

GIS in het Web: hoe werkt het?

Barend Köbben en Edward Mac Gillavry

Barend Köbben (kobben@itc.nl) is docent bij de divisie Geoinformatica, Kartografie en Visualisatie van het ITC, Postbus 6, 7500 AA te Enschede. Edward Mac Gillavry (macgillavry@itc.nl) volgt de MSc. Cartography opleiding, een gezamenlijke inspanning van de disciplinegroep Kartografie van de Universiteit Utrecht en de divisie Geoinformatica, Kartografie en Visualisatie van het ITC.

Dit artikel is bedoeld voor wie geïnteresseerd is in het publiceren van GIS data op het World Wide Web. Er wordt duidelijk gemaakt hoe het Web werkt en hoe het gebruikt kan worden om de gebruiker met kartografische informatie in een GIS omgeving te laten werken. Een viertal commerciële oplossingen om GIS data op het Web te publiceren wordt kort voorgesteld en er wordt verslag gedaan van een onderzoek dat onlangs gedaan is naar de geschiktheid ervan voor kartografische toepassingen.

Keywords

gebruik van geografische informatie, programmatuur, GIS/LIS, overzicht

Abstract

This article addresses those who want to publish their GIS data on the World Wide Web. A short overview is given of how the Web sprouted from the Internet and of the file formats and communication protocols that are its building blocks. It will become clear how the basic set-up allows the publication only of static maps with very limited interactivity and how the possibilities can be broadened by adding server-side applications or client-side plugins and applications such as Java applets or scripts. Nowadays, most GIS software vendors market a solution to offer mapping and GIS functionality on the Web. Four of these are briefly introduced and a recent benchmark to assess their usefulness for cartographic applications is presented.

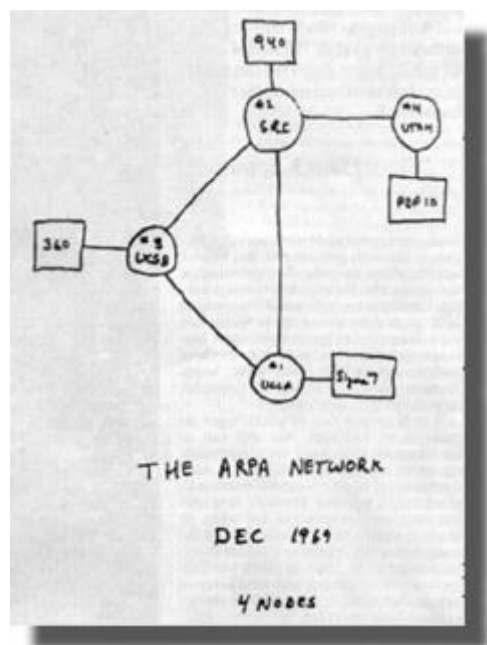
Inleiding

Het World Wide Web (WWW of kortweg het Web) is een relatief nieuw medium en nu al worden er naar schatting elke dag zo'n 50 miljoen kaarten op geraadpleegd (Peterson, 1999). Hoeveel daarvan afkomstig zijn uit of onderdeel zijn van een GIS, is niet gespecificeerd, maar het zal een aanzienlijk deel zijn. De gebruikers van die enorme hoeveelheid kaarten zijn waarschijnlijk voor het grootste deel niet geïnteresseerd in hoe ze nu feitelijk geproduceerd worden, en terecht! Maar voor diegenen die zelf hun GIS-data willen publiceren zijn de technische achtergronden wel degelijk van belang. Inmiddels bieden zo'n beetje alle leveranciers van GIS-software een oplossing om GIS-data en functionaliteit op het Web te realiseren.

Internet en het World Wide Web

Om te begrijpen wat het World Wide Web nu precies is, moeten we als eerste inzien dat het in feite een sub-set is van het zogenaamde Internet. Dat Internet is op zijn beurt ontstaan uit een bescheiden eerste computer netwerk, ARPAnet geheten. Dit is indertijd gecreëerd door het Advanced Research Projects Agency en werd gefinancierd door het Amerikaanse ministerie van Defensie. Het diende twee doelen: het maakte het mogelijk dat wetenschappers die op verschillende plekken werkten hun digitale informatie konden uitwisselen. Tegelijkertijd zorgde het ervoor dat die belangrijke informatie verspreid werd over een aantal locaties, waardoor de kwetsbaarheid voor nucleaire aanvallen verminderd werd. Het ARPAnet was opgebouwd als een gedistribueerd,

gedecentraliseerd netwerk waarin elk knooppunt even belangrijk was. Om de computers met elkaar te laten communiceren was het nodig dat ze allemaal dezelfde 'taal' spraken. Daarvoor werd een zogenaamd *communicatieprotocol* gebruikt, Network Communication Protocol (NCP) geheten. In zo'n protocol is vastgelegd hoe de systemen informatie van elkaar kunnen vragen en die vervolgens naar elkaar kunnen sturen. Eind 1969 verbond deze 'grootmoeder van het Internet' vier computers in de Verenigde Staten. Zoals in figuur 1 te zien is, werd dit netwerk ook in kaart gebracht.



Figuur : Schets van het eerste ARPAnet (uit: Salus, 1995).

Langzamerhand groeide ARPAnet uit tot een veelomvattend netwerk en kwamen ook anderen tot ontwikkeling, met als belangrijkste USENET en EUnet. In 1983 schakelden deze netwerken over op een nieuw communicatieprotocol, genaamd TCP/IP, wat staat voor Transmission Control Protocol/Internet Protocol. Het sterke punt van dit protocol is, dat het gelaagd is en daarom geschikt om weer andere protocollen als het ware 'in te pakken en op te sturen'. Op die manier ondersteunde dit TCP/IP bestaande protocollen als TelNet (voor terminal besturing), FTP (voor het verzenden van bestanden) en SMTP (voor het versturen van email) en kon het uitgebreid worden met allerlei nieuwe protocollen. Dit wordt algemeen gezien als de geboorte van het Internet (URL 1).

Naast het Internet bestaat er dan nog, ietwat verwarrend, dat 'World Wide Web'. Dit is oorspronkelijk opgezet door Tim Berners-Lee van CERN (het Centre Européen de Recherches Nucléaires) om de uitwisseling van onderzoeksresultaten via het Internet te verbeteren en vergemakkelijken. Daarvoor was het niet voldoende om bestanden uit te kunnen wisselen, er moesten ook standaarden komen om de informatie op te slaan en af te beelden. Daartoe werd de HyperText Markup Language (HTML) ontworpen. Het formaat van HTML en een nieuw communicatieprotocol HTTP (HyperText Transfer Protocol), speciaal voor het zoeken en ophalen van HTML bestanden van verschillende locaties, werden vastgelegd in 1994 (URL 2). Ook dit HTTP werkte weer 'ingepakt in' TCP/IP. Daarom kan het Web gezien worden als een sub-set van het Internet: dat deel ervan dat gebruik maakt van het HTTP communicatieprotocol en de bestandsformaten die in de Web standaarden zijn vastgelegd.

Standaard opzet

Eén van de sterke punten van het Web is de zogeheten *platformafhankelijkheid* van de informatie. De data wordt gedistribueerd door 'server' software op computers die draaien op bijvoorbeeld AppleOS, Unix, Linux, BeOS of WindowsNT, om er maar een paar te noemen. Om die data te kunnen ophalen en bekijken, zijn er zogeheten 'browsers' beschikbaar voor vrijwel elk computertype, inclusief spelcomputers en mobiele telefoons. De meest bekende browsers zijn Netscape navigator en Microsoft Internet Explorer. Het hiervoor beschreven TCP/IP protocol zorgt er voor dat al die systemen met elkaar kunnen communiceren. Maar om het mogelijk te maken dat die informatie ook door elke browsers kan worden geïnterpreteerd en correct afgebeeld, is het ook nodig dat de data wordt opgeslagen in een aantal gestandaardiseerde *bestandsformaten*.

HTML

Het belangrijkste Web formaat is het eerder genoemde HTML. Dit is een niet-commercieel, zogeheten *open formaat*, dat opgeslagen is als leesbare tekst. Daardoor kan het bewerkt worden met allerlei software, van eenvoudige tekstverwerkertjes tot speciale 'HTML-editors'. Vanuit de meeste standaard 'office' programma's (zoals Microsoft Office of WordPerfect Office) kunnen documenten tegenwoordig rechtstreeks naar HTML geëxporteerd worden. HTML is feitelijk standaard tekst, afgewisseld met zogeheten 'tags'

die aanwijzingen geven voor de opmaak van de tekst én de functionaliteit.
Zo wordt bijvoorbeeld de HTML code:

```
<P>Hallo <B>daar</B><HR>Hoe gaat  
het met <I>jou</I>?<P>  
<A HREF= "anderbestand.html">Klik  
hier</A>om verder te gaan.
```

afgebeeld als:

Hallo **daar**

Hoe gaat het met *jou*?

[Klik hier](#) om verder te gaan.

Bij het klikken op de onderstreepte woorden "Klik hier" zal de browser een ander HTML bestand laden (opgeslagen als "anderbestand.html" en dat op zijn beurt afbeelden.

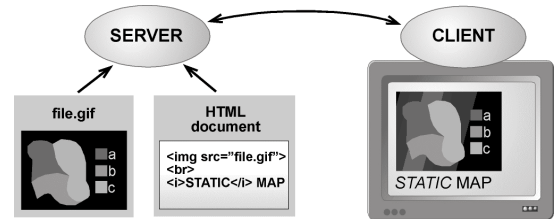
Plaatjes: GIF en JPEG

Om plaatjes te kunnen afbeelden wordt een zogenaamde 'tag' opgenomen, die verwijst naar een bestand dat het plaatje bevat, dat vervolgens wordt geladen en in het momenteel afgebeelde HTML bestand wordt opgenomen (vandaar dat deze 'inline images' worden genoemd). Momenteel worden volgens de standaard alleen rasterplaatjes ondersteund, opgeslagen in GIF formaat (vooral geschikt voor plaatjes met grotere vlakken in volle kleuren, zoals grafieken en kaarten) of JPEG formaat (vooral geschikt voor foto's en andere halftoon grafiek).

Standaard Web publicatie in werking

De combinatie van de server software met de browsers samen met de standaard bestandsformaten HTML, GIF en JPEG, beperkt de publicatie van informatie op het Web tot *statische informatie*. De gebruiker kan alleen data bekijken in de vorm van tekst en rasterplaatjes, zoals die opgeslagen staan op één of meerdere servers. Als iemand bijvoorbeeld een thematische kaart van statistische data wil publiceren, zal hij die kaart eerst moeten construeren, vervolgens moeten omzetten naar GIF formaat, ernaar moeten verwijzen in een HTML bestand, en vervolgens deze twee bestanden op een server zetten. Daardoor zijn elementen als de lay-out, de classificatie en de kleurkeuze vastgelegd en kan de gebruiker deze niet aanpassen. Als er nieuwe gegevens beschikbaar komen, moet het hele proces herhaald

worden om de geactualiseerde kaart te kunnen publiceren. In figuur 2 zien we deze standaard opzet.



Figuur 2: Opzet van standaard Web publicatie

Deze standaard opzet werkt in principe met een serie vragen en antwoorden. Het begint met de gebruiker die het adres invoert in de browser van de informatie die hij wil zien. Dat adres is in de vorm van een zogeheten 'Universal Resource Locator' (URL), zoals <http://kartoweb.itc.nl/index.html>. De URL kan direct ingetypt zijn of afkomstig uit het klikken op een hyperlink in een eerder HTML document. Het resultaat is dat de browser een verzoek verzendt naar de server op het adres 'kartoweb.itc.nl' om het bestand 'index.html' op te sturen. De server zal reageren door het gevraagde bestand terug te sturen, waarna de browser het HTML file zal lezen en de 'tags' die erin staan zal interpreteren en het document zal afbeelden. Wanneer, zoals in het voorbeeld in figuur 2, de 'tag' voor een plaatje wordt gevonden, wordt een aanvullend verzoek voor het betreffende raster bestand (te vinden op het locatie in de SRC= parameter) gedaan. Dit wordt op zijn beurt verstuurd, ontvangen en afgebeeld, waarna de interpretatie van de HTML code weer verder gaat.

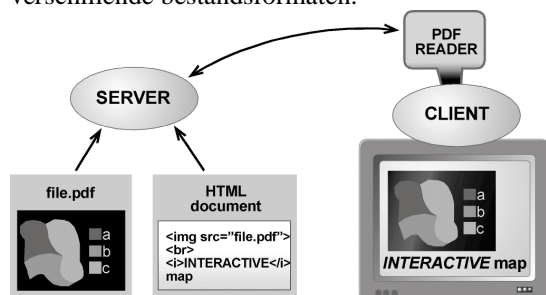
Uitbreiden van de functionaliteit

De in de vorige paragraaf beschreven configuratie voldoet volledig aan de officiële standaarden voor het Web, zoals beschreven door het World Wide Web Consortium (W3C, zie URL 3). Voor kartografische en GIS toepassingen is die opzet behoorlijk beperkend, vooral omdat voor grafische toepassing slechts de rasterformaten JPEG en GIF beschikbaar zijn. Daardoor kunnen kaarten niet interactief zijn (behalve dat bepaalde locaties in de kaart links naar andere Webpagina's kunnen bevatten), is de grafische kwaliteit zeer beperkt en kunnen kaartbeelden niet uit databases gegenereerd worden. Om deze beperkingen te overwinnen, zijn er vele mogelijkheden: aan de kant van de browser ('client-side') kan

extra functionaliteit worden ingebouwd, of aan de kant van de Webserver ('server-side') of er kan een combinatie van beide worden gebruikt. Het is wel belangrijk je te realiseren dat alle hierna beschreven oplossingen uitbreidingen zijn van de standaarden en daarom de platformafhankelijkheid ondermijnen en daarom de gepubliceerde informatie minder algemeen bruikbaar maken.

Plugins

Eén van de eenvoudigste en meest gebruikte manieren om de Webfunctionaliteit uit te breiden is het gebruik van *plugins*. Een plugin is een stukje software dat samenwerkt met de browser software en er voor zorgt dat de browser extra bestandsformaten kan interpreteren naast de standaard HTML, GIF en JPEG files. Er zijn vele duizenden plugins beschikbaar voor bijna net zoveel verschillende bestandsformaten.



Figuur 3: Opzet van Web publicatie met gebruik van een plugin voor PDF bestanden

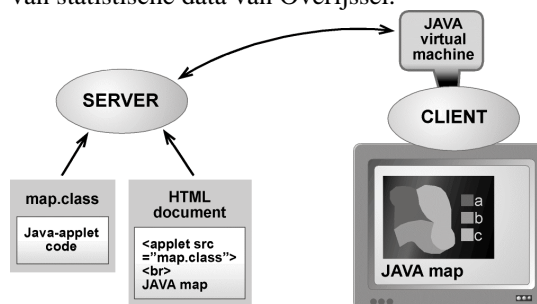
Het voorbeeld in figuur 3 laat het gebruik zien van de plugin die Adobe heeft ontwikkeld om haar Portable Document Format (PDF, een populair op Postscript gebaseerd vectorformaat) in browsers te kunnen gebruiken. Vergelijking van de figuren 2 en 3 maakt duidelijk dat het enige verschil met de standaardopzet de toevoeging van de plugin aan de browser is en de afwijkende extensie van de bestandsnaam in de "SRC=" parameter van de 'tag'. Door deze extensie kan de browser dat bestand herkennen als 'niet standaard' en de controle over het afbeelden van en de interactie met de inhoud overdragen aan de PDF-plugin. Deze kan vervolgens alle extra functionaliteit van het PDF-formaat bieden aan de gebruiker: zooming, panning, printen in hoge kwaliteit, selecteren van en zoeken naar tekst, enzovoort.

Het voordeel van het plugin systeem is de verkrijgbaarheid van plugins voor veel bestaande grafische bestandsformaten, meestal gratis beschikbaar gesteld door de makers van het betreffende bestandsformaat. Het belangrijkste nadeel

is de noodzakelijkheid van het 'downloaden' en installeren van steeds nieuwe (updates van) plugins. Bovendien is elke plugin alleen geschikt is voor één bepaald besturingssysteem en meestal ook maar voor één bepaalde browser.

Java applets en scripts

Het gebruik van Java op het Web is in korte tijd zeer populair geworden. Een belangrijke reden daarvoor is het feit dat Java echt platformafhankelijk is en dat de meeste browsers standaard voorzien zijn van Java interpreters (zogenaamde 'virtual machines'). Zo kan de functionaliteit van een Web pagina eenvoudig worden uitgebreid door een klein Java programmaatje (dat wordt een 'applet' genoemd) aan te roepen in de HTML code. Dit wordt vervolgens van de server gehaald en uitgevoerd door de 'virtual machine' van de browser (zie figuur 4). Ook kan een variant van Java code, een zogeheten 'JavaScript', direct worden opgenomen in een HTML document. De functionaliteit van Java kan variëren van eenvoudige interactiviteit van knoppen, zodat bijvoorbeeld de grafische weergave ervan verandert als je de muis erover beweegt, tot een uitgebreide kartografische omgeving. Een fraai voorbeeld van dat laatste is het Descartes systeem, dat exploratieve grafische analyse van ruimtelijke data mogelijk maakt, zoals in het voorbeeld in URL 4 van statistische data van Overijssel.

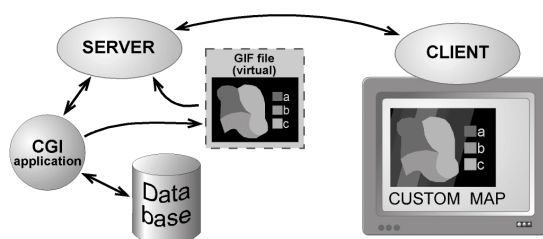


Figuur 4: Opzet van Web publicatie met gebruik van een Java applet

'Server-side' functionaliteit

De hiervoor beschreven oplossingen zijn voorbeelden van het uitbreiden van de functionaliteit van de browser, de zogeheten 'client-side' oplossingen. Maar ook aan de 'server-side' kunnen programma's worden ingezet, die samenwerken met de Webserver software. Die samenwerking wordt in de meeste gevallen gerealiseerd door gebruik te maken van het Common Gateway Interface (CGI) protocol. Dit definieert de communicatie tussen de communicatie tussen webserver en de applicatie, maar ook de manier waarop de browser informatie kan doorspelen aan de appli-

catie via de Webserver. Dat laatste wordt bereikt door aan de URL commando's en parameters toe te voegen. Een vraagteken wordt gebruikt om het CGI deel van de rest van de URL te scheiden en ampersands om de parameters van elkaar te scheiden (bijvoorbeeld: `http://kartoweb .itc.nl?doe_iets¶meter1¶meter2`). CGI programma's kunnen gebruikt worden voor een enorme verscheidenheid aan taken, bijvoorbeeld toegang bieden tot databases en het genereren van kaarten op basis van gegevens uit diezelfde databases, zoals in de opzet in figuur 5.



Figuur 5: Opzet van Web publicatie met gebruik van een CGI applicatie

Hier wordt door de CGI applicatie op basis van de invoer van de gebruiker uit een database een kaart gegenereerd en verstuurd in de vorm van een GIF plaatje (dat nooit als bestand opgeslagen wordt, vandaar de term 'virtual GIF file'). De 'National Atlas of Canada Online' (URL 5) gebruikte in vroegere edities een dergelijke opzet met volledig zelfgeschreven software, omdat dit in 1993, toen het systeem werd ontworpen, de enige mogelijkheid was. In de huidige 6e versie wordt gebruik gemaakt van ESRI technologie (Frappier & Williams, 1999).

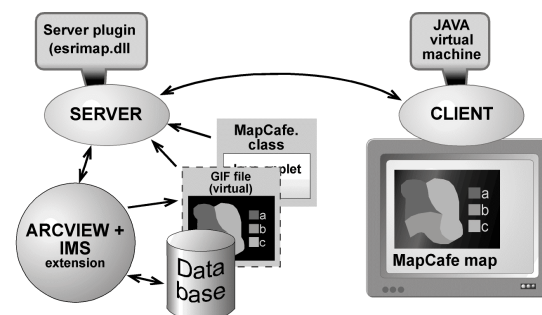
Het grote voordeel van dergelijke 'server-side' oplossingen is dat de gebruiker genoeg heeft aan een standaard browser en toch over uitgebreide kartografische en/of GIS functies kan beschikken, welke browser op welk besturingssysteem hij ook gebruikt. Het grote nadeel is de belasting van de server. Op populaire sites kan het aantal verzoeken aan de server ('hits') oplopen tot honderden per minuut. De Webserver software heeft dan de verhoudingsgewijs makkelijke taak om daarop te reageren met het verzenden van honderden bestanden, maar de CGI software moet even zo vele keren een complete kaart genereren uit de database!

GIS in het Web: benchmark van vier applicaties

Door de enorme populariteit van het Web als middel om ruimtelijke informatie te verspreiden, konden de leveranciers van software voor kartografie en GIS er natuurlijk niet onderuit om daar mogelijkheden voor aan te bieden. Vrijwel alle bekende firma's hebben dan ook hun oplossingen op de markt gebracht. In het kader van een doctoraalscriptie is onlangs een en vergelijkend onderzoek vanuit kartografisch oogpunt gedaan naar een viertal van de belangrijkste 'WebGIS' oplossingen, nl.:

- ArcView Internet Map Server 1.0 (ESRI)
- MapXtreme 2.0 (MapInfo Corporation)
- MapGuide 4.0 (Autodesk)
- Geomedia Webmap 3.0 (Intergraph)

Om de extra functionaliteit van een GIS omgeving op het Web te kunnen bieden maken alle vier gebruik van de in eerdere paragrafen genoemde technieken, vaak in een combinatie van 'client-side' uitbreidingen zoals plugins en Java met 'server-side' applicaties. Als voorbeeld is in figuur 6 de opzet van ArcView Internet Map Server (ArcView IMS) te zien.



Figuur 6: Opzet van ESRI's ArcView Internet Map Server

Hierbij fungeert het standaard ArcView programma, uitgebreid met de IMS extensie, als CGI applicatie. Op die manier kan ArcView 'op afstand worden bestuurd' door vanuit de browser commando's in CGI formaat te geven. Die commando's worden gegenereerd door een Java applet (MapCafé geheten), die de gebruiker een ArcView-achtige interface biedt. Als de gebruiker bijvoorbeeld op de zoom-knop klikt, verstuurt de applet het juiste CGI commando, dat door ArcView wordt uitgevoerd. De resulterende, ingezoomde, kaart wordt vervolgens omgezet naar een virtueel GIF plaatje dat weer naar de browser wordt teruggestuurd en door de applet wordt af-

gebeeld. Het laatste onderdeel is een speciale plugin (esrimap.dll) in de Webserver software, die ervoor zorgt dat vanaf één Webserver meerdere ArcView servers kunnen worden aangesproken, om zo de server belasting te kunnen spreiden.

De aangeboden functionaliteit bij standaard installatie verschilt nogal tussen de vier applicaties. Uitbreiden van de zogenaamde 'out of the box' opzet is in alle gevallen mogelijk, variërend van eenvoudig aanpassen van instellingen, via het verfijnen van de mogelijkheden door HTML codering tot het vergaand (her)programmeren van Java applets. Alle oplossingen hebben gemeen dat er (nog) weinig aandacht voor kartografie in de pakketten is. Zo is er een beperkte kleurcontrole: geen 'web-save' kleurenpalet, grijswaarde-trappen die niet zelf te definiëren zijn, geen CYMK-kleuren, et cetera.

Benchmark taken

Om de pakketten te kunnen vergelijken werd een twaalfstal taken ('benchmarks') opgesteld die betrekking hadden op drie fictieve toepassingen. Als eerste was er een algemene GIS toepassing, waarin standaard GIS bewerkingen, zoals in- en uitvoer van ruimtelijke data in een aantal veel gebruikte bestandsformaten, projectie, topografische en thematische visualisatie op het scherm en op hardcopy.

De tweede toepassing was een meer kartografische, gericht op communicatie naar het brede publiek. Hier werd vooral getest op de navigatie door, oriëntatie in en identificatie van ruimtelijke bestanden, met taken als tekstplaatsing, zoekfuncties, ondersteunde kaarttypen, kleurkeuze, en dergelijke. Hierbij werd er speciaal gelet in hoeverre de functionaliteit beschikbaar was voor een breed publiek, dus in standaard browsers op vele besturingssystemen, zonder gebruik van plugins.

Als laatste was er een toepassing voor gebruik in onderzoek en beleid. Hier werd getest op meer exploratieve kartografische mogelijkheden, waarbij zaken als toegang tot databases, queries, kaart- en datavergelijking, integratie in een onderzoeksomgeving en dergelijke onderzocht werden. Voor een dergelijke, meer gespecialiseerde gebruikersgroep werd het noodzakelijk zijn van een meer uitgebreide gebruikersconfiguratie, met plugins, specifieke hard- en software, et cetera, niet als bezwaar gezien.

Voor testen van de taken was een speciale gestandaardiseerde dataset opgezet met daarin ruimtelijke en statistische data van Overijssel.

Benchmark resultaten

De resultaten van de tests laten zien dat de vier pakketten duidelijk hun eigen sterke en zwakke punten hebben. Niet één ervan is duidelijk als beste uit de bus gekomen, maar bepaalde applicaties zijn wel beter geschikt voor bepaalde taken. In de algemene GIS toepassingen scoren ArcView IMS en MapGuide het beste, vooral omdat ze gebruikersvriendelijker zijn in deze taken dan GeoMedia en MapXtreme. Voor de communicatie naar een breed publiek blijken ArcView IMS en MapXtreme het meest geschikt. GeoMedia scoort hier wat minder en MapGuide is door het noodzakelijk gebruik van plugins hiervoor feitelijk niet geschikt. In de meer exploratieve taken waren het juist GeoMedia en MapGuide die de beste cijfers kregen. ArcView IMS werd ook redelijk bevonden, zolang de applicaties niet al te zware eisen stellen. MapXtreme scoorde hier minder goed.

Samenvattend kan gesteld worden dat er niet al te grote verschillen gevonden zijn. Als in een organisatie al GIS applicaties van één van de leveranciers gebruikt worden, zal de keuze voor de WebGIS oplossing van diezelfde fabrikant het meest voor de hand liggen. In andere gevallen kan worden gesteld dat de architectuur van ArcView IMS en MapXtreme zich het beste leent voor toepassingen op het algemeen toegankelijke Internet, waar GeoMedia en MapGuide vooral geschikt lijken voor Web toepassingen binnen organisaties (het zogenaamde 'Intranet').

Wie geïnteresseerd is in de volledige benchmark resultaten kan contact opnemen met de auteurs.

URLs

- URL 1: *A Brief History of the Internet*,
<http://www.isoc.org/internet-history/brief.html>
- URL 2: *A Short History of Internet Protocols at CERN*,
<http://www.info.cern.ch/pdp/ns/ben/tcphist.html>
- URL 3: *W3C-The World Wide Web Consortium*,
<http://www.w3.org/>
- URL 4: *Explorative maps of Overijssel using the Descartes system*,
<http://allanon.gmd.de/and/java/iris/app/itc/indexm.html>

URL 5: *National Atlas of Canada Online*,
<http://atlas.gc.ca/english/>

References

- Frappier, J. and Williams, D., 1999, An overview of the National Atlas of Canada. In: Proceedings of the 19th International Cartographic Conference ICC99, Ottawa, edited by Keller, C.P. (Ottawa: Canadian Institute of Geomatics), pp. 261-267. (See also: http://atlas.gc.ca/english/about_us/index.html#overview)
- Peterson, M.P., 1996, Cartographic animation on the Internet.. Proceedings of the joint ICA Commissions Seminar on Teaching Animated Cartography, Edited by Ormeling, F.J., Köbben, B and Perez Gomez, R (Utrecht: International Cartographic Association), pp. 11-14.
- Salus, P.H., 1995, Casting the Net: From ARPAnet to Internet and beyond, (Reading, MA: Addison-Wesley Publishing Company).