



UvA-DARE (Digital Academic Repository)

Formalizing the concepts of crimes and criminals

Elzinga, P.G.

Publication date
2011

[Link to publication](#)

Citation for published version (APA):

Elzinga, P. G. (2011). *Formalizing the concepts of crimes and criminals*. [Thesis, externally prepared, Universiteit van Amsterdam].

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

SAMENVATTING

In het kader van de leerstoel "Knowledge Discovery in Databases" hebben de Katholieke Universiteit Leuven en de Regiopolitie Amsterdam-Amstelland de afgelopen jaren een aantal nieuwe analysemethodes ontwikkeld. Deze analysemethodes hebben betrekking op domeinen als huiselijk geweld, mensenhandel en terreur.

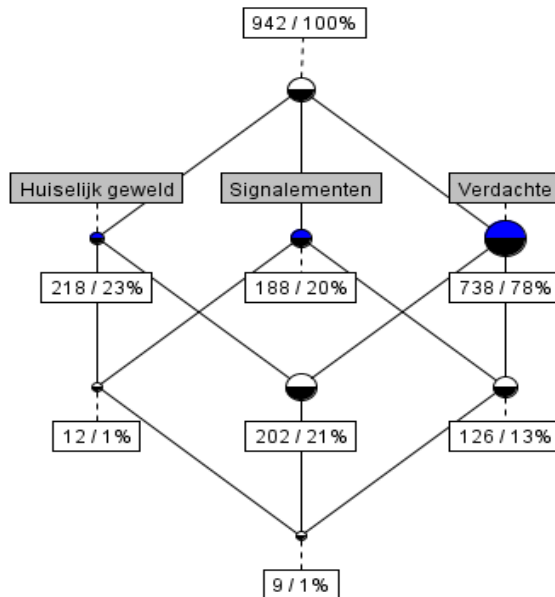
De ontwikkeling van de "Informatie of Intelligence Gestuurde Politie" heeft geleid tot een jaarlijkse toename van het aantal aandachtsvestigingen, algemene mutaties en overige meldingen binnen de BVH, het bedrijfsprocessensysteem dat bij alle politiekorpsen in Nederland in gebruik is. Het gaat hier om rapportages met eigen waarnemingen van de mensen op straat (het 'blauw') die worden opgeslagen als ongestructureerde tekst binnen de BVH. Tot nu toe werd er relatief weinig gedaan met de mogelijkheden die deze steeds groeiende, ongestructureerde, gegevensverzamelingen bieden om uit deze verzamelingen nieuwe gestructureerde informatie te genereren om het politiewerk beter te ondersteunen. Het hoofddoel van de samenwerking tussen de Regiopolitie Amsterdam-Amstelland en de Katholieke Universiteit Leuven werd het ontwikkelen van een nieuwe, efficiënte en operationeel inzetbare methode om bruikbare kennis uit deze grote hoeveelheden ongestructureerde informatie te onttrekken en toe te passen. Deze methodes moeten leiden tot een betere en snellere herkenning van (nieuwe) potentiële daders en slachtoffers. Voor dit doel is de afgelopen drie jaar gewerkt aan een drietal projecten: huiselijk geweld, mensenhandel (sexuele uitbuiting) en terrorisme (moslim radicalisering). Gedurende dit onderzoek is een toolbox ontwikkeld, Concept Relation Discovery and Innovation Enabling Technology (CORDIET). Aan de basis van deze toolbox ligt de C-K theorie van Hachtuell et al. (1999, 2002 en 2004) welke transitiestappen bevat voor het verkennen van bestaande en ontdekken en toepassen van nieuwe kennis. Belangrijk bij de transitiestappen is de rol van de onderzoeker. Deze moet bij elke stap de waarde van de informatie beoordelen en beslissingen nemen welke informatie meegenomen moet worden naar de volgende transitiestap. De transitieprocessen kunnen gezien worden als kennisexploratiestappen waarbij elke stap leidt tot het concretiseren en het operationaliseren van de verworven kennis. Deze werkwijze sluit nauw aan bij het proces van informatiegestuurde politie.

2. Huiselijk geweld

Het eerste project ging van start in 2007 en had als doel om een automatische detectie van huiselijk geweld binnen de BVH database mogelijk te maken. De oorspronkelijk uit de wiskunde afkomstige techniek "formele conceptanalyse" (Wille 1982, Ganter et al. 1999) waarin data geanalyseerd worden met behulp van conceptgrafën, werd gebruikt om interactief de onderliggende concepten en eigenschappen van huiselijk geweld (Van Dijk 1997) af te bakenen. De eigenschappen van huiselijk geweld werden weergegeven in de vorm van indicatoren die bestaan uit woorden en/of combinaties van woorden. De open source tool Lucene werd gebruikt om de tekstuele rapporten te indexeren met deze termen

SAMENVATTING

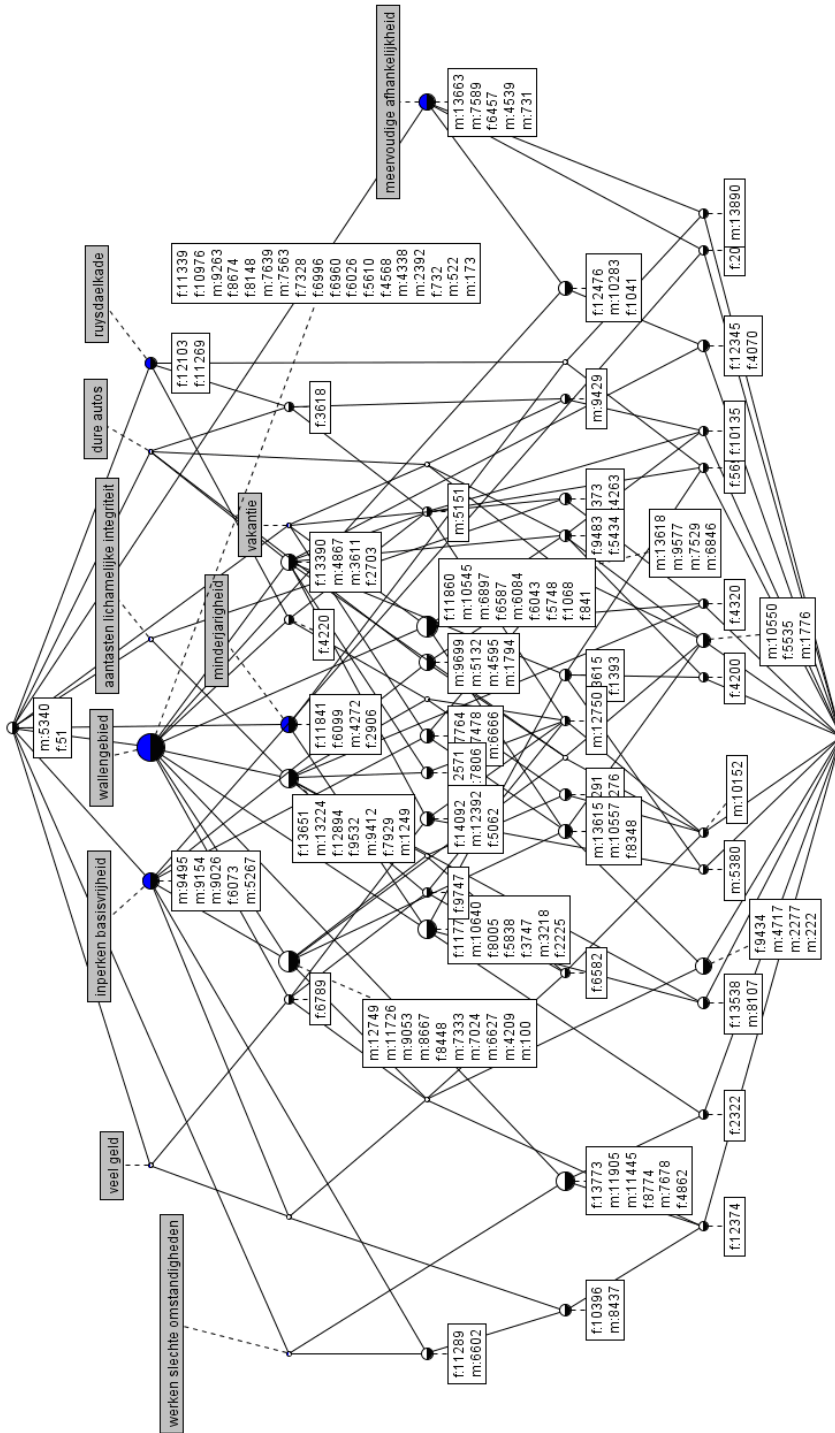
en zinnen. Met behulp van de visualisatie van de conceptgrafen op basis van de indicatoren en BVH-zaken werd het mogelijk kennisregels te ontdekken. Het proces van samenstellen van de indicatoren en kennisregels had tot gevolg dat de definitie van huiselijk geweld verder verfijnd kon worden. Zo konden situaties ontdekt worden die door rapporteurs als verwarrend werden beschouwd. Ook kwamen talloze foutief als huiselijk geweld aangemerkte zaken boven water. Dit onderzoek heeft geresulteerd in een nieuw kennisregel-gebaseerd systeem dat zaken met huiselijk geweld uit de BVH selecteert (Poelmans et al. 2009, Elzinga et al. 2009). Op dit moment wordt binnen de Regiopolitie Amsterdam-Amstelland dit kennisregel-gebaseerde systeem toegepast in combinatie met nTrueblue, het landelijke beheersysteem voor gegevenskwaliteit. Dit kennisregel-gebaseerde systeem kan overigens ook worden toegepast om andere zaken te selecteren, zoals in dit onderzoek is gedaan voor terrorisme en mensenhandel. Onderstaand figuur geeft een voorbeeld van een visualisatie van een “formele conceptanalyse” van mogelijk foutief geclassificeerde zaken van huiselijk geweld in de vorm van een conceptgraaf. De knopen in de graaf geven de concepten weer. Elk concept bestaat uit twee delen: een objecten- en een attributenverzameling. De cijfers in de witte kaders geven het aantal objecten weer dat tot dat concept behoort. De attributen staan vermeld in de grijze kaders. Een concept heeft een attribuut als we vertrekkend van de bijhorende knoop, enkel de lijnen naar boven volgen en bij dit attribuut kunnen uitkomen. De graaf in de onderstaande figuur kunnen we bijvoorbeeld op de volgende manier aflezen. Neem de knoop helemaal onderaan, dit concept bevat 9 politierapporten. Volgen we de lijnen naar boven, dan komen we uit bij de attributen “huiselijk geweld”, “signalementen” en “verdachte”.



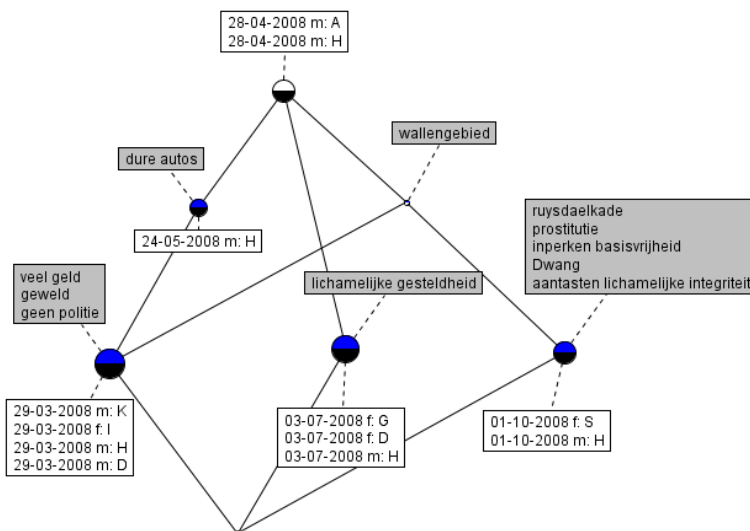
Van de 218 huiselijk geweld zaken zijn er 202 zaken (rechts naar beneden) waarbij een verdachte genoemd wordt. Verder is te zien dat er 9 zaken zijn die als huiselijk geweld gelabeld zijn, waarbij zowel een verdachte genoemd wordt als een signalement aanwezig is. Nader onderzoek leert dat van deze verdachten geen vaste woon- en/of verblijfplaats bekend is en dat een opsporingsbericht is uitgegaan. Dan blijven er nog 3 zaken van huiselijk geweld over waar een signalement voorkomt en geen verdachte wordt genoemd. Al deze 3 zaken bleken foutief als huiselijk geweld aangemerkt te zijn. Uit deze analyse kan een kennisregel afgeleid worden: dat van geweldszaken waarbij een signalement voorkomt, maar er geen verdachte wordt genoemd er met bijna 100% zekerheid gezegd kan worden dat het geen huiselijk geweld kan zijn.

3. Mensenhandel

De volgende stap is het toepassen van de kennisexploratietechniek “formele conceptanalyse” om (nieuwe) potentiële verdachten en slachtoffers te herkennen en te profileren. Het eerste domein was mensenhandel met als motief sexuele uitbuiting van het slachtoffer, een veel voorkomend misdrijf waar de aangiftebereidheid zeer laag ligt (Poelmans et al. 2010a, Highes 2000). Nadat de fase van het samenstellen van de relevante indicatoren is doorlopen, kan met deze methode van een potentiële verdachte of slachtoffer een gedetailleerd profiel gegenereerd worden met daarin de datum van observatie, de indicatoren en de contacten met andere betrokkenen. De eerste stap is het herkennen van potentiële verdachten en slachtoffers. In deze figuur zijn de namen geanonimiseerd en is voor de leesbaarheid een aantal indicatoren weggelaten.



De personen (f = female en m = male) onderin de figuur komen het eerst in aanmerking als potentiële verdachte of slachtoffer aangezien personen lager in de graaf aan meer indicatoren voldoen. Van elke persoon uit deze figuur kan een afzonderlijke analyse gemaakt worden. Een selectie van een van de mannen links onderin de figuur levert de volgende “formele conceptanalyse” op:



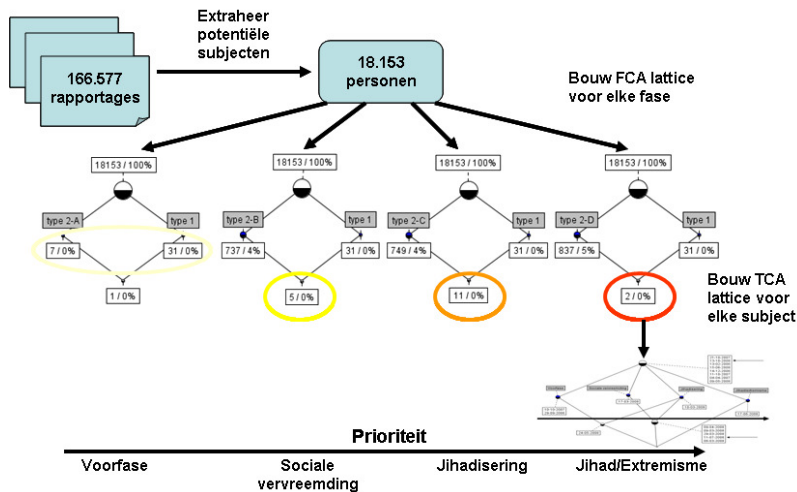
In deze figuur zijn verschillende tijdstippen van de waarnemingen weergegeven bij de indicatoren. De variant van “formele conceptanalyse” die gebruik maakt van temporele gegevens is de “temporele conceptanalyse” (Wolff 2005). Uit de figuur blijkt dat man D (4^e links onderin) mogelijk verantwoordelijk is voor de logistiek, omdat deze in een dure auto rijdt waarin de inzittenden gedrag vertonen dat ze liever niet met de politie in contact willen komen. De man H (in alle objecten voorkomend), is de mogelijke pooier, waarbij de vrouw S (1^e rechts bovenin) zijn vermoedelijke slachtoffer is, omdat hier sprake is van prostitutie onder dwang. Aan de hand van deze figuur kan in combinatie met de bijbehorende rapporten worden beoordeeld of een *27 constructie*, een document op basis van artikel 273a van het Wetboek van Strafrecht (Staatscourant 2006, 58) omtrent beleidregels opsporing/bevoegdheden mensenhandel, kan worden samengesteld. Dit is een document dat voorafgaat aan eventueel verder strafrechterlijk onderzoek tegen de man H.

4. Terrorisme

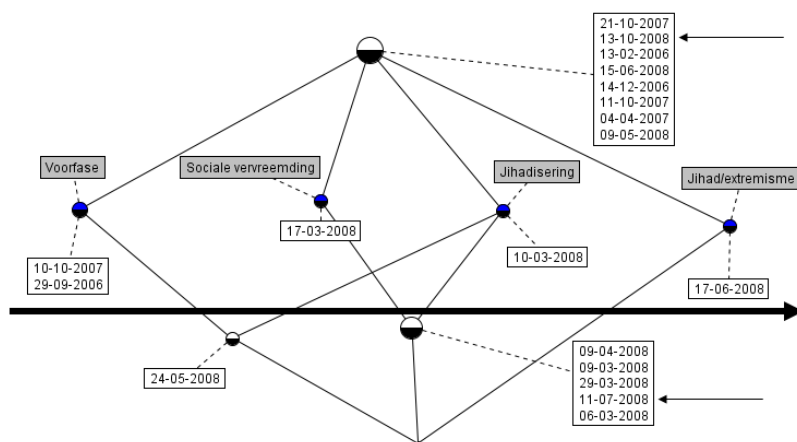
In het laatste project is samengewerkt met het projectteam Kennis in Modellen (KiM) van het Korps Landelijke Politie Diensten (KLPD). Daarbij werd de kennisexploratietechniek ingezet om het moslim radicaliseringmodel van KiM te gebruiken voor het actief opsporen van potentiële terreurverdachten (Elzinga et al. 2010, AIVD 2006). Ook hier bleek het toepassen van de kennisexploratietechniek van de formele conceptanalyse zeer bruikbaar. Waar bij mensenhandel gezocht werd naar profielen in de tijd, is er bij moslim radicalisering sprake van een groeimodel,

SAMENVATTING

waarbij een potentiële verdachte verschillende fasen van radicalisering doorloopt. Het projectteam van KiM heeft op basis van expertonderzoeken een verzameling van 35 indicatoren samengesteld op grond waarvan een persoon in een bepaalde fase kan worden gepositioneerd. Samen met de KLPD is intensief gezocht naar kenmerkende woorden en woordcombinaties die de verschillende indicatoren kenmerken. Het verschil met de voorgaande modellen is dat het KiM-model een extra dimensie toevoegt in de vorm van het aantal verschillende indicatoren waaraan een radicaliseringeniveau dient te voldoen.



De analyse is uitgevoerd op de verzameling waarnemingen uit de Basis Voorziening Handhaving (BVH) van de Regiopolitie Amsterdam-Amstelland over de jaren 2006, 2007 en 2008 met als resultaat dat uit 166.577 rapporten 18.153 personen werden gevonden die aan minimaal 1 indicator voldoen. Uit deze 18.153 personen werden 38 personen gevonden die voldeden aan de 1^e fase van de radicalisering. Nadere analyse brengt aan het licht dat 19 terecht geselecteerd waren, waarbij 3 personen niet bij de Regiopolitie Amsterdam-Amstelland als zodanig bekend waren, maar wel bij de KLPD. Van deze 19 personen bleken er uiteindelijk 2 te voldoen aan minimale voorwaarden van de extremistische fase. Van een van deze personen is een profiel gemaakt van alle indicatoren verspreid over de tijd.



Uit deze figuur is af te leiden dat de betrokken persoon de extremistische fase heeft bereikt op 17 juni 2008 en na die tijd nog 2 keer is waargenomen door surveillanten (de 2 pijlen rechtsboven en rechtsonder in de figuur) op 11-07-2008 en 13-10-2008.

5. CORDIET

Steeds meer bedrijven beschikken over grote hoeveelheden ongestructureerde gegevens, veelal in tekstuele vorm. De weinige analyse-instrumenten die zich richten op dit probleemgebied bieden onvoldoende functionaliteit voor de specifieke behoeften van veel van deze organisaties. In het kader van het onderzoekswerk verricht in het doctoraatsonderzoek van Jonas Poelmans (Aspirant FWO²¹) werd in september 2010 gestart met de ontwikkeling van de toolset Concept Relation Discovery and Innovation Enabling Technology (CORDIET) in samenwerking met de Moscow Higher School of Economics. Onder toezicht van Prof. dr. Sergei Kuznetsov, drs. Paul Elzinga en dr. Jonas Poelmans werd een projectplan opgesteld, waar 20 master studenten, 2 doctoraatsonderzoekers, 2 postdoctorale onderzoekers en 2 professoren, allen uit Rusland afkomstig, actief aan deelnemen. Het resultaat van deze samenwerking zal de compleet ingerichte toolset CORDIET zijn, waaronder de succesvolle toepassing van deze toolset op ongestructureerde rapportages van de Regiopolitie Amsterdam-Amstelland en medische verslagen van de GZA ziekenhuizen.

Deze toolset zal ingezet worden in de doorlopende projecten voor de proactieve opsporing van mogelijk potentiële verdachten van terrorisme en mensenhandel in de politieregio Amsterdam-Amstelland. In Elzinga et al. (2010) werd al een proof of concept uitgevoerd die de kracht van onze aanpak met conceptgrafen en andere visualisatietechnieken zoals “emergent self organising maps” heeft aangetoond voor de opsporing van individuen die radicaliserend gedrag vertonen. Gedurende dit onderzoek werd een aantal mogelijke verdachten en slachtoffers van mensenhandel

²¹ FWO: Fonds voor Wetenschappelijk Onderzoek - Vlaanderen

SAMENVATTING

geanalyseerd en geprofileerd (Poelmans et al. 2010c). Deze toolset biedt de mogelijkheid om veel sneller en gedetailleerder data analyses uit te voeren en relevante personen uit politiegegevens te distilleren. De werkwijze van deze toolset past niet alleen in de filosofie van de Informatie Gestuurde Politie maar past ook binnen een ziekenhuiscontext waar de behandelingsgegevens van borstkankerpatiënten werden geanalyseerd om de verstrekte zorg te verbeteren (Poelmans et al. 2010d). Ook in de GZA ziekenhuisgroep zal deze toolset in een project ingezet worden om de meer dan 43 actieve zorgpaden voor 75 zorgprocessen te verbeteren. Over dit thema is door de Katholieke Universiteit Leuven en de Moscow Higher School of Economics in de zomer van 2011 een workshop georganiseerd met als titel “Concept Discovery in Unstructured Data”²². Samen met de Regiopolitie Amsterdam-Amstelland zal worden onderzocht of CORDIET kan worden ingezet voor het voorspellen van criminele carrières van potentiële beroepscriminelen.

De architectuur van de CORDIET toolset bevat 3 lagen. De database laag bevat zowel de data opslag, alsook de ontologie: de tekstdocumenten worden geïndexeerd met Lucene en de ontologie elementen in xml formaat worden vertaald naar Lucene syntax. In de middelste laag worden de “formele conceptanalyse”, “temporele conceptanalyse”, “emergent self organizing maps”, “hidden Markov modellen” en tekstanalysecomponenten gebruikt om visuele modellen te genereren op basis van de data en de ontologie. De derde laag bevat de presentatielaag met de grafische gebruikersinterface.

De grafische gebruikersinterface wordt op een manier ontwikkeld die het toelaat om complexe analyses uit te voeren door mensen met weinig kennis van statistiek en data analyses. In de ontologie kunnen tekstmining attributen gedefinieerd worden om de documenten te analyseren. Temporele attributen kunnen helpen bij het ontdekken van verbanden over de tijd. Samengestelde attributen laten toe om complexe attributen te creëren uit de tekstmining en temporele attributen met behulp van eerste orde logica. Voor deze specifieke ontologische structuren en de bijhorende persistentie (data-opslag) worden nieuwe XML structuren gedefinieerd. Parsers dienen ontwikkeld te worden om de werkomgeving te verbinden aan traditionele data-opslag (SQL databases) en datawarehouse systemen. De modellen gegenereerd met de componenten uit de middelste laag zullen als volgt gebruikt worden:

- “formele conceptanalyse” conceptgrafen: opsporen van verdachten van mensenhandel, terreur, huiselijk geweld etc.
- “temporele conceptanalyse” conceptgrafen: visueel profiel van potentiële daders creëren en interessante patiënten
- “hidden Markov modellen”: in kaart brengen van zorgpaden en criminele carrières
- “emergent self organizing maps”: in combinatie met “formele

²² Concept Discovery in Unstructured Data 2011: <http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-757/>

conceptanalyse” de data exploreren

Wij willen nog als bijzonderheid vermelden dat elk van deze vier technieken afzonderlijk weliswaar in één of meerdere statistische omgevingen zijn geïmplementeerd zoals Matlab en SPSS, maar deze technieken zijn nog nooit eerder samen in één omgeving gecombineerd en geïntegreerd. Het gevolg is dat analyses met CORDIET op een veel grotere schaal, veel sneller en efficiënter kunnen worden toegepast. De user interface maakt mogelijk de ontologie elementen visueel te wijzigen via een graaf. Ook de modellen kunnen eenvoudig gegenereerd en geanalyseerd worden. Bovendien zullen verschillende uitbreidingen voor FCA worden opgenomen, vooral metrieken zoals concept stability, etc.

6. Conclusies

De drie projecten die uitgevoerd zijn in het kader van de leerstoel geven de potentie aan van de kennisexploratietechniek “formele conceptanalyse”. Voornamelijk de intuïtief interpreteerbare visuele voorstelling werd van groot belang gevonden door de informatiespecialisten binnen de politie op zowel de strategische als de tactische en operationele niveaus. Deze visualisatie liet niet alleen toe om interactief de data te verkennen en te analyseren maar ook om de onderliggende concepten van de probleemdomen in kaart te brengen. Zo werden onder andere nieuwe concepten, anomalieën, verwarrende situaties en foutief gelabelde zaken ontdekt, maar ook bij de politie niet bekende subjecten die mogelijk betrokken zijn bij mensenhandel of terroristische activiteiten. Ook de temporele variant van de “formele conceptanalyse” bleek van groot nut te zijn bij het profileren van verdachten en hun evolutie over tijd. Niet eerder werden ongestructureerde informatiebronnen ontsloten op een wijze waarop nieuwe inzichten, nieuwe verdachten en slachtoffers zichtbaar werden. Om deze reden zal de “formele conceptanalyse” in de nabije toekomst een belangrijk instrument gaan vormen voor de informatiespecialisten binnen de politie en zal essentieel gaan bijdragen aan de vorming van Intelligence binnen de Nederlandse politie.