



Stichting NIOC

Stichting NIOC en de NIOC kennisbank

Stichting NIOC (www.nioc.nl) stelt zich conform zijn statuten tot doel: het realiseren van congressen over informatica onderwijs en voorts al hetgeen met een en ander rechtstreeks of zijdelings verband houdt of daartoe bevorderlijk kan zijn, alles in de ruimste zin des woords.

De stichting NIOC neemt de archivering van de resultaten van de congressen voor zijn rekening. De website www.nioc.nl ontsluit onder "Eerdere congressen" de gearchiveerde websites van eerdere congressen. De vele afzonderlijke congresbijdragen zijn opgenomen in een kennisbank die via dezelfde website onder "NIOC kennisbank" ontsloten wordt.

Op dit moment bevat de NIOC kennisbank alle bijdragen, incl. die van het laatste congres (NIOC2025, gehouden op donderdag 27 maart 2025 jl. en georganiseerd door Hogeschool Windesheim). Bij elkaar zo'n 1500 bijdragen!

We roepen je op, na het lezen van het document dat door jou is gedownload, de auteur(s) feedback te geven. Dit kan door je te registreren als gebruiker van de NIOC kennisbank. Na registratie krijg je bericht hoe in te loggen op de NIOC kennisbank.

Het eerstvolgende NIOC vindt plaats op 18 maart 2027 in Arnhem en wordt georganiseerd door HAN University of Applied Sciences.

Reacties over de NIOC kennisbank en de inhoud daarvan kun je richten aan de beheerder:

R. Smedinga kennisbank@nioc.nl.

Vermeld bij reacties jouw naam en telefoonnummer voor nader contact.

Open data ontsluiten

De SunnyApp – Open data ontsluiten via het dataspaceprincipe

Door: Peter Conradie, docent/onderzoeker, Creating 010; Hogeschool Rotterdam.

Met medewerking van: Judith Lemmens, Tony Busker, Sunil Choenni, Creating 010; Hogeschool Rotterdam.

Trefwoorden: data space, database, database management, open data.

Dit projectpaper beschrijft de ‘SunnyApp’ die als basis dient voor de ontwikkeling van een prototype dataspacestelsel waarmee open data kan worden ontsloten en gekoppeld. Het dataspaceprincipe kan complementair zijn aan de huidige catalogussystemen en voorziet in flexibele dataopslag- en verwerkingsmogelijkheden. Voor de ontwikkeling van de open data space wordt het dataspace principe gehanteerd en de SunnyApp als pilot case gebruikt. Deze applicatie geeft aan waar vrije zonnige plekken op Rotterdamse terrassen zijn en is als pilot case gekozen vanwege de verwerking van diverse soorten data.

1. Inleiding

Steeds meer overheden geven publieke-sector-informatie vrij voor hergebruik door derden (ook wel open data genoemd). Iedereen kan deze data kosteloos gebruiken, hergebruiken en verspreiden, op voorwaarden dat de bron wordt vermeld en dat de gegevens met een gelijksoortige licentie worden verspreid. Deze nieuwe openbare informatiebronnen zijn heel divers van inhoud, grootte en formaat. Denk bijvoorbeeld aan bestanden met statistische geografische gegevens, continu geactualiseerde verkeersdata, of historische meteorologische gegevens.

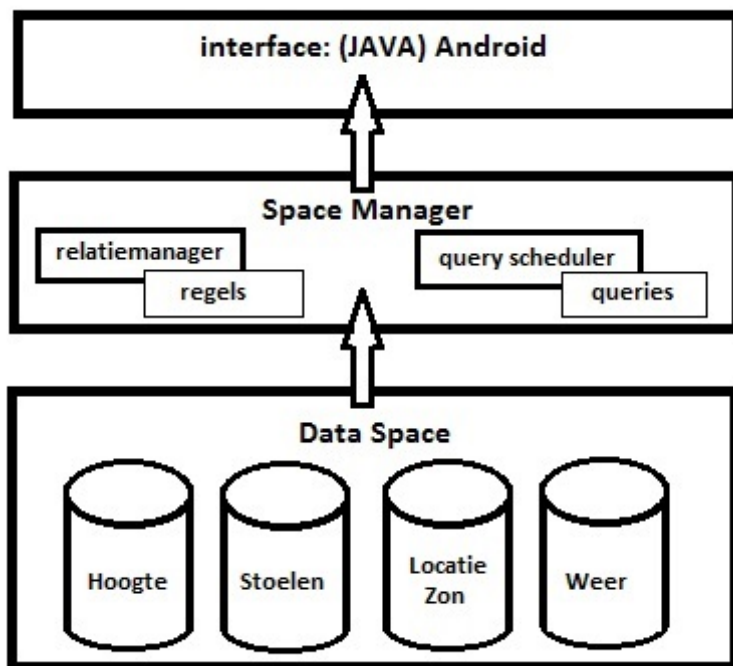
Publicatie van deze gegevens gebeurt veelal via online catalogussystemen, zoals Comprehensive Knowledge Archive Network (CKAN) of Microsoft Azure. De huidige versies van deze systemen verwijzen naar of indexerend databestanden. Ze fungeren als een catalogus, waarbij de bron van de data in veel gevallen bij de eigenaar blijft. Hierdoor zijn de opslag- en verwerkingsmogelijkheden beperkt. Om deze mogelijkheden te vergroten, lijkt het conceptueel aantrekkelijk om alle bestanden met elkaar te koppelen en in één groot datawarehouse onder te brengen. Dit is echter om technische redenen niet altijd uitvoerbaar. Om een datawarehouse op te zetten moet je kunnen koppelen aan unieke sleutels; sleutels die er niet altijd zijn. Als deze sleutels er wel zijn, dan kunnen ze alsnog niet altijd gebruikt worden om privacy redenen of omdat het principe van ‘trias politica’ (leer van de scheiding der machten) gerespecteerd dient te worden.

De Hogeschool Rotterdam doet onderzoek naar het benutten van data afkomstig uit heterogene bronnen die ogenschijnlijk niet te relateren zijn. Dit vergt een complexe manier van dataopslag. In dit projectpaper wordt beschreven hoe de Hogeschool Rotterdam, in samenwerking met de gemeente Rotterdam, de applicatie SunnyApp ontwikkelt als basis voor een prototype dataspace systeem voor de opslag en de ontsluiting van publieke-sector-informatie. Deze applicatie geeft aan waar vrije zonnige plekken op Rotterdamse terrassen zijn en is als pilot case gekozen vanwege de verwerking van diverse soorten data, zoals (gesimuleerde) sensor-, statistische- en dynamische data. Met behulp van een dataspace systeem kunnen verschillende databestanden aan elkaar gerelateerd worden, zodat nieuwe inzichten ontstaan en innovatieve toepassingen kunnen worden ontwikkeld.

Kenmerkend voor een dataspace systeem is dat het relaties tussen databases vastlegt, terwijl een databasesysteem zich meer richt op de relaties binnen een database. In de rest van dit artikel worden het principe en de voordelen van een dataspace systeem beschreven. Vervolgens wordt de SunnyApp toegelicht en sluit het paper af met een overzicht van onderwerpen voor toekomstig onderzoek.

2. Dataspace systemen

Een dataspace systeem bestaat uit drie lagen, (1) een data space, (2) een space manager en (3) een interfacelaag (figuur 1).



Figuur 1. Dataspace systeem in drie lagen.

De eerste laag, de data space, bestaat uit een verzameling van databases. In principe kunnen de databases van verschillende typen zijn, zoals XML-databases, relationele databases, en meer. De data space kan zowel survey- als registratiedata bevatten, ten aanzien van de soort data worden geen restricties opgelegd. Deze eigenschap maakt een data space geschikt voor open data.

In de tweede laag bevindt zich de space manager. De space manager heeft als taak om een informatievraag van een gebruiker op een adequate manier af te handelen. We onderscheiden twee hoofdcomponenten binnen een space manager: een query-scheduler en relatiemanager. De query-scheduler vertaalt een informatievraag in een aantal uitvoerbare query's en bepaalt welke database in de data space een query zal verwerken. Hiertoe maakt de query-scheduler gebruik van de informatie die is opgeslagen in de relatiemanager. In de relatiemanager worden de relaties tussen de verschillende databases vastgelegd. Als twee databases verschillende definities hanteren met betrekking tot een fenomeen, dan wordt deze informatie vastgelegd. Stel dat bijvoorbeeld een bepaald terras door de gemeente wordt geregistreerd onder variabele X en door de KVK onder variabele Y dan worden deze verschillende duidingen van het object in de relatiemanager vastgelegd. Een relatiemanager is erop gericht inconsistenties in aangeboden informatie aan gebruikers te voorkomen en de samenhang tussen data in verschillende bronnen te bewerkstelligen. De resultaten

van de query's worden aangeboden aan de spacemanager, die vervolgens de samenhang tussen de verkregen resultaten aanbrengt.

De derde, en laatste, laag betreft de interface. Hier vindt de presentatie van de informatie op een door de gebruiker gewenste manier plaats. De interfacelaag is bedoeld om de interactie tussen gebruikers en het systeem op een effectieve manier te laten verlopen. Gebruikers moeten op een eenvoudige wijze standaardrapportages kunnen genereren en vragen kunnen stellen aan het systeem. Het systeem moet de antwoorden op beknopte en toegankelijke manier aan een gebruiker tonen. De interfacelaag voor de SunnyApp zal gemaakt worden in een Android-omgeving, om te profiteren van het onafhankelijke karakter van JAVA, waarin de Android-applicatie is geschreven. Het gebruik van een dataspace aanpak biedt een groot aantal voordelen, zoals onafhankelijkheid van lagen waardoor uitbreiding en onderhoud van data eenvoudiger wordt. Als bijvoorbeeld nieuwe relaties tussen twee databases bekend worden, kan deze ingevoegd worden in de relatiemanager zonder dat er iets op de laag van de dataspace hoeft te veranderen. Veranderingen op de interfacelaag hebben in principe geen impact op de onderste lagen. Tenslotte maakt een data space het mogelijk om relaties binnen ketens inzichtelijk te maken.

3. Een toepassing: SunnyApp

Om het dataspace principe uit te werken, wordt de SunnyApp als basis gebruikt. De SunnyApp is een applicatie, een pilot case in het kader van het SIA-RAAK-project van Creating 010 (onderzoeksproject Professionals Supported – Rotterdam Open Data), waarmee geïllustreerd wordt hoe gegevens van verschillende soorten bronnen kunnen worden gecombineerd in een dataspace systeem.

Met de SunnyApp kan een gebruiker binnen een aangegeven straal een terras zoeken waar vrije zitplaatsen in de zon te vinden zijn. De gebruiker kan zelf invoeren binnen welke straal hij een terras zoekt. Deze optie bestaat uit het kiezen van een straal op de plaats van de locatie waar de gebruiker zich op dat moment bevindt of vanuit een andere locatie in de stad waar hij een plek in de zon zoekt. De applicatie geeft vervolgens via de interfacelaag aan waar de grootste kans is op een terras in de zon, binnen de aangegeven straal, en hoeveel zitplaatsen er vrij zijn op dit terras.

Om dit te realiseren wordt gebruikgemaakt van (1) een simulatie van sensordata om de terrasbezetting aan te geven, (2) gegenereerde gegevens over de stand van de zon om, in combinatie met, (3) een hoogtebestand van Rotterdam en (4) KNMI-data over het weer, aan te tonen of een terras in de zon staat of niet. Het hoogtebestand en de KNMI-data zijn beschikbaar als open data. Deze vier bestanden vormen de basis van het systeem, waarbij de spacemanager de relaties tussen de onderliggende bestanden vastlegt en beheert. De interfacelaag, in de vorm van een Android-applicatie, vormt de gebruikersinterface.

3.1. Dynamische data: terras-simulatiemodel en KNMI-gegevens

De simulatie genereert data (in .csv- of tab-delimited-formaat) met de volgende velden: terrasID, n_vrij, timestamp. Om de terrasbezetting zo accuraat mogelijk te simuleren, maakt het simulatiemodel gebruik van een aantal dynamische processen. Als eerste wordt het weer (temperatuur, zon en neerslag) gebruikt. Hierbij kan als alternatief ook gebruik worden gemaakt van historische data, mits er voldoende resolutie in de tijd is. Verder wordt het aanbod van terrasbezoekers per buurt, per tijdstip van de dag, voor elke dag van het jaar aangegeven. Een aanbod dat ook afhankelijk is van het weer. De laatste variabele voor het simulatiemodel is de gecalculerde bezettingsgraad van een willekeurig aantal terrasjes.

Er is een stochastisch simulatiemodel ontwikkeld waarmee een populatie wordt gesimuleerd, bestaande uit individuen in drie verschillende toestanden: de rates zijn afhankelijk van de dag, het tijdstip, het weer, en meer. Ook bestaat er een tabel met gegevens per terras, waarin de exacte terraslocatie, openingstijden en populariteit beschreven staan. De 'huidige' toestand is de meest recente update voor elk van de terrassen (en de toestand op een bepaald moment kan worden samengesteld uit de meest recente updates voor alle terrassen voor een bepaald tijdstip). Daarnaast is ook de gemiddelde verblijftijd op een terras afhankelijk van het weer, de dag van de week, het tijdstip van de dag van belang. Om drukte te simuleren worden individuen gewogen en willekeurig verdeeld over terrassen. De 'ontvankelijkheid' van een terras voor bezoekers is afhankelijk van: de locatie (buurt, veel of weinig winkels), populariteit (willekeurige eigenschappen van een terras) en het aantal stoelen. Door bij de plaatsing van individuen op een terras wel of geen rekening te houden met aantal vrije stoelen en het percentage zon, kan het effect van het gebruiken van de SunnyApp worden geïllustreerd.

De achterliggende gedachte van de simulatie is om het gebruik van dynamische sensordata in een dataspace omgeving te beproeven. Zodoende is het uiteindelijk mogelijk om in plaats van gesimuleerde data, live data te gebruiken. Denk hierbij aan actuele files, of de bezettingsgraad van parkeergarages.

Naast terrasbezetting bevat de dataspace ook de huidige gegevens over het weer. Deze data is beschikbaar als open data en is onafhankelijk van de huidige stand van de zon, bepalend voor of de zon überhaupt schijnt op een terras. Deze data is ook van belang om de terrassimulatie te realiseren.

3.2. Statische data: stand van de zon, hoogtebestand

Om de waarschijnlijkheid te berekenen van een zonnig terras, bevat de data space ook statische gegevens. Deze data betreft de locaties van terrassen, de stand van de zon, en het hoogtebestand. Het hoogtebestand is afkomstig van het gemeentelijk cluster Stadsbeheer in .csv-formaat en bevat de xyz-coördinaten van de stad. Deze gegevens zijn tot 20 centimeter nauwkeurigheid genoteerd. De stand van de zon is met behulp van een formule berekend.

3.3. Relatiemanager

In de relatiemanager wordt de kern van de SunnyApp vastgelegd, de relatie tussen de verschillende bestanden:

- welke coördinaten uit het hoogtebestand worden gebruikt bij een bepaalde stand van de zon;
- welke verwerkingen zijn er mogelijk met de data;
- de relatie tussen de terrassen en het hoogtebestand; de delen van het hoogtebestand die betrekking hebben op een terras.

4. Conclusie en toekomstig onderzoek

De huidige implementatie van de data space betreft slechts vier bestanden. Verder onderzoek en implementatie zouden het mogelijk moeten maken om meerdere opendata-bronnen op te slaan in de data space. Bij vragen waarbij grote hoeveelheden data aangevraagd worden, kan de snelheid van het systeem een probleem zijn. Dit vraagstuk nodigt ook uit tot verder onderzoek.

Wilt u reageren op dit artikel? Neem dan contact op met de auteur:

P.D. Conradie; docent/onderzoeker; Creating 101 / Hogeschool Rotterdam.

p.d.conradie@hr.nl