Semi-automated knowledge discovery: identifying and profiling human trafficking

Jonas Poelmans [a] [b] , Paul Elzinga [c] , Dmitry I. Ignatov [b] & Sergei O.
Kuznetsov [b]

[a] Faculty of Business and Economics, KU Leuven, Naamsestraat 69,
3000, Leuven, Belgium

[b] National Research University Higher School of Economics (HSE),
Pokrovskiy bd. 11, 101000, Moscow, Russia

[c] Amsterdam-Amstelland Police, James Wattstraat 84, 1000 CG,
Amsterdam, The Netherlands
Version of record first published: 17 Sep 2012.

PLEASE SCROLL DOWN FOR ARTICLE

Semi-automated knowledge discovery: identifying and profiling human trafficking

Jonas Poelmans[a,b]*, Paul Elzinga[c], Dmitry I. Ignatov[b] and Sergei O. Kuznetsov[b]

[a]*Faculty of Business and Economics, KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium;*
[b]*National Research University Higher School of Economics (HSE), Pokrovskiy bd. 11, 101000 Moscow, Russia;* [c]*Amsterdam-Amstelland Police, James Wattstraat 84, 1000 CG Amsterdam, The Netherlands*

We propose an iterative and human-centred knowledge discovery methodology based on formal concept analysis. The proposed approach recognizes the important role of the domain expert in mining real-world enterprise applications and makes use of specific domain knowledge, including human intelligence and domain-specific constraints. Our approach was empirically validated at the Amsterdam-Amstelland police to identify suspects and victims of human trafficking in 266,157 suspicious activity reports. Based on guidelines of the Attorney Generals of the Netherlands, we first defined multiple early warning indicators that were used to index the police reports. Using concept lattices, we revealed numerous unknown human trafficking and loverboy suspects. In-depth investigation by the police resulted in a confirmation of their involvement in illegal activities resulting in actual arrestments been made. Our human-centred approach was embedded into operational policing practice and is now successfully used on a daily basis to cope with the vastly growing amount of unstructured information.

Keywords: formal concept analysis; semi-automated knowledge discovery; human trafficking

Mathematics Subject Classification: 06B23; 06B35

1. Introduction

Traditional fully automated knowledge discovery methods proved to be useful in many areas. However, the major drawback of all automated and supervised machine learning techniques, including decision trees, is that these algorithms assume that the underlying concepts of the data are clearly defined, which is often not the case. These techniques allow almost no interaction between the human actor and the tool, and fail at incorporating valuable expert knowledge into the discovery process (Keim 2002), which is needed to go beyond the fool's gold (Smyth, Pregibon, and Faloutsos 2002). In the paper by Hollywood, Strom, and Pope (2009) these problems were clearly addressed in the context of terrorist threat assessment. The central question was whether it is possible to find terrorists with traditional fully automated data mining techniques and the answer was no. Because of the nature of the available data and the presence of human expertize which is hard to formalize, one needs in certain cases an interactive methodology. Human-centred knowledge discovery in databases (KDD) refers to the constitutive nature of human interpretation for the discovery of knowledge and stresses the complex interactive process

*Corresponding author. Email: jonas.poelmans@gmail.com

of KDD as being led by human thought (Brachman and Anand 1996). The aim is to make it easy, practical, and convenient to explore very large databases for organizations and users with vast amount of data but without years of training as data analysts (Fayyad and Uthurusamy 2002). A significant part of the art of data mining is the user's intuition with respect to the tools (Marchionini 2006). In this study, we propose a semi-automated and human-centred knowledge discovery methodology based on formal concept analysis (FCA; Ganter and Wille 1999). The proposed approach recognizes the important role of the domain expert in mining real-world enterprise applications and makes use of specific domain knowledge, including human intelligence and domain-specific constraints. To illustrate its effectiveness, we report on a real-life case study on using the methodology at the Amsterdam-Amstelland police in the Netherlands, which is aimed at identifying and profiling human trafficking.

In 2009, Amsterdam was shocked by the brute murder of a Hungarian 19-year-old woman who was forced to work in prostitution but resisted to her pimps. In countries such as the Netherlands, prostitution is legalized but severe penalties can be given to criminals who force a girl to have sex for money. Girls of Dutch nationality who were forced to work in prostitution in Amsterdam typically fell prey to a loverboy. The loverboy is a relatively new phenomenon (Bovenkerk, Van San, Boone, Van Solinge, and Korf 2004) in the Netherlands. A loverboy is a man, mostly with Moroccan, Antillean, or Turkish roots, who makes a girl fall in love with him and then uses her emotional dependency to force her to work as a prostitute. Forcing girls and women into prostitution through a loverboy approach is seen as a special kind of human trafficking in the Netherlands (article 250a of the code of criminal law).

Human trafficking is the fastest growing criminal industry in the world, with the total annual revenue for trafficking in persons estimated to be between $5 and $9 billion (United Nations, Economic and Social Council 2004). The council of Europe states that 'people trafficking has reached epidemic proportions over the past decade, with a global annual market of about $42.5 billion' (Equality Division, Directorate General of Human Rights of the Council of Europe 2006). Rough estimates suggest that 700,000 to 2 million women and girls are trafficked across international borders every year (O'Neill Richard 1999; US Department of State 2008). The impoverished former Eastern block countries such as Albania, Moldova, Romania, Hungary, Bulgaria, Russia, Belarus, and Ukraine have been identified as major trafficking source countries for women and children (Levchenko 1999; Dettmeijer-Vermeulen, Boot-Mattthijssen, Van Dijk, De Jonge van Ellemeet, and Smit 2008). Since the fall of the Iron curtain starting in 1989 in Poland, millions of Central and Eastern European girls and women have been victims of human trafficking and were forced to work in the European sex industry (estimated 175,000–200,000 yearly[1]).

Because of the overload of mostly textual information in police databases and a lack of adequate supporting instruments to make these data more accessible, it becomes increasingly difficult to identify potential suspects and gather all available information about them (Poelmans, Elzinga, Dedene, Viaene, and Kuznetsov 2011). In this study, we aimed at describing the new investigation procedures we developed with the Amsterdam-Amstelland police for identifying and profiling potential suspects from this large amount of textual reports. Since the introduction of intelligence-led policing (Collier 2006) in 2005, a management paradigm, for police organizations which aims at gathering and using information to allow for pro-active identification of suspects, police officers are required to write down everything suspicious they noticed during motor vehicle inspections, police patrols, etc. These observational reports, 34,817 in 2005, 40,703 in 2006, 53,583 in 2007, 69,470 in 2008, and 67,584 in 2009, may contain indications that can help reveal

individuals who are involved in human trafficking, forced prostitution, terrorist activities, etc. However, till date almost no analyses were carried out on these documents.

Concept lattices are used to display the persons found in the available police reports and the early warning indicators observed for each of them. Police officers can then extract persons in whom they are interested and create a detailed profile for them. This profile can be represented by a concept lattice which displays all available information about this suspect, including social structure and temporal information, in one appealing visual picture. Our approach promotes efficient decision making and significantly outperformed the currently employed manual investigation methods. The concept lattices revealed some cases in which there were sufficient indications for starting an in-depth investigation. We applied FCA and its temporal variant to zoom in on some real-life cases and suspects, resulting in actual arrestments being made and/or illegal prostitution locations closed down.

The remainder of this paper is composed as follows. In Section 2, we give background information on human trafficking, forced prostitution, and the guidelines that were developed by the Attorney Generals of the Netherlands to help detect trafficking and loverboy suspects. In Section 3, we describe FCA and in Section 4, the data-set. In Section 5, we describe our analysis method to detect and profile potential suspects. In Section 6, we describe some real-life cases in which the suspects were found with FCA. Section 7 contains a discussion of our work. Finally, Section 8 concludes the paper.

2. Human trafficking and forced prostitution

The most popular destinations for trafficked women are countries where prostitution is legal such as the Netherlands (Hughes 2000). According to Shelley (1999), most of these women are in conditions of slavery. A survey of women from central and Eastern Europe in the Netherlands found that 80% of them were kept in isolation and forced to work long hours for no pay and were physically and emotionally abused by pimps, traffickers, and male buyers (Hyde and Denisenko 1997). Human trafficking and illegal forced prostitution are typically organized by international crime networks that make large amounts of money through the exploitation of young women and children. The money made by the criminal networks does not stay in poor communities but is laundered through bank accounts of criminal bosses in financial centres such as the USA, Western Europe, and off-shore accounts (Savona 1998). In Amsterdam, in particular Bulgarian and Hungarian criminals are active. Women who have been forced into prostitution can keep little or nothing of the money they earned. If they manage to escape they will return home in poverty and physically and emotionally damaged for life (Farley and Barkan 1998). One of her only ways to escape the unwanted sex with multiple men each day is becoming a perpetrator herself. Women who fell prey to traffickers sometimes return home to recruit new victims. According to Hughes and Denisova (2003), 70% of pimps in Ukraine are women. A recruiter gets US $2000–$5000 for each woman recruited. Pimps can make 5–20 times as much from a woman as they paid for her in a short time.

2.1 *Human trafficking model*

Victims of human trafficking rarely make an official statement to the police. The human trafficking team of the Amsterdam-Amstelland police is installed to proactively search police databases for any signals of human trafficking. Unfortunately, this turns out to be a laborious task. The investigators have to manually read and analyse the police reports, one by one, because only an estimated 15% of the information containing human trafficking

indications have been labelled as such by police officers. As soon as the investigators find sufficient indications against a person, a document based on Section 273f of the code of criminal law is composed for the person under scrutiny. Based on this report, a request is sent to the public prosecutor to start an in-depth investigation against the potential suspects. After permission is received from the public prosecutor, the use of special investigation techniques such as phone taps and observation teams is allowed.

The Attorney Generals of the Netherlands developed a set of guidelines based on which police forces can gather evidence of human trafficking and forced prostitution against potential suspects. These guidelines mention indications of human trafficking and forced prostitution and define in which cases pro-active intervention by police may be necessary. This information had not yet been used to actively search police databases for suspicious activity reports containing human trafficking indicators. Table 1 contains the five main types of indicators contained in these guidelines and Table 2 illustrative examples for each of them. The full list of indicators can be found in Appendix A.

2.2   *Loverboy model*

We discuss another model that was developed by Bullens and Van Horn (2000) for the identification of loverboys who typically force girls of Dutch nationality into prostitution. Loverboys use their love affair with a woman to force her to work in prostitution. Forcing girls and women into prostitution through a loverboy approach is seen as a special kind of human trafficking in the Netherlands (article 250a of the code of criminal law). This model is a resource used by the Amsterdam-Amstelland police during the trainings of police officers about this topic. A typical loverboy approach consists of three main phases. Table 2 contains the four main types of indicators and two illustrative examples for each of them. The full list of indicators can be found in Appendix B.

3.   **Formal concept analysis**

FCA arose 25 years ago as a part of applied lattice theory (Wille 1982; Ganter and Wille 1999) and has over the years grown into a powerful tool for data analysis (Lakhal and Stumme 2005; Priss 2006; Poelmans, Elzinga, Viaene, and Dedene 2010c), data visualization (Doerfel, Jäschke, and Stumme 2012), and information retrieval (Carpineto

Table 1.   Human trafficking indicators.

Dependency of the exploiter
- The woman has a fake or counterfeit passport
- The woman does not know properly what her working address is

Deprivation of liberty
- The victim does not receive necessary medical treatment
- The victim does not carry her own identity papers

Being forced to work under bad circumstances
- The victim receives an unusually low wage compared to the market
- The victim has to work under all circumstances and unreasonably long

Violation of bodily integrity of the victim
- Threatened or confronted with violence
- Certain things that may indicate the dependence on the exploiter such as tattoos or voodoo material

Non-incidental pattern of abuse by suspect(s)
- Working at different places from time to time
- Tips of reliable third parties

Table 2.   Loverboy indicators.

Preparatory activities to recruit girls
  • Actual recruitment and arranging residence and shelter locations for the girls
  • During the first meeting, they estimate how vulnerable a girl is to attention and flattery. Their
    sensitivity to attention, presents, etc. made her fall in love with the pimp
Forcing her into prostitution
  • Deflowering and forcible rape: in particular for Islamic girls, deflowering and the threat of being
    brought back home increase their anxiety to say no to the pimp's demands, because it can result
    in her abandonment by her family
  • Blackmailing: if the girls do not want to work in prostitution, the pimps threaten to bring her
    back to her parents
Keeping the girl in prostitution
  • Emotional dependence: feelings of love, nobody else to support her, the pimp is the father of her
    child, etc.
  • Social isolation: she becomes isolated from the outside world and only meets people from the
    prostitution circuit
The pimp will also try to protect his organization
  • Internal protection measurements: he will make sure that the girls are constantly under
    surveillance and with the threat of physical violence he completely dominates her life
  • External protection: the pimp will threaten, bribe, interrogate, etc. the girls who have been in
    contact with the police

---

and Romano 2004; Poelmans, Ignatov, Viaene, Dedene, and Kuznetsov 2012b). The usage
of FCA for browsing text collections has been suggested before, e.g. by Cole (2000). Here,
we make a stress on using FCA in an actionable environment for discovering different
types of knowledge in unstructured text. FCA has been applied in a wide range of domains,
including medicine (Schnabel 2002; Belohlávek, Sigmund, and Zacpal 2011), biology
(Motameny, Versmold, and Schmutzler 2008), social sciences, linguistics (Priss 2004),
ontology (Cimiano, Hotho, Stumme and Tane 2004), and software engineering
(Eisenbarth, Koschke, and Simon 2003). For instance, FCA has been applied to analysing
data of children with diabetes (Scheich, Skorsky, Vogt, Wachter, and Wille 1993), for
duplicate detection in web search results (Ignatov and Kuznetsov 2009), for developing a
recommender system in Internet advertisement (Ignatov and Kuznetsov 2008), and for an
Information Technology (IT) security management system (Becker, Stumme, Wille,
Wille, and Zickwolff 2000). In Eklund, Ducrou, and Brawn (2004) and Domingo and
Eklund (2005), FCA was used as a visualization technique that allows human actors to
quickly gain insight by browsing through information. We previously applied FCA to a
police data-set containing domestic violence cases and were able to establish its practical
usefulness (Poelmans, Elzinga, Viaene, and Dedene 2010a). FCA is particularly suited for
exploratory data analysis because of its human-centredness (Hereth, Stumme, Wille, and
Wille 2003; Valtchev, Missaoui, and Godin 2004). It is a fundamental principle that the
generation of knowledge from information is promoted by representations that make the
inherent logical structure of the information transparent. FCA builds on the model that
concepts are the fundamental units of human thought. Hence, the basic structures of logic
and logical structure of information are based on concepts and concept systems (Stumme,
Wille, and Wille 1998; Stumme 2002). Consequently, FCA uses the mathematical
abstraction of the concept lattice to describe systems of concepts to support human actors
in their information discovery and knowledge creation practice (Wille 2002).

A *formal context* is a triple of sets (G, M, I), where G is interpreted as a set of *objects*,
M is interpreted as a set of *attributes*, and binary *incidence relation* defines attributes
describing particular objects, i.e. $(x, y) \in I$ means object x has attribute y. A formal

Table 3. Example of a formal context.

	Prostitution	Loverboy	Violence	Expensive cars	Large amount of money	Bulgarian
Report 1: 13 June 2007	×	×				×
Report 2: 26 July 2008			×	×	×	
Report 3: 28 September 2008	×	×	×	×	×	
Report 4: 05 February 2009						×
Report 5: 22 February 2009				×	×	

context can be represented by a *cross table* with set of rows *G*, set of columns *M*, where each cross corresponds to an element of relation *I*, i.e. say that a particular object has a particular attribute. An example of a cross table is displayed in Table 3. Here, suspicious activity reports (they make objects of the context) are related to a number of terms (which make attributes of the context): a report is related to a term if the report contains this term. The data-set in Table 3 is an excerpt from the one we used in our research.

The central notion of FCA is that of a (*formal*) *concept*. The way one looks at concepts in FCA is in line with the international standard ISO 704, which gives the following definition. A concept is considered to be a unit of thought constituted of two parts: its *extent* and its *intent* (Wille 1982; Ganter and Wille 1999). The *extent* consists of all objects subsumed by the concept, while the *intent* consists of all attributes shared by those objects. Let us illustrate the notion of a formal concept for the formal context in Table 3. For a set of objects $O \subseteq G$, the set of their common attributes, denoted by O', is given by the following formula:

$$A = O' = \{m \in M | \forall g \in O : (g, m) \in I\}.$$

Take, for example, the attributes that describe report 5 in Table 3: 'expensive cars' and 'large amount of money'. By collecting all reports of this context that share these attributes, we get the set $O \subseteq G$ consisting of reports 2, 3, and 5. Formally, this operation is given by the following formula:

$$O = A' = \{g \in G | \forall m \in A : (g, m) \in I\}.$$

In other words, O is the set of all objects sharing all attributes of A, and A is the set of all attributes shared by all objects contained in O. Each such pair (O, A) is called a *formal concept* (or just *concept*) of the given context such that $A = O'$ and $O' = A$. The set A is called the intent, while O is called the extent of the (formal) concept (O, A).

There is a natural hierarchical ordering relation defined on the concepts of a given context that is called the subconcept–superconcept relation.

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

A concept $C_1 = (O_1, A_1)$ is called a *subconcept* of a concept $C_2 = (O_2, A_2)$ (or equivalently, C_2 is called a *superconcept* of a concept C_1) if and only if $C_1 \leq C_2$, i.e. the extent of C_1 is a subset of the extent of C_2 (or equivalently, the intent of C_1 is a superset of the intent of C_2). For example, the concept with intent 'expensive cars', 'large amount of money', and 'violence' is a subconcept of the concept with intent 'expensive cars' and

'large amount of money'. With reference to Table 3, the extent of the latter is composed of reports 2 and 3, while the extent of the former is composed of reports 2, 3, and 5.

The set of all concepts of a formal context ordered by the subconcept–superconcept relations makes a mathematical structure known as a *complete lattice*. The complete lattice of concepts is called the *concept lattice* of the context. A concept lattice is visualized by its *(labelled) line diagram*, which is based on the *covering relation* associated with the ordering relation. Recall that for a partial order $(P, \geq)$ and two elements $x,y \in P$ one has x C y (*x covers y*) if $x > y$ and there is no $z \in P$ such that $x > z, z > y$. A diagram uniquely determines a lattice, so one often uses these terms interchangeably. For example, the line diagram in Figure 1 represents the concept lattice of the formal context given by Table 3. The circles or nodes in this line diagram represent the formal concepts. 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 3 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 '*g*' to the node named '*m*.' For example, report 5 is described by the attributes 'expensive cars' and 'large amount of money'.

Retrieving the extent of a formal concept from a line diagram such as the one in Figure 1 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 intent 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
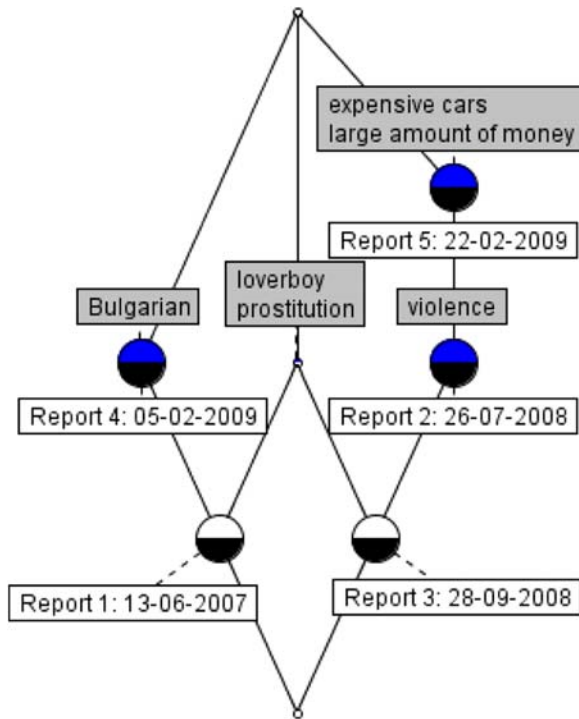

Figure 1.   Concept lattice diagram.

row 3 is defined by the attributes 'violence', 'expensive cars', and 'large amount of money'. The top and bottom concepts in the lattice are special: the top concept contains all objects in its extent, whereas the bottom concept contains all attributes in its intent. A concept is a subconcept of all concepts that can be reached by travelling upward and it will inherit all attributes associated with these superconcepts.

4. Dataset

Our data-set consists of 266,157 suspicious activity police reports, 34,817 in 2005, 40,703 in 2006, 53,583 in 2007, 69,470 in 2008, and 67,584 in 2009. These police reports are stored in the police databases as unstructured text documents and have the following associated structured data fields: title of the incident, project code assigned by the responsible officer, location of the incident, and optionally a formally labelled suspect, victim and/or other involved persons. The unstructured part of these suspicious activity reports describes observations made by police officers during motor vehicle inspections, during a police patrol, when a known person was seen at a certain place, etc. These reports were extracted from the database and turned into html documents that were indexed using the open source engine Lucene. An example of a report is displayed in Figure 2.

5. Method

Our semi-automated investigation procedure consists of multiple iterations through the square of Figure 3. For background information on FCA and its applications in KDD we refer the reader to Poelmans *et al.* (2010c). The guidelines of Section 2 contain a non-limitative list of indications, and the indications can be subdivided into several main
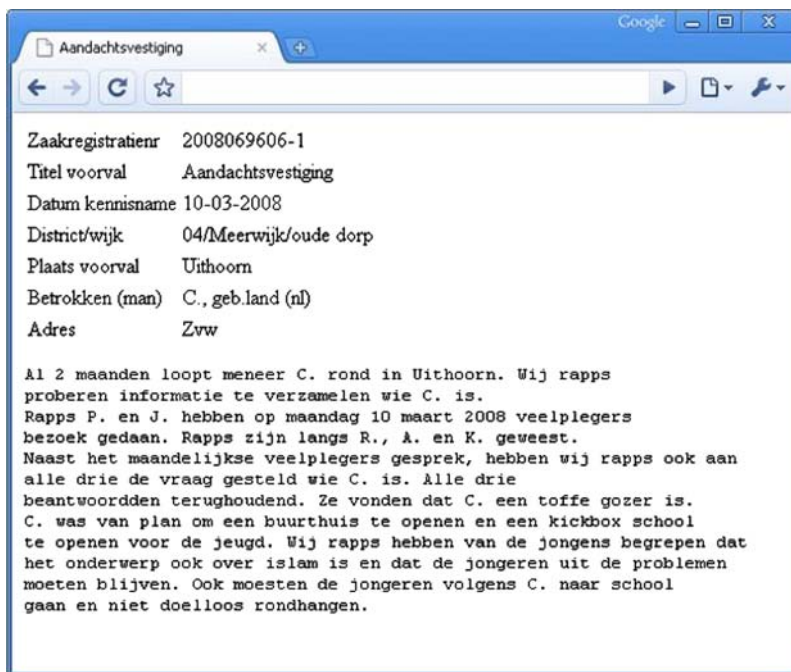

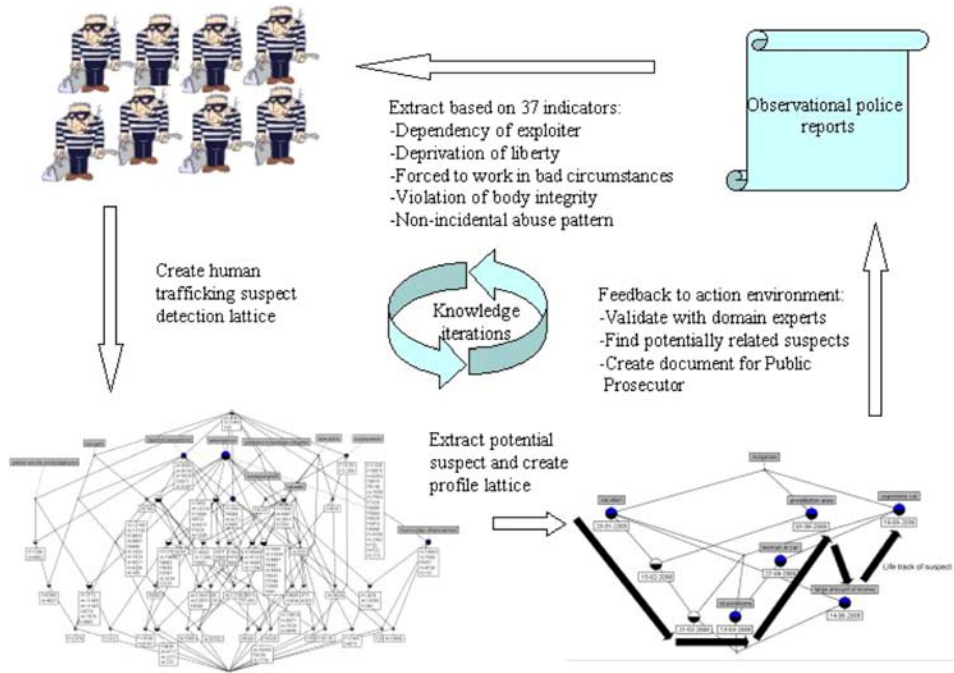Figure 2. Example of suspicious activity police report.

Figure 3.   Criminal intelligence process.

categories. If at least one of the thesaurus elements corresponding to one of these indications is present for a person or a group of persons, we might be dealing with a case of human trafficking or forced prostitution. These early warning indicators are cheap and reliable indicators that may indicate involvement of a person in illegal activities but may result in some false positives remaining, i.e. persons not involved in human trafficking. They serve to reduce the search space effectively without losing suspects. Then, in the reduced search space, concept lattices based on early and late indicators are created. The presence of a (set of) late indicator(s) is a strong hint that a person might be involved in illegal activities. Sometimes a combination of early indicators also presents an interesting challenge for further analysis. The concept lattice visualization allows the human expert to zoom in on aspects of the reduced search space and to interactively explore the data. He can steer the KDD process and the lattice partial ordering gives him clues on where to look first.

From the 266,157 reports in our data-set, the relevant reports which contain at least one indicator are selected. Then, the persons mentioned in these reports are extracted and concept lattices are created, showing all the indications observed for each person. From these lattices containing persons, potential suspects or victims can be distilled and they can be further analysed in detail with FCA and temporal concept lattices. If sufficient indications are available, a document based on article 273f of the code of criminal law can be created and sent to the public prosecutor with the request for using advanced intelligence gathering instruments such as observation teams, phone taps, and so on. If the suspects are indeed involved in human trafficking and forced prostitution they can be taken into custody.

5.1 *FCA-based analysis*

Our method based on FCA consists of four main types of analysis which are carried out as follows:

- Concept exploration of the forced prostitution problem of Amsterdam: in Poelmans *et al*. (2010a) and Poelmans, Elzinga, Viaene, and Dedene (2010b)), our FCA-based approach for automatically detecting domestic violence in unstructured text police reports is described in detail. We not only improved the domestic violence definition but also found multiple niche cases, confusing situations, faulty case labellings, etc. that were used among others to improve police training. Part of the research reported on in this study such as the construction of the thesaurus consisted of repeating the procedures described in our domestic violence case study papers.

- Identifying potential suspects: concept lattices allow for the detection of potentially interesting links between independent observations made by different police officers. When grouping suspicious activity reports on a per person basis, the available information about the individuals is displayed in one intuitive and understandable picture that facilitates efficient decision making on where to look. In particular, persons lower in the lattice can be of interest since they combine multiple early warning indicators.

- Visual suspect profiling: some FCA-based methods such as temporal concept analysis (TCA) were developed to visually represent and analyse data with a temporal dimension (Wolff 2005). Temporal concept lattices were used in Elzinga, Poelmans, Viaene, Dedene, and Morsing (2010) to create visual profiles of potentially interesting terrorism subjects. Elzinga, Wolff, Poelmans, Viaene, and Dedene (2012) used TCA in combination with nested line diagrams to analyse pedophile chat conversations. Scharfe, Oehrstrom, and Gyori (2009) used a model of branching time in which there are alternative plans for the future corresponding to any possible choice of a person and used it as the basis of an information and communication technology toolset for supporting autism diagnosed teenagers. For creating the temporal profile of individual suspects, we use traditional FCA lattices and the timestamps of the police reports on which these lattices are based are used as object names. The nodes of the concept lattice can then be ordered chronologically.

- Social structure exploration: concept lattices may help expose interesting persons related to each other, criminal networks, the role of certain suspects in these networks, etc. With police officers we discussed and compared various FCA-based visualization methods of criminal networks. Individual police reports mentioning network activity and the timestamps of these police reports together with each suspect name mentioned in these reports make object names.

5.2 *Thesaurus*

The thesaurus constructed for this research contains the terms and phrases used to detect the presence or absence of indicators in these police reports. This thesaurus consists of two levels: the individual search terms and the term cluster level that was used to construct the lattices in this work. We used a semi-automated approach as described in Poelmans *et al*. (2010a). Search terms and term clusters were defined in collaboration with experts of the anti-human trafficking team and gradually improved by validating their effectiveness on subsets of the available police reports. Each of these search terms was thoroughly analysed for being sufficiently specific. The quality of the term clusters was determined based on

their completeness. The validation of the quality of the thesaurus and the improvements were done by us and in conjunction with members of the anti-human trafficking team. Concept structures were created on multiple randomly selected subsets of the data. It was manually verified whether all relevant indicators were found in these reports and no indicators were falsely attributed to these reports. For example, the term cluster 'prostitute' in the end contained more than 20 different terms such as 'prostituee', 'dames van lichte zeden', 'prosti', 'geisha' and so on used by officers to describe a prostitute in their textual reports. To create the formal contexts in this paper, the term clusters in the thesaurus were used as attributes and the police reports as objects. A prototype of the FCA-based toolset COncept Relation Discovery and Innovation Enabling Technology (CORDIET) was used during the analysis process (Poelmans, Elzinga, Viaene, and Dedene 2010d; Poelmans *et al.* 2012a).

6.   Analysis and results

Traditional data mining techniques often focus on automating the knowledge discovery process as much as possible. Since the detection of actual suspects in large amounts of unstructured text police reports is still a process in which the human expert should play a central role, we did not want to replace him, but rather empower him in his knowledge discovery task. We were looking for a semi-automated approach and in this section we try to illustrate the main reasons why FCA was ideal for this type of police work. With FCA at the core, we were able to offer police officers an approach which they could use to interactively explore and gain insight into the data to find cases of interest to them on which they could zoom in or out. Section 6.1 shows two lattices which were of significant interest to investigators of the anti-human trafficking team. For the first time, the overload of observational reports was transformed into visual artefacts that first showed them a set of 4895 persons and a subset of 1255 eastern Europeans potentially of interest to the police and the indicators observed for each of them. The lattices visually summarize the data and make the data more easily accessible for officers who want to efficiently explore it and extract unknown suspects. When zooming in on the nodes on the left of the lattice in Figure 4, we found a concept with two underaged girls in its extent and with suspicious loverboy indications in its intent. Section 6.2 describes the analysis of the first girl which led us to the discovery of the first loverboy suspect. We showcase how the discovery of a potential victim was followed by querying our data-set for reports about this girl using CORDIET and analysing the found textual reports which led us to our pimp for whom a lattice summarizing available evidence in the data was created. Twelve reports and indicators found in them were used to create a lattice. This lattice showed that there was sufficient evidence for the officers to compose a document to obtain permission for special investigative techniques from the public prosecutor. Section 6.3 also shows how a concept lattice diagram can give insight into the evolution of a person over time, in this case of our second loverboy suspect. We then chose to highlight the case of the Turkish human trafficking network in Section 6.4. From the lattice in Figure 5, two potential suspects were distilled since they were regularly spotted carrying out illegal activities. We found that the name of a bar was mentioned a couple of times and used this information to build the concept lattice of Section 6.4. This lattice was of particular interest to police officers since FCA quickly gave them a concise overview of the persons who were observed to be involved around a suspicious location and the lattice structure helped them to identify the most important suspects in this network. In particular, the visualization of persons in a lattice was helpful during their exploration. The partial ordering on concepts gave them
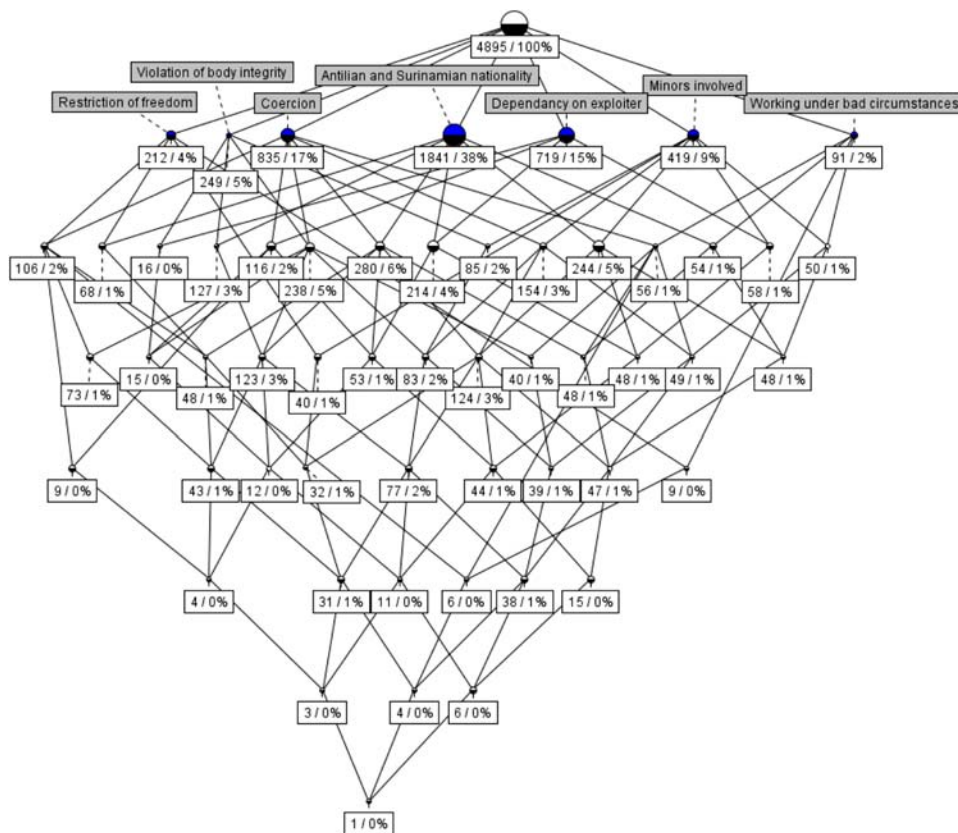
Figure 4.   Human trafficking suspect detection lattice diagram.

clues on where to look first. The lower a person appears in the lattice, the more indicators he has. Section 6.5 showcases how the FCA visualization was used to combine temporal and social structure information in one easy-to-interpret picture. Such profile lattices were of significant interest to police officers since they allow for quick decision making on whether a person might be involved in illegal activities. Moreover, the lattices may help infer the roles of the persons mentioned in the network. The fifth case in Section 6.6 is of interest, since it shows how a concept lattice can give insight into the evolution of a person over time, in this case, how to detect the special case of a woman who was the first victim and then became a suspect. The remaining part of this section describes cases of human trafficking and forced prostitution which were further investigated with FCA; two of them were identified in the lattice in Figure 4 and three of them in the lattice in Figure 5. Note that real names were replaced by false names because of privacy reasons.

6.1   *Detection of suspects of human trafficking and forced prostitution*

Multiple concept lattices were created for detecting human trafficking suspects in a set of persons. Each of these concept lattices contained over 200 concepts which were based on different sets of attributes. Since the format of this study does not allow to visualize the entire lattices in a readable way, we chose to simplify these lattices and zoomed in on their most important aspects. Our first lattice described the behaviour of 4895 persons extracted
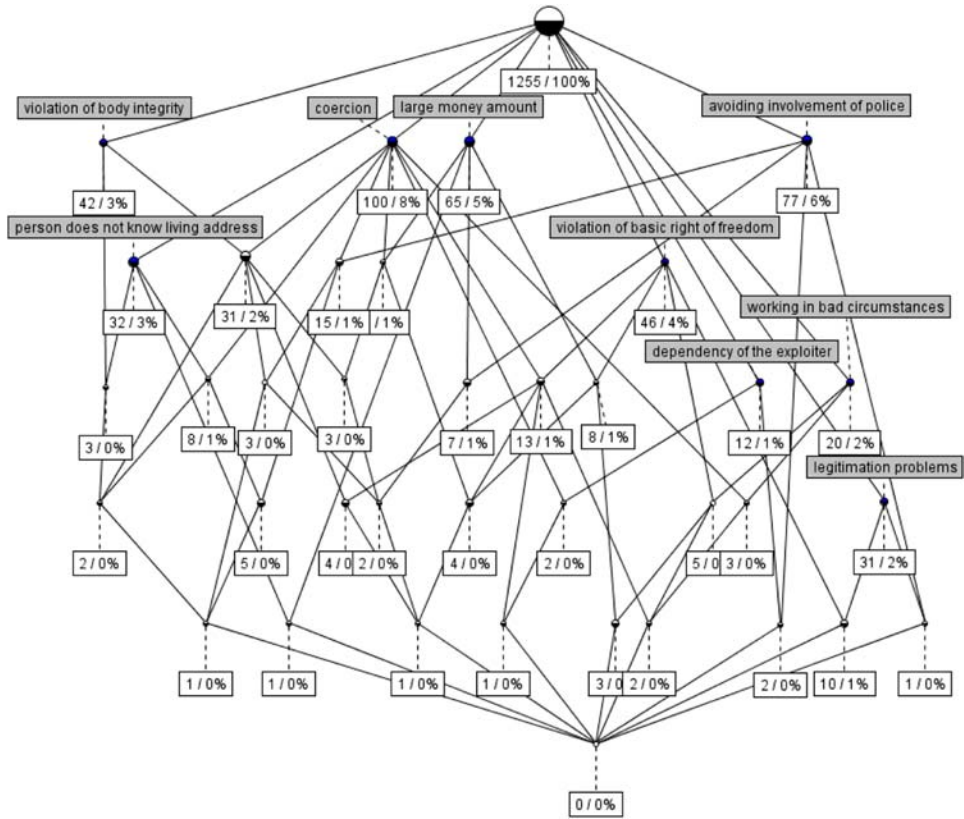
Figure 5.   Human trafficking suspect detection lattice.

from the police reports in our data-set. Each of them had at least one indicator. The attributes are based on the five types of indicators discussed in the models of Sections 2.1 and 2.2. Additional attributes can be selected and deselected during analysis. Analysis of the node containing 11 persons and attributes 'violation of body integrity', 'minors involved', 'dependency on exploiter', 'coercion', and 'Antillean and Surinamian nationality' in the lattice diagram revealed two girls of Dutch nationality who were younger than 18 years of age. This led us to the discovery of an unknown loverboy suspect described in Section 6.2. The second loverboy suspect was found by zooming in on the concept with four persons and attributes 'coercion', 'restriction of freedom', 'violation of body integrity', 'Antillean and Surinamian nationality', and 'dependency on exploiter', and is described in Section 6.3.

Figure 5 presents a diagram of the lattice for 1255 Bulgarian, Hungarian, and Romanian persons. The concepts related to some of the suspects of Section 6.4 were found on the right and bottom part of the lattice diagram and have 10 persons in its extent and attributes 'legitimation problems' and 'violation basic right of freedom', 12 objects in the extent and attribute 'dependency on exploiter', etc. The concept related to the main suspect of Section 6.5 was found on the left and bottom part of the lattice diagram, and has one object in its extent and attributes 'violation of body integrity', 'coercion', 'violation basic right of freedom', 'avoiding involvement of police', and 'large money amount'. The woman of Section 6.6 was found in the concept with one object in the extent and attributes

'coercion', 'violation of body integrity', 'person does not know living address', and 'avoiding involvement of police'. In the following two sections, we describe and profile each of these suspects in detail.

6.2 Case 1: Loverboy suspect B

In this section, we first describe the analysis of the underaged girl found in the lattice diagram of Figure 4. In our data-set, there were three reports available about girl H. The reports about this girl led us to the discovery of loverboy suspect B. Figure 6 shows the case numbers of the found reports in the 'search results' field and the first report with highlighting of thesaurus elements about this girl in the 'selected report' field.

The first report (26 November 2008) contains the notification of the police by a youth aid organization in Alkmaar about girl H. They report a suspicious tattoo on her wrist containing the name 'B'. This 'B' refers to her boyfriend who carries the same first name, is 30 years old, and is of Surinamian origin.

Figure 6. Possible loverboy victim.

Figure 7.   Observation of abnormal injuries and long working days.

The second report was written by a police officer who works in the red light district and knows many women working in brothels or behind the windows. During a patrol he saw H working as a prostitute, made a chat with her, and observed suspicious facts which made him write the report. The report in Figure 7 shows four suspicious facts reported by the officer. First, an unbelievable story why she works as a prostitute: a bet between girlfriends if someone would dare to work as a prostitute. Second, the tattoos of which one tattoo is mentioned in the document of Figure 6 ('B') and a new one on her belly. Third, the injuries, she has scratches on her arm (possibly from a fight) and burns on her leg. According to the victim, she has dropped a hot iron on her leg and had an accident with a gourmet set. Fourth is the observation of making long working days. The third document in Figure 8 (21 December 2008) shows an observation of the victim walking with the possible suspect.

In the report from Figure 8 the police officer reports that he saw the victim and a man walking close to each other. The police officer knows the man and knows that he is active in the world of prostitution. When the man saw the officer, he immediately took some distance of the victim. As soon as they have passed the officer, they walk close together and into a well-known street where prostitutes work behind the windows. The first name of the person is B, the same name which is tattooed on the victim's wrist, and the description of the person is about the same as described by the youth aid organization. This information signals that the man is the possible loverboy of the victim. The three reports together give serious presumptions of B being a loverboy with H being the victim. The next step is investigating B. We need serious indications that B is really involved in forced prostitution. Twelve observational reports were found for B and the resulting lattice is shown in Figure 9.

Figure 8.   Possible loverboy suspect.

Figure 9. Lattice diagram of B.

Investigating these reports shows that he frequently visits the red light district and has strong relationships with other pimps. One of these pimps is the suspect of the loverboy case in Section 6.3. From the six observations where B was seen in the red light district, four are violence related, including the observation of H's suspicious burn wounds. The other violence-related observations are situations of fights with customers who are unwilling to leave or pay. Such violence-related observations are related to pimps who want to protect their prostitutes from customers and competing gangs. In the Netherlands, prostitution is legal, so each prostitute has the right to ask the police to protect her. The violence observations of the suspect strengthened the suspicion of B being the pimp of H. Moreover, we found another girl R who was also a potential victim of him. These indications were enough to create a summary report and send a request for using special investigation techniques to the public prosecutor.

6.3 *Case 2: Loverboy suspect A*

In this section, we describe the case of a loverboy, which we exposed by gathering evidence from multiple observational reports. This person was found by analysing the lattice in Figure 4 by zooming in on Antillean, Moroccan, and Turkish persons. Victim V is a girl of Dutch nationality who officially lived in the Netherlands but fell prey to a loverboy of originally Antillean nationality. We found multiple indications in filed suspicious activity reports that referred to elements of the model in Section 2. The lattice of suspect A and victim V is displayed in Figure 10.

On 27 April 2006, suspect A and victim V were noticed for the first time on the streets during a police patrol. They had a serious argument with each other and suspect A took the cell phone with force out of V's hand. When the police intervened they claimed that
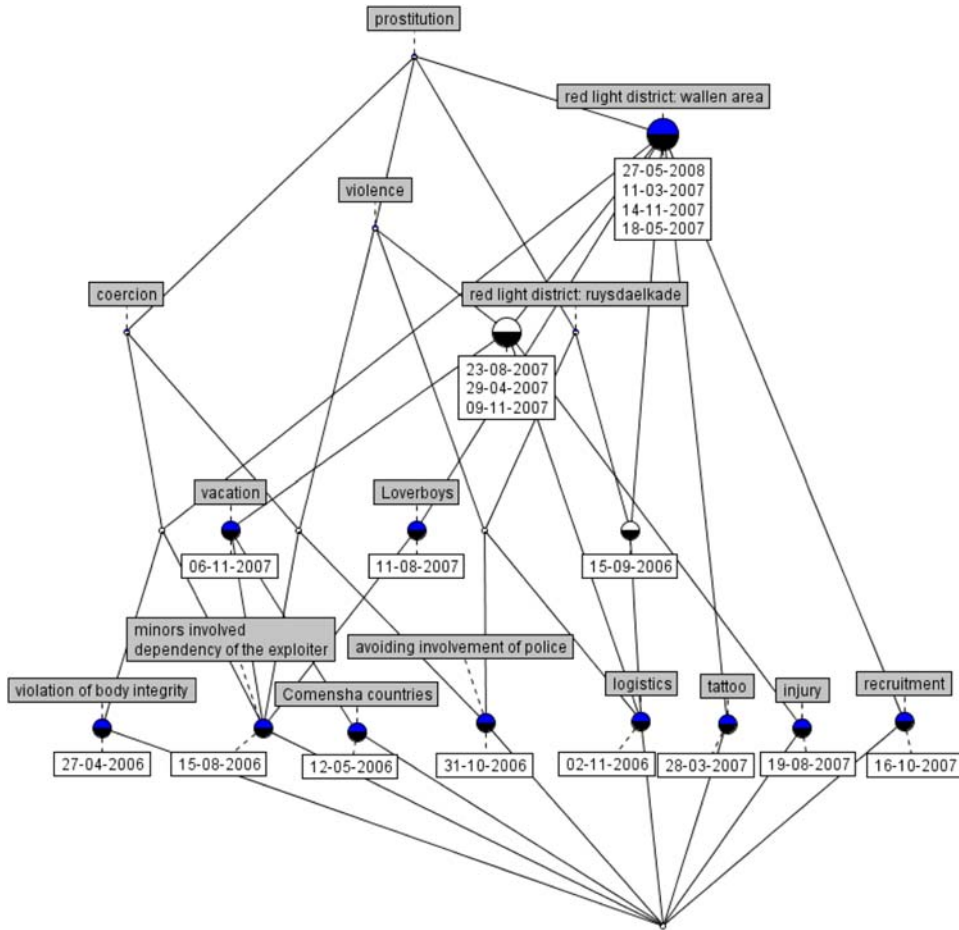
Figure 10.   Profile of loverboy suspect.

nothing had happened. In the police station, she declared that she works voluntarily in prostitution although her words were not convincing to the officer. On 15 August 2006, an Amsterdam citizen sent an email to the police about young Antillean men who constantly surveillance some women in the red light district. Among others suspect A brings food and drinks to the women who are not allowed to leave their rooms. On 31 October 2006 during a police patrol, victim V was noticed while she got out of a car and quickly ran inside. The driver of the car was suspect A. She told the police later on that she was brought to and picked up every day at this apartment by her boyfriend suspect A. The police noticed her dismayed and timid attitude and asked again whether she was forced to work in prostitution. In a non-convincing way she responded that she did her job voluntarily. On 15 September 2006, suspect A had to stay in jail for 6 h because of illegal weapon possession. When the police asked about his income he told that he earned good money thanks to his girlfriend who works in prostitution. On 2 November 2006, officers noticed that the car of victim V was parked on the road and two Negroid men were inside. The driver, suspect A, got out of the car and yelled to the girl he was picking up at her apartment that she had to hurry up. The whole scene looked very intimidating to the police and it turned out that the girl was victim V. Suspicious was that the car was registered on the name of V while V had

no driver license. On 28 March 2007, victim B came to the police office to ask whether she was allowed to work with a badly damaged id document or whether she had to wait for a new one. She mentioned that suspect A was her ex-boyfriend and that she and victim V were victims of extortion but she did not dare to make an official statement to the police. Afterwards, the police checked a home where they found two women: victims V and B. Victim V had a big tattoo on her right shoulder and a smaller tattoo on her upper arm. On 19 August 2007, suspect A was involved in a knifing incident in the red light district between three men and one of these men got seriously injured. This man wanted sex with victim V, but suspect A did not allow this because of the man's ethnicity, which caused the fight. On the camera surveillance videos, victim V was observed to accompany suspect A all the time. On 16 October 2007, officers observed that suspect A who walked over the streets said 'hi' to all women who passed by.

6.4   *Case 3: Turkish human trafficking network*

By analysing the diagram of the concept lattice (Figure 5) based on observational reports, we were able to expose a criminal network operating in Amsterdam, involved in illegal and forced prostitution. The concept lattice in Figure 11 contains 61 persons and indicators found in the police reports mentioning activity around a bar in Amsterdam that played a central role in the network's activities and was closed down in 2009. Multiple suspects operating in this network were found and some of the observations are described in this section. The most important suspects are the persons with indication 'legitimation problems', since they were carrying the id papers of the girls. The police reports contained many indications of illegal and forced prostitution taking place, activities that were run by the owners or acquaintances of the owners of the bar. We found out that the bar was used as a central hub, where mostly Turkish men met up with Bulgarian girls who had been forced into prostitution and took them to another location. We found at least two pimps who have multiple girls working for them.
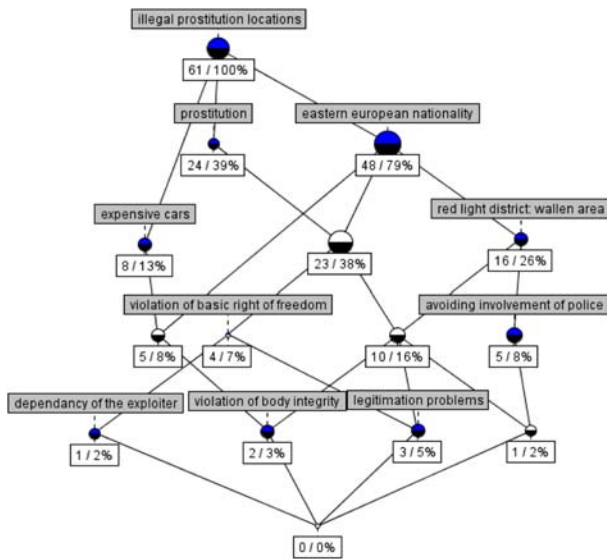

Figure 11.   Concept lattice of human trafficking network.

Starting in 2007, the first observations were made that hinted at illegal and forced prostitution being organized from within this bar. On 2 June 2008, victim H declared to the police that she was forced to work as a prostitute in the bar and did not get any money for that. She was never allowed to leave the house alone and the door of her apartment was locked from the outside such that she could not leave. On 12 December 2008, suspect A came out of the bar with a girl, their statements to the police did not match and moreover the girl was dressed in sexy clothing. Most likely the girl works as a prostitute and the driver is her pimp. On 25 January 2009, police officers stopped a car and behind the wheel was suspect B and next to him the victim E. We found woman E often sitting at the bar and also the car being regularly parked in front of the bar. Suspect B gave the passport of victim E to the police and afterwards he placed it back in his pocket. Moreover, suspect B was carrying a large amount of cash money, 1000€ in his pocket. On 26 January 2009, police did a check-up on the guests in the bar. One girl was new and told she just arrived by train, she had no train tickets with her and she did not know her living address. Suspect B was also there and told the police that he is a car trader so he travels a lot between Bulgaria and Netherlands. An excuse typically used by criminals responsible for the logistics of a trafficking network. Also victim E and two other girls, victims F and G, were there. On 20 February 2009, police officers saw suspect A talking to the driver of a car with Bulgarian licence plate. Afterwards he forced a girl to follow him and when the police asked about their relationship they told that they had been friends for 3 months. The girl did not have her id papers with her and the police went to her living address. In the house, there were many mattresses and another girl. Both of them told that they have no job. Most likely the house serves as an illegal prostitution location for the criminal gang.

Sufficient indications were found and on 17 June 2009 an observation team observed the bar during the evening. Eastern European women were sitting at the bar and mostly Turkish, Moroccan, and Eastern European men at the tables. During the evening, the team saw multiple girls who were taken out of the bar by a customer to a hotel, house, etc. and were brought back to the bar afterwards. On 15 July 2009, sufficient evidence was gathered that illegal prostitution was organized from within this bar and authorities closed down the bar.

6.5   *Case 4: Bulgarian male suspect*

In this section, we describe a profile of a Bulgarian suspect who was also operating in Amsterdam. The lattice in Figure 12 shows that on 3 October 2007, suspect A was observed for the first time during a police patrol. An officer told the driver of a BMW car with Bulgarian licence plate to turn right instead of left, the driver, however, ignored the instructions he received and quickly drove to the left with squeaking tyres. The officer went after and in the end stopped the car. There were three men and one woman in the car. Suspect B was the driver and suspect A was sitting next to him. On the backseat of the car were woman F and man K. They told the officer that they arrived only 3 days ago in the Netherlands and are a couple. Suspect A and suspect B were taken to the police office, the man and the woman walked away and were followed by a second officer. He saw that K was strongly holding the hand of F and forced her into a home at the corner of a street in central Amsterdam. In the police office, suspect B was not able to tell the address of the apartment he was going to rent. Suspect A was carrying a large amount of cash money in his pocket.

On 30 June 2009, woman J went to the police to ask whether they could supervise the undersigning of a tenancy agreement of an apartment by man M who promised her
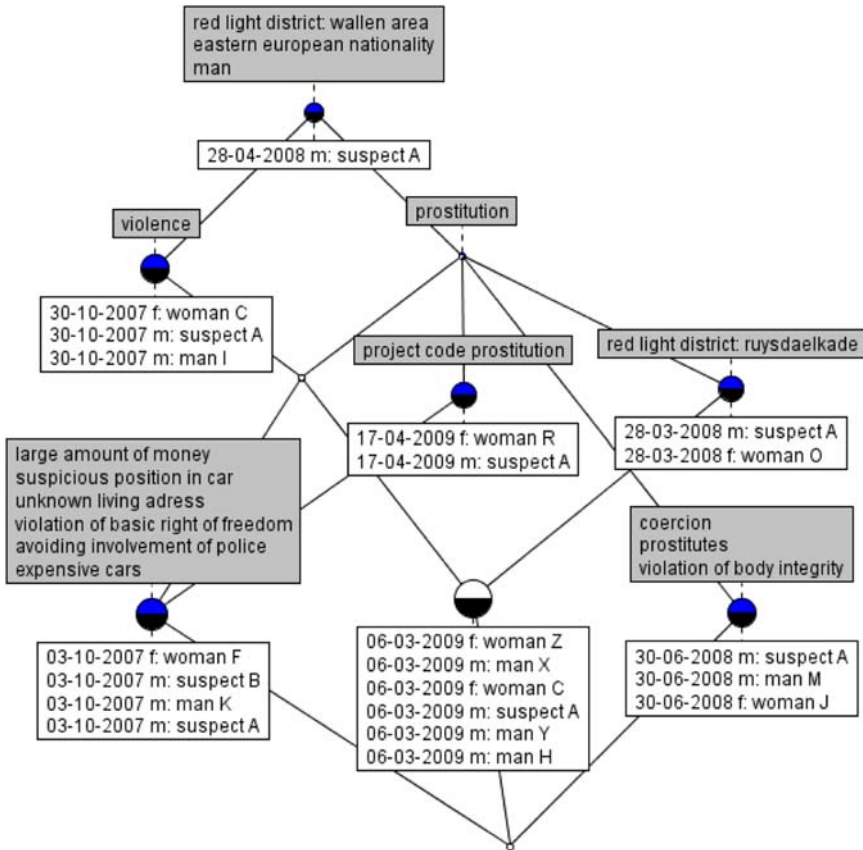
Figure 12. Profile lattice of individual suspect and his network.

accommodation. She told that suspect A was intimidating and trying to scare away man M because suspect A wanted to rent the apartment for prostitution purposes. She was very afraid of suspect A and the officer noted that she might have been forced in prostitution by him. On 30 October 2007, the police did a routine inspection of two individuals who were waiting with two motorcycles in a street that had been plagued by street robberies. This was the second observation of suspect A by the police and his motorcycle was registered by the name of woman C who had been involved in human trafficking activities as a victim. On 6 March 2009, the police received a tip that a fugitive Colombian criminal might be living at a certain address owned by professional criminal H. When they entered the apartment they found two men and two women of Bulgarian nationality. Man X and woman C declared to be on holiday and would go back to Bulgaria, although we found that suspect A was driving around with a scooter registered at C's name in 2007. Man Y declared that he exports expensive cars to Bulgaria and regularly drives back and forth between Netherlands, an excuse typically used by suspects taking care of logistics of a human trafficking gang. Woman Z declared to work in prostitution in Groningen. When the officers left the apartment they found a motorcycle registered on the name of suspect A. The last observation dates back to 17 April 2009 when the police saw suspect A call somebody while standing in the entrance hall of prostitute R. He tells the police that he has

nothing to do with prostitution and owns a restaurant in Bulgaria. After his phone call he gives the cell phone to the prostitute.

To conclude, suspects A and B are most likely involved in human trafficking and there were sufficient signals found to request the use of special investigation techniques. Permission was granted, suspicions were confirmed, and both A and B were arrested by the police in 2010. Moreover, these lattices showed some other people who are involved in the same gang and could be monitored.

6.6 *Case 5: Hungarian woman both victim and suspect*

In this section, we describe a girl who was first a victim and then became a suspect of human trafficking. The concept lattice in Figure 13 contains indications that SV1 has been forced to work in prostitution but now also takes part in criminal activities such as 'facilitating' new girls in the prostitution circuit.

On 16 March 2006, woman SV1 was observed for the first time by the police in the red light district. She did not speak Dutch, English, or any other language spoken by police officers in the Netherlands. She had some indications of a woman who was lured into prostitution in her home country and trafficked to the Netherlands by a criminal gang. On 18 June 2006, the id papers of SV1 and another girl I were checked and both pictures were
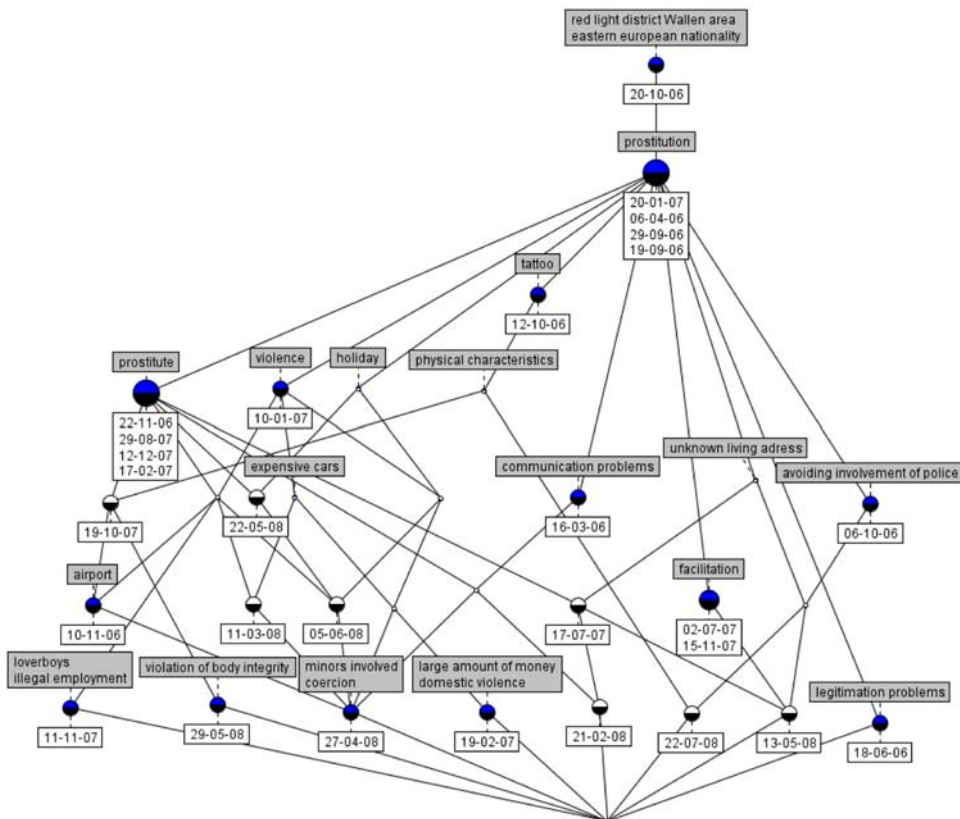

Figure 13. Profile lattice of women who were the first victim and then became suspect.

very similar and had almost nothing in common with SV1 or I. Their id cards were counterfeit, something regularly done by criminal gangs who took away their real identity papers. On 19 February 2007, prostitute Q declared to the police that she had to give all her money to a Hungarian pimp who worked for a large criminal network. She also told that SV1 works for one of the pimps of this network and most likely undergoes the same treatment. On 19 October 2007, SV1 was observed with a new tattoo. Tattoos are regularly used by gangs to clearly show whose property the girl is. On 29 May 2008, officers saw that SV1 underwent a breast enlargement.

From 2007 onwards, police officers started to see more and more indications that SV1 is becoming a perpetrator herself by facilitating girls in the prostitution circuit. On 02 July 2007, officers noticed that SV1 always pays the rent of the prostitution room for a new Hungarian girl L. On 17 July 2007, the police asked for the id card of a woman unknown to them who was working as a prostitute and resided in the Netherlands since 14 days. She did not know her living address, she lives with SV1 and was brought every day from and to her working place by SV1. The police asked whether she likes her job, but she had a very despairing look and could not answer their question. On 11 November 2007, police went to the lodging-house keeper of a room often rented by a Dutch girl D who worked in prostitution but mysteriously disappeared for multiple weeks. She told that she was threatened by a group of loverboys whom she met through SV1. They were trying to force her to work for them and give the money she earns away, among others through blackmailing, threatening, and emotional manipulation. Among others on 15 November 2007, police saw SV1 having long conversations with Hungarian men for whom she most likely worked. She was granted more liberty than the other girls and seemed to function as a kind of supervisor over the new girls who came into the business. On 13 May 2008, police did a routine inspection of three girls in the red light district but they only spoke Hungarian, and SV1 was asked to translate their questions. When the police asked the girls about the place where they lived, they became very nervous, tried to invent the name of a hotel, etc. In the end they asked to SV1 whether they could tell their real address, but SV1 answered no and if the police would try to force them they must first call the men of their network to ask for permission. On 22 July 2008, officers did a routine inspection in the red light district. Woman C was found to live together with SV1 and when the police asked her about their living address, C turned to SV1 who said in Hungarian 'say whatever you want but don't tell the address'.

SV1 has many indications of a former victim of forced prostitution who had no better choice than becoming part of the criminal activities herself. She was part of a big network of Hungarian criminals who might be of interest to the police. In the beginning of 2011, the police gathered enough evidence against her, and then she was arrested and now serves her time in a Dutch prison.

7.   Discussion

Human-centred data mining focuses on making the human expert efficiently interact with the data by supporting him instead of trying to replace him. We wanted to help him in the laborious task of searching through the police reports and coming up with potential suspects but did not want to decide for him who should be investigated. The main goal of our semi-automated KDD in unstructured text approach is the active involvement of the human expert who steers the knowledge discovery process, sifts through the data and is supported in his decision making by visualizations that make the massive amount of data that used to numb domain experts accessible again.

Our real-life validation setting allowed us to understand which aspects of FCA were particularly interesting to users, in this case police officers:

- Summarization of conceptual structure of data in one picture: the lattices of Section 6.1 were used to showcase this appealing aspect of FCA. The overload of reports was turned into an intuitively analysable artefact.
- An effective means to zoom in and out of the data: from the lattices in Section 6.1, multiple persons were picked out and analysed in detail in the subsequent sections.
- Intuitive visualization with a partial ordering of the persons based on the indicators observed: Police officers were guided by the partial ordering of concepts when analysing the lattices in Section 6.1. From the analysis it was indeed revealed that they had more evidence to start a case against suspects lower in the lattice than suspects higher in the lattice.
- Conceptual relationships between individual documents, persons, timestamps, etc. became visible, whereas they often stay hidden when individual documents are analysed one by one: the lattice of Section 6.4 was used to showcase how a criminal network operating in Amsterdam was exposed. Multiple independent observations contained indications that illegal network activity was carried out around one central location.
- Visualization of temporal evolution of a person: the lattices in Sections 6.2, 6.3, and 6.5 showed the evidence that became available over time against several human trafficking and loverboy suspects. Section 6.6 showed how a woman was first a victim and later on became a human trafficking suspect.

The literature on data mining describes many fully automated approaches for thesaurus building, classification, visualization, etc. Fully automated approaches have proven their usefulness for the analysis of certain crimes and criminals such as the identification of a serial killer's living address (VICLAS, Collins, Johnson, Choy, Davidson, and Mackay 1998). The algorithm is based on a domain with clear underlying rules and concepts and takes as input a carefully prepared large amount of structured information about the suspect. The powerful pattern matching and computational capabilities of the computer clearly outperform the human expert in this task.

Unfortunately, in complex domains such as the domain described in this paper, it is very difficult if not impossible to be successful with pure automated analysis techniques. Many of these automated techniques may have serious drawbacks for complex domains with one or more of the following properties:

- Black-box classification is not acceptable: police officers need insight into the reasons behind a decision, behind an assigned label, etc. Each decision to label a suspect should be grounded in evidence and be accompanied by a detailed report of the indications observed. False positives and false negatives, i.e. real criminals who remained undetected, are unacceptable given the severity of the crimes in which the persons are potentially involved and the penalties they may receive.
- Texts are short, of equal length, and written by authors with different writing styles: this makes it impossible and useless to apply term extraction techniques such as frequency analysis. The terms we obtained through software packages such as DataDetective and Clementine were not satisfying either. Advanced Natural Language Processing (NLP) techniques were tried out in the past but failed because of the shortness of textual reports. Relationships between persons, documents, and networks play an essential role but are hard or impossible to automatically distil

from the texts. An essential element to the success of a text mining approach is a high-quality thesaurus. We chose for a semi-automated thesaurus-building approach and complemented it with following automated methods to maintain quality: word stemming, using synonym lists, spell checking, etc. We also used Named Entity Recognition for extracting licence plates, suspect names, etc.

- Contexts of words and phrases are essential for interpretation of the data: the interpretation of words, phrases, etc. is often strongly dependent on the context in which they are used. For example, during a police patrol, an officer checks a new prostitute and asks her about the scars on her legs. He wrote down that she told that during her childhood she was sexually abused and beaten but then suddenly their conversation was interrupted by the pimp who brought her food. The attributes 'sexually abused', 'pimp', 'bring food', and 'scars' may lead to a false positive, although this document alone is far from sufficient to start a forced prostitution case. Moreover, multiple persons are mentioned in many reports and their roles such as suspect, victim, or both are difficult to distil from these reports, even with advanced NLP instruments. Also some attributes should be solely attributed to one person but often it is impossible to automatically infer to which one. Human decision making remains necessary.

- Only little information is available per person and the target group is a small fraction of the total population. The information we have is naturally incomplete since the reports written by officers describe only a part of the reality, namely that part observed by them during their work. The police does only have information about fragments of these persons' lives based on which they decide whether this person might be interesting. Given the incompleteness of the information, one should take caution with fully automated decision making and leave this critical task to specialized and trained police officers who can judge whether sufficient evidence is available and slightly vary their decision criteria based on their years of experience in the field. The focus of our approach lies on the development of an early warning system that helps to reduce the pool of potential suspects, gather all information about them in one visual picture that supports the officers in efficient decision making on which case should or should not receive special attention.

- There have been no labels assigned to individuals or reports: our data did not contain any labelled individuals. Moreover, the target group is a small fraction of the total population. Training an automated classifier became impossible. To identify phrases referring to forced prostitution during thesaurus construction we had to rely on expert knowledge.

- The underlying concepts of the domain are unclear: the conceptual relationships between persons, documents, locations, etc. were of significant importance and had to be made visible to officers since they are essential to decision making. Many visualization techniques such as self-organizing maps only give a distribution of the persons, documents, etc. but the relationships between them are not explicitly shown.

A potential issue and avenue for future research is scalability of the approach. Lattice diagrams are only readable until a certain number of concepts. Therefore, in each lattice we must limit the number of attributes and/or objects. This was, however, not a serious problem in our case since we were working with a stationary data-set, in which only a small part of the individuals were of interest. For other types of crimes such as credit card fraud detection, where we are dealing with massive amounts of fast changing data, FCA

should be complemented with other visualization techniques such as emergent self-organizing maps. Another issue is the potential evolvement over time of the terms and phrases used by officers to describe their observation. Our thesaurus may become incomplete and maintenance methods should be developed to keep our system up to date on the long term.

8. Conclusions

Textual documents contain a lot of useful information that is rarely turned into actionable knowledge by the organizations that own these data repositories. In this study, we proposed an approach to knowledge discovery from unstructured texts using FCA. The semi-automated exploration process is essentially human-centred. In this study, we argued for the discovery capabilities of FCA acting as an information browser in the hands of human analysts. The tool was shown to help analysts proceed with knowledge expansion by progressively looping through four main phases. We demonstrated the method using a real-life case study with data from the Amsterdam-Amstelland police. Police forces in the Netherlands dispose of a large amount of such textual reports that may contain early warning indicators that can help to proactively identify persons involved in illegal activities. Since the observations of one suspect are typically made by different officers who are not aware of each other's work, spread over multiple databases, etc. automated analysis techniques such as FCA can be of significant importance for police forces who are interested in the proactive identification of perpetrators. FCA is one of the few techniques that can be used to interactively expose, investigate, and refine the underlying concepts and relationships between them in a large amount of data. In this study, we described our successful application of FCA to find suspects of human trafficking and forced prostitution in the Amsterdam-Amstelland police district. From 266,157 observational reports we distilled multiple suspicious cases of which 5 have been described in this paper. For each of these persons and networks a document was composed, containing all the indicators and evidence available. Permission to use special investigation techniques was obtained by the anti-human trafficking team based on the identified indications. For each case we exposed, phone-taps, observation teams, etc. indeed confirmed the suspect's involvement in human trafficking and forced prostitution. We believe that in making the shift from reactive police work, where action is only undertaken when a victim comes to talk directly to the police, to the pro-active identification of suspects, FCA can play an important role. Interesting avenues for future research include applying concept selection techniques such as stability indices (Kuznetsov and Ignatov 2007). We will also investigate the potential of automated regression and classification techniques for assisting in the identification of relevant suspects (Dejaeger, Verbeke, Martens, and Baesens 2012a; Dejaeger, Verbraken, and Baesens 2012b).

Note

1. Eerste rapportage Nationaal Rapporteur Mensenhandel, http://www.bnrm.nl/Images/Rapportage%201%20(Ned)_2002_tcm63-83113.pdf.

Notes on contributors

Dr Jonas Poelmans graduated in 2007 as a 'Master in Computer Science' with distinction at the 'Katholieke Universiteit Leuven'. In 2010, he obtained his degree of 'Doctor in Applied Economics' at the 'Katholieke Universiteit Leuven'. In 2012 he was conference chair of the 10th International Conference on Formal Concept Analysis. He authored more than 45 papers published in international peer-reviewed conferences and journals. He won twice the best paper award at the 'Industrial Conference on Data Mining' and edited several conference and workshop proceeding volumes. His main interests include Formal Concept Analysis and Data Mining.

Dr Paul Elzinga received his MSc degree in Econometrics from the University of Groningen in 1984 and his MSc degree in Knowledge and Information Engineering from the Middlesex University of London in 1995. He received his PhD degree in Economics and Business science at the University of Amsterdam in 2011. He started his career with developing information systems for the Groningen Police Department in 1983. He developed many information systems of various kinds from dispatching systems to knowledge-based systems for monitoring potential criminals. He was information architect at the national police department from 2002 to 2005 and designed the XML-based national full text enquiry database in which all incidents and all criminal records of the Dutch police are made available for information analysis. In 2005, he continued his career at the Amsterdam-Amstelland Police Department where he worked on knowledge-based information systems and started his PhD entitled 'Formalizing the concepts of crimes and criminals' in 2007. At present, he is working as a project leader in several innovative data- and text-mining projects within different police organizations.

Dr Dmitry Ignatov works as Seniour Lecturer for State University Higher School of Economics at the chair of Artificial Intelligence and Data Analysis. Dr. Dmitry Ignatov graduated in 2004 as a "Specialist in Physics and Mathematics" with distinction at the "Kolomna Teachers' Training Institute" (Russia, Kolomna) and in 2008 as a "Master of Applied Mathematics and Information Sciences" at the "State University Higher School of Economics" (Russia, Moscow). In 2010 he obtained his degree of "Candidate of sciences in Mathematical Modeling, Numerical Methods, and Software Systems" at the "National Research University Higher School of Economics". He did his PhD (Candidate of science in Russian) research in All-Russian Institute for Scientific and Technical Information specializing in Theoretical Computer Science. He also was a guest researcher as a PhD student of the Postgraduate Program "Specification of Discrete Processes and Systems of Processes by Operational Models and Logics", Department of Computer Science, Dresden University of Technology. He is an author of more than 25 papers published in peer reviewed conferences, workshops and journals. His main interests include Formal Concept Analysis, Data Mining and Machine Learning, especially biclustering and multimodal clustering. He was a co-organizer of several international conferences and workshops: ICCS 2009, RSFDGrC 2011, PReMI 2011, CDUD 2011, SCAKD 2011.

Prof Dr Sergei O. Kuznetsov is the head of Applied Mathematics and Information Science Department at the National Research University Higher School of Economics in Moscow, Russian Federation since 2006. In 1985, he graduated from the Moscow Institute for Physics and Technology, Department of Applied Mathematics and Control. In 1990, he obtained his Candidate of Science Degree (PhD equivalent), dissertation 'Improving Performance of Expert Systems'. In 2002, he obtained his Doctor of Science Degree (Habilitation) for 'Theory of Learning in Concept Lattices'. Between 1985 and 2006, he was (Senior) Researcher at the All-Russia Institute for Scientific and Technical Information (VINITI), Moscow. Between 1999 and 2004, he was Humboldt Fellow and invited professor at the Dresden Technical University.

References

Becker, K., Stumme, G., Wille, R., Wille, U., and Zickwolff, M. (2000), 'Conceptual Information Systems Discussed Through an IT-Security Tool,' *In*: *Proceedings of 12th European Workshop on Knowledge Acquisition, Modeling and Management*, EWKAMM, LNAI 1937, London: Springer-Verlag, pp. 352–365.

Belohlávek, R., Sigmund, E., and Zacpal, J. (2011), 'Evaluation of IPAQ Questionnaires Supported by Formal Concept Analysis,' *Information Sciences*, 181, 10, 1774–1786.

Bovenkerk, F., Van San, M., Boone, M., Van Solinge, T.B., and Korf, D.J. (2004), *'Loverboys' of Modern Pooierschap in Amsterdam*, Utrecht: Willem Pompe Instituut voor Strafwetenschappen.

Brachman, R., and Anand, T. (1996), 'The Process of Knowledge Discovery in Databases: A Human-Centered Approach,' in *Advances in Knowledge Discovery and Data Mining*, AAAI/MIT Press, pp. 37–58.

Bullens, R.A.R., and Van Horn, J.E. (2000), 'Daad uit Liefde: Gedwongen Prostitutie van Jonge Meisjes,' *Justitiele Verkenningen*, 26, 6, 25–41.

Carpineto, C., and Romano, G. (2004), 'Exploiting the Potential of Concept Lattices for Information Retrieval with CREDO,' *Journal of Universal Computing*, 10, 8, 985–1013.

Cimiano, P., Hotho, A., Stumme, G., and Tane, J. (2004), 'Conceptual Knowledge Processing with Formal Concept Analysis and Ontologies,' *In*: *Proceedings of International Conference on Formal Concept Analysis*, LNAI 2961, Berlin: Springer, pp. 189–207.

Cole, R.J. (2000), '*The Management and Visualization of Document Collections Using Formal Concept Analysis*,' Ph.D. thesis, Griffith University.

Collier, P.M. (2006), 'Policing and the Intelligent Application of Knowledge,' *Public Money & Management*, 26, 2, 109–116.

Collins, P.I., Johnson, G.F., Choy, A., Davidson, K.T., and Mackay, R.E. (1998), 'Advances in Violent Crime Analysis and Law Enforcement: The Canadian Violent Crime Linkage Analysis System,' *Journal of Government Information*, 25, 3, 277–284.

Dejaeger, K., Verbeke, W., Martens, D., and Baesens, B. (2012a), 'Data Mining Techniques for Software Effort Estimation: A Comparative Study,' *IEEE Transactions on Software Engineering*, 38, 2, 375–397.

Dejaeger, K., Verbraken, T., and Baesens, B. (2012b), 'Towards Comprehensible Software Fault Prediction Models Using Bayesian Network Classifiers,' *IEEE Transactions on Software Engineering*, doi: 10.1109/TSE.2012.20.

Dettmeijer-Vermeulen, C.E., Boot-Mattthijssen, M., Van Dijk, E.M.H., De Jonge van Ellemeet, H., and Smit, H. (2008), *Mensenhande. Aanvullende kwantitatieve gegevens, Zesde rapportage van de Nationaal Rapporteur*. http://www.bnrm.nl/publicaties/zesde/.

Doerfel, S., Jäschke, R., and Stumme, G. (2012), 'Publication Analysis of the Formal Concept Analysis Community,' *Lecture Notes in Computer Science*, 7278, 77–95.

Domingo, S., and Eklund, P. (2005), 'Evaluation of Concept Lattices in a Web-Based Mail Browser,' *In*: *Proceedings of International Conference on Conceptual Structures*, LNAI 3596, Heidelberg: Springer-Verlag, pp. 281–294.

Eisenbarth, T., Koschke, R., and Simon, D. (2003), 'Locating Features in Source Code,' *IEEE Transactions on Software Engineering*, 29, 3, 210–224.

Eklund, P., Ducrou, J., and Brawn, P. (2004), 'Concept Lattices for Information Visualization: Can Novices Read Line Diagrams,' *In*: *Proceedings of International Conference on Formal Concept Analysis*, LNAI 2961, Berlin:Springer-Verlag, pp. 57–73.

Elzinga, P., Poelmans, J., Viaene, S., Dedene, G., and Morsing, S. (2010), 'Terrorist Threat Assessment with Formal Concept Analysis,' *In*: *Proceedings of IEEE International Conference on Intelligence and Security Informatics 2010*, Vancouver, Canada, 23–26 May pp. 77–82.

Elzinga, P., Wolff, K.E., Poelmans, J., Viaene, S., and Dedene, G. (2012), 'Analyzing Chat Conversations of Pedophiles with Temporal Relational Semantic Systems,' *In*: *Contributions to 10th International Conference on Formal Concept Analysis*, Leuven, Belgium, 6–10 May pp. 82–97.

Equality Division, Directorate General of Human Rights of the Council of Europe (2006), 'Action Against Trafficking in Human Beings: Prevention, Protection and Prosecution,' *In*: *Proceedings of the Regional Seminar*, Bucharest, Romania, 4–5 April.

Farley, M., and Barkan, H. (1998), 'Prostitution, Violence Against and Posttraumatic Stress Disorder,' *Women and Health*, 27, 37–49.

Fayyad, U., and Uthurusamy, R. (2002), 'Evolving Data Mining into Solutions for Insights,' *Communications of the ACM*, 45, 8, 28–31.

Ganter, B., and Wille, R. (1999), *Formal Concept Analysis: Mathematical Foundations*, Berlin: Springer-Verlag.

Hereth, J., Stumme, G., Wille, R., and Wille, U. (2003), 'Conceptual Knowledge Discovery – A Human-Centered Approach,' *Applied Artificial Intelligence*, 17, 281–302.

Hollywood, J., Strom, K., and Pope, M. (2009), 'Can Data Mining Turn Up Terrorists?' *OR/MS Today*.

Hughes, D.M. (2000), 'The "Natasha" Trade: The Transnational Shadow Market of Trafficking in Women,' *Journal of International Affairs*, 53, 2.

Hughes, D.M., and Denisova, T. (2003), 'Trafficking in Women from Ukraine,' US Department of Justice Research Report.

Hyde, L., and Denisenko, M. (1997), 'Modern-Day Slavery Traps Local Women,' *Kyiv Post*, 9 October.

Ignatov, D.I., and Kuznetsov, S.O. (2008), 'Concept-Based Recommendations for Internet Advertisement,' *In Proceedings of the Sixth International Conference Concept Lattices and Their Applications*, CLA Olomouc: Palacky University, pp. 157–166.

Ignatov, D.I., and Kuznetsov, S.O. (2009), 'Frequent Itemset Mining for Clustering Near Duplicate Web Documents,' *In Proceedings of the 17th International Conference on Conceptual Structures*, ICCS, LNAI 5662, Berlin: Springer-Verlag, pp. 185–200.

Keim, D.A. (2002), 'Information Visualization and Visual Data Mining,' *IEEE Transactions on Visualization and Computer Graphics*, 8, 1, 1–8.

Kuznetsov, S.O., and Ignatov, D.I. (2007), 'Concept Stability for Constructing Taxonomies of Web-site Users,' *In*: *Proceedings of Satellite Workshop 'Social Network Analysis and Conceptual Structures: Exploring Opportunities' at the 5th International Conference Formal Concept Analysis (ICFCA'07)*, Clermont-Ferrand, France pp. 19–24.

Lakhal, L., and Stumme, G. (2005), 'Efficient Mining of Association Rules Based on Formal Concept Analysis,' in *Formal Concept Analysis*, LNAI 3626, Berlin: Springer-Verlag, pp. 180–195.

Levchenko, K. (1999), *Combat of Trafficking in Women and Forced Prostitution: Ukraine, Country Report*, Vienna: Ludwig Boltzmann Institute of Human Rights.

Marchionini, G. (2006), 'Exploratory Search: From Finding to Understanding,' *Communications of the ACM*, 49, 4, 41–46.

Motameny, S., Versmold, B., and Schmutzler, R. (2008), 'Formal Concept Analysis for the Identification of Combinatorial Biomarkers in Breast Cancer,' *In*: *Proceedings of International Conference on Formal Concept Analysis*, pp. 229–240.

O'Neill Richard, A. (1999), *International Trafficking to the United States: A Contemporary Manifestation of Slavery and Organized Crime*, DCI Exceptional Intelligence Analyst Program, An Intelligence Monograph. Central Intelligence Agency.

Poelmans, J., Elzinga, P., Viaene, S., and Dedene, G. (2010a), 'Curbing Domestic Violence: Instantiating C-K Theory with Formal Concept Analysis and Emergent Self Organizing Maps,' *Intelligent Systems in Accounting, Finance and Management*, 17, 167–191.

Poelmans, J., Elzinga, P., Viaene, S., and Dedene, G. (2010b), 'Formally Analyzing the Concepts of Domestic Violence,' *Expert Systems with Applications*, 38, 3116–3130.

Poelmans, J., Elzinga, P., Viaene, S., and Dedene, G. (2010c), 'Formal Concept Analysis in Knowledge Discovery: A Survey,' *In*: *Proceedings of International Conference on Conceptual Structures*, ICCS, LNAI, 6208, 26–30 July Kuching, Sarawak, Malaysia: Springer, pp. 139–153.

Poelmans, J., Elzinga, P., Viaene, S., and Dedene, G. (2010d), 'Concept Discovery Innovations in Law Enforcement: A Perspective,' in *IEEE Computational Intelligence in Networks and Systems Workshop (INCos 2010)*, Thesalloniki, Greece.

Poelmans, J., Elzinga, P., Dedene, G., Viaene, S., and Kuznetsov, S.O. (2011), 'A Concept Discovery Approach for Fighting Human Trafficking and Forced Prostitution,' *In*: *Proceedings of International Conference on Conceptual Structures*, ICCS, LNAI Vol. 6828 Berlin: Springer-Verlag, pp. 201–214.

Poelmans, J., Elzinga, P., Neznanov, A., Dedene, G., Viaene, S., and Kuznetsov, S.O. (2012b), 'Human-Centered Text Mining: A New Software System,' *In*: *Proceedings of the 12th Industrial Conference on Data Mining*, ICDM, LNAI 7377, Berlin: Springer-Verlag, pp. 258–272.

Poelmans, J., Ignatov, I., Viaene, S., Dedene, G., and Kuznetsov, S.O. (2012b), 'Text Mining Scientific Papers: A Survey on FCA-Based Information Retrieval Research,' *In*: *Proceedings of the 12th Industrial Conference on Data Mining*, ICDM, LNAI 7377 Springer-Verlag, pp. 273–287.

Priss, U. (2004), 'Modeling Lexical Databases with Formal Concept Analysis,' *Journal of Universal Computer Science*, 10, 8, 967–984.

Priss, U. (2006), 'Formal Concept Analysis in Information Science,' in *Annual Review of Information Science and Technology* (Vol. 40), Medford, NJ: Information Today, Inc. pp. 521–543.

Savona, E.U. (1998), 'The Organizational Framework of European Crime in the Globalisation Process,' in *108th International Seminar on Current Problems in the Combat of Organized Crime*, 27 February, Tokyo, Japan.

Scharfe, H., Oehrstrom, P., and Gyori, M. (2009), 'A Conceptual Analysis of Difficult Situations – Developing Systems for Teenagers With ASD,' in *Supplement Proceedings Of the 17th International Conference On Conceptual Structures*, Moscow, Russia, CEUR Workshop Proceedings.

Scheich, P., Skorsky, M., Vogt, F., Wachter, C., and Wille, R. (1993), 'Conceptual Data Systems,' in *Information and Classification*, Heidelberg: Springer-Verlag, pp. 72–84.

Schnabel, M. (2002), 'Representing and Processing Medical Knowledge Using Formal Concept Analysis,' *Methods of Information in Medicine*, 41, 33–48.

Shelley, L. (1999), 'Human Trafficking: Defining the Problem,' *Organized Crime Watch-Russia*, 1, 2, 1–2.

Smyth, P., Pregibon, D., and Faloutsos, C. (2002), 'Data-Driven Evolution of Data Mining Algorithms,' *Communications of the ACM*, 45, 8, 33–37.

Stumme, G. (2002), 'Efficient Data Mining Based on Formal Concept Analysis,' in *Lecture Notes in Computer Science*, Vol. 2453 Berlin: Springer-Verlag, pp. 3–22.

Stumme, G., Wille, R., and Wille, U. (1998), 'Conceptual Knowledge Discovery in Databases Using Formal Concept Analysis Methods,' *In*: *Proceedings of the 2nd European Symposium on Principles of Data Mining and Knowledge Discovery*, LNAI 1510, Berlin: Springer-Verlag, pp. 450–458.

US Department of State (2008), 'Trafficking in Persons Report,' http://www.state.gov/g/tip/rls/tiprpt/2008, retrieved on 26 December 2010.

United Nations, Economic and Social Council (2004), 'Economic Causes of Trafficking in Women in the Unece Region,' in *Regional Preparatory Meeting for the 10-year Review of Implementation of the Beijing Platform for Action*, 14–15 December.

Valtchev, P., Missaoui, R., and Godin, R. (2004), 'Formal Concept Analysis for Knowledge Discovery and Data Mining: The New Challenges,' *In Proceedings of the International Conference on Formal Concept Analysis*, LNAI 2961 Springer-Verlag, pp. 352–371.

Wille, R. (1982), 'Restructuring Lattice Theory: An Approach Based on Hierarchies of Concepts,' In
I. Rival (Ed.), *Ordered Sets*, Dordrecht-Boston: Reidel, pp. 445–470.

Wille, R. (2002), 'Why Can Concept Lattices Support Knowledge Discovery in Databases?' *Journal
of Experimental and Theoretical Artificial Intelligence*, 14, 2, 81–92.

Wolff, K.E. (2005), 'States, Transitions and Life Tracks in Temporal Concept Analysis,' in *Formal
Concept Analysis*, LNAI 3626, Berlin: Springer-Verlag, pp. 127–148.

Appendix A: Human trafficking indicators

1. *Dependency on the exploiter*: Typically in human trafficking the housing, clothing, and
   transportation of the woman are arranged through the exploiter, the woman will often have debts
   towards the exploiter and will be forced to earn the money back:

   - The woman did not arrange the travel, visa, etc. herself
   - The woman has a fake or counterfeit passport
   - The woman resides/works illegally in the Netherlands
   - The woman fears for maltreatment and being set out of the country
   - The woman sleeps over at the workplace
   - The woman has no proper living address in the Netherlands
   - The woman does not know properly what her working address is
   - The woman is socially isolated by the exploiter
   - The woman is in debt with a third party such as the exploiter
   - The exploiter of the woman paid a take-over price.

2. *Deprivation of liberty*: Often the victim is not allowed to have contact with the outside world.
   They will typically not have their passports with them which are carried by the pimps. Also
   suspicious is when the victim cannot freely dispose of the money she earns.

   - The victim does not receive necessary medical treatment
   - The victim is not allowed to move around freely
   - The victim does not carry her own identity papers
   - The victim cannot freely dispose of her own money she earns
   - The victim has to give an unreasonably large sum of her income to someone else

3. *Being forced to work under bad circumstances*:

   - The victim receives an unusually low wage compared to the market
   - The victim works under dangerous circumstances
   - The victim works exceptionally long
   - The victim has to work under all circumstances and unreasonably long
   - The family of the victim is threatened and blackmailed
   - Indications of smuggling of single women
   - The combination of a non-European nationality, a marriage or stay with a partner and shortly
     after working in prostitution
   - Relationships with persons with relevant antecedents and locations associated with human
     trafficking
   - The woman is forced to earn a minimum amount of money each day
   - The woman has a slavish attitude towards exploiter
   - The woman lives and/or works in buildings with internal cameras, hiding places, fake
     decoration, bodyguards, etc.

4. *Violation of bodily integrity of the victim:*

   - Giving away organs
   - Involuntarily employed in prostitution
   - Threatened or confronted with violence
   - Carrying traces of bodily maltreatment
   - Certain things that may indicate the dependence of the exploiter such as tattoos or voodoo
     material
   - Being forced to perform sexual deeds

5. *Non-incidental pattern of abuse by suspect(s)*:

- Working at different places from time to time
- Tips of reliable third parties

Appendix B: Loverboy indicators

1. *Preparatory activities to recruit girls*: Actual recruitment and arranging residence and shelter locations for the girls. Sometimes a girl is both a prostitute and a recruiter of other girls. Sometimes loverboys recruit girls for each other. During the first meeting, they estimate how vulnerable a girl is to attention and flattery. Their sensitivity to attention, presents, etc. made her fall in love with the pimp. They are not critical anymore and do not wonder where the money comes from and what the pimps intentions are.

2. *Forcing her into prostitution*: Pimps use a number of techniques to force the girls into prostitution:

- Deception: They promise the girls they can keep the money or the money will be used for a vacation or a house
- Deflowering and forcible rape: In particular Islamic girls, deflowering and the threat of being brought back home increase their anxiety to say no to the pimp's demands, because it can result in her abandonment by her family
- Blackmailing: If the girls do not want to work in prostitution, the pimps threaten to bring her back to her parents
- Physical violence and threats: This is seen as the most effective technique to force the girl into prostitution

3. *Keeping the girl in prostitution*:

- Emotional dependence: Feelings of love, nobody else to support her, the pimp is the father of her child, etc
- Deception: in combination with the naivety and emotional dependence of girls
- Fear: the fear to be maltreated and the fear that her parents will be informed. In Islamic culture, virginity of the girl is a matter of family honor. If this girl is no longer virgin and the family finds it out, she might not be welcome anymore
- Social isolation: She becomes isolated from the outside world and only meets people from the prostitution circuit
- Pride: by hiding the fact that they have to give away all the prostitution money, by acting as if they have a better life than many others, the girls justify for themselves the abuse they suffer and apparently have something they can be proud of
- Police as an enemy: In particular under-aged girls start seeing the police as their enemy
- Competition and intermittent reinforcement: The pimp introduces competition between the girls and the girl who earns most will not be punished, but gets all attention and compliments from the pimp

4. *The pimp will also try to protect his organization*:

- Internal protection measurements: He will make sure that the girls are constantly under surveillance and with the threat of physical violence he completely dominates her life
- External protection: The pimp will threaten, bribe, interrogate, etc. the girls who have been in contact with the police. He may also force her to place a tattoo, to change her working address, etc. The tattoos are used in the prostitution world to trace girls who run away and are a powerful psychological instrument to make her consent to exploitation. When a girl runs away, the pimp may threaten to maltreat her or her family.