

Van applicatieland naar in

Bij de Amsterdamse politie wordt onder de naam TopicView een applicatie voor kennisintegratie ontwikkeld. TopicView zorgt dat echt belangrijke kennis binnen de Politie gedeeld kan worden. Daarbij wordt gebruik gemaakt van de semantische modelleringstechniek Topic Maps. Omdat deze techniek om informatie en kennis toegankelijk te maken, in Nederland nog vrij onbekend is, geeft Gabriel Hopmans ook een algemene inleiding op deze techniek. Voorbeelden daarin zijn aan de Nederlandse politiek ontleend.

Gabriel Hopmans

Een aantal jaren geleden is bij de Politie Amsterdam Amstelland het project TopicView gestart, omdat de wens bestond om over meer systemen heen te kunnen zoeken en dan niet applicatiespecifiek, maar gebaseerd op onderwerp. De laatste tijd werd al in het nieuws gemeld dat kennis tussen politieregio's zeer lastig te delen is. Er zijn veel applicaties en in feite geldt: kennis is macht. De wens leeft om voor bijna ieder informatie-element aan te geven wie het wel mag zien en delen en wie niet. Informatie rond allerlei onderwerpen is gefragmenteerd in veel verschillende systemen terug te vinden: personen zijn vaak lid van meerdere groepen, spelen verschillende rollen en kunnen soms bij gebeurtenissen betrokken zijn die achteraf toch meer informatiewaarde kunnen

verschaffen. Er bestaan heel veel losse tekstbeschrijvingen die waardevol kunnen zijn, maar niet zijn 'veredeld'. In een politierapportage is bijvoorbeeld gemeld dat een persoon relaties heeft met leden van een bepaalde doelgroep, maar dat staat niet zo geassocieerd in het registratiesysteem. Bovendien wordt veel informatie op verschillende plaatsen apart beheerd. Analisten moeten beter en sneller kunnen zoeken om bijvoorbeeld managementrapportages te maken: een half uur werken aan een zaak die versnipperd staat in meerdere systemen, is inefficiënt als dat werk vooral kopieer- en plakwerk is. Daarnaast bestaat bovendien de wens meer met kennis te kunnen doen, bijvoorbeeld beweringen of vermoedens kwijt te kunnen in een systeem en deze in een

Topic Maps uitgelegd

Topic Maps is een internationale standaard waarmee kennis en informatie wordt vastgelegd aan de hand van concepten en relaties. De standaardisatie

van Topic Maps wordt verzorgd door een werkgroep³ van ISO (International Organization for Standardization). Oorspronkelijk is de standaard begin jaren negentig opgesteld⁴ om een masterindex voor documentatie te kunnen maken, waarin men verschillende indexen met elkaar moest mengen. Er was toen al grote behoefte om meerdere bronnen te ontsluiten en een overzicht op het geheel te creëren. In 2001 kwam er dan ook een webstandaard met de naam XML Topic Maps. De groep van XML-specialisten is later gaan werken aan diverse andere uitwisselingsformaten op basis van hetzelfde datamodel. Dit datamodel blijft het hart van de standaard, het is de 'map' – als metafoor van landkaart – waarin de aspecten van een kennisorganisatie in computer-

Wie is...



Gabriel Hopmans is afgestudeerd bij de Universiteit Maastricht als kennistechnoloog. In 1999 startte hij samen met Peter-Paul Kruijssen het kennistechnologiebedrijf Morpheus. Hier worden onder andere voor de politie kennisoplossingen met Topic Maps ontwikkeld.

schap formatielandschap

.....

bepaalde groep collega's te kunnen delen, om ze op te volgen of vermoedens te versterken. Beweringen moeten makkelijk bevestigd of verworpen kunnen worden (gebaseerd op bronnen en in relatie met de wet).

Oplossing met Topic Maps

Deze behoeftes zouden met Topic Maps deels of helemaal vervuld kunnen worden. De Semantische Standaard Topic Maps wordt onderaan dit artikel uitgelegd. De benaming TopicView voor het nieuwe systeem is hier uiteraard op geïnspireerd.¹ De toegevoegde waarde die TopicView bij de Politie bovenop bestaande systemen moet bieden is:

> dwarsverbanden maken;

- > semantische bevestigingen mogelijk maken;
- > integreren met externe systemen en bronnen;
- > verrijkingen aanbrenge;
- > zorgen dat alleen de juiste mensen bepaalde informatie zien.

Met semantische bevestigingen wordt bedoeld dat ook de betekenis van wat je zoekt, wordt meegenomen in de bevestiging. Bij de bevestiging 'Hopmans in Amsterdam tijdens periode t1-t2' wordt dan meegenomen dat Hopmans mogelijk een persoon is, Amsterdam een plaats, die ook relaties heeft met hoofdstad, Noord-Holland, et cetera.

De situatie bij de politie was eerst als volgt: als een analist informatie ging opzoeken over alle gebeurtenissen rond een

'De wens leeft om voor bijna elk informatie-element aan te geven wie het wel mag zien en delen en wie niet'

bepaald persoon, welke personen bij dezelfde gebeurtenissen betrokken waren geweest, welke voorwerpen of voertuigen een relatie met deze personen hadden enzovoort, dan was de analist genooddacht drie verschillende systemen te raadplegen en de resultaten handmatig samen te voegen. Dit komt uiteraard voort uit het feit

.....

interpreteerbare vorm kunnen worden beschreven.

De basis

De representatie van de kennis en informatie van een organisatie krijgt vorm door de specificaties van de Topic Maps-standaard te volgen: het wordt een map van concepten en relaties en wordt aangeduid als een topic map.⁵ De basis (the TAO)⁶ van Topic Maps bestaat uit:

- > topics, de belangrijkste onderwerpen;
- > relaties tussen topics: Associations;
- > occurrences: 'voorkomens' in de data, dit kunnen feiten zijn, en links naar webpagina's, documenten, velden in een database enzovoort.

Als we nu een topic map toepassen op de Nederlandse politiek, dan kunnen bijvoorbeeld 'Pechtold' en 'D66' topics zijn, 'Lid van Kamerfractie' is een mogelijke associatie hiertussen en de webpagina van D66 is een occurrence van het topic 'D66'.

Vanuit deze basis wordt gewerkt, maar daarnaast zijn er nog diverse andere handvaten om de kennis flexibel in de map te kunnen representeren. Topics zijn surrogaten, of letterlijk eigenlijk 'proxies' voor onderwerpen waarover gesproken kan worden. Deze onderwerpen kunnen willekeurige onderwerpen zijn in de breedste zin van het woord: bijvoorbeeld het dossier 'AOW' is een topic, zoals gezegd is 'D66' een topic, maar ook minder grijpbare topics zoals het onderwerp 'Financiële crisis' of de toezegging door een bewindspersoon dat 'Brief aan de Kamer of de huidige ICT-infrastructuur van de overheid klaar is voor web 2.0' voor een bepaalde datum zal worden voldaan.

Eigenschappen van topics

Naast de basis (Topics, Associations, Occurrences) is er binnen de standaard het concept van Types; types leveren zogenaamde klasse-instantiesemantiek, het geeft aan tot welke klasse of klassen een topic be-

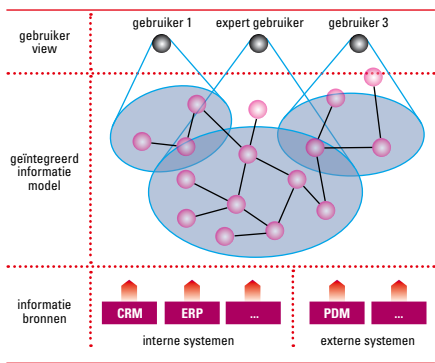
hoort (waar het een 'instantie' van is). In ons voorbeeld geven we onder andere aan: Balkenende is een persoon, CDA is een partijfractie.

Verder duiden we rond topics nog de volgende eigenschappen aan:

- > topics hebben identiteit. Balkenende kunnen we bijvoorbeeld identificeren met zijn geboortedatum, zijn Burger Service Nummer of zijn e-mailadres. (Op identiteit komen we nog terug.)
- > topics hebben karakteristieken:
 - nul of meer namen: naam balkenende maar ook JP is een bijnaam;
 - nul of meer occurrences;
 - nul of meer associations.

Associatief modelleren

Als illustratie hebben we zojuist de Topic Types 'Persoon' en 'Fractie' genoemd en hebben we het gehad over de Associatie 'lid-van-kamerfractie'. Deze associatie heeft echter ook een Type, net zoals bij het Topic Type.



Figuur 1. Topic Maps als informatiemodel op verschillende bronnen en verschillende views op dit geïntegreerde model

dat de systemen met een bepaald doel zijn opgezet, een bepaalde reikwijdte hebben en vaak ook verschillende eigenaren.

Navigeren over verschillende systemen

Binnen TopicView worden informatie-elementen zoals personen, gebeurtenissen en auto's met Topic Maps uniek geïdentificeerd, in een ontologie gemodelleerd en zodanig getypeerd, dat hierna de samenhang in het geheel snel naar voren komt. In systeem A bij de Politie staan veel soorten documenten en ook tekstuele meldingen, waarin betrokken personen genoemd

worden. Maar er is geen directe link naar bijvoorbeeld een registratie van een persoon, die op een bepaalde dag behandeld is. Die informatie staat in systeem B en daar ontbreekt weer de relatie met de termen die bij deze case van toepassing zijn. Binnen de topic map zijn al deze relaties echter onderling al aanwezig.

Nadat in topic maps een ontologie is opgesteld en bronnen zijn ontsloten, is het eerste resultaat dat gebruikers al heel veel informatie uit verschillende systemen in samenhang zien. In figuur 1 is dit het geïntegreerde informatie model. De rode bolletjes zijn topics met daartussen de associaties. Iedere topic zelf heeft hiernaast een eenduidige verwijzing naar de voorkomens in de onderliggende systemen. Hoe deze relaties semi-automatisch tot stand kunnen komen, wordt verderop in dit artikel geïllustreerd. Het geïntegreerde informatiemodel biedt zo een soort super-index waarbij over alle topics, relaties en voorkomens valt te navigeren. Hierna worden voor de gebruiker verschillende views op het model ontwikkeld. Door middel van samengestelde queries (met de Topic Maps Query Taal: TMQL) is op te vragen welke informatie-elementen rond een bepaalde zaak of een persoon te vinden zijn, welke directe relaties er zijn

'TopicView geeft ons iedere dag alle informatie uit alle politiebesteden'

en in welke context we dit kunnen zien. Het grote voordeel hierbij is dat één query kan worden losgelaten op één datamodel, terwijl normaal gesproken meerdere queries op verschillende systemen ontwikkeld hadden moeten worden, waarna met reguliere databasetechnieken een aantal moeilijke zogenaamde Joins- en Groupby-constructies hadden moeten worden toegepast.

In het geval van openstaande meldingen op de top 10 meest gezochte personen, kan bijvoorbeeld worden getoond hoe eerder door een analist beweringen zijn toegevoegd en hoe hierop is gereageerd, of hoe later bepaalde uitingen gerelateerd

Om in een topic map de associatie: lid-van-kamerfractie(pechtold : persoon, d66 : fractie) te kunnen toevoegen, zullen we vastleggen dat deze associatie mogelijk is tussen de Topic Types 'Persoon' en 'Fractie'. Op 'hoger kennisniveau' wordt in de representatie gedefinieerd dat lid-van-kamerfractie een association-type is tussen persoon en fractie.

lid-van-kamerfractie(\$P : persoon, \$F : fractie)

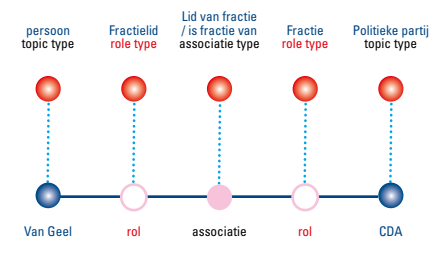
De variabelen \$P en \$F mogen later worden ingevuld door topics van het type persoon en fractie, zoals 'Pechtold' en 'D66'. Het is vooral op dit onderdeel dat veel semi-automatisch kan worden ontsloten uit bestaande bronnen. Het idee dat alles in topic maps handmatig moet worden gemodelleerd, is op dit instantieniveau dus niet juist.

Rollen

In topic maps bezitten associaties van

nature geen richting: in associations beweren we iets over relaties tussen onderwerpen en uit de bewering Van Geel is lid van Fractie CDA volgt automatisch de bewering CDA is de fractie van Van Geel. Het idee van richting wordt veroorzaakt door het gebruik van natuurlijke taal. In topic maps wordt het idee van richting vervangen door associatie rollen. Bijvoorbeeld:

De topic Van Geel speelt de rol van ka-



Figuur 3. De Associatie Van Geel is lid van de fractie CDA

merlid in de relatie Lid van Fractie waarbij CDA de rol van fractie speelt.

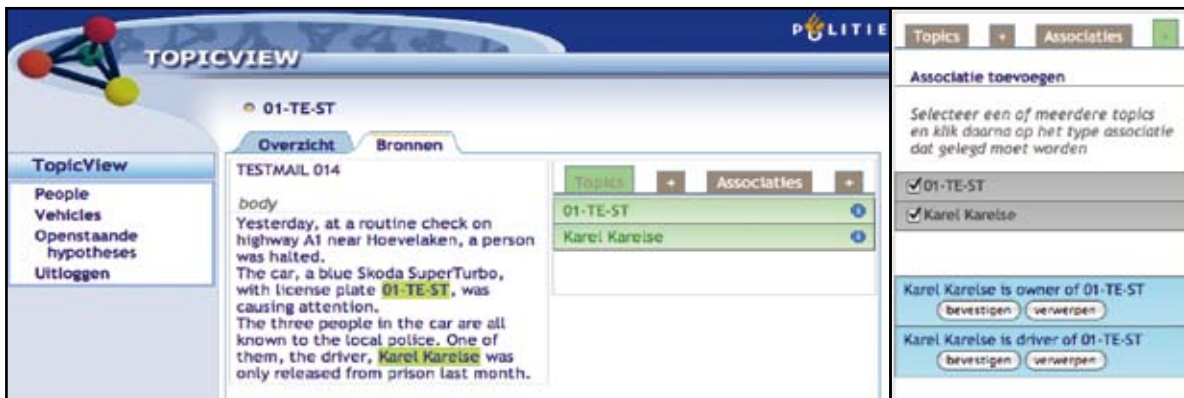
Ten slotte zijn er ook nog Occurrence-types:

Webpagina, e-mailadres, Leeftijd, Beschrijving, Kamerstuknummer zijn mogelijke occurrence-types. Ze geven aan wat de aard van de occurrence is. De waardes hiervan: <http://www.cda.nl>; 45, enzovoort zijn de occurrences zelf.

Twee andere belangrijke elementen bij Topic Maps zijn Merging en Scoping. Deze kunnen ons bijvoorbeeld helpen om kamerleden per kabinet en per periode op te vragen.

Merging

In Topic Maps representeert iedere topic een onderwerp en wordt uitgegaan van precies één topic per onderwerp. Meerdere mensen kunnen ieder tegelijk een topic map uitwerken en later bij elkaar voegen. Als twee topic maps gemerged worden, moeten topics die



Figuur 2. Voorbeeld uit TopicView, waarbij uit een mailbericht mogelijke relaties tussen een kenteken en een persoon worden afgeleid

zijn met de uiteindelijke actie die hierop is genomen. Naast Topic Maps zijn ook andere semantische technologieën bekeken en gebruikt, die op ontologieën zijn gebaseerd, zoals de combinatie RDF-OWL (twee standaarden van het WWW Consortium – W3C). Uiteindelijk is echter gekozen voor Topic Maps, omdat daarmee zo eenvoudig beweringen aan relaties kunnen worden toegevoegd.

Kennisverrijkingen in TopicView

Het ontwikkelen van één overzicht op meerdere bronnen is, na een goede onto-

logiemodellering, nog een relatief gemakkelijke klus. Een veel lastiger onderdeel is om kennisverrijkingen te kunnen creëren en semi-automatisch de ongestructureerde bronnen te benutten.

In figuur 2 vinden we een simpel voorbeeld waarin een mail wordt geanalyseerd en vervolgens wordt afgeleid dat hier waarschijnlijk een relatie te vinden is tussen een nummerplaat van een auto en een persoon.

De gebruiker hoeft alleen nog te bevestigen of te verwerpen dat deze relatie klopt. Tevens wordt standaard meegenomen op basis van welke bron hij dit wil bevestigen. Dit semi-automatisch vanuit het systeem aanbieden van een classificatie waar-

uit de gebruiker kan kiezen, illustreert de toepassing van de in de uitleg van Topic Maps beschreven scoping en TMQL. Merk op dat de Politie benadrukt dat alle hypothesen moeten worden gecontroleerd en niets automatisch gebeurt.

In dit vereenvoudigde voorbeeld gaat het om de topics ‘Persoon’ en ‘Auto’ maar er zijn ook zogenaamde ‘Early warning’-indicatoren. Zulke indicatoren moeten vanuit een kennisgebied worden opgesteld en kunnen op de ongestructureerde bronnen worden toegepast. Denk bijvoorbeeld aan indicatoren rond het radicaliseringproces met indelingen (en voorkomens) die te maken hebben met:

hetzelfde onderwerp representeren gemerged worden tot één enkele topic. Als twee topics gemerged worden, krijgt het resulterende topic de combinatie van karakteristieken van de twee oorspronkelijke topics. Kortom, als u meerdere bestanden samenvoegt waarin zaken dubbel staan, dan worden deze samengevoegd.

Stel dat we alle kamerleden in een topic map hebben vastgelegd, dan moeten we uiteraard 150 personen zien. Als we een topic map hebben van de kamerleden onder het kabinet-Balkenende IV en ook een oudere topic map van bijvoorbeeld twee jaar eerder, dan is daar uiteraard een verschil tussen. Zo vinden we in de oudere topic map nog Arend Jan Boekesteijn terug. Stel dat dit het enige verschil zou zijn en je zou de twee lijsten met kamerleden bij elkaar voegen, dan zouden we zonder filtering een lijst met 300 kamerleden krijgen waarbij 298 namen er dubbel in staan en twee namen niet: namelijk die

van Boekesteijn en die van zijn vervanger (Mark Harbers).

Met het merging-karakter van topic maps en sterke identificatie zouden we 151 kamerleden terugkrijgen, omdat dubbelingen eruit worden gefilterd. De kracht van merging wordt pas echt duidelijk door het zelf toe te passen. Met de opensourcesoftware van Ontopia (www.ontopia.net) kunt u relatief eenvoudig zelf snel aan de slag om bijvoorbeeld te ervaren hoe u uw eigen adresadministratie kunt ontdebellen.

Ontdebellen is uiteraard nog niet voldoende. Daarom willen we ook zorgen dat per kabinet en/of per tijdsperiode wordt bijgehouden wanneer een persoon kamerlid is. Dit kan gedaan worden met Scope.

Scoping

Scope maakt het mogelijk om:

- > de subjectiviteit van kennis uit te drukken;

- > vanuit verschillende gezichtspunten naar een enkele kennisbron te kijken;
- > gepersonaliseerde gezichtspunten te bieden voor verschillende gebruikersgroepen.

Met behulp van scope kunnen we zorgen dat bijvoorbeeld de relatie lid-van alleen geldig is in een bepaalde periode of in een bepaald kabinet. In figuur 4 willen we aangeven dat A.J. Boekesteijn kamerlid was onder het kabinet-Balkenende IV tot 24 november 2009.

Met scope is het ook mogelijk om op bestaande relaties gemakkelijk beweringen toe te voegen. In de kennislaag is het dus mogelijk om een persoonlijke blik op kennis toe te voegen, nuttig in het geval:

- > verschillende gebruikers verschillende eisen over kennis hebben, en
- > een persoonlijke tint, gebaseerd op persoonlijke voorkeur, vaardigheden, beveiligingsniveau, toegevoegd dient te worden.

Een bepaald wetsvoorstel kan bijvoor-

- > identiteit en uiterlijk (traditionele kledij);
 - > ideologie (bekeringsactiviteiten, ronselen);
 - > afnemende tolerantie (jegens vrouwen).
- Voor deze indicatoren wordt tevens scope gebruikt als mechanisme dat niet zomaar iedereen de informatie rond een bepaald subject mag zien.

TopicView in gebruik

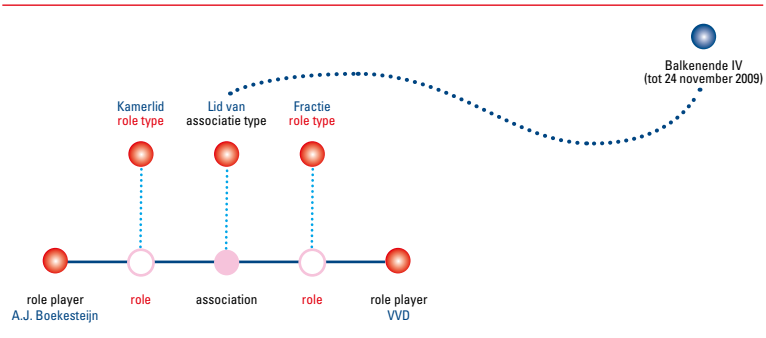
TopicView wordt nu nog gebruikt door verschillende kleinere gebruikersgroepen en gaat binnenkort bij grotere doelgroepen in productie. De gebruikersgroep die zich bezighoudt met Risico Jeugdgroepen in probleemwijken is erg enthousiast. Zo geeft Remko de Boer districtscoördinator veelplegers aan: ‘Het geeft ons iedere dag alle informatie uit alle politiebestanden. Niet alleen uit officiële stukken, maar ook uit Word-bestanden die de politie even snel heeft ingevoerd.’ TopicView bevat ook ‘zachte informatie’. ‘Iemand rijdt in een dure auto, maar het gezin heeft schulden. Vroeger zouden we iets aan de schulden doen. Nu gaan we op zoek naar de illegale financieringsbron. Een informatiespecialist verwerkt de informatie en op basis daarvan zetten

we een informatieopdracht uit.’² In kennisintensieve organisaties is de behoefte om belangrijke informatie uit verschillende systemen te halen en in samenhang te tonen, erg groot geworden. Er is geen tijd meer om afzonderlijk in verschillende systemen te gaan zoeken en resultaten met kennis te gaan samenvoegen: semantiek is nodig. Uitdaging is om ook relaties met niet-eenduidige gebeurtenissen in de wereld te kunnen toevoegen of semi-automatisch te kunnen aanbrengen. Life events zoals ‘Explosie Stockholm’, ‘Politie onder minister Opstelten’ of ‘bewering dat persoon A waarschijnlijk bij groepering X hoort op basis van bron Z’ zijn namelijk niet digitaal en gestructureerd vastgelegd. Met Topic Maps is enerzijds vrij snel en makkelijk, door middel van een goede ontologie, een overzichtelijke kenniskaart van de gestructureerde bronsystemen te maken. Anderzijds is het de gestructureerde kapstok om ook niet-gestructureerde informatie-elementen en life events te associëren met de gestructureerde elementen in de topic map. Aan een relatie tussen persoon A en een gebeurtenis kan eenvoudig nog een relatie worden toegevoegd waar tegelijk een bewering en een bron bij kunnen horen. Die bewering

kan later worden versterkt, afgezwakt of verwijderd met ook weer relatie naar een bron of een link naar een overtreden wetsartikel. Als na analyse van collega-analisten blijkt dat persoon A inderdaad geassocieerd mag worden met die gebeurtenis, dan kan direct zichtbaar worden dat de vermoedelijke contacten van persoon A inderdaad met hem in relatie staan. Dit alleen door het mechanisme van Scope en de versterking van de vermoedelijk relatie: kortom echt praktisch kennis delen! Met Topic Maps hoeft kennismanagement niet gecentraliseerd te worden om kennis deelbaar te maken in een organisatie. In het geval bij de Politie (en bij veel grote organisaties) is dit zeer wenselijk omdat veel huidige systemen de komende jaren in gebruik zullen blijven. Een kennislaag die hier ‘als een deken overheen ligt’, zorgt voor informatieverrijking van deze bronsystemen. Men kan immers pas goed zoeken en vinden nadat meer betekenis is toegekend. Bovendien willen de meeste mensen niet alleen iets vinden maar er daarna ook iets mee doen. Topic Maps brengt informatie dus in samenhang. Het is een open standaard waarmee op innovatieve manier valt te integreren met andere systemen en kennis kan worden uitgewisseld. <

beeld als erg positief worden bestempeld voor een bepaalde partij of standpunt, maar misschien is er ook een aantal aspecten in het wetsvoorstel dat dubieus of juist negatief voor dezelfde partij is. In een aantal projecten waarin Topic Maps is toegepast, is bovenstaand voorbeeld van scope soortgelijk toegepast door na de ontsluiting van meerdere systemen een functionaliteit aan te bieden

waarbij semi-automatisch een classificatie vanuit het systeem wordt aangeboden (zie voorbeelden hiervan bij de Politie in de beschrijving van TopicView). Naast het gebruik van de bovenstaande Topic Maps-concepten wordt daarbij ook gebruik gemaakt van de Topic Maps Query Taal, TMQL.⁷ Hiermee kunnen complexe bevragingen geconstrueerd worden, enigszins vergelijkbaar met wat in gestructureerde databases met SQL mogelijk is. <



Figuur 4. Gebruik van Scope als extra filtermechanisme

Noten

- 1] Naar een idee van Paul Elzinga, Onderzoeker, projectleider bij Politie Amsterdam Amstelland.
- 2] Zie het rapport *Samen effectief, de praktijk als inspiratie*. Uitgave ter gelegenheid van de realisatie van het landelijk dekkend netwerk Veiligheidshuizen, december 2009, Ministerie van Justitie www.veiligheidshuizen.nl/doc/publicaties/Samen_effectief.pdf.
- 3] www.isotopicmaps.org.
- 4] Bij SOFABED = Standard Open Formal Architecture for Browsible Electronic Documents.
- 5] Met ‘Topic Maps’ (meervoud, hoofdletters) wordt de methode aangegeven; een ‘topic map’ (enkelvoud, kleine letters) is een op basis van die methode gebouwd informatiesysteem.
- 6] Paper van Steve Pepper: The Tao of Topic Maps, www.ontopia.net/topicmaps/materials/tao.html.
- 7] Voor een introductie zie bijvoorbeeld www.xml.com/pub/a/2005/06/01/tmq1.html.