

Op zoek naar de bekende dader in het HKS

Een onderzoek naar de mogelijkheden voor
een nieuw zoekinstrument voor het HKS

Eindrapport

Ben Rovers
BTVO- Bureau voor Toegepast Veiligheidsonderzoek
's-Hertogenbosch, december 2003

B T V O

Inhoudsopgave

Vooraf/dankwoord	5
1. Op zoek naar de bekende dader in het HKS	7
1. Inleiding	7
1.1 Uitgangspunt van het onderzoek	7
1.2 Doel van het onderzoek	7
1.3 Onderzoeksvragen	7
1.4 Opzet van het rapport	8
2. Beschrijving van de huidige zoekroutines in het HKS en hun beperkingen	8
2.1 Formuleren van de zoekvraag	9
2.2 Uitvoeren van een zoekopdracht	10
2.3 Presenteren van het zoekresultaat	14
3 Samenvatting: beperkingen van de huidige praktijk	15
2. Uitgangspunten voor een nieuw zoekinstrument	19
1. Uitgangspunten voor het zoekproces	19
2. Overzicht van zoekdoelen	23
3. Welke gegevens zijn beschikbaar en wat kunnen we ermee?	24
4. Overzicht van zoekvragen	28
5. Overzicht van zoekmethoden	33
5.1 Zoekmethode 1: Matchen op geselecteerde kenmerken (huidige praktijk)	33
5.2 Matchen met toegevoegde gegevens	35
5.3 Zoekmethode 2: Matchen op een combinatie van kenmerken	35
5.4 Zoekmethode 3: Gecombineerd matchen	37
5.5 Zoekmethoden 4a-4c: Matchen van geclusterde gegevens	37
5.6 Samenstelling van delictprofielen bij het matchen van geclusterde gegevens	38
6. Samenvatting	40
3. Opzet van de toetsing	41
1. Inleiding	41
2. Selectie van gegevens	41
3. Veredelen van informatie uit geselecteerde gegevensvelden	44
4. Uitgangspunten en opzet van de toetsing (algemeen)	45
5. Simulatie van zoekmethode 1	47
6. Uitwerking van zoekmethode 3	53
7. De rangordepositie van de werkelijke dader bepalen	54
8. Opzet van de toetsing per groep van zoekvragen	55
8.1 Groep 1: matchen van individuele delicten	55
8.2 Groep 2: matchen van individuele delicten met delictprofielen	56
8.3 Groep 3: matchen van signalen	57
8.4 Groep 4: matchen van delict/signalelementcombinaties	58

4. Onderzoeksresultaten	61
1. Groep 1: matchen van individuele delicten.....	61
2. Groep 2: matchen van individuele delicten met delictprofielen	68
3. Groep 3: matchen van individuele signalementen	73
4. Groep 4: matchen van delict/signalementcombinaties.....	78
5. Nadere beschouwing over zoekvragen voor tactisch gebruik.....	81
6. Bruikbaarheidsindicatoren	83
7. Samenvatting: de belangrijkste onderzoeksresultaten op een rij	84
5. Samenvatting en conclusie.....	91
Bijlagen.....	97
Bijlage 1 Overzicht van geselecteerde gegevensvelden aan de aangiftenkant.....	99
Bijlage 2 Overzicht van geselecteerde gegevensvelden aan de personenkant.....	101
Bijlage 3 Overzicht van delictkenmerken die gebruikt zijn bij zoekmethode 1	103

Vooraf/dankwoord

Het onderzoeksproject waarvan in dit rapport verslag wordt gedaan is financieel mogelijk gemaakt door het programmabureau 'Politie en Wetenschap' in Apeldoorn in het kader van de 'Call 2001' (onderzoeksovereenkomst pw/oc/2002/3). De heer drs Cees Loef trad op als projectbegeleider namens 'Politie en Wetenschap'.

De Politie Haaglanden, afdeling Analyse & Research, heeft aan de onderzoeker gastvrijheid verleend om het project aldaar uit te voeren.

De begeleidingscommissie die het onderzoek inhoudelijk heeft begeleid was als volgt samengesteld:

Voorzitter:

- Dhr. drs Peter Versteegh (Politieregio Haaglanden, afdeling Analyse & Research, Den Haag)

Overige leden:

- Dhr. drs Cees Loef (projectbegeleider namens Politie en Wetenschap, Apeldoorn)
- Dhr. drs Huib van Nierop (Politieregio Haaglanden, Bureau Recherche Informatie, Den Haag)
- Dhr. Jan van Nieuwamerongen (Politieregio Haaglanden, Bureau Recherche Informatie, Den Haag)
- Mw. mr. Pauline van Panhuis (Politieregio Haaglanden, afdeling Analyse & Research, Den Haag)
- Mw. Astrid Patty-Hüttman (Politieregio Haaglanden, afdeling Analyse & Research, Den Haag)
- Dhr. Leen Prins, (Korps Landelijke Politiediensten, Dienst Nationale Rechercheinformatie, Afdeling Onderzoek & Analyse, Zoetermeer)

Ik wil alle bovengenoemde personen en instellingen bedanken voor hun bijdragen aan het mogelijk maken van het onderzoek en het realiseren van dit rapport.

Daarnaast wil ik ook de respondenten bedanken die zich in de beginfase van dit project hebben laten interviewen: Rob de Jongh, Joke van Kan, Simon Rotteveel & Art Wijnants, allen werkzaam bij de politie Haaglanden. Met name Rob de Jongh ben ik dank verschuldigd voor zijn herhaalde hulp en aanwijzingen in de eerste fase van het onderzoek.

Ten slotte wil ik alle medewerkers van de afdeling Analyse & Research bedanken voor hun aangename gezelschap en hun betrokkenheid bij het project en 'aanverwante zaken'. Met name Janine Jansen, Astrid Patty en Mark van de Vin ben ik dankbaar voor hun herhaalde hulp bij allerlei praktische zaken/problemen die tijdens de looptijd van dit project aan de orde waren. In één woord: super!

Ben Rovers

1. Op zoek naar de bekende dader in het HKS

1. Inleiding

1.1 Uitgangspunt van het onderzoek

Het uitgangspunt van dit onderzoek is gelegen in de ervaring (van zowel opsporingsonderzoekers als wetenschappers) dat een relatief kleine groep daders verantwoordelijk is voor een relatief groot deel van de misdrijven. Dit gegeven is van evident belang voor opsporingsonderzoek. Op grond hiervan kunnen we immers stellen dat bij een willekeurig misdrijf met onbekende dader (OD), de kans groot is dat de politie al eerder werd geconfronteerd met delicten van deze dader (lees hier ook: *dadergroep*). Ook is de kans aanwezig dat de dader zelf al eerder met de politie in aanraking kwam. Met andere woorden, bij (de voorbereiding van) een willekeurig opsporingsonderzoek is de kans groot dat de politie al beschikt over gegevens omtrent delicten en/of daders die in verband kunnen worden gebracht met het delict dat men wil ophelderen. Het (snel) boven water krijgen van deze gegevens kan een belangrijke bijdrage leveren aan het opsporingsonderzoek. Hiervoor is een efficiënt en effectief zoekinstrument onontbeerlijk.

Er zijn doorgaans verschillende gegevensbronnen beschikbaar voor het verrichten van zoekactiviteiten zoals hier bedoeld. In het kader van dit onderzoek beperken we ons tot het opsporingsregister van de herkenningdienst, het zogenaamde HKS. Hierin bevinden zich gegevens omtrent misdrijven die ter kennis zijn gekomen van de politie en gegevens omtrent verdachten en hun *criminal record*. Het HKS wordt vaak, maar niet structureel, gebruikt voor het verrichten van zoekactiviteiten in het kader van opsporingsonderzoek. De wijze waarop dit doorgaans gebeurt, biedt wellicht mogelijkheden voor verbetering in termen van efficiency en effectiviteit.

1.2 Doel van het onderzoek

Het doel van dit onderzoek is de haalbaarheid te onderzoeken van een nieuw zoekinstrument dat, gebruikmakend van gegevens uit het HKS, in staat is op een efficiënte en effectieve wijze gegevens (omtrent verdachten, antecedenten en aangiften) te koppelen aan gegevens die beschikbaar zijn in een (te starten) opsporingsonderzoek, opdat aanknopingspunten worden verkregen voor (verder) opsporingsonderzoek. Concreet wordt hiermee beoogd de beschikbare gegevens in het HKS optimaler te benutten voor tactische en operationele onderzoeksdoelen dan nu -in de huidige praktijk- het geval is.

De concrete realisatie van het zoekinstrument (in termen van het programmeren van een gebruiksklaar softwareproduct) valt buiten het bestek van dit onderzoek.

1.3 Onderzoeksvragen

Het optimaler benutten van gegevens uit het HKS voor het beantwoorden van zoekvragen, kan op verschillende manieren plaatsvinden, niet alleen door het ontwikkelen van nieuwe zoekmethoden. We stellen onszelf in dit verband de volgende vragen:

1. De huidige zoekmodaliteit in het HKS wordt alleen gebruikt voor operationele doeleinden. Is het mogelijk en zinvol om een zoekinstrument te ontwikkelen dat ook voor tactische doeleinden gebruikt kan worden, bijvoorbeeld om vast te stellen wat de opsporingsmogelijkheden zijn van een specifieke aangifte of groep van aangiften of van een (groep van) verdachte(n)? Kortom, is het mogelijk de zoekdoelen uit te breiden?
2. Is het mogelijk en zinvol om nieuwe zoekvragen te creëren, bijvoorbeeld door gebruik te maken van het feit dat bepaalde gegevens in het HKS geclusterd zijn binnen eenheden (zoals delicten binnen verdachten)? Kortom, is het mogelijk de zoekvragen uit te breiden?
3. Heeft het veredelen van de beschikbare gegevens in het HKS (door het aggregeren en combineren van bestaande gegevens tot nieuwe gegevens) meerwaarde voor het zoekproces?
4. Geven alternatieve zoekmethoden een beter zoekresultaat dan de huidige zoekmodaliteit in het HKS?
5. Op dit moment bestaan er geen criteria waarmee de bruikbaarheid van een verkregen zoekresultaat kan worden beoordeeld. Is het mogelijk dergelijke bruikbaarheidscriteria te ontwikkelen?

1.4 Opzet van het rapport

De opzet van dit rapport is verder als volgt: hierna, in paragraaf 2, wordt de huidige zoekmodaliteit in het HKS besproken, het zogenaamde ‘bevragen van HKS’. Hierbij worden mogelijke en gesignaleerde tekortkomingen en beperkingen op een rij gezet. Deze kennis vormt de basis voor de ontwikkeling van het nieuwe zoekinstrument, waarvan we de uitgangspunten formuleren in hoofdstuk 2. In dat hoofdstuk zetten we nieuwe zoekdoelen, nieuwe zoekvragen en nieuwe zoekmethoden op een rij. Om deze op hun bruikbaarheid en toegevoegde waarde te toetsen moeten verschillende simulaties en analyses ontwikkeld worden. In hoofdstuk 3 wordt beschreven welke gegevens voor deze analyses zijn gebruikt en op welke wijze de toetsing plaatsvindt. In hoofdstuk 4 worden de onderzoeksresultaten besproken. In het vijfde en laatste hoofdstuk wordt een en ander samengevat en worden conclusies getrokken.

2. Beschrijving van de huidige zoekroutines in het HKS en hun beperkingen

Het zoekproces kan uiteengelegd worden in 3 fasen: 1) formulering van de zoekvraag, 2) uitvoering van de zoekopdracht in het HKS en 3) presentatie van het resultaat. We zullen hierna per fase kort de huidige praktijk schetsen en daarbij ingaan op (mogelijke) beperkingen.¹

Het is van belang hierbij het volgende in het oog te houden. Een gedetailleerde opsomming van ‘mogelijke’ beperkingen in de huidige zoekroutine van het HKS wekt wellicht de indruk dat er weinig deugt van de huidige praktijk. Dit is echter geenszins het geval. De huidige

¹ We spreken hier en daar over ‘mogelijke’ beperkingen, om aan te geven dat het om theoretische uitspraken gaat. In de praktijk kan blijken dat de gesignaleerde beperkingen niet of nauwelijks van invloed zijn op het zoekresultaat. Dit is op voorhand niet duidelijk, vandaar de aanduiding: ‘mogelijke’ beperkingen.

zoekroutine werkt, voorzover wij hebben kunnen beoordelen, volledig naar wens als het gaat om zoekvragen waarop het antwoord vooraf bekend is. Dat wil zeggen, de vrager weet van tevoren wat hij wil weten. Hij is bijvoorbeeld op zoek naar een bepaald type delict of naar een bepaald soort buit, een bepaald type auto, etc. Dergelijke zoekvragen kunnen met de huidige bevragsingsmodule in het HKS uitstekend worden beantwoord. In dit onderzoek zijn we echter geïnteresseerd in andersoortige vragen, zoekvragen waarop het antwoord vooraf niet bekend is. Als er een delict is waarvan de dader onbekend is en we willen weten of deze dader mogelijk in het HKS voorkomt, dan hebben we vooraf geen toetsingskader waaraan het zoekresultaat moet voldoen. Het is vanuit deze optiek dat we de huidige zoekroutines in het HKS bespreken. Bovendien, een kritische beschouwing van de bestaande situatie is vereist om de thans onbenutte mogelijkheden van het systeem zo goed mogelijk boven water te krijgen alsmede de beperkingen die verbonden zijn aan het ontwikkelen van nieuwe zoekroutines.

2.1 Formuleren van de zoekvraag

Zoekvragen kunnen voorafgaand aan of tijdens een concreet opsporingsonderzoek aan de orde zijn. Indien ze voorafgaand aan een opsporingsonderzoek worden gesteld, dienen ze feitelijk een tactisch doel: om te onderzoeken in hoeverre één of meer aangiften ‘mogelijkheden’ bieden voor opsporingsonderzoek, bijvoorbeeld doordat een aangifte ‘past’ bij het delictgedrag van een bekende verdachte of ‘past’ bij andere aangiften die recentelijk zijn gedaan. In de huidige praktijk bestaat geen systematische aanpak om aangiften op deze wijze te controleren op ‘opsporingswaardigheid’.

Incidenteel kan het voorkomen dat een medewerker van de herkenningdienst (HKD) bij de invoer van recente aangiften stuit op misdrijven die een verband met elkaar lijken te hebben, bijvoorbeeld omdat ze een bijzondere modus operandi (MO) delen. In zo’n geval zal deze medewerker de verantwoordelijke leidinggevende binnen de recherche op de hoogte stellen. Dit proces is echter afhankelijk van vele toevalsfactoren: het verband zal alleen worden opgemerkt indien de betrokken aangiften één of meer opvallende kenmerken delen, en wanneer ze in korte tijd aan hetzelfde bureau door dezelfde medewerker in het HKS worden ingevoerd. In alle andere gevallen zal het verband niet worden opgemerkt.

In zijn algemeenheid kan daarom worden opgemerkt dat hierdoor potentieel waardevolle informatie uit het HKS onvoldoende benut wordt.

Zoekvragen komen in veruit de meeste gevallen voort uit opsporingsonderzoek. De werkwijze hierbij is doorgaans de volgende: een rechercheur (of in een enkel geval een analist) beschikt over gegevens uit het onderzoek en wil controleren of hij deze kan koppelen aan gegevens in het HKS. Het kan hierbij gaan om zeer uiteenlopende soorten informatie en om zoekvragen die kunnen variëren van eenvoudig tot complex. Omdat een rechercheur zelf geen autorisatie heeft om het HKS te bevragen, moet hij zijn zoekvraag voorleggen aan een medewerker van de herkenningdienst (HKD). Deze HKD-er is geautoriseerd om het systeem te bevragen. Bovendien beschikt deze medewerker over de noodzakelijke kennis om het naar huidige maatstaven ingewikkelde commandogestuurde programma te kunnen bedienen.²

² Er zijn ook andere functionarissen die de bevoegdheid hebben om het HKS te bevragen, zoals analisten en ploegchefs. Deze laatste groep heeft maar een beperkte toegang en kan alleen gegevens omtrent concrete personen opvragen indien ze beschikken over een naam of adres of iets dergelijks (het zogenaamde ‘beperkt bevragen’ in HKS). Daarnaast zijn aan rechteamts soms ‘administratieve medewerkers’ verbonden die net als medewerkers van de HKD een autorisatie hebben om gegevens in het HKS in te voeren en het systeem

Vanuit het oogpunt van het formuleren van een goede zoekvraag is deze situatie niet optimaal. Een goede zoekvraag komt tot stand door zowel de beschikbare zoekgegevens (waarmee men zoekt) als de beschikbare doelgegevens (waarnaar men zoekt) zo optimaal mogelijk te benutten. Een rechercheur beschikt doorgaans over heel veel kennis uit het onderzoek, maar is meestal niet goed op de hoogte van de mogelijkheden van het HKS, waardoor vaak niet alle bruikbare gegevens uit het onderzoek worden benut voor de zoekvraag. De HKD-er, die het HKS wel van binnen en van buiten kent, kan hierbij een adviserende rol spelen, en proberen de zoekvraag van de rechercheur in sommige gevallen te ‘optimaliseren’. Zijn handicap is echter dat hij niet beschikt over de gegevens uit het opsporingsonderzoek en derhalve ook niet altijd in staat zal zijn bij de relevante gegevens uit te komen.

Hierbij kan nog worden opgemerkt dat de HKD-er in de huidige situatie niet geprikkeld wordt om mee te denken teneinde de zoekvraag te verbeteren. Daartoe ontbreekt zijn belang (en soms wellicht ook de tijd of ervaring). Als het bij wijze van spreken erg druk is, zal hij een willekeurige zoekvraag uitvoeren, maar ook niet veel meer dan dat. De inspanning is er dan vooral op gericht een aanvaardbare hoeveelheid output te krijgen (niet te veel en niet te weinig), zodat de rechercheur verder kan met zijn onderzoek.

Een bijkomend probleem is dat het zoekstelsel in het HKS commandogestuurd is. Dit prikkelt de gebruiker niet om zichzelf de vraag te stellen of hij alle mogelijk bruikbare gegevens ook heeft benut. In een menugestuurd zoekstelsel kan de gebruiker geattendeerd worden op alle mogelijkheden die hij heeft om te zoeken, waardoor de kans kleiner is dat potentieel relevante gegevens onbenut blijven.

Samenvattend kunnen we opmerken dat er in de huidige praktijk enkele mechanismen werkzaam zijn die niet bevorderlijk zijn voor het formuleren van de meest optimale zoekvragen, waardoor beschikbare en mogelijk relevante gegevens (uit zowel het opsporingsonderzoek als het HKS) ongebruikt blijven. De praktijk is nu vaak dat de HKD-er relatief eenvoudige zoekvragen krijgt voorgelegd, gebaseerd op min of meer geïsoleerde gegevens uit het opsporingsonderzoek.

2.2 Uitvoeren van een zoekopdracht

Het HKS biedt de gebruiker, als deze daartoe geautoriseerd is, de mogelijkheid om het systeem ‘beperkt’ of ‘uitgebreid’ te bevragen.

Beperkt bevragen houdt in dat de vrager beschikt over bepaalde personalia zoals naam, adres en dergelijke en wil controleren of de betrokken persoon voorkomt in het systeem. Als dit het geval is, kan een belangrijk deel van de gegevens over deze persoon worden ingezien. Om redenen die verband houden met het privacyreglement kan de vrager bij deze optie geen gegevens inzien over medeverdachten.

Het HKS kan uitgebreid bevraagd worden op aangiften of personen (verdachten en vermiste personen). Bij deze optie hoeft de vrager niet te beschikken over specifieke gegevens om het systeem te kunnen bevragen (zoals hiervoor bij het ‘beperkt bevragen’). Hij kan nu

beperkt en uitgebreid te bevragen. Deze medewerkers zijn echter niet in alle recherche-onderzoeken beschikbaar. In die gevallen is een rechercheur aangewezen op de ‘externe’ HKD-medewerker.

bijvoorbeeld ook naar personen zoeken op basis van bepaalde signalementkenmerken of op basis van delictgegevens.

Afhankelijk van de informatiebehoefte van de vrager, zal de zoekvraag dus worden gesteld aan de ‘aangiftenkant’ van de database of aan de ‘personenkant’. Zo kan iemand bijvoorbeeld een selectie willen zien van alle aangiften van diefstal met geweld in een bepaalde periode waarvan een verdachte bekend is (bevraging aan de aangiftenkant). Omgekeerd is het ook mogelijk dat iemand een selectie wil zien van alle personen die in een bepaalde periode verdacht zijn geweest van diefstal met geweld (bevraging aan de personenkant). Behalve dat de invalshoek verschillend is, zal ook het zoekresultaat in deze gevallen niet helemaal identiek zijn. Het voert echter te ver om deze kwestie hier verder te verkennen.

Zonder in te willen gaan op de details van het zoekproces³, willen we toch een globaal beeld schetsen van de huidige zoekmethode in het HKS. De HKD-er kan een groot aantal velden in het HKS bevragen, zoals de locatie van het strafbare feit, de naam van de aangever, specifieke kenmerken van het delict, specifieke signalementkenmerken van de (onbekende) dader, de personalia van een bekende verdachte, etc. Bijna alle individuele velden waarin opsporingsinformatie is opgeslagen, kunnen worden bevraagd. Bij het formuleren van de zoekvraag kunnen één of meer van deze velden worden gespecificeerd. Per veld kunnen vervolgens één of meer waarden worden opgegeven waaraan het zoekresultaat moet voldoen. Het gebruik van enkele (logische en relationele) operatoren maakt het mogelijk de zoekvraag zodanig te ‘richten’ dat precies dát resultaat wordt verkregen dat wordt gewenst. Een voorbeeld van een zoekvraag aan de aangiftenkant zou bijvoorbeeld kunnen zijn: selecteer alle aangiften die voldoen aan de volgende kenmerken:

- de aangifteperiode is vanaf 1 januari 2002 tot en met 1 maart 2002 en
- de plaats van het misdrijf is Zoetermeer of Leidschendam of het wetsartikel is Wsr300 of Wsr312

Bij deze vraag worden dus alle aangiften geselecteerd die zijn gedaan in de periode van 1-1-2002 tot en met 1-3-2002, die bovendien voldoen aan de volgende kenmerken: ze hebben betrekking op misdrijven in Zoetermeer of Leidschendam (ongeacht het type misdrijf) OF ze hebben betrekking op één van de genoemde wetsartikelen (ongeacht de plaats van het misdrijf).⁴

Met betrekking tot het uitvoeren van een zoekopdracht in de huidige praktijk kunnen we verschillende (mogelijke) beperkingen signaleren. De eerste groep van beperkingen ligt in de beperkte beschikbaarheid van gegevens in het HKS.

- De eerste factor die hierbij een rol speelt is dat het HKS regionaal is opgezet. Er zijn dus alleen regionale gegevens beschikbaar, dat wil zeggen gegevens over verdachten woonachtig in de regio en gegevens over misdrijven gepleegd in de regio (inclusief gegevens over eventuele verdachten van buiten de regio). Deze geografische begrenzing begrenst ook de mogelijkheden om gegevens over delicten respectievelijk personen te koppelen (denk hierbij bijvoorbeeld aan vaak verhuizende verdachten of aan verdachten die op zeer uiteenlopende plaatsen hun misdrijven plegen). Weliswaar kan via de centrale verwijzingsindex (CVI) informatie worden opgevraagd uit het HKS van een andere regio, maar dit gebeurt ‘handmatig’ via een speciaal daartoe ingerichte procedure. Bovendien is de mogelijkheid om vervolgens verdere zoekopdrachten uit te voeren in het HKS van de andere regio zeer beperkt, laat staan

³ Daarvoor zij verwezen naar de handleidingen en andere publicaties over het HKS van het ITO (afd. Applicaties) in Driebergen.

⁴ De operatoren die in de huidige bevragsingsmodule van HKS gebruikt kunnen worden, zijn: ‘en’, ‘of’, ‘en-niet’ (logische operatoren) en ‘=’, ‘<’, ‘>’, ‘IN’ (relationele operatoren).

dat het mogelijk is de gegevens uit de andere regio te kunnen koppelen aan de gegevens die gevonden zijn in de eigen regio, teneinde de gevonden informatie te kunnen veredelen. Het wordt nog ingewikkelder als de antecedenten van bijvoorbeeld één verdachte over veel verschillende regio's zijn 'uitgesmeerd'. Weliswaar worden de antecedentgegevens van een verdachte van buiten de regio in sommige gevallen doorgestuurd naar de regio waar de verdachte woonachtig is, maar deze informatie is om verschillende redenen zeer beperkt.⁵ Al met al gaan er dus mogelijkheden tot matching verloren door de geografische compartimentering van het opsporingsregister.

- Een tweede factor die we in dit verband kunnen noemen, is de achterstand die vaak bestaat in de invoer van gegevens in het HKS. In de regio Haaglanden lagen in mei 2002 circa 2000 zaken (processen-verbaal van oplossing) op de plank die nog ingevoerd moesten worden. Het gaat hierbij vooral om zaken waarbij verdachten gekoppeld moeten worden aan aangiften die al in het systeem opgenomen zijn. Deze achterstand beperkt de mogelijkheden om het HKS optimaal te kunnen benutten voor tactische en/of operationele (zoek-)doeleinden.
- Een derde factor waarmee altijd rekening gehouden moet worden als het om databases gaat, zijn de beperkingen die voortvloeien uit een gebrekkige kwaliteit van gegevens. Denk hierbij aan het ontbreken van gegevens of aan foutieve invoer van gegevens. Gelet op de kwaliteitscontrole die doorgaans plaatsvindt ten aanzien van de invoer van gegevens in het HKS, mogen we vooralsnog aannemen dat foutieve invoer tot een aanvaardbaar minimum is beperkt. In hoeverre ontbrekende gegevens de mogelijkheden tot matching beperken, kan vooraf niet helemaal beoordeeld worden. Wel is duidelijk dat met name de signalementgegevens van zowel bekende verdachten als onbekende daders in het HKS in dit opzicht beperkt zijn. Van alle bekende verdachten in het HKS beschikt zo'n 13% over signalementgegevens (meetpunt: mei 2002). Het gaat om personen die (op enig moment) in verzekering zijn gesteld. De signalementgegevens van onbekende daders zijn om begrijpelijke redenen ook beperkt. Bij de meeste aangiften zijn in het geheel geen signalementgegevens van de dader(s) bekend en bij de aangiften waar dit wel het geval is (minder dan 3% van het totaal aantal aangiften), zijn de gegevens meestal summier.⁶
- De vierde factor die genoemd kan worden in verband met de beperkte beschikbaarheid van gegevens is dat het HKS mogelijkheden biedt voor informatieveredeling die nu niet benut worden. Door gegevens uit verschillende velden van de database met elkaar te combineren, kan 'nieuwe' informatie worden gecreëerd waarop ook gezocht en gematcht kan worden. Het gaat er tenslotte om de beschikbare gegevens zo optimaal mogelijk te benutten. Van deze mogelijkheid maakt de huidige zoekroutine in het HKS geen gebruik.
- De laatste factor die in dit verband genoemd kan worden, heeft betrekking op het *updaten* van bepaalde persoonsgegevens van verdachten wanneer er nieuwe antecedenten bekend zijn. Deze gegevens worden aangepast wanneer hierin veranderingen zijn opgetreden (zoals het woonadres, eventueel uiterlijke kenmerken en dergelijke als er een nieuw signalement wordt aangemaakt). De 'oude' gegevens worden hierbij overschreven. De beperking die hieruit voortvloeit is dat de mogelijkheid wordt weggenomen om de historie van een verdachte, in termen van zijn

⁵ De eerste reden is dat dit doorsturen heel vaak niet meer gebeurt, waardoor het overzicht verre van compleet is. Daarnaast is de doorgestuurde informatie zeer summier (alleen pv-nummer, datum en wetsartikel van het betreffende misdrijf).

⁶ De HKD hanteert wel als richtlijn dat signalementgegevens bij een aangifte OD pas worden ingevoerd als er minimaal 3 kenmerken bekend zijn.

persoonsgegevens, in kaart te brengen (dit geldt vooral in Den Haag). Door deze gegevens te overschrijven en niet te bewaren, gaat eveneens de mogelijkheid verloren deze specifieke persoonskenmerken te koppelen aan het delictgedrag van de verdachte. Een simpel voorbeeld kan zijn dat de verdachte altijd delicten pleegt in de buurt van zijn woonadres. De combinatie van deze twee gegevens, pleegadres en woonadres, genereert extra informatie om op te zoeken. Deze informatie kan vanuit het huidige HKS echter niet gegenereerd worden, omdat woonadressen niet gekoppeld zijn aan antecedenten.

Een andere mogelijke beperking van de huidige praktijk kan worden gevonden in de toegepaste zoekmethode (die hiervoor kort beschreven is). Gegevens uit het opsporingsonderzoek worden gematcht met gegevens uit het HKS op het niveau van (combinaties van) kenmerken van individuele aangiften of verdachten. Dit gebeurt op grond van aannames vooraf omtrent de kenmerken (en de waarden van die kenmerken) die van belang worden geacht. De uitkomst van het zoekproces is derhalve altijd overeenkomstig de vooraf geformuleerde verwachtingen: men zoekt naar aangiften of personen met bepaalde kenmerken en zo ziet het resultaat van de zoekopdracht er ook uit. Aan deze methode kleven diverse beperkingen.

- Er worden aannames gemaakt omtrent datgene wat constant is in bijvoorbeeld het delictgedrag van de dader; men heeft te maken met een serie inbraken en gaat in het HKS zoeken naar inbrekers. Of men heeft te maken met bepaalde modus operandi en gaat daarnaar op zoek. Wat echter constant is in het gedrag van de dader zal van persoon tot persoon, van delict tot delict, van tijdstip tot tijdstip, van plaats tot plaats, etc., sterk kunnen verschillen. In plaats van aannames hierover te formuleren, is het ook mogelijk een zoekmethode te gebruiken die het 'constante' in het gedrag van de dader niet veronderstelt, maar constateert (door gebruik te maken van de patronen - lees: constantes- die in de gegevens van het HKS kunnen worden aangetroffen).
- Een tweede beperking van de genoemde zoekmethode ligt in de operationalisering van wat constant verondersteld wordt in het gedrag van de dader. Hiervoor werd het voorbeeld genoemd van de serie inbraken en het zoeken naar inbrekers in het HKS. De veronderstelling hierbij is dat iemand die nu inbraken pleegt dat wellicht ook eerder gedaan heeft en als zodanig gevonden moet kunnen worden in het HKS. Het delictgedrag wordt dus in de tijd constant verondersteld. Het is echter ook mogelijk om niet zozeer het gedrag van de verdachte/dader in de tijd constant te veronderstellen, maar de ontwikkeling van dit gedrag in de tijd. Wie naar de delictprofielen van verdachten kijkt, ziet dat er zelden sprake is van specialisme. Veel vaker is sprake van een constante in de *ontwikkeling* van het delictgedrag: men begint met winkeldiefstal, dan volgt diefstal met geweld, etc. Kortom, niet alleen in het gedrag zelf, maar ook in de ontwikkeling ervan kan een patroon (een constante) waarneembaar zijn. De huidige zoekmethode houdt hiermee geen rekening en is als zodanig beperkt.
- Een derde beperking die aan deze zoekmethode kleeft is dat niet alle gegevens die mogelijk beschikbaar zijn in het opsporingsonderzoek worden benut voor de zoekvraag, zodat mogelijk relevante gegevens onbenut blijven. Om het voorbeeld van hiervoor aan te houden: men zoekt naar inbrekers en/of naar daders met bepaalde modus operandi, maar de plaats en het tijdstip zijn misschien wel de constanten in het gedrag van de dader. Door die gegevens niet mee te nemen in de zoekvraag, zal in dit geval geen goed resultaat worden verkregen. Vanuit de optiek van een effectieve zoekmethode verdient het derhalve aanbeveling om alle beschikbare gegevens uit het onderzoek te benutten voor een zoekvraag.

- Een vierde beperking van de hiervoor beschreven zoekmethode vloeit voort uit het feit dat geen gebruik wordt gemaakt van het feit dat bepaalde gegevens in het HKS geclusterd zijn, waardoor er extra informatie uit de database kan worden verkregen. Simpel gezegd gaat het hierbij om delicten die geclusterd zijn binnen personen of personen die geclusterd zijn binnen delicten. Meest in het oog springende voorbeeld is de ‘verzameling antecedenten’ per verdachte (clustering van delicten binnen personen). Van veel verdachten is meer dan één strafbaar feit bekend. Door te kijken naar het totaal (en de combinatie) van de gepleegde feiten, komen we meer te weten over het delictgedrag van een specifieke verdachte. Dit biedt een extra mogelijkheid om te matchen. Een ander voorbeeld is de clustering van personen binnen delicten in de vorm van samenwerkingsverbanden. Gegevens over combinaties van personen (onbekende daders of bekende verdachten) die betrokken zijn bij een delict, kunnen behulpzaam zijn bij het matchen. Bij de huidige zoekmethode blijven deze mogelijkheden onbenut.

Tenslotte willen we hier nog wijzen op het feit dat het uitvoeren van met name complexere zoekopdrachten gehinderd wordt door het commandogestuurde zoekstelsel dat het HKS kent. Met name bij deze complexere zoekopdrachten loopt de gebruiker een groter risico op het maken van (syntax-)fouten, met alle gevolgen van dien. Zeker bij dit soort zoekopdrachten zal het voor een buitenstaander, zoals een onderzoeker, moeilijk zijn de eventuele fout te ontdekken. Een menugestuurde ‘omgeving’ kan dit probleem ondervangen.

Samenvattend kunnen we opmerken dat het uitvoeren van een zoekopdracht in het HKS gekenmerkt wordt door beperkingen van uiteenlopende aard. Om verschillende redenen is de beschikbaarheid van gegevens in het HKS beperkt te noemen. Dit belemmert een optimale zoekprocedure. Daarnaast hebben we opgemerkt dat de gehanteerde zoekmethode in een aantal opzichten voor verbetering vatbaar is. Tenslotte zagen we dat er foutenrisico’s voortvloeien uit het feit dat de zoekvragen in een commandogestuurde ‘omgeving’ worden geformuleerd.

2.3 Presenteren van het zoekresultaat

Als de zoekvraag is geformuleerd en de zoekopdracht is uitgevoerd, verschijnt het resultaat op het beeldscherm. Omdat de gevraagde informatie zelden op één scherm past, zijn er verschillende mogelijkheden om door te klikken naar andere schermen waar zich detailinformatie bevindt of informatie over andere aangiften respectievelijk verdachten. Deze output kent de volgende beperkingen.

De eerste beperking is dat de (indicatie van) bruikbaarheid van het gevonden resultaat voor het opsporingsonderzoek onbekend is. Dit is weinig efficiënt, omdat misschien veel tijd wordt gestoken in het doorrecherchen op een zoekresultaat met ‘geringe potentie’.

Er zijn twee oorzaken die verband houden met de onbekende bruikbaarheid van het resultaat. De eerste oorzaak heeft te maken met de huidige zoekmethode. Die leidt ertoe dat het resultaat per definitie voldoet aan de eisen die de onderzoeker hieraan vooraf heeft gesteld (hij zoekt bijvoorbeeld inbrekers met bepaalde modus operandi uit een bepaald deel van de stad en verkrijgt een resultaat dat precies aan deze kenmerken voldoet). Bij de huidige zoekmethode is de bruikbaarheid van het zoekresultaat derhalve gekoppeld aan de aannames die ten grondslag liggen aan de zoekvraag. We weten echter niet hoe vaak en in welke mate

een gevonden zoekresultaat aanleiding geeft tot stappen in het opsporingsonderzoek die betekenisvol zijn voor het succesvolle verloop van een onderzoek.

Een tweede oorzaak die verband houdt met de onbekende bruikbaarheid van de verkregen zoekresultaten, is gelegen in het feit dat er geen systematische feedback plaatsvindt vanuit het opsporingsonderzoek: de HKD-er voert een zoekopdracht uit en probeert dat zo goed mogelijk te doen, maar hij hoort zelden of nooit –en zeker niet systematisch- wat het resultaat was van zijn operatie. Aldus heeft hij geen mogelijkheid om van zijn (zoek-)ervaringen in het verleden te leren.

In de huidige praktijk wordt de bruikbaarheid van het zoekresultaat vooral geformuleerd in termen van de omvang van de output: te weinig output is niet okay en teveel ook niet. Het laatste betekent immers dat het doorrechercheren teveel tijd gaat kosten. Bruikbaarheid van het zoekresultaat is in de huidige praktijk derhalve niet een inhoudelijk, maar vooral een pragmatisch criterium, gebaseerd op de beperkingen van het systeem.

Het HKS kent ook technische beperkingen bij het presenteren van de output. De eerste is dat het systeem niet meer dan ‘1000 records’ output kan produceren. Hierdoor worden voor een rechercheur (of analist) de mogelijkheden beperkt om nadere analyses of rechercheactiviteiten uit te voeren op grotere gegevensverzamelingen. De output wordt bovendien alleen op papier geleverd in de vorm van rapporten die een weergave vormen van de informatie die op de verschillende ‘resultaatschermen’ zichtbaar is (d.w.z. *casewise*). De aard en vorm van de output beperken de rechercheur (of analist) sterk in zijn mogelijkheden de gegevens nader te analyseren. Hij moet immers alle gegevens handmatig verwerken en dat is niet alleen inefficiënt, maar wellicht ook ineffectief (omdat hij bij grotere gegevensverzamelingen sneller verbanden over het hoofd zal zien). Deze problemen treden vooral op wanneer de hoeveelheid informatie groter is. Om deze reden zijn rechercheurs in de huidige praktijk meestal gefocust op een beperkte output (bijvoorbeeld hooguit 20-30 aangiften of hooguit 10-15 personen).

Samenvattend kunnen we opmerken dat het HKS bij de presentatie van het zoekresultaat enkele beperkingen kent. Zo is de bruikbaarheid van het verkregen resultaat voor het opsporingsonderzoek onbekend. Bovendien kent het HKS technische beperkingen, waardoor de vorm van de output niet optimaal is (in termen van efficiency en effectiviteit) voor verdere verwerking door een rechercheur of analist.

3 Samenvatting: beperkingen van de huidige praktijk

In de voorgaande paragrafen hebben we in grove lijnen de huidige praktijk geschilderd van het bevragen van het HKS voor operationele en tactische doeleinden waar het gaat om zoekvragen waarop het antwoord vooraf niet bekend is. Hierbij hebben we vooral gekeken naar beperkingen van het systeem, in termen van bevraging, uitvoering van de zoekopdracht en presentatie van het resultaat. Deze beperkingen vatten we hier nog eens samen.

- *de mogelijkheden om het HKS te gebruiken voor zoekvragen van tactische aard zijn tot op heden on(der)benut gebleven*
- *operationele zoekvragen worden niet optimaal gesteld (d.w.z. gebruikmakend van alle beschikbare gegevens)*
 - o omdat de vrager (meestal een rechercheur) geen kennis heeft van het HKS en de uitvoerder (de HKD-er) geen kennis heeft van het opsporingsonderzoek
 - o omdat de HKD-er geen direct belang heeft bij een goed zoekresultaat

- omdat het zoeken commandogestuurd is, waardoor de vrager en gebruiker niet worden gestimuleerd alle beschikbare gegevens over een bepaalde zaak of persoon in te voeren (i.t.t. een menugestuurd zoekstelsel)
- *zoekopdrachten worden niet optimaal uitgevoerd*
 - omdat de gegevens in het HKS beperkt zijn
 - door regionale compartimentering van het stelsel
 - door achterstand bij de invoer van gegevens
 - door foutief ingevoerde of ontbrekende gegevens
 - door gegevens uit verschillende velden niet te combineren en zo extra informatie te genereren
 - door het *updaten* van sommige persoonsgegevens, waardoor historische informatie over verdachten verloren gaat en tevens de mogelijkheid tot het creëren van extra informatie (koppeling van persoons- aan delictgegevens)
 - omdat de gehanteerde zoekmethode beperkt is
 - doordat gebruik wordt gemaakt van aannames omtrent de aan te treffen patronen in de gegevens (in plaats van uit te gaan van geconstateerde patronen)
 - doordat het constante in het gedrag van de dader slechts op één manier wordt uitgelegd (namelijk door individuele delictkenmerken in de tijd constant te veronderstellen)
 - doordat niet alle beschikbare gegevens uit het opsporingsonderzoek worden benut
 - doordat geen gebruik wordt gemaakt van het feit dat bepaalde gegevens in het HKS geclusterd zijn (zoals delicten binnen een verdachte en verdachten binnen een delict)
 - omdat het commandogestuurde karakter van het zoekprogramma vooral bij complexere zoekopdrachten leidt tot een grotere kans op het maken van (syntax-)fouten
- *zoekresultaten worden niet optimaal gepresenteerd*
 - omdat er geen indicatie is van de bruikbaarheid van het zoekresultaat voor het opsporingsonderzoek
 - dit vloeit voort uit de gehanteerde zoekmethode
 - en uit het ontbreken van feedbackmechanismen in de praktijk
 - omdat technische beperkingen de mogelijkheden voor verdere verwerking van het zoekresultaat belemmeren
 - doordat de output in omvang beperkt is tot maximaal 1000 records
 - doordat de output niet digitaal is (maar op papier)
 - doordat de aard van de output onoverzichtelijk en niet ‘verwerkingsvriendelijk’ is (gegevens worden per individuele aangifte of verdachte gepresenteerd)

De conclusie mag zijn dat in de huidige ‘zoekpraktijk’ beperkingen van uiteenlopende aard kunnen worden aangetroffen. Dit laat onverlet dat de huidige mogelijkheden van het bevragen van het HKS nu en in de toekomst hun nut (zullen) bewijzen als het gaat om zoekvragen waarbij het gewenste zoekresultaat vooraf kan worden geformuleerd.

Hier staat de vraag centraal in hoeverre de gesignaleerde beperkingen ruimte geven aan het ontwikkelen van nieuwe, uitgebreidere of verbeterde zoekmogelijkheden. De voornaamste

betekenis van dit overzicht is dan ook om zowel de mogelijkheden als beperkingen voor verdere ontwikkeling van de zoekfunctie in beeld te krijgen. Als zodanig vormt dit overzicht een startpunt en leidraad voor de ontwikkeling van een nieuw zoekinstrument, waarvan de contouren in het volgende hoofdstuk worden geschetst.

2. Uitgangspunten voor een nieuw zoekinstrument

1. Uitgangspunten voor het zoekproces

Zoals in het vorige hoofdstuk reeds werd opgemerkt zijn de uitgangspunten voor het nieuwe zoekinstrument in eerste instantie afgeleid van de beperkingen die we hebben aangetroffen in de huidige praktijk. Een deel van deze beperkingen zal echter ook van toepassing zijn op een nieuw te ontwikkelen zoekinstrument (zoals bijvoorbeeld het feit dat het HKS regionaal georganiseerd is). Dit deel laten we daarom verder buiten beschouwing. Voor wat betreft de andere gesignaleerde beperkingen, kunnen we constateren dat deze deels betrekking hebben op organisatorische en juridische kwesties en op technische tekortkomingen van het systeem. Het werken aan dit type beperkingen valt buiten het kader van dit onderzoek. In het kader van dit onderzoek zullen we ons richten op het wegnemen van mogelijke beperkingen ten aanzien van:

1. de toepassing van het zoekinstrument voor tactische doeleinden
2. de toepassing van het zoekinstrument voor meer soorten zoekvragen
3. het optimaal gebruiken van de beschikbare gegevens in het HKS
4. het hanteren van optimale zoekmethoden
5. het verschaffen van bruikbaarheidscriteria bij de presentatie van een zoekresultaat

De eerste 5 (groepen van) uitgangspunten die hierna worden geformuleerd, zijn gebaseerd op de beperkingen die we ten aanzien van deze zaken hebben gesignaleerd in de huidige praktijk. Ze worden aangevuld met 3 algemene uitgangspunten. Dit levert de volgende lijst van uitgangspunten op:

Ad 1) Onderzoek de mogelijkheid om het zoekinstrument voor zowel operationele als tactische zoekdoelen te gebruiken

De bedoeling hiervan is te onderzoeken in hoeverre zoekmethoden voor operationele doeleinden ook geschikt zijn of eenvoudig geschikt gemaakt kunnen worden voor tactische doeleinden. Dit zou een uitbreiding van de toepassingsmogelijkheden kunnen betekenen. Onder operationele doeleinden verstaan we hier het matchen van gegevens uit een individueel opsporingsonderzoek teneinde een dader op het spoor te komen dan wel een bekende verdachte te koppelen aan delicten waarvan de dader nog niet bekend is. Onder tactische doeleinden verstaan we hier het matchen van gegevens van verschillende delicten of personen teneinde vast te kunnen stellen welke van deze delicten of personen de meeste aanknopingspunten bieden voor een te starten opsporingsonderzoek.

Ad 2) Onderzoek welke zoekvragen aan het HKS kunnen worden voorgelegd

In de huidige praktijk worden vaak zoekvragen gesteld waarbij de vrager van tevoren aangeeft aan welke criteria het resultaat moet voldoen. Dit zijn vaak simpele zoekvragen gebaseerd op een beperkt aantal gegevens omtrent een delict of persoon. Wat zijn echter de mogelijkheden om te zoeken wanneer we de zoek- en doelgegevens compleet in kaart brengen? Welke zoekvragen zijn dan mogelijk? Het is van belang om hierin inzicht te krijgen, omdat het wellicht zicht biedt op nieuwe, tot op heden onbenutte, mogelijkheden om in het HKS te zoeken.

Ad 3a) Benut zoveel mogelijk alle gegevens uit het opsporingsonderzoek voor de zoekopdracht

Bij de huidige werkwijze wordt vooraf een selectie gemaakt van kenmerken (van een aangifte of persoon) waarin men geïnteresseerd is. Op die kenmerken gaat men zoeken in het HKS. Aldus blijven andere kenmerken (van de betrokken aangifte of persoon), die bekend zijn en mogelijk ook relevant, buiten beeld in de zoekprocedure. Het algemene idee is dat, in een overigens vergelijkbare situatie, het gebruik van méér gegevens bij een zoekopdracht de kans op een goed zoekresultaat zal vergroten.

Ad 3b) Onderzoek de mogelijkheden om extra informatie te creëren door gegevens uit bestaande velden van de database te combineren

Dit is in het vorige hoofdstuk al aan de orde geweest: door bestaande gegevens in het HKS te combineren tot nieuwe gegevens kan extra informatie worden verkregen waarop ook gezocht kan worden.

Ad 3c) Maak gebruik van het feit dat bepaalde gegevens binnen eenheden geclusterd zijn om zodoende extra informatie te creëren

De huidige zoekmethode maakt geen gebruik van de situatie dat bepaalde gegevens in het HKS geclusterd zijn binnen eenheden, zoals delicten/antecedenten geclusterd zijn binnen individuele personen. Dit maakt het bijvoorbeeld mogelijk op basis van een serie delicten/antecedenten een delictprofiel van een verdachte te bepalen. Het informatiegehalte van een dergelijk profiel zal veel hoger liggen dan van de afzonderlijke antecedenten, waardoor er wellicht betere mogelijkheden ontstaan om bijvoorbeeld nieuwe aangiften aan bekende verdachten te koppelen. Een ander voorbeeld is dat bekende verdachten/onbekende daders geclusterd kunnen zijn binnen delicten/aangiften (in de vorm van samenwerkingsverbanden). Ook dit gegeven biedt extra mogelijkheden tot matching.

Ad 4a) Maak zo weinig mogelijk aannames omtrent de selectie van gegevens in het zoekresultaat

Hoe meer aannames je vooraf maakt omtrent de selectie van gegevens in het zoekresultaat, des te groter is de kans dat er één onjuiste aanname bij zit, waardoor het zoekresultaat als geheel waardeloos wordt. Bij de ontwikkeling van zoekmethoden moet hiermee rekening worden gehouden door methoden te ontwikkelen waarbij vooraf zo weinig mogelijk aannames worden gemaakt omtrent de selectie van gegevens in het zoekresultaat. Het algemene idee is dat bepaalde kenmerken (in het gedrag) van de dader constant zullen zijn en bepaalde kenmerken zullen variëren. Het matchen van persoons- en delictgegevens is gebaseerd op het vinden van constanten in de waarden van deze kenmerken. Welke deze constanten zijn, is van tevoren niet duidelijk: deze zullen variëren tussen individuele daders, tussen plaatsen, tussen tijdstippen, etc. Uitgangspunt bij het ontwikkelen van nieuwe zoekmethoden moet derhalve zijn, dat deze methoden zodanig zijn dat de gegevens in het HKS een optimale ‘kans krijgen’ om eventuele constanten aan de gebruiker te tonen. Door vooraf aannames te formuleren over de selectie van gegevens in het zoekresultaat, ontnemt de gebruiker zichzelf het zicht op constanten in het gedrag van de dader die hij niet vermoedt maar die er wel zijn.

Ad 4b) Onderzoek op welke wijze(n) constantes in het gedrag van een dader kunnen worden vastgesteld

De bestaande zoekmethode is gebaseerd op gedragsconstanten in de tijd. Dat wil zeggen, iemand die ooit via methode X een inbraak heeft gepleegd en als zodanig geregistreerd staat in het HKS, komt in aanmerking als verdachte van een aangifte van inbraak via methode X. In

de vorige paragraaf hebben we aangegeven dat de constante in het gedrag van de dader ook op een andere manier vastgesteld kan worden, door bijvoorbeeld te kijken naar de ontwikkeling van zijn delictgedrag in de tijd (gerelateerd aan de ontwikkeling daarvan bij andere, 'soortgelijke' verdachten). De bedoeling is om inzicht te krijgen in de verschillende manieren waarop naar gedragspatronen in de gegevens kan worden gezocht en vervolgens vast te stellen in welke mate deze aansluiten bij de patronen die in de gegevens zelf kunnen worden aangetroffen (via simulaties/analyses op bestaande HKS-gegevens).

Ad 5) Onderzoek de mogelijkheden om bruikbaarheidscriteria te formuleren voor het zoekresultaat

Welke zoekvraag ook wordt gesteld en welke methode hierbij ook wordt gehanteerd, er zal altijd een 'uitkomst' zijn in de vorm van een rijtje personen of aangiften. De vraag is natuurlijk hoe bruikbaar dit 'rijtje' is. In de huidige praktijk bestaan geen criteria waaraan de bruikbaarheid van het verkregen zoekresultaat kan worden getoetst. Toch zou het prettig zijn om een dergelijk criterium te kunnen formuleren, omdat daarmee beoordeeld kan worden hoe zinvol het is om tijd, geld en energie te steken in het nader onderzoeken van het zoekresultaat. In een beperkt aantal gevallen zal zonneklaar zijn dat het zoekresultaat zeer bruikbaar is, bijv. omdat het zonder al te veel omwegen leidt tot een directe en sterke aanwijzing voor een dader. In een beperkt aantal andere gevallen zal meteen duidelijk zijn dat het zoekresultaat hoegenaamd niets oplevert. Daar tussenin ligt echter een grijs gebied, waarin het zoekresultaat niet op het blote oog uitsluitend zal geven over de kwaliteit van de gevonden match en waarbij de analist dus over een hulpmiddel zou moeten kunnen beschikken om te beoordelen of het de moeite waard is het zoekresultaat nader te analyseren. Dit hulpmiddel zou bij voorkeur een kwantitatieve maat moeten zijn die op enigerlei wijze de kwaliteit van de gevonden match aanduidt. Als de waarde op deze maat beneden een bepaald niveau komt, zou men kunnen overwegen het zoekresultaat niet nader te onderzoeken, omdat de kans op het succesvol koppelen van gegevens dan heel klein is. In dat geval kunnen kostbare tijd en energie worden bespaard, die dan beschikbaar komen voor andere, nuttiger zaken.

6) Algemeen uitgangspunt: onderzoek het effect van kansgrootheden op het zoekresultaat

Om tot enige match te kunnen komen van een willekeurige zoekvraag moet aan bepaalde noodzakelijke voorwaarden worden voldaan. Stel een rechercheur beschikt over gegevens van een bepaald delict en hij wil weten of er in het HKS wellicht nog andere delicten vóórkomen die mogelijk van dezelfde dader zijn. De noodzakelijke voorwaarde om in dit geval tot een match te komen is dat één of meer andere delicten van de betreffende dader in het HKS vóórkomen.

Stel nu dat de rechercheur beschikt over signalementgegevens van een dader. Om in dit geval tot een match op persoonsgegevens te komen, moet worden voldaan aan de noodzakelijke voorwaarde dat de dader zelf reeds in het HKS voorkomt: als bekende verdachte (mét een signalement) of als signalement bij een aangifte onbekende dader (OD).

In beide gevallen moet dus worden voldaan aan een voorwaarde om überhaupt tot een match te kunnen komen. De kans echter dat een dader al in het systeem vóórkomt is hoogstwaarschijnlijk een stuk kleiner dan de kans dat zijn (andere) delicten al in het systeem vóórkomen. Jaarlijks worden in Haaglanden circa 100.000 aangiften gedaan, waarvan hooguit 15% wordt opgehelderd. Een deel van de niet-opgehelderde delicten zal gepleegd zijn door daders die al vóórkomen in het HKS (m.n. de veelplegers). In een groot aantal andere gevallen zullen de daders niet bekend zijn, althans niet vóórkomen in het HKS. Hoe deze verhoudingen precies liggen is onbekend, maar in zijn algemeenheid mag je hieruit concluderen dat de kans op een 'personenmatch' kleiner is dan de kans op een 'delictmatch'.

Hierbij moet opgemerkt worden dat wanneer wordt voldaan aan de noodzakelijke voorwaarde dat de dader vóórkomt in het HKS (als signalement bij een aangifte OD of als persoon aan de personenkant van de database), dit automatisch betekent dat ook wordt voldaan aan de voorwaarde dat één of meer van zijn delicten vóórkomen in het HKS.⁷ Immers persoonsgegevens in het HKS zijn altijd gekoppeld aan delictgegevens. Omgekeerd is dit niet het geval: bij delictgegevens in het HKS zijn niet altijd persoonsgegevens bekend. Hiervoor werd reeds opgemerkt dat aan minder dan 3% van de aangiften OD een signalement is gekoppeld van een onbekende dader (= circa 3000 signalementen per jaar). Aan naar schatting 15% van de aangiften is een bekende verdachte gekoppeld. Maar van deze groep is slechts in 1 op de 7 à 8 gevallen een persoonsbeschrijving (in de vorm van een signalement) beschikbaar. Als we voor het gemak aannemen dat 15% van de 100000 aangiften is opgehelderd en dat dit gelijk staat aan 15000 gehoorde verdachten (zijnde unieke personen).⁸ Dan is op jaarbasis gerekend dus van circa 2000 bekende verdachten een signalement bekend. Op jaarbasis beschikken we derhalve over delictgegevens van circa 100000 aangiften en over persoonsgegevens van circa 5000 personen (ca. 3000 signalementen OD en ca. 2000 signalementen BD (grove schatting)).

Uit het voorgaande volgt dat de kans op enige match bij het koppelen van persoonsgegevens kleiner is dan de kans op een match bij het koppelen van delictgegevens. We zien onszelf hier geconfronteerd met een dilemma: algemeen geldt dat hoe meer gegevens we benutten voor het stellen van de zoekvraag, des te groter is de kans op een goede match (vanwege het unieker worden van de combinatie van gegevens). Echter, wanneer er persoonsgegevens aan te pas komen, is sprake van een afnemende kans op een match vanwege de hiervoor genoemde factoren.

Dit dilemma doet zich vooral voor wanneer we beschikken over zowel delict- als persoonsgegevens. De vraag is dan of het gecombineerd matchen van deze gegevens een meerwaarde heeft boven het afzonderlijk matchen van de delictgegevens. Immers, de aanwezigheid van de dader als persoon in het systeem vormt in dit geval een extra voorwaarde om tot een match te komen. In zo'n geval kan het zinvol blijken om alleen op de delictgegevens te matchen. Omgekeerd is dat niet het geval: wanneer we beschikken over zowel delict- als persoonsgegevens is het niet verstandig om alleen te matchen op persoonsgegevens. Dit heeft te maken met het feit dat persoonsgegevens in alle gevallen gekoppeld zijn aan delictgegevens. Het gebruik van delictgegevens bij het matchen levert hier dus geen nadere voorwaarde voor enig zoekresultaat op. Het zou dan onverstandig zijn om van de beschikbaarheid van deze gegevens geen gebruik te maken.

Nog kleiner dan de kans op het succesvol matchen van individuele personen is de kans om groepen van personen succesvol te matchen, zoals dadercombinaties bij een bepaald misdrijf. De aanwezigheid van elke afzonderlijke dader in het HKS geldt immers als een (extra) noodzakelijke voorwaarde voor enige match. Bovendien geldt hier nog als extra noodzakelijke voorwaarde dat ze ook in *combinatie* bij één of meer delicten in het HKS vóórkomen. Dus om een samenwerkingsverband van 2 daders op persoonskenmerken te kunnen koppelen moet aan 3 voorwaarden worden voldaan: elke afzonderlijke dader moet

⁷ In vóórkomende gevallen kan het echter gaan om verdachten die louter handmatige antecedenten hebben (bijvoorbeeld verdachten uit de regio die elders delicten hebben gepleegd). De beschikbare informatie over de delicten is dan heel summier.

⁸ In werkelijkheid kan één persoon als verdachte aan meer dan één delict gekoppeld zijn. Omgekeerd is het ook mogelijk dat één delict door meer personen is gepleegd. Het gaat ons hier echter niet om de precieze verhoudingen, maar om een ruwe schatting.

vóórkomen in het HKS (= 2 voorwaarden) én ze moeten ook als dadercombinatie in het HKS vóórkomen. Bij een dadercombinatie van 4 praten we over 5 noodzakelijke voorwaarden waaraan voldaan moet worden alvorens tot een match te kunnen komen. Het lijkt, gelet op het grote aantal noodzakelijke voorwaarden waaraan voldaan moet worden om tot enige match te kunnen komen, niet zinvol om (grotere) dadercombinaties te matchen. In het onderhavige onderzoek zullen we wel onderzoeken of er resultaten geboekt kunnen worden met het matchen van individuele dadergegevens. Mocht dit niet het geval zijn, dan heeft het zeker geen zin om nog naar combinaties van personen te kijken.

Samenvattend: uit bovenstaande analyse kunnen we op voorhand afleiden dat bepaalde zoekvragen vaker tot een succesvol antwoord zullen leiden dan andere, simpelweg omdat de noodzakelijke voorwaarden voor succes in het ene geval minder stringent zijn dan in het andere geval. Uitgangspunt voor dit onderzoek is om niet bij voorbaat bepaalde zoekvragen uit te sluiten, omdat de voorwaarden voor succes te stringent zijn. De analyses moeten uitwijzen welke resultaten bij de verschillende zoekvragen geboekt kunnen worden en met welke zoekmethoden. Daarna kunnen we pas onderbouwde uitspraken doen over nut en onnut van bepaalde zoekvragen.

7) Algemeen uitgangspunt: baseer het zoekproces enkel op de in het HKS beschikbare gegevens

Door gegevens uit het HKS te koppelen aan andere (nu niet in het HKS beschikbare) gegevens, kan de informatie omtrent aangiften en verdachten wellicht verder worden verrijkt en kunnen de mogelijkheden om tot succesvolle matches te komen misschien wel worden vergroot. Deze werkwijze zullen wij echter niet toepassen. Het argument hiervoor is dat een nieuw zoekinstrument –indien in aanleg zinvol- zonder drempels moet kunnen worden geïmplementeerd in een bestaande software-omgeving (zoals bijvoorbeeld in DEX2000). Het gebruik van aanvullende gegevens vormt hiervoor een belemmering, doordat extra eisen worden gesteld aan het aanleveren van gegevens door de korpsen.

8) Algemeen uitgangspunt: eenvoudige oplossingen verdienen de voorkeur boven ingewikkelde oplossingen

Ook dit is een voor de hand liggend uitgangspunt. Ingewikkelde oplossingen kosten meer energie (zijn niet zuinig), vergroten de kans op het maken van fouten en zijn minder inzichtelijk, waardoor het moeilijker wordt de resultaten van zo'n oplossing te interpreteren. Deze argumenten pleiten voor eenvoud als leidend principe.

2. Overzicht van zoekdoelen

De zoekvragen die we hierna in paragraaf 4 formuleren, geven richting aan de problemen die het nieuwe te ontwikkelen instrument moet kunnen oplossen. Om bij deze zoekvragen uit te komen, moeten we eerst een antwoord formuleren op de volgende twee vragen:

- Waarom zoeken we / wat zijn mogelijke zoekdoelen?
- Welke gegevens staan tot onze beschikking om *mee* te zoeken (*zoekgegevens*) en om *op* te zoeken (*doelgegevens*) en op welke wijze kunnen we deze gegevens aan elkaar koppelen?

Dit levert een overzicht op van alle logisch mogelijke en praktisch denkbare zoekvragen. In deze paragraaf zetten we de mogelijke zoekdoelen uiteen. In de volgende paragraaf (3) gaan we in op de beschikbare zoek- en doelgegevens.

We kunnen 3 zoekdoelen onderscheiden, welke parallel lopen aan de ‘organisatiedoelen’ voorafgaand aan en tijdens een opsporingsonderzoek.

Het *eerste zoekdoel* is te bepalen welke criminaliteit in aanmerking komt voor opsporingsonderzoek. Naast andere zaken zal vooral de ‘opsporingswaardigheid’ van aangiften hierbij een rol spelen: welke aanknopingspunten zijn er om ‘nieuwe’ aangiften te koppelen aan bestaande aangiften of aan bekende verdachten? Behalve het vaststellen van de opsporingswaardigheid van (nieuwe) aangiften is het ook mogelijk de opsporingswaardigheid van personen/verdachten vast te stellen. Door bijvoorbeeld personen die eerder verdacht zijn van strafbare feiten te koppelen aan nog niet opgehelderde aangiften is het mogelijk per verdachte vast te stellen in hoeverre deze ‘past’ bij de nog niet opgehelderde aangiften. In de praktijk zal deze werkwijze wellicht minder vaak gevolgd worden en zullen de niet opgehelderde aangiften veelal het startpunt vormen voor een opsporingsonderzoek.

Het *tweede zoekdoel* is het opsporen van één of meer daders van één of meer delicten die onderwerp zijn van een opsporingsonderzoek.

Het *derde zoekdoel* is om meer kennis te vergaren over de ‘criminele achtergrond’ van verdachten die in beeld zijn gekomen in een opsporingsonderzoek.

Samengevat kunnen we derhalve de volgende drie zoekdoelen definiëren:

- 1) Het voorafgaand aan een opsporingsonderzoek vaststellen van opsporingsmogelijkheden van verschillende delicten of personen (tactisch zoekdoel);
- 2) Het opsporen van 1 of meer onbekende daders in een opsporingsonderzoek (operationeel zoekdoel);
- 3) Het doorrechercheren op 1 of meer concrete verdachten in een opsporingsonderzoek (operationeel zoekdoel).

3. Welke gegevens zijn beschikbaar en wat kunnen we ermee?

De volgende onderscheidingen zijn van belang voor het beantwoorden van de vraag welke gegevens beschikbaar zijn voor het beantwoorden van zoekvragen in het HKS:⁹

- er zijn *zoekgegevens* beschikbaar (gegevens uit het opsporingsonderzoek) en er zijn *doelgegevens* waaraan deze gekoppeld kunnen worden (gegevens in het HKS);
 - o de beschikbaarheid van de doelgegevens is bepalend voor de relevantie van de zoekgegevens. Met andere woorden: het heeft geen zin te zoeken op gegevens die niet in het HKS beschikbaar zijn;
- de doelgegevens in het HKS hebben betrekking op *kenmerken van delicten*, *kenmerken van personen* (hierna gemakshalve signalementen genoemd) of op een *combinatie hiervan*;
 - o bij delictgegevens kan een nader onderscheid worden gemaakt naar delicten van onbekende daders (OD) en delicten van bekende verdachten (BD);
 - o bij persoonsgegevens kan een nader onderscheid worden gemaakt naar persoonsgegevens van onbekende daders en persoonsgegevens van bekende verdachten;
- de aard van de zoekgegevens bepaalt de *kenmerken waarop gematcht wordt*: er wordt gematcht op persoonskenmerken indien de zoekgegevens enkel signalementinformatie bevatten; er wordt gematcht op delictkenmerken indien de zoekgegevens enkel

⁹ Er zijn uiteraard andere indelingen denkbaar. Maar deze biedt voor het onderhavige onderzoek het meeste gemak.

delictinformatie bevatten (van al of niet bekende personen); en er wordt gematcht op combinaties van beide indien de zoekgegevens zowel delict- als signalementinformatie bevatten;

- de gegevens over delicten en personen zijn verder te onderscheiden naar *aantal*: je beschikt over gegevens betreffende één of meer delicten respectievelijk één of meer personen. Het is in dit verband alleen zinvol om gebruik te maken van samengevoegde gegevens (over ‘meer’ delicten respectievelijk ‘meer’ personen) indien deze gegevens geclusterd zijn binnen ‘natuurlijke’ eenheden, zoals *delicten die geclusterd zijn binnen individuele personen* (de gekende strafbare feiten van een bekende verdachte; hierna aangeduid als delictprofielen) of *personen die geclusterd zijn binnen individuele delicten* (de onbekende daders of bekende verdachten binnen een gekend strafbaar feit; hierna signalementencombinatie OD en BD genoemd). In alle andere gevallen berust clustering van gegevens op het veronderstellen van enig verband tussen de individuele gegevens. Aangezien dit nu juist het doel is van het zoekproces, kan een dergelijke werkwijze niet als uitgangspunt worden gehanteerd.¹⁰

Op grond van deze onderscheidingen kunnen we tot een eerste globale ordening van zoek- en doelgegevens komen. In schema 1 is een dergelijke ordening weergegeven, waarbij de inhoud van de gegevens (delictgegevens, persoonsgegevens of een combinatie van beide) leidend is geweest voor de ordening.

Schema 1 Globale combinaties van zoek- en doelgegevens bij zoekvragen in het HKS

Beschikbare zoekgegevens	Doelgegevens in het HKS		
	<i>alleen delictgegevens</i>	<i>alleen persoonsgegevens</i>	<i>combinatie van beide</i>
<i>alleen delictgegevens</i>	(1) aangiften OD delicten BD delictprofielen	(2)	(3)
<i>alleen persoonsgegevens</i>	(4)	(5) ind. signalementen OD ind. signalementen BD signalementencombi. OD signalementencombi. BD	(6)
<i>combinatie van beide</i>	(7)	(8)	(9) aangiften OD met signalement(en) OD delicten BD met signalement(en) BD delictprofielen met signalement BD

Toelichting op schema 1:

- Algemeen: de grijze vlakken vormen onmogelijke of onnodige combinaties. De vlakken 2 en 4 zijn logisch onmogelijk, omdat delictgegevens niet aan persoonsgegevens kunnen worden gekoppeld en omgekeerd. De vlakken 3,6,7 en 8 zijn onnodig, omdat hier afzonderlijke delict- of persoonsgegevens worden gekoppeld aan gegevens waarin een combinatie van beide voorkomt. Deze matching voegt niets

¹⁰ Een voorbeeld hiervan zijn gegevens omtrent ‘meerdere delicten van één of meer onbekende daders’. Deze clustering is gebaseerd op aannames omtrent het verband tussen de delicten (meestal gaat het om delicten die qua aard, MO, tijd en plaats sterk overeenkomen). Het gebruik van zogenaamde ‘chronolijsten’ door operationeel analisten is op dit idee gebaseerd: het op een rij zetten en vergelijken van ‘vergelijkbare’ delicten. Dit is conceptueel een problematische activiteit, omdat het constante en het variabele in de betrokken delicten volledig wordt bepaald door de keuzes die de analist vooraf zelf heeft gemaakt.

toe aan de mogelijkheden die elders in het schema worden genoemd (in respectievelijk de vlakken 1 en 5).

- T.a.v. vlak 1: de zoekgegevens bestaan enkel uit delictinformatie; bij de delictgegevens in het HKS (de doelgegevens) zijn soms ook persoonsgegevens bekend. Daarvan wordt in dit geval geen gebruik gemaakt.
- T.a.v. vlak 5: de zoekgegevens bestaan enkel uit persoonsinformatie; bij de persoonsgegevens in het HKS (de doelgegevens) zijn altijd ook delictgegevens bekend. Daarvan wordt in dit geval geen gebruik gemaakt.

Uit schema 1 kunnen we 10 soorten zoek- respectievelijk doelgegevens afleiden, namelijk gegevens betreffende:

1. delictkenmerken van 1 delict van onbekende dader (= *aangifte OD*)
2. delictkenmerken van 1 delict van bekende verdachte (= *delict BD*)
3. delictkenmerken van meer delicten van 1 bekende verdachte (= *delictprofiel*)
4. persoonskenmerken van 1 onbekende dader (= *signalement OD*: signalement bij aangifte onbekende dader)
5. persoonskenmerken van 1 bekende verdachte (= *signalement BD*: signalement bekende dader)
6. persoonskenmerken van groep onbekende daders (binnen gekend strafbaar feit) (= *signalementencombinatie OD*)
7. persoonskenmerken van groep bekende verdachten (binnen gekend strafbaar feit) (= *signalementencombinatie BD*)
8. delict- en persoonskenmerken van 1 delict van 1 onbekende dader (= *aangifte OD + signalement OD*)¹¹
9. delict- en persoonskenmerken van 1 delict van 1 bekende verdachte (= *delict BD + signalement BD*)¹²
10. delict- en persoonskenmerken van meer delicten van 1 bekende verdachte (= *delictprofiel + signalement BD*)

De gegevens die binnen één vakje van schema 1 genoemd worden, kunnen met elkaar gematcht worden. Zo kunnen delictkenmerken van een aangifte OD niet alleen gekoppeld worden aan andere aangiften OD, maar ook aan delicten BD en aan delictprofielen. Dit levert een nadere verfijning op van alle mogelijk combinaties van zoek- en doelgegevens. Deze zijn in schema 2 op de volgende pagina weergegeven.

¹¹ Hier zou eventueel nog een uitsplitsing gemaakt kunnen worden naar delict- en persoonskenmerken van 1 delict van meerdere onbekende daders. Deze uitwerking laten we hier voorsnog achterwege, omdat een succesvolle match op dergelijke gegevens pas mogelijk is indien aan 2 extra noodzakelijke voorwaarden is voldaan (zie tekst bij uitgangspunt 6 in paragraaf 1). We proberen eerst deze meer eenvoudige variant uit.

¹² Zie redenering in vorige noot.

Schema 2 Gedetailleerde combinaties van zoek- en doelgegevens

Doelgegevens in het HKS									
delictkenmerken van:			persoonskenmerken van:				delict- en persoonskenmerken van:		
1 aangifte OD	1 delict BD	1 delictprofiel	1 signalement OD	1 signalement BD	1 signalementcombinatie OD	1 signalementcombinatie BD	1 aangifte OD + signalement OD	1 delict BD + signalement BD	1 delictprofiel + signalement BD
<i>delictkenmerken van:</i>									
x	X	x							
x	X	x							
x	X	x							
<i>persoonskenmerken van:</i>									
			x	x	1	1			
			x	2	1	1			
			1	1	x	x			
			1	1	x	2			
<i>delict- en persoonskenmerken van:</i>									
							x	x	x
							x	3	3
							x	3	3

Toelichting op schema 2:

- De donkergrijze vlakken vormen onmogelijke of onnodige verbindingen (zie ook toelichting bij schema 1);
- De lichtgrijze vlakken met nummer 1, 2 en 3 leveren logischerwijs geen zinvolle informatie op (die iets toevoegt aan de andere mogelijkheden in het schema). Bij de velden met nummer 1 gaat het om het matchen van individuele signalementgegevens met gegevens over combinaties van signalementen. Deze optie voegt niets toe aan de mogelijkheid om individuele signalementen te matchen. Bij de velden met nummer 2 en 3 gaat het om het matchen van (onder andere) signalementen van 2 (groepen van) bekende verdachten. Ook dit is uiteraard geen zinvolle activiteit.
- Alle andere velden, de 20 velden met een kruisje, kunnen in beginsel zinvolle informatie genereren.

Samenvattend: om inzicht te verkrijgen in de gegevens die beschikbaar zijn om ‘mee’ en ‘op’ te zoeken, hebben we een categorisering van zoek- en doelgegevens gemaakt en vervolgens alle logisch mogelijke combinaties op een rij gezet en de bij voorbaat onnodige combinaties verwijderd. Dit levert een overzicht op van alle mogelijke combinaties van zoek- en doelgegevens waarmee we in de praktijk te maken kunnen krijgen. Of het hierbij in alle gevallen ook om praktisch zinvolle combinaties gaat, valt in dit stadium niet te zeggen. In paragraaf 4 wordt nagegaan welke zoekvragen we aan de hand van dit overzicht kunnen beantwoorden.

4. Overzicht van zoekvragen

Nu de zoekdoelen zijn bepaald en ook de gegevens bekend zijn waarmee en waarop we kunnen zoeken, kunnen we een overzicht produceren van alle mogelijke zoekvragen die we kunnen stellen.

Elk niet-grijs vakje in schema 2 kan worden beschouwd als een mogelijke zoekvraag. We kunnen derhalve 20 verschillende soorten zoekvragen formuleren. Dit levert een zeer gedetailleerd overzicht op dat we in schema 3 hebben weergegeven. Het gaat hier dus om alle mogelijke combinaties van zoek- en doelgegevens. De zoekvragen in het schema zijn geordend naar de gegevens die beschikbaar zijn om mee te zoeken. In het schema wordt tevens aangegeven welke zoekgegevens beschikbaar (moeten) zijn om de betreffende zoekvraag te beantwoorden, welke doelgegevens nodig zijn, wat het beoogde resultaat van de zoekvraag is en waarom de vraag wordt gesteld, ofte wel welke zoekdoelen kunnen worden gediend.

Neem als voorbeeld de eerste zoekvraag: we beschikken over gegevens (delictkenmerken) van een aangifte met onbekende dader (OD). Deze delictgegevens zijn onze zoekgegevens. We vergelijken de aangifte met andere aangiften OD (de doelgegevens). De vraag die we onszelf hierbij kunnen stellen is of er in de verzameling van de doelgegevens aangiften voorkomen die ‘passen’ bij de aangifte waarmee we zoeken. Het resultaat van deze operatie is een overzicht van delicten die mogelijk door dezelfde dader zijn gepleegd als het zoekdelict. We kunnen de hier beschreven zoekvraag tactisch of operationeel invullen. Als we de vraag voor tactische doeleinden gebruiken, willen we weten in hoeverre het zoekdelict aanknopingspunt biedt voor opsporingsonderzoek (opsporingswaardig is). Als we de zoekvraag voor operationele doeleinden gebruiken, zijn we op zoek naar de onbekende dader van het zoekdelict. Op deze wijze kan ook de rest van het schema worden begrepen.

Niet alle zoekvragen zijn voor dit onderzoek relevant. Deze zijn lichter weergegeven. Zie de uitleg hieromtrent achter schema 3.

Schema 3 Overzicht van mogelijke zoekvragen in het HKS

<i>Beschikbare zoekgegevens</i>	<i>Mogelijke doelgegevens</i>	<i>Zoekvraag</i>	<i>Beoogd resultaat</i>	<i>Mogelijke zoekdoelen</i>
1) delictkenmerken van 1 aangifte OD	Delictkenmerken van 1 aangifte OD	Kunnen we dit delict van een onbekende dader koppelen aan andere delicten van onbekende daders in het HKS?	een overzicht van delicten die mogelijk van dezelfde dader(s) zijn	opsporingswaardigheid van delict bepalen; onbekende dader(s) opsporen
2) delictkenmerken van 1 aangifte OD	delictkenmerken van 1 delict BD	Kunnen we dit delict van een onbekende dader koppelen aan delicten van bekende verdachten in HKS?	een overzicht van delicten die mogelijk van dezelfde dader(s) zijn én van mogelijke verdachten op grond van hun individuele delictkenmerken	opsporingswaardigheid van delict bepalen; onbekende dader(s) opsporen
3) delictkenmerken van 1 aangifte OD	delictkenmerken van 1 delictprofiel	Kunnen we dit delict van een onbekende dader koppelen aan delictprofielen in HKS?	een overzicht van mogelijke verdachten op grond van hun delictprofielen	opsporingswaardigheid van delict bepalen; onbekende dader(s) opsporen
4) delictkenmerken van 1 delict BD	delictkenmerken van 1 aangifte OD	Kunnen we dit delict van een bekende verdachte koppelen aan delicten met een onbekende dader in HKS?	een overzicht van delicten die mogelijk van dezelfde dader zijn	opsporingswaardigheid van dader bepalen; doorrechercheren op concrete verdachte
5) delictkenmerken van 1 delict BD	delictkenmerken van 1 delict BD	Kunnen we dit delict van een bekende verdachte koppelen aan delicten van andere bekende verdachten in HKS?	een overzicht van delicten die mogelijk van dezelfde dader zijn én van mogelijke mededaders van dit delict op grond van hun individuele delictkenmerken	opsporingswaardigheid van dader bepalen; onbekende dader(s) opsporen; doorrechercheren op verdachte
6) delictkenmerken van 1 delict BD	delictkenmerken van 1 delictprofiel	Kunnen we dit delict van een bekende verdachte koppelen aan delictprofielen van andere bekende verdachten in HKS?	een overzicht van mogelijke mededaders van dit delict op grond van hun delictprofielen	opsporingswaardigheid van dader bepalen; onbekende dader(s) opsporen; doorrechercheren op verdachte
7) delictkenmerken van 1 delictprofiel	delictkenmerken van 1 aangifte OD	Kunnen we dit delictprofiel van een bekende verdachte koppelen aan delicten met een onbekende dader in HKS?	een overzicht van delicten die mogelijk van dezelfde dader zijn	opsporingswaardigheid van dader bepalen; doorrechercheren op concrete verdachte
8) delictkenmerken van 1 delictprofiel	delictkenmerken van 1 delict BD	Kunnen we dit delictprofiel van een bekende verdachte koppelen aan delicten van andere bekende verdachten in HKS?	een overzicht van mogelijk mededaderschap van deze verdachte bij andere delicten op grond van zijn individuele delictkenmerken	opsporingswaardigheid van dader bepalen; onbekende dader(s) opsporen; doorrechercheren op verdachte
9) delictkenmerken van 1 delictprofiel	delictkenmerken van 1 delictprofiel	Kunnen we dit delictprofiel van een bekende verdachte koppelen aan delictprofielen van andere bekende verdachten in HKS?	een overzicht van mogelijke samenwerkingsverbanden op grond van delictprofielen	deze zoekoptie past niet binnen de genoemde zoekdoelen
10) persoonskenmerken van 1 signalement OD	persoonskenmerken van 1 signalement OD	Kunnen we dit signalement van een onbekende dader koppelen aan signalementen van andere onbekende daders in HKS?	een overzicht van delicten die mogelijk van dezelfde dader zijn	opsporingswaardigheid van delict bepalen; onbekende dader opsporen

11) persoonskenmerken van 1 signalement OD	persoonskenmerken van 1 signalement BD	Kunnen we dit signalement van een onbekende dader koppelen aan signalementen van bekende verdachten in HKS?	een overzicht van mogelijke verdachten op grond van hun persoonskenmerken	opsporingswaardigheid van delict bepalen; onbekende dader opsporen
12) persoonskenmerken van 1 signalement BD	persoonskenmerken van 1 signalement OD	Kunnen we dit signalement van een bekende verdachte koppelen aan signalementen van onbekende daders in HKS?	een overzicht van delicten die mogelijk van dezelfde dader zijn	opsporingswaardigheid van dader bepalen; doorrecheren op concrete verdachte
13) persoonskenmerken van signalement-combinatie OD	persoonskenmerken van signalement-combinatie OD	Kunnen we de signalementcombinatie van deze groep onbekende daders koppelen aan de signalementcombinatie van andere groepen onbekende daders in HKS?	een overzicht van delicten die mogelijk van dezelfde dadercombinatie zijn	opsporingswaardigheid van delict bepalen; onbekende daders opsporen
14) persoonskenmerken van signalement-combinatie OD	persoonskenmerken van signalement-combinatie BD	Kunnen we de signalementcombinatie van deze groep onbekende daders koppelen aan de signalementcombinatie van groepen bekende verdachten in HKS?	een overzicht van mogelijke verdachtencombinaties op grond van hun persoonskenmerken (in samenwerkingsverbanden)?	opsporingswaardigheid van delict bepalen; onbekende daders opsporen
15) persoonskenmerken van signalement-combinatie BD	persoonskenmerken van signalement-combinatie OD	Kunnen we de signalementcombinatie van deze groep bekende verdachten koppelen aan een signalementcombinatie van groepen onbekende daders in HKS?	een overzicht van delicten die mogelijk van dezelfde verdachtencombinatie zijn	opsporingswaardigheid van daders bepalen; doorrecheren op verdachten
16) delict-en persoonskenmerken van 1 aangifte OD + sign. OD	delict-en persoonskenmerken van 1 aangifte OD + signalement OD	Kunnen we de (combinatie van) delict- en signalementgegevens van deze aangifte met onbekende dader koppelen aan andere aangiften met onbekende dader in HKS?	een overzicht van delicten die mogelijk van dezelfde dader(s) zijn	opsporingswaardigheid van delict bepalen; onbekende dader(s) opsporen
17) delict-en persoonskenmerken van 1 aangifte OD + sign. OD	delict-en persoonskenmerken van 1 delict BD + signalement BD	Kunnen we de (combinatie van) delict- en signalementgegevens van deze aangifte met onbekende dader koppelen aan individuele delicten van bekende verdachten in HKS?	een overzicht van delicten die mogelijk van dezelfde dader(s) zijn én van mogelijke verdachten op grond van hun individuele delict- en signalementkenmerken	opsporingswaardigheid van delict bepalen; onbekende dader(s) opsporen
18) delict-en persoonskenmerken van 1 aangifte OD + sign. OD	delict-en persoonskenmerken van 1 delictprofiel + signalement BD	Kunnen we de (combinatie van) delict- en signalementgegevens van deze aangifte met onbekende dader koppelen aan delictprofielen van bekende verdachten in HKS?	een overzicht van mogelijke verdachten op grond van hun delictprofielen en signalementkenmerken	opsporingswaardigheid van delict bepalen; onbekende dader(s) opsporen
19) delict-en persoonskenmerken van 1 delict BD + signalement BD	delict-en persoonskenmerken van 1 aangifte OD + signalement OD	Kunnen we de (combinatie van) individuele delict- en signalementgegevens van deze bekende verdachte koppelen aan aangiften met onbekende dader in HKS?	een overzicht van delicten die mogelijk van dezelfde dader zijn	opsporingswaardigheid van dader bepalen; doorrecheren op concrete verdachte
20) delict-en persoonskenmerken van 1 delictprofiel + signalement BD	delict-en persoonskenmerken van 1 aangifte OD + signalement OD	Kunnen we dit delictprofiel met het signalement van deze bekende verdachte koppelen aan aangiften met onbekende dader in HKS?	een overzicht van delicten die mogelijk van dezelfde dader zijn	opsporingswaardigheid van dader bepalen; doorrecheren op concrete verdachte

We kunnen de gegevens uit schema 3 vereenvoudigen door de meest belangwekkende zoekvragen te selecteren en door het schema ‘in te klappen’ op één of meer kenmerken.

Selectie van zoekvragen

Aan de zoekvragen 9, 13,14 en 15, en 18 en 20 zullen wij hierna geen aandacht meer besteden.

- Zoekvraag 9 is de enige die niet in het schema lijkt te passen. Het koppelen van delictprofielen aan delictprofielen dient voorzover het zich laat aanzien geen operationeel of tactisch (opsporings)doel. Deze zoekvraag zullen we derhalve verder buiten beschouwing laten.
- De zoekvragen 13,14 en 15 laten we verder buiten beschouwing, omdat we eerder in dit hoofdstuk hebben geconstateerd dat de kans om goede matchresultaten te behalen hierbij erg klein is, vanwege het grote aantal voorwaarden waaraan voldaan moet worden om tot een match te komen.
- De zoekvragen 18 en 20 zullen we om dezelfde reden hier niet verder behandelen: het aantal voorwaarden waaraan voldaan moet worden om tot (goede) matchresultaten te komen, is groter dan bij de andere zoekvragen, zodat deze zoekvragen niet de eerste prioriteit hebben.

Samenvoegen en groeperen van de zoekvragen

De overblijvende zoekvragen (1-8, 10-12, 16,17,19) zullen we gebruiken voor de analyses. We kunnen deze vragen groeperen op grond van de structuur van de zoek- en doelgegevens. Deze structuur bepaalt immers de wijze waarop de gegevens straks gematcht kunnen worden. Aldus verkrijgen we 4 groepen van zoekvragen:

Groep 1: matchen van individuele delicten	→ zoekvragen 1,2, 4,5
Groep 2: matchen van individuele delicten met delictprofielen	→ zoekvragen 3,6,7,8
Groep 3: matchen van individuele signalementen	→ zoekvragen 10,11,12
Groep 4: matchen van delict/signalementcombinaties	→ zoekvragen 16,17,19

In schema 3 zijn de zoekvragen uitgesplitst naar zowel zoek- als doelgegevens. In de praktijk zal de uitsplitsing naar doelgegevens vaak niet relevant zijn. Vergelijk in dit verband bijvoorbeeld de zoekvragen 1 & 2; om de onbekende dader van een delict op te sporen zal men zich bij het matchen niet willen beperken tot aangiften OD of delicten BD, maar de matchoperatie op beide verzamelingen tegelijk willen uitvoeren. Een vergelijkbare redenering gaat op voor de zoekvragen 4 & 5, 7 & 8, 10 & 11, 16 & 17.

Ten slotte kunnen we nog opmerken dat een aantal zoekvragen in schema 3 spiegelbeeldig geformuleerd is. Dit geldt bijvoorbeeld voor de zoekvragen 2 & 4. Vanuit het oogpunt van matching betekent dit dat ze tot identieke resultaten zullen leiden (als de verzameling waarin de matching plaatsvindt in beide gevallen gelijk is). Matching is immers een symmetrisch proces. Vanuit de optiek van het stellen van zoekvragen hebben we uiteraard van doen met zeer verschillende situaties. In het eerste geval (zoekvraag 2) zijn we op zoek naar een verdachte, in het tweede geval (zoekvraag 4) is reeds een verdachte in beeld en proberen we daar andere delicten aan te koppelen.

Als we de hiervoor beschreven selecties toepassen en schema 3 ‘inklappen’ op de doelgegevens verkrijgen we het overzicht van geselecteerde zoekvragen in schema 4.

Schema 4 Overzicht van geselecteerde zoekvragen voor analyses (ingeklapt op doelgegevens, gegroepeerd naar matchprocedure)

<i>Beschikbare zoekgegevens</i>	<i>Mogelijke doelgegevens</i>	<i>Zoekvraag</i>	<i>Beoogd resultaat</i>	<i>Mogelijke zoekdoelen</i>
Groep 1: Matchen van individuele delicten				
Vragen 1&2) delictkenmerken van 1 aangifte OD	Delictkenmerken van 1 aangifte OD/ 1 delict BD	Kunnen we dit delict van een onbekende dader koppelen aan andere delicten in het HKS?	een overzicht van delicten die mogelijk van dezelfde dader(s) zijn	opsporingswaardigheid van delict bepalen; onbekende dader(s) opsporen
Vragen 4&5) delictkenmerken van 1 delict BD	delictkenmerken van 1 aangifte OD/ 1 delict BD	Kunnen we dit delict van een bekende verdachte koppelen aan andere delicten in HKS?	een overzicht van delicten die mogelijk van dezelfde dader zijn	opsporingswaardigheid van dader bepalen; doorrechercheren op concrete verdachte
Groep 2: Matchen van individuele delicten met delictprofielen				
Vraag 3) delictkenmerken van 1 aangifte OD	delictkenmerken van 1 delictprofiel	Kunnen we dit delict van een onbekende dader koppelen aan delictprofielen in HKS?	een overzicht van mogelijke verdachten op grond van hun delictprofielen	opsporingswaardigheid van delict bepalen; onbekende dader(s) opsporen
Vraag 6) delictkenmerken van 1 delict BD	delictkenmerken van 1 delictprofiel	Kunnen we dit delict van een bekende verdachte koppelen aan delictprofielen van andere bekende verdachten in HKS?	een overzicht van mogelijke mededaders van dit delict op grond van hun delictprofielen	opsporingswaardigheid van dader bepalen; onbekende dader(s) opsporen; doorrechercheren op verdachte
Vragen 7&8) delictkenmerken van 1 delictprofiel	delictkenmerken van 1 aangifte OD / 1 delict BD	Kunnen we dit delictprofiel van een bekende verdachte koppelen aan delicten in HKS?	een overzicht van delicten die mogelijk van dezelfde dader zijn	opsporingswaardigheid van dader bepalen; doorrechercheren op concrete verdachte
Groep 3: Matchen van individuele signalementen				
Vragen 10 & 11) persoonskenmerken van 1 signalem. OD	persoonskenmerken van 1 signalement OD /BD	Kunnen we dit signalement van een onbekende dader koppelen aan andere signalementen in HKS?	een overzicht van delicten die mogelijk van dezelfde dader zijn	opsporingswaardigheid van delict bepalen; onbekende dader opsporen
Vraag 12) persoonskenmerken van 1 signalement BD	persoonskenmerken van 1 signalement OD	Kunnen we dit signalement van een bekende verdachte koppelen aan signalementen van onbekende daders in HKS?	een overzicht van delicten die mogelijk van dezelfde dader zijn	opsporingswaardigheid van dader bepalen; doorrechercheren op concrete verdachte
Groep 4: Matchen van delict/signalementcombinaties				
Vragen 16& 17) delict- en persoonskenmerken van 1 aangifte OD+sig. OD	delict- en persoonskenmerken van 1 aangifte OD/BK + signalement OD/BK	Kunnen we de (combinatie van) delict- en signalementgegevens van deze aangifte met onbekende dader koppelen aan andere delicten in HKS?	een overzicht van delicten die mogelijk van dezelfde dader(s) zijn	opsporingswaardigheid van delict bepalen; onbekende dader(s) opsporen
Vraag 19) delict- en persoonskenmerken van 1 delict BK + signalement BK	delict- en persoonskenmerken van 1 aangifte OD + signalement OD	Kunnen we de (combinatie van) individuele delict- en signalementgegevens van deze bekende verdachte koppelen aan aangiften met onbekende dader in HKS?	een overzicht van delicten die mogelijk van dezelfde dader zijn	opsporingswaardigheid van dader bepalen; doorrechercheren op concrete verdachte

5. Overzicht van zoekmethoden

Wanneer we willen onderzoeken op welke wijze de bestaande zoekpraktijk kan worden verbeterd, kunnen we op verschillende niveaus aan de slag. Hieraan geven we in dit onderzoek feitelijk op de volgende manieren vorm:

1. door te onderzoeken wat de meerwaarde kan zijn van het stellen van *aanvullende zoekdoelen*
2. door te onderzoeken wat de meerwaarde kan zijn van het stellen van *aanvullende zoekvragen*
3. door te onderzoeken wat de meerwaarde kan zijn van het *intensiever benutten van de bestaande gegevens* in het HKS
4. door te onderzoeken wat de meerwaarde kan zijn van het gebruik van *alternatieve zoekmethoden*
5. door te onderzoeken wat de meerwaarde kan zijn van het ontwikkelen van *bruikbaarheidscriteria voor het zoekresultaat*.

NB: in deze opsomming moet de term ‘meerwaarde’ worden beschouwd in relatie tot de mogelijkheden die de huidige zoekpraktijk reeds biedt.

De eerste twee kwesties, van de zoekdoelen en zoekvragen, zijn in de voorgaande 2 paragrafen reeds besproken. De vijfde kwestie zal verderop in dit rapport terugkomen. In deze paragraaf concentreren we ons op de kwesties genoemd bij 3 en 4. Het ontwikkelen en uittesten van nieuwe zoekmethoden behelst voor een deel het optimaler gebruiken van de beschikbare gegevens in het HKS. Vandaar dat we deze zaken hier in samenhang bespreken.

Deze paragraaf is verder als volgt opgebouwd. We tonen eerst een schematische weergave van de huidige zoekmethode en bespreken kort de voors en tegens hiervan (par. 5.1). Vervolgens bespreken we de alternatieve zoekmethoden (par. 5.2- par. 5.5). In paragraaf 5.6, tenslotte, bespreken we enkele bijzonderheden die samenhangen met het matchen van geclusterde gegevens.

5.1 Zoekmethode 1: Matchen op geselecteerde kenmerken (huidige praktijk)

We kunnen de zoekmethoden het beste illustreren door de gegevens die gekoppeld moeten worden, voor te stellen als matrices. Een matrix voor een delict ziet er dan als volgt uit (hierna zullen we telkens het voorbeeld van een delict aanhouden):¹³

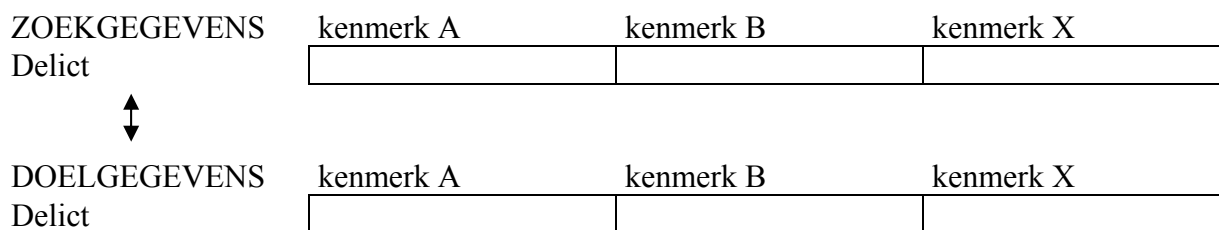
	kenmerk A	kenmerk B	kenmerk X
Delict			

Deze matrix bestaat uit 1 rij en een X-aantal kolommen.

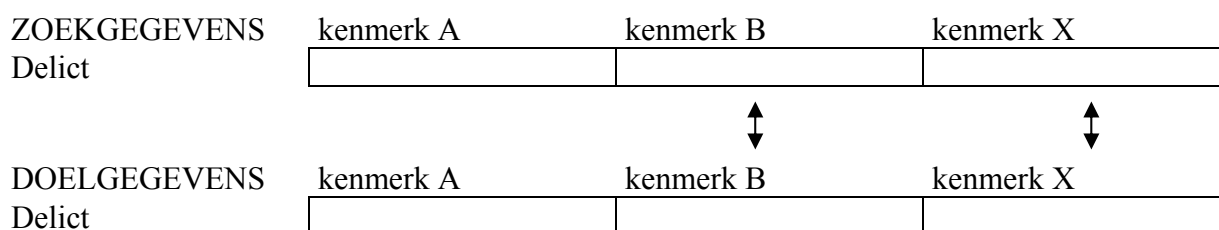
- De rij representeert 1 object (in dit geval 1 delict; het kan ook 1 persoon zijn);
- de kolommen representeren de kenmerken die van dit object bekend zijn (velden in het HKS, bijv. plaats, tijd, MO, etc.);
- en in de vakken worden de waarden van dit object op de gespecificeerde kenmerken weergegeven (bijv. ‘buurt X’, ‘20.00 uur’, ‘d.m.v. verbreking’).

¹³ Het matchen van persoonsgegevens geschiedt op identieke wijze.

Het matchen van 2 delicten kan derhalve worden voorgesteld als het telkens paarsgewijs koppelen van een matrix met zoekgegevens (zoekmatrix) aan matrices met doelgegevens. Schematisch:



Nu zijn er verschillende methoden denkbaar om tot een overzicht van best passende doelmatrices te komen. De methode die nu in de praktijk wordt toegepast kan worden aangeduid als ‘het matchen op geselecteerde kenmerken’. Deze methode ziet er schematisch als volgt uit:



Er zijn gegevens bekend over een delict en een rechercheur besluit op zoek te gaan naar vergelijkbare aangiften door enkele kenmerken te selecteren die bekend zijn van dit delict en vervolgens de aangiften in het HKS te zoeken die op *alle geselecteerde* kenmerken matchen. (Matchen betekent in dit geval dat de gespecificeerde waarden van deze kenmerken overeenkomen). De selectie van de kenmerken kan tot stand komen op grond van gegevens die beschikbaar zijn in het onderzoek, de ervaring van de betrokken functionaris, fingerspitzengefühl, (klets)natte vingers in de lucht, etc.

Het grote voordeel van deze methode is dat hij simpel is (iedereen begrijpt wat er gebeurt) en dat er veel zoekvragen zijn waarvoor deze methode de meest geëigende is. Het gaat hierbij vooral om zoekvragen waarbij de rechercheur (of enige andere vragende partij) van tevoren weet waarnaar hij op zoek is. Een rechercheur die (om welke reden ook) vooraf geïnteresseerd is in delicten met een bepaalde MO, of in een persoon met bepaalde signalementkenmerken of in bepaalde voorwerpen die gestolen zijn bij woninginbraken, etc., heeft aan deze methode meer dan genoeg voor het oplossen van zijn vraag.

Als het er echter om gaat zoekvragen te beantwoorden waarop de rechercheur van tevoren het antwoord niet weet, dan kent deze zoekmethode enkele tekortkomingen. Hierbij gaat het om zoekvragen waarbij personen aan delicten of aan elkaar gekoppeld kunnen worden op grond van ‘gelijksoortigheid’ (met als doel de opsporingswaardigheid van een delict of dader te bepalen, een dader op te sporen of door te rechercheren op een verdachte).

Een tekortkoming die in de praktijk vaak wordt aangetroffen is dat niet alle beschikbare gegevens (uit het opsporingsonderzoek) worden benut, waardoor mogelijkheden om te matchen verloren gaan. Echter, zou men de huidige praktijk hierop aanpassen, zodat alle beschikbare gegevens wel worden gebruikt voor de zoekvraag, dan zou deze zoekmethode waarschijnlijk spaak lopen, omdat er een *pay off* bestaat tussen het aantal kenmerken waarop

een zoekresultaat moet matchen en de hoeveelheid output die wordt verkregen: hoe meer kenmerken worden gespecificeerd des te kleiner het zoekresultaat en omgekeerd. Centrale vraag in deze methode is dan ook welke kenmerken wel en welke niet worden geselecteerd. Achter deze keuze gaan aannames schuil over het gedrag van de betrokken dader(s). De rechercheur veronderstelt immers dat bepaalde gedragingen constant zullen zijn (zoals MO of type delict) en andere niet. Dit is de tweede beperking van deze methode: aannames kunnen juist zijn, maar ook onjuist. Er hoeft bij deze methode maar één aanname onjuist te zijn (één verkeerd gekozen kenmerk) om het zoekresultaat waardeloos te maken.

Als het dus gaat om vragen waarbij het delictgedrag van enige dader in het spel is (en dat is hier het geval) en de rechercheur vooraf het antwoord niet weet, dan moet hij ‘kiezen’ welke gedragskenmerken hij constant veronderstelt. Als hij er teveel kiest, is zijn output nul. Als hij er te weinig kiest, ziet hij zich geconfronteerd met een zeer omvangrijke output. Als hij de verkeerde kenmerken kiest verkrijgt hij geen bruikbaar zoekresultaat, omdat de gedragskenmerken die constant geacht worden in werkelijkheid variabel blijken te zijn en er mogelijk geen gebruik is gemaakt van gegevens omtrent andere kenmerken omdat deze variabel geacht worden, terwijl deze in feite juist constant zijn.

5.2 Matchen met toegevoegde gegevens

In paragraaf 3 stelden we dat de beschikbaarheid van de doelgegevens bepalend is voor de relevantie van de zoekgegevens. Met andere woorden, hoe meer doelgegevens we hebben, des te meer mogelijkheden hebben we ook om op te zoeken.

Een mogelijkheid om dit zonder nadere gegevensverzameling te bereiken is het combineren van bestaande gegevens in het HKS tot nieuwe gegevens. Dit ziet er schematisch als volgt uit:

ZOEKGEGEVENS	kenmerk A	kenmerk B	kenmerk X	kenmerk AB
Delict				
		↕	↕	↕
DOELGEGEVENS	kenmerk A	kenmerk B	kenmerk X	kenmerk AB
Delict				

In combinatie met de huidige zoekmethode zal deze werkwijze echter weinig meerwaarde opleveren, omdat de kwaliteit van het zoekresultaat nog steeds bepaald wordt door de selectie van (delict-) kenmerken. Het enige verschil met de huidige zoekmethode is dat we de beschikbare gegevens in het HKS ‘diepgaander’ benutten om aldus meer mogelijkheden te creëren om delicten (of personen of combinaties van beide) te matchen.

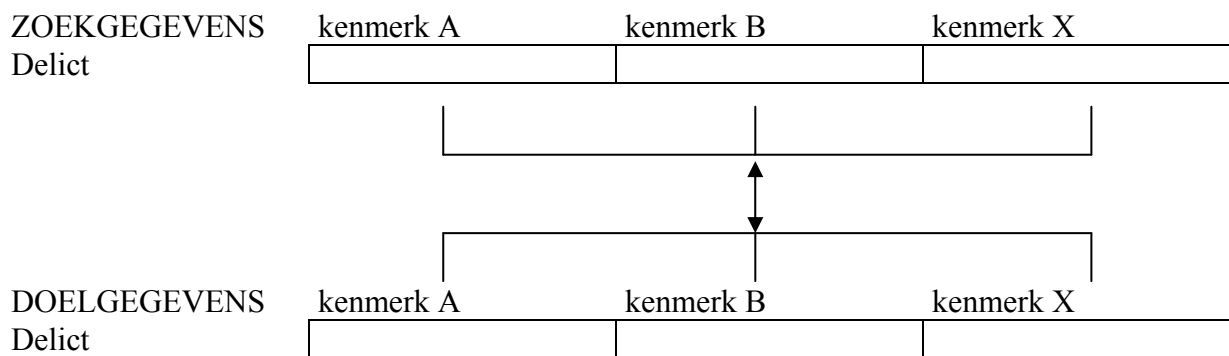
Het creëren van extra gegevensvelden door het aggregeren of combineren van bestaande gegevensvelden zal wel een meerwaarde opleveren wanneer we een andere zoekmethode gebruiken, zoals de zoekmethode die we nu gaan beschrijven in paragraaf 5.3.

5.3 Zoekmethode 2: Matchen op een combinatie van kenmerken

We hebben geconstateerd dat in de huidige praktijk vaak niet alle beschikbare zoekgegevens worden aangewend voor het zoekproces. Dit is in strijd met ons uitgangspunt om zoveel mogelijk de beschikbare gegevens te benutten voor de zoekvraag. We hebben echter ook gezien dat de huidige zoekmethode zich niet goed leent voor het gebruikmaken van alle

beschikbare zoekgegevens. Als we dat namelijk doen, creëren we in een aantal gevallen een dermate groot aantal voorwaarden waaraan het zoekresultaat moet voldoen, dat de kans groot is dat de output nul wordt (zie par. 5.1).

De vraag is nu of er een methode denkbaar is waarbij wel gebruik gemaakt kan worden van alle beschikbare zoekgegevens, echter zonder daarmee de kans op een (goed) zoekresultaat te verkleinen, zoals bij de huidige methode het geval is. Het matchen op een combinatie van kenmerken is zo'n methode. Deze ziet er schematisch als volgt uit:



Deze methode komt erop neer dat een rechercheur alle beschikbare informatie over een delict gebruikt, dat wil zeggen dat hij alle bekende waarden invult in de zoekmatrix en vervolgens op zoek gaat naar doelmatrices die op zoveel mogelijk waarden overeenkomen. Hij zoekt dus niet op een selectie van gegevens, maar op alle gegevens die hij beschikbaar heeft en daarbinnen zoekt hij naar zoveel mogelijk overeenkomsten op kenmerken. Deze methode heeft ten opzichte van de vorige methode(n) de volgende voordelen:

- alle beschikbare zoekgegevens worden benut;
- er worden vooraf geen aannames gemaakt over wat constant is in het gedrag van de dader (waardoor foutieve aannames worden voorkomen);
- de data zelf laten zien welke aangiften op welke kenmerken matchen;
- dit laatste biedt weer mogelijkheden om bruikbaarheidcriteria te ontwikkelen voor het zoekresultaat (bijv. hoeveel en welke kenmerken moeten minimaal matchen om van een bruikbaar resultaat te kunnen spreken?).

Uit dit overzicht blijkt ook de meerwaarde van het creëren van extra gegevensvelden bij toepassing van deze zoekmethode. Immers, hoe meer zoekgegevens beschikbaar zijn, hoe fijnmaziger de doelgegevens gematcht kunnen worden.

Het verschil tussen deze methode en methode 1 blijkt duidelijk uit het volgende voorbeeld: stel dat een rechercheur gegevens beschikbaar heeft over een bepaald delict en hij wil weten of er andere delicten in het HKS te vinden zijn die mogelijk van dezelfde dader zijn. Stel dat van dit delict (waarmee hij gaat zoeken) de waarden op 7 delictkenmerken bekend zijn. Bij de huidige methode (methode 1) zullen alleen die aangiften eruit rollen die op alle 7 kenmerken matchen. Als dit geen output oplevert, zal de rechercheur zijn zoekvraag verbreden door 1 of meer kenmerken weg te laten uit zijn zoekvraag. Net zo lang tot dit een (bevredigende hoeveelheid) output oplevert. Bij methode 3 zal hij altijd alle beschikbare gegevens gebruiken en zo bijvoorbeeld kunnen zien dat er geen delicten zijn die op alle 7 kenmerken matchen, maar dat er wel delicten zijn die op 5 of 6 kenmerken matchen. Doordat hij geen aannames maakt -zoals bij methode 1- met betrekking tot de kenmerken die hij handhaaft en de kenmerken die hij weglaat uit zijn zoekvraag, kan hij zodoende op aangiften stuiten die hij anders niet op het spoor gekomen zou zijn.

5.4 Zoekmethode 3: Gecombineerd matchen

Rekening houden met de Lewis-factor

Deze factor, vernoemd naar de bekende rechterhand van Morse, die nauwelijks werd gehoord maar heel vaak gelijk had, verwijst naar de menselijke factor bij het zoekproces. De aannames die rechercheurs maken in het huidige zoekproces zijn soms goed onderbouwd of ‘aangevoeld’, maar soms ook niet. Hoe het ook zij, bij het verder ontwikkelen van het zoekproces moet hiermee rekening worden gehouden door enerzijds te voorkomen dat onjuiste aannames leiden tot foutieve zoekresultaten en anderzijds te bevorderen dat juiste en intelligente aannames een kans krijgen getoetst te worden. Methode 2 voorkomt weliswaar het gebruik van foutieve aannames, maar deze biedt geen mogelijkheid om intelligente aannames te toetsen (wel mogelijk bij methode 1). Bijvoorbeeld door nadere voorwaarden aan het zoekresultaat te stellen (bijv. het delict moet per se matchen op kenmerk x of y) of door het toekennen van gewichten aan het zoekresultaat (bijv. indien het te matchen delict kenmerk x of kenmerk y heeft, dan moet het een extra punt krijgen in het zoekresultaat).

In de praktijk zullen rechercheurs (en andere vragende partijen), op grond van welke overweging ook, de mogelijkheid willen (en ook moeten kunnen) hebben om hun kennis, ervaring of gevoel een plaats te geven in het zoekproces. Dit kan zoekmethoden opleveren die een combinatie vormen van de hier genoemde methoden 1 en 2.

In dit onderzoek zullen we hieraan vormgegeven door enkele varianten uit te werken, waarbij de resultaten van zoekmethode 2 worden ‘gewogen’ naar het al of niet voldoen aan bepaalde nadere voorwaarden.

5.5 Zoekmethoden 4a-4c: Matchen van geclusterde gegevens

Tot op heden hebben we telkens gesproken over het matchen van enkelvoudige gegevens: het koppelen van individuele delicten of personen. We gaan echter ook geclusterde gegevens koppelen, namelijk delictprofielen van bekende verdachten. Schematisch ziet dit er als volgt uit:

ZOEKGEGEVENS	kenmerk A	kenmerk B	kenmerk X
Delict			

↕

DOELGEGEVENS	kenmerk A	kenmerk B	kenmerk X
delict 1			
delict 2			
delict N			

Hier gaat het om het koppelen van 1 delict aan een cluster van delicten. De matrix van een cluster heeft minimaal 2 en maximaal N rijen. Bij de uitgangspunten hebben we opgemerkt dat de beschikbaarheid van meer gegevens de kwaliteit van het zoekresultaat misschien zal bevorderen, met andere woorden: een betere match zal opleveren. Dit werkt als volgt. Het matchen is gebaseerd op het vinden van constantes in het gedrag van de dader. Wanneer we (doel)gegevens hebben over 1 delict kunnen we nog heel weinig zeggen over dergelijke constantes. Daarvoor hebben we bij voorkeur meer delicten nodig. Bij de 1-op-1 match (methoden 1 t/m 3) zoek je derhalve naar zoveel mogelijk overeenkomsten tussen het zoeken het doeldelict in de hoop dat je delicten aan elkaar kunt koppelen op grond van een

opvallend aantal gemeenschappelijke kenmerken. Wanneer we beschikken over een cluster van delicten, kunnen we veel fijnmaziger vaststellen in hoeverre een zoekdelict past bij een bepaalde dader. Hoe groter de matrix van delicten in het profiel des te beter kan de 'handtekening' van de dader worden achterhaald.

Ook bij geclusterde gegevens zijn verschillende methoden denkbaar om de constante in het gedrag van de dader te vinden. We zullen deze methoden hierna bespreken.

Zoekmethode 4a: Matchen van geselecteerde kenmerken

Deze methode is verwant aan methode 1 bij het zoeken in enkelvoudige gegevens. Per delict in een delictprofiel wordt bepaald of dit volgens methode 1 matcht met het zoekdelict. Deze procedure wordt voor alle delicten in het profiel herhaald, waarna een samenvattende maat wordt gecreëerd. We zullen drie samenvattende maten berekenen:

- 1) het gemiddelde matchresultaat van de delicten in het delictprofiel die matchen met het zoekdelict;
- 2) het percentage delicten in het delictprofiel dat matcht met het zoekdelict;
- 3) een combinatie van de voorgaande 2 maten (percentage delicten dat matcht * gemiddeld matchresultaat).

Zoekmethode 4b: Matchen op een combinatie van kenmerken

Deze methode is verwant aan methode 2 bij het zoeken in enkelvoudige gegevens. Van elk delict in een delictprofiel wordt bepaald op hoeveel kenmerken het matcht met het zoekdelict. Deze procedure wordt voor alle delicten in het profiel herhaald, waarna een gemiddeld matchresultaat voor het profiel wordt berekend (het gemiddelde percentage match van alle delicten in het profiel).

Zoekmethode 4c: Gecombineerd matchen

Deze methode is verwant aan methode 3 bij het zoeken in enkelvoudige gegevens. De delictprofielen worden eerst geordend naar de matchresultaten die bij methode 4b zijn behaald (gemiddeld percentage match van alle delicten in het profiel). Daarna wordt een extra voorwaarde gesteld waaraan de delicten in het profiel moeten voldoen. Voor elk afzonderlijk delict in het profiel wordt getoetst of dit aan de nadere voorwaarde voldoet. Het aantal delicten in het profiel dat voldoet aan de gestelde voorwaarde wordt vervolgens uitgedrukt als een percentage van het totaal aantal delicten in het profiel. Dit percentage wordt opgeteld bij het matchresultaat dat werd behaald op grond van methode 4b en het resultaat gemiddeld. Aldus wordt het matchresultaat bij methode 4b 'gewogen' naar de aanwezigheid van delicten in het delictprofiel die voldoen aan de gestelde voorwaarde.

5.6 Samenstelling van delictprofielen bij het matchen van geclusterde gegevens

Bij het matchen van delictprofielen doen zich twee bijzonderheden voor die niet aan de orde zijn bij het matchen van enkelvoudige gegevens zoals individuele delicten. De volgende twee vragen zijn hierbij aan de orde:

- Hoe kunnen we rekening houden met de factor tijd in een delictprofiel?
- Hoe kunnen we rekening houden met verschillen in omvang van de delictprofielen?

De eerste kwestie is hiervoor al aan de orde geweest. Het gaat om de vraag of het delictgedrag van de dader in de tijd min of meer constant is of zich ontwikkelt in een bepaalde richting. Laten we even ervan uitgaan dat we de dader van een bepaald zoekdelict kennen en zijn delictprofiel hiermee willen matchen. Als het delictgedrag in de tijd constant is zullen alle

delicten in het profiel ceteris paribus een even grote kans hebben om te matchen met het zoekdelict. Als het delictgedrag echter in de tijd verandert zullen delicten in het delictprofiel die in de tijd verder verwijderd liggen van het zoekdelict een kleinere kans hebben om te matchen dan delicten die in de tijd dichterbij het zoekdelict liggen. Met deze mogelijkheid moeten we rekening houden omdat deze het matchresultaat negatief kan beïnvloeden.

De tweede kwestie, de variërende omvang van de delictprofielen, hangt hier deels mee samen. Als het delictgedrag in de tijd sterk verandert, zullen grotere delictprofielen minder ‘samenhang’ vertonen dan kleine delictprofielen. Hiervoor spraken we al over het fenomeen dat specialisatie van delictgedrag, zeker bij veelplegers, zelden voorkomt. Dit zou betekenen dat veelplegers (personen met delictprofielen waarin veel delicten zitten) moeilijker te matchen zijn dan personen met ‘kleine’ profielen of met hooguit 1 delict in hun ‘profiel’. Ook deze mogelijkheid kan het zoekresultaat negatief beïnvloeden.

Om met deze twee factoren rekening te houden, zullen we bij het matchen van geclusterde gegevens (de delictprofielen) enkele aanvullende analyses uitvoeren, waarbij de factor tijd en de variërende omvang van de delictprofielen worden verdisconteerd. In schema 5 wordt hiervan een overzicht gegeven. Op de rijen vinden we de 3 zoekmethoden bij geclusterde gegevens die hiervoor beschreven zijn. De kolommen geven aan hoe het delictprofiel telkens is samengesteld.

Schema 5 Overzicht van aanvullende analyses bij geclusterde gegevens

<i>Zoekmethoden</i>	<i>Samenstelling van delictprofiel</i>			
	alle delicten	Alle delicten gewogen naar tijd	Selectie van delicten binnen 1 jaar van zoekdelict	Selectie van beste 3 delicten
4a) geselecteerde kenmerken	x	nvt	nvt	nvt
4b) combinatie van kenmerken	x	x	X	x
4c) gecombineerd matchen	x	x	X	x

In de tweede kolom van dit schema (‘alle delicten’) worden de analyses weergegeven die we in de vorige paragraaf hebben besproken: het gaat om 3 zoekmethoden waarbij telkens alle delicten in het delictprofiel meetellen voor het zoekresultaat.

In kolom 3 worden deze analyses herhaald, maar nu worden de delicten in het delictprofiel gewogen naar de tijdsafstand die bestaat tussen deze delicten en het zoekdelict. De idee hierachter is uiteraard dat we rekening willen houden met de factor tijd en met veranderingen in delictgedrag. We gaan ervan uit dat delicten die in de tijd dichterbij het referentiedelict liggen relevanter zijn dan delicten die in tijd verder weg liggen. Dit kunnen we tot uitdrukking brengen door de delicten in het profiel te wegen. Hierbij worden de volgende gewichten toegekend:

Tijdsafstand tussen delict uit profiel en zoekdelict:	weegfactor:
0 < 7 dagen	1
8 < 14 dagen	.9
15 < 31 dagen	.8
1 maand < 3 maanden	.7
3 maanden < 6 maanden	.6
6 maanden < 1 jaar	.5
1 jaar of langer geleden	.4

Hoe verder een delict in de tijd verwijderd is van het zoekdelict, hoe groter de correctie die plaatsvindt op het gevonden matchresultaat. Aldus zullen delicten die in de tijd dichterbij het zoekdelict liggen ceteris paribus een beter matchresultaat kunnen behalen dan delicten die daar ver vandaan liggen.

Een alternatieve methode om rekening te houden met de factor ‘tijd’ is gegeven in kolom 4 van schema 5. Wederom worden de analyses met de 3 zoekmethoden herhaald, maar nu tellen alleen die delicten uit het delictprofiel mee die in de tijd maximaal 1 jaar (= 365 dagen) van het zoekdelict liggen. Deze werkwijze is eenvoudiger dan de vorige en heeft als mogelijk voordeel dat delicten die in de tijd ver weg liggen van het zoekdelict niet meetellen. Een mogelijk nadeel is dat niet alle gegevens uit het profiel worden benut.

De laatste kolom in schema 4 is een methode om verschillen in omvang van de delictprofielen te verdisconteren. Dit doen we door alleen de 3 best matchende delicten uit het profiel (als het profiel groter of gelijk is aan 3 delicten) mee te tellen voor het matchresultaat. Als het profiel bestaat uit minder delicten tellen alle delicten mee (maximaal 1 of 2).

De kruisjes in de gearceerde velden van schema 5 geven de aanvullende analyses aan. Deze zijn alleen van toepassing op de zoekmethoden 4b en 4c. Het is niet mogelijk bovengenoemde selecties toe te passen bij zoekmethode 4a (velden ‘nvt’), omdat we bij deze zoekmethode telkens te maken hebben met individuele delicten die wel of niet voldoen aan gestelde voorwaarden. Deze werkwijze kan niet gecombineerd worden met de nadere selecties die we hiervoor beschreven hebben.

6. Samenvatting

In het vorige hoofdstuk hebben we een aantal (mogelijke) tekortkomingen gesignaleerd in de huidige zoekmodaliteit van het HKS. Deze hebben we als uitgangspunt genomen voor het ontwikkelen van een nieuw zoekinstrument, waarvan we de contouren in dit hoofdstuk hebben geschetst.

De belangrijkste kenmerken van dit nieuwe zoekinstrument zijn dat het in staat moet zijn om naast zoekvragen voor operationeel gebruik ook zoekvragen voor tactisch gebruik te beantwoorden. Door de beschikbare zoek- en doelgegevens op een systematische manier met elkaar te vergelijken is een volledig overzicht verkregen van alle mogelijke zoekvragen die door het nieuwe zoekinstrument kunnen worden gesteld. Door gebruik te maken van het feit dat gegevens in het HKS geclusterd zijn binnen eenheden (zoals bij delictprofielen: delicten binnen een dader) is het ook mogelijk zoekvragen te formuleren die in de huidige zoekmodaliteit van het HKS niet voorkomen. Het nieuwe zoekinstrument bevat aanvullende zoekmethoden die mogelijk tot betere matchresultaten leiden dan de huidige zoekmodaliteit. Deze zoekmethoden kunnen ook gebruik maken van extra informatie die ontstaat uit het aggregeren en combineren van gegevens in het HKS. Bij de nieuwe zoekmethoden is het ook mogelijk de kwaliteit van het zoekresultaat in een kwantitatieve maat uit te drukken.

In het volgende hoofdstuk beschrijven we welke gegevens we hebben gebruikt voor de analyses die moeten uitwijzen of het hier voorgestelde zoekinstrument een meerwaarde heeft boven de bestaande zoekpraktijk in het HKS. We beschrijven ook hoe de toetsing is opgezet om een zo kritisch mogelijke vergelijking tussen de verschillende zoekmethoden mogelijk te maken.

3. Opzet van de toetsing

1. Inleiding

In het vorige hoofdstuk hebben we de contouren geschetst voor een nieuw zoekinstrument. In dit hoofdstuk beschrijven we de wijze waarop we zullen toetsen of dit zoekinstrument een toegevoegde waarde kan hebben voor de huidige praktijk. In eerste instantie geven we aan welke gegevens we hiervoor gebruiken en op welke wijze we bestaande informatie uit het HKS veredelen door gegevens uit verschillende velden te aggregeren en/of te combineren (par. 2 & 3). In paragraaf 4 beschrijven we de uitgangspunten en de opzet van de toetsing. In paragraaf 5 geven we aan hoe we de huidige zoekroutine nabootsen. In paragraaf 6 beschrijven we op welke wijze zoekmethode 3 wordt uitgewerkt. In paragraaf 7 wordt een korte opmerking gemaakt over de wijze waarop bij de verschillende zoekmethoden de matchresultaten worden berekend en in de laatste paragraaf (par. 8) worden de bijzonderheden van de toetsing per groep van zoekvragen besproken.

2. Selectie van gegevens

Voor de toetsing is gebruik gemaakt van een extractie van gegevens uit het regionale HKS van de politieregio Haaglanden. Het gaat om 2 extracties:

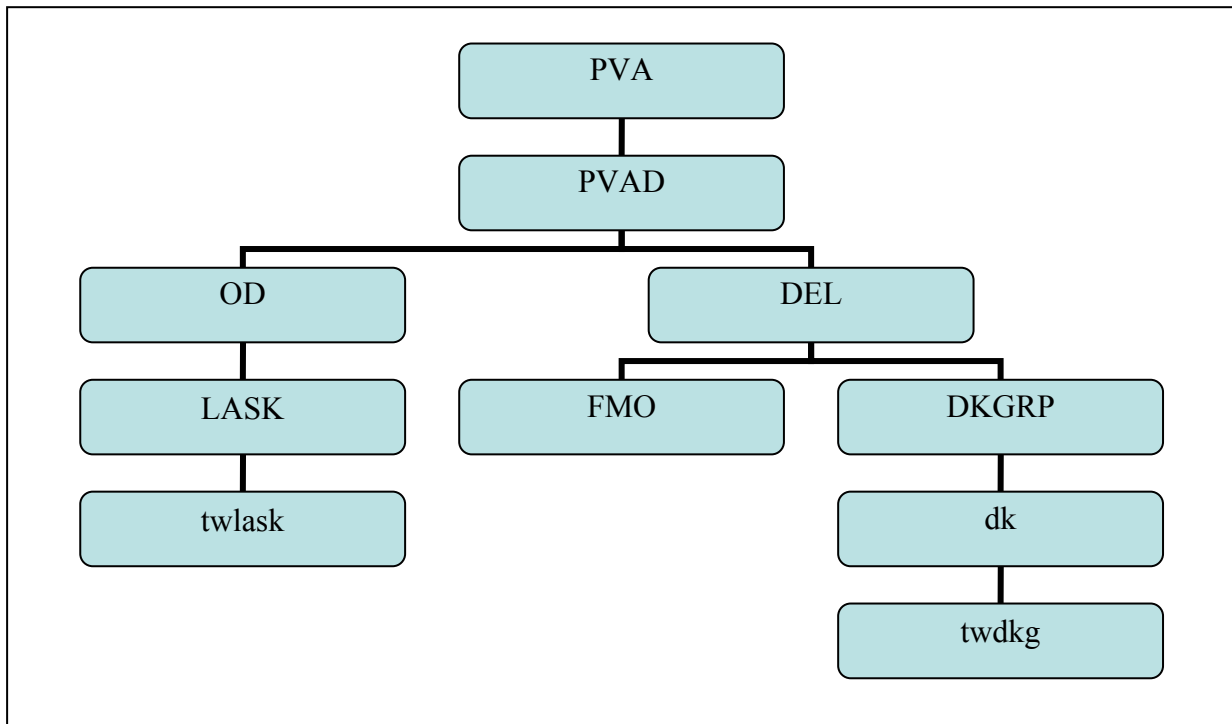
- een extractie van alle aangiften met een aangiftdatum vanaf 1-1-1999 tot en met 31-12-2001 en
- een extractie van alle personen die in dezelfde periode voorkomen met 1 of meer antecedenten. Deze selectie is gebaseerd op antecedentdatum en bevat alle personen met antecedenten in deze periode. Dus ook personen die alleen handmatige, voorlopige of geseponeerde antecedenten hebben. Van deze personen beschikken we over hun volledige delictgeschiedenis voorzover deze in het HKS is opgeslagen. Het gaat hierbij om antecedenten van vóór 1-1-1999.

Het HKS is een database die bestaat uit circa 40 afzonderlijke gegevenstabellen die via allerhande sleutels aan elkaar gekoppeld zijn. Deze tabellen kunnen zowel horizontaal als verticaal (dus hiërarchisch) met elkaar in verbinding staan. Niet al deze tabellen zijn voor het onderhavige onderzoek even interessant. Derhalve is een selectie gemaakt van tabellen die de voor dit onderzoek noodzakelijk informatie bevatten. In schema 6 op de volgende pagina is weergegeven welke tabellen aan de aangiftenkant zijn gebruikt. Hetzelfde hebben we in schema 7 gedaan voor de tabellen aan de personenkant.

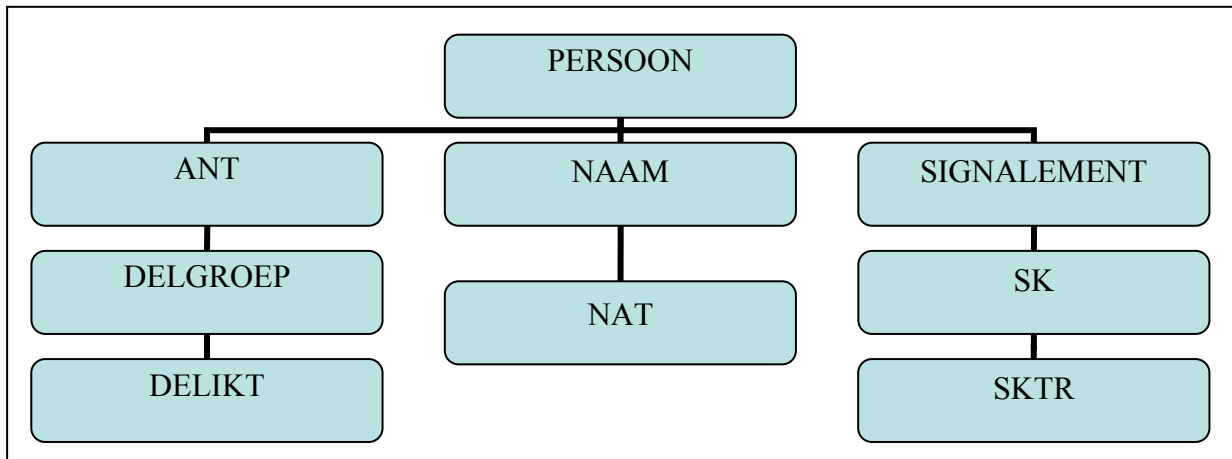
In schema 6 gaat het aan de linkerkant om 3 tabellen waarin informatie ligt opgeslagen over signalementgegevens van onbekende daders (die gekoppeld zijn aan aangiften OD) en aan de rechterkant om 5 tabellen met delictinformatie uit de aangiften. Signalementgegevens van onbekende daders betreffen onder meer lengte en leeftijd (geboortjaar) (tabel OD), lichaamsafhankelijke signalementkenmerken zoals gecodeerd op het LMF-formulier (tabel LASK) en trefwoorden die eventueel aan deze kenmerken zijn toegevoegd (tabel TWLASK). Aan de aangiftenkant gaat het om wetsartikelen en feit-modus-operandicode's (tabel FMO), om allerhande MO-kenmerken zoals gecodeerd op het LMF-formulier (tabel DK) en trefwoorden die eventueel hieraan zijn toegekend (tabel TWDKG). De bestanden DEL en DKGRP zijn tellerbestanden die nodig zijn om de verschillende bestanden aan elkaar te kunnen koppelen. De bovenste 2 tabellen in schema 6 (PVA en PVAD) bevatten algemene

informatie over het proces-verbaal (PVA) en de tijd- en adresgegevens van afzonderlijke gebeurtenissen in het proces-verbaal (PVAD).

Schema 6 Selectie van HKS-tabellen aan de aangiftenkant



Schema 7 Selectie van HKS-tabellen aan de personenkant



De gegevens over delicten zijn in dit onderzoek geanalyseerd op het niveau van tabel DEL. Dit zijn de afzonderlijke delicten welke geclusterd kunnen zijn binnen afzonderlijke gebeurtenissen (PVAD) binnen een proces-verbaal (PVA). De informatie uit de lager gelegen tabellen (zoals FMO, DKGRP, DK en TWDKG) is geaggregeerd naar dit niveau.¹⁴

¹⁴ Het 'platslaan' van gegevens zodat alle informatie over 1 delict op 1 record van een database past, heeft als gevolg dat de bestanden snel erg omvangrijk worden, in voorkomende gevallen té omvangrijk om er nog analyses op te kunnen verrichten. Om dit probleem te ondervangen is ervoor gekozen uit de lagere tabellen niet alle informatie te aggregeren, maar een maximaal informatieverlies van maximaal 0,5% toe te staan. Dit relatief

De tabellen in schema 6 gebruiken we om analysebestanden te maken van enerzijds *delicten* en anderzijds *signalementen van onbekende daders*. In bijlage 1 is een overzicht opgenomen van de gegevensvelden die we uit deze tabellen hebben gebruikt.

In schema 7 zijn de geselecteerde tabellen van de personenextractie weergegeven. De 3 tabellen links bevatten gegevens omtrent delicten, de twee tabellen in het midden bevatten persoonsgegevens en de 3 tabellen aan de rechterkant bevatten signalementgegevens van bekende verdachten. De delictgegevens in de 3 tabellen links zijn niet volledig. Bij een extractie van personen komen ook delicttabellen mee zoals weergegeven in schema 6. Het gaat hierbij om aangiften die gekoppeld zijn aan de geëxtraheerde verdachten. Deze tabellen zijn inhoudelijk identiek aan de tabellen die we hiervoor besproken hebben. Om die reden zijn ze hier niet nogmaals weergegeven.¹⁵ De delicttabellen aan de personenkant die wel in schema 7 zijn weergegeven (ANT, DELGROEP en DELIKT) bevatten informatie over alle antecedenten. De tabellen aan de aangiftenkant (die hier niet zijn weergegeven) bevatten alleen informatie over automatische (gekoppelde) antecedenten.

Voor het samenstellen van de delictprofielen is alleen gebruik gemaakt van de gekoppelde aangiften. De reden hiervoor is dat de overige antecedenten (de zogenaamde voorlopige, handmatige en geseponeerde antecedenten) te weinig informatie bevatten om zinvol mee te kunnen zoeken. Van deze antecedenten zijn alleen wetsartikelen en fmo-codes bekend (tabel DELIKT). Het bestand ANT bevat gegevens over antecedenten, zoals antecedentdatum, soort antecedent en een zogenaamde pvo-sleutel (proces-verbaal van oplossing). Eén antecedent kan meer delicten bevatten. Het bestand DELGROEP is een tellerbestand waarin alle delicten van een bekende verdachte zijn opgeteld. Dit bestand wordt ook gebruikt om een koppeling te leggen met de delictgegevens aan de aangiftenkant (via de tabel BD, hier niet weergegeven). De bestanden NAAM en NAT bevatte enkele persoonsgegevens over de verdachten, zoals hun geslacht, hun geboortedatum, het geboorteland en hun nationaliteit. In de bestanden rechts vinden we weer signalementgegevens, ditmaal van bekende verdachten. Ofschoon de inhoud van deze tabellen goeddeels gelijk is aan wat hiervoor beschreven is bij de signalementen van onbekende daders, kan worden opgemerkt dat deze gegevens kwalitatief en kwantitatief aanzienlijk beter zijn. Een verdachte kan meer signalementen hebben, hoewel dit niet heel erg vaak voorkomt (het beleid van de HKD-en is om dit zoveel mogelijk te voorkomen). In voorkomende gevallen hebben wij gekozen voor het eerste signalement dat bij een verdachte genoteerd staat. Dit is het meest recente respectievelijk meest complete signalement (opgemaakt door de TOD).

De tabellen in schema 7 gebruiken we om analysebestanden te maken van *delictprofielen* en *signalementen van bekende verdachten*. In bijlage 2 is een overzicht opgenomen van de gegevensvelden die uit de hier besproken tabellen zijn geselecteerd.

geringe informatieverlies biedt voldoende besparing om de omvang van de analysebestanden hanteerbaar te maken. Een voorbeeld kan deze werkwijze verduidelijken. Neem de delictkenmerkcodes zoals deze op het LMF-formulier voorkomen. Bij veruit de meeste aangiften zijn hooguit enkele van deze codes ingevuld. Zo zijn bij ruim driekwart van de aangiften hooguit 5 van deze codes ingevuld. Er zijn echter ook aangiften waar 50, 80 of zelfs meer dan 100 van deze codes zijn ingevuld. Dit zijn echter uitzonderingen. In ons onderzoek hebben we alleen de eerste 25 codes meegenomen. Daarmee is onze informatie in 99,6% van de gevallen compleet. Bij 0,4% van de aangiften missen we 1 of meer codes en mogelijk ook trefwoorden die hieraan gekoppeld zijn. Deze ingreep heeft geen enkel effect op de onderzoeksresultaten waar het gaat om het vergelijken van verschillende zoekmethoden. Deze worden immers allemaal toegepast op dezelfde gegevens. Het effect op de uitkomsten per zoekresultaat is verwaarloosbaar. Het informatieverlies is in kwantitatieve zin zeer beperkt. Bovendien is er bij het matchen geen meerwaarde te verwachten van aangiften waarover uitzonderlijk veel gegevens bekend zijn; van veruit de meeste aangiften zijn veel minder gegevens bekend zijn en deze 'zwakkere schakels' bepalen het zoekresultaat.

¹⁵ Alleen de sleutels waarop de tabellen worden gekoppeld zijn hier anders, zie bijlage 2.

3. Veredelen van informatie uit geselecteerde gegevensvelden

Onderzocht is of er mogelijkheden zijn om gegevens uit afzonderlijke gegevensvelden van het HKS te aggregeren of combineren tot nieuwe gegevens.

Bij de signalementgegevens is het niet gelukt om zinvolle nieuwe gegevensvelden te creëren. Deze gegevens laten zich door hun aard moeilijk veredelen.

Bij de delictgegevens hebben we enkele nieuwe gegevensvelden kunnen creëren door het samenvoegen of aggregeren van bestaande gegevens. In schema 8 zijn deze weergegeven. Zo hebben we de locatie van het delict aangevuld met de velden ‘regio’ en ‘wijk’. Ten aanzien van het tijdstip van het delict hebben we extra velden gecreëerd, zoals het seizoen waarin het delict gepleegd is, de dag van de week, en delen van de dag (dagdelen van 6 en 3 uur). Ten aanzien van de aard van het delict hebben we op basis van de wetsartikelen 2 grovere delictindelingen gemaakt. De eerste (cbs-cat) komt overeen met de cbs-indeling, de tweede (hfd-cat) is daar weer een samenvatting van, met categorieën als geweld, zeden, vermogen met geweld, vermogen, openbare orde & gezag, verkeer, milieu en opiumwetdelicten. Tenslotte is geprobeerd op basis van de delictkenmerkcodes op het LMF-formulier enkele samenhangende modus-operandicategorieën te construeren. Negen van dergelijke velden zijn gecreëerd: het gebruik van wapens, het gebruik van geweld, in contact treden met het slachtoffer, het gebruik van bijzondere hulpstukken (indicatie van voorbereiding), het gebruik van misleiding, het betreden van (afgesloten) privé-terrein, en 3 buitcategorieën. De gedachte achter de constructie van al deze velden is dat hierin mogelijk een constante kan liggen in het gedrag van de dader, die we op deze manier eerder op het spoor zullen komen dan bij het gebruik van de afzonderlijke velden in het HKS.

Schema 8 Creatie van nieuwe gegevensvelden uit bestaande gegevensvelden in het HKS

<i>Bestaande gegevensvelden:</i>	<i>Nieuw gegevensveld:</i>	<i>Verkregen door:</i>
sleutel proces-verbaal	regio waar delict is gepleegd ?	bestaand veld
Wijk-buurt-subbuurt	wijk waar delict is gepleegd	aggregatie
Delictdatum	seizoen waarin delict is gepleegd*	aggregatie
Delictdatum	dag van de week ?	bestaand veld
Delicttijd	dagdelen van 6 uur*	aggregatie
Delicttijd	dagdelen van 3 uur*	aggregatie
Wetsartikelcode	grovere indeling delict (cbs-cat)	aggregatie
Wetsartikelcode	grovere indeling delict (hfd-cat)	aggregatie
<i>Delictkenmerken LMF[#]</i>	<i>modus operandi cluster</i>	<i>combinatie</i>
a11-16,d8,9,11,k46	mo: gebruik van wapens*	combinatie
a18,24,c6,14,15,24,33,34,d19	mo: gebruik van geweld*	combinatie
a7,8,c1,2,9,10,c18,d18,f1,2,4,5		
f7-9,f11-14	mo: contact met slachtoffer*	combinatie
c7,8,17,26-28,d2,5-7,10	mo: gebruik van hulpstukken*	combinatie
c1,29,d15,16	mo: gebruik van misleiding*	combinatie
a20,21,e2,3,17-19,24,32,40,42	mo: betreden van privé-terrein*	combinatie
g5,11,29,k6,17,25,36,45	mo: buit is geld*	combinatie
g9,21,22,25,k12,16,18,21	mo: buit is elektr. Apparaten*	combinatie
g12,24,k20,23,28	mo: buit is kunst*	combinatie

[#] voor de informatie uit de A-velden is ook gebruik gemaakt van de fmo-codes

Overigens zijn niet alle in schema 8 genoemde velden een daadwerkelijke toevoeging aan de mogelijkheden die de HKD-er nu beschikbaar heeft. Alleen de velden met een * zijn op geen enkele wijze bevroegbaar in de huidige zoekprocedure. De overige door ons gecreëerde velden kunnen ook door de HKD-er worden 'benut' door bij het zoeken in de bestaande velden gebruik te maken van wildcards en dergelijke. Er staan 2 nieuwe velden in het schema (regio, dag van de week) waar een vraagteken achter staat. Het is ons niet duidelijk of deze velden bevroegbaar zijn in het huidige HKS. Ten aanzien van de dagdelen moeten we, ten slotte, nog opmerken dat ook in de huidige zoekmodaliteit van het HKS een mogelijkheid bestaat om op dagdelen te zoeken, zij het iets minder verfijnd dan in het schema aangegeven.

4. Uitgangspunten en opzet van de toetsing (algemeen)

Het doel van de toetsing is om na te gaan of de zaken die in het vorige hoofdstuk zijn voorgesteld iets toevoegen aan de bestaande praktijk. Concreet gaat het hierbij om een antwoord op de volgende vragen:

- Heeft het meerwaarde om naast zoekvragen voor operationeel gebruik ook zoekvragen voor tactisch gebruik aan te wenden?
- Heeft het meerwaarde om gebruik te maken van extra gecreëerde gegevensvelden?
- Heeft het meerwaarde om gebruik te maken van andere zoekmethoden?
- Heeft het meerwaarde om delicten te matchen met delictprofielen (in plaats van delicten)?
- Heeft het meerwaarde om delict- en signalementgegevens tegelijkertijd te matchen (in plaats van afzonderlijk)?
- Is het mogelijk om in kaart te brengen of er delicten en personen zijn die zich beter laten matchen dan andere?
- Is het mogelijk om bruikbaarheidscriteria te ontwerpen voor (nieuwe) zoekmethoden?

Om deze vragen te kunnen beantwoorden moet de toetsing enerzijds goed interpreteerbare zoekresultaten opleveren en anderzijds een mogelijkheid bieden om de resultaten uit verschillende tests met elkaar te vergelijken. Dit zijn de belangrijkste uitgangspunten voor de toetsing die we hierna zullen bespreken.

We moeten dus eerst een analyse ontwerpen die het resultaat van de verschillende zoekvragen en zoekmethoden onomstotelijk kan vaststellen. Dit is mogelijk door gebruik te maken van simulatie. In het HKS bevinden zich aangiften die zijn opgehelderd. Van deze delicten kennen we dus één of meer daders. Door deze aangiften te behandelen alsof het aangiften OD (onbekende dader) zijn en vervolgens op zoek te gaan naar de bekende dader, kunnen we een indicatie verkrijgen van de kwaliteit van het verkregen zoekresultaat. We kennen immers de dader van het delict en we kunnen nagaan waar deze in het zoekresultaat voorkomt. Dit geeft een indicatie van de kwaliteit van het zoekresultaat.

Dit beginsel kunnen we ook toepassen bij het matchen van persoonsgegevens (signalementen), omdat zich de bijzondere omstandigheid voordoet dat de eventueel aanwezige signalementen OD niet altijd uit het HKS worden verwijderd wanneer een aangifte gekoppeld wordt aan een verdachte. Als er van deze verdachte ook een signalement beschikbaar is aan de personenkant van het HKS (signalement BD), zijn we aldus in staat om met het 'overgebleven' signalement OD een matchprocedure uit te voeren. We kunnen nagaan hoe goed de verschillende zoekmethoden in staat zijn om dit signalement OD te koppelen aan het 'bijbehorende' signalement BD. Net als hiervoor bij de delicten geeft de positie van de werkelijke dader in het zoekresultaat een indicatie van de kwaliteit van de zoekmethode. Dit

is echter alleen maar het geval bij operationele zoekvragen. Daarom beperken we de toetsing hiertoe.¹⁶

We gaan dus op zoek naar de bekende dader (hetzij via een delict, hetzij via een signalement) door gebruik te maken van gegevens over delicten en signalementen waarvan we de dader al kennen. Die kennis gebruiken we vervolgens om de kwaliteit van het zoekresultaat te toetsen. In zijn algemeenheid werkt deze procedure als volgt: stel we beschikken over een delict waarvan we de dader kennen, maar we doen alsof dit niet het geval is. We stellen onszelf een operationele zoekvraag: we willen weten of de dader van dit delict in het HKS voorkomt en zo ja, wie het is. Dit doen we door het delict op kenmerken (via de verschillende zoekmethoden) te koppelen aan andere delicten waarvan we al een dader kennen. Dit levert een bepaald zoekresultaat op, waarbij de delicten die het best matchen bovenaan eindigen en de delicten die het slechtst matchen onderaan. Als we deze zoekvraag in de praktijk zouden stellen, zouden we geen indicatie hebben van de kwaliteit van het zoekresultaat. We zouden weliswaar weten welke delicten beter matchen en welke slechter, maar daarmee is nog niet gezegd dat de dader van het zoekdelict zich bevindt bij de best matchende delicten. Door gebruik te maken van simulatie kunnen we dit echter toetsen. Hierdoor zijn we niet alleen in staat om de bruikbaarheid (de kwaliteit) van het zoekresultaat vast te stellen, maar door dezelfde procedure bij verschillende zoekvragen en zoekmethoden toe te passen, kunnen we ook te weten komen hoe de resultaten van de verschillende zoekvragen en zoekmethoden zich tot elkaar verhouden en welke vragen en methoden dus de beste resultaten opleveren.

Om vergelijking tussen de verschillende procedures mogelijk te maken moeten we 2 zaken standaardiseren. In de eerste plaats moeten we (zoveel mogelijk) gebruik maken van identieke steekproeven (van delicten of signalementen) om de matchresultaten van verschillende zoekvragen met elkaar te kunnen vergelijken. In de tweede plaats moeten de uitkomsten van de analyses van verschillende zoekmethoden op een vergelijkbare manier gepresenteerd worden (om na te kunnen gaan welke methode de beste resultaten oplevert).

Voor het toetsen van de verschillende zoekvragen en zoekmethoden maken we telkens gebruik van een *toetsbestand* en een *steekproef*. Het toetsbestand bevat de doelgegevens. Hieruit trekken we een (willekeurige) steekproef van bijvoorbeeld delicten die we één voor één gaan matchen met de doelgegevens. De steekproef bevat dus onze zoekgegevens. Het resultaat van een enkele matchopdracht (het matchen van één delict uit de steekproef met alle gegevens in het doelbestand) levert een rangorde op van (in dit voorbeeld) delicten of delictprofielen naar de mate waarin deze matchen met het zoekdelict. Ergens in die rangorde bevinden zich delicten of bevindt zich het delictprofiel van 'onze dader'. Hoe hoger deze rangordepositie des te beter is onze zoekmethode. Immers, we zijn op zoek naar de bekende dader en ons streven is erop gericht deze zo hoog mogelijk in de rangorde terug te vinden. De hoogste rangordepositie van de werkelijke dader in het zoekresultaat is de uitkomst van die specifieke matchprocedure.

Door deze procedure te herhalen voor alle cases uit de steekproef verkrijgen we een gemiddeld zoekresultaat. Tabel 1 geeft aan hoe een dergelijk resultaat eruit zal zien (de getallen zijn geheel fictief). In deze tabel hebben we de rangordeposities die zijn verkregen uit de afzonderlijke matchopdrachten in 6 klassen of clusters ondergebracht. Voor elke individuele matchopdracht hebben we de hoogste rangordepositie van de werkelijke dader vastgesteld. Dit hebben we 500 keer gedaan, voor elk delict uit de steekproef. De fictieve resultaten hiervan vinden we terug in de tabel. We zien dat in 5% van de gevallen de dader werd aangetroffen bij de 10 best matchende delicten. In 70% van de gevallen werd de dader

¹⁶ Het resultaat van operationele zoekvragen zal altijd uitgedrukt worden in de mogelijkheid om (direct of indirect) een concrete verdachte aan een bepaald zoekgegeven (zoals een delict) te koppelen. Bij zoekvragen met een tactisch doel hoeft dit niet altijd het geval te zijn.

bij de 500 best matchende delicten aangetroffen. In 30% van de gevallen werd de dader niet bij de 500 best matchende delicten aangetroffen.

De kwaliteit van het zoekresultaat wordt enerzijds bepaald door het cumulatief percentage van delicten dat überhaupt gematcht kan worden in één van de 6 clusters. Anderzijds is het van belang dat de kleinste clusters zoveel mogelijk gevuld zijn. Immers, een dader die zich bevindt bij de 10 best matchende delicten is eenvoudiger op te sporen dan een dader die zich bij de eerste 500 delicten bevindt. We hebben ervoor gekozen het grootste cluster te laten eindigen bij rangordepositie 500. Een zoekresultaat van meer dan 500 lijkt ons minder praktisch.¹⁷

Het *format* van tabel 1 levert weliswaar enig informatieverlies op ten opzichte van het gebruik van ‘pure’rangordecijfers, maar het voordeel ervan is dat de resultaten van alle zoekmethoden op deze manier kunnen worden weergegeven, zodat we hun prestaties probleemloos kunnen vergelijken.

Tabel 1 Voorbeeldtabel van een zoekresultaat na matching van een steekproef van 500 delicten

<i>Kleinste cluster (van doelgegevens) waarin 'werkelijke' dader is aangetroffen</i>	<i>Aantal delicten in steekproef (%)</i>	<i>% cumulatief</i>
1: Cluster 1-10	25 (5%)	5%
2: Cluster 11-25	25 (5%)	10%
3: Cluster 26-50	50 (10%)	20%
4: Cluster 51-100	50 (10%)	30%
5: Cluster 101-250	100 (20%)	50%
6: Cluster 251-500	100 (20%)	70%
dader niet aangetroffen in voorgaande clusters	150 (30%)	100%
TOTAAL	500 (100%)	

Nadat het zoekresultaat van tabel 1 is verkregen, is het mogelijk om na te gaan of en zo ja, op welke kenmerken steekproefdelicten met een *goed* matchresultaat afwijken van delicten met een *slecht(er)* matchresultaat. Dit kan ons inzicht verdiepen in de waarde en het toepassingsbereik van een zoekvraag of zoekmethode. Deze correlatieanalyse zal na elke matchprocedure worden uitgevoerd.

5. Simulatie van zoekmethode 1

De waarde van de toetsing staat of valt bij de mogelijkheid om goed in kaart te brengen wat de huidige zoekpraktijk aan resultaten oplevert. Deze resultaten vormen immers het ijkpunt van de toetsing. In hoofdstuk 1 werd al opgemerkt dat er nauwelijks inzicht bestaat in wat de huidige zoekmethode precies oplevert. Daarnaast is uit gesprekken met verschillende HKD-medewerkers gebleken dat de zoekvragen die in het onderhavige onderzoek aan de orde zijn, dat wil zeggen vragen waarop het antwoord vooraf niet bekend is, in de praktijk nauwelijks voorkomen. Rechercheurs selecteren doorgaans kenmerken van een delict en/of signalement en zijn vervolgens geïnteresseerd in de vraag of er meer delicten of signalementen met de betreffende kenmerken in het HKS kunnen worden aangetroffen. Een en ander betekent dat we geen inzicht hebben in de resultaten van de huidige zoekmethode (zoekmethode 1).

Er zijn twee manieren om de zoekresultaten van de huidige zoekmodule in het HKS in kaart te brengen. De ene is om een onderzoek uit te voeren onder HKD-medewerkers en hen

¹⁷ In heel veel gevallen zal een cluster dat groter is dan 100 al weinig praktisch zijn. Er zijn echter ook opsporingsonderzoeken denkbaar waarbij het de moeite zal lonen grotere clusters te onderzoeken. Om die reden hebben we de clusters 101-250 en 251-500 hier vermeld.

cases van bijvoorbeeld delicten voor te leggen met de vraag hier aangiften bij te zoeken die zo goed mogelijk passen. Door te werken met steekproeven van delicten waarvan de dader al bekend is, kan aldus inzicht worden verkregen in de opbrengst van deze zoekmethode. Het alternatief is om de resultaten van de huidige zoekmethode (zoekmethode 1) door middel van computersimulatie in beeld te brengen. Dit is de optie waarvoor wij in dit onderzoek hebben gekozen. Hieraan ligt in eerste instantie een negatieve overweging ten grondslag. Het opzetten van een empirisch onderzoek onder HKD-medewerkers op een zodanige wijze dat daardoor een betrouwbaar beeld kan worden verkregen van de resultaten van de huidige zoekmethode, vereist een te grote investering voor de onderzochten (de HKD-medewerkers). Tenminste een aantal HKD-medewerkers zou bereid gevonden moeten worden om mee te werken (om mogelijke verschillen in zoekresultaten tussen personen in kaart te krijgen). Deze medewerkers zouden vervolgens een groot aantal, in ieder geval meer dan 1000 cases (delicten, signalementen), moeten gaan matchen. Om de resultaten van elke casus in een zodanig format te krijgen dat de resultaten vergeleken kunnen worden met de andere zoekmethoden, zou de HKD-er per casus diverse zoekopties moeten uitproberen. Dit zou hem bij elkaar waarschijnlijk enkele weken gaan kosten. Dit is geen realistische optie.

Aan het gebruik van computersimulatie om de bestaande zoekpraktijk na te bootsen zijn voor- en nadelen verbonden (zoals aan alles in het leven). Deze zullen we aan het einde van deze paragraaf bespreken. We beschrijven eerst op welke wijze de simulatie zal worden uitgevoerd. We zullen het voorbeeld uitwerken van een delictmatching (individueel delict matchen met andere individuele delicten).

In hoofdstuk 2 hebben we uiteengezet hoe de huidige zoekmethode werkt: er worden kenmerken van (in dit geval) een delict geselecteerd en vervolgens gaat men in de doelgegevens op zoek naar delicten die aan deze kenmerken voldoen. Hierbij moet men ervoor waken dat de omvang van het zoekresultaat niet al te omvangrijk is. Deze zoekmethode maakt dat (delict)kenmerken met uitzonderlijke waarden bruikbaar zijn dan kenmerken met veel voorkomende waarden. Die laatste hebben immers onvoldoende discriminerende werking: het enkel zoeken op delicten in de gemeente Den Haag levert in de regio Haaglanden (waar dit een veel voorkomende waarde is) een veel te grote output op. Hoe unieker de waarde, des te bruikbaar is het kenmerk om mee te zoeken. In de praktijk worden daarom vooral kenmerken met (relatief) unieke waarden gebruikt om mee te matchen. Dit uitgangspunt hebben we overgenomen in onze simulatie.

Bij het matchen van delicten hebben we ook nog een tweede uitgangspunt overgenomen. Uit gesprekken met HKD-medewerkers is gebleken dat zij bij delictmatching één kenmerk altijd willen laten matchen, namelijk het wetsartikel in engere of ruimere zin. Dit is altijd de eerste voorwaarde waaraan het zoekresultaat moet voldoen. Vaak gebruikt men hiervoor niet het gedetailleerde wetsartikel zelf, maar een delicttypering die qua abstractie net een stapje hoger ligt. Als het wetsartikel van een zoekdelict bijvoorbeeld SR311.3.4.5 is zal men beginnen met te zoeken naar delicten die voldoen aan SR311. Wij volgen in de simulatie dezelfde aanpak door gebruik te maken van de CBS-delictcategorieën. In sommige gevallen zijn deze veel ruimer geformuleerd dan het onderliggende wetsartikel, in andere gevallen is het verschil tussen beide kleiner (d.w.z. beide waarden zijn dan ongeveer even uniek). Hierbij dient opgemerkt te worden dat het wetsartikel gewoon mee blijft doen in de zoekopdracht als een zelfstandig kenmerk waarop gematcht kan worden naast de cbs-categorie.¹⁸ Aan een delict kunnen meer wetsartikelen, dus ook meer cbs-delictcategorieën ‘hangen’. Een delict uit het doelbestand wordt verondersteld te matchen met het zoekdelict indien tenminste één van de wetsartikelen of cbs-delictcategorieën van dat delict overeenkomen met tenminste één van

¹⁸ Als het wetsartikel een veel uniekere waarde heeft dan de cbs-categorie zal het bij het matchen vanzelf weer boven komen drijven. Als de waarde van het wetsartikel en de cbs-categorie ongeveer even uniek zijn, heeft het wetsartikel geen toegevoegde waarde bij het zoeken.

de wetsartikelen of cbs-delictcategorieën van het zoekdelict. Deze procedure is overeenkomstig de werkwijze in de huidige praktijk.

Afhankelijk van de omvang van het zoekresultaat dat wordt verkregen nadat aan de voorwaarde van de delictsoort is voldaan, kan het resultaat indien nodig worden verkleind door één of meer delictkenmerken toe te voegen aan de zoekopdracht. Hierbij geldt het uitgangspunt dat we hiervoor hebben besproken: hoe unieker de waarde op een kenmerk, des te bruikbaar is het om mee te zoeken. Dit laat onverlet dat men in de praktijk in voorkomende gevallen goede redenen kan hebben om ook te zoeken met kenmerken die minder goed discrimineren (algemenere waarden hebben). Voor het onderhavige onderzoek, en de simulatie die wij uitvoeren, is dit gegeven niet van belang, omdat de keus van dergelijke kenmerken het uiteindelijke zoekresultaat zoals wij dat in tabel 1 hebben weergegeven, niet of nauwelijks beïnvloedt. Immers, de waarden van deze kenmerken discrimineren nauwelijks. Alleen (sterk) discriminerende waarden verkleinen het zoekresultaat zodanig dat het uiteindelijk een maximale omvang van 500 bereikt (maximum van zoekresultaat in tabel 1).

Aan de twee uitgangspunten die hiervoor zijn genoemd, kan nog een derde worden toegevoegd. Als we denken aan de omvang van een zoekresultaat (het aantal aangiften dat matcht), kunnen we dit bereiken door het zoekresultaat op meer of minder waarden te laten matchen. Het is mogelijk een zoekresultaat van 10 aangiften te bereiken door het gebruik van één enkel kenmerk dat sterk discrimineert (bijvoorbeeld een heel bijzondere MO), maar het is ook mogelijk dit resultaat te bereiken door het zoekresultaat te laten voldoen aan vier kenmerken die elk afzonderlijk minder sterk discrimineren, maar in combinatie tot hetzelfde kleine zoekresultaat van 10 aangiften leiden. Het is ons niet bekend hoe hiermee doorgaans in de praktijk wordt omgesprongen, maar het verdient de voorkeur om zo weinig mogelijk kenmerken te gebruiken om tot een zoekresultaat van bepaalde omvang te komen. Hoe meer kenmerken worden gebruikt, des te groter het aantal voorwaarden waaraan het zoekresultaat moet voldoen en des te kleiner de kans op een goed matchresultaat.¹⁹

Het laatste uitgangspunt dat we formuleren voor de simulatie van zoekmethode 1 is het volgende: de simulatie moet zodanig van opzet zijn dat het maximale uit de bestaande zoekmethode wordt gehaald. Met andere woorden, er moet gestreefd worden naar een maximaal haalbaar zoekresultaat bij deze zoekmethode. De bedoeling hiervan is om een zo kritisch mogelijke toets voor de alternatieve zoekmethoden te creëren. Als we het maximale uit zoekmethode 1 hebben gehaald en de andere zoekmethoden blijken toch nog betere resultaten te geven, dan kunnen we behoorlijk zeker zijn van de meerwaarde van deze zoekmethoden.

Aldus kunnen we de volgende 4 uitgangspunten formuleren voor de computersimulatie van zoekmethode 1:

- 1) kenmerken met unieke waarden verdienen de voorkeur bij het matchen boven kenmerken met minder unieke waarden;
- 2) bij het matchen van delicten is het matchen van de waarde op het kenmerk delictsoort (volgens de cbs-categorisering) een noodzakelijke voorwaarde;
- 3) zoekresultaten die zijn gebaseerd op een minimaal aantal gespecificeerde (voor)waarden genieten de voorkeur boven zoekresultaten die zijn gebaseerd op een groter aantal (voor)waarden;
- 4) de simulatie moet zodanig worden vormgegeven dat het maximaal haalbare uit de bestaande zoekmethode wordt gehaald.

¹⁹ Er komt weliswaar altijd een output van bepaalde omvang uit, maar dat zegt nog niks over de kwaliteit van die output.

De simulatieprocedure van zoekmethode 1 kan nu worden beschreven. Om de beschrijving nietodeloos ingewikkeld te maken, laten we het tweede uitgangspunt (delictsoort als noodzakelijke voorwaarde voor matching bij delicten) vooralsnog achterwege. De procedure die we hier beschrijven is toegepast bij het matchen van signalen. We gebruiken deze als voorbeeld. Het matchen van delicten wijkt hierin alleen af op de kwestie genoemd bij uitgangspunt 2. Dit aspect wordt aansluitend toegelicht.

Het streven van de simulatie is erop gericht zoveel mogelijk van de klassen die in tabel 1 genoemd worden, te vullen met verzamelingen van matchende zoekgegevens. Houd echter in de gaten dat we hierna de procedure van een individuele matching beschrijven, terwijl tabel 1 uiteindelijk het gemiddelde resultaat van 500 matchprocedures weergeeft. Om deze tabel echter te kunnen samenstellen moeten we eerst per individuele matchprocedure proberen alle klassen in deze tabel gevuld te krijgen. Hoe dit in zijn werk gaat wordt hierna beschreven.

- *Stap 1: (eerste zoekronde)* eerst wordt bepaald welke kenmerken van het zoeksignalement waarden bevatten;
- *Stap 2:* vervolgens wordt bepaald hoe uniek of algemeen deze waarden zijn door voor elk kenmerk de cluster grootte te bepalen, dit is het aantal signalen in het doelbestand dat dezelfde waarde op het kenmerk heeft;
- *Stap 3:* de 6 kenmerken die de kleinste clusters vormen, worden ‘apart’ gezet. Indien één of meer van deze 6 clusters kleiner of gelijk is aan 500, worden de waarden hiervan weggezet in de klasse(n) van tabel 1. Te beginnen met het kleinste cluster en zo oplopend. Stel dat het kleinste cluster 240 signalen groot is, dan komt het resultaat terecht in rij 5: het cluster 101-250. Als het tweede kleinste cluster bijvoorbeeld 450 groot is, kan het resultaat worden geplaatst in rij 6 (cluster 251-500). Als het tweede kleinste cluster groter is dan 500 wordt er niks mee gedaan. De mogelijkheid kan zich voordoen dat er meer clusters zijn die in één klasse kunnen worden geplaatst. In dat geval wordt het kleinste cluster in de betreffende klasse geplaatst en het andere cluster in de bovenliggende klasse.²⁰ De uitkomst van deze stap kan zijn dat geen enkele klasse gevuld kan worden, omdat alle clusters groter zijn dan 500. Het is ook mogelijk dat alle klassen gevuld worden omdat de 6 kleinste clusters allemaal kleiner zijn dan 500. Waarschijnlijker is het dat hooguit 1 of enkele klassen gevuld worden, omdat de discriminerende werking van afzonderlijke kenmerken doorgaans niet zo groot is dat daarmee meteen hele kleine zoekresultaten worden verkregen.
- *Stap 4: (tweede zoekronde)* het kleinste cluster dat is gevormd op grond van een afzonderlijk kenmerk in zoekronde 1 wordt gebruikt om het zoekresultaat verder te verkleinen. Opnieuw worden cluster groottes bepaald voor alle kenmerken van het signalement dat we willen matchen, maar nu moeten de doelsignalen niet alleen matchen op het betreffende kenmerk, maar ook op het kenmerk dat in de eerste ronde het kleinste cluster vormde. De clusters die aldus worden verkregen verkleinen het zoekresultaat uit de eerste ronde.
- *Stap 5:* Opnieuw kunnen we de 6 kleinste clusters, die nu worden gevormd door signalen die op telkens 2 waarden matchen, op een rij zetten. We herhalen de procedure uit stap 3: opnieuw wordt bekeken of er clusters klein genoeg zijn om in de klassen van tabel 1 geplaatst te kunnen worden. Bedenk dat deze clusters verfijningen zijn van het kleinste cluster uit de eerste ronde. Dus de zoekresultaten zijn hier altijd kleiner dan het kleinste resultaat dat in de eerste ronde werd verkregen. Als het kleinste cluster in de eerste ronde groter dan 500 was dan kunnen in theorie alle

²⁰ Deze procedure kan herhaald worden als er meer dan 2 clusters in een bepaalde klasse vallen, maar dit komt zelden voor.

klassen van tabel 1 in deze ronde alsnog gevuld worden. Als het kleinste cluster in de eerste ronde bijvoorbeeld 80 groot was (en het resultaat dus in rij 4 is geplaatst: cluster 51-100), dan kunnen in deze ronde alleen de eerste 3 klassen nog worden gevuld (tot aan clustergrootte 50). Een eenmaal gevulde klasse kan in een volgende ronde niet overschreven worden, want dit is strijdig met uitgangspunt 4: een resultaat verkregen op basis van 1 (voor)waarde geniet de voorkeur boven een resultaat verkregen op basis van 2(voor)waarden. Als in de eerste ronde het kleinste cluster klein genoeg is om in rij 1 van tabel 1 geplaatst te kunnen worden (clustergrootte 1-10), heeft het dus geen zin meer om een tweede zoekronde uit te voeren.

- *Stap 6: (derde zoekronde)* dit is een herhaling van stap 4, maar nu moeten de gegevens uit het doelbestand matchen op 3 kenmerken, waaronder het kenmerk dat in de eerste ronde het kleinste cluster vormde en het kenmerk dat in de tweede ronde het kleinste cluster vormde. Het spreekt voor zich dat deze stap alleen wordt uitgevoerd en zinvol is indien er klassen (van kleine clusters) in tabel 1 nog niet gevuld zijn. Het streven is er immers op gericht zoveel mogelijk klassen te vullen.
- *Stap 7:* dit is een herhaling van stap 5.
- *Stap 8: (vierde zoekronde)* idem stap 6, maar nu moeten de doelgegevens op 4 kenmerken matchen, waaronder de kenmerken die in de voorgaande 3 rondes de kleinste clusters opleverden. Deze stap wordt alleen uitgevoerd als er nog klassen (van kleine clusters) niet gevuld zijn.
- *Stap 9:* idem stap 8.
- *Stap 10: (bepalen van de rangordepositie van de dader)* we hebben nu 4 matchrondes gedaan, waarbij we in elke ronde een kenmerk hebben toegevoegd waarop de doelgegevens moesten matchen. In elke ronde hebben we de kleinste clusters bepaald en indien deze klein genoeg waren om in tabel 1 te worden geplaatst hebben we dit gedaan. Als er aan de onderkant van tabel 1 (bij de kleinste clusters) nog klassen niet ingevuld waren zijn we verder gegaan met het verkleinen van het zoekresultaat. In de laatste stap wordt het kleinste cluster gevormd door zoekgegevens die voldoen aan 4 kenmerken die in combinatie het kleinst denkbare cluster opleveren. In veel gevallen zal deze verzameling leeg zijn, omdat al in eerdere rondes het kleinst denkbare resultaat (zijnde groter dan 0) werd bereikt. Het resultaat van de operatie tot nu toe is een tabel waarin de klassen gevuld zullen zijn met een cluster dat we gevonden hebben in één van bovenstaande zoekrondes. Het is overigens mogelijk dat niet alle klassen gevuld zijn. We kunnen nu de tabel gaan doorlopen, te beginnen bij de kleinste klasse, en kijken in welke verzameling (in welk cluster) we de werkelijke dader voor het eerst aantreffen. Dit is dan het kleinste cluster waarin de dader zich bevindt en *dit cluster vormt het uiteindelijke zoekresultaat van deze simulatie.*²¹

Door de hier beschreven procedure te herhalen voor alle cases in de steekproef verkrijgen we het gemiddelde zoekresultaat zoals getoond in tabel 1.

Bij het matchen van delicten worden de clusters in de eerste ronde niet gevormd door delicten die enkel matchen op het betreffende kenmerk, maar door delicten die matchen op het betreffende kenmerk én op enige cbs-delictcategorie. Immers, de cbs-delictcategorie moet

²¹ Voor de volledigheid hier nog het volgende: in de simulatieprocedure is beschreven dat het kan voorkomen dat er twee clusters in één klasse van tabel 1 passen en dat in dat geval het grootste cluster opschuift naar de klasse erboven. Dit kan betekenen dat een cluster van 8 delicten terechtkomt in de klasse van 11-25. Als de ons bekende dader zich nu juist in dit cluster bevindt zou het niet goed zijn om als zoekresultaat van deze match het cluster 11-25 aan te geven. Derhalve wordt het matchresultaat uitgedrukt als het kleinste cluster waarin de dader is aangetroffen, in dit geval dus het cluster 1-10.

altijd matchen. Het verschil met het matchen van signalementen bestaat er derhalve uit dat de clusters in ronde 1 gebaseerd zijn op 2 kenmerken in plaats van 1 kenmerk. De clusters in zoekronde 2 zijn gebaseerd op 3 in plaats van 2 kenmerken, enzovoort.

Blijft over de vraag hoe de uitkomsten van deze simulatie zich verhouden tot wat het zoekresultaat zou zijn indien we de zoekopdracht zouden laten uitvoeren door HKD-medewerkers. Een simulatie zoals deze heeft als evident nadeel dat op mechanische wijze een bepaalde procedure wordt uitgevoerd. Het is niet mogelijk om rekening te houden met eventuele bijzonderheden in de zoekgegevens. Een HKD-medewerker zal met zijn kennis en ervaring in voorkomende gevallen tot andere afwegingen komen omtrent de te matchen kenmerken. Afwegingen die misschien ook tot betere zoekresultaten zullen leiden. Dat is uiteraard niet te controleren. Echter, hiertegenover staan ook voordelen en deze wegen naar ons idee uiteindelijk ruimschoots op tegen het nadeel. Als we de simulatie vergelijken met de praktijk kunnen we als eerste voordeel noteren dat alle beschikbare gegevens van het zoekdelict of het zoeksignalement worden benut. In de praktijk selecteren rechercheurs vaak bepaalde kenmerken waarvan ze op voorhand vermoeden dat ze een bepaalde relevantie hebben. Een tweede voordeel van de simulatie is dat van alle beschikbare waarden exact en zonder fouten berekend wordt hoe groot de clusters zijn, dus hoeveel delicten of signalementen matchen op het gespecificeerde kenmerk. In de praktijk is dit niet mogelijk, omdat de HKD-medewerker weliswaar een globaal inzicht heeft in welke waarden uniek zijn en welke algemeen, maar deze kennis heeft uiteraard ook zo zijn beperkingen. Een derde voordeel van deze simulatie boven de bestaande praktijk is dat de simulatie zodanig is ontworpen dat parallelle zoekresultaten zijn toegestaan. Dit gebeurt bij het vullen van de klassen in tabel 1 tijdens de individuele matchprocedure. Parallelle zoekresultaten zijn verzamelingen die niet hiërarchisch geordend zijn. Dit betekent dat het zoekresultaat in klasse 1 geen deelverzameling hoeft te zijn van het zoekresultaat in klasse 2 of 3. Hierdoor kunnen we in voorkomende gevallen te maken hebben met 1 of meer parallelle verzamelingen. Feitelijk betekent dit dat we in zo'n situatie niet met één maar met meer zoekresultaten te maken hebben. Dit vergroot uiteraard de kans op goede matchresultaten. Deze mogelijkheid is bewust ingebouwd om het resultaat van deze zoekmethode te optimaliseren. Ter toelichting: parallelle zoekresultaten kunnen zich voordoen wanneer het kleinste cluster dat op enig moment wordt aangetroffen in een matchprocedure in tabel 1 kan worden geplaatst, maar niet in de grootste klasse (251-500). Stel dat dit cluster 14 (bijv. delicten) groot is, dan wordt het geplaatst in de tweede rij (11-25). Dit creëert een mogelijkheid om de minder grote clusters in de betreffende zoekronde in de bovengelige klassen te plaatsen. Echter, die clusters zijn elk weer op andere voorwaarden (op grond van andere waarden) tot stand gekomen. Derhalve kunnen we in zo'n geval spreken van parallelle zoekresultaten (in tegenstelling tot zoekresultaten die in elkaars verlengde liggen en deelverzamelingen van elkaar vormen, zoals bij zoekmethode 2 het geval is). Overigens betekent plaatsing van het kleinste cluster in een kleinere klasse niet noodzakelijkerwijs dat in de bovengelige klassen parallelle zoekresultaten worden geplaatst. Dit is afhankelijk van de omvang van de betreffende clusters. Als het kleinste cluster op enig moment 40 is en de andere zijn allemaal groter dan 500. Dan komen er in het zoekresultaat geen parallelle zoekresultaten voor (en worden dus de klassen 51-100 en hoger ook niet gevuld). Immers, in de volgende zoekronde wordt het zoekresultaat verfijnd op basis van het kenmerk dat het kleinste cluster heeft opgeleverd.

Op grond van de hier genoemde voordelen achten wij de kans dat de simulatie slechtere zoekresultaten oplevert dan de praktijk zo goed als uitgesloten. Wij achten daarentegen de kans dat de simulatie betere resultaten oplevert dan de praktijk groot.

Om ons inzicht in de resultaten van zoekmethode 1 verder te verdiepen, hebben we een drietal alternatieven uitgewerkt voor de selectie van de kenmerken. Bij het eerste alternatief wordt het aantal kenmerken beperkt dat in de zoekprocedure wordt betrokken. Kenmerken die vaak hele kleine clusters zullen opleveren, zoals gedetailleerde adresgegevens of trefwoorden, worden in deze procedure weggelaten. Bij het tweede alternatief wordt wel weer met alle beschikbare kenmerken van het zoekdelict gezocht, maar nu hoeft de cbs-delictcategorie niet per se te voldoen. Het derde alternatief is een combinatie van de vorige 2 alternatieven: niet alle kenmerken worden gebruikt bij het matchen en de cbs-delictcategorie hoeft niet per se te voldoen. In bijlage 3 zijn de kenmerken opgesomd waarvan gebruik is gemaakt bij het matchen volgens zoekmethode 1.

6. Uitwerking van zoekmethode 3

In het voorgaande hoofdstuk hebben we beschreven dat zoekmethode 3 een combinatie is van zoekmethode 1 en zoekmethode 2. Het resultaat van zoekmethode 2 laat zich eenvoudig samenvatten: de rangordeposities in het matchresultaat komen tot stand op grond van het aantal kenmerken (van het delict, het signalement of het delictprofiel) dat matcht met het zoekgegeven. In de vorige paragraaf hebben we in enig detail de uitwerking van zoekmethode 1 beschreven. Het resultaat komt daar tot stand doordat de doelgegevens matchen op één of meer kenmerken met de meest unieke waarden (al of niet in combinatie).

Zoekmethode 3 bewandelt een middenweg. Het zoekresultaat wordt eerst geordend naar de uitkomsten van zoekmethode 2, maar vervolgens gewogen naar de voorwaarden die bij zoekmethode 1 aan de orde zijn. We hebben hiervan enkele varianten uitgewerkt, om het zoekresultaat bij verschillende voorwaarden te kunnen bestuderen. Het verschil in zoekmethode 1 ten aanzien van het matchen van delicten en signalementen, komt ook terug bij zoekmethode 3.

Bij het matchen van signalementen werkt zoekmethode 3 als volgt: de signalementen in het doelbestand worden geordend naar de resultaten van zoekmethode 2. Vervolgens gaan we na of deze signalementen matchen met de kenmerken die in de eerste ronde bij zoekmethode 1 de drie kleinste clusters hebben opgeleverd. We noemen deze kenmerken voor het gemak A (kenmerk dat kleinste cluster vormt), B (kenmerk dat op één na kleinste cluster vormt) en C (kenmerk dat op twee na kleinste cluster vormt). Als aan de voorwaarde niet wordt voldaan, verdwijnen de signalementen uit de rangorde en worden van de overgebleven signalementen opnieuw rangordeposities berekend. Het zoekresultaat komt tot stand door na te gaan op welke rangordepositie de werkelijke dader zich bevindt. De voorwaarden waaraan moet worden voldaan zullen we als volgt variëren:

- het (doel)signalement matcht met kenmerken A en B en C
- het (doel)signalement matcht met kenmerken A en B
- het (doel)signalement matcht met kenmerken A en C
- het (doel)signalement matcht met kenmerken B en C
- het (doel)signalement matcht met kenmerken A en (B of C)
- het (doel)signalement matcht met kenmerken B en (A of C)
- het (doel)signalement matcht met kenmerken C en (A of B)
- het (doel)signalement matcht met kenmerken A of B
- het (doel)signalement matcht met kenmerken A of C
- het (doel)signalement matcht met kenmerken B of C
- het (doel)signalement matcht met kenmerken A of B of C

De voorwaarden lopen van streng naar soepel. De eerste voorwaarden zijn mogelijk strenger dan de voorwaarden waaraan het resultaat bij zoekmethode 1 moest voldoen. De laatste voorwaarden zullen wellicht soepeler zijn. Hierin zijn we dan ook vooral geïnteresseerd als het gaat om een mogelijk toegevoegde waarde van zoekmethode 3: een combinatie van zoekmethode 2 (alle gegevens meetellen) met zoekmethode 1 (voorwaarden stellen aan de kenmerken die moeten matchen, maar niet heel erg stringent). De eerste voorwaarden hebben we toegevoegd om te onderzoeken op welke wijze deze het zoekresultaat beïnvloeden.

Elk van bovenstaande voorwaarden wordt afzonderlijk gekoppeld aan het resultaat van zoekmethode 2 en levert aldus een aparte ordening van gegevens op waarin we telkens de rangordepositie van de ons bekende dader kunnen bepalen (vgl tabel 1). Deze zoekmethode levert dus 11 rangordeposities op van de ons bekende dader (en dus 11 zoekresultaten!).

De procedure bij het matchen van delicten is nagenoeg identiek, met dit verschil dat de nadere voorwaarden hier niet worden gevormd door het matchen op de kenmerken die de drie kleinste clusters vormen, maar door het matchen op de volgende kenmerken:

A: CBS-delictcategorie

B: het kenmerk dat in combinatie met CBS-delictcategorie het kleinste cluster vormt

C: het kenmerk dat in combinatie met CBS-delictcategorie het op één na kleinste cluster vormt.

7. De rangordepositie van de werkelijke dader bepalen

De kwaliteit van het zoekresultaat wordt telkens uitgedrukt in de rangordepositie van de werkelijke dader in dat zoekresultaat. Doordat we gebruik maken van een simulatiemodel (d.w.z. kennis hebben van de werkelijke dader) hebben we een middel in handen om te beoordelen hoe goed de verschillende zoekmethoden presteren.

Bij zoekmethode 1 wordt de rangordepositie van de werkelijke dader vastgesteld op grond van zijn aanwezigheid in enig cluster (van matchende gegevens). De rangordepositie wordt dan bepaald door de omvang van dat cluster. Het kleinste cluster waarin de werkelijke dader voorkomt, bepaalt welke rangordepositie aan de deze dader wordt toegekend.

Bij de zoekmethoden 2 en 3 worden de doelgegevens geordend naar de mate waarin deze matchen met het zoekgegeven. Hierbij kunnen knopen optreden in het zoekresultaat. Dit zijn doelgegevens met dezelfde matchwaarde. Stel dat bij enige matchprocedure blijkt dat 1 aangifte voor 70% matcht, dat 2 aangiften voor 60% matchen, 10 aangiften voor 50%, 100 aangiften voor 40%, etc. Indien de werkelijke dader zich in de groep bevindt van doelgegevens die voor 40% matchen met het zoekgegeven, zal deze als rangordepositie 113 krijgen. Met andere woorden, als rangordepositie wordt telkens de ‘onderkant’ van de knoop aangehouden. Hoe groter de knopen bij een bepaalde matchprocedure, des te lager zullen de gevonden rangordeposities van de werkelijke dader uitvallen. Dit resultaat hangt samen met de hoeveelheid gegevens die kan worden gematcht: hoe minder gegevens er zijn om te matchen, des te groter de kans op omvangrijke knopen aan de bovenkant van het zoekresultaat (daar waar de beste matches zich bevinden). Hoe groter het aantal gegevens om te matchen, des te kleiner de kans op omvangrijke knopen, omdat er veel fijnmaziger gematcht kan worden.

8. Opzet van de toetsing per groep van zoekvragen

In schema 4 van hoofdstuk 2 hebben we de zoekvragen geordend naar matchprocedure. Dit leverde de volgende 4 groepen op:

Groep 1: matchen van individuele delicten	→ zoekvragen 1,2, 4,5
Groep 2: matchen van individuele delicten met delictprofielen	→ zoekvragen 3,6,7,8
Groep 3: matchen van individuele signaleringen	→ zoekvragen 10,11,12
Groep 4: matchen van delict/signalementcombinaties	→ zoekvragen 16,17,19

Hierna zullen we per groep van zoekvragen de details van de toetsing bespreken.

8.1 Groep 1: matchen van individuele delicten

Uitgaande van de zoekgegevens kunnen we hier 2 soorten analyses onderscheiden:

1. we beschikken over delicten OD die we willen koppelen aan andere delicten (OD en/of BD)
2. we beschikken over delicten BD die we willen koppelen aan andere delicten (OD en/of BD)

De zoekdoelen en zoekvragen die aan deze analyses ten grondslag liggen kunnen uiteraard verschillen, maar vanuit de optiek van het matchen gaat het hier om identieke (want spiegelbeeldige) analyses. Om de matchresultaten te toetsen kunnen we derhalve volstaan met het uitvoeren van één van bovengenoemde analyses. De resultaten van deze toetsing zijn identiek aan de resultaten die behaald zouden worden als we de andere analyse zouden uitvoeren. We kiezen ervoor de eerste analyse uit te voeren.

Deze analyse kunnen we aanwenden voor zowel tactische als operationele doeleinden, maar hiervoor hebben we aangegeven dat alleen de toepassing van een analyse voor operationeel gebruik ons een middel in handen geeft om de matchresultaten te beoordelen op kwaliteit. Dit werkt als volgt: door gebruik te maken van simulatie hebben we kennis van de werkelijke dader bij het matchen. Om deze kennis optimaal te kunnen benutten moeten we ons beperken tot zoekvragen die bedoeld zijn om een concrete verdachte/dader in beeld te krijgen. Dit zijn operationele zoekvragen. Bij deze vragen zijn de doelgegevens enger samengesteld dan bij tactische zoekvragen, ze beperken zich tot informatie over bekende verdachten. Deze informatie hebben we juist nodig om de simulatie goed uit te kunnen voeren. Toevoeging aan het doelbestand van gegevens die niet aan concrete personen kunnen worden gekoppeld (zoals delicten OD), maken het minder goed mogelijk de prestaties van de verschillende zoekmethoden te meten. Als we ons beperken tot doelgegevens die gekoppeld zijn aan bekende verdachten, kan de rangordepositie van de werkelijke dader in het zoekresultaat worden gebruikt als een zuivere indicator van de kwaliteit van de zoekmethode. Dat betekent in dit geval dat we de zoekdelicten alleen zullen matchen met delicten BD, we willen immers op zoek naar de bekende dader. *Formeel gaan we dus delicten OD matchen met delicten BD.* Omdat we deze situatie gaan simuleren maken we bij de zoekgegevens echter geen gebruik van delicten OD, maar van delicten BD die we zullen behandelen als zijnde delicten OD. *Feitelijk gaan we dus delicten BD matchen met delicten BD.*

Voor deze analyse maken we gebruik van het volgende toetsbestand: uit alle delicten met aangiftegegevens tussen 1-1-1999 en 31-12-2001 zijn die delicten geselecteerd waaraan één of meer verdachten zijn gekoppeld. Het gaat in totaal in deze periode om 291.310 delicten,

waarvan er 47.798 gekoppeld zijn aan tenminste één verdachte. Deze 47.798 delicten vormen het toetsbestand. De structuur van dit bestand is simpel: elk record bevat de delictkenmerken van 1 delict en de identiteit van de daaraan gekoppelde verdachte(n) (zie schema 9).

Schema 9 Bestandsstructuur van toetsbestand bij matches van individuele delicten

Delict 1	Kenmerken delict 1	Identiteit verdachte(n) delict 1
Delict 2	Kenmerken delict 2	Identiteit verdachte(n) delict 2
Delict N	Kenmerken delict 3	Identiteit verdachte(n) delict 3

De steekproef van delicten die we gaan matchen bestaat uit 500 willekeurig gekozen delicten uit dit bestand. Elk delict uit de steekproef gaan we matchen met alle andere delicten uit het toetsbestand met uitzondering van de andere steekproefdelicten.²² Om te voorkomen dat we voor de hand liggende matchresultaten verkrijgen doordat er in een proces-verbaal identieke delicten zijn opgenomen, zijn delicten uit hetzelfde proces-verbaal en zelfs delicten die op dezelfde dag zijn gepleegd uitgesloten van de matchprocedure. De uitkomst van een individuele matchprocedure wordt bepaald door de hoogste rangordepositie van een delict waaraan (onder meer) de werkelijke dader is gekoppeld.

8.2 Groep 2: matchen van individuele delicten met delictprofielen

Uitgaande van de zoekgegevens kunnen we hier 3 soorten analyses onderscheiden:

1. we beschikken over delicten OD die we willen koppelen aan delictprofielen
2. we beschikken over delicten BD die we willen koppelen aan delictprofielen
3. we beschikken over delictprofielen die we willen koppelen aan delicten OD of BD

De analyses vermeld bij 1 en 2 zijn weer spiegelbeeldig aan de analyse bij 3. Ook hier geldt dus weer dat het niet zoveel uitmaakt welke analyse we gebruiken voor de toetsing, omdat de resultaten identiek zullen zijn. We kiezen ervoor de eerste analyse uit te voeren, omdat dit ons de mogelijkheid geeft de resultaten te vergelijken met de resultaten van de analyses in groep 1. De vraag die hierdoor beantwoord kan worden is of het matchen van delictprofielen een meerwaarde heeft boven het matchen van individuele delicten. *Formeel matchen we delicten OD met delictprofielen.* Omdat we deze situatie simuleren *matchen we feitelijk delicten BD met delictprofielen.*

Voor deze analyse maken we gebruik van het volgende toetsbestand: we hebben alle personen geselecteerd die delicten hebben gepleegd met een antecedentdatum die ligt tussen 1-1-1999 en 31-12-2001. Van elk van deze personen hebben we een delictprofiel samengesteld. Hierbij hebben we niet alle delicten geselecteerd: we hebben alleen de zogenaamde automatische antecedenten gebruikt, de gekoppelde aangiften (omdat de overige antecedenten te weinig informatie bevatten om zinvol op te kunnen zoeken). Verder hebben we van deze personen alleen delicten gebruikt met een antecedentdatum tussen 1-1-1999 en 31-12-2001. De reden voor deze afbakening is enerzijds praktisch: het analysebestand zou anders onhandelbaar groot worden. Een inhoudelijk argument voor deze afbakening is dat hierdoor gebruik wordt gemaakt van dezelfde verzameling delicten als in groep 1, alleen nu geclusterd binnen daders. Dit geeft een goede mogelijkheid om te beoordelen of het matchen van delictprofielen meerwaarde heeft boven het matchen van individuele delicten. Voor het

²² Omdat het om een willekeurige steekproef gaat kunnen hierin meer delicten van één dader vóórkomen. Dit kan ongewenste effecten veroorzaken in de matchresultaten. Derhalve worden de overige steekproefdelicten bij het matchen buiten beschouwing gelaten.

resultaat van de analyses betekent dit dat het wellicht minder optimaal is dan bij het gebruik van de complete delictprofielen. Voor de onderhavige toetsing is dit echter geen ongunstig uitgangspunt: als het zoekresultaat met deze profielen betere resultaten oplevert dan bij individuele delicten, dan mogen we wellicht nog betere resultaten verwachten bij het gebruik van complete delictprofielen. De 25.735 daders met automatische antecedenten tussen 1999 en 2001 hebben samen 61.782 delicten gepleegd, dat is gemiddeld 2,4 delicten per dader.²³ De bestandstructuur van het toetsbestand is weergegeven in schema 10.

Schema 10 Bestandsstructuur van toetsbestand bij matches van delicten met delictprofielen

Dader 1	Kenmerken delict 1
Dader 1	Kenmerken delict 2
Dader 2	Kenmerken delict 1
Dader 2	Kenmerken delict 2
Dader 2	Kenmerken delict 3
Dader N	Kenmerken delict 1
Dader N	Kenmerken delict N

De steekproef bestaat uit dezelfde 500 willekeurig gekozen delicten als bij groep 1 (om de zoekresultaten te kunnen vergelijken). Elk delict uit de steekproef wordt gematcht met alle andere delicten uit het toetsbestand, met uitzondering van de andere steekproefdelicten en delicten die in hetzelfde proces-verbaal voorkomen of op dezelfde datum zijn gepleegd. Deze delicten worden ook niet meegeteld bij het vaststellen van de gemiddelde matchwaarden per profiel.²⁴ Door de gevonden matchwaarden per delict te middelen en andere bewerkingen uit te voeren, worden de matchwaarden per delictprofiel verkregen (zie uitleg hierover in par. 5.5 van het vorige hoofdstuk). De uitkomst van een individuele matchprocedure wordt bepaald door de rangordepositie van het delictprofiel van de werkelijke dader.

8.3 Groep 3: matches van signalementen

Uitgaande van de zoekgegevens kunnen we hier 2 soorten analyses onderscheiden:

1. we beschikken over signalementgegevens van onbekende daders (OD) en willen deze koppelen aan signalementgegevens van bekende verdachten of andere onbekende daders (BD/OD)
2. we beschikken over signalementgegevens van bekende verdachten (BD) en willen deze koppelen aan signalementgegevens van onbekende daders (OD)

Ook deze analyses zijn spiegelbeeldig. We kiezen ervoor de eerste analyse uit te voeren. Omdat we op zoek zijn naar de bekende dader beperken we het doelbestand tot signalementen BD. We hoeven hier geen simulatie uit te voeren, omdat we gebruik kunnen maken van de bijzondere omstandigheid dat in voorkomende gevallen signalementen OD (via een aangifte) gekoppeld zijn aan signalementen BD van dezelfde persoon (zie par. 4 van dit hoofdstuk). Door het koppelen van deze signalementen OD aan de signalementen BD, kunnen we via de

²³ Het aantal delicten is hier weliswaar groter dan bij de analyses in groep 1, maar het gaat om dezelfde verzameling. In het onderhavige toetsbestand kan een delict meermalen voorkomen als het aan verschillende verdachten gekoppeld is.

²⁴ Immers, anders zou de dader van het steekproefdelict altijd een delict in zijn profiel hebben dat 100 procent matcht. Dat levert geen goede vergelijking met de andere profielen op.

rangordepositie van de werkelijke dader de prestaties van de verschillende zoekmethoden in kaart brengen. *Formeel en feitelijk matchen we dus signaleringen OD met signaleringen BD.*

Voor deze analyse maken we gebruik van het volgende toetsbestand: aan 47.798 van de 291.310 delicten in de periode 1999-2001 zijn één of meer daders gekoppeld. Het gaat in totaal om 25.735 personen. 7302 van deze personen beschikken ook over een signalement (BD) aan de personenkant van het HKS. Deze 7302 signaleringen BD vormen ons toetsbestand (zie structuur van bestand in schema 11).

Schema 11 Bestandsstructuur van toetsbestand bij matchen van signaleringen OD met signaleringen BD

Verdachte 1	Kenmerken signal. BD verdachte 1	-
Verdachte 2	Kenmerken signal. BD verdachte 2	Kenmerken sign. OD verdachte 2 (steekproefsignalement)
Verdachte N	Kenmerken signal. BD verdachte N	-

De steekproef bestaat uit aangiften waaraan zowel een signalement OD als een signalement BD is gekoppeld. Het gaat in totaal om 560 aangiften in de periode 1999-2001. Hieruit hebben we die aangiften geselecteerd waaraan hooguit één signalement OD is gekoppeld en hooguit één signalement BD (om eventuele verwarring bij de vergelijking te voorkomen). Dit zijn 410 aangiften. De delictgegevens van deze aangiften hebben we verder niet gebruikt, alleen de signalementgegevens. De 410 aangiften OD vormen onze steekproef. In het toetsbestand, zie verdachte 2 in schema 11, hebben dus 410 personen niet alleen een signalement BD, maar ook een signalement OD. We gaan nu onderzoeken in hoeverre we in staat zijn om de signaleringen OD te matchen aan de ‘juiste persoon’ in de verzameling signaleringen BD. De uitkomst van een individuele matchprocedure wordt bepaald door de rangordepositie van het signalement BD dat toebehoort aan de persoon van wie ook het signalement OD is.

8.4 Groep 4: matchen van delict/signalementcombinaties

Uitgaande van de zoekgegevens kunnen we hier 2 soorten analyses onderscheiden:

1. we beschikken over delict- en signalementgegevens van delicten OD en willen deze koppelen aan andere delict- en signalementgegevens (OD/BD);
2. we beschikken over delict- en signalementgegevens van delicten BD en willen deze koppelen aan delict- en signalementgegevens van delicten OD.

Wederom gaat het om spiegelbeeldige analyses. Wij beperken ons ook hier tot de eerste analyse. Omdat we op zoek zijn naar de bekende dader, beperken we het doelbestand tot delict/signalementcombinaties van delicten BD. Een belangrijk doel van deze analyse is om vast te stellen hoe het zoekresultaat zich verhoudt tot de resultaten van zoekvragen waarin enkel gebruik wordt gemaakt van delict- respectievelijk signalementgegevens (analyses in groep 1 en 3). Achterliggend idee is dat persoonsgegevens in HKS veel schaarser zijn dan delictgegevens. Wat is nu het effect wanneer we beide groepen van (zoek)gegevens in combinatie gebruiken? Levert dat betere of juist slechtere zoekresultaten op dan bij afzonderlijk gebruik van deze gegevens?

Het toetsbestand ziet er als volgt uit: we beschikken over 7302 signaleringen BD van verdachten die in de periode 1999-2001 één of meer delicten hebben gepleegd. Deze signaleringen hebben we gekoppeld aan de delicten van deze verdachten (aan de aangiftenkant van het HKS), waarbij we ons beperkt hebben tot delicten die plaatsvonden

tussen 1-1-1999 en 31-12-2001. Dit levert een bestand op van 23.950 delict/signalementcombinaties. Eén delict kan gekoppeld zijn aan meer signaleringen BD (het delict is aan meer dan één verdachte gekoppeld en deze beschikken deels of allemaal over een signalement BD) en één signalement BD kan gekoppeld zijn aan meer delicten (de verdachte in kwestie is gekoppeld aan meer dan 1 delict). Dit bestand kan worden beschouwd als een kruising van het toetsbestand bij groep 1 (matchen van delicten) en groep 3 (matchen van signaleringen). In schema 12 is de bestandsstructuur weergegeven. Elk record bevat een unieke combinatie van een signalement BD dat gekoppeld is aan een delict. Zowel signaleringen als delicten kunnen hierdoor dubbel voorkomen, de combinaties zijn echter altijd uniek.

Schema 12 Bestandsstructuur van toetsbestand bij matchen van delict/signalement-combinaties

Verdachte 1	Kenmerken sign. BD verdachte 1	Kenmerken delict 1	
Verdachte 1	Kenmerken sign. BD verdachte 1	Kenmerken delict 2	
Verdachte 2	Kenmerken sign. BD verdachte 2	Kenmerken delict 3	Kenmerken sign. OD verdachte 2 (steekproefsignalement)
Verdachte 2	Kenmerken sign. BD verdachte 2	Kenmerken delict 4	
Verdachte 3	Kenmerken sign. BD verdachte 3	Kenmerken delict 1	
Verdachte N	Kenmerken sign. BD verdachte N	Kenmerken delict	

De steekproef bestaat net als bij de analyses hiervoor in groep 3 uit delicten waaraan zowel een signalement OD als BD is gekoppeld. We zien in schema 12 dat aan delict 3 van verdachte 2 een signalement OD is gekoppeld. De steekproef bestaat hier uit 400 delicten waaraan tegelijkertijd een signalement OD en BD is gekoppeld (10 van de 410 cases uit de steekproef in groep 3 konden hier niet gekoppeld worden). De toetsing is een combinatie van de toetsing in groep 1 en 3: we matchen de 400 delict/signalement OD combinaties met de delict/signalement BD combinaties in het toetsbestand. Delicten uit de steekproef, delicten uit hetzelfde proces-verbaal en delicten die op dezelfde dag zijn gepleegd zijn weer uitgesloten van deze procedure. De uitkomst van een individuele matchprocedure wordt bepaald door de hoogste rangordepositie van een delict/signalement BD combinatie van de werkelijke dader.

Zoals hiervoor al aangegeven kunnen we de resultaten van deze analyse enerzijds vergelijken met de resultaten van de analyses in groep 3 (matchen van signaleringen) en anderzijds met de resultaten van de analyses in groep 1 (matchen van delicten). Een mogelijk probleem bij deze laatste vergelijking is dat de steekproef in groep 1 er anders uitziet dan in deze analyse en dat het toetsbestand daar ook omvangrijker is (het bevat ook delicten waaraan géén signalement BD kan worden gekoppeld). Om mogelijke verschillen in resultaten op deze grond uit te sluiten voeren we hier nogmaals een delictmatching uit zoals in groep 1, maar nu met de steekproef en het toetsbestand van groep 4. Om deze matching goed uit te kunnen voeren moeten delicten die dubbel voorkomen in het toetsbestand (omdat hieraan meer signaleringen BD gekoppeld zijn) worden verwijderd. Het aantal unieke delicten waarop de delictmatching wordt uitgevoerd bedraagt 20.079.

4. Onderzoeksresultaten

In dit hoofdstuk presenteren we de resultaten van de uitgevoerde analyses. In de paragrafen 1 tot en met 4 wordt verslag gedaan van de matchprocedures in de vier groepen. In paragraaf 1: het matchen van individuele delicten, in paragraaf 2: het matchen van individuele delicten met delictprofielen, in paragraaf 3: het matchen van individuele signaleringen, en in paragraaf 4: het matchen van delict/signaalcombinaties. In paragraaf 5 wijden we een nadere beschouwing aan de toepassing van de verschillende zoekvragen voor tactisch gebruik en in paragraaf 6 bespreken we de mogelijkheden om de resultaten van de verschillende zoekmethoden te voorzien van bruikbaarheidsindicatoren. In de 7^e en laatste paragraaf worden de onderzoeksresultaten samengevat.

1. Groep 1: matchen van individuele delicten

Van de delicten die tussen 1-1-1999 en 31-12-2001 in de regio Haaglanden zijn aangegeven hebben we die delicten geselecteerd waarvan één of meer daders bekend zijn (we spreken hier gemakshalve telkens over daders, het gaat feitelijk om verdachten). Dit zijn bijna 48.000 delicten. Uit dit bestand hebben we willekeurig 500 delicten geselecteerd die we elk hebben gematcht met de overige delicten in het bestand. Het resultaat van elke individuele matchprocedure wordt uitgedrukt in de hoogste rangordepositie van een delict waaraan de werkelijke dader is gekoppeld. De werkelijke dader is de dader van het zoekdelict. Deze procedure is herhaald voor alle delicten uit de steekproef. (De details van deze procedure zijn beschreven in het vorige hoofdstuk.)

In de grijze kolommen van tabel 2 op de volgende pagina zijn de zoekresultaten van de huidige zoekmethode weergegeven (zoekmethode 1), zoals wij deze middels computersimulatie hebben verkregen. We zien hier dat bij ruim een kwart van de delicten uit de steekproef (25,6%) de werkelijke dader kon worden aangetroffen bij de beste 500 delicten in het zoekresultaat. In driekwart van de gevallen is de dader dus niet aangetroffen bij de 500 beste delicten in het zoekresultaat (74,4%). Bij één op de twintig delicten uit de steekproef (5,2%) werd de dader aangetroffen bij de beste 10 delicten in het zoekresultaat en bij één op de vijf delicten uit de steekproef (20,8%) werd de dader aangetroffen bij de beste 100 in het zoekresultaat.

In het vorige hoofdstuk hebben we aangegeven dat we het inzicht in de mogelijkheden van de huidige zoekmethode willen verdiepen door enkele alternatieven uit te werken voor de selectie van kenmerken (waarop doeldelicten moeten matchen met het zoekdelict). Bij alternatief 1 is geen gebruik gemaakt van kenmerken die doorgaans hele kleine clusters opleveren, zoals trefwoorden die gekoppeld zijn aan delictkenmerkcodes of gedetailleerde adresgegevens (zoals wbs-code en straatcode). Bij alternatief B is de voorwaarde losgelaten dat de doeldelicten per se moeten matchen op delictcategorie. (Bij dit alternatief zijn wel weer alle kenmerken gebruikt om mee te zoeken.) Alternatief 3 is een combinatie van de voorgaande 2: bij deze optie is geen gebruik gemaakt van de kenmerken die doorgaans hele kleine clusters opleveren. Bovendien hoeft de cbs-delictcategorie hier niet per se te matchen.

Wanneer we de resultaten van deze alternatieven in oenschouw nemen (in tabel 2), kunnen we vaststellen dat deze weinig toevoegen aan de 'standaardprocedure' waarbij alle kenmerken van een zoekdelict worden gebruikt en de doeldelicten per se moeten matchen op delictcategorie. Er zijn twee indicatoren waarop we de resultaten van de alternatieven kunnen beoordelen: het cumulatief percentage van delicten in de steekproef waarbij de dader is aangetroffen bij de beste 500 delicten in het zoekresultaat en het percentage delicten in de steekproef waarbij de dader in de kleinste clusters is aangetroffen. We zien dat alternatief 1 en

2 marginaal beter scoren als het gaat om het cumulatief percentage van steekproefdelicten waarbij de dader is aangetroffen bij de eerste 500 delicten in het zoekresultaat (respectievelijk 28,4% en 27% versus 25,6% bij de standaardprocedure). Als we vervolgens kijken naar het percentage steekproefdelicten waarbij de dader is aangetroffen bij de beste 10 of 25 delicten in het zoekresultaat, zien we dat de alternatieven marginaal slechter scoren dan de standaardprocedure. Bij de standaardprocedure kan de dader in 5,2 respectievelijk 10,6% van de gevallen worden aangetroffen bij de 10 respectievelijk 25 beste delicten in het zoekresultaat. De alternatieven scoren hier 4,2 en 7,6% (alternatief 1) en 4,6 en 9%.

We concluderen derhalve dat de resultaten van de alternatieve uitwerkingen van zoekmethode 1 niet noemenswaardig afwijken van de resultaten van de standaardprocedure. We zullen hierna dan ook de resultaten van de standaardprocedure gebruiken als ijkpunt voor de alternatieve zoekmethoden.

Tabel 3 (A,B en C) bevat de resultaten van de zoekmethoden 2 en 3. De resultaten van zoekmethode 2 (in tabel 3A) laten zien dat bij 29,6% van de delicten de dader werd aangetroffen bij de 500 beste delicten in het zoekresultaat. Dit is iets beter dan bij zoekmethode 1 (25,6%). We zien echter ook dat de verdeling van de steekproefdelicten over de verschillende klassen hier duidelijk afwijkt van de verdeling die werd verkregen bij zoekmethode 1. Bij zoekmethode 2 is in 13,4% van de gevallen de dader aangetroffen bij de beste 10 delicten in het zoekresultaat. Bij zoekmethode 1 is dit slechts 5,2%. Dit is een duidelijk beter resultaat.

In de overige kolommen van tabel 3A en in de tabellen 3B en 3C zijn de resultaten van zoekmethode 3 weergegeven. Hiervoor hebben we aangegeven dat het resultaat van deze zoekmethode tot stand komt door de rangordeposities die zijn verkregen op basis van zoekmethode 2 te wegen naar het al of niet voldoen aan bepaalde voorwaarden van zoekmethode 1. Deze voorwaarden zijn dat de doeldelicten moeten matchen op:

A: CBS-delictcategorie;

B: het kenmerk dat in combinatie met CBS-delictcategorie het kleinste cluster vormt in het Zoekresultaat;

C: het kenmerk dat in combinatie met CBS-delictcategorie het op één na kleinste cluster Vormt in het zoekresultaat.

We hebben verschillende varianten voor deze voorwaarden uitgewerkt, die lopen van streng naar soepel. Bij de eerste variant moet het zoekresultaat voldoen aan alle hier genoemde kenmerken (A+B+C; zie tabel 3a). Bij de meest soepele variant moeten de doeldelicten voldoen aan A of B of C (laatste kolommen van tabel 3D). Wanneer we de resultaten van de verschillende varianten van zoekmethode 3 beschouwen, kunnen we constateren dat deze beter worden naarmate de voorwaarden soepeler zijn geformuleerd. Wanneer de doeldelicten moeten voldoen aan zowel kenmerk A als B als C, kan bij slechts 5,8% van de steekproefdelicten de dader worden aangetroffen bij de beste 500 delicten in het zoekresultaat. We zien dat een aantal klassen zelfs niet gevuld wordt, omdat de voorwaarden zo streng zijn dat slechts enkele delicten uit de steekproef hieraan kunnen voldoen. Deze scoren dan wel relatief goed, dat wil zeggen dat de werkelijke dader in deze gevallen veelal kan worden aangetroffen bij de beste 10 of 25 delicten in het zoekresultaat. Het versoepelen van de voorwaarde zodat de doeldelicten nog maar op 2 geselecteerde kenmerken hoeven matchen (A&B, A&C, B&C) zorgt voor enige verbetering in het resultaat, maar het cumulatief percentage van de steekproefdelicten waarbij de dader wordt aangetroffen bij de eerste 500 delicten in het zoekresultaat blijft onder het niveau van de zoekmethoden 1 en 2.

Tabel 2 Matchen van individuele delicten: resultaten van zoekmethode 1 (volledige steekproef N=500)

	Zoekmethode 1		Zoekmethode 1 (alternatief 1)		Zoekmethode 1 (alternatief 2)		Zoekmethode 1 (alternatief 3)	
	%	cumulatief	%	cumulatief	%	cumulatief	%	cumulatief
<i>Kleinste cluster waarin werkelijke dader is aangetroffen</i>								
1: Cluster 1-10	5,2	5,2	4,2	4,2	4,6	4,6	5,0	5,0
2: Cluster 11-25	5,4	10,6	3,4	7,6	4,4	9,0	4,2	9,2
3: Cluster 26-50	6,4	17,0	7,6	15,2	3,4	12,4	4,6	13,8
4: Cluster 51-100	3,8	20,8	3,2	18,4	5,0	17,4	2,4	16,2
5: Cluster 101-250	3,4	24,2	5,6	24,0	6,6	24,0	3,8	20,0
6: Cluster 251-500	1,4	25,6	4,4	28,4	3,0	27,0	2,6	22,6
7: Dader niet in vorige clusters	74,4	100,0	71,6	100,0	73,0	100,0	77,4	100,0
TOTAAL N=500	100,0		100,0		100,0		100,0	

Tabel 3A Matchen van individuele delicten: resultaten van zoekmethode 2 en 3 (volledige steekproef N=500)

	Zoekmethode 2		Zoekmethode 3 Voorwaarde: A+B+C		Zoekmethode 3 Voorwaarde: A+B		Zoekmethode 3 Voorwaarde A+C	
	%	cumulatief	%	cumulatief	%	cumulatief	%	cumulatief
<i>Kleinste cluster waarin werkelijke dader is aangetroffen</i>								
1: Cluster 1-10	13,4	13,4	5,4	5,4	8,0	8,0	9,8	9,8
2: Cluster 11-25	2,6	16,0	4,4	5,8	8,8	8,8	1,0	10,8
3: Cluster 26-50	3,6	19,6			4,4	9,2	6,4	11,4
4: Cluster 51-100	2,6	22,2			2,2	9,4	4,4	11,8
5: Cluster 101-250	4,4	26,6						
6: Cluster 251-500	3,0	29,6						
7: Dader niet in vorige clusters	70,4	100,0	94,2	100,0	90,6	100,0	88,2	100,0
TOTAAL N=500	100,0		100,0		100,0		100,0	

Tabel 3B Matchen van individuele delicten: resultaten van zoekmethode 3-ervolg 1 (volledige steekproef N=500)

	Zoekmethode 3 Voorwaarde: B+C		Zoekmethode 3 Voorwaarde: A+(B of C)		Zoekmethode 3 Voorwaarde: B+(A of C)		Zoekmethode 3 Voorwaarde: C+(A of B)	
	%	cumulatief	%	cumulatief	%	cumulatief	%	cumulatief
<i>Kleinste cluster waarin werkelijke dader is aangetroffen</i>								
1: Cluster 1-10	6,2	6,2	11,0	11,0	8,6	8,6	10,0	10,0
2: Cluster 11-25	,4	6,6	1,8	12,8	1,0	9,6	1,4	11,4
3: Cluster 26-50	,2	6,8	,8	13,6	,6	10,2	,8	12,2
4: Cluster 51-100			,6	14,2			,4	12,6
5: Cluster 101-250	,2	7,0	,2	14,4	,4	10,6	,2	12,8
6: Cluster 251-500			,2	14,6				
7: Dader niet in vorige clusters	93,0	100,0	85,4	100,0	89,4	100,0	87,2	100,0
TOTAAL N=500	100,0		100,0		100,0		100,0	

Tabel 3C Matchen van individuele delicten: resultaten van zoekmethode 3-ervolg 2 (volledige steekproef N=500)

	Zoekmethode 3 Voorwaarde: A of B		Zoekmethode 3 Voorwaarde: A of C		Zoekmethode 3 Voorwaarde: B of C		Zoekmethode 3 Voorwaarde: A of B of C	
	%	cumulatief	%	cumulatief	%	cumulatief	%	cumulatief
<i>Kleinste cluster waarin werkelijke dader is aangetroffen</i>								
1: Cluster 1-10	16,8	16,8	16,8	16,8	11,4	11,4	16,8	16,8
2: Cluster 11-25	3,6	20,4	3,6	20,4	3,0	14,4	3,6	20,4
3: Cluster 26-50	2,4	22,8	2,2	22,6	1,8	16,2	2,4	22,8
4: Cluster 51-100	2,4	25,2	2,2	24,8	1,6	17,8	2,4	25,2
5: Cluster 101-250	3,8	29,0	4,2	29,0	1,4	19,2	4,0	29,2
6: Cluster 251-500	2,2	31,2	2,4	31,4	,4	19,6	2,2	31,4
7: Dader niet in vorige clusters	68,8	100,0	68,6	100,0	80,4	100,0	68,6	100,0
TOTAAL N=500	100,0		100,0		100,0		100,0	

Dit geldt ook voor de varianten waarbij de doeldelicten moeten matchen op 2 kenmerken, maar waarbij enige soepelheid is ingebouwd ten aanzien van de kenmerken waarop ze moeten matchen. Wel zien we dat de zoekresultaten hierbij weer iets beter zijn dan bij de vorige groep (matchen op 2 vooraf geselecteerde kenmerken). In tabel 3C zijn de meest soepele varianten van zoekmethode 3 weergegeven. Hier zien we ineens aanzienlijke betere matchresultaten, in 3 gevallen ook weer iets beter dan het resultaat dat werd verkregen op grond van zoekmethode 2. Als de doeldelicten moeten voldoen aan kenmerk A of B of C, zien we dat bij 31,4% van de steekproefdelicten de dader werd aangetroffen bij de beste 500 delicten in het zoekresultaat (tegen 25,6% bij de standaardprocedure). In 16,8% van de gevallen werd de dader aangetroffen bij de eerste 10 delicten in het zoekresultaat (tegen 5,2% bij de standaardprocedure).

De verschillende varianten van zoekmethode 3 laten zien hoezeer het toevoegen van voorwaarden bij het matchen van delicten ten koste gaat van de kwaliteit van het zoekresultaat. Alleen het stellen van hele soepele voorwaarden levert geen verslechtering van het resultaat op en kan in combinatie met het zoekresultaat dat is verkregen op grond van zoekmethode 2 zelfs tot een nog iets betere uitkomst leiden.

De zoekresultaten zoals gepresenteerd in de tabellen 2 en 3 zijn gebaseerd op de matchresultaten van alle delicten in de steekproef (N=500). Als zodanig geven ze een goed beeld van de uitkomsten die verwacht kunnen worden wanneer we een willekeurig delict van een onbekende dader matchen met delicten van bekende daders. Bij alle zoekmethoden hebben we kunnen waarnemen dat in de meerderheid van de gevallen de werkelijke dader niet werd aangetroffen bij de eerste 500 delicten in het zoekresultaat. Bij zoekmethode 1 gaat het om 74,4% van de delicten in de steekproef, bij zoekmethode 3 (de meest soepele variant) gaat het om 68,6%. Deze verzameling van steekproefdelicten, waarbij de dader niet kan worden aangetroffen bij de beste 500 delicten in het zoekresultaat, kan uiteengelegd worden in 2 deelverzamelingen:

- het kan hierbij enerzijds gaan om delicten gepleegd door een dader die weliswaar (vaker) voorkomt in het HKS, maar waarbij de diverse zoekmethoden niet in staat zijn gebleken om deze persoon hoog genoeg in de rangorde te krijgen;
- het kan hierbij anderzijds gaan om delicten gepleegd door een dader die niet (vaker) voorkomt in het HKS, waardoor het niet mogelijk is om deze persoon te matchen.

Bij het zoeken in de praktijk weten we niet of de dader van het zoekdelict zich in het HKS bevindt of niet. Omdat wij deze kennis wel hebben, kunnen we nagaan in welke gevallen de dader (nogmaals) voorkomt in het HKS.²⁵ We kunnen nu de zoekresultaten presenteren van die delicten waarvan bekend is dat de dader (nogmaals) in het HKS voorkomt. Dit geeft een meer zuivere vergelijking van de prestaties van de verschillende zoekmethoden.²⁶ Bovendien vergemakkelijkt een dergelijke presentatie de vergelijking tussen zoekresultaten die gebaseerd zijn op verschillende steekproeven.

In tabel 4 zijn de zoekresultaten van zoekmethode 1 en zoekmethode 3 (de variant met de beste uitkomst) nogmaals naast elkaar gezet, maar nu is het zoekresultaat enkel gebaseerd op de steekproefdelicten waarvan bekend is dat de dader (nogmaals) voorkomt in het HKS. Dit zijn er 331 (van de 500). We zien in deze tabel dat zoekmethode 1 er in slaagt om de

²⁵ De aanduiding 'nogmaals' verwijst naar het feit dat elke dader uit de steekproef met tenminste één delict in het HKS voorkomt. Dit vloeit voort uit de selectie van de steekproef. Het gaat er dus om dat ze NOGMAALS in het HKS voorkomen (met één of meer andere delicten).

²⁶ Een groot aantal delicten dat niet kan worden gematcht omdat de dader zich niet in het HKS bevindt, kan het verschil in uitkomsten tussen zoekmethoden flatteren.

werkelijke dader in 38,7% van de gevallen bij de beste 500 delicten in het zoekresultaat te krijgen (als de dader dus –nogmaals- voorkomt in het HKS!). Bij zoekmethode 3 lukt dat in 47,4% van de gevallen. Opvallend is het verschil bij het kleinste cluster: als de dader voorkomt in het HKS, zal deze bij zoekmethode 1 in 7,9% van de gevallen worden aangetroffen bij de beste 10 delicten van het zoekresultaat. Bij zoekmethode 3 ligt dit percentage ruim drie keer zo hoog: 25,4%; dus in meer dan een kwart van de gevallen zal bij zoekmethode 3 de dader bij de beste 10 delicten in het zoekresultaat worden aangetroffen.

Tabel 4 Matchen van individuele delicten: resultaten van zoekmethoden 1 en 3 (selectie van daders die (nogmaals) vóórkomen in het HKS; steekproef N=331)

<i>Kleinste cluster waarin werkelijke dader is aangetroffen</i>	<i>Zoekmethode 1</i>		<i>Zoekmethode 3 Voorwaarde: A of B of C</i>	
	<i>%</i>	<i>% cumulatief</i>	<i>%</i>	<i>% cumulatief</i>
1: Cluster 1-10	7,9	7,9	25,4	25,4
2: Cluster 11-25	8,2	16,0	5,4	30,8
3: Cluster 26-50	9,7	25,7	3,6	34,4
4: Cluster 51-100	5,7	31,4	3,6	38,1
5: Cluster 101-250	5,1	36,6	6,0	44,1
6: Cluster 251-500	2,1	38,7	3,3	47,4
7: Dader niet in vorige clusters	61,3	100,0	52,6	100,0
TOTAAL N=331	100,0		100,0	

Correlatieanalyse

Uitgaande van de gegevens in tabel 4 kunnen we nagaan of delicten met goede matchresultaten afwijken van delicten met minder goede of slechte matchresultaten. Dit kan ons inzicht verdiepen in de mogelijkheden en het toepassingsbereik van verschillende zoekmethoden. De analyses waarvan hierna verslag wordt gedaan zijn soms gebaseerd op kleine aantallen. Gevonden verbanden worden alleen besproken indien ze statistisch significant zijn op het $p < .05$ niveau.

Ten aanzien van verschillen tussen delicten uit de steekproef die goed matchen en delicten die minder goed of slecht matchen, hebben we de volgende waarnemingen gedaan (we zullen ze hier op een rij zetten en daarna bespreken, zm=zoekmethode):

- hoe groter het aantal daders van een delict, des te beter is het matchresultaat
 - o ($r = -.21^{**}$ bij zm1, $r = -.26^{**}$ bij zm3)
- hoe groter het aantal gegevens dat beschikbaar is van een delict, des te beter is het matchresultaat bij zoekmethode 3 (en 2)
 - o ($r = -.14^{**}$ bij zm3)
- geweldsdelicten laten zich bij zoekmethode 3 slechter dan gemiddeld matchen
 - o ($r = .17^{**}$ bij zm3)
- vermogensdelicten laten zich beter dan gemiddeld matchen. Dit effect is bij zoekmethode 3 sterker dan bij zoekmethode 1
 - o ($r = -.12^*$ bij zm1, $r = -.22^{**}$ bij zm3)
- delicten tegen de openbare orde en het openbaar gezag laten zich bij zoekmethode 1 beter dan gemiddeld matchen
 - o ($r = -.16^{**}$ bij zm1)

- verkeersdelicten laten zich slechter dan gemiddeld matchen
 - o ($r = .20^{**}$ bij zm1, $r = .22^{**}$ bij zm3)
- delicten waarbij de dader in contact komt met zijn slachtoffer laten zich bij zoekmethode 3 slechter dan gemiddeld matchen
 - o ($r = .19^{**}$ bij zm3)
- delicten waarbij de dader gebruik maakt van specifieke hulpstukken laten zich beter dan gemiddeld matchen
 - o ($r = -.14^{**}$ bij zm1, $r = -.21^{**}$ bij zm3)
- delicten waarbij de dader elektrische apparatuur heeft buitgemaakt laten zich bij zoekmethode 3 beter dan gemiddeld matchen
 - o ($r = -.15^{**}$ bij zm3)

Let op: de meeste correlatiecoëfficiënten ogen weliswaar zwak, maar de gevonden verbanden hebben in de meeste gevallen een substantieel effect op het zoekresultaat.

Toelichting op de gevonden verbanden:

Indien een delict door meer dan één dader is gepleegd, is de kans groter dat er in het HKS goed matchende delicten worden gevonden, omdat elke afzonderlijke dader in het HKS kan vóórkomen en aldus kan bijdragen aan een goed matchresultaat. Het gemiddeld aantal daders per delict in de steekproef bedraagt 1.6. Alleen bij de delicten in het kleinste cluster (de best matchende delicten dus) ligt het gemiddeld aantal daders royaal boven de 2, namelijk 2.5 daders per delict (bij zoekmethode 3). In de andere clusters bij zoekmethode 3 varieert het aantal daders per delict tussen 1.2 en 1.6.

Dat het aantal waarden waarmee gezocht kan worden van invloed is op het matchresultaat bij zoekmethode 2 en 3 ligt enigszins voor de hand: hoe meer waarden beschikbaar zijn om mee te zoeken, des te fijnmaziger –en dus ook beter- zal het zoekresultaat zijn. De delicten in het kleinste cluster van zoekmethode 3 hebben gemiddeld 27,4 waarden. In de andere clusters ligt het aantal waarden 1 tot 3 lager.²⁷ Het aantal beschikbare waarden heeft –om voor de hand liggende redenen- geen invloed op het zoekresultaat bij zoekmethode 1; deze methode is immers gebaseerd op het selectief gebruik van kenmerken bij het zoeken.

Verschillende delictsoorten scoren boven- of juist beneden het gemiddelde van de matchresultaten. De matchresultaten van verkeersdelicten en in mindere mate ook van geweldsdelicten liggen beneden het gemiddelde, terwijl vermogensdelicten en in mindere mate ook openbare orde & gezag-delicten boven het gemiddelde scoren. Om een indruk te geven van de omvang van deze verschillen: de gemiddelde waarde van de klasse waarin vermogensdelicten gematcht zijn bedraagt bij zoekmethode 3: 4.2. (= cluster 51-100), terwijl de gemiddelde waarde van alle overige (niet-vermogens) delicten 5.4 is (dus ruim een klasse hoger). Het verschil tussen de matchresultaten van verkeersdelicten en overige delicten is nog iets groter: de gemiddelde klasse waarin verkeersdelicten zijn gematcht is 6.3, de overige delicten scoren gemiddeld in klasse 4.6. Het is voorsnog onduidelijk waarom bepaalde delictsoorten beter matchen dan andere.

Ook enkele MO-kenmerken correleren met de gevonden matchresultaten. Delicten waarbij geweld wordt gebruikt laten zich bij zoekmethode 3 slechter dan gemiddeld matchen. Een identiek resultaat vonden we hiervoor op basis van de delictcategorie (vandaar dat we het resultaat niet nogmaals hebben weergegeven). Verder zien we bij zoekmethode 3 dat delicten waarbij de dader in contact is gekomen met het slachtoffer slechter dan gemiddeld matchen, terwijl delicten waarbij de dader elektrische apparatuur heeft buitgemaakt beter dan gemiddeld matchen. De verschillen in matchresultaten bedragen in beide gevallen ruim een

²⁷ Het aantal waarden is gebaseerd op de velden die in het HKS beschikbaar zijn om mee te zoeken plus de door ons aangemaakte extra velden (zie schema 8 in het vorige hoofdstuk).

klasse. Beide zoekmethoden laten zien dat het gebruik van specifieke hulpmiddelen bij een delict het matchresultaat van dit delict ten goede komt. Het verschil in matchresultaat bedraagt bij zoekmethode 3 circa anderhalve klasse (het gemiddeld matchresultaat is klasse 3.6, voor delicten die waarden hebben op dit MO-veld, tegen 5.1 voor delicten die op dit MO-veld geen waarde hebben).

De zoekresultaten die zijn verkregen op grond van de in tabel 4 gepresenteerde zoekmethoden 1 en 3 correleren $r=.8^{**}$ met elkaar.²⁸ Dit betekent weliswaar dat de meeste delicten die goed matchen op de ene zoekmethode ook goed matchen op de andere, maar er zullen ook delicten zijn die goed matchen op de ene methode en slecht op de andere. Deze delicten kunnen ons inzicht verdiepen in het soort delicten waarvoor de ene zoekmethode blijkbaar gevoeliger is dan de andere. Om hierin inzicht te krijgen hebben we enerzijds delicten geselecteerd die hoog scoren op zoekmethode 1 en laag op zoekmethode 3. Anderzijds hebben we delicten geselecteerd die hoog scoren op zoekmethode 3 en laag op zoekmethode 1. Ofschoon de vergelijking betrekking heeft op kleine aantallen, kunnen we toch enkele significante verschillen constateren. Delicten die hoog scoren op zoekmethode 1 en laag op zoekmethode 3 zijn vaker dan gemiddeld geweldsdelicten (of delicten met een gewelds-MO). Vermogensdelicten komen in deze groep minder dan gemiddeld voor. Verder zijn het delicten waarbij veel vaker dan gemiddeld sprake is van contact tussen dader en slachtoffer. Delicten die hoog scoren op zoekmethode 3 en laag op zoekmethode 1 hebben een bovengemiddeld aantal waarden om mee te zoeken, zijn vaker dan gemiddeld vermogensdelicten en zijn vaker dan gemiddeld delicten die een waarde hebben op één van de nieuw samengestelde MO-categorieën (met uitzondering van de MO-categorie: contact van dader met slachtoffer). Twee voorzichtige conclusies laten zich hieruit trekken:

1. zoekmethode 1 lijkt het iets beter te doen bij delicten met meer unieke waarden (vooral in de sfeer van de delictcategorie), terwijl zoekmethode 3 beter lijkt te scoren bij veel voorkomende delictsoorten;
2. zoekmethode 3 profiteert van de extra gegevensvelden die zijn gecreëerd op basis van de delictkenmerkcodes op het lmf-formulier. De steekproefdelicten die hoog scoren op zoekmethode 3 en laag op zoekmethode 1 hebben veel vaker dan gemiddeld waarden op deze kenmerken.

2. Groep 2: matchen van individuele delicten met delictprofielen

Om te onderzoeken of er bij het matchen van delicten een meerwaarde kan worden bereikt door de gegevens van delicten te clusteren binnen verdachten, hebben we een toetsbestand gecreëerd van delicten die in de periode 1999-2001 zijn aangegeven in de regio Haaglanden. Het gaat delicten waaraan tenminste één verdachte is gekoppeld. Deze delicten hebben we geclusterd binnen de betrokken verdachten. Aldus hebben we van 25.735 verdachten een delictprofiel samengesteld dat bestaat uit de delicten die zij tussen 1-1-1999 en 31-12-2001 hebben gepleegd. De steekproef van delicten die we met deze delictprofielen matchen is dezelfde als die we bij de vorige analyse hebben gebruikt.

In tabel 5 zijn de matchresultaten van de drie varianten van zoekmethode 1 weergegeven. Let op, het gaat hierbij niet om de huidige zoekmethode in het HKS, want die kan geen geclusterde gegevens matchen. Het zijn door ons ontwikkelde varianten die verwant

²⁸ ** Correlatie is significant op het $P < 0.01$ niveau (2-zijdig); * Correlatie is significant op het $P < 0.05$ niveau (2-zijdig).

Tabel 5 Matchen van individuele delicten met delictprofielen: resultaten van zoekmethode 1 (volledige steekproef N=500)

	<i>Zoekmethode 1 A: Gemiddelde cluster grootte</i>		<i>Zoekmethode 1 B: percentage delicten dat matcht</i>		<i>Zoekmethode 1 C: combinatie van A & B</i>	
	%	% cum	%	% cum	%	% cum
<i>Kleinste cluster waarin werkelijke dader is aangetroffen</i>						
1: Cluster 1-10	5,6	5,6	1,2	1,2	5,2	5,2
2: Cluster 11-25	4,4	10,0	3,6	4,8	3,8	9,0
3: Cluster 26-50	4,8	14,8	4,0	8,8	4,4	13,4
4: Cluster 51-100	3,0	17,8	4,6	13,4	3,8	17,2
5: Cluster 101-250	4,4	22,2	7,8	21,2	5,2	22,4
6: Cluster 251-500	1,4	23,6	2,2	23,4	1,2	23,6
7: Dader niet in vorige clusters	76,4	100,0	76,6	100,0	76,4	100,0
TOTAAL N=500	100,0		100,0		100,0	

zijn aan deze methode, waardoor we gemakshalve toch spreken van zoekmethode 1. Bij de eerste variant zijn de profielen gematcht op de gemiddelde cluster grootte van die delicten in het profiel die voldoen aan de gestelde voorwaarden. Bij variant 2 is het zoekresultaat tot stand gekomen door het percentage delicten in het profiel te tellen dat matcht met het zoekdelict. Het zoekresultaat bij variant 3 is tot stand gekomen op basis van een combinatie van de voorgaande 2 maten (percentage delicten dat matcht maal de gemiddelde cluster grootte van de matchende delicten in het profiel).

We zien in tabel 5 dat de resultaten van de verschillende varianten van zoekmethode 1 niet heel sterk uiteenlopen. De beste zoekresultaten worden behaald bij het gebruik van variant A. Deze slaagt erin om het delictprofiel van de werkelijke dader in 5.6% van de gevallen bij de beste 10 profielen in het zoekresultaat te matchen. In totaal wordt de werkelijke dader in 23,6% van de gevallen aangetroffen bij de eerste 500 delictprofielen in het zoekresultaat.

In tabel 6 hebben we de zoekresultaten van zoekmethoden 2 en 3 weergegeven. We presenteren hier niet meer alle varianten van zoekmethode 3, maar alleen de variant die de beste matchwaarden oplevert. We hebben hiervoor gezien dat dit het geval is wanneer de aanvullende voorwaarden zo soepel mogelijk worden geformuleerd. In deze tabel zien we dat zoekmethode 2 tot betere matchresultaten leidt dan de beste variant van zoekmethode 3. Zoekmethode 2 slaagt erin om in 19% van de gevallen de dader bij de beste 500 profielen in het zoekresultaat te matchen. Bij zoekmethode 3 gebeurt dit in 13% van de gevallen. Ook bij de kleinste clusters scoort zoekmethode 2 iets beter dan zoekmethode 3. Als we de resultaten van deze tabel echter vergelijken met die uit tabel 5, kunnen we vaststellen dat de verschillende varianten van zoekmethode 1 hier beter scoren dan de zoekmethoden 2 en 3.

Tabel 6 Matchen van individuele delicten met delictprofielen: resultaten van zoekmethoden 2 en 3-beste variant (volledige steekproef N=500)

<i>Kleinste cluster waarin werkelijke dader is aangetroffen</i>	<i>Zoekmethode 2</i>		<i>Zoekmethode 3 Voorwaarde: A of B</i>	
	%	% cumulatief	%	% cumulatief
1: Cluster 1-10	4,6	4,6	3,2	3,2
2: Cluster 11-25	2,6	7,2	2,2	5,4
3: Cluster 26-50	1,4	8,6	1,4	6,8
4: Cluster 51-100	1,8	10,4	,8	7,6
5: Cluster 101-250	5,2	15,6	2,4	10,0
6: Cluster 251-500	3,4	19,0	3,0	13,0
7: Dader niet in vorige clusters	81,0	100,0	87,0	100,0
TOTAAL N=500	100,0		100,0	

Bij het gebruik van geclusterde delictgegevens is het mogelijk om de factor tijd en verschillen in omvang van de clusters te verdisconteren door niet alle delicten uit het delictprofiel mee te laten doen aan de matchprocedure of door een tijdweging uit te voeren op de gevonden matchwaarden. Concreet hebben we 3 varianten uitgewerkt: A) matchwaarden van individuele delicten *wegen naar tijd*: hoe verder ze van het zoekdelict liggen, hoe minder zwaar ze meetellen in het gemiddelde matchresultaat van het profiel, B) alleen delicten die zijn gepleegd *binnen 1 jaar van het zoekdelict* tellen mee bij het vaststellen van de gemiddelde matchwaarde van het profiel en C) alleen de *3 best matchende* delicten tellen mee bij het bepalen van de gemiddelde matchwaarde van het profiel.

In tabel 7 zijn opnieuw de resultaten van zoekmethode 2 weergegeven, maar nu gebaseerd op de hiervoor besproken weging en selecties van delicten in het delictprofiel. Wanneer we de resultaten vergelijken met de resultaten van zoekmethode 2 bij het gebruik van ‘ongewogen’ profielen, kunnen we vaststellen dat het wegen van delicten naar tijd weinig verandert aan het zoekresultaat. Het selecteren van de 3 beste delicten in het delictprofiel leidt zelfs tot een verslechtering van het zoekresultaat. Alleen de tweede optie, het selecteren van delicten uit het delictprofiel die gepleegd zijn binnen 1 jaar van het zoekdelict, leidt tot een licht beter zoekresultaat.

Tabel 7 Matchen van individuele delicten met delictprofielen: resultaten van zoekmethode 2 bij alternatieve samenstellingen van de delictprofielen (volledige steekproef N=500)

<i>Kleinste cluster waarin werkelijke dader is aangetroffen</i>	<i>Zoekmethode 2 A: tijdweging</i>		<i>Zoekmethode 2 B: delicten < 1 jaar</i>		<i>Zoekmethode 2 C: beste 3 delicten</i>	
	%	% cum	%	% cum	%	% cum
1: Cluster 1-10	5,6	5,6	6,4	6,4	1,4	1,4
2: Cluster 11-25	2,6	8,2	3,4	9,8	1,0	2,4
3: Cluster 26-50	1,6	9,8	1,8	11,6	1,4	3,8
4: Cluster 51-100	2,4	12,2	2,4	14,0	1,0	4,8
5: Cluster 101-250	2,8	15,0	5,2	19,2	3,2	8,0
6: Cluster 251-500	4,6	19,6	3,2	22,4	2,0	10,0
7: Dader niet in vorige clusters	80,4	100,0	77,6	100,0	90,0	100,0
TOTAAL N=500	100,0		100,0		100,0	

Wanneer we de alternatieve matchwaarden van zoekmethode 2 als uitgangspunt nemen, kunnen we op basis hiervan opnieuw de zoekresultaten voor zoekmethode 3 berekenen. De resultaten hiervan zijn weergegeven in tabel 8. We zien in deze tabel dat de resultaten in alle gevallen duidelijk onder die van zoekmethode 2 liggen (in tabel 7). Dit ligt ook voor de hand omdat hiervoor gebleken is dat zoekmethode 3 bij het matchen van geclusterde gegevens niet in staat is de zoekresultaten van zoekmethode 2 te verbeteren.

Tabel 8 Matchen van individuele delicten met delictprofielen: resultaten van zoekmethode 3 bij alternatieve samenstellingen van de delictprofielen (volledige steekproef N=500)

	<i>Zoekmethode 3</i> <i>Voorwaarde: A of B</i> <i>A: tijdweging</i>		<i>Zoekmethode 3</i> <i>Voorwaarde: A of B</i> <i>B: delicten < 1 jaar</i>		<i>Zoekmethode 3</i> <i>Voorwaarde: A of B</i> <i>C: beste 3 delicten</i>	
	%	% cum	%	% cum	%	% cum
<i>Kleinste cluster waarin werkelijke dader is aangetroffen</i>						
1: Cluster 1-10	4,2	4,2	4,2	4,2	,2	,2
2: Cluster 11-25	2,6	6,8	2,4	6,6	,4	,6
3: Cluster 26-50	,6	7,4	1,8	8,4		
4: Cluster 51-100	1,0	8,4	1,2	9,6		
5: Cluster 101-250	1,8	10,2	3,6	13,2	,2	,8
6: Cluster 251-500	3,0	13,2	4,4	17,6		
7: Dader niet in vorige clusters	86,8	100,0	82,4	100,0	99,2	100,0
TOTAAL N=500	100,0		100,0		100,0	

Op basis van deze analyses kunnen we concluderen dat variant A van zoekmethode 1 (gemiddelde clustergrootte) en variant B van zoekmethode 2 (selectie van delicten < 1 jaar van het zoekdelict) de beste zoekresultaten opleveren. Om de resultaten hiervan te kunnen vergelijken met de resultaten van het matchen van individuele delicten, presenteren we hier nogmaals de zoekresultaten, maar nu weer gebaseerd op een selectie van delicten uit de steekproef waarvan de dader (nogmaals) voorkomt in het HKS.²⁹ We zien in tabel 9 dat de resultaten van beide zoekmethoden elkaar niet veel ontlopen. In beide gevallen slagen de zoekmethoden erin om in 42 tot 44% van de gevallen de dader bij de beste 500 in het zoekresultaat te matchen. In 10 tot 12% van de gevallen kan de dader bij de beste 10 in het zoekresultaat worden aangetroffen.

Wanneer we deze resultaten vergelijken met de resultaten die zijn behaald bij het matchen van individuele delicten, kunnen we concluderen dat het clusteren van gegevens leidt tot een lichte verbetering in het zoekresultaat indien we dit vergelijken met de uitkomsten van de huidige zoekmethode bij het matchen van individuele delicten (zie tabel 4). Echter, wanneer we de resultaten vergelijken met de beste presterende zoekmethode, is géén sprake van een verbetering in het resultaat. Integendeel, er is dan sprake van een verslechtering. Zoekmethode 3, toegepast bij het matchen van individuele delicten, slaagt erin ruim 25% van de daders uit de steekproef bij de beste 10 in het zoekresultaat te matchen. Hier gebeurt dat slechts in 10 tot 12% van de gevallen. Ook het cumulatieve resultaat is beter: bij zoekmethode 3 kan 47,7% van de daders (als ze in het HKS vóórkomen) bij de eerste 500 in het zoekresultaat worden gematcht. Hier gebeurt dat in 42 respectievelijk 44% van de gevallen. We mogen derhalve concluderen dat het matchen van geclusterde gegevens weliswaar bevredigende resultaten oplevert, maar niet leidt tot een verbetering van de zoekresultaten die zijn verkregen bij het matchen van individuele delicten.

²⁹ Deze steekproef is om technische redenen kleiner dan bij de delictmatching.

Tabel 9 Matchen van individuele delicten met delictprofielen: resultaten van de beste 2 zoekmethoden (selectie van daders die (nogmaals) vóórkomen in het HKS; steekproef N=268)

<i>Kleinste cluster waarin werkelijke dader is aangetroffen</i>	<i>Zoekmethode 1</i> <i>A: Gemiddelde clustergrootte</i>		<i>Zoekmethode 2</i> <i>B: delicten < 1 jaar</i>	
	%	% cumulatief	%	% cumulatief
1: Cluster 1-10	10,4	10,4	11,9	11,9
2: Cluster 11-25	8,2	18,7	6,3	18,3
3: Cluster 26-50	9,0	27,6	3,4	21,6
4: Cluster 51-100	5,6	33,2	4,5	26,1
5: Cluster 101-250	8,2	41,4	9,7	35,8
6: Cluster 251-500	2,6	44,0	6,0	41,8
7: Dader niet in vorige clusters	56,0	100,0	58,2	100,0
TOTAAL N=268	100,0		100,0	

Correlatieanalyse

We beperken de analyses hier tot de resultaten van de zoekmethoden 1 en 2 zoals gepresenteerd in tabel 9. We hebben in dit geval onderzocht op welke kenmerken delictprofielen die goed matchen afwijken van delictprofielen die minder goed of slecht matchen. We hebben hieromtrent de volgende waarnemingen gedaan (zm=zoekmethode):

- hoe groter het aantal delicten dat van een bepaalde dader vóórkomt in het HKS, hoe beter delicten van deze dader kunnen worden gematcht
 - o ($r=-.36^{**}$ bij zm1, $r=-.34^{**}$ bij zm2)
- hoe meer gegevens bekend zijn van delicten in het HKS des te beter kan de dader van deze delicten worden gematcht
 - o ($r=-.14^{**}$ bij zm1, $r=-.24^{**}$ bij zm2)
- daders die *naar verhouding* vaak geweldsdelicten of gewelddadige vermogensdelicten of verkeersdelicten of opiumwetdelicten hebben gepleegd, laten zich slechter dan gemiddeld matchen
 - o (geweldsdelicten $r=.16^*$ bij zm1, $r=.17^{**}$ bij zm2)
 - o (gewelddadige vermogensdelicten: $r=.13^*$ bij zm2)
 - o (verkeersdelicten: $r=.18^*$ bij zm1, $r=.18^{**}$ bij zm2)
 - o (opiumwetdelicten: $r=.16^{**}$ bij zm2)
- daders die naar verhouding vaak vermogensdelicten of delicten tegen de openbare orde/het openbaar gezag hebben gepleegd laten zich beter dan gemiddeld matchen
 - o (vermogensdelicten: $r=-.12^*$ bij zm1, $r=-.16^{**}$ bij zm2)
 - o (delicten tegen de op. orde/op. gezag: $r=-.16^*$ bij zm1, $r=-.17^{**}$ bij zm2)
- daders die naar verhouding vaak wapens of geweld gebruiken laten zich bij zoekmethode 2 slechter dan gemiddeld matchen
 - o (MO gebruik van wapens $r=.13$ bij zm2)
 - o (MO gebruik van geweld $r=.13$ bij zm2)

- daders met relatief veel delicten die een waarde hebben op de door ons gecreëerde MO-gegevensvelden, laten zich beter dan gemiddeld matchen. Dit geldt in sterkere mate voor zoekmethode 2.
 - o (gebruik van speciale hulpmiddelen: $r = -.24^{**}$ bij zm2)
 - o (gebruik van misleiding: $r = -.13^*$ bij zm1, $r = -.15^*$ bij zm2)
 - o (betreden van –afgesloten- privé-plaatsen: $r = -.15^*$ bij zm1, $r = -.26^{**}$ bij zm2)
 - o (geld als buit: $r = -.13^*$ bij zm1, $r = -.23^{**}$ bij zm2)
 - o (elektrische apparatuur als buit: $r = -.23^{**}$ bij zm1, $r = -.23^{**}$ bij zm2)
 - o (kunst als buit: $r = -.15^*$ bij zm2)

Toelichting op de gevonden verbanden:

Dat daders beter kunnen worden gematcht wanneer ze a) vaker vóórkomen in het HKS en er b) meer gegevens bekend zijn over hun delicten, is geruststellend. Dit gegeven vormt immers het uitgangspunt van het matchen. Ook hier zien we weer dat bepaalde delicttypen beter matchen dan andere. Het is onduidelijk waarom dit het geval is. Mogelijk worden deze correlaties veroorzaakt door andere factoren die wij hier niet in beeld hebben. Verder zien we ook hier weer dat daders van delicten met veel waarden op de door ons gecreëerde MO-categorieën, vaker dan gemiddeld een goed matchresultaat behalen. Dit kan deels worden toegeschreven aan het feit dat delicten die waarden hebben op deze velden uiteraard ook waarden hebben op de onderliggende delictkenmerkcodes, waardoor er een bovengemiddeld aantal gegevens van deze delicten beschikbaar is. Daarnaast zien we echter ook een onafhankelijk toegevoegd effect van deze waarden op het resultaat bij zoekmethode 2 (en 3).

De resultaten van de zoekmethoden 1 en 2 in tabel 9 correleren $r = .66^{**}$. Dit betekent dat delictprofielen van daders die hoog scoren op methode 1 dit in veel gevallen ook zullen doen op zoekmethode 2 (en omgekeerd), maar het verband is niet heel sterk. We hebben derhalve onderzocht of delictprofielen die goed matchen op zoekmethode 1 en slecht op 2 (en omgekeerd) in bepaalde opzichten afwijken. De belangrijkste resultaten hiervan zijn dat profielen die hoog scoren op zoekmethode 2 en laag op zoekmethode 1 delicten bevatten waarvan gemiddeld meer gegevens bekend zijn en die (dus) ook vaker waarden hebben op de door ons aangemaakte MO-velden.

3. Groep 3: matchen van individuele signalementen

De toetsing is hier gebaseerd op het matchen van signalementen OD met signalementen BD op basis van een steekproef van signalementen OD waarvan de dader in het HKS bekend is en ook een signalement aan de personenkant heeft (signalement BD). Dit zijn 410 signalementen OD in de periode 1999-2001.³⁰

Alvorens over te gaan tot de toetsing van de verschillende zoekmethoden, hebben we eerst onderzocht in hoeverre de signalementen OD uit de steekproef overeenkomen met de hieraan gekoppelde signalementen BD. Dit geeft een indicatie van de kwaliteit van de betreffende signalementen OD. Om überhaupt tot resultaten bij het matchen te kunnen komen is een minimale overeenkomst tussen beide vereist. Deze hebben we eerst onderzocht. In tabel 10 zijn hiervan de resultaten weergegeven. In deze tabel zien we in de eerste kolom de signalementen OD uit de steekproef uitgesplitst naar het aantal kenmerken waarop ze waarden hebben. Deze lopen van 1 tot 12. Merk op dat het aantal waarden hier aanzienlijk geringer is

³⁰ Gebaseerd op aangiften waaraan hooguit 1 signalement OD is gekoppeld en hooguit 1 signalement BD.

dan bij de delicten, die gemiddeld op 25 kenmerken waarden hebben. In de laatste kolom zien we het aantal signalementen OD met een specifiek aantal waarden: 1 signalement heeft waarden op 1 kenmerk, 13 signalementen hebben waarden op 2 kenmerken, 37 op 3 kenmerken, enzovoort. Gemiddeld hebben de signalementen OD op iets minder dan 6 kenmerken waarden. We hebben per groep van signalementen OD (de rijen in tabel 10) in kaart gebracht hoeveel van deze waarden ook aangetroffen kunnen worden in de ‘bijbehorende’ signalementen BD. Deze hebben we als rijpercentages weergegeven. Dus als we in de linkerkolom signalementen OD selecteren die op bijvoorbeeld 3 kenmerken waarden hebben, zien we in 5,4% van de gevallen geen enkele waarde terug in het bijbehorende signalement BD, in 16,2% van de gevallen zien we 1 van de 3 waarden terug, in 56,8% van de gevallen zien we 2 van de 3 waarden terug en in 21,6% van de gevallen zien we alle waarden uit het signalement OD terug in het signalement BD. De licht gearceerde vlakken in de tabel geven aan waar de meeste waarnemingen per rij zich bevinden (sommigen zijn gebaseerd op hele kleine aantallen). Uit het patroon kunnen we enerzijds afleiden dat het aantal waarden waarop de signalementen BD overeenkomen met de signalementen OD door de bank genomen toeneemt naarmate de signalementen OD op meer kenmerken waarden hebben. Anderzijds kunnen we ook zien dat de relatieve kwaliteit van de match afneemt naarmate het aantal waarden op het signalement OD toeneemt; de afstand tussen de klasse met de meeste waarnemingen en de klasse waarin alle waarden matchen wordt steeds groter. Het maximum aantal waarden waarop enig signalement BD overeenkomt met het bijbehorende signalement OD bedraagt 8 (om die reden zijn de kolommen 9 tot en met 12 hier niet weergegeven). In de onderste rij van de tabel zien we het aantal kenmerken waarop signalementen BD overeenkomen met signalementen OD, ongeacht het aantal waarden dat van een signalement OD beschikbaar is. We zien hier dat in ruim een kwart van de gevallen (25,6%) de signalementen BD op 4 kenmerken overeenkomen met de bijbehorende signalementen OD. Het aantal signalementen BD dat op geen enkel kenmerk overeenkomt met het signalement OD is heel beperkt, namelijk 1,2%. Gemiddeld matchen de signalementen BD op 3,5 kenmerken met het bijbehorende signalement OD (dat gemiddeld op 5,9 kenmerken waarden heeft). Onze conclusie is derhalve dat de kwantiteit en kwaliteit van de signalementen OD weliswaar niet heel erg goed is, maar vooralsnog voldoende lijkt om een matchprocedure uit te voeren.

Tabel 10 Mate waarin signalen OD uit de steekproef overeenkomen met signalen BD van dezelfde dader

<i>Aantal kenmerken waarop signalement OD waarden heeft</i>	<i>Aantal kenmerken van signalement BD dat overeenkomt met 'bijbehorend' signalement OD (rij %)</i>								Rij % (N)	
	0	1	2	3	4	5	6	7		8
1 kenmerken	100,0%									3,2% (1)
2 kenmerken		53,8%	46,2%							3,2% (13)
3 kenmerken	5,4%	16,2%	56,8%	21,6%						9,0% (37)
4 kenmerken		12,8%	31,9%	38,3%	17,0%					11,5% (47)
5 kenmerken		7,9%	14,5%	30,3%	36,8%	10,5%				18,5% (76)
6 kenmerken		7,2%	13,3%	24,1%	32,5%	14,5%	8,4%			20,2% (83)
7 kenmerken		3,6%	10,8%	15,7%	31,3%	20,5%	14,5%	3,6%		20,2% (83)
8 kenmerken	2,7%		5,4%	13,5%	21,6%	32,4%	18,9%	5,4%		9,0% (37)
9 kenmerken	4,5%		9,1%		13,6%	36,4%	27,3%	4,5%	4,5%	5,4% (22)
10 kenmerken			25,0%	25,0%	50,0%					1,0% (4)
11 kenmerken			20,0%	20,0%	20,0%	40,0%				1,2% (5)
12 kenmerken					100,0%					(2)
Kolom %	1,2%	8,3%	19,3%	21,7%	25,6%	13,9%	8,3%	1,5%	2,2%	100%
(N)	(5)	(34)	(79)	(89)	(105)	(57)	(33)	(6)	(2)	(410)

In tabel 11 zijn de matchresultaten van zoekmethode 1, 2 en 3 (beste variant) weergegeven. Bij zoekmethode 1 is het signalement BD van de werkelijke dader in 21,7% van de gevallen aangetroffen bij de beste 500 signalementen in het zoekresultaat. Bij zoekmethode 2 is dit het geval bij 24,6% en bij zoekmethode 3 bij 27,8%. We kunnen derhalve stellen dat zoekmethode 3 in dit opzicht een marginaal beter zoekresultaat oplevert dan de huidige zoekmethode 1. Als we kijken naar de verdeling van de matchresultaten over de klassen, zien we bij alle zoekresultaten hetzelfde beeld: naar verhouding weinig daders zijn aangetroffen in de kleinste zoekclusters. Voorzover de werkelijke daders zijn aangetroffen bij de beste 500 signalementen in het zoekresultaat bevinden ze zich voornamelijk in de grotere en dus minder waardevolle clusters. We mogen daarom stellen dat zoekmethode 3 weliswaar iets betere resultaten laat zien dan zoekmethode 1, maar over de gehele linie is het resultaat mager.

De opzet van de toetsing is hier zodanig dat alle signalementen (OD) in de steekproef in beginsel gematcht kunnen worden met het signalement (BD) van de betreffende dader. Dit betekent dat we het resultaat in tabel 11 kunnen vergelijken met de uitkomst van tabel 4 om te zien hoe de zoekresultaten van het matchen van signalementen zich verhouden tot de resultaten die zijn verkregen bij het matchen van delicten. Deze vergelijking laat zien, dat het matchen van delicten aanzienlijk betere resultaten oplevert dan het matchen van signalementen. Als we ons beperken tot de beste zoekmethoden, zien we dat bij het matchen van delicten de dader in 47,8% van de gevallen werd aangetroffen bij de beste 500 delicten in het zoekresultaat, hier is dat het geval bij 27,8%. Bovendien is de verdeling van de zoekresultaten bij het matchen van delicten aanzienlijk gunstiger dan bij het matchen van signalementen. Bij het matchen van delicten troffen we in 25,4% van de gevallen de dader aan bij de beste 10 in het zoekresultaat, bij het matchen van signalementen bedraagt dit percentage slechts 0,5.

Tabel 11 Matchen van individuele signalementen: resultaten van zoekmethoden 1, 2 en 3 (volledige steekproef N=410)

<i>Kleinste cluster waarin werkelijke dader is aangetroffen</i>	<i>Zoekmethode 1</i>		<i>Zoekmethode 2</i>		<i>Zoekmethode 3 Voorw.: A of B</i>	
	<i>%</i>	<i>% cum</i>	<i>%</i>	<i>% cum</i>	<i>%</i>	<i>% cum</i>
1: Cluster 1-10	,2	,2	,5	,5	,5	,5
2: Cluster 11-25	1,0	1,2	,5	1,0	,5	1,0
3: Cluster 26-50	2,2	3,4	2,4	3,4	3,4	4,4
4: Cluster 51-100	1,7	5,1	2,9	6,3	2,7	7,1
5: Cluster 101-250	6,1	11,2	7,3	13,7	9,0	16,1
6: Cluster 251-500	10,5	21,7	11,0	24,6	11,7	27,8
7: Dader niet in vorige clusters	78,3	100,0	75,4	100,0	72,2	100,0
TOTAAL N=410	100,0		100,0		100,0	

De relatief magere matchresultaten worden hier niet veroorzaakt door het feit dat personen zich moeilijker laten matchen dan delicten of doordat relatief weinig personen signalementen (BD) in het HKS hebben. We hebben onze toetsing immers beperkt tot dié signalementen OD waarvan we de ‘eigenaar’ op voorhand kenden en ook zijn signalement dat zich aan de personenkant van het HKS bevindt. Waarom de resultaten slechter zijn dan bij het matchen van delicten kunnen we niet met zekerheid zeggen, maar vermoedelijk spelen de volgende factoren hierbij een rol:

1. Het aantal waarden dat beschikbaar is om mee te zoeken is bij signalementen OD nogal beperkt, aanzienlijk beperkter dan het aantal waarden dat gemiddeld beschikbaar is bij het matchen van delicten;
2. De kwaliteit van de signalementen OD is matig, wellicht te matig om zinvol mee te kunnen zoeken;
3. Bij het matchen van delicten is het mogelijk om meer dan één delict van een bepaalde dader te matchen, dit geeft de mogelijkheid om alleen het beste matchresultaat van de betreffende dader mee te tellen in het zoekresultaat. Deze optie is niet beschikbaar bij het matchen van signalementen.

Correlatieanalyse

Ten aanzien van de verschillen tussen signalementen OD die goed matchen en signalementen OD die minder goed of slecht matchen, hebben we de volgende waarnemingen gedaan (we zullen ze hier op een rij zetten en daarna bespreken, zm=zoekmethode):

- Hoe groter het aantal gegevens dat beschikbaar is van een signalement OD, des te beter is het matchresultaat
 - o ($r=-.16^{**}$ bij zm1, $r=-.17^{**}$ bij zm3)
- Signalementen van vrouwen laten zich beter matchen dan signalementen van mannen
 - o ($r=-.16^{**}$ bij zm1, $r=-.13^*$ bij zm3)
- Signalementen waarvan trefwoorden beschikbaar zijn die gekoppeld zijn aan signalementcodes, laten zich marginaal beter dan gemiddeld matchen
 - o ($r=-.09$ bij zm1, $r=-.10^{**}$ bij zm3)

Toelichting op de gevonden verbanden:

Dat het aantal waarden dat beschikbaar is om mee te zoeken van invloed is op het matchresultaat bij zoekmethode 3 zijn we eerder tegengekomen en het ligt ook voor de hand, omdat deze zoekmethode gevoelig is voor de hoeveelheid gegevens die gebruikt wordt in het zoekproces. Het is opvallend dat ook het resultaat bij zoekmethode 1 hierdoor wordt beïnvloed. Dit was hiervoor niet het geval. Dit zou er op kunnen duiden dat het aantal kenmerken dat hier beschikbaar is om mee te zoeken zich rond het minimum beweegt dat voor zoekmethode 1 nodig is om tot (goede) zoekresultaten te komen.

De best gematchte signalementen (in de klassen 1-3) beschikken over meer waarden om mee te zoeken. Gemiddeld ligt het aantal beschikbare waarden bij deze signalementen ca. 2.5 hoger dan bij signalementen die in klasse 7 gematcht zijn.

Signalementen van vrouwen worden iets beter gematcht dan signalementen van mannen en hetzelfde geldt voor signalementen waarvan trefwoorden beschikbaar zijn. Het gaat hierbij echter niet om hele grote verschillen.

De resultaten van zoekmethode 1 en 3 in tabel 11 correleren $r=.6^{**}$. Dit betekent dat signalementen die hoog scoren op de ene zoekmethode vaak ook hoog zullen scoren op de andere zoekmethode. De samenhang is echter niet zo sterk. Dit betekent dat er genoeg signalementen zijn die het goed doen bij zoekmethode 1 en niet bij zoekmethode 3 (en omgekeerd). We hebben gekeken naar eventuele verschillen tussen deze signalementen. Het enige significante verschil dat we hierbij hebben aangetroffen is dat signalementen die het goed doen bij zoekmethode 1 en niet goed doen bij zoekmethode 3 ($N=22$) vaker dan gemiddeld signalementen van vrouwen zijn. Dit bevestigt een vermoeden dat hiervoor al geuit werd dat zoekmethode 1, in tegenstelling tot zoekmethode 3, het wellicht iets beter doet bij

gegevens waarin unieke waarden voorkomen. Vrouwen vormen immers een grote minderheid in het HKS.

4. Groep 4: matchen van delict/signalementcombinaties

De toetsing vindt plaats door 400 delicten waarvan zowel delict- als signalementgegevens (OD) bekend zijn te matchen met delicten waarvan ook signalement-BD gegevens bekend zijn (N=23.950 delict/signalementcombinaties). De 400 delicten uit de steekproef komen overeen met de delicten waarvan in de vorige paragraaf gebruikt is gemaakt voor het matchen van signalementen. We kunnen aldus de resultaten hier vergelijken met die uit de vorige paragraaf.

Om de resultaten ook te kunnen vergelijken met het matchen van enkel delictgegevens hebben we bovengenoemde steekproef van 400 ook gebruikt om nogmaals een delictmatching á la groep 1 uit te voeren, maar nu op het hier genoemde bestand van delicten waaraan signalementen van bekende verdachten zijn gekoppeld. Alvorens de resultaten van het matchen met gecombineerde delict/signalementgegevens te bespreken, geven we hier eerst de resultaten weer van de (afzonderlijke) delictmatching. Deze zijn te vinden in tabel 12. Deze tabel toont de resultaten van de zoekmethoden 1, 2 en 3 (de beste variant). De zoekresultaten zijn enkel gebaseerd op de steekproefdelicten waarvan bekend is dat de dader (nogmaals) voorkomt in het HKS. Dit is gedaan om een betere vergelijking met de andere analyses mogelijk te maken. Het gaat om 375 van de 400 steekproefdelicten.

Tabel 12 Matchen van individuele delicten: resultaten van zoekmethoden 1, 2 en 3 (selectie van daders die –nogmaals- vóórkomen in het HKS; steekproef N=375)

	Zoekmethode 1		Zoekmethode 2		Zoekmethode 3 Vw: A of B of C	
	%	% cum	%	% cum	%	% cum
<i>Kleinste cluster waarin werkelijke dader is aangetroffen</i>						
1: Cluster 1-10	25,6	25,6	29,3	29,3	37,1	37,1
2: Cluster 11-25	9,9	35,5	9,3	38,7	8,8	45,9
3: Cluster 26-50	7,2	42,7	7,7	46,4	6,4	52,3
4: Cluster 51-100	5,3	48,0	8,0	54,4	5,1	57,3
5: Cluster 101-250	5,3	53,3	10,1	64,5	8,8	66,1
6: Cluster 251-500	1,1	54,4	5,9	70,4	5,9	72,0
7: Dader niet in vorige clusters	45,6	100,0	29,6	100,0	28,0	100,0
TOTAAL N=375	100,0		100,0		100,0	

We zien in deze tabel dat alle zoekmethoden goede resultaten laten zien. Bij zoekmethode 1 wordt in meer dan de helft van de gevallen (54,4%) de werkelijke dader aangetroffen bij de beste 500 delicten in het zoekresultaat (als de dader dus in het HKS vóórkomt!). Bij zoekmethode 2 ligt dit percentage zelfs op 70% en bij de beste variant van zoekmethode 3 op 72%. Bij zoekmethode 1 wordt de werkelijke dader in 25,6% van de gevallen aangetroffen bij de beste 10 delicten in het zoekresultaat, bij zoekmethoden 2 en 3 is dit respectievelijk 29,3 en 37,1%. We kunnen derhalve concluderen dat de analyse op basis van deze steekproef de resultaten bevestigen die we eerder vonden bij de analyses in groep 1: met name zoekmethode 3 levert een substantiële verbetering op ten opzichte van zoekmethode 1.

Opvallend is verder dat de resultaten van alle zoekmethoden hier significant beter zijn dan bij de analyses op basis van de steekproef in groep 1 (vgl. de resultaten in tabel 4). De vraag is waardoor dit verschil wordt veroorzaakt. Een mogelijke verklaring kan worden

gevonden in de verschillende samenstelling van het toetsbestand. Het toetsbestand bij de analyses in groep 1 bevat bijna 2,5 keer zoveel delicten als het toetsbestand hier. Daardoor is de toetsing in groep 1 wellicht kritischer. Verder zijn er enkele opvallende verschillen tussen de gebruikte steekproeven. De delicten in de hier gebruikte steekproef hebben gemiddeld op ruim 29 kenmerken waarden. Bij de steekproef in groep 1 ligt het gemiddelde op 'slechts' 25. Dit kan een beter matchresultaat hooguit ten dele verklaren, omdat zoekmethode 1 niet zo gevoelig is voor de hoeveelheid beschikbare waarden. Verder bevat de hier gebruikte steekproef veel meer gewelddadige vermogensdelicten (wetsartikelen 312/317): 38 tegen 4 procent in de steekproef bij groep 1. De hier gebruikte steekproef wijkt daarmee aanzienlijk af van de populatiegegevens (aangiftengegevens) in de regio Haaglanden, waar hooguit enkele procenten van de aangegeven delicten gewelddadige vermogensdelicten betreffen. We weten echter niet of deze factor van invloed is op het zoekresultaat en zo ja, in welke mate. Een laatste verschil tussen de hier gebruikte steekproef en de steekproef van groep 1 is dat de delicten in bovenstaande steekproef gespreid zijn over de gehele periode waarop de matching betrekking heeft (1-1-1999 tot en met 31-12-2001). De delicten in de steekproef van groep 1 zijn weliswaar via een random procedure gekozen maar hebben alle plaatsgevonden tussen januari en augustus 1999. Het is niet uitgesloten dat dit verschil in spreiding van invloed is op de gevonden matchresultaten.

Ten slotte hebben we nog gekeken naar mogelijke verschillen tussen delicten met goede matchresultaten en delicten met minder goede en slechte matchresultaten. De resultaten van deze analyse wijken niet sterk van die bij groep 1. De belangrijkste bevindingen:

- vermogensdelicten en gewelddadige vermogensdelicten (SR 312/317) matchen beter dan gemiddeld, terwijl geweldsdelicten en delicten tegen de openbare orde en het openbaar gezag slechter dan gemiddeld matchen;
- delicten die waarden hebben op de door ons gecreëerde MO-velden matchen beter dan gemiddeld. Dit geldt vooral bij zoekmethode 3.

Dit laatste resultaat bevestigt dat het gebruik van deze samenvattende MO-velden een meerwaarde heeft boven het gebruik van de afzonderlijke gegevensvelden in het HKS.

Nu we inzicht hebben in de resultaten die zijn verkregen op grond van het afzonderlijk matchen van delict- respectievelijk signalementgegevens kunnen we de uitkomsten presenteren van het matchen van gecombineerde delict- en signalementgegevens. In tabel 13 zijn deze weergegeven voor de zoekmethoden 1, 2 en 3 (de beste variant). Ook hier zijn de zoekresultaten weer gebaseerd op een selectie van steekproefdelicten waarvan bekend is dat de dader (nogmaals) voorkomt in het HKS.

We zien in tabel 13 dat de resultaten van zoekmethode 1 opvallend slechter zijn dan die van de alternatieve zoekmethoden: bij deze methode wordt in ruim 6% van de gevallen de dader aangetroffen bij de beste 10 in het zoekresultaat, bij de andere methoden gebeurt dat in ruim 35 en 40% van de gevallen. Ook de cumulatieve zoekresultaten zijn veel slechter: in nog geen 16% van de gevallen wordt de dader aangetroffen bij de eerste 500 in het zoekresultaat, terwijl dit bij de andere methoden bijna 73 en 71% is. Wanneer we deze resultaten vergelijken met het matchen van enkel delictgegevens (tabel 12) of signalementgegevens (tabel 11) kunnen we de volgende zaken constateren:

- het matchen van delicten levert bij alle zoekmethoden aanzienlijk betere resultaten op dan het matchen van signalementen;
- het toevoegen van signalementgegevens aan delictgegevens zorgt voor :
 - o een lichte *verbetering* in het resultaat van zoekmethode 2;
 - o een min of meer *ongewijzigd* zoekresultaat van zoekmethode 3;
 - o een aanzienlijk *slechter* zoekresultaat van zoekmethode 1.

Tabel 13 Matchen van delict/signalementcombinaties: resultaten van zoekmethoden 1, 2 en 3 (selectie van daders die –nogmaals- vóórkomen in het HKS; steekproef N=377)

	Zoekmethode 1		Zoekmethode 2		Zoekmethode 3 Vw: A of B of C	
	%	% cum	%	% cum	%	% cum
<i>Kleinste cluster waarin werkelijke dader is aangetroffen</i>						
1: Cluster 1-10	6,4	6,4	35,5	35,5	40,1	40,1
2: Cluster 11-25	4,0	10,3	9,8	45,4	8,2	48,3
3: Cluster 26-50	2,1	12,5	5,0	50,4	4,0	52,3
4: Cluster 51-100	1,1	13,5	6,9	57,3	7,4	59,7
5: Cluster 101-250	2,1	15,6	9,5	66,8	7,4	67,1
6: Cluster 251-500	,3	15,9	6,1	72,9	3,7	70,8
7: Dader niet in vorige clusters	84,1	100,0	27,1	100,0	29,2	100,0
TOTAAL N=377	100,0		100,0		100,0	

De verbetering van zoekmethode 2 laat zich verklaren op grond van het feit dat bij deze methode het toevoegen van extra gegevens alleen maar kan leiden tot verfijndere en daardoor betere zoekresultaten. Ook als de toegevoegde gegevens van zichzelf weinig ‘matchpotentie’ hebben, zullen ze het zoekresultaat niet (kunnen) negatief beïnvloeden. Dit is wel het geval bij zoekmethode 1: daar leidt het toevoegen van gegevens met een ‘geringe matchpotentie’ tot aanzienlijk slechtere resultaten. De waarden op de signalementkenmerken zijn vaak unieker dan de waarden op de delictkenmerken, waardoor ze kleinere clusters vormen in de doelgegevens. De methode selecteert derhalve vooral deze kenmerken om te matchen, terwijl ze zich juist slecht laten matchen, zoals we gezien hebben in de vorige paragraaf. De resultaten van zoekmethode 1 bij het matchen van gecombineerde delict/signalementgegevens zijn daardoor nog slechter dan bij het matchen van enkel signalementgegevens. Zoekmethode 3 is de middenweg van de zoekmethoden 1 en 3 en gedraagt zich derhalve ook als zodanig: het grotere aantal waarden om mee te zoeken verbetert hier het zoekresultaat, maar dit wordt deels weer teniet gedaan door het feit dat de nadere voorwaarden vaker betrekking zullen hebben op kenmerken die slecht matchen (namelijk signalementkenmerken).

Correlatieanalyse

We beperken de analyses hier tot de resultaten van de zoekmethoden 1 en 3. Ten aanzien van de verschillen tussen delict/signalementcombinaties die goed respectievelijk minder goed of slecht matchen, hebben we de volgende waarnemingen gedaan (zm=zoekmethode):

- hoe groter het aantal daders van een delict, des te beter is het matchresultaat
 - o ($r=-.20^{**}$ bij zm1 en $r=-.19^{**}$ bij zm3)
- hoe groter het aantal gegevens dat beschikbaar is van het delict en het signalement OD, des te beter is het matchresultaat bij zoekmethode 3
 - o ($r=-.19^{**}$ bij zm3)
- gewelddadige vermogensdelicten laten zich beter dan gemiddeld matchen
 - o ($r=-.10^{*}$ bij zm1 en $r=-.15^{**}$ bij zm3)
- delicten tegen de openbare orde/het openbaar gezag laten zich slechter dan gemiddeld matchen bij zoekmethode 3
 - o ($r=.20^{**}$ bij zm3)

- delicten die waarden hebben op een aantal door ons aangemaakte MO-gegevensvelden scoren beter dan gemiddeld bij zoekmethode 3
 - o ($r = -.17^{**}$ bij gebruik van bijzondere hulpmiddelen)
 - o ($r = -.11^*$ bij gebruik van misleiding)
 - o ($r = -.12^*$ bij betreding van afgesloten- privé-plaatsen)
 - o ($r = -.18^{**}$ bij geld als buit)
- signalementen (OD) die géén waarde hebben op geslacht laten zich bij zoekmethode 1 beter matchen dan signalementen die op dit kenmerk wel een waarde hebben
 - o ($r = .22^{**}$ bij zm1)
- signalementen (OD) van vrouwen laten zich bij zoekmethode 1 beter matchen dan signalementen van mannen
 - o ($r = .31^{**}$ bij zm1)

Toelichting op de gevonden verbanden:

Deze bevindingen komen in grote lijnen overeen met wat we hiervoor al vonden. Heel duidelijk is hier de toegevoegde waarde van de extra gecreëerde MO-gegevensvelden voor zoekmethode 3. Opvallend is verder dat signalementen die geen waarde hebben op geslacht bij zoekmethode 1 beter matchen dan signalementen die hierop wel een waarde hebben. Misschien heeft dit te maken met het feit dat de afwezigheid van waarden op signalementkenmerken de kans vergroot dat de kleinste clusters in de doelgegevens worden gevormd door delictkenmerken in plaats van signalementkenmerken (en deze leveren betere matchresultaten). Dat signalementen van vrouwen zich beter dan gemiddeld laten matchen hebben we hiervoor ook al gezien: Het verschil is omvangrijk: het gemiddelde matchresultaat van delict/signalementcombinaties van vrouwen is klasse 4.3, dat van mannen is gemiddeld klasse 6.6. Dus ruim 2 klassen hoger.

De resultaten van de zoekmethoden 1 en 3. correleren $r = .26^{**}$. Dit betekent weliswaar dat delict/signalementcombinaties die hoog scoren op methode 1 dit in een aantal gevallen ook zullen doen op zoekmethode 3 (en omgekeerd), maar het verband is matig, zeker niet wanneer we het vergelijken met de voorgaande analyses. We hebben derhalve nog een blik geworpen op de delict/signalementcombinaties die wel goed matchen bij zoekmethode 3 maar niet bij zoekmethode 1 (omgekeerd komt hier niet voor). Combinaties die hoog scoren op zoekmethode 3 en laag op 1 beschikken gemiddeld over meer zoekwaarden, het gaat vaker dan gemiddeld over vermogensdelicten en minder vaak dan gemiddeld over geweldsdelicten of delicten tegen de openbare orde/het openbaar gezag. Verder hebben ze ook vaker dan gemiddeld waarden op de door ons aangemaakte MO-gegevensvelden. Deze resultaten bevestigen wat we ook bij de andere analyses al vonden.

5. Nadere beschouwing over zoekvragen voor tactisch gebruik

De toetsing in de voorgaande paragrafen is gebaseerd op zoekvragen met een operationeel doel. Het operationele komt onder meer naar voren in de gebruikte toetsbestanden (met alleen delicten en/of signalementen van bekende verdachten) en in de wijze waarop de resultaten van de verschillende zoekmethoden zijn beoordeeld (door de rangordepositie van de werkelijke dader in het zoekresultaat te bepalen).

Hoe zit het nu met de toepassing van de eerder besproken zoekvragen voor tactisch gebruik? Allereerst kunnen we hierover opmerken dat hiervoor dezelfde zoekprocedures kunnen worden gebruikt als die hiervoor getoetst zijn. De verschillen tussen tactische en operationele analyses doen zich voor ten aanzien van de volgende aspecten:

- de toetsbestanden zullen bij tactische analyses niet alleen bestaan uit gegevens die betrekking hebben op bekende verdachten, maar ook gegevens bevatten over delicten of signalementen waarvan de dader nog onbekend is. Immers, deze analyses zijn met name gericht op het in kaart brengen van de opsporingsmogelijkheden van bepaalde delicten of personen. Daarvoor willen we gebruik maken van alle mogelijke gegevens, niet alleen gegevens waaraan al een verdachte is gekoppeld;
- Het zoekresultaat bij tactische analyses wordt niet beoordeeld op grond van de rangordepositie van een dader waarnaar we op zoek zijn, maar simpelweg op grond van de matchwaarden zelf. Met andere woorden, bij tactische analyses zijn we alleen geïnteresseerd in de best matchende waarden in het zoekresultaat, de top-X van een bepaalde matchprocedure.

Ook bij tactische analyses wordt de grondslag van het matchen gevonden in het bij elkaar plaatsen van gegevens die mogelijk een zelfde dader toebehoren. Het is bij deze analyses echter niet noodzakelijk (wel mogelijk!) om hieraan meteen concrete namen van verdachten te koppelen. Deze grondslag maakt dat we de resultaten uit de voorgaande paragrafen ook kunnen benutten om de waarde van de verschillende zoekmethoden voor tactisch gebruik vast te stellen. Het is zelfs de enige grondslag die we hebben om de resultaten van een tactische analyse op bruikbaarheid te beoordelen. We kunnen bij tactische analyses namelijk geen gebruik maken van een simulatiemodel om de resultaten te beoordelen (doordat een deel van de resultaten betrekking heeft op onbekende daders).

Kortom, wanneer we geïnteresseerd zijn om van een bepaald delict of signalement te weten welke de opsporingsmogelijkheden zijn (op basis van een uitgevoerde matchprocedure), kunnen we simpelweg gebruik maken van de hiervoor ontwikkelde en beschreven zoekmethoden en kunnen we de resultaten uit de voorgaande paragrafen gebruiken om de bruikbaarheid van de verschillende zoekmethoden voor tactische toepassing te bepalen.

De reikwijdte van een tactische analyse kan echter ook het individuele delict of signalement ontstijgen door voor een groep van delicten of signalementen te bepalen welke de meeste opsporingsmogelijkheden bieden. Dit zou een aanzienlijke uitbreiding van de toepassingsmogelijkheden van het zoekinstrument bieden. Wij zijn er in dit onderzoek echter niet in geslaagd hieraan op enige manier vorm te geven. We zullen kort aangeven welk probleem zich hierbij aandient dat door ons niet kan worden opgelost.

Bij het uitvoeren van een tactische analyse zoals bedoeld in de vorige alinea kan niet volstaan worden met het ordenen van de zoekresultaten van individuele matchprocedures om vervolgens de best matchende delicten of signalementen te selecteren als zoekresultaat. Dit heeft te maken met het feit dat we hier de (beste) matchresultaten van verschillende delicten of signalementen met elkaar moeten vergelijken en hierin schuilt een probleem, omdat de resultaten die worden verkregen bij een individuele matchprocedure afhankelijk zijn van de kenmerken waarop het zoekdelict of –signalement waarden heeft en de inhoud van die waarden. Concreet: hoe groter het aantal kenmerken waarop bijvoorbeeld een delict waarden heeft, des te lager zullen doorgaans de (beste) matchwaarden uitvallen. Immers, hoe groter het aantal waarden, des te moeilijker de doelgegevens op alle waarden kunnen voldoen en des te lager het matchresultaat zal uitvallen. Hetzelfde geldt voor de inhoud van de waarden: naarmate deze unieker zijn zal het delict moeilijker gematcht kunnen worden en zullen de matchwaarden dus lager uitvallen. Schematisch kunnen we dit als volgt voorstellen (zie schema 13):

Schema 13 Matchresultaten van delicten met variërende waarden

<i>Delict heeft:</i>	Algemene waarden	Unieke waarden
Weinig waarden	++	+ -
Veel waarden	+ -	--

Delicten (en hetzelfde geldt voor signalementen) die weinig waarden hebben om mee te zoeken zullen bij zoekmethode 2 (en ook bij 3) relatief hoge matchwaarden laten zien, terwijl delicten met veel waarden naar verhouding lagere matchresultaten laten zien. Hetzelfde geldt bij zoekmethode 1 voor delicten met algemene en unieke waarden. Zoekmethode 3, als middenweg tussen de zoekmethoden 1 en 2, zal door beide mechanismen tegelijk beïnvloed worden. Dit maakt het problematisch om de matchresultaten tussen delicten (of signalementen) te vergelijken.

Let op: over deze kwestie kan wellicht verwarring ontstaan omdat in de voorgaande paragrafen telkens werd aangegeven dat delicten met meer zoekwaarden ook betere matchresultaten laten zien. Het gaat hier echter om een vergelijking *tussen* delicten die hiervoor niet aan de orde was: per delict worden nu de beste matchresultaten bepaald en deze resultaten worden vervolgens *over* delicten met elkaar vergeleken. Bij *deze* vergelijking treedt het probleem naar voren dat we hier beschrijven.

Als we aan dit probleem geen aandacht zouden besteden en simpelweg de delicten of signalementen met de beste matchwaarden zouden selecteren, dan zou het resultaat telkens bestaan uit een verzameling van delicten of signalementen die bij elkaar passen op grond van het feit dat ze op relatief weinig kenmerken waarden hebben en de waarden bovendien vrij algemeen zijn. Dit zijn uiteraard geen gegevens met veel opsporingswaarde. Nu zou men kunnen beweren dat in deze situatie dus niet de best matchende maar juist de slechtst matchende delicten of signalementen het meest waardevol zijn. Immers, delicten of signalementen met relatief veel waarden en/of relatief unieke waarden zullen doorgaans juist lagere matchwaarden opleveren. Deze redenering gaat weliswaar op, maar lost niet de vraag op hoe we vervolgens moeten omgaan met het ordenen van de zoekresultaten uit individuele matchprocedures. Immers, ook de ‘onderkant’ van het zoekresultaat wordt beïnvloed door de kenmerken waarop het delict of signalement waarden heeft en de mate waarin deze waarden algemeen of uniek zijn. Alleen door middel van een weegprocedure of iets dergelijks zou het mogelijk zijn de resultaten van de individuele matchprocedures naar ‘opsporingsmogelijkheden’ te ordenen, maar deze werkwijze is erg omslachtig. Bovendien wordt het hierdoor wel erg moeilijk om nog te begrijpen wat het zoekresultaat voorstelt en op grond van welke criteria een bepaald zoekresultaat is verkregen. De vraag is ook of de praktijk met een dergelijke procedure voldoende uit de voeten kan.

Vooralsnog concluderen we daarom dat een tactische analyse die is gebaseerd op het selecteren (uit enige verzameling) van delicten of signalementen als zijnde ‘meest opsporingswaardig’, niet tot bruikbare resultaten zal leiden.

6. Bruikbaarheidsindicatoren

In de huidige praktijk vindt er geen toetsing plaats van de bruikbaarheid van een bepaald zoekresultaat. Dit wordt in veel gevallen wellicht ook niet gemist, omdat het soort zoekvragen dat in dit onderzoek aan de orde is, zelden wordt voorgelegd in de huidige praktijk. Echter, ook al hebben de zoekvragen in de huidige praktijk vaak een ander karakter, in veel gevallen ligt er wel dezelfde informatiebehoefte aan ten grondslag als aan onze zoekvragen, namelijk (direct of indirect) inzicht verkrijgen in de kenmerken of de persoon van de (mogelijke) dader

door het bij elkaar plaatsen van gegevens over delicten en/of personen. In de huidige praktijk kan de bruikbaarheid van een zoekresultaat niet worden vastgesteld, omdat daarvoor een maat ontbreekt; de doelgegevens voldoen eenvoudig wel of niet aan bepaalde vooraf opgestelde criteria. Of het uiteindelijke zoekresultaat bijdraagt aan de opsporingsactiviteit is niet bekend.

De analyses die we in dit hoofdstuk hebben uitgevoerd geven een indicatie van de bruikbaarheid van de huidige zoekpraktijk bij verschillende zoekvragen. Wanneer we deze kennis (wellicht nog verder verfijnd naar verschillende typen delicten etc.) in een zoekinstrument verwerken als expertkennis, kan op basis hiervan een indicatie van bruikbaarheid worden verschaft aan elk zoekresultaat dat op basis van deze zoekmethode wordt verkregen. Deze werkwijze kan uiteraard ook worden gevolgd bij de andere zoekmethoden. Bij deze laatste methoden is het echter ook mogelijk het verkregen zoekresultaat als zodanig op bruikbaarheid te beoordelen, omdat hier ook matchwaarden kunnen worden berekend. Onafhankelijk van expertkennis geven deze waarden inzicht in de mate waarin de doelgegevens overeenkomen met de zoekgegevens.

We concluderen derhalve dat er verschillende mogelijkheden zijn om een zoekinstrument van bruikbaarheidsindicatoren te voorzien. Enerzijds kunnen deze gebaseerd worden op expertkennis die via simulaties is verkregen (zoals in dit onderzoek). Anderzijds kan bij de zoekmethoden 2 en 3 gebruik worden gemaakt van de berekende matchwaarden.

7. Samenvatting: de belangrijkste onderzoeksresultaten op een rij

In dit hoofdstuk hebben we 4 groepen van analyses uitgevoerd. Het gaat hierbij om:

- het matchen van individuele delicten
- het matchen van individuele delicten met delictprofielen
- het matchen van individuele signalementen
- het matchen van delict/signalementcombinaties

De resultaten van deze analyses worden hier samengevat. Aansluitend bespreken we welke delicten, signalementen en delictprofielen zich beter respectievelijk slechter dan gemiddeld laten matchen. Ook geven we aan welke delicten, etc. goede matchresultaten laten zien op de ene zoekmethode, terwijl ze juist slecht matchen op de andere zoekmethode. Dit geeft enig inzicht in de gevoeligheid van de zoekmethoden voor het matchen van verschillende soorten delicten, etc. Enkele korte opmerkingen over het toepassen van zoekvragen voor tactisch gebruik en het hanteren van bruikbaarheidscriteria bij het interpreteren van de zoekresultaten, besluiten deze paragraaf.

Het matchen van individuele delicten

Een willekeurige steekproef van delicten (waarvan de dader bekend is) is gematcht met circa 48000 delicten die in de periode 1999-2001 in de regio Haaglanden zijn aangegeven. Voor deze delicten geldt verder dat er tenminste 1 verdachte aan gekoppeld is. De matchprocedure is erop gericht de delicten van de bekende dader (van het steekproefdelict) een zo hoog mogelijke rangordepositie in het zoekresultaat te bezorgen.

De huidige zoekmethode (zoekmethode 1) slaagt erin om 5,2% van de bekende daders bij de beste 10 delicten in het zoekresultaat te matchen. In ruim een kwart van alle gevallen wordt bij deze methode de dader aangetroffen bij de beste 500 delicten in het zoekresultaat. De alternatieve zoekmethoden (2 en 3) laten betere resultaten zien. De beste variant van zoekmethode 3 slaagt erin om niet 5,2% van de daders bij de beste 10 delicten te matchen, maar 16,8%. In totaal kan 31,4% van de daders worden gematcht bij de beste 500 delicten (tegen 25,6% bij zoekmethode 1). Als we ons beperken tot delicten uit de steekproef waarvan de dader (nogmaals) vóórkomt in het HKS, zien we dat zoekmethode 3 erin slaagt om 25,4%

van de daders bij de beste 10 delicten in het zoekresultaat te matchen (tegen 7,9% bij zoekmethode 1). In totaal kan bijna de helft van alle daders (47,4%) bij de eerste 500 delicten worden gematcht (=38,7% bij zoekmethode 1).

We concluderen derhalve dat het matchen van delicten goede resultaten oplevert en dat de alternatieve zoekmethode 3 beter scoort dan de huidige zoekmethode.

NB De verschillende varianten die zijn uitgewerkt voor zoekmethode 3 laten zien hoezeer het toevoegen van voorwaarden aan de zoekprocedure ten koste gaat van een goed zoekresultaat. Wanneer de gegevens moeten matchen op drie vooraf gespecificeerde kenmerken kan hooguit 5,8% van de daders bij de eerste 500 delicten gematcht worden! Dit onderstreept de zwakte van de huidige zoekmodaliteit in het HKS. (We komen hierop terug in de conclusie.)

Het matchen van individuele delicten met delictprofielen

Om na te gaan of het matchen van geclusterde gegevens een meerwaarde heeft boven het matchen van enkelvoudige gegevens, hebben we een willekeurige steekproef van 500 delicten gematcht met 25.735 delictprofielen van verdachten die in de periode 1999-2001 tenminste 1 delict hebben gepleegd dat in de regio Haaglanden is geregistreerd. De steekproef van delicten is dezelfde als die we bij de vorige analyse hebben gebruikt. De matchprocedure is erop gericht het delictprofiel van de bekende dader (van het steekproefdelict) een zo hoog mogelijke rangordepositie in het zoekresultaat te bezorgen.

Van de verschillende zoekmethoden die we ontwikkeld hebben voor het matchen van geclusterde gegevens blijkt de methode die het meest verwant is aan de huidige zoekmethode (voor enkelvoudige gegevens) de beste resultaten te genereren (=zoekmethode 1). Zoekmethode 2 blijkt bij de geclusterde gegevens, in tegenstelling tot de analyses hiervoor, tot betere zoekresultaten te leiden dan (de beste variant van) zoekmethode 3. We hebben ook onderzocht in hoeverre het zoekresultaat wordt beïnvloed door de factor tijd (de tijd die is verstreken tussen het zoekdelict en de doelgegevens) en de variërende omvang van de delictprofielen. Hiervoor hebben we diverse varianten van zoekmethode 2 uitgewerkt waarvan we vervolgens de zoekresultaten hebben onderzocht. Eén van die varianten leidt tot een verbetering: wanneer we bij zoekmethode 2 alleen die delicten uit het profiel meetellen die binnen 1 jaar van het zoekdelict gepleegd zijn, zijn we in staat het zoekresultaat van zoekmethode 2 te verbeteren. Uiteindelijk lijkt dit zoekresultaat op het resultaat dat we hebben verkregen op grond van onze variant van zoekmethode 1.

Als we ons beperken tot delicten uit de steekproef waarvan de dader (nogmaals) vóórkomt in het HKS, zien we dat zoekmethode 1 erin slaagt om 10,4% van de daders bij de beste 10 delicten in het zoekresultaat te matchen (tegen 11,9% bij zoekmethode 2). In totaal kan 44% van alle daders bij de eerste 500 profielen worden gematcht (=41,8% bij zoekmethode 2). Deze resultaten leveren een verbetering op ten opzichte van het resultaat dat werd verkregen bij het matchen van individuele delicten op grond van de huidige zoekmethode. Echter, de door ons ontwikkelde zoekmethode 3 voor het matchen van individuele delicten scoort beter dan de beste zoekmethoden voor geclusterde gegevens.

We concluderen derhalve dat het matchen van geclusterde gegevens bevredigende resultaten oplevert, die ook beter zijn dan de resultaten van het matchen van individuele delicten op basis van de huidige zoekmethode. Als we het resultaat echter vergelijken met het zoekresultaat bij het matchen van individuele delicten, moeten we constateren dat het matchen van delictprofielen geen meerwaarde oplevert ten opzichte van het matchen van individuele delicten.

Het matchen van individuele signalementen

Voor het matchen van signalementen OD met signalementen BD is gebruik gemaakt van een steekproef van 410 signalementen OD die bewaard zijn gebleven nadat de aangifte gekoppeld werd aan een verdachte (met een signalement BD). Het gaat om aangiften in de periode 1999-2001 (in de regio Haaglanden). Deze steekproef stelt ons in staat te toetsen in hoeverre signalementen OD bruikbaar zijn om (d.m.v. matching met signalementen BD) uit kunnen komen bij de persoon van de dader. Het toetsbestand bestaat uit alle signalementen (BD) van personen die in de periode 1999-2001 tenminste 1 delict hebben gepleegd. Dit zijn er 7302.

Alvorens de toetsing uit te voeren hebben we onderzocht in hoeverre de signalementen OD uit de steekproef overeenkomen met de signalementen BD van de betreffende dader. Een minimale overeenkomst is een noodzakelijke (maar geen voldoende!) voorwaarde om enig resultaat van het matchen te mogen verwachten. Uit deze analyse bleek dat de signalementen uit de steekproef gemiddeld op bijna 6 kenmerken waarden hebben waarvan er gemiddeld 3,5 terug te vinden zijn in het 'bijbehorende' signalement BD. Naarmate een signalement OD over meer waarden beschikt neemt ook het aantal waarden toe dat in het bijbehorende signalement is terug te vinden. De relatieve kwaliteit van de 'match' tussen signalement OD en BD neemt echter iets af naarmate het aantal waarden op het signalement OD toeneemt. We concluderen daarom dat de kwaliteit van de signalementen OD niet erg hoog is, maar vooralsnog voldoende om de matchprocedure uit te voeren.

De resultaten van de matchprocedure zijn voor alle zoekmethoden aan de magere kant. Zoekmethode 3, die net als bij het matchen van delicten ook hier de beste resultaten levert, slaagt er slechts in een half procent van de daders uit de steekproef bij de 10 beste signalementen in het zoekresultaat te matchen (=0,2% bij zoekmethode 1). In totaal kan zoekmethode 3 27,8% van de daders bij de eerste 500 signalementen matchen (=21,7% bij zoekmethode 1). Aangezien het hier gaat om een steekproef van signalementen waarvan de daders zich allemaal in het HKS bevinden, mogen we dit resultaat als mager kwalificeren.

We concluderen derhalve dat het matchen van signalementen niet veel oplevert, tenminste veel minder dan het matchen van delicten. We zien ook hier dat de alternatieve zoekmethode 3 (iets) beter scoort dan de huidige zoekmethode.

NB Een verklaring voor de magere zoekresultaten kan mogelijk gevonden worden in het beperkte aantal waarden dat bij signalementen OD beschikbaar is om mee te zoeken (gemiddeld 5.9 tegen gemiddeld ruim 25 waarden bij delicten). Verder is de kwaliteit van de signalementen OD matig (veel waarden zijn niet terug te vinden in de bijbehorende signalementen BD). Ten slotte: bij delicten is het mogelijk om meer dan één delict van een verdachte te matchen. Alleen het beste matchresultaat van de betreffende verdachte telt mee. Dus hoe meer delicten een verdachte heeft, des te groter de kans op een goed matchresultaat. Bij het matchen van signalementen is deze optie niet beschikbaar.

Het matchen van delict/signalementcombinaties

In hoofdstuk 2 is betoogd dat het toevoegen van signalementgegevens aan delictgegevens onduidelijke gevolgen heeft voor het zoekresultaat van een matchprocedure; de toename van gegevens suggereert betere matchresultaten, maar het feit dat persoonsgegevens in het HKS schaarser zijn dan delictgegevens suggereert slechtere matchresultaten. De bedoeling van de onderhavige analyse was dan ook te onderzoeken welk effect optreedt bij de verschillende zoekmethoden. Hierbij is gebruik gemaakt van dezelfde steekproef als hiervoor bij de signalementen. Alleen is nu ook gebruik gemaakt van de delictgegevens. Deze delicten met de daaraan gekoppelde signalementen OD zijn gematcht met delicten waaraan een signalement BD is gekoppeld. In totaal gaat het om 23.950 unieke delict / signalementcombinaties in de periode 1999-2001 (in de regio Haaglanden).

Alvorens deze toetsing uit te voeren hebben we nogmaals een matching van individuele delicten uitgevoerd, maar nu op basis van de hier beschreven steekproef en het hier beschreven toetsbestand. Dit hebben we gedaan om een meer zuivere vergelijking mogelijk te maken tussen de verschillende zoekresultaten (matchen van delicten versus matchen van signalementen versus matchen van delict/signalementcombinaties). *De resultaten van deze analyse bevestigen het beeld dat eerder naar voren kwam bij het matchen van delicten, namelijk dat het matchen van delicten goede resultaten oplevert en dat de alternatieve zoekmethode 3 een substantiële verbetering oplevert ten opzichte van de huidige zoekmethode.*

NB Opvallend bij deze analyse is dat de resultaten zelfs aanzienlijk beter zijn dan bij de analyse die hiervoor werd beschreven. Als de dader (nogmaals) vóórkomt in het HKS, slaagt zoekmethode 3 erin om deze in 37,1% van de gevallen bij de beste 10 delicten te matchen (=25,6% bij zoekmethode 1). In totaal wordt de dader in 72% van de gevallen bij de beste 500 delicten gematcht (=54,4% bij zoekmethode 1). Mogelijk is het resultaat hier beter omdat het toetsbestand kleiner is dan bij de analyse in groep 1, waardoor de toetsing minder kritisch is. Ook in de steekproeven hebben we opvallende verschillen aangetroffen: de hier gebruikte steekproef bevat per delict gemiddeld meer waarden om mee te zoeken (29 tegen 25 in de andere steekproef) en aanzienlijk meer gewelddadige vermogensdelicten (38 tegen 4 procent). Ook de periode die de delicten bestrijken, verschilt in beide steekproeven. Het is echter onduidelijk in hoeverre deze zaken het zoekresultaat beïnvloeden.

Hierna hebben we de resultaten gepresenteerd van het gecombineerd matchen van delict- en signalementgegevens. Uit deze analyse komt naar voren dat zoekmethode 1 opvallend slechter scoort dan de zoekmethoden 2 en 3. Zoekmethode 3 slaagt erin om –als de dader (nogmaals) vóórkomt in het HKS, deze in ruim 40% van de gevallen bij de beste 10 delicten in het zoekresultaat te matchen (=6,4% bij zoekmethode 1). In totaal slaagt deze methode erin om bijna 71% van de daders bij de beste 500 delicten in het zoekresultaat te matchen (=15,9% bij zoekmethode 1). Zoekmethode 2 scoort in dit opzicht zelfs nog beter; in totaal bijna 73% van de daders bij de beste 500 delicten in het zoekresultaat.

We concluderen derhalve dat het matchen van gecombineerde delict- en signalementgegevens goede resultaten oplevert indien gebruik wordt gemaakt van zoekmethode 2 of 3. Indien zoekmethode 1 wordt gebruikt, zijn de resultaten erg mager. Het toevoegen van signalementgegevens aan delictgegevens zorgt voor: a) een lichte verbetering in het resultaat van zoekmethode 2, b) een min of meer ongewijzigd zoekresultaat van zoekmethode 3 en c) een aanzienlijk slechter zoekresultaat van zoekmethode 1.

NB de verschillen in prestaties tussen de zoekmethoden kunnen hier worden verklaard uit de wijze waarop per methode de matchresultaten tot stand komen. Voor zoekmethode 2 leveren de toegevoegde waarden van de signalementen OD een bijdrage aan het betere zoekresultaat, terwijl de relatief unieke waarden op de signalementen bij zoekmethode 1 juist zorgen voor een verslechtering van het zoekresultaat. Zoekmethode 3 profiteert van het aantal extra waarden om mee te zoeken, maar dit voordeel wordt deels ongedaan gemaakt door het feit dat de nadere voorwaarden vaker betrekking hebben op relatief unieke signalementkenmerken (die slecht matchen).

Delicten, signalementen en delictprofielen die beter of slechter dan gemiddeld matchen

Het inzicht in de zoekresultaten kan worden verdiept door na te gaan of delicten etc. die goed matchen afwijken van delicten etc. die niet goed matchen. De belangrijkste bevindingen hieromtrent worden hieronder samengevat:

- hoe meer gegevens bekend zijn in het HKS (van een delict, een signalement of een delictprofiel), des te beter kan de dader worden gematcht (vooral van belang voor zoekmethoden 2 en 3);
- hoe groter het aantal delicten van een dader in het HKS, des te beter kunnen delicten van deze dader worden gematcht;
- hoe groter het aantal daders dat betrokken is bij een delict, des te beter kan dit delict worden gematcht;
- sommige delictsoorten laten zich beter dan gemiddeld matchen (zoals vermogensdelicten), terwijl andere zich slechter dan gemiddeld laten matchen (zoals geweldsdelicten en verkeersdelicten). Ten aanzien van de andere delictsoorten is het beeld diffuus;
- delicten met waarden op de door ons gecreëerde MO-velden matchen vooral bij zoekmethode 2 en 3 beter dan gemiddeld (met uitzondering van de MO-categorieën die op het gebruik van geweld duiden);
- signalementen van vrouwen laten zich beter matchen dan signalementen van mannen.

Delicten, signalementen en delictprofielen die goed matchen bij de ene zoekmethode maar slecht bij de andere

Doorgaans bestaat er een matig tot sterk verband tussen de resultaten van de verschillende zoekmethoden. Niettemin kan het inzichtelijk zijn om te kijken welke delicten, etc. bij de ene zoekmethode goed matchen, terwijl ze bij de andere methode juist slecht matchen. Deze gegevens verschaffen ons informatie over de gevoeligheid van de verschillende zoekmethoden voor specifieke kenmerken van zoekgegevens.

Gegevens die doorgaans goed gematcht worden bij de huidige zoekmethode (zoekmethode 1), maar minder goed of slecht bij zoekmethode 2 of 3, zijn vaker dan gemiddeld geweldsdelicten of delicten met een gewelds-MO en signalementen van vrouwen. Gegevens die goed gematcht worden bij zoekmethode 2 of 3 (en dus niet goed bij de huidige zoekmethode) zijn vaker dan gemiddeld delicten of signalementen waarvan relatief veel gegevens beschikbaar zijn. Ook gaat het vaker dan gemiddeld om vermogensdelicten en om delicten die waarden hebben op de door ons aangemaakte MO-velden.

We kunnen concluderen dat zoekmethode 1 gevoeliger is voor het matchen van unieke gegevens, terwijl de zoekmethoden 2 en 3 het beter doen bij delicten, signalementen en delictprofielen waarover veel gegevens bekend zijn.

Zoekvragen toepassen voor tactisch gebruik

De procedure bij het matchen van gegevens voor tactisch gebruik (om de opsporingswaardigheid vast te stellen) wijkt niet van die bij operationele zoekvragen. De resultaten van de in dit hoofdstuk uitgevoerde analyses kunnen derhalve ook gebruikt worden om de toepassing van verschillende zoekvragen en –methoden voor tactisch gebruik te evalueren. De tactische analyses wijken in twee opzichten af van de operationele analyses: de toetsbestanden bestaan niet alleen uit gegevens waaraan concrete verdachten gekoppeld zijn, maar ook uit gegevens waarvan nog geen dader bekend is. Daarnaast komt het zoekresultaat anders tot stand: bij tactische analyses zijn we alleen geïnteresseerd in de doelgegevens met de hoogste matchwaarden, de top-X van een bepaalde matchprocedure.

Wanneer we voor een groep van delicten, signalementen of delictprofielen de ‘opsporingswaardigheid’ willen beoordelen, stuiten we op het probleem dat het niet goed mogelijk is de gevonden matchwaarden tussen delicten, etc. te vergelijken. Om die reden hebben we deze analyse niet uitgewerkt.

De toepassing van bruikbaarheidsindicatoren om de zoekresultaten te beoordelen

Om de zoekresultaten op hun nut te kunnen beoordelen is het handig om gebruik te maken van indicatoren die aangeven hoe bruikbaar een zoekresultaat is, welke waarde eraan toegekend kan worden. We hebben aangegeven dat de kennis die hier is ontwikkeld op grond van de simulaties (eventueel nog verder verfijnd) in een zoekinstrument verwerkt kan worden als expertkennis op basis waarvan zoekresultaten voorzien kunnen worden van een bruikbaarheidsindicatie. Dit kan bij elke zoekmethode worden toegepast. Bij de zoekmethoden 2 en 3 is het bovendien mogelijk om de berekende matchwaarden als indicatie te nemen voor de bruikbaarheid van het zoekresultaat.

5. Samenvatting en conclusie

Uitgangspunt

Het uitgangspunt van dit onderzoek is dat een relatief kleine groep daders verantwoordelijk is voor een relatief groot deel van de (veel voorkomende) criminaliteit. Dit gegeven is van evident belang voor opsporingsonderzoek. Op grond hiervan kunnen we immers stellen dat bij een willekeurig misdrijf met onbekende dader (OD), de kans groot is dat de politie al eerder werd geconfronteerd met delicten van deze dader. Ook is de kans aanwezig dat de dader zelf al eerder met de politie in aanraking kwam. Met andere woorden, bij (de voorbereiding van) een willekeurig opsporingsonderzoek is de kans groot dat de politie al beschikt over gegevens omtrent delicten en/of daders die in verband kunnen worden gebracht met het delict dat men wil ophelderen. Het (snel) boven water krijgen van deze gegevens kan een belangrijke bijdrage leveren aan het opsporingsonderzoek. Hiervoor is een efficiënt en effectief zoekinstrument onontbeerlijk.

Doel

Het doel van dit onderzoek is de haalbaarheid te onderzoeken van een nieuw zoekinstrument dat, gebruikmakend van gegevens uit het HKS, in staat is op een efficiënte en effectieve wijze gegevens (van delicten of personen) te koppelen aan gegevens die beschikbaar zijn in een (te starten) opsporingsonderzoek, opdat aanknopingspunten worden verkregen voor (verder) opsporingsonderzoek. Concreet wordt hiermee beoogd de beschikbare gegevens in het HKS optimaler te benutten voor tactische en operationele onderzoeksdoelen dan nu -in de huidige praktijk- het geval is.

Onderzoeksvragen

Het optimaler benutten van gegevens uit het HKS voor het beantwoorden van zoekvragen, kan op verschillende manieren plaatsvinden. We hebben onszelf in dit verband de volgende onderzoeksvragen gesteld:

1. De huidige zoekmodaliteit in het HKS wordt alleen gebruikt voor operationele doeleinden. Is het mogelijk en zinvol om een zoekinstrument te ontwikkelen dat ook voor *tactische doeleinden* gebruikt kan worden, bijvoorbeeld om vast te stellen wat de opsporingsmogelijkheden zijn van een specifieke aangifte of groep van aangiften of van een (groep van) verdachte(n)? Kortom, is het mogelijk de zoekdoelen uit te breiden?
2. Is het mogelijk en zinvol om *nieuwe zoekvragen* te creëren, bijvoorbeeld door gebruik te maken van het feit dat bepaalde gegevens in het HKS geclusterd zijn binnen eenheden (zoals delicten binnen verdachten)? Kortom, is het mogelijk de zoekvragen uit te breiden?
3. Heeft het *veredelen van de beschikbare gegevens* in het HKS (door het aggregeren en combineren van bestaande gegevens tot nieuwe gegevens) meerwaarde voor het zoekproces?
4. Geven *alternatieve zoekmethoden* een beter zoekresultaat dan de huidige zoekmodaliteit in het HKS?
5. Op dit moment bestaan er geen criteria waarmee de bruikbaarheid van een verkregen zoekresultaat kan worden beoordeeld. Is het mogelijk dergelijke *bruikbaarheidscriteria* te ontwikkelen?

Onderzoeksopzet

We hebben eerst in kaart gebracht hoe de huidige zoekprocedure in het HKS werkt en welke (mogelijke) beperkingen de huidige zoekpraktijk kent bij het beantwoorden van zoekvragen zoals die in dit onderzoek aan de orde zijn. Hierbij werden beperkingen van uiteenlopende aard aangetroffen. Beperkingen van technische, organisatorische en juridische aard vallen buiten de focus van dit onderzoek. We hebben ons beperkt tot zaken die we kunnen beïnvloeden. Dit zijn de 5 zaken die aan de orde zijn in de onderzoeksvragen. Om deze vragen te kunnen beantwoorden hebben we in kaart gebracht welke mogelijke zoekdoelen, zoekvragen en zoekmethoden we zouden kunnen ontwikkelen. Vervolgens hebben we onderzocht op welke wijze en met behulp van welke gegevens we deze nieuwe mogelijkheden zouden kunnen toetsen op bruikbaarheid en toegevoegde waarde voor de huidige praktijk. Hiervoor hebben we gebruik gemaakt van simulaties; door delicten en signalen te matchen van bekende verdachten kunnen we nagaan hoe goed de verschillende zoekmethoden in staat zijn om de ‘juiste persoon’ in het zoekresultaat boven water te krijgen.

Onderzoeksresultaten

Kan het zoekinstrument ook voor tactische doeleinden worden gebruikt?

Het antwoord op deze vraag is ‘ja’ en ‘nee’. Het is mogelijk om voor individuele delicten, signalen of delictprofielen de opsporingswaardigheid na te gaan door hiervan de best matchende gegevens in kaart te brengen. Wanneer we dit voor een groep van delicten, etc. willen doen, stuiten we op het probleem dat het niet goed mogelijk is de gevonden matchwaarden tussen delicten (respectievelijk signalen, delictprofielen) te vergelijken. Om die reden hebben we deze analyse niet uitgewerkt.

Is het mogelijk om nieuwe zoekvragen te creëren?

We hebben onderzocht of het meerwaarde heeft om delicten binnen verdachten te clusteren en deze (delictprofielen) vervolgens te matchen met afzonderlijke delicten. De analyses laten zien dat de zoekresultaten verbeteren ten opzichte van de huidige zoekmethode bij het matchen van individuele delicten. Als we de resultaten echter vergelijken met de beste zoekmethoden die we hebben voor het matchen van individuele delicten, wordt geen verbetering bereikt ten opzichte van het matchen van afzonderlijke delicten. Dit hoeft overigens niet te betekenen dat het matchen van gegevens uit delictprofielen geen toepassing kan vinden. De resultaten van de analyses zijn op zichzelf bevredigend.

Heeft het meerwaarde voor het zoekproces om gegevens in het HKS te veredelen?

Met name de door ons ontwikkelde zoekmethoden blijken erg gevoelig voor het aantal gegevens dat beschikbaar is om te matchen. Hoe meer gegevens er beschikbaar zijn, des te beter de kwaliteit van het zoekresultaat. De extra gegevensvelden die we door aggregatie en combinatie van bestaande velden in het HKS hebben verkregen, bewijzen aldus goede diensten voor de ons ontwikkelde zoekmethoden. Uit de analyses blijkt dat met name de door ons gecreëerde MO-velden (waarin delictkenmerkcodes zijn samengevoegd tot geaggregeerde MO-categorieën) een belangrijke bijdrage leveren aan het verkrijgen van een goed matchresultaat. Het antwoord op de vraag luidt derhalve dat het inderdaad meerwaarde heeft voor het zoekproces om (zoveel mogelijk) nieuwe gegevens te creëren uit de bestaande gegevens.

Produceren alternatieve zoekmethoden betere zoekresultaten dan de huidige zoekmethode?

De analyses tonen aan dat de door ons ontwikkelde alternatieve zoekmethoden in alle gevallen beter scoren dan de bestaande zoekmethode. Bij het matchen van delicten leveren de

alternatieve zoekmethoden substantieel betere zoekresultaten. Bij het matchen van signalementen gaat het om licht betere resultaten. Bij het gelijktijdig matchen van delict- en signalementgegevens zijn de verschillen in zoekresultaat tussen de alternatieve zoekmethoden en de huidige methode zelfs extreem groot (in het voordeel van de alternatieve zoekmethoden). Bij het matchen van delictprofielen kan geen vergelijking met de huidige praktijk worden gemaakt.

Hoe moeten we dit resultaat evalueren? De beoordeling hangt af van de wijze waarop we de uitkomsten van de huidige zoekmethode beschouwen. In hoofdstuk 3 hebben we uiteengezet dat de resultaten van zoekmethode 1 in dit onderzoek gebaseerd zijn op een computersimulatie. We hebben geprobeerd de opzet hiervan zodanig vorm te geven dat daarmee een zo kritisch mogelijke toets van de alternatieve zoekmethoden zou worden verkregen. De vraag of we hierin geslaagd zijn, bepaalt in belangrijke mate de waarde van de alternatieve zoekmethoden. Ons antwoord op deze vraag is dat de resultaten van de uitgevoerde analyses in dit onderzoek ons nog meer aanleiding geven om te denken dat de resultaten die wij verkregen hebben uit de computersimulatie gunstig tot misschien wel zeer gunstig afsteken tegen de zoekresultaten die in de huidige zoekpraktijk worden behaald. We hebben hiervoor 2 argumenten:

- 1) In hoofdstuk 3 hebben we uitgelegd dat we, om een kritische toets mogelijk te maken, zogenaamd parallelle zoekresultaten hebben toegestaan bij zoekmethode 1. Feitelijk betekent dit dat we per matchprocedure meerdere (niet geneste) matchresultaten tegelijk in beschouwing hebben genomen en ook verwerkt hebben in het uiteindelijke zoekresultaat. Doordat in onze simulatie alle zoekgegevens optimaal benut worden, komt het vaak voor (waarschijnlijk veel vaker dan in de praktijk) dat al in de eerste ronde van de matchprocedure vrij kleine clusters ontstaan van doelgegevens die matchen op enig kenmerk (in combinatie met bijvoorbeeld cbs-delictcategorie). Dit gegeven vergroot de kans op parallelle zoekresultaten en zorgt voor betere 'prestaties' van zoekmethode 1 in dit onderzoek (ten opzichte van de praktijk waar doorgaans 1 zoekresultaat wordt geproduceerd).
- 2) Het tweede argument ontleen we aan de zoekresultaten die we hebben verkregen bij bepaalde varianten van zoekmethode 3. In paragraaf 1 van hoofdstuk 8 hebben we laten zien wat het effect is op de matchresultaten wanneer je voorwaarden toevoegt aan de zoekprocedure (de feitelijke werkwijze in de huidige praktijk). Elke waarde van een kenmerk die wordt gespecificeerd als een voorwaarde waaraan het zoekresultaat moet voldoen, zorgt ervoor dat het matchresultaat sterk achteruit gaat. We hebben in tabel 3 van hoofdstuk 8 kunnen zien dat bij het formuleren van slechts twee van dergelijke voorwaarden het matchresultaat al schrikbarend slecht wordt. De aldaar getoonde zoekresultaten tonen hoe de uitkomst van zoekmethode 1 eruit ziet wanneer geen parallelle matchresultaten zijn toegestaan. Als zodanig zouden deze uitkomsten wel eens representatiever kunnen zijn voor de huidige zoekpraktijk dan onze simulatieresultaten.

Al met al concluderen we daarom dat de verschillen in prestaties tussen de huidige zoekmethode en de door ons ontwikkelde zoekmethoden op zijn minst reëel zijn en in de praktijk nog wel eens (aanzienlijk) groter zouden kunnen blijken, omdat wij de resultaten van zoekmethode 1 in dit onderzoek erg gunstig hebben ingeschat.

Is het mogelijk om bruikbaarheidscriteria op te stellen om de zoekresultaten te kunnen beoordelen?

Om de zoekresultaten op hun nut te kunnen beoordelen is het handig om gebruik te maken van indicatoren die aangeven hoe bruikbaar een zoekresultaat is, welke waarde eraan

toegekend kan worden. We hebben aangegeven dat de kennis die hier is ontwikkeld op grond van de simulaties in een zoekinstrument verwerkt kan worden als expertkennis op basis waarvan zoekresultaten voorzien kunnen worden van een bruikbaarheidsindicatie. Dit kan bij elke zoekmethode worden toegepast. Bij de alternatieve zoekmethoden is het bovendien mogelijk om de berekende matchwaarden als indicatie te nemen voor de bruikbaarheid van het zoekresultaat.

Conclusie: voegt een nieuw zoekinstrument iets toe?

De resultaten van de vier zoekvragen die we in dit onderzoek getoetst hebben, zijn indicatief voor de mogelijkheden om het zoekinstrument te gebruiken voor vier soorten matchprocedures: het matchen van individuele delicten, het matchen van delicten met delictprofielen, het matchen van signalementen en het matchen van gecombineerde delict- en signalementgegevens. In hoofdstuk 2 hebben we uiteengezet welke zoekvragen hierbij zoal aan de orde kunnen zijn (weer onderverdeeld naar zoekvragen voor tactisch en operationeel gebruik). Dit geeft een breed overzicht van mogelijke zoekvragen die in uiteenlopende praktijksituaties bij de politie aan de orde kunnen zijn. De analyses hebben aangetoond dat de meeste van deze zoekvragen zich ook bevredigend laten beantwoorden, waarbij we de alternatieve en beter presterende zoekmethoden als maatstaf nemen.

Het matchen van delicten biedt, zeker bij het gebruik van de alternatieve zoekmethoden, een meerwaarde boven de bestaande praktijk. De zoekresultaten die wij hebben verkregen kunnen waarschijnlijk nog verder verbeterd worden door bijvoorbeeld de tijdperiode van de doelgegevens nader af te bakenen. Bij het matchen van delictprofielen hebben we gezien dat de zoekresultaten toenemen wanneer deze gebaseerd worden op een selectie van delicten die binnen 1 jaar van het zoekdelict gepleegd zijn. Al onze andere analyses in dit onderzoek zijn gebaseerd op toetsbestanden die 3 jaar beslaan. We verwachten dan ook dat de zoekresultaten nog verder verbeterd kunnen worden bij een nadere afbakening van de tijdperiode.

Het matchen van delictprofielen heeft niet de verwachte meerwaarde opgeleverd ten opzichte van het matchen van individuele delictgegevens. Dit neemt niet weg dat deze optie voor de praktijk in vóórkomende gevallen interessant kan zijn. In opsporingsonderzoeken kan het bijvoorbeeld nuttig zijn om het delictprofiel van een verdachte die in beeld is te matchen met delicten OD. De zoekresultaten van deze procedure zijn naar ons idee bevredigend.

Het matchen van signalementen levert bij alle zoekmethoden relatief magere zoekresultaten op. De hoeveelheid gegevens waarop de match gebaseerd moet worden en de kwaliteit van deze gegevens zijn hier debet aan. Wij vermoeden dan ook dat deze optie weinig toegevoegde waarde heeft voor de praktijk.

Het matchen van gecombineerde delict/signalementgegevens levert bij de huidige zoekmethode hele slechte zoekresultaten op. De alternatieve zoekmethoden laten echter sterke resultaten zien. Deze methoden profiteren van het feit dat extra zoekgegevens bijna per definitie tot een verbetering van het zoekresultaat leiden, ongeacht de aard van de (extra) gegevens. Om die reden lijkt het ons ook nuttig om een dergelijke zoekoptie toe te voegen. Deze zal zeker een meerwaarde opleveren ten opzichte van de huidige praktijk, omdat de bestaande zoekmethode niet kan profiteren van het feit dat er in een bepaalde situatie veel of extra gegevens voorradig zijn. De huidige zoekmethode is hiervoor immers niet gevoelig.

Al met al geloven we daarom dat een nieuw zoekinstrument een toegevoegde waarde voor de praktijk kan hebben, vanwege:

- uitbreiding van zoekdoelen en zoekvragen
- betere zoekresultaten
- inzicht in de bruikbaarheid van het zoekresultaat
- betere benutting van beschikbare gegevens in de praktijk

- overzichtelijke en verwerkingsvriendelijke output
- betere toegang voor recherchemedewerkers en analisten

De eerste 3 voordelen komen voort uit de resultaten van het onderhavige onderzoek, de laatste 3 voordelen kunnen worden behaald door bij de implementatie van een nieuw zoekinstrument de gesignaleerde beperkingen in de huidige zoekpraktijk op te heffen.

Aan de implementatie van een nieuw zoekinstrument zoals hier beschreven, zitten uiteraard ook organisatorische, juridische, en andersoortige aspecten die van belang zijn om te overwegen. We noemen er enkele. De regionale compartimentering van het HKS, bijvoorbeeld, beperkt de mogelijkheden van een nieuw zoekinstrument. Een andere kwestie: in de huidige praktijk zijn de gegevens van het HKS slechts voor een beperkt aantal functionarissen toegankelijk. Dit beperkt eveneens de mogelijkheden om het zoekinstrument te benutten. Overigens, de wijze waarop in de huidige praktijk met zoekvragen wordt omgegaan zoals die hier aan de orde zijn, is ook van belang. HKD-medewerkers hebben meermalen bevestigd dat bijvoorbeeld rechercheurs zich vaak niet bewust zijn van het feit dat de resultaten van zoekprocedures kunnen worden verbeterd door gebruik te maken van alle gegevens die over een bepaald feit of een bepaalde persoon beschikbaar zijn. De reikwijdte van de zoekvragen is nu vaak heel beperkt; veel gegevens die wel beschikbaar zijn worden niet optimaal benut. Ingesleten patronen kunnen in dit opzicht belemmerend werken voor de benutting van het hier voorgestelde zoekinstrument. Ook is het van belang een eventuele implementatie te koppelen aan een bestaande tool, zoals bijvoorbeeld DEX2000. Dit vergroot de kansen op inburgering en toepassing van het zoekinstrument. Kortom, het succes van een eventuele implementatie is ook afhankelijk van deze (en wellicht ook andere) randvoorwaarden.

Bijlagen

Bijlage 1 Overzicht van gegevensvelden in geselecteerde HKS-tabellen aan de aangiftenkant.

Bijlage 2 Overzicht van gegevensvelden in geselecteerde HKS-tabellen aan de personenkant.

Bijlage 3 Overzicht van delictkenmerken die gebruikt zijn bij zoekmethode 1

Bijlage 1 Overzicht van geselecteerde gegevensvelden aan de aangiftenkant

Tabel: PVA (proces-verbaal)	
Pva_key	Sleutel proces-verbaal
Korps_code	Korpscode (pva-sleutel)
Korps_afd	Korpsafdeling (pva-sleutel)
Jaar	Pva-jaar (pva-sleutel)
Pva_nummer	Pva-nummer (pva-sleutel)

Tabel: PVAD (proces-verbaal detail)	
Pva_key	Sleutel proces-verbaal
Pvad_id	Sleutel proces-verbaal detail
Tijd_v_hh	Delicttijd: begintijd uur
Tijd_v_mm	Delicttijd: begintijd minuut
Tijd_t_hh	Delicttijd: tot tijd uur
Tijd_t_mm	Delicttijd: tot tijd minuut
Delp_v_dd	Delictdatum: begin dag
Delp_v_mm	Delictdatum: begin maand
Delp_v_yy	Delictdatum: begin jaar
Delp_t_dd	Delictdatum: eind dag
Delp_t_mm	Delictdatum: eind maand
Delp_t_yy	Delictdatum: eind jaar
Land_code	Landcode
Plts_code	Plaatscode (dit is niet de gemeentecode)
Str_code	Straatcode
Wbs_code	Wijk-buurt-subbuurtcode
Pva_v_dag	Delictdag: begin dag van de week
Pva_t_dag	Delictdag: einde dag van de week

Tabel: OD (onbekende dader)	
Pva_key	Sleutel proces-verbaal
Pvad_id	Sleutel proces-verbaal detail
Od_id	Sleutel onbekende dader identiteit
Lengte	Geschatte lengte van onbekende dader
gebjr	Geschatte geboortedatum van onbekende dader (afgeleid van leeftijdschatting)

Tabel: LASK (lichaamsafhankelijke signalementkenmerken van OD)	
Pva_key	Sleutel proces-verbaal
Pvad_id	Sleutel proces-verbaal detail
Od_id	Sleutel onbekende dader identiteit
Lask_id	Sleutel lichaamsafhankelijk signalementkenmerk
Lask_code	Code lichaamsafhankelijk signalementkenmerk
L_r_indik	Links/rechts indicatie
B_o_indik	Boven/onder indicatie

Tabel: TWLASK (trefwoorden van lichaamsafhankelijke signalementkenmerken)	
Pva_key	Sleutel proces-verbaal
Pvad_id	Sleutel proces-verbaal detail
Od_id	Sleutel onbekende dader identiteit
Lask_id	Sleutel lichaamsafhankelijk signalementkenmerk
Twlask_id	Sleutel trefwoord lichaamsafhankelijk signalementkenmerk
trefwoord	trefwoord

Tabel: DEL (tellerbestand voor delicten binnen proces-verbaal)	
Pva_key	Sleutel proces-verbaal
Pvad_id	Sleutel proces-verbaal detail
Del_id	Sleutel afzonderlijk delict binnen proces-verbaal detail

Tabel: FMO (wetsartikelen en feit modus-operandicodes per afzonderlijk delict)	
Pva_key	Sleutel proces-verbaal
Pvad_id	Sleutel proces-verbaal detail
Del_id	Sleutel afzonderlijk delict binnen proces-verbaal detail
Fmo_id	Sleutel wetsartikel/fmo-code
Wa_code	Wetsartikel
Fmo_code	Feit modus-operandicode

Tabel: DKGRP (tellerbestand voor delictkenmerkgroepen binnen afz. delicten)	
Pva_key	Sleutel proces-verbaal
Pvad_id	Sleutel proces-verbaal detail
Del_id	Sleutel afzonderlijk delict binnen proces-verbaal detail
Dkgrp_id	Sleutel delictkenmerkgroepen binnen afzonderlijke delicten

Tabel: DK (delictkenmerken)	
Pva_key	Sleutel proces-verbaal
Pvad_id	Sleutel proces-verbaal detail
Del_id	Sleutel afzonderlijk delict binnen proces-verbaal detail
Dkgrp_id	Sleutel delictkenmerkgroepen binnen afzonderlijke delicten
Dk_id	Sleutel delictkenmerken
Dk_code	Code delictkenmerk

Tabel: TWDKG (trefwoorden delictkenmerken)	
Pva_key	Sleutel proces-verbaal
Pvad_id	Sleutel proces-verbaal detail
Del_id	Sleutel afzonderlijk delict binnen proces-verbaal detail
Dkgrp_id	Sleutel delictkenmerkgroepen binnen afzonderlijke delicten
Dk_id	Sleutel delictkenmerken
Twdkg_id	Sleutel trefwoord delictkenmerk
Trefwoord1	Trefwoord 1 (object)
Trefwoord2	Trefwoord 2 (merk)
Trefwoord3	Trefwoord 3 (type)
Trefwoord4	Trefwoord 4 (kleur)

Bijlage 2 Overzicht van geselecteerde gegevensvelden aan de personenkant

Tabel: PERSOON (identiteit bekende verdachten)

Pers_id	Sleutel identiteit verdachte
---------	------------------------------

Tabel: ANT (antecedenten per bekende verdachte)

Pers_id	Sleutel identiteit verdachte
Ant_id	Sleutel antecedent
Ant_dat_yy	Datum antecedent: jaar
Ant_dat_mm	Datum antecedent: maand
Ant_dat_dd	Datum antecedent: dag
Soort_ant	Soort antecedent: automatisch, handmatig, voorlopig of sepot
Korps_code	Korpscode (pvo-sleutel)
Afdeling	Korpsafdeling (pvo-sleutel)
Jaar	Pva-jaar (pvo-sleutel)
Nummer	Pva-nummer (pvo-sleutel)

Tabel: DELGROEP (tellerbestand afzonderlijke delicten binnen antecedenten)

Pers_id	Sleutel identiteit verdachte
Ant_id	Sleutel antecedent
Delg_id	Sleutel afzonderlijk delict
Aantal_mal	Aantal malen dat delict gepleegd is

Tabel: DELIKT (wetsartikelen en feit modus-operandicode per afzonderlijk delict)

Pers_id	Sleutel identiteit verdachte
Ant_id	Sleutel antecedent
Delg_id	Sleutel afzonderlijk delict
Dlk_id	Sleutel wetsartikel/fmo-code
Wa_code	Wetsartikel
Fmo_code	Feit modus-operandicode

Tabel: NAAM (persoonsgegevens van bekende verdachten)

Pers_id	Sleutel identiteit verdachte
Naam_id	Sleutel naam
Geslacht	Geslacht
Geb_dat_yy	Geboortejaar
Geb_dat_mm	Geboortemaand
Geb_dat_dd	Geboortedag
Geb_land	Geboorteland

Tabel: NAT (nationaliteit bekende verdachten)

Pers_id	Sleutel identiteit verdachte
Naam_id	Sleutel naam
Nat_id	Sleutel nationaliteit
Nat_code	Code nationaliteit

Tabel: SIGNALEMEN (signalementen van bekende verdachten)

Pers_id	Sleutel identiteit verdachte
Sig_id	Sleutel signalement
Lengte	Lengte
Schoenmaat	Schoenmaat

Tabel: SK (signalementkenmerken)	
Pers_id	Sleutel identiteit verdachte
Sig_id	Sleutel signalement
Sk_id	Sleutel signalementkenmerk
Sk_code	Code signalementkenmerk

Tabel: SKTR (trefwoorden signalementkenmerken)	
Pers_id	Sleutel identiteit verdachte
Sig_id	Sleutel signalement
Sk_id	Sleutel signalementkenmerk
Sktr_id	Sleutel trefwoord signalementkenmerk
Trefwoord	Trefwoord signalementkenmerk

LET OP: bij een extractie aan de personenkant komen ook delictgegevens mee van de aangiftenkant. Het gaat om aangiften die gekoppeld zijn aan verdachten. De tabellen zijn inhoudelijk identiek aan de tabellen die hiervoor in bijlage 1 zijn besproken, alleen de sleutels zijn anders. Aan de personenkant zijn alle 'delicttabellen' gesleuteld op de velden 'pers_id' (identiteit van bekende verdachte) en 'bd_id' (teller van gekoppelde delicten per verdachte). De gegevens aan de personenkant worden via de tabel BD gekoppeld aan de delictkenmerken (deze tabel is hier niet weergegeven). De koppelingsprocedure wordt hier niet beschreven.

Bijlage 3 Overzicht van delictkenmerken die gebruikt zijn bij zoekmethode 1

- cbs-delictcategorie
- pleegplaats (\neq gemeente)
- wijk
- wijk-buurt-subbuurt
- straat
- tijdvak (= dagdeel volgens HKS-categorisering)
- maximaal 3 wetsartikelen en bijbehorende fmo-codes
- maximaal 25 delictkenmerkcodes
- trefwoorden die eventueel gekoppeld zijn aan deze delictkenmerkcodes

Bij de alternatieven 1 en 3 voor zoekmethode 1 is géén gebruik gemaakt van de volgende kenmerken:

- wijk-buurt-subbuurt
- straat
- trefwoorden verbonden aan delictkenmerkcodes