuudet tekstinä.

Jos esimerkiksi tiedostossa on vain kaivoja ja putkia, se voidaan jakaa osiin laittamalla jakotoiminnon kohtaan Koodi arvo #TYPE. Se jakaa tiedoston tyyppin mukaan eli toiseen elementtiin siirtyvät hajapisteet ja toiseen viivat. Syntyneet kaksi tiedostoa voidaan sitten valita elementtistasta ja kirjoittaa pikavalikon toiminnolla Kirjoita. Listasta valitaan Shapefile-muunnin, jonka asetuksiin on laitettu yllä oleva ohjaintiedosto ja kirjoitusgeometriaksi valittu Automaattinen.

Muunnin kirjoittaa molemmat tiedostot ja valitsee kummallekin automaattisesti käytettävät geometrian ja ohjaintiedoston osion. Tuloksena esimerkiksi kirjoitettujen putkien ominaisuuksien dbf-taulu luettuna Exceliin:

B3				239
	A	B	C	
1	T3	DIAMETER	MATERIAL	
2	409	239,0	PVC	
3	409	239,0	PVC	
4	409	239,0	PVC	
5	409	239,0	PVC	
6	409	239,0	PVC	
7	409	239,0	PVC	
8	409	239,0	PVC	

Kullakin rivillä on yksittäisen viivan ominaisuudet. Excel säätää sarakkeiden tyytit, leveydet ja desimaalien määrän ohjaintiedoston kirjoittamien asetusten mukaan.

Muuntimen ohjesivu: [Shape vektoriformaatti](#)

Tiedoston jako osiin: [Tiedoston jako](#)

Lue lisää: [GIS-ominaisuudet](#)

Tekninen dokumentaatio: [ESRI Shapefile Technical Description](#)