Formally analysing the concepts of domestic violence

Jonas Poelmans[1], Paul Elzinga[3], Stijn Viaene[1,2], Guido Dedene[1,4]

[1]K.U.Leuven, Faculty of Business and Economics, Naamsestraat 69,
3000 Leuven, Belgium
[2]Vlerick Leuven Gent Management School, Vlamingenstraat 83,
3000 Leuven, Belgium
[3]Amsterdam-Amstelland Police, James Wattstraat 84,
1000 CG Amsterdam, The Netherlands
[4]Universiteit van Amsterdam Business School, Roetersstraat 11
1018 WB Amsterdam, The Netherlands
{Jonas.Poelmans, Stijn.Viaene, Guido.Dedene}@econ.kuleuven.be
Paul.Elzinga@amsterdam.politie.nl

Abstract. The types of police inquiries performed these days are incredibly diverse. Often data processing architectures are not suited to cope with this diversity since most of the case data is still stored as unstructured text. In this paper Formal Concept Analysis (FCA) is showcased for its exploratory data analysis capabilities in discovering domestic violence intelligence from a dataset of unstructured police reports filed with the regional police Amsterdam-Amstelland in the Netherlands. From this data analysis it is shown that FCA can be a powerful instrument to operationally improve policing practice. For one, it is shown that the definition of domestic violence employed by the police is not always as clear as it should be, making it hard to use it effectively for classification purposes. In addition, this paper presents newly discovered knowledge for automatically classifying certain cases as either domestic or non-domestic violence is. Moreover, it provides practical advice for detecting incorrect classifications performed by police officers. A final aspect to be discussed is the problems encountered because of the sometimes unstructured way of working of police officers. The added value of this paper resides in both using FCA for exploratory data analysis, as well as with the application of FCA for the detection of domestic violence.

Keywords: Formal Concept Analysis (FCA), domestic violence, knowledge discovery in databases, text mining, exploratory data analysis, knowledge enrichment, concept discovery

1 Introduction

Concept discovery is a relatively new approach for discovering knowledge from textual information [10]. At the core of the method is the visualization of the underlying concepts of the data by means of Formal Concept Analysis (FCA) lattices [8, 9] which are interpreted, analysed and discussed by domain experts. FCA arose twenty-five years ago as a mathematical theory [14] and has over the years grown into a powerful framework for data analysis, data visualization [15], information retrieval and text mining [16, 17, 20]. In this paper FCA is for the first time used as an exploratory data analysis and knowledge enrichment technique for police data. Compared to traditional black-box data mining techniques, this human-centred approach has the advantage of actively engaging expert knowledge in the discovery process.

The goal of Intelligence Led Policing (ILP) is to complement intuition led police actions with information coming from analyses on aggregated operational data, such as crime figures and criminal characteristics [25, 26, 39]. While over 80% of all information available to police organizations

resides in textual form, analysis has to date been primarily focused on the structured portion of the available data. Though text mining has been identified as a promising area in the formal framework for crime data mining by Chen et al. [27], this work has hardly found its way into mainstream scientific literature. One of the notorious exceptions is the paper by Ananyan [28] in which historical police reports were analysed to identify hidden patterns.

According to the Ministry of Justice of the Netherlands, 45% of the population once fell victim to non-incidental domestic violence and for 27% of the population, the incidents even occurred on a weekly or daily basis [22]. These gloomy statistics brought this topic to the centre of the political agenda and made it to one of the pivotal projects of the Balkenende administration when it took office in 2003[1] and the Amsterdam-Amstelland police in the Netherlands [32]. Sufficient insight into the nature of domestic violence, being able to swiftly recognise suspicious cases and label reports accordingly is of the utmost importance. However, in the past intensive audits of the police databases related to filed reports established that many reports tended to be wrongly labelled as domestic or as non-domestic violence cases.

In this paper we shall demonstrate the effectiveness of concept discovery methods for distilling new knowledge from the unstructured text in police reports. FCA amongst others helped us to improve the definition, the understanding by police officers and the management of the notion of domestic violence. Additionally, we aim at automating detection of domestic violence from the unstructured text in police reports. The very first steps taken in this direction are described in [37] and in [38] an independent research track pursued in parallel with the work presented in this paper based on Emergent Self Organizing Maps is described. Although the usage of FCA for browsing text collections has been suggested before by Cole et al. [18, 35], almost none of these papers have focused on how FCA can be used for knowledge enrichment and for discovering different types of knowledge in unstructured text. Neither has it been thoroughly discussed in the literature how FCA can be used to incrementally construct and refine a high-quality domain-specific thesaurus (which is a prerequisite for developing an effective information retrieval system). Moreover, only minor attention has been paid to the possibilities offered by FCA to incorporate prior knowledge in the knowledge discovery

[1] http://www.regering.nl/Het_kabinet/Eerdere_kabinetten/Kabinet_Balkenende_II/Regeerakkoord#internelink4

process. Finally, some of the aspects of this paper have already been discussed in the literature in a fragmented way (e.g. information retrieval, knowledge browsing), but an integrated approach has never been pursued.

FCA is particularly suited for exploratory data analysis because of its human-centredness. Representations that expose the underlying conceptual structure of the information promote the creation of new knowledge. What makes FCA an especially appealing technique for knowledge discovery in databases from a practitioner's point of view is the compactness of its information representation and the minimal need for users to tune (hyper-) parameters to distill a useful, actionable picture of the mining exercise. Concepts are the elementary units of human reasoning and this notion of concept is central to FCA [23, 24]. The underlying structure of the information is considered to be a concept system and FCA concept lattices are used to visualize the concepts and their interrelationships. These visual representations support human actors in their information discovery and knowledge creation exercise.

This paper is composed as follows. In section 2 we describe the current situation of the domestic violence reporting procedure and previous attempts to improve the situation. In section 3 we cover the essentials of FCA theory, introducing the pivotal FCA notions of concept and concept lattice and describing the process of FCA for knowledge discovery. Section 4 elaborates on the dataset used in our research, while section 5 focuses on how this dataset was analysed and discusses the results of the application of FCA for exploratory analysis of domestic violence cases using this dataset. In section 6 the results of the domain exploration are validated. Finally, section 7 presents a number of concluding remarks.

2 Domestic violence discovery

According to the U.S. Office on Violence against Women, domestic violence is a "*pattern of abusive behavior in any relationship that is used by one partner to gain or maintain power and control over another intimate partner*" [1]. Domestic violence can take the form of physical violence, which includes biting, pushing, maltreating, stabbing or even killing the victim. Physical violence is often accompanied by mental or emotional abuse, which includes insults and verbal threats of physical violence towards the victim, the self or others, including children. Domestic violence occurs all over

the world, in various cultures [2] and affects people throughout society, irrespective of economic status [3].

2.1 Current situation

The XPol database – the database of the Amsterdam-Amstelland police – contains most of the documents with regard to criminal offences. Documents related to certain types of crime receive corresponding labels. It is of the utmost importance that a correct label is assigned to each of the filed police reports. First, there are some legal consequences. If the police judged an incident to be domestic violence, the public prosecutor can accuse the offender of committing a domestic violence crime. This is taken into account by the judge as an aggravating circumstance, often resulting in a more severe penalty. Second, police officers will be able to better assess new incidents between the perpetrator and the victim, resulting in a more effective way of tackling the problem. Finally, if a domestic violence label was incorrectly assigned to a case, this will result in a waste of the valuable time of the police officers assigned to the case.

Immediately after the reporting of a crime, police officers are given the possibility to judge whether or not it is a domestic violence case. If they believe it is, they can indicate this by assigning the label "domestic violence" to the report. However, not all domestic violence cases are recognised as such by police officers. This may have several reasons, for example, because of a lack of training, a lack of prior experience or new types of domestic violence occurring. As a consequence, many documents are lacking the appropriate label, which put on the agenda the need for a more efficient and effective case triage software program to automatically filter out suspicious cases for in-depth, manual inspection and classification. The in-place case triage system has been configured to filter out these reports for in-depth manual inspection and classification, with the aim of substantially reducing the number of domestic violence cases that are not recognised as such. It retrieves suspicious cases that lack the label of domestic violence and sends them back to the data quality management team. At present, each case retrieved by the in-place case triage system is subjected to an in-depth manual inspection by one of the co-workers of the quality control department. If analysis reveals that a case was wrongly classified as non-domestic violence, it is sent back to the police officer responsible for the case, who is obliged to re-examine and reclassify the police report. It is obvious that this is a very time-consuming and, by

consequence, costly procedure. Given that it takes an individual at least five minutes to read and classify a case, it is clear that more accurate triage will result in major savings.

Currently the triage is based on either one or both of the following two criteria being met. The first criterion is whether the perpetrator and the victim live at the same address. The second criterion is whether any or a combination of the following expressions appear in the case documents: "ex-boyfriend", "ex-girlfriend", "ex-husband", "ex-wife", "domestic", "stalk", "lived together", "live together", "son and scared", "child and scared", "child and threat", "son and threat", "daughter and threat" or "daughter and scared".

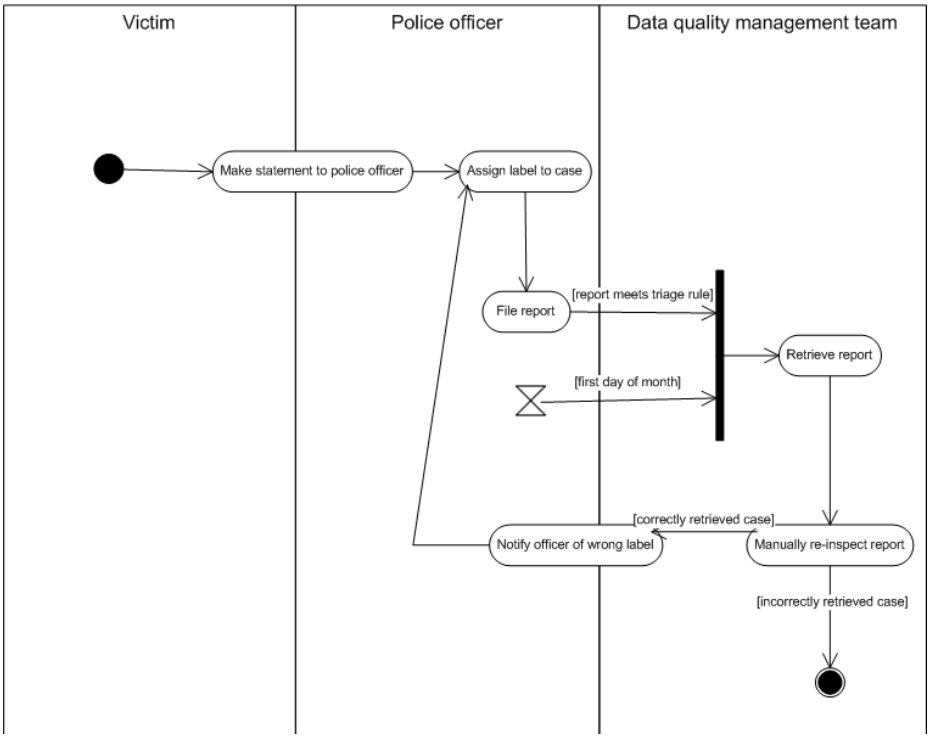

Fig. 1. Current domestic violence reporting procedure

A summary of the current domestic violence reporting procedure is displayed in Figure 1. There are several problems associated with this process. First, recent audits have confirmed that many of the retrieved cases are wrongly selected for in-depth manual inspection. Going back to 2006, the system retrieved 1157 cases, 80% of which actually turned out to be non-domestic violence cases. For example, going back to 2007, the triage system retrieved 1091 of such cases in which the victim made

a statement to the police. Second, because of a lack of manpower the data management quality team was not able to analyse each retrieved police report. Third, audits of the police databases revealed that not all domestic violence cases lacking the appropriate label were retrieved by the case triage system. Fourth, no actions have yet been undertaken to address the issue of the filed reports that were wrongly classified as domestic violence.

2.2 Previous attempts to resolve situation

Previous attempts have mainly focused on developing a machine learning classifier that automatically classified cases as domestic or as non-domestic violence. Previously developed systems were mainly multi-layer perceptions that were trained on a dataset consisting of cases that were labelled by police officers as domestic or as non-domestic violence. These systems did not provide any insight into the problem, since they are black-boxes and their performance was around 80% only [31]. As a consequence, these systems never made it into operational policing practice. We found that a critical error was that the developers never performed an in-depth exploration of the data. They overlooked the complexity of the notion of domestic violence, were unaware that different people have different visions about the nature and scope of it and did not pay attention to niche cases. Moreover, the correctness of the labels assigned to cases by police officers was never verified. We found that different police officers regularly assigned different labels to the same situation. Finally, the developers did not dispose of a high-quality domain-specific thesaurus that contained sufficient discriminant terms for accurately classifying cases.

3 FCA knowledge discovery process

This section introduces the main ideas of FCA and how it was used during the knowledge discovery process. According to R.S. Brachman and T. Anand [29], much attention and effort has been focused on the development of data mining techniques, but only a minor effort has been devoted to the development of tools that support the analyst in the overall discovery task. They argue for a more human-centred approach. Human-centred KDD refers to the constitutive character of human interpretation for the discovery of knowledge, and stresses the complex, interactive process of KDD as

being led by human thought. In most real-world knowledge discovery applications, an indispensable part of the discovery process is that the analyst explores and sifts through the raw data to become familiar with it and to get a feel for what the data may cover. Often an explicit specification of what one is looking for only arises during an interactive process of data exploration, analysis and segmentation. R.S. Brachman et al. [30] introduce the notion of data archeology for KDD tasks in which a precise specification of the discovery strategy, the crucial questions and the basic goals of the task have to be elaborated during an unpredictable exploration of the data. Data archeology can be considered as a highly human-centred process of asking, exploring, analysing, interpreting and learning by interacting with the underlying database. Comprehensible support should be provided to the analyst during the KDD process. According to Brachman et al. [29] this should be embedded into a knowledge discovery support environment. How the process of human-centred KDD can be supported by Formal Concept Analysis (FCA) was for the first time investigated by Stumme et al. [12].

Smyth et al. [33] already stated that the algorithm designer and the scientist should be able to bring in prior knowledge so the data mining algorithm does not just rediscover what is already known. Moreover, the scientist should be able to "get inside" and "steer" the direction of the data mining algorithm. FCA fulfils these requirements. Starting from initial knowledge on the problem area, it provides the user with a visual display of the relevant concepts available in the dataset and their relationships. Additionally, the user can visually interact with the concept lattice and thereby steer the knowledge discovery process.

What makes FCA into an especially appealing technique for knowledge discovery in databases is that it meets the important requirement stated by, amongst others, Fayyad et al. [34] that data mining should be primarily concerned with making it easy, convenient and practical to explore very large databases for organizations and users with vast amounts of data but without years of training as data analysts. FCA offers the user an intuitive visual display of different types of structures available in the dataset and guides the user in the exploration of the dataset. This end-user-friendly interface also makes the data mining more transparent to the user.

When compared to other, more traditional, techniques such as associates rules, FCA has a larger explanatory power because of its underlying non-hierarchical structure [36]. While traditional

association rules are flat, FCA provides an order of significance, which makes its representation richer and more intuitive to use.

3.1 FCA essentials

Formal Concept Analysis is a recent mathematical technique that can be used as an unsupervised clustering technique [11, 13]. Police reports containing terms from the same term clusters are grouped in concepts. The starting point of the analysis is a database table consisting of rows M (i.e. objects), columns F (i.e. attributes) and crosses $T \subseteq M \times F$ (i.e. relationships between objects and attributes). The mathematical structure used to represent such a cross table is called a formal context (*T, M, F*). An example of a cross table is displayed in Table 1. In this table reports of domestic violence (i.e. the objects) are related (i.e. the crosses) to a number of terms (i.e. the attributes); here a report is related to a term if the report contains this term. The dataset in Table 1 is an excerpt of the one we used in our research. Given a formal context, FCA then derives all concepts from this context and orders them according to a subconcept-superconcept relation, which results in a line diagram (a.k.a. lattice).

Table 1. Example of a formal context

	kicking	dad hits me	stabbing	cursing	scratching	maltreating
report 1	X	X				X
report 2			X	X	X	
report 3	X	X	X	X	X	
report 4						X
report 5				X	X	

The notion of concept is central to FCA. The way FCA looks at concepts is in line with the international standard ISO 704, which formulates the following definition. A concept is considered to be a unit of thought constituted of two parts: its extension and its intension, [14, 16]. The extension consists of all objects belonging to the concept, while the intension comprises all attributes shared by those objects. Let us illustrate the notion of concept of a formal context using the data in Table 1. For a set of objects $O \subseteq M$, the common features, written $\sigma(O)$, can be identified via the following formula:

$$A = \sigma(O) = \{ f \in F \mid \forall o \in O : (o, f) \in T \}$$

Take the attributes that describe report 5 in Table 1, for example. By collecting all reports of this context that share these attributes, we get to a set $O \subseteq M$ consisting of reports 2, 3 and 5. This set O of objects is closely connected to set A consisting of the attributes "cursing" and "scratching."

$$O = \tau(A) = \{i \in M \mid \forall f \in A : (i, f) \in T\}$$

That is, O is the set of all objects sharing all attributes of A, and A is the set of all attributes that are valid descriptions for all the objects contained in O. Each such pair (O, A) is called a formal concept (or concept) of the given context. The set $A = \sigma(O)$ is called the intent, while $O = \tau(A)$ is called the extent of the concept (O, A).

There is a natural hierarchical ordering relation between the concepts of a given context that is called the subconcept-superconcept relation.

$$(O_1, A_1) \subseteq (O_2, A_2) \Leftrightarrow (O_1 \subseteq O_2 \Leftrightarrow A_2 \subseteq A_1)$$

A concept $d = (O_1, A_1)$ is called a subconcept of a concept $e = (O_2, A_2)$ (or equivalently, e is called a superconcept of a concept d) if the extent of d is a subset of the extent of e (or equivalently, if the intent of d is a superset of the intent of e). For example, the concept with intent "cursing", "scratching" and "stabbing" is a subconcept of a concept with intent "cursing" and "scratching." With reference to Table 1, the extent of the latter is composed of reports 2, 3 and 5, while the extent of the former is composed of reports 2 and 3.

The set of all concepts of a formal context combined with the subconcept-superconcept relation defined for these concepts gives rise to the mathematical structure of a complete lattice, called the concept lattice of the context, which is made accessible to human reasoning by using the representation of a (labelled) line diagram. The line diagram in Figure 1, for example, is a compact representation of the concept lattice of the formal context abstracted from Table 1. The circles or nodes in this line diagram represent the formal concepts. It displays only concepts that describe objects and is therefore a subpart of the concept lattice. The shaded boxes (upward) linked to a node represent the attributes used to name the concept. The non-shaded boxes (downward) linked to a node represent the objects used to name the concept. The information contained in the formal context of Table 1 can be distilled from the line diagram in Figure 1 by applying the following reading rule: an object "g" is

described by an attribute “m” if and only if there is an ascending path from the node named by “g” to the node named by “m”. For example, report 5 is described by the attributes “cursing” and “scratching”.

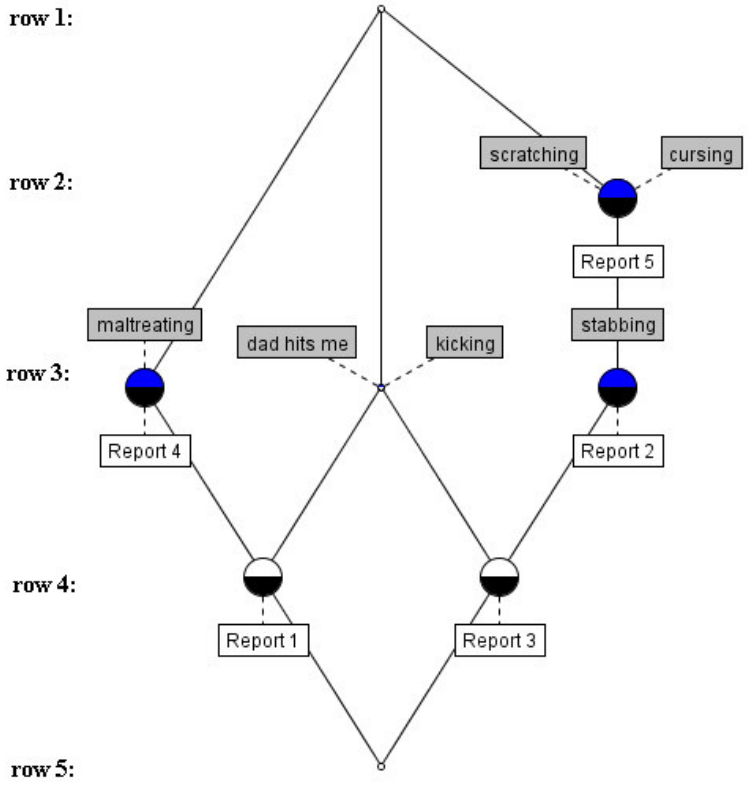

Fig. 2. Line diagram corresponding to the context from Table 1

Retrieving the extension of a formal concept from a line diagram such as the one in Figure 2 implies collecting all objects on all paths leading down from the corresponding node. In this example, the objects associated with the third concept in row 3 are reports 2 and 3. To retrieve the intension of a formal concept, one traces all paths leading up from the corresponding node in order to collect all attributes. In this example the third concept in row 3 is defined by the attributes “stabbing”, “cursing” and “scratching”. The top and bottom concepts in the lattice are special: the top concept contains all objects in its extension, whereas the bottom concept contains all attributes in its intension. A concept is a subconcept of all concepts that can be reached by travelling upward. This concept will inherit all

attributes associated with these superconcepts. Note that the extension of the concept with attributes "kicking" and "dad hits me" is empty. This does not mean that there is no report that contains these attributes. However, it does mean that there is no report containing only these two attributes.

3.2 Human-centred knowledge discovery with FCA

In contrast to most data mining algorithms, the discovery process using FCA is human-centred. It is definitely not a black-box that runs and optimises without intervention beyond specifying initial model choices and parameters. During the mining process two persons, an exploratory data analyst and a domain expert, were the driving force behind the exploration and collaborated intensively. There was a continuous process of iterating back and forth between the FCA lattices and the police reports. This knowledge discovery process is summarised in Figure 3. It is an abstract description of the methodology that is displayed here, but this process will be exemplified in the results section.

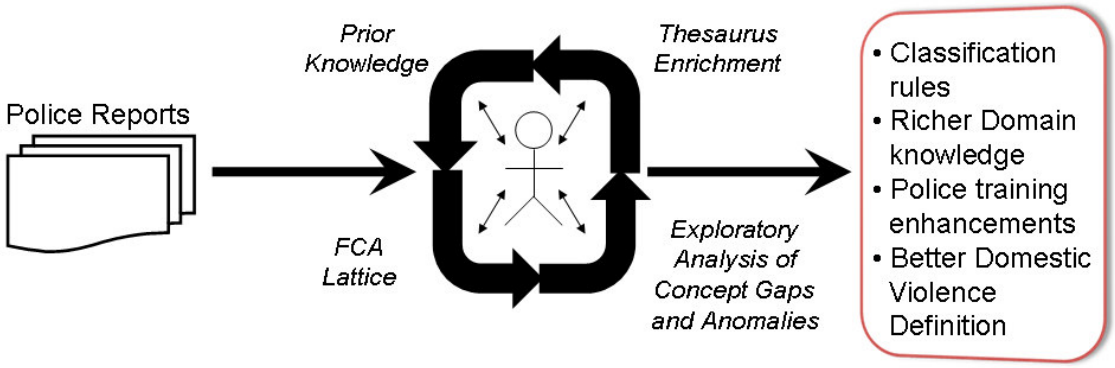

Fig. 3. Abstract human-centered FCA knowledge discovery process

The process of using FCA for exploratory data analysis consists basically of iteratively applying the following process. A lattice is constructed by the exploratory data analyst based on the domain expert's prior knowledge of the problem area, the police reports contained in the dataset and the terms contained in the thesaurus. The lattice provides a reduced search space to the domain expert, who then visually inspects and analyses this lattice paying special attention to anomalies and counter-intuitive facts. The latter provide a clear guideline to the exploratory data analyst and the domain expert in order for them to pursue their data exploration. The obtained results, together with the relevant prior

knowledge of the domain expert, are then incorporated into the existing visual representation, resulting in a new lattice.

The FCA lattice can be considered as a knowledge browser. Our contention is that it allows for an effective interaction between the human actors and the underlying information. The focus of the use of the FCA technique is on truly gaining incremental insight into the problem area by optimally incorporating prior domain knowledge in learning cycles. This insight encompasses an enrichment as well as a validation of the correctness and the practical usefulness of existing prior knowledge. Additionally, FCA is used to enrich and refine the domain-specific thesaurus. This thesaurus plays a key role in the incremental knowledge discovery germane to our research. FCA is also used to discover missing values and inconstancies from police reports. Finally, FCA is used to investigate some important, significant aspects of operational policing practice concerning domestic violence cases and to discover accurate and comprehensible classification rules. Each of these aspects of the process will be described in more detail in section 5, where we comment on the empirical analysis and results.

4 Dataset

The dataset we report on in this paper consists of a selection of 4814 police reports describing a whole range of violent incidents from the year 2007. The domestic violence cases for that period are a subset of this dataset. The 1091 cases selected by the in-place case triage system for 2007 are a subset of this dataset too. This latter selection came about by, amongst other things, filtering from a larger set those police reports that did not contain the reporting of a crime by a victim, which is necessary for establishing domestic violence. This happens, for example, when a police officer is sent to an incident and later on writes a report in which he/she mentions his/her findings, while the victim has not made an official statement to the police. The follow-up reports referring to previous cases were also removed from the initial set of reports. Ultimately, this gave rise to a set of 4814 reports that were used as input for our investigation. From these reports, the person who reported the crime, the suspect, the persons involved in the crime, the witnesses, the project code and the statement made by the victim

to the police were extracted. Of the 4814 reports, 1657 were classified as domestic violence; the others were not. An example of a report is displayed in Figure 4.

Title of incident	Violent incident xxx
Reporting date	26-11-2007
Project code	Domestic violence against seniors (+55)
Crime location	Amsterdam Keizersgracht yyy
Suspect (male) Suspect (18-45yr)	zzz
Address	Amsterdam Keizersgracht yyy
Involved (male) Involved (18-45yr)	Neighbours
Address	Amsterdam Keizersgracht www
Victim (female) Victim (older than 45yr)	uuu
Address	Amsterdam Keizersgracht vvv

Reporting of the crime

Last night I was attacked by my husband. I was watching television in the living room when he suddenly attacked me with a knife. I fell on the floor. Then he tried to kick me in my stomach. I tried to escape through the back door while I was yelling for help. I ran to the neighbours for help. They called the emergency services. Meanwhile my son ran away. My leg was bleeding; my head was bouncing, etc.

Fig. 4. Example police report

The validation set consists of a selection of 4738 cases describing a whole range of violent incidents from the year 2006 where the victim made a statement to the police. Again, the follow-up reports were first removed. Of these 4738 cases 1734 were classified as domestic violence by police officers. In 2006 the in-place case triage system retrieved 1157 police reports containing a statement made by the victim that had to be manually classified by police officers. 318 were classified as domestic violence, while 839 were classified as non-domestic violence.

In addition to the set of reports, we had an initial thesaurus – a collection of 123 domain-specific terms – at our disposal, which was obtained by performing frequency analyses on the set of police reports. The terms that occurred most often were retrieved and added to the initially empty thesaurus. Each police report was then searched for each of these terms. The result was a cross table in which a cross indicated that the corresponding police report contained the corresponding term.

5 Analyses and results

In this section, we showcase the possibilities of FCA as a knowledge discovery and knowledge enrichment technique. The knowledge discovery process using FCA is summarised and displayed in Figure 5.

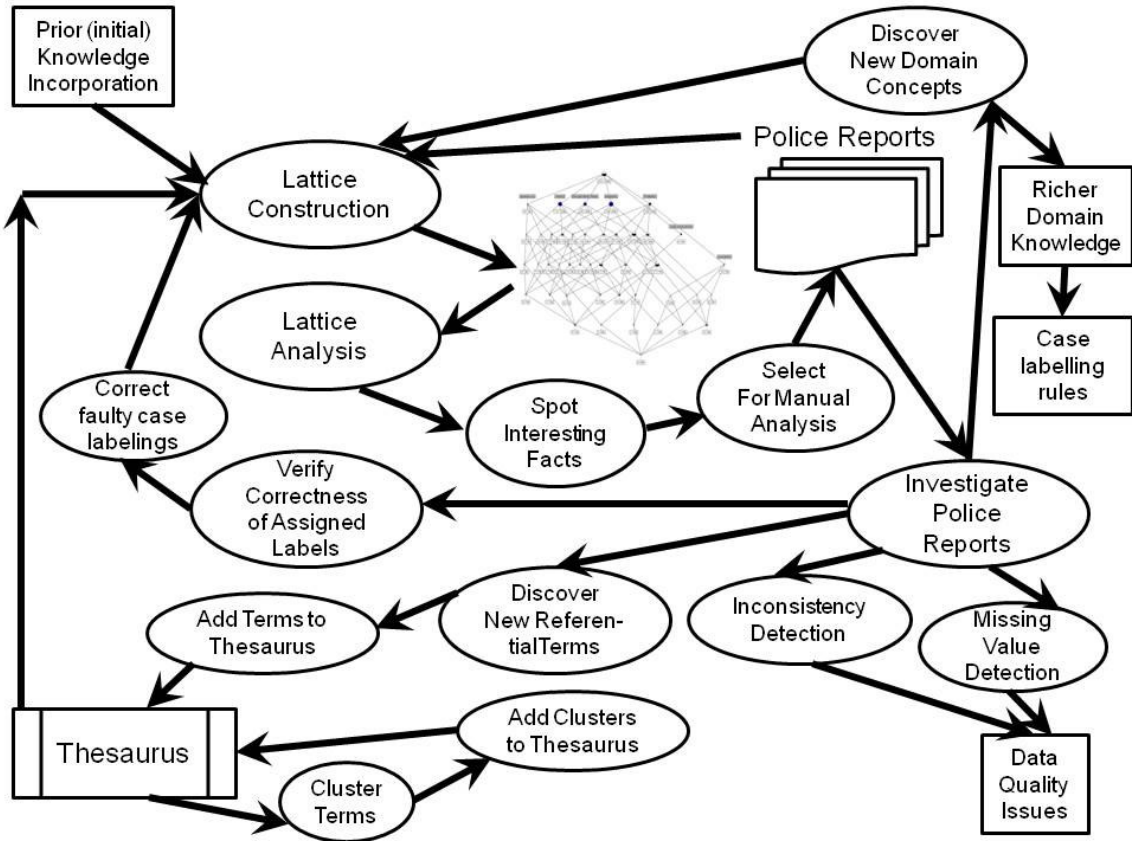

Fig. 5. Detailed human-centered knowledge discovery process using FCA

It is clear that the process displayed in Figure 5 contains an iterative learning loop. Initially, an FCA lattice is constructed based on expert prior knowledge, the terms contained in the thesaurus and the police reports contained in the dataset. Then, the FCA lattice is analysed by the exploratory data analyst and domain expert. Based on the results obtained through the analysis process, which is described in the subsequent paragraphs and demonstrated in detail in the next subsections, a new lattice can be constructed.

The FCA lattices are used as an instrument to discover new case labelling rules and to enrich, test and refine expert prior knowledge. Furthermore, the FCA lattices are used to browse and annotate the collection of police reports and efficiently select representative reports for in-depth manual inspection.

The first major aspect of the process consists in searching these reports for new attributes that can be used to discriminate between the domestic and non-domestic violence reports or that may lead to an enrichment of existing domain knowledge. New referential terms were not acquired and selected using a term extractor, but they were obtained by carefully reading some representative reports and then selecting relevant terms as attributes. We built in the necessary validation mechanisms such as using synonym lists, spelling checking, etc. to ensure the completeness of the thesaurus. During the research the thesaurus was under constant evolution: when new terms and concepts were discovered, the terms were added to the thesaurus. Because of the large number of police reports in the dataset, it was not possible to visually analyse concept lattices containing more than 14 attributes. Therefore, terms with a similar semantic meaning or referring to the same domain concept were clustered by the domain experts. When these term clusters were used to create an FCA lattice, they were considered as attributes. This approach ensured that the thesaurus remained at all times a reflection of the already gained knowledge.

The second major aspect of the process consists of verifying the correctness of the labels assigned by police officers to the selected cases and searching the reports for missing values and inconsistencies. This allowed for the discovery of faulty case labellings and situations that were often not recognised by police officers as domestic or as non-domestic violence. This information was used by the data quality management team to significantly improve the quality of the data contained in the police databases and to improve the way police officers handle domestic violence cases. The information was also useful for the domestic violence programme manager to improve the training of police officers. We also found some regularly occurring confusing situations that could not be uniquely classified as domestic or non-domestic violence based on the domestic violence definition. These situations were presented to the programme manager and were used to enrich, improve and refine the concept and definition of domestic violence.

The third major aspect of the process consists in discovering accurate and comprehensible case labelling rules to automatically classify cases as domestic or as non-domestic violence. In the past this turned out to be impossible. We found that this was largely due to the incorrect labels assigned by police officers to cases, to the vagueness of the domestic violence definition and to the lack of a high-

quality thesaurus. We managed to resolve many of these problems using FCA, resulting in a set of highly accurate and comprehensible classification rules. All these different aspects of the process, which have only been briefly introduced so far, are discussed more extensively in the next sections.

5.1 Domain exploration starting from expert prior knowledge

In this section it is illustrated how we used prior knowledge to start and guide the exploration of the data. We based our initial lattice on the domestic violence definition, by clustering the terms contained in the thesaurus into term clusters associated with one of the two components of the definition (i.e. prior knowledge incorporation). The definition of domestic violence employed by the police organization of the Netherlands is as follows: "Domestic violence can be characterised as serious acts of violence committed by someone in the domestic sphere of the victim. Violence includes all forms of physical assault. The domestic sphere includes all partners, ex-partners, family members, relatives and family friends of the victim. The notion of family friend includes persons that have a friendly relationship with the victim and (regularly) meet with the victim in his/her home [6]."

We intended to verify whether a report can be classified as domestic violence by checking it for the occurrence of one or more terms related to each of the two components of the domestic violence definition. That is, a case can be labelled as domestic violence if the following two conditions are fulfilled. First, a criminal offence has occurred. This may range from verbal threats over pushing and kicking to even killing the victim. To verify whether a criminal offence has occurred, the report is searched for terms such as "hit", "stab" and "kick". These terms are grouped into the term cluster "acts of violence". Second, a person in the domestic circle of the victim is involved in the crime. It should be noted that a report is always written from the point of view of the victim and not from the point of view of the officer. A victim always adds "my", "your", "her" and "his" when referring to the persons involved in the crime. Therefore, the report is searched for terms such as "my dad", "my mom" and "my son". These terms are grouped into the term cluster "family members". The report is also searched for terms such as "my ex-boyfriend", "my ex-husband", and "my ex-wife". These terms are grouped into the term cluster "ex-partners". Furthermore, the report is searched for terms such as "my nephew", "her uncle", "my aunt", "my step-father" and "his step-daughter". These terms are grouped

under the term cluster “relatives.” Then the report is searched for terms such as “family friend” and “co-occupant”. These terms are grouped into the term cluster “family friends”.

Reports that were assigned the label “domestic violence” have been classified as such by police officers. The remaining reports were classified as non-domestic violence. This results in the lattice displayed in Figure 6.

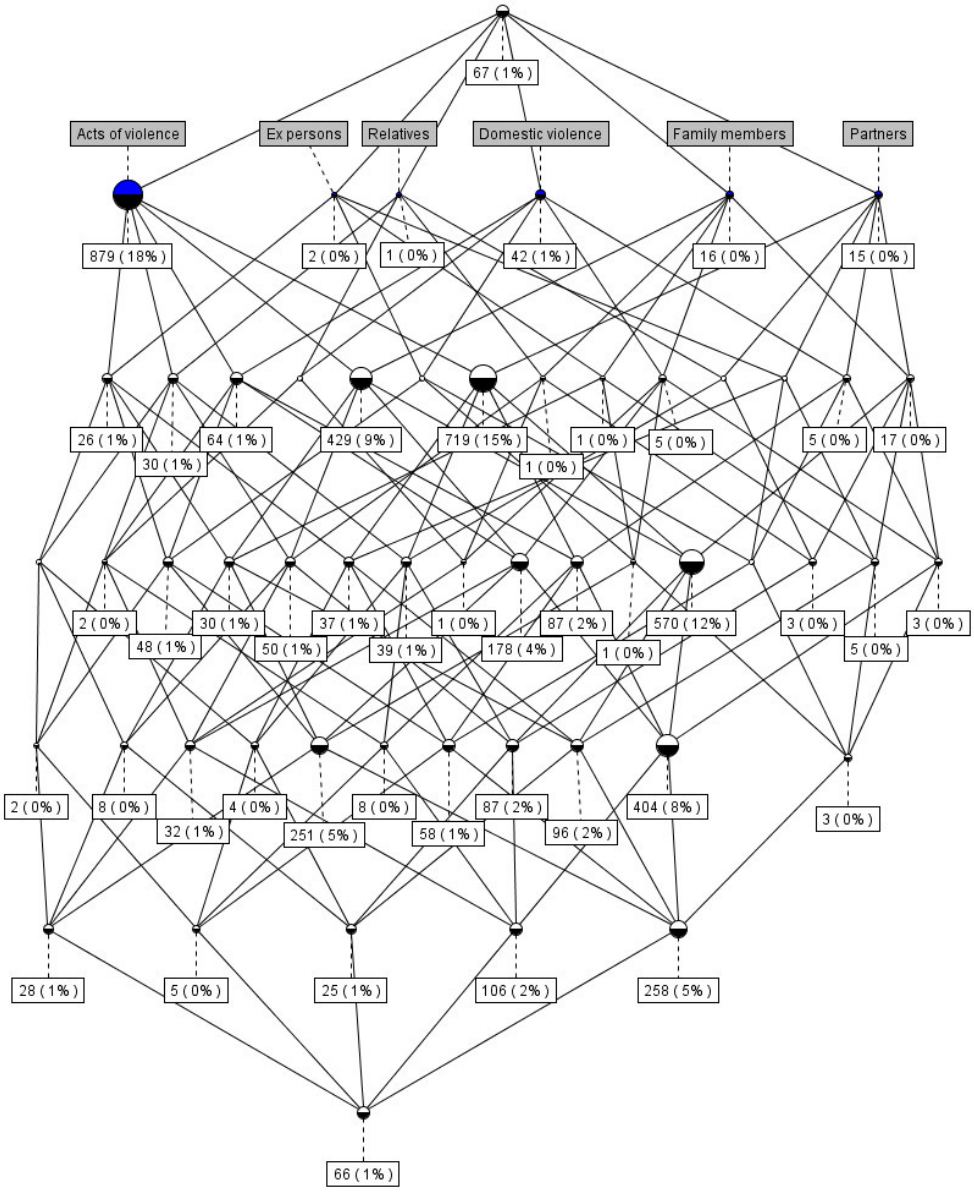

Fig. 6. Initial lattice based on the police reports from 2007

From an initial inspection of the lattice in Figure 6 it quickly became clear that a lattice containing only term clusters based on the starting definition of domestic violence would not discriminate sufficiently between domestic and non-domestic violence reports (i.e. knowledge enrichment). Many non-domestic violence reports seemed to also contain terms attributed to one or more of the term clusters (i.e. prior knowledge validation). Still, some interesting findings emerged from this lattice and triggered further investigation. These findings are discussed in the next section. The lattice structure also made it possible for us to discover the most frequently occurring types of domestic violence cases for 2007. These are summarised in Table 2.

Table 2. Most frequently occurring types of domestic violence in 2007

	% of all domestic violence cases of 2007
"Acts of violence" and "family members" and "partners"	25%
"Acts of violence" and "family members" and "partners" and "ex-persons"	16%
"Acts of violence" and "family members" and "ex-persons"	15%
"Acts of violence" and "family members"	10%
"Acts of violence" and "family members" and "partners" and "relatives"	6%
"Acts of violence" and "partners"	5%

5.2 Prior knowledge testing and referential term discovery

In this section it is demonstrated how we used prior knowledge to guide the exploration of the data. In contrast to what the domain expert initially thought, not all cases labelled as domestic violence by police officers contained terms associated with the two components of the definition (i.e. prior knowledge testing). This led to the discovery of cases that were assigned a wrong label by police officers (i.e. detection of faulty case labellings), to new domain-specific terms that were lacking in the original thesaurus (i.e. referential term discovery) and to a labelling error that was regularly made by police officers (i.e. improvement of training of police officers).

Table 3. Interesting observations from the lattice in Figure 6

	Non-domestic violence	Domestic violence
No "acts of violence"	67	42
No "acts of violence" and one or more of the persons clusters	61	19

Only “acts of violence”	879	64

As can be seen from Table 3, a total of 61 (i.e. 42 and 19) domestic violence cases did not contain a term from the “acts of violence” term cluster. Of these 61 cases 19 contained a term from one of the clusters containing terms referring to a person in the domestic sphere of the victim. After in-depth manual inspection of these 19 cases, it turned out that they contained other violence terms, such as “abduction”, “strangle” and “deprivation of liberty”, which were lacking in the initial thesaurus. The remaining 42 cases, on the other hand, turned out to be wrongly classified as domestic violence.

Interestingly, some 28% (i.e. 879) of the non-domestic violence reports only contain terms from the “acts of violence” cluster, while there are only 64 domestic violence reports in the dataset that share that characteristic. Manual inspection, again, revealed that more than two thirds of these reports were wrongly classified as domestic violence. For some unknown reason, police officers regularly seem to misclassify burglary, car theft, bicycle theft and street robbery cases as domestic violence. Therefore, terms such as “street robbery”, burglary” and “car theft” were combined into a new term cluster called “burglary cases”.

5.3 Term clustering and concept discovery

In this section it is shown how new domain-specific terms, discovered by careful analysis of police reports, and terms with a similar semantic meaning, proposed by the domain expert, were clustered together in term clusters (i.e. term clustering). These term clusters led to the discovery of new concepts that were lacking in the domain expert’s conception of the problem area (i.e. concept discovery). Additionally, two new term clusters based on prior knowledge were introduced (i.e. prior knowledge incorporation). These new term-clusters were used to construct the second FCA lattice.

When browsing a sample of the remaining police reports, we spotted some interesting terms that led to the discovery of two new and important concepts that were lacking in the domain expert’s conception of the problem area. The reports contained terms such as “I had a relationship with”, “relational problems” and “marriage problems”. These terms typically refer to the concept of a broken relationship, which is why they were brought together into the cluster “relational problems”. A

distinction was made between a broken relationship and an ongoing relationship. Terms such as “I have a relationship with” and “live together” were brought together in the cluster “in a relationship”.

According to the literature, domestic violence is a phenomenon that mainly occurs inside the house [4, 5, 6, 21]. Therefore, an attribute called “private locations” was introduced. This term cluster contained terms such as “bathroom”, “living room” and “bedroom”. An attribute called “public locations” was also introduced. To summarise, although the lattice in Figure 5 could not be used to effectively distinguish domestic violence reports from non-domestic violence reports, it could be used to detect cases that were wrongly classified as domestic violence. Also, it helped in discovering new attributes that turned out to be missing in the user’s understanding of the problem area.

The redefined lattice structure, taking into account the above analyses, is displayed in Figure 7. In order to keep the lattice comprehensible, the terms belonging to the clusters “family members”, “relatives”, “partners”, “ex-partners” and “family friends” have been lumped into a cluster “persons”.

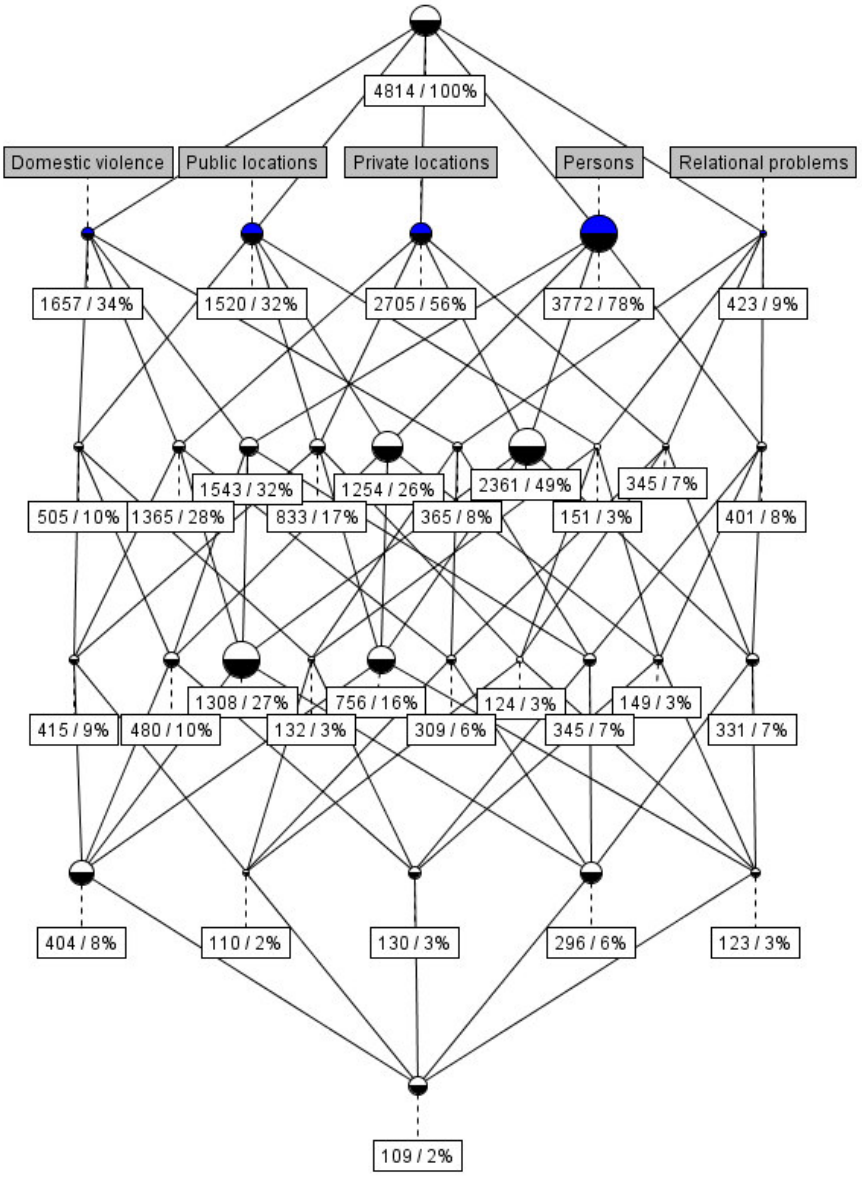

Fig. 7. First refined lattice based on the police reports from 2007

5.4 Detecting faulty case labellings and confusing situations

In this section, we demonstrate how FCA was used to detect faulty case labellings and situations that are confusing to police officers. This was used to improve the training of police officers, to enrich and refine the domestic violence definition and to improve the quality of the data contained in the police databases. We also discovered new referential terms and clustered them based on their semantic meaning, leading to a further enrichment of the thesaurus and the existing domain knowledge.

It should be clear from the lattice in Figure 7 that the terms contained in the cluster "relational problems" tend to be associated with domestic violence cases. Some of the more interesting observations from this lattice are displayed in Table 4.

Table 4. Results from the lattice in Figure 7

	Non-domestic violence	Domestic violence
"relational problems"	58	365
"private locations"	1340	1365
"public locations"	1015	505

Apparently, only 58 non-domestic violence reports contained one or more terms from the "relational problems" cluster. Further investigation revealed that a startling 95% of these cases had been wrongly classified as non-domestic violence. Moreover, about 70% of these cases had in common that a third person made a statement to the police for someone else. For example, one case described a father who made a statement to the police about the sexual abuse of his daughter by her stepfather. This is a clear case of domestic violence. But since it was not the victim who made the statement to the police, the police officer did not recognise it as such.

Analysis of the remaining 30% of these misclassified cases led to the discovery of a new and important concept that was initially lacking from the domain expert's understanding of domestic violence. Many of the reports turned out to contain terms such as "I was attacked by the new boyfriend of my ex girlfriend" and "I was maltreated by the new girlfriend of my ex boyfriend". These terms were grouped into the cluster "attack by new friend of ex-person". Police officers and policy makers confirmed that this type of situation was to be seen as domestic violence, mainly because the perpetrator often aims at emotionally hurting the ex-partner. Consequently, the expectation was for the terms contained in this cluster to frequently occur in domestic violence reports. However, this turned out to be incorrect. It became clear from the investigation that this type of situation in general was very confusing to police officers. A quick scan revealed that more than 50% of police officers actually had trouble with this. The ensuing investigation and discussions with police officers and policy makers revealed that this situation needed to be addressed during the training of police officers. Several interesting cases like the previous one were picked up during the data exploration. All of them gave rise to a clearer insight into the nature of domestic violence.

5.5 Prior knowledge incorporation and testing

In this section we demonstrate how expert prior knowledge was incorporated into the FCA knowledge discovery process. It is also made clear how we used FCA to verify the correctness and the practical usefulness of this prior knowledge.

Most of the domestic violence cases under scrutiny (1365 cases or 82%) contained one or more terms from the "private locations" term cluster. However, 1340 (42%) of the non-domestic violence cases also contained one or more terms from this same term cluster. In addition, a hypothesis that was formulated prior to the data exploration was that almost no domestic violence case was expected to have taken place on the street. Surprisingly, this hypothesis was proven incorrect by the data. In about one-fourth of the domestic violence cases there had been an incident at a public location. While scrutinising these police reports, we discovered that this was often the case when ex-partners were involved. It became apparent that it was not possible to distinguish domestic from non-domestic violence reports by means of the type of locations mentioned in the reports. Combining the clusters "private locations" and "public locations" with clusters such as "family members" or "ex-persons", for example, did not yield the expected results in terms of discriminatory power.

5.6 Definition refinement: niche cases

In this section we focus on how FCA was used to enrich and refine the operationally employed domestic violence definition. Using FCA, we discovered multiple niche cases, which were presented to the domestic violence programme manager. This resulted in an enrichment of the domain knowledge, a refinement of the domestic violence definition and an improvement of the training of police officers.

We continued our knowledge discovery exercise in search of additional attributes to help us distinguish domestic violence from non-domestic violence reports. We noticed that in a large number of the domestic violence cases (416 cases or 28%) the perpetrator and the victim happened to live at the same address at the time the victim made their statement to the police. Most of these cases (379 cases or 91%) were classified as domestic violence. When studying the remaining 37 non-domestic violence cases more carefully, we found, much to our surprise, that the perpetrator and the victim often lived together in the same institution (e.g. a youth institution, a prison or a retirement home). It

turned out that of the 41 cases where the perpetrator and the victim lived in the same institution only 30 actually had been classified as cases of domestic violence.

This finding brought about a lively discussion amongst the police officers of the Amsterdam police force. More importantly, it exposed the discord amongst police officers on how to classify such cases. We took note of all their reflections and presented them to the board members responsible for the domestic violence policy. After intensive debate the following classification guidelines, displayed in Table 5, were obtained.

Table 5. Classification guidelines for incidents involving inhabitants of the same institution

Perpetrator	Victim	Classification
Caretaker	Inhabitant	Domestic violence
Inhabitant	Caretaker	Non-domestic violence
Inhabitant younger than 18y	Inhabitant younger than 18y	Domestic violence
Inhabitant older than 18y	Inhabitant older than 18y	Non-domestic violence
Inhabitant of prison older than 18y	Inhabitant of prison older than 18y	Individual evaluation
Inhabitant older than 18y	Inhabitant younger than 18y	Domestic violence
Inhabitant younger than 18y	Inhabitant older than 18y	Individual evaluation

The presence or absence of a dependency relationship between the perpetrator and the victim was in the end the decisive factor for classifying a case as either domestic or as non-domestic violence.

The non-domestic violence cases where the perpetrator and the victim lived at the same address and were not inhabitants of an institution turned out to be wrongly classified as non-domestic violence. Therefore, a new attribute called “institution” was introduced.

5.7 Missing values detection

In this section, we demonstrate how FCA was used to detect missing values and inconsistencies in police reports. We also show how we exposed inefficiencies in the overall domestic violence policy employed by the police, using FCA..

Another interesting finding that emerged in our search for novel and potentially interesting classification attributes was that some 34% of the reports (1623 cases) did not mention a suspect. According to the domestic violence definition (which specifies that the perpetrator must belong to the domestic circle of the victim), the offender has to be known in domestic violence cases. Naturally, we

had assumed that these reports described non-domestic violence cases. Nevertheless, when looking into these cases, we found that 181 of them turned out to describe domestic violence cases after all.

Analysis revealed that this was a result of police officers' rather haphazard ways of registering victims for these cases. Apparently, while some officers immediately registered a suspect at the moment the victim mentioned this person as a suspect, others preferred to first interrogate them before casting the label of suspect. In the latter cases, the person then would just be added to the list of persons who were said to be involved in or witnessed the crime. Because such lists included friends, family members or bystanders, they could potentially be very extensive and diverse, which is why suspects easily got lost in these lists. When we inquired about the proper policy regarding the labelling of suspects, we were told there simply was none. Our analysis made a strong case for the need of such a policy. In the end, the quick-win proposal that could be implemented to solve this issue involved a relatively simple change to the registration software: an additional data entry field would need to be introduced for police officers to register the persons that were mentioned by the victim as offenders.

Classification of police reports can only be performed on the basis of comprehensible and correct rules that do not inflate the false negative rate, while minimizing the false positive rate. Automatically assigning the non-domestic violence label to a case that does not mention a suspect is thus unacceptable because of the high false negative rate. Nevertheless, we found out that some 44% of the reports (711 cases) that lacked a labelled suspect did contain a description of the actual suspect. Of these 711 cases, only 16 reports were classified as domestic violence. After studying these 16 reports, we discovered that the majority of them were wrongly classified as domestic violence. Classifying cases as non-domestic violence because they lack a labelled suspect and contain a description of the suspect was thus acceptable.

All of this newly discovered knowledge can once again be added to the lattice in Figure 7. When we introduce the attributes "same address", "no suspect" and "description of suspect" to this lattice, this results in the refined lattice structure displayed in Figure 8.

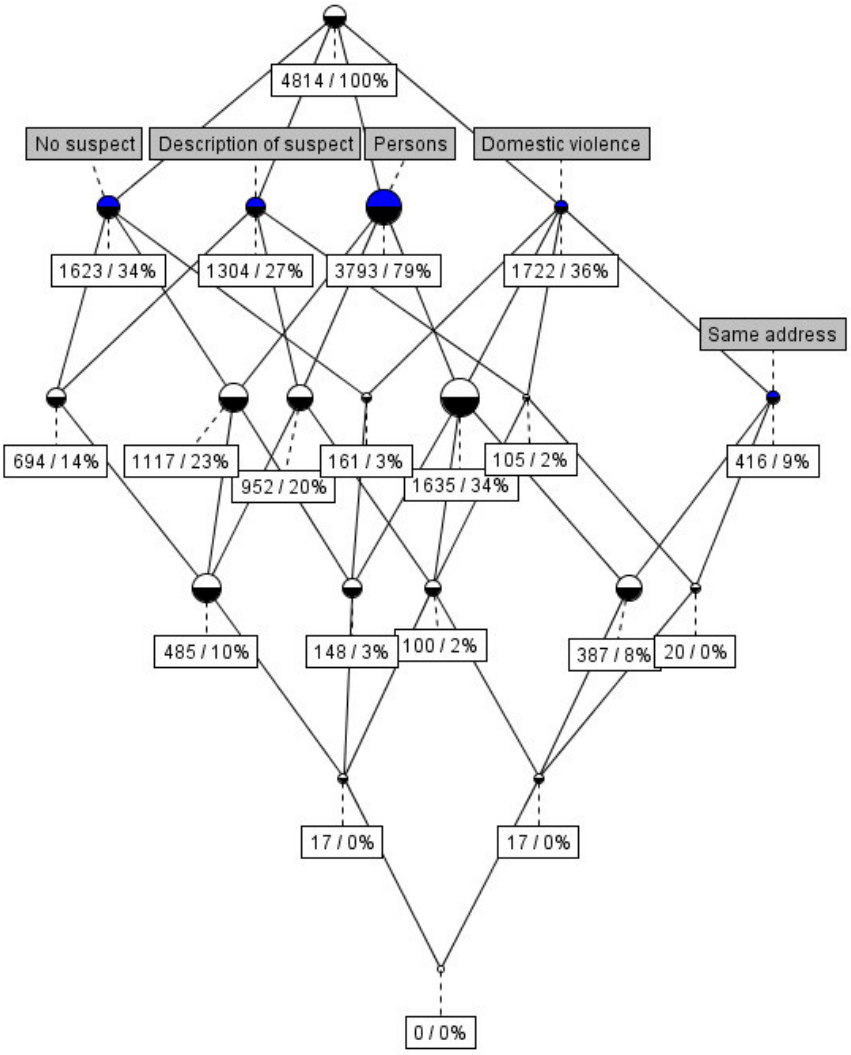

Fig. 8. Second refined lattice based on the police reports from 2007

The lattice in Figure 8 proved to be of much more use for discriminating domestic from non-domestic violence reports. We summarised some of the most interesting findings embedded in that lattice structure in Table 6.

Table 6. Results from the lattice in Figure 8

	Non-domestic violence	Domestic violence
Acts of violence and same address	37	379
Acts of violence and no suspect and description of suspect	695	16
Acts of violence and no suspect	1442	181

5.8 Discovering accurate and comprehensible classification rules

In this section we focus on how we used FCA to discover accurate and comprehensible classification rules. We also illustrate how FCA can play a key role in detecting faulty case labellings.

While further exploring the domestic violence reports, it became apparent that in many cases the victim made statements such as “I want to institute legal proceedings against my husband” and “I want to institute legal proceedings against my brother”. These sentences were brought together into the cluster “legal proceedings against domestic sphere”. Another type of phrasing that was regularly used by victims of domestic violence was, for example, “the crime was committed by my dad” or “the crime was committed by my ex-boyfriend”. These sentences were brought together into the cluster “committed by domestic sphere”. Yet another type of wording that was also frequently used by a victim was phrases such as “I was maltreated by my husband” and “I was threatened by my ex-partner”. These sentences in turn were brought together into the cluster “threatened by domestic sphere”. Finally, neighbourhood quarrels (non-domestic violence) often made reference to phrases such as “I want to institute legal proceedings against my neighbour” and “committed by the man next door”, so these sentences were combined into the cluster “neighbours”. Thus, the lattice was further refined and the result is displayed in Figure 9, with some of the most interesting facts summarised in Table 7.

Table 7. Results from the lattice displayed in Figure 9

	Non-domestic violence	Domestic violence
“legal proceedings against domestic sphere”	19	266
“committed by domestic sphere”	5	81
“threatened by domestic sphere”	4	98
“neighbors"	67	5

After browsing the 19 non-domestic violence cases in which the victim used one or more terms from the “legal proceedings against domestic sphere” cluster, it turned out that these reports should have been classified as domestic violence. The same observation was made when the 5 non-domestic violence reports containing a term from the “committed by domestic sphere” cluster and the 4 non-domestic violence cases containing a term from the “threatened by domestic sphere” cluster were analysed. In-depth investigation of the 5 domestic violence cases in which a term from the

"neighbours" cluster occurred, showed that these reports should have been classified as non-domestic violence.

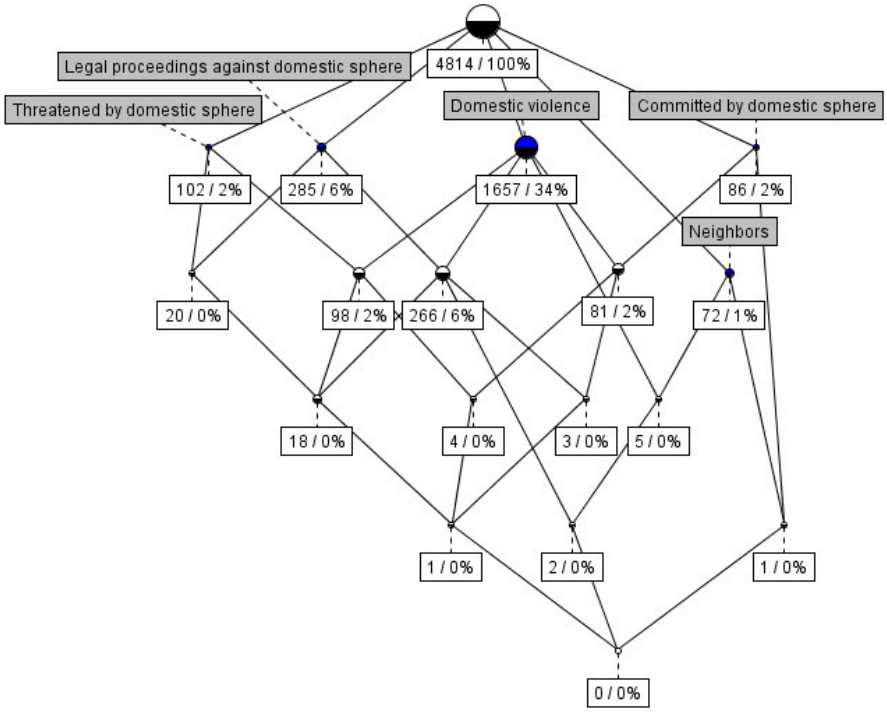

Fig. 9. Third refined lattice based on the police reports from 2007

5.9 Operational validation

In this section, we clarify how FCA was used for the validation of some aspects of operational policing practice. For some specific situations it was verified whether police officers disposed of sufficient knowledge about the problem area to recognise these cases as domestic violence.

Some very important special domestic violence situations were considered, including incest and honour-related violence. For the first type of situation, reports were searched for terms such as "incest" and "sexual abuse by my father". For the second type of situation, reports were searched for terms such as "marriage of convenience" and "marry off". The resulting lattice after incorporating these special cases is displayed in Figure 10. Table 8 summarises the classification.

Table 8. Results from the lattice displayed in Figure 10

	Non-domestic violence	Domestic violence
"incest"	7	8
"honor-related violence"	2	18

Careful inquiry into these cases taught us that police officers regularly misclassified incest cases as non-domestic violence. On the other hand, even for insiders it was quite surprising to observe how almost all honour-related violent incidents ended up being correctly classified as domestic violence. The latter was probably attributable to the intensive sensitisation campaigns organised to inform police officers of this important societal problem.

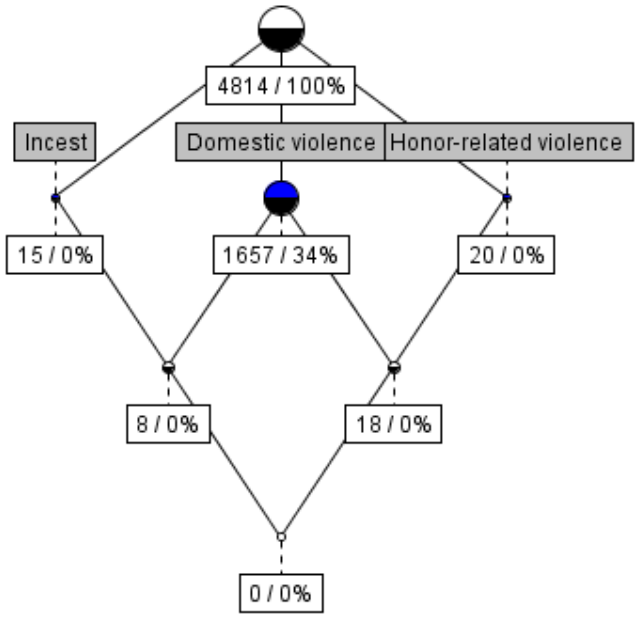

Fig. 10. Fourth refined lattice based on the police reports from 2007

6 Validation experiment

In this section we elaborate on the run-time power of the distilled knowledge. We start by mapping the proposed lattice structures obtained during discovery of the 2007 police data on the police reports from 2006. We demonstrate that the findings obtained through in-depth analysis of the 2007 police data are also valid for the police reports from 2006. Then, we apply the discovered knowledge to automatically classify the output of the in-place case triage system. Finally, we demonstrate how the newly discovered knowledge was used to detect and reclassify filed reports that were incorrectly labelled by police officers.

For the classification rules discovered in section 5, we verified how many domestic and non-domestic violence reports correspond to each rule. The rules and these counts are represented in Table 8. For the first eight rules, the non-domestic violence cases turned out to be incorrect labellings performed by police officers. For rules 9 up to 13, the domestic violence cases turned out to be incorrect labellings performed by police officers. Using rule 14, we found that in 160 cases that were classified as domestic violence by police officers a formally labelled suspect was lacking.

Table 8. Discovered knowledge applied to police reports from 2006

	Non-domestic violence	Domestic violence
Domestic violence rules		
1. "legal proceedings against domestic sphere"	24	237
2. "committed by domestic sphere"	9	101
3. "threatened by domestic sphere "	11	106
4. "incest"	0	3
5. "attack by new friend of ex-person"	6	12
6. "relational problems"	61	364
7. "same address" and not in "institution"	23	299
8. "honor-related violence"	1	16
Non-domestic violence rules		
9. "burglary cases"	32	24
10. "neighbors"	13	6
11. "no suspect" and "description of suspect"	504	15
12. no "acts of violence"	30	38
13. "acts of violence" and no "persons"	865	94
Data quality check extra		
14. "no suspect"	1074	160

For classification, the protocol is as follows. When a case comes in for labelling, the first step consists in verifying whether one of the domestic violence rules is satisfied. If this is the case, the case is classified as domestic violence. If the "no suspect" or one of the non-domestic violence rules turns out to be also satisfied, the case is sent to the data quality management team, because there probably is a data quality problem. Otherwise, it is verified whether one of the non-domestic violence rules is satisfied. If this is the case, the case is classified as non-domestic violence. Otherwise, the case is left unclassified. By applying the first thirteen rules in Table 8, 50% of the dataset of 2006 could be automatically correctly classified).

A further validation encompassed the application of the discovered knowledge to automatically classify the output of the in-place case triage system. For example, going back to 2006, the system retrieved 1157 cases, 80% of which actually turned out to be non-domestic violence cases. It is to deal

with these shortcomings in the current system that the rules in Table 8 will prove to be extremely useful.

Some 9% of the cases contained terms from the "committed by domestic sphere", "threatened by domestic sphere" or "legal proceedings against domestic sphere" clusters and could be automatically classified as domestic violence. About 10% of the cases contained one or more terms from the "relational problems" cluster and could for that reason be automatically classified as domestic violence. A further 11% of the retrieved cases could be classified as domestic violence simply because the perpetrator and the victim lived at the same address, which was not an institution. About 18% of the retrieved cases did not mention a suspect. If the policies we proposed had been implemented, these could all have been classified as non-domestic violence. Some 5% of the cases lacked a formally designated suspect but contained a description of a suspect. These cases could be classified as non-domestic violence. Another 14% of the cases retrieved by the triage system in 2006 could immediately be classified as non-domestic violence. They all contained one or more terms from the "acts of violence" cluster and none from the "persons" cluster. In sum, 514 of the 1157 cases retrieved by the triage system in 2007 could be correctly classified in an automated way when making use of the newly discovered knowledge. These findings are displayed in Table 9.

Table 9. % of the 2006 cases classified automatically

	% retrieved cases classified automatically
Current situation	0%
Applying first 13 discovered rules from Table 8	44%
Adding data field for suspect mentioned by victim to police registration form	54%

The proposal that could be implemented to solve this issue involves a rather small change to the triage software: incorporating the first thirteen rules from Table 8 into the existing triage model. As a result, about 44% of the retrieved cases will be automatically classified correctly. Moreover, if an additional data field for the suspect mentioned by the victim is added to the police registration form, the fourteenth rule of Table 8 can also be integrated into the triage model. This would result in an automatic and correct classification of about 54% of the retrieved cases. An additional result is that a large number of the filed reports that were wrongly classified can now be automatically detected and corrected, the results of which are displayed in Table 10.

Table 10. Number of filed reports that were incorrectly classified, but corrected by means of the 13 rules

	Non-domestic corrected to Domestic	Domestic corrected to Non-domestic	Total
Year 2006	135	110	245
Year 2007	124	88	212
First quarter 2008	54	24	78

Using the newly discovered rules, many of these incorrectly classified police reports can be automatically detected and reclassified. For example, for the year 2007, we found 212 filed police reports that were incorrectly classified.

7 Conclusions

Domestic violence is one of the top priorities of the Amsterdam-Amstelland police. When a victim makes a statement to the police, police officers are given the possibility to indicate whether it is a domestic violence case. Still, this has proven to be problematic. The use of FCA , however, can play a significant role in overcoming some of the hurdles encountered when dealing with domestic violence cases. This paper specifically showcased the possibilities of using FCA for knowledge discovery from police reports.

The FCA lattices prove to be very useful as knowledge browsers. The construction of an initial lattice containing term clusters created by a domain expert on the basis of the domestic violence definition and the incremental refinement of this lattice was shown to provide a powerful framework for exploring unstructured data. First, it was shown that the domestic violence definition is too vague, making it hard to use it effectively for classification purposes. Moreover, the scope of terms such as ex-partners and violence, was nowhere communicated in the definition. Second, we exposed that there exists a considerable amount of confusion amongst police officers about the nature and scope of domestic violence. Regularly occurring domestic violence situations such as incest or an ex-boyfriend attacking the new boyfriend of a girl were often not recognised as such by police officers. Third, using FCA, we were able to discover some essential characteristics that discriminate domestic from non-domestic violence reports. These characteristics include phrasings, words and word combinations that typically occur in either domestic or non-domestic violence cases.

This newly discovered knowledge was then used to automatically assign a label to the cases retrieved by the in-place case triage system. It turned out to be possible to automatically and correctly

classify about 44% of the cases that used to be set aside for manual inspection. Moreover, a large part of the filed reports that were incorrectly classified, could be automatically detected and reclassified.

Acknowledgements

The authors would like to thank the police of Amsterdam-Amstelland for granting them the liberty to conduct and publish this research. In particular, we are most grateful to Deputy Police Chief Reinder Doeleman and Police Chief Hans Schönfeld for their continued support. Jonas Poelmans is aspirant of the Fonds Voor Wetenschappelijk Onderzoek – Vlaanderen or Research Foundation – Flanders.

References

[1] Office on Violence against Women (2007) About Domestic Violence (http:www.usdoj.gov/ovw/domviolence.htm). Retrieved on 2007-10-22
[2] Watts, C., Timmerman, C. (2002) Violence against women: global scope and magnitude. The Lancet 359 (9313): pp.1232-1237. RMID 1155557
[3] Waits, K. (1985). The criminal Justice System's response to Battering: Understanding the problem, forging the solutions. Washington Law Review 60: pp. 267-330
[4] Vincent, J.P., Jouriles, E.N. (2000) Domestic violence. Guidelines for research-informed practice. Jessica Kingsley Publishers Londen and Philadelphia
[5] Black, C.M. (1999) Domestic violence: Findings from a new British Crime Survey self-completion questionnaire. London: Home Office Research Study.
[6] Keus, R., Kruijff, M.S. (2000) Huiselijk geweld, draaiboek voor de aanpak. Directie Preventie, Jeugd en Sanctiebeleid van de Nederlandse justitie.
[7] Yevtushenko, S.A. (2000). System of data analysis "Concept Explorer." Proceedings of the 7th national conference or Artificial Intelligence. KII-2000. 127-134, Russia
[8] Ganter, B., Wille, R. (1999), Formal Concept Analysis: Mathematical Foundations. Springer, Heidelberg.
[9] Wille, R. (1982), Restructuring lattice theory: an approach based on hierarchies of concepts, I. Rival (ed.). Ordered sets. Reidel, Dordrecht-Boston, 445-470.
[10] Poelmans, J., Elzinga, P., Viaene, S., Dedene, G. (2010), Formal Concept Analysis in knowledge discovery: a survey. Lecture Notes in Computer Science, 6208, 139-153, 18th international conference on conceptual structures (ICCS): from information to intelligence. 26 - 30 July, Kuching, Sarawak, Malaysia**.** Springer.
[11] Wille, R. (2002), Why can concept lattices support knowledge discovery in databases?, Journal of Experimental & Theoretical Artificial Intelligence, 14: 2, 81-92.
[12] Stumme, G., Wille, R., Wille, U. (1998), Conceptual Knowledge Discovery in Databases Using Formal Concept Analysis Methods, In: J.M. Zytkow, M. Quafofou (eds.): Principles of Data Mining and Knowledge Discovery, Proc. 2 nd European Symposium on PKDD '98, LNAI 1510, Springer, Heidelberg, 1998, 450-458.
[13] Stumme, G. (2002) Efficient Data Mining Based on Formal Concept Analysis. Lecture Notes in Computer Science Vol. 2453, Springer, Heidelberg, 3-22
[14] Stumme, G. (2002), Formal Concept Analysis on its Way from Mathematics to Computer Science. Proc. 10th Intl. Conf. on Conceptual Structures (ICCS 2002). LNCS, Springer, Heidelberg 2002.
[15] Priss, U. (2000), Lattice-based information Retrieval. Knowledge Organization, 27, 3, 132-142.
[16] Godin, R., Gescei, J., Pichet, C. (1989), Design of browsing interface for information retrieval. In: N.J.Belkin, C.J. van Rijsbergen (Eds.), Proc. SIGIR '89, 32-39.
[17] Carpineto, C., Romano, G. (2005), Using concept lattices for text retrieval and mining. In Formal Concept Analysis-State of the Art, Proc. of the first International Conference on Formal Concept Analysis, Berlin, Springer.
[18] Cole, R. , Eklund, P. (2001), Browsing Semi-structured Web Texts Using Formal Concept Analysis. In H. Delugach, G., Stumme (Eds.), Conceptual Structures: Broadening the Base, LNAI 2120, Berlin, Springer, 319-332.

[19] Eklund, P., Ducrou, J., Brawn, P. (2004), Concept Lattices for Information Visualization: Can Novice Read Line Diagrams? In P. Eklund (Ed.), Concept lattices: Second International Conference on Formal Concept Analysis, LNCS 2961, Berlin, Springer, 14-27.
[20] Priss, U. (1997), A Graphical Interface for Document Retrieval Based on Formal Concept Analysis. In: E. Santos (Ed.), Proc. of the 8th Midwest Artificial Intelligence and Cognitive Science Conference. AAAI Technical Report CF-97-01, 66-70.
[21] Beke, B.M.W.A., Bottenberg, M. (2003) De vele gezichten van huiselijk geweld. In opdracht van Programma Bureau Veilig / Gemeente Rotterdam. Uitgeverij SWP Amsterdam.
[22] T. van Dijk, Huiselijk geweld, aard, omvang en hulpverlening (Ministerie van Justitie, Dienst Preventie, Jeugd-bescherming en Reclassering, oktober 1997).
[23] Peirce, Ch. S. (1992), Reasoning and the logic of Things : The Cambridge Conferences lectures of 1898 edited by K. L. Ketner and H. Putman, Cambridge: Harvard University Press.
[24] Arnauld, A., Nicole, P. (1985), la logique ou l'Art de penser. Edition Gallimard .
[25] Collier, P.M. (2006) Policing and the intelligent application of knowledge. Public money & management. Vol. 26, No. 2, pp. 109-116.
[26] Collier, P.M., Edwards, J.S. and Shaw, D. (2004) Communicating knowledge about police performance. International Journal of Productivity & Performance Management. Vol. 53, No. 5, pp. 458-467
[27] Chen, H., Chung, W., Xu, J.J., Wang, G., Qin, Y., Chau, M. (2004) Crime data mining: a general framework and some examples. IEEE Computer, April 2004.
[28] Ananyan, S. (2002) Crime Pattern Analysis Through Text Mining. Proceedings of the Tenth Americas Conference on Information Systems, New York, New York, August 2004.
[29] Brachman, R., Anand, T. (1996) The process of knowledge discovery in databases: a human-centered approach. In advances in knowledge discovery and data mining, ed. U. Fayyad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy. AAAI/MIT Press
[30] Brachman, R.J., Selfridge, P.G., Terveen, L.G., Altman, B., Borgida, A., Halper, F., Kirk, T., Lazar, A., Mc Guinnes, D.L. and Resnick, L.A. (1993) Integrated support for data archaeology. International Journal of In telligent and Cooperative Information Systems, 2:159-185.
[31] Raaijmakers, S.A., Kraaij, W., Dietz, J.B. (2007) Automatische detectie van huiselijk geweld in processenverbaal. TNO-rapport 34293.
[32] Politie Amsterdam-Amstelland (2008) http://www.politie-amsterdam-amstelland.nl/get.cfm?id=86, Retrieved on 2008-02-22.
[33] Smyth, P., Pregibon, D., Faloutsos, C. (2002) Data-driven evolution of data mining algorithms. Communica tions of the ACM, Vol. 45, no. 8.
[34] Fayyad, U., Uthurusamy, R. (2002) Evolving data mining into solutions for insights. Communications of the ACM, Vol. 45, no. 8
[35] Cole, R.J. (2000) The management and visualization of document collections using Formal Concept Analysis. Ph. D. Thesis, Griffith University.
[36] Christopher, A. (1965) A city is not a tree. Architectural Forum, Vol 122, No 1, April 1965, pp 58-62 (Part I) and Vol 122, No 2, May 1965, pp 58-62 (Part II)
[37] Poelmans, J., Elzinga, P., Viaene, S., Dedene, G. (2008). An exploration into the power of formal concept analysis for domestic violence analysis, Lecture Notes in Computer Science, 5077, 404 – 416, Advances in Data Mining. Applications and Theoretical Aspects, 8th Industrial Conference (ICDM), Leipzig, Germany, July 16-18, 2008, Springer.
[38]Poelmans, J., Elzinga, P., Viaene, S., Van Hulle, M. & Dedene G. (2009). Gaining insight in domestic violence with emergent self organizing maps, Expert systems with applications, 36, (9), 11864 – 11874.
[39] Viaene S., De Hertogh S., Lutin L., Maandag A., den Hengst S., Doeleman R. (2009). Intelligence-led policing at the Amsterdam-Amstelland police department: operationalized business intelligence with an enterprise ambition. Intelligent systems in accounting, finance and management. 16 (4) : 279 -292.