Analyzing chat conversations of pedophiles with temporal relational semantic systems

Paul Elzinga
Department of Business Information
Police Amsterdam-Amstelland
Amsterdam, The Netherlands
Paul.Elzinga@amsterdam.politie.nl

Karl Erich Wolff
Ernst Schröder Center
Darmstadt University of Technology
Darmstadt, Germany
Karl.erich.wolff@t-online.de

Jonas Poelmans
Faculty of Business and Economics
Katholieke Universiteit Leuven
Leuven, Belgium
Jonas.Poelmans@econ.kuleuven.be

Abstract— **Grooming is the process by which pedophiles try to find children on the internet for sex-related purposes. In chat conversations they may try to establish a connection and escalate the conversation towards a physical meeting. Till date no effective methods exist for quickly analyzing the contents, evolution over time, the present state and threat level of these chat conversations. In this paper we propose a novel method based on Temporal Relational Semantic Systems, the main structure in the temporal and relational version of Formal Concept Analysis. For rapidly gaining insight into the topics of chat conversations we combine a linguistic ontology for chat terms with conceptual scaling and represent the dynamics of chats by life tracks in nested line diagrams. To showcase the possibilities of our approach we used chat conversations of a private American organization which actively searches for pedophiles on the internet.**

Keywords; Formal Concept Analysis, Temporal Concept Analysis, Conceptual Scaling, Relational Systems, Nested Line Diagrams, Transition Diagrams

I. INTRODUCTION

Several years ago, one of the strategic goals formulated by the Amsterdam-Amstelland police was to make the shift from a reactive intuition-led management style to an Intelligence Led-Policing (ILP) approach. ILP is “a business model and managerial philosophy where data analysis and crime intelligence are pivotal to an objective decision-making framework that facilitates crime and problem reduction and prevention through both strategic management and effective enforcement strategies that target prolific and serious offenders” [16]. Criminal intelligence is the “product of gathering, evaluation and synthesis of raw data on individuals or activities suspected of being or known to be, criminal in nature. Intelligence is information that has been analyzed to determine its meaning and relevance” [9]. Since over 80% of information available to police organizations is in textual form being able to distill valuable knowledge from these data is a major step towards becoming an ILP organization. Although text mining was identified as a promising area in the formal framework for crime data mining by [2], only few successful applications have been reported in the literature.

A. *Grooming*

In this paper we focus on the analysis of chat conversations of pedophiles who try to find children online for sex related purposes [21]. This process is also called grooming and consists of several phases [7]. In a first step the online sexual predator chooses a virtual location which is likely to attract children such as a chat box, MSN, Bebo etc. and becomes particularly interested in a child. The offender may then try to form a bond with the child and sexually escalate the conversation towards a physical meeting. Gottschalk (2011) [7] identified three types of groomers according to the level of danger they pose to society: the distorted attachment offender, the adaptable online groomer and the hyper-sexualized type. The distorted attachment offender has offence supportive believes that contact with minors can be seen as a relationship. The adaptable online groomer has offence supportive believes and acts to fulfill his own needs and sees minors as mature and capable. The hyper-sexualized groomer mostly has large collections of child pornography and significant online contact with other pedophiles.

B. *Analysis of chat conversations*

1) Goals

The goal is to develop a system which crawls the internet, can recognize suspicious chat conversations as early as possible and indicate the threat level to a child’s safety. For that purpose the following sub goals have to be reached.

- The first sub goal is to develop suitable structures for representing chat conversations such that the main content of a chat conversation can be quickly analyzed by a human, for example a police officer. For that purpose the evolution of the chat conversation over time has to be represented such that the states of the chat conversation are clearly recognizable.

- The second sub goal is to develop a suitable scale of threat levels for chat conversations such that each chat conversation can be mapped automatically to its threat level.
- The third sub goal is the construction of a practically applicable procedure which classifies huge amounts of chat conversations with respect to their threat levels.

2) Previous research

The analysis of chat conversations as described in this paper builds further on previous research. In the first case study, with the Amsterdam-Amstelland police which started in 2007, we used FCA to analyze statements made by victims to the police. We iteratively enriched and refined the concept of domestic violence, resulting in an improved definition and highly accurate automated labeling of new incoming cases [11,12,13]. Later on we made a shift to the observational reports from which we extracted persons involved in human trafficking and terrorism. One of our concept lattices allowed for the detection of several suspects involved in human trafficking or showing radicalizing behavior. Some of them were mentioned in multiple reports and a detailed profile of one suspect depicted as a lattice, with timestamps of the observations as objects and indications as attributes helped to gain insight into their threat to society [5,14]. The goal in each of these papers was to make an overload of information available in an intuitive visual format that may speed up and improve decision making by police investigators on where and when to act. Nevertheless, there are several differences with analyzing traditional police reports, i.e. observational reports, reports containing a statement made by a victim, etc. Whereas visual representation of the topics described in such an individual report was not necessary (since they were rather short) this becomes a mandatory requirement for (sometimes very long, e.g. hundreds of pages) chat conversations. They do not contain a summary of facts; rather several topics emerge between two or more persons. We should judge the severity and crime committed and distinguish between several types of suspects, e.g. is this person someone who only fantasizes about abusing children, did he actually exchange child pornography, did he abuse any children, etc.

3) Application of Temporal Concept Analysis

Besides standard FCA, which we used for representing a huge collection of chat conversations (objects) and their topics (attributes), we now employ the recently developed Temporal Relational Semantic Systems (TRSSs) [28,29] for the representation of temporal relational aspects of chat conversations. As opposed to the relational representation by power context families the TRSSs use many-valued contexts and conceptual scales and combine it with the temporal and relational structures in a simple way. The conceptual scales are a powerful instrument to look at the data at different levels of granularity. The keywords used to index the chats are grouped based on their semantic meaning according to the threat level they pose to the child's safety. These groups are structured with respect to their semantic meaning in a conceptual scale which is employed to form meaningful (nested) line diagrams used to represent chat conversations in an intuitive and readable visible format such that suspicious conversations can easily be identified. To show the dynamical behavior of a chat conversation the temporal tools for TRSSs were used to visualize a chat conversation by its life track in a nested line diagram showing the state space of the chat conversation. These animated line diagrams help to reveal the behavior of the groomer.

The remainder of this paper is composed as follows. In section II we introduce the chat data used in this paper. In section III we describe only the main ideas of the theory of Temporal Relational Semantic Systems. In section IV we show the results of our approach. Finally section V concludes the paper and presents promising avenues for further research.

II. CHAT DATA

The purpose of this project is to support the Dutch police force organizations in their investigations of criminal subjects, by semi-automatically evaluating chats of pedophiles with children. Because the original chat data collected by the Dutch police force organizations are restricted by law, they may not be used in this paper.

A. Perverted Justice Chat Data

To demonstrate our methods using TRSSs we chose for chat data collected by a private American organization, Perverted Justice, which actively searches for pedophiles on the internet. We downloaded 544 chat files with 544 different suspects.

In these chats a pedophile is chatting with a victim, but the pedophile does not know that the victim is a member of the Perverted Justice organization playing the role of a child of preteen age or younger. If the suspect tries to make contact and wants to make an appointment, he will be directed to a bust house where the police will wait for him. Figure 1 shows an excerpt of an original chat file. The second line shows the nickname "Blue Bonnet" of the Perverted Justice member who busted the suspect. He is acting in the chat under the name "rubyslipper013" while the suspect uses the name "mesadash8pilot". This chat conversation starts with telling each other age and sex. The Perverted Justice member with the nickname "Blue Bonnet" has written the comments in the chat file. It took Blue Bonnet four months to make an appointment and finally the suspect was trapped in the bust house, as the last comment within brackets states.

```
 He had only one thing on his scummy mind... raping a child.
 Bust by Blue Bonnet @ 4/8/2011 2:06 PM PST
 [ Conviction ] Perverted Justice mark: Scott Andrew Erb , 23
Yahoo IM: mesadash8pilot [mesadash8pilot@yahoo.com]
Location: Las Vegas, Nevada
[send mesadash8pilot a message ]
This wannabe pedo tried to solicit Rubyslippers013, a 13 year old girl ... or so they thought!

 mesadash8pilot (08/17/08 12:28:23 AM): hey
 mesadash8pilot (08/17/08 12:28:25 AM): asl
rubyslippers013 (08/17/08 12:32:44 AM): hi
 mesadash8pilot (08/17/08 12:32:51 AM): whats up?
rubyslippers013 (08/17/08 12:32:53 AM): im ruby
rubyslippers013 (08/17/08 12:32:56 AM): 13 f ut
rubyslippers013 (08/17/08 12:33:00 AM): whats ur asl?
 mesadash8pilot (08/17/08 12:33:10 AM): 21/m/nv, I'm scott
...
...
...
 rubyslippers013 (12/28/08 3:20:20 PM): ok
rubyslippers013 (12/28/08 3:20:25 PM): r u leavin now?
 mesadash8pilot (12/28/08 3:20:29 PM): yeah
rubyslippers013 (12/28/08 3:20:32 PM): yay
 mesadash8pilot (12/28/08 3:20:43 PM): just remember what we talked about
rubyslippers013 (12/28/08 3:20:48 PM): i do
 mesadash8pilot (12/28/08 3:20:58 PM): ok bye bye call me in 20 minutes
rubyslippers013 (12/28/08 3:21:06 PM): k! luvs u!
 mesadash8pilot (12/28/08 3:21:13 PM): love you too
 mesadash8pilot (12/28/08 3:21:14 PM): bye bye
 ( Can we say that I'm surprised he managed to show up at the bust house???
   I was giving him exact directions but he didn't listen long enough to understand them.
   He is such an idiot. )
```

Figure 1. An excerpt of an original chat file

B. *Preparing the Perverted Justice Chat Data*

For the purpose of developing suitable methods for the selection of relevant chat files from a huge amount of chat files we first introduced a classification of the terms occuring in the chat files. The main idea is to classify the terms according to the meaning of the following seven states of the grooming process as described in Table 1. The seven states are developed by using the CORDIET toolset [14], which is based on the C/K design methodology as described in [12].

TABLE I. THE STATES OF THE GROOMING PROCESS

1 Sweet greetings	By calling each other with sweet names, like “hey sweety” or “hey love”, an intimate sphere is created.
2 Compliments	Many young girls and boys are very uncertain about their looks, getting compliments about their looks is another step in getting an intimate sphere.
3 Intimate parts	In this state, the suspects tries to get the victim to the intimate parts or calling intimate parts by their popular names.
4 exual handlings	The chat is becoming more explicit. The suspect is fantasizing how he will have sex with the victim.
5 Cam and pictures	The cam and pictures makes the contact more intimate and gives the suspect the opportunity to show his intimate parts and even his sexual handlings.
6 Where	If the suspect aims at meeting the victim, he'll try to get the victim's address.
7 When	Here the suspect actually tries to find a time for a meeting with the victim.

To use these seven states for the selection of interesting chat files we indexed the huge set of all terms occuring in the chat files with these seven states. For example, the terms “hey sweety”, “hey love” and many others are indexed by the state “1 Sweet greetings”. The resulting classification $\mathfrak{C} = \{C_1,\ldots, C_7\}$ of all occuring chat terms in the 544 downloaded chat files will be used in the following extensively. We found 215 chat files meeting all classes (states). The selection of these 215 chat files was done using the CORDIET system [14].

III. TEMPORAL RELATIONAL SEMANTIC SYSTEMS

The chat data evaluated in this paper contain the three main aspects of Temporal Relational Semantic Systems (TRSSs) as introduced in [28], namely the temporal, the relational and the semantic aspect, the last one referring to the meaning preserving power of conceptual scaling. In this section we only give a short overview over the main ideas in Temporal Concept Analysis and its central structure, the Temporal Relational Semantic Systems. For the theory of Temporal Relational Semantic Systems the reader is referred to [28].

A. *Relational Aspects: Conceptual Graphs and Concept Graphs*

The TRSSs are based on two research directions, namely first the theory of Conceptual Graphs as introduced by John Sowa [17,18] and further developed by Chein and Mugnier [1]. Rudolf Wille replaced the undefined notion of a ‘concept’ in the theory of Conceptual Graphs by the mathematically defined notion of a ‘formal concept’ in the sense of Formal Concept Analysis (FCA) [19,6]. He described a set of relations of arbitrary arities by the notion of a ’power context family’ for which he introduced the notion of a ‘concept graph’. For the theory of concept graphs the reader is referred to [20,15,3].

B. *Temporal Concept Analysis*

The second research direction on which TRSSs are based is Temporal Concept Analysis (TCA) introduced in [22]. TCA describes temporal systems in terms of FCA such that the classical notions of discrete and continuous systems are generalized. For example, the spacio-temporal systems in physics as well as the discrete systems in Automata Theory can be described in TCA. For an introduction the reader is referred to [25] where the notions of a ‘state’ of a ‘temporal object’ at a certain ‘time granule’ (for example: a day, or a minute) is defined with respect to a chosen ‘view’ (describing for example: a map). If the state of a temporal object has at each time granule at most one concept, then the set of all its states forms the ‘life track’ of the temporal object. For a precise definition the reader is referred to [25,28,29]. Line diagrams with arrows representing transitions of temporal objects are called ‘transition diagrams’ (see Figure 3, 4 and 5).

C. *Conceptual Semantic Systems*

The main motivation for the introduction of the notion of a Conceptual Semantic System (CSS) by Wolff was the aim to find a common conceptual understanding of the notions of ‘particles’ and ‘waves’ in physics. This aim was successfully reached as shown in [23,24]. For that purpose the idea of representing ‘objects in reality’ by formal objects of a formal context had to be generalized. The main idea was to represent the usually many kinds of ‘objects in reality’ as values of a many-valued context and interpret the formal objects of a many-valued context as ‘observation units’. The values of the many-valued context are then represented as formal concepts of given ‘semantic scales’. Conceptual Semantic Systems offer the new possibility to represent ‘distributed objects’, as for example waves. Using Conceptual Semantic Systems it is no longer necessary in applications to search for some kind of ‘objects in reality’ which can be used as formal objects of the intended many-valued context.

D. Temporal Relational Semantic Systems

The main idea in the construction of Relational Semantic Systems [26,27] is to represent relational statements explicitly in the rows of a data table. That is done mainly by introducing a specific many-valued attribute whose values denote ‘relational expressions’ of arbitrary arity. In the example of the chat data we will use for example the binary relational expression ‘person x asks person y for his address’. In Temporal Relational Semantic Systems the relational expressions can be time-dependent. For example, ‘at chat time t person x asks person y for his address’. In Temporal Relational Semantic Systems all tools for Temporal Conceptual Semantic Systems can be used. For the mathematical definitions the reader is referred to [28] and for applications to [29,30].

IV. Conceptual Evaluation of Chat Data

A. FCA programs: Workflow

For the conceptual evaluation of chat data we first generate a data table, usually an Excel-file which should be in the format CSV (MS-DOS) to be importable into the program CERNATO. There we generate suitable conceptual scales in the form of views. The resulting data file is exported as an XML-file which is imported in the program SIENA (a program in TOSCANAJ). In SIENA we generate line diagrams of the views and also nested line diagrams of two views. Using the TEMPORAL CONCEPT ANALYSIS menu in SIENA we can decide for which temporal object (for example: a person or a chat conversations) we would like to generate its life track with respect to a chosen time attribute.

For each of these (nested) line diagrams we can draw in SIENA a ‘transition diagram’ which shows in the given (nested) line diagram life tracks of temporal objects. These life tracks can be visualized in an animation where the transition arrows are shown one at a time or even several successive arrows at a time.

B. A data table for the chat data

In this paper we concentrate on 13 selected suspects s1,...,s13 chatting with their victims v1,..., v13, respectively. For each of the suspects the chats between the suspect and his victim are collected in a chat log-file (with labels between 104 and 513) together with the time (in the form: year, month, day, hour, minute) when each single chat had been sent. In Table 2 we show some rows of one of our data tables for the chat data which represents in each row (labeled from 0 to 518) some information about a single chat. For example, in the rows from 0 to 30 we have all chats of chat log 104, in row 31 chat log 194 starts. There the chat time for chat log 194 starts again with 0. Hence the chat time just labels the single chats in a chat log, starting with 0. In Table 2 we do not show the information about the “real” time (in the form: year, month, day, hour, minute). The three columns following ‘chat time’ show for each chat the names of the person making the chat, receiving the chat and the chat text. For example: in row 26 suspect s5 makes the chat, which is received from his victim v5; the chat text is “what is your zip”. In the column ‘state’ we classify each chat text with respect to the seven classes described in Table 1. That will be discussed in detail later in this paper. In the last column ‘relation’ we use a ‘relational’ classification of chat texts which will be discussed in the following.

TABLE II. A PART OF THE DATA TABLE FOR THE CHAT CONVERSATIONS

row	chatlog	chat time	makes chat	receives chat	chat text	state	relation
0	104	0	s5	v5	hey	1	/
...	...	...	...	...	...	...	...
26	104	26	s5	v5	what is your zip	6	asks for address
...	...	...	...	...	...	...	...
30	104	30	s5	v5	ok how many times can u cum	4	describes about sex
31	194	0	v4	s4	nice to meet you	1	/
...	...	...	...	...	...	...	...
512	513	63	v1	s1	we can have sex	4	asks about sex
513	513	64	v1	s1	when you coming	7	asks
...	...	...	...	...	...	...	...
518	513	69	s1	v1	and i not going stop tell i cum ok	4	describes about sex

C. Content analysis for chat texts

The problem of classifying chat texts belongs to the large linguistic field of Content Analysis [10]. Chat texts have the additional problem that special slangs with many abbreviations are used. We employ the well-known technique of indexing a huge amount of terms by keywords - which corresponds to a mapping and its partition of the inverse images of the keywords. It is obvious that automated keyword extraction often fails in finding the ‘correct’ keyword which an expert would assign.

D. Scaling the states of the chats

In our investigation of the chat data we made several approaches to extend a given nominal scale which was obtained from an indexing of a huge amount of terms by relatively few keywords. As an example we take again the indexing of terms in chats with the seven keywords ‘1 Sweet greetings, 2 Compliments,..., 7 When’ as mentioned in Table 1.

The corresponding partition of the keywords can be described as a nominal scale. In this paper we will show examples where we embed such a nominal scale into a richer scale by introducing further attributes which lead to a better structured concept lattice with respect to our interpretation of the classes of the given partition.

The concept lattice of a scale for the seven states is shown in Figure 2 where the central chain from ‘>=Sweet greetings’ to ’>=Sexual handlings’ shows a hierarchy of increasing sexual meaning. The seven classes are represented by the seven object concepts in this scale. By the way, the attribute ‘Dating’ was introduced to represent a common super-concept for ‘When’ and ‘Where’.

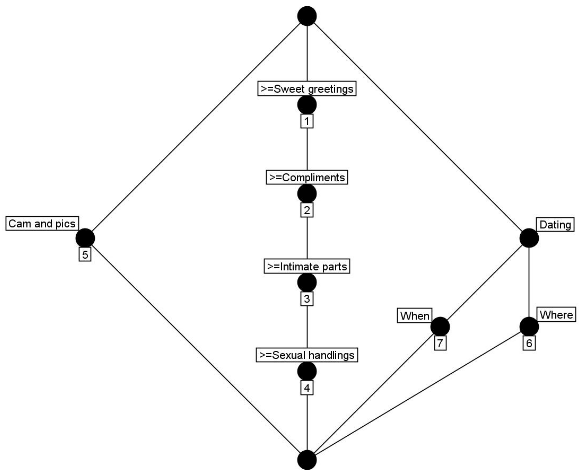

Figure 2. A diagram of a scale of the 7 states

E. Transition diagrams of chat conversations

Figure 3 shows a transition diagram of the chat conversation 451 based on the line diagram of the scale shown in Figure 2. We explain Figure 3 first intuitively, and then we connect it with the mathematical definitions. Figure 3 is constructed by restricting the data table shown in Table 2 to the rows where chat log = 451, which are the 22 rows from 300 to 321. The chat time runs in these rows from 0 to 21. The many-valued attribute 'state' has in row 300, that is at time 0, the value '2' which means that the conversation 451 is in the state '2 Compliments'; in the next row 301, at time 1, the conversation 451 is in the state '5 Cam and pics'. This transition is graphically represented in Figure 3 by the arrow from the object concept of 300 to the object concept of 301. Clearly, the direction of the arrow is induced from the fact that time 0 is the predecessor of time 1 (in the natural ordering of integers). Therefore, in the formal definition of a 'transition of a temporal object $\mathfrak{o}$ we use a binary relation $\mathfrak{R}_{\mathfrak{o}}$ whose elements are called 'base transitions'.

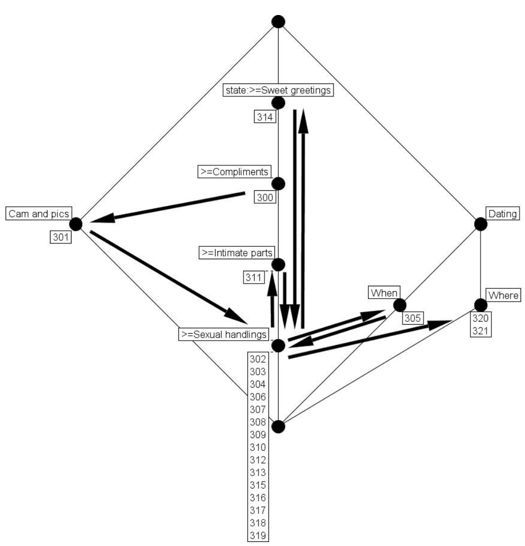

Figure 3. A transition diagram for chat conversation 451

In Figure 3 the reader can now follow all transitions of chat conversation 451 by searching for the next row label. For example, the row labels 302, 303, 304 belong to the state '6 Sexual handlings' which is the most frequent state of chat conversation 451 while the row label 305 belongs to the state 'When'. The 'life track' of 451 ends in the state '6 Where' which is the object concept of the rows 320 and 321 in the chosen part of the 'semantically derived context' of the given data.

These observations show the following main ideas in Temporal Concept Analysis:

- a state is defined as a set of object concepts (in our examples each state consists of a single object concept)
- a state, which will be denoted as '$\gamma_Q(\sigma(\mathfrak{o},\mathfrak{t}))$', is defined with respect to some given 'temporal object $\mathfrak{o}$', some 'time granule $\mathfrak{t}$', some 'view *Q*' and some chosen 'selection σ'
- the definition of a 'transition of a temporal object $\mathfrak{o}$ should make explicit not only two states, but also the corresponding 'time granules' when the temporal object $\mathfrak{o}$' is in these states.

By the way, the 7 'states' of the grooming process as introduced in Table 1 are in Figure 3 indeed represented as the states of the temporal object 'chat conversation 451' which are object concepts in the chosen view and the chosen selection.

It is obvious that transition diagrams are very useful for a quick understanding of processes, for example of chat conversations.

F. Animation of transitions in (nested) line diagrams

In Figure 3 we can not recognize which one of the two persons in the chat conversation just makes the chat. To show this we draw a nested line diagram where the inner diagram distinguishes whether the suspect or the victim makes the chat. To visualize in Figure 4 several effects we change from chat conversation 451 to 513 where subject s1 and victim v1 are chatting. We also change to the standard situation in the conceptual evaluation of chat conversations where the investigated data table has all rows (from 0 to 518). Then one can switch easily from one chat conversation to another.

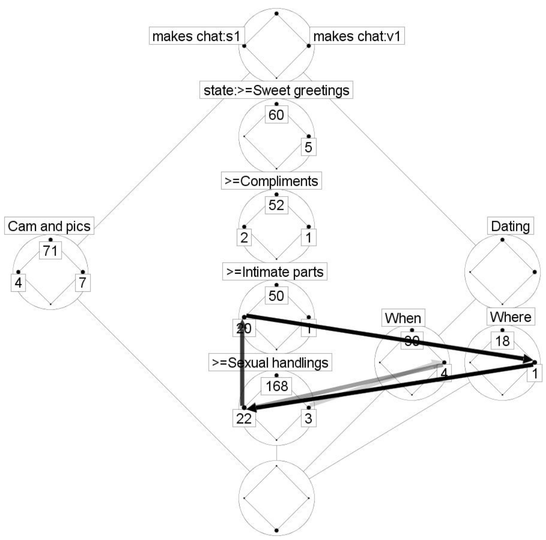

Figure 4. A snapshot of an animated diagram showing a trap for a pedophile

In Figure 4 we show a nested line diagram where the coarse structure represents the scale for the many-valued attribute 'state' shown in Figure 2, while the fine structure represents a scale for the many-valued attribute 'makes chat'. Its two scale attributes 's1' and 'v1' yield the two attributes 'makes chat: s1' and 'makes chat: v1' in the derived context. There the attribute 'makes chat: s1' selects by the usual database selection all rows *g* such that s1 occurs in the cell determined by row *g* and column 'makes chat'. The object label at an object concept in Figure 4 shows the number of formal objects having this object concept. Hence Figure 4 shows that subject s1 during chat conversation 513 never mentioned something about a dating while victim v1 chatted 4 times something about 'When' and once about 'Where'. That all happened during the last 6 chats of chat conversation 513; the corresponding 5 transitions are shown in Figure 4. The arrows of these last 5 transitions are generated graphically using a snapshot from the animation of the transitions of this chat conversation. We have chosen the parameters of the animation in the Temporal Concept Analysis menu in SIENA in such a way that only about five arrows are visible, the previous ones are slowly vanishing. For example, the arrow from the concept labeled with 3 to the concept labeled with 4 is just vanishing and therefore the earliest among the five shown arrows. Clearly, for any top concept in the fine structure its object label is the number of rows where someone different from s1 and v1 makes the chat. For example, there are exactly 18 rows among all 519 rows where neither s1 nor v1 is chatting about something classified by 'Where'.

1) A trap for a pedophile

Figure 4 tells us the following interesting story of a trap for suspect s1: the victim v1 who is a Perverted Justice member playing the role of a child chats something about sexual handlings (with the original chat text: 'we can have sex' as shown in Table 2, row 512), then he writes something about the time of a date (with the original chat text: 'when you coming'), then the suspect s1 chats something about sexual handlings and intimate parts, followed by a chat of victim v1 about location ('Where') and finally the suspects chats something about sexual handlings again, but s1 never chatted something about dating during this chat conversation. That looks really like a trap for s1.

G. *Using the relational part in Temporal Relational Semantic Systems*

The special feature of a Relational Semantic System [28] is the explicit description of a set of relations by listing all their tuples in the rows of a single data table combined with the possibility to scale the values in these tuples. These values as well as the relation names (or for example relational formulas like 'x asks y' or '*-expressions' like '*asks*') are introduced as formal concepts of given formal contexts, called 'semantic scales'. That allows for treating the relations in the same way as the values of their tuples by conceptual scaling. As an example, Figure 5 shows a nested line diagram where the coarse diagram is a scale for the relational attribute of a Temporal Relational Semantic System which was constructed from Table 2 by attaching a further many-valued attribute whose values are those '*-expressions' which are shown as attribute labels in the coarse diagram of Figure 5. The fine diagram has the same meaning as in Figure 4. As an example we express the meaning of the transition of chat conversation 513 from chat time 26 to 27 as show by the arrow in Figure 5: At chat time 26 the victim v1 makes the chat and describes consequences of sex with subject s1; then, at chat time 27 the subject s1 asks about sexual behavior of victim v1.

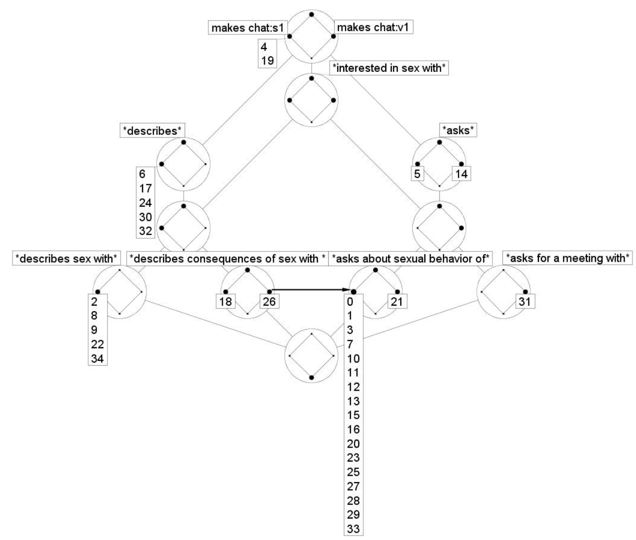

Figure 5. Relations as attributes: the transition from chat time 26 to 27 for chat conversation 513

This example shows that using TRSSs quite complicated temporal relational statements can be treated and visualized in a clear manner.

It seems to be very promising to automate this 'hand made' indexing of chat texts of chat conversation 513. A first result of such an indexing is shown in the column for the many-valued attribute 'relation', the last column in Table 2. We are working to improve these indexing methods.

V. CONCLUSION AND FUTURE WORK

Since children who actively use the internet can be exposed to unknown individuals and inappropriate sexual materials, they require safeguarding from sexual predators [4]. In this paper we proposed a novel method based on Temporal Concept Analysis using Temporal Relational Semantic Systems, conceptual scaling and nested line diagrams to analyze chat conversations. This method yields a quick understanding of chat conversations and represents the states of the chat conversation clearly. Therefore, we have attained the first sub goal described in section 1.2. It is not yet clear, how this method can be used to attain the second and the third mentioned sub goals. The most difficult part is to index chat texts in a meaningful way to suitable concepts.

In the past a couple of technical reports were published which proposed methods to identify pedophile keywords in search engine queries, to classify chats as dangerous or not, etc. These approaches typically generated many false positives and filtering these out was a very time intensive task. The unique novelty of our approach compared to existing methods is the combination of techniques which allow visualizing the whole collection of chats, identifying the state of the relationship between offender and victim and the evolution over time of the conversation in an easy to interpret manner. Our method was empirically validated on a real life dataset from the website www.perverted-justice.com.

Potential avenues for further research include:

- Developing an automated keyword extraction technique that may help reveal new emerging terms and phrases used by pedophiles who are chatting with young children to maintain acceptable recall and precision over time.
- Relational Concept Analysis [8] may become useful in case there are several interrelated offenders. Not only individual conversations need to be analyzed in this case but also the interpersonal links between different offenders.
- The result of these analyses will be implemented in the CORDIET toolset which is being further developed by researchers at the Amsterdam-Amstelland police [14].

ACKNOWLEDGMENT

Jonas Poelmans is Aspirant of the "Fonds voor Wetenschappelijk Onderzoek – Vlaanderen" (FWO) or Research Foundation - Flanders.

REFERENCES

[1] Chein, M., Mugnier, M.-L.: Graph-based Knowledge Representation. Computational Foundations of Conceptual Graphs. Springer-Verlag London Limited (2009)

[2] Chen, H., Chung, W., Xu, J.J., Wang, G., Qin, Y., Chau, M.: Crime data mining: a general framework and some examples. IEEE Computer, April (2004)

[3] Dau, F.: The Logic System of Concept Graphs with Negation And Its Relationship to Predicate Logic. LNAI 2892, Springer, Heidelberg (2003).

[4] Dombrowski, S.C., Gischlar, K.L., Durst, T.: Safeguarding young people from cyber pornography and cyber sexual predation: a major dilemma of the internet. Child abuse review, Vol. 16, pp. 153--170 (2007).

[5] Elzinga, P., Poelmans, J., Viaene, S., Dedene, G., Morsing, S. : Terrorist threat assessment with Formal Concept Analysis. Proc. IEEE International Conference on Intelligence and Security Informatics. May 23-26, 2010 Vancouver, Canada. ISBN 978-1-42446460-9/10, pp.77--82. (2010)

[6] Mathematical Foundations. Springer, Heidelberg (1999); German version: Springer, Heidelberg (1996)

[7] Gottschalk, P.: A dark side of computing and information sciences: characteristics of online groomers. The Journal of Emerging Trends in Computing and Information Sciences, Vol. 2, No. 9, pp. 447--455, September (2011).

[8] Huchard, M., Rouane-Hacene, M., Cyril Roume, Valtchev, P.: Relational concept discovery in structured datasets. Ann. Math. Artif. Intell. 49(1--4), pp. 39--76.(2007)

[9] IALEIA: Law Enforcement Analytic Standards. Richmond, VA: Global Justice Information Sharing Initiative.(2004)

[10] Krippendorf, K.: The Content analysis Reader. With M. A. Bock (Eds.). Thousand Oaks, CA: Sage, 481 pp.(2008)

[11] Poelmans, J., Elzinga, P., Viaene, S., Dedene, G. . A case of using formal concept analysis in combination with emergent self organizing maps for detecting domestic violence, Lecture Notes in Computer Science, 5633, pp. 247 –- 260, Advances in Data Mining. Applications and Theoretical Aspects, 9th Industrial Conference (ICDM), Leipzig, Germany, July 20--22, 2009, Springer (2009)

[12] Poelmans, J., Elzinga, P., Viaene, S., Dedene, G. : Curbing domestic violence: Instantiating C-K theory with Formal Concept Analysis and Emergent Self Organizing Maps. Intelligent Systems in Accounting, Finance and Management 17, (3--4), pp. 167--191. Wiley and Sons, Ltd. doi 10.1002/isaf.319 (2010)

[13] Poelmans, J., Elzinga, P., Viaene, S., Dedene, G., Kuznetsov, S. : A concept discovery approach for fighting human trafficking and forced prostitution. Lecture Notes in Computer Science 6828, pp. 201--214, 19th International conference on conceptual structures, July 25--29, Derby, England. Springer (2011)

[14] Poelmans, J., Elzinga, P., Neznanov, A., Dedene, G., Viaene, S., Kuznetsov, S.: Human-Centered Text Mining: a New Software System. P. Perner (Ed.): Lecture Notes in Artificial Intelligence 7377, 258–272, 12th Industrial Conference on Data Mining. Springer (2012)

[15] Prediger, S.: Kontextuelle Urteilslogik mit Begriffsgraphen. Ein Beitrag zur Restrukturierung der mathematischen Logik. Dissertation, TU Darmstadt 1998. Shaker, Aachen (1998)

[16] Ratcliffe, J.: Intelligence-Led Policing. Collumpton, UK Willan Publishing (2008)

[17] Sowa, J.F.: Conceptual structures: information processing in mind and machine. Adison-Wesley, Reading (1984)

[18] Sowa, J.F.: Knowledge representation: logical, philosophical, and computational foundations. Brooks Cole Publ. Comp., Pacific Grove, CA (2000)

[19] Wille, R.: Restructuring Lattice Theory: an Approach based on Hierarchies of Concepts. In: Rival, I. (ed.): Ordered Sets. pp. 445--470, Reidel, Dordrecht-Boston (1982). Reprinted in: Ferré, S., Rudolph, S. (eds.): Formal Concept Analysis. ICFCA 2009. LNAI 5548, pp. 314--339. Springer, Heidelberg (2009) M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[20] Wille, R.: Conceptual Graphs and Formal Concept Analysis. In: D. Lukose, H. Delugach, M. Keeler, L. Searle, J.F. Sowa (eds.): Conceptual Structures: Fulfilling Peirce's Dream. LNAI 1257, pp. 290--303. Springer, Heidelberg (1997)

[21] Wolak, J., Finkelhor, D., Mitchell, K.J., Ybarra, M.L.: Online predators and their victims - myths, realities and implications for prevention and treatment. American Psychologist Vol. 63, No. 2, pp. 111--128 (2008).

[22] Wolff, K.E.: Temporal Concept Analysis. In: E. Mephu Nguifo et al.(eds.): ICCS-2001 International Workshop on Concept Lattices-Based Theory, Methods and Tools for Knowledge Discovery in Databases, Stanford University, Palo Alto, CA, 91--107 (2001)

[23] Wolff, K.E.: 'Particles' and 'Waves' as Understood by Temporal Concept Analysis. In: K.E. Wolff, H.D. Pfeiffer, H.S. Delugach (eds.): Conceptual Structures at Work. LNAI 3127, pp. 126--141. Springer, Heidelberg (2004)

[24] Wolff, K.E.: States of Distributed Objects in Conceptual Semantic Systems. In: F. Dau, M.L.Mugnier, G.Stumme (eds.): Common Semantics for Sharing Knowledge. LNAI 3596, pp. 250--266. Springer, Heidelberg (2005)

[25] Wolff, K.E.: States, Transitions, and Life Tracks in Temporal Concept Analysis. In: B. Ganter, G. Stumme, R. Wille (eds.): Formal Concept Analysis – State of the Art. LNAI 3626, pp. 127--148. Springer, Heidelberg (2005)

[26] Wolff, K.E.: Relational Semantic Systems, Power Context Families, and Concept Graphs. In: Wolff, K.E. et al (eds.): Contributions to ICFCA 2009, pp. 63--78. Verlag Allgemeine Wissenschaft, Darmstadt (2009)

[27] Wolff, K.E.: Relational Scaling in Relational Semantic Systems. In: Rudolph, S. et al (eds.): Conceptual Structures: Leveraging Semantic Technologies. LNAI 5662, pp. 307--320. Springer-Verlag, Heidelberg (2009)

[28] Wolff, K.E.: Temporal Relational Semantic Systems. In: Croitoru et al (eds.): Conceptual Structures: From Information to Intelligence. LNAI 6208, pp. 165--180. Springer-Verlag, Heidelberg (2010)

[29] Wolff, K.E.: Applications of Temporal Conceptual Semantic Systems. In: Wolff, K.E. et al (eds.): Knowledge Processing and Data Analysis. LNAI 6581, pp. 59--78. Springer-Verlag, Heidelberg (2011)

[30] Wollbold, J., Wolff, K.E., Huber, R., Kinne, R.: Conceptual Representation of Gene Expression Processes. In: Wolff et al (eds.): Knowledge Processing and Data Analysis. LNAI 6581, pp. 79--100. Springer, Heidelberg (2011)