



Ô[} &^] oÁ^|æā } Åã &[ç^i^ Áæ åÁ } [çæā } Á
^} æā|ā * Á&@ [[* ^ ÁÔUÜÖØVD

R } æ Á[^|{ æ • ÉÚæ |À: ā * æÛcā) Áæ } ^Áæ åÁ Õ~ ā[Ä^å^ } ^



DEPARTMENT OF DECISION SCIENCES AND INFORMATION MANAGEMENT (KBI)

Concept Relation Discovery and Innovation Enabling Technology

(CORDIET)

Jonas Poelmans¹, Paul Elzinga³, Stijn Viaene^{1,2}, Guido Dedene^{1,4}

¹K.U.Leuven, Faculty of Business and Economics, Naamsestraat 69,
3000 Leuven, Belgium

²Vlerick Leuven Gent Management School, Vlamingenstraat 83,
3000 Leuven, Belgium

³Amsterdam-Amstelland Police, James Wattstraat 84,
1000 CG Amsterdam, The Netherlands

⁴Universiteit van Amsterdam Business School, Roetersstraat 11
1018 WB Amsterdam, The Netherlands

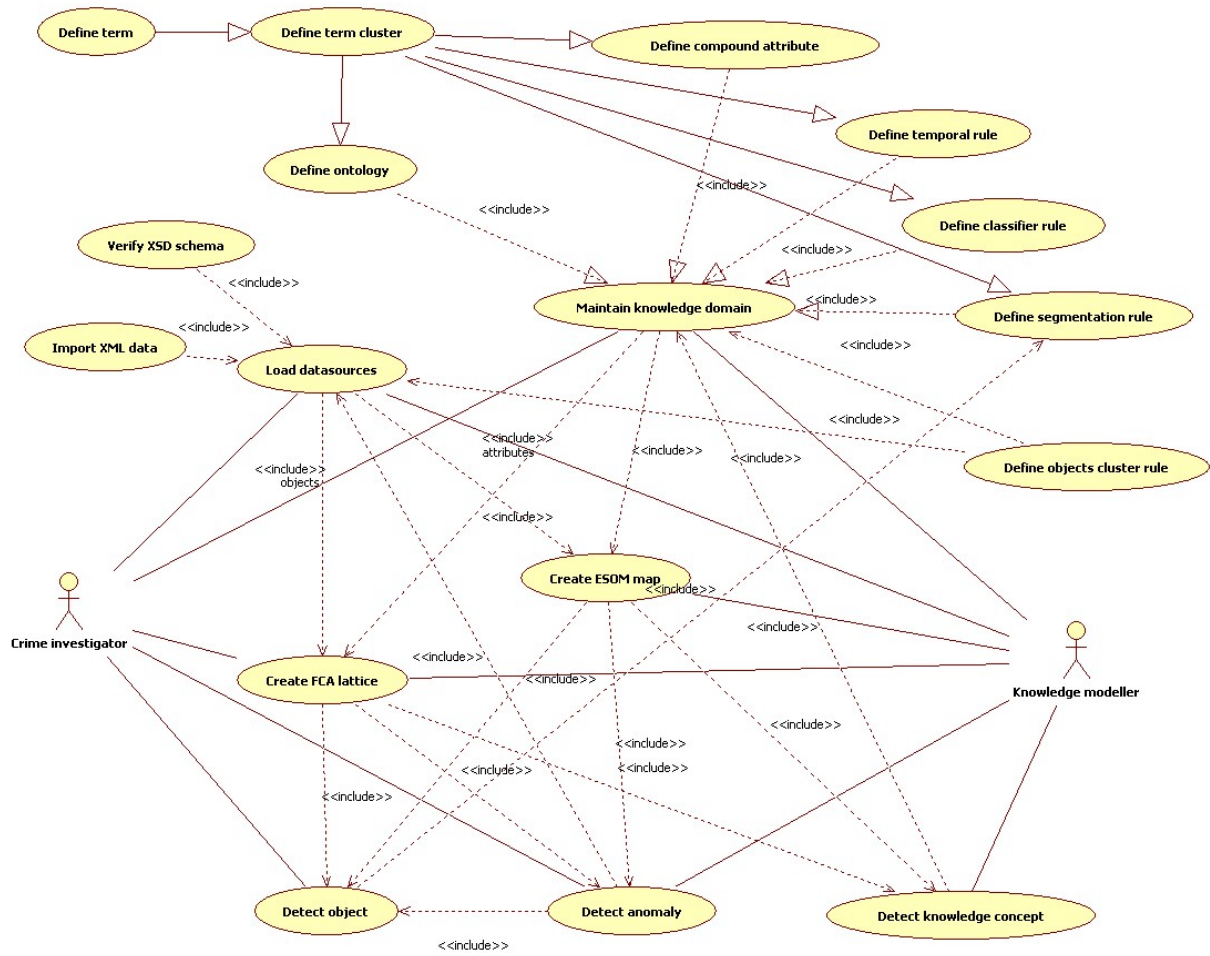
{Jonas.Poelmans, Stijn.Viaene, Guido.Dedene}@econ.kuleuven.be
Paul.Elzinga@amsterdam.politie.nl

1. What is Cordiet?

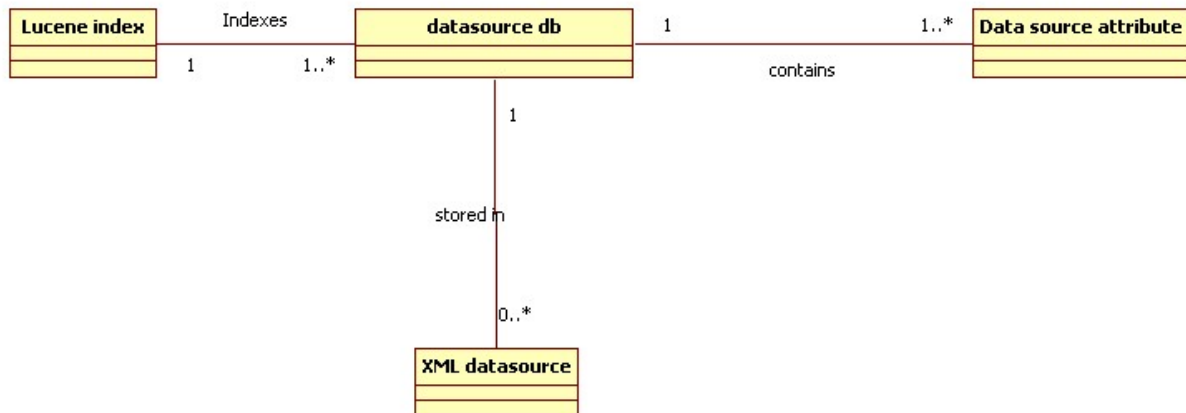
Concept Relation Discovery and Innovation Enabling Technology (CORDIET), is a toolbox for gaining new knowledge from unstructured text data. At the core of CORDIET is the C-K theory which captures the essential elements of innovation. The tool uses Formal Concept Analysis (FCA), Emergent Self Organizing Maps (ESOM) and Hidden Markov Models (HMM) as main artifacts in the analysis process. The user can define temporal, text mining and compound attributes. The text mining attributes are used to analyze the unstructured text in documents, the temporal attributes use these document's timestamps for analysis.

The compound attributes are XML rules based on text mining and temporal attributes. The user can cluster objects with object-cluster rules and can chop the data in pieces with segmentation rules. The artifacts are optimal zed for efficient data analysis, object labels in the FCA lattice and ESOM map contain an URL on which the user can click to open the selected document.

2. Business use cases



3.1 Loading data sources in CORDIET



Business event description:

Collect data sources for the investigation and store the results in a relational database and index the data source in a Lucene index.

Trigger:

A user wants to load the data to make it available for data analysis

Preconditions

The data sources must be in XML format

An ontology describing the data source architecture must be available for the database.

Active stakeholders:

Crime investigator

Interested stakeholders:

Project leader

Normal business flow

1. The user creates a new empty investigation
2. The user selects a XML data source
3. The user imports the XML data source into database
4. The user is finished with loading the data source.

Alternate business flow

1a If the user wants to add new data sources to an existing investigation

1a1 the user indicates that he wants to add a new data source to an existing investigation

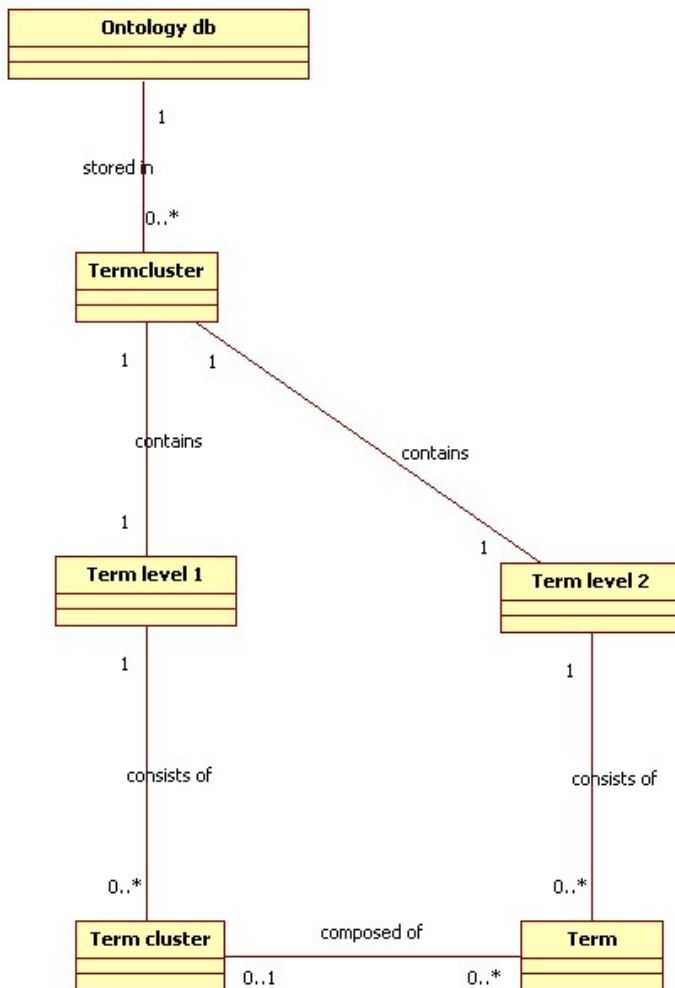
1a2 the user selects an existing data source. The flow returns to step 2.

Exception business flow

3a. If the XML data source is not wellformed

3a1. The user receives an error message “XML data source is not valid” and the use case ends

3.2 Creating an ontology for the investigation



Business event description:

Create an ontology containing text mining attributes.

Trigger:

The user wants to create a text mining ontology

Preconditions

Expert knowledge about the domain must be available

Active stakeholders:

Knowledge worker

Interested stakeholders:

Crime investigator

Normal business flow

1. The user creates a new initially empty text mining ontology
2. The user enters a name for the ontology
3. The user chooses to work on level 2 of the ontology
4. The user adds zero or more terms of level 2
5. The user removes zero or more search terms of level 2.
6. The user indicates she is finished with ontology maintenance work

Alternate business flow

1-2a. If the user chooses to open an existing ontology.

1-2a1. The user signals that she wants to open an existing ontology

1-2a2. The user selects an existing ontology file and opens the file. The flow returns to step 3.

3-5b. If the user chooses to work on level 1 of the ontology

3-5b1. The user indicates she wants to work on level 1 of the ontology

3-5b2. The user adds 1 or more term clusters of level 1 to the ontology

3-5b3. The user indicates one or more terms of level 2

3-5b4. The user adds the selected terms to the contents of a term cluster of level 1

3-5b5. The user removes 0 or more terms of level 2 from a term cluster

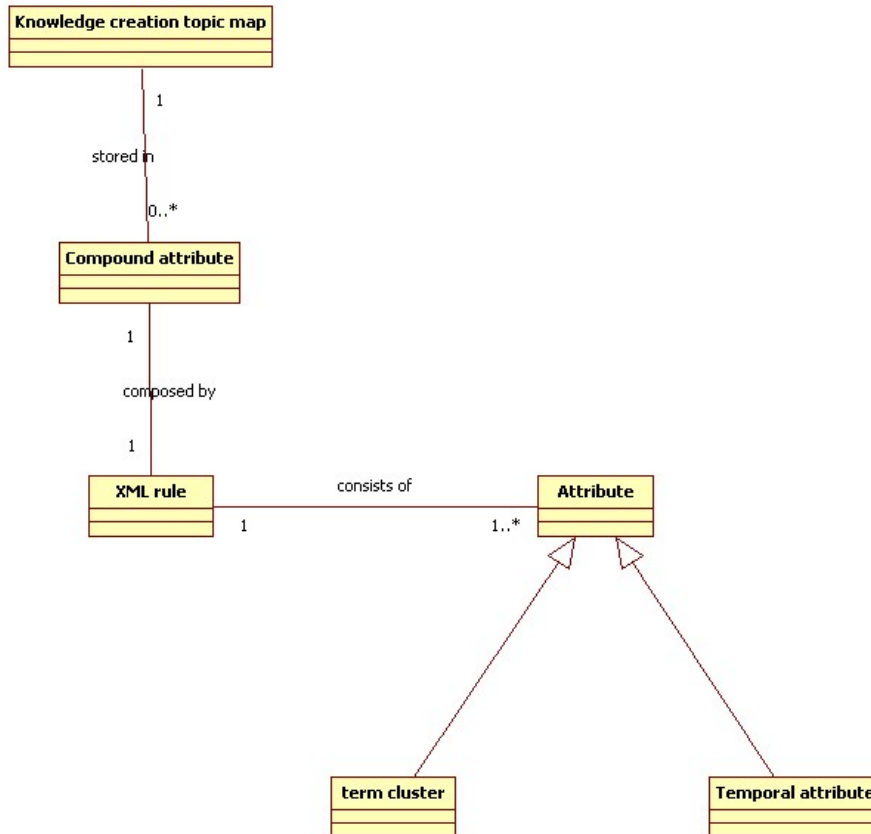
3-5b6. The user removes 0 or more term clusters of level 1 from the ontology. The flow returns to step 6.

6a. If the user chooses to continue ontology maintenance work.

6a1. The user indicates she wants to continue ontology maintenance work. The flow

returns to step 3.

3.3 Create a compound attribute



Business event description:

Create a compound attribute

Trigger:

A user wants to create a compound attribute

Preconditions

The knowledge creation topic map must be loaded including ontology and compound attribute rulebase.

Active stakeholders:

Knowledge worker

Interested stakeholders:

Crime investigator

Normal business flow

1. the user wants to create a new compound attribute
2. the user enters a name for the compound attribute
3. the user creates a new XML rule by including use case *create XML compound attribute rule*
4. the user store the compound attribute

Alternate business flow

1a If the user wants to modify an existing compound attribute

1a1 the user indicates he wants to modify an existing compound attribute

1a2 the user selects a compound attribute

1a3 the user modifies the compound attribute by including use case *modify compound attribute*. The flow returns to step 4.

1b If the user wants to delete an existing compound attribute

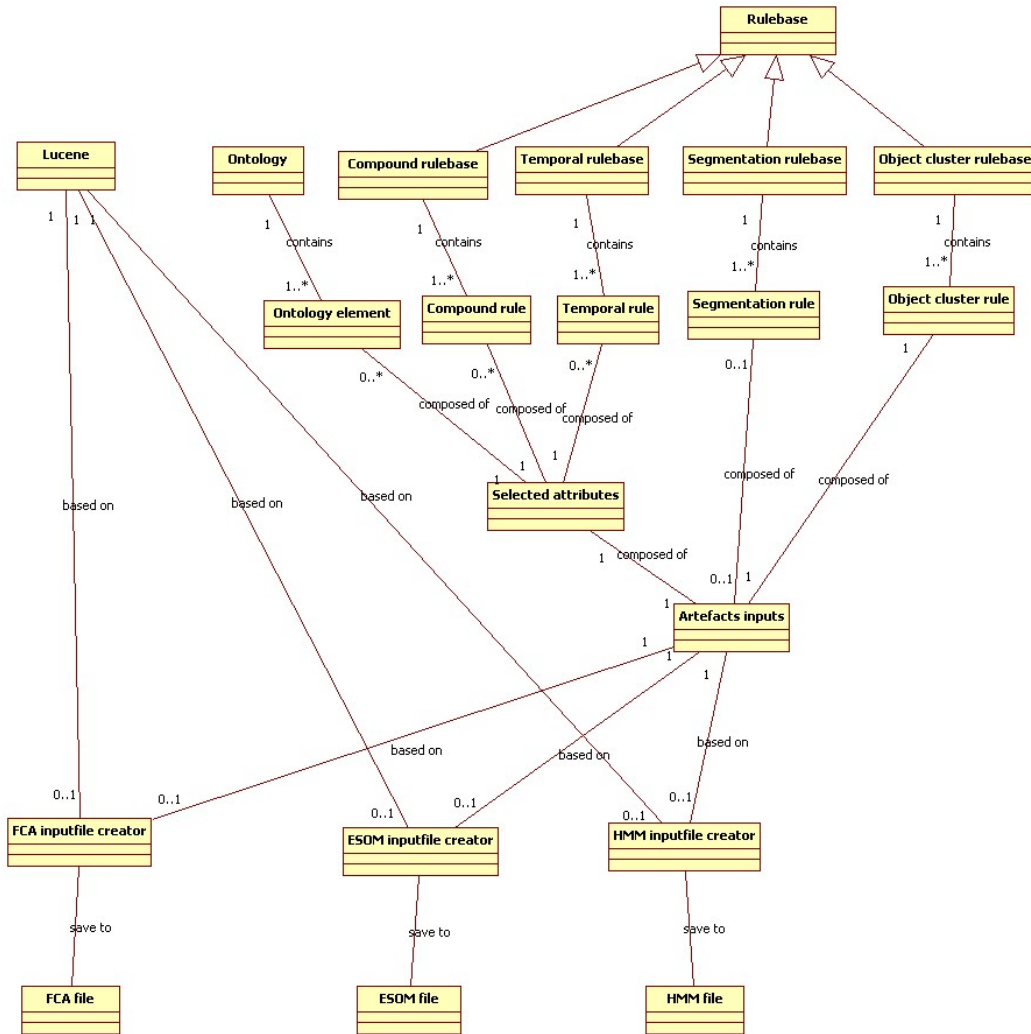
1b1 the user indicates he wants to delete an existing compound attribute

1b2 the user selects a compound attribute

1b3 the user deletes the compound attribute from the rulebase. The use case ends here.

Exception business flow

3.4 *Create an artifact input file*



Business event description:

Create an artefact.

Trigger:

The user wants to create an artefact

Preconditions

Datasource should be loaded

Ontology must be available

Active stakeholders:

Crime investigator

Knowledge worker

Interested stakeholders:

Project leader

Normal business flow

1. the user indicates that he wants to create an artifact
2. the user selects an available data source
3. the user selects an ontology
4. the user selects the FCA artifact by including use case *create FCA artifact*

Alternate business flow

4a. If the user wants to select an ESOM artifact

4a1. The user indicates he wants to create a ESOM artifact by including use case *create ESOM artifact*

4b. If the user wants to select a HMM artifact

4b1. The user indicates he wants to create a HMM artifact by including use case *create HMM artifact*

3.5 Create an FCA artifact input file

Business event description:

Create a FCA artefact.

Trigger:

The user has chosen to create a FCA artefact

Preconditions

A data source has been loaded

An ontology has been selected

The user has chosen to create a FCA artifact

Active stakeholders:

Crime investigator

Knowledge worker

Interested stakeholders:

Project leader

Normal business flow

1. the user selects the relevant text mining attributes of level 1
2. the users selects the relevant temporal attributes
3. the user selects the relevant segmentation rule
4. the user selects the desired object clustering rule
5. the user signals he wants to create the FCA input file
6. the user saves the FCA input file

Alternate business flow

1a. If the user wants to select all text mining attributes at once

1a1. the user indicates he wants to select all text mining attributes at once. The flow returns to step 2

1b. If the user does not want to use text mining attributes

1b1. The user indicates the wants to deselect all text mining attributes. The flow returns to step 2.

2a. If the user wants to select all temporal attributes at once

2a1. the user indicates he wants to select all temporal attributes at once. The flow returns to step 3

2b. If the user does not want to use temporal attributes

2b1. The user indicates the wants to deselect all temporal attributes. The flow returns to step 3

3a. If the user does not want to use a segmentation rule

3a1. The user indicates he does not want to use a segmentation rule. The flow returns to step 4.

4a. If the user does not select a object cluster rule

4a1. The user indicates he want to use the default object cluster rule. The flow returns to step 5

3.6 Create an ESOM artifact input file

Business event description:

Create an ESOM artefact.

Trigger:

The user has chosen to create an ESOM artefact

Preconditions

A data source has been loaded

An ontology has been selected

The user has chosen to create an ESOM artifact

Active stakeholders:

Crime investigator

Knowledge worker

Interested stakeholders:

Project leader

Normal business flow

1. the user select the text mining attribute level 2
2. the user selects the relevant text mining attributes

3. the users selects the relevant temporal attributes
4. the user selects the relevant segmentation rule
5. the user selects the desired object clustering rule
6. the user selects the desired classification attributes (more than one and always level 1, each attribute has its own color)
7. the user signals he wants to create the ESOM input file
8. the user saves the ESOM input file

Alternate business flow

1a. If the user wants to select the text mining attribute level 1

1a1. The user indicates he wants to select the text mining attribute level 1. The flow returns to step 2.

2a. If the user wants to select all text mining attributes at once

2a1. the user indicates he wants to select all text mining attributes at once. The flow returns to step 3

2b. If the user does not want to use text mining attributes

2b1. The user indicates the wants to deselect all text mining attributes. The flow returns to step 3.

3a. If the user wants to select all temporal attributes at once

3a1. the user indicates he wants to select all temporal attributes at once. The flow returns to step 4

3b. If the user does not want to use temporal attributes

3b1. The user indicates the wants to deselect all temporal attributes. The flow returns to step 4

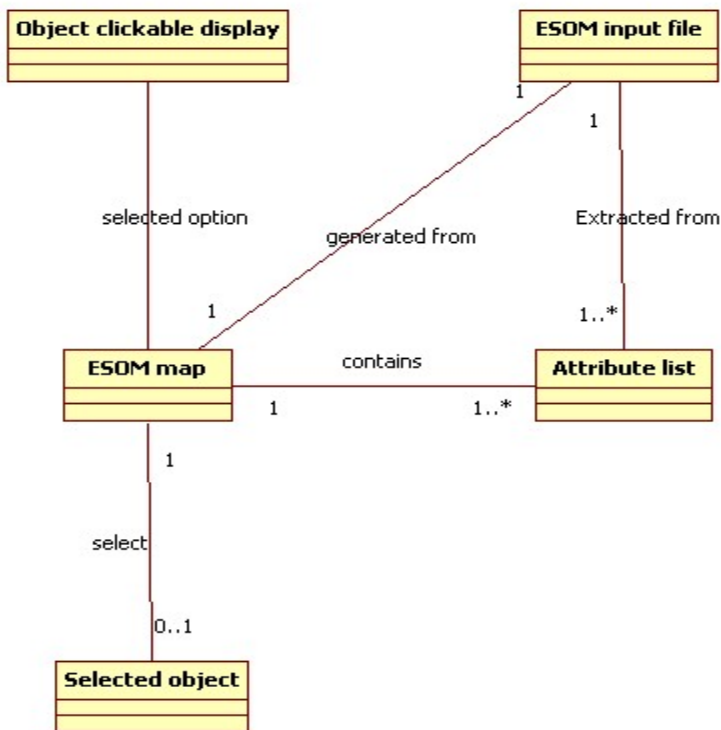
4a. If the user does not want to use a segmentation rule

4a1. The user indicates he does not want to use a segmentation rule. The flow returns to step 5.

5a. If the user does not select a object cluster rule

5a1. The user indicates he want to use the default object cluster rule. The flow returns to step 6q1

3.7 Browse additional functionality of ESOM map



Business event description:

Browse additional functionality of the ESOM map

Trigger:

A user wants to analyse an ESOM map

Preconditions

The ESOM input file must have been created.

The input file contains at least one attribute

The ESOM map must have been generated

Active stakeholders:

Knowledge worker

Crime investigator

Interested stakeholders:

Project leader

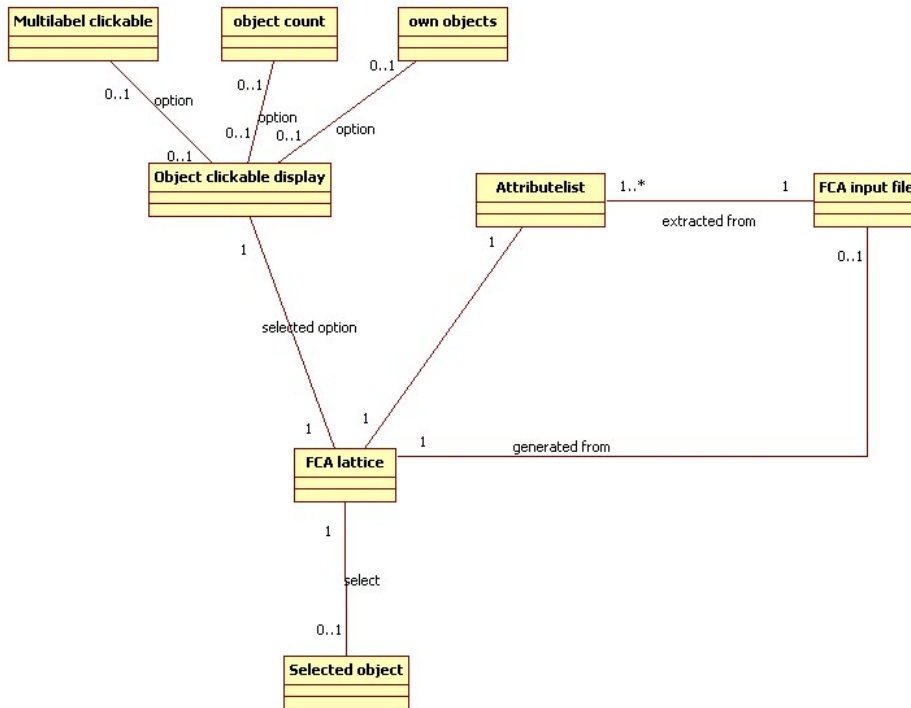
Normal business flow

5. The user selects a clickable object display option
6. The user selects an object by clicking on the hyperlink by including use case *highlighting document*.

Alternate business flow

Exception business flow

3.8 Browse additional functionality of FCA lattice



Business event description:

Browse additional functionality of the FCA lattice

Trigger:

A user wants to analyse a FCA lattice

Preconditions

The FCA input file must have been created.

The input file contains at least one attribute

All attributes from the attribute list are deselected, except the first one in the list

The FCA lattice must have been generated

Active stakeholders:

Knowledge worker

Crime investigator

Interested stakeholders:

Project leader

Normal business flow

7. The user selects attributes from the attribute list
8. The user selects a clickable object display option
9. The user selects an object by clicking on the hyperlink by including use case *highlighting document*.
10. The user deselects attributes from the attribute list

Alternate business flow

1a If the user wants to select all attributes

1a1 the user indicates he wants to select all attributes. The flow returns to step 2

4a If the user wants to deselect all attributes

4a1 the user indicates he wants to deselect all attributes. The use case ends here.

Exception business flow

3. Functionalities and modules

Loading data sources:

Data files used as input for the CORDIET software package should be in XML format. This XML file has an identifier and a number of structured data fields. For example an XML input file corresponding to a police report, contains the name of the suspect, the location of the incident and the textual report. Each of these fields has a value. The XML document also has a timestamp. For police reports this is the time of the incident. Finally the XML document contains the unstructured text. For a police report this is the statement made by the victim or the observations made by the police officer. Each XML document contains only 1 data document, for example one police report or one patient and all the activities performed to this patient. For our dataset of 4814 domestic violence reports. We transform these reports in 4814 XML

documents and input them in a PostgreSQL database.

PostgreSQL database:

With the PostgreSQL database a data ontology is associated which contains the structure of the input data so that the data can be read and stored into the database. This file is an XSD file and used to verify the well-formedness of the XML files using a SAX or DOM parser. The parsing procedure goes as follows. The unique identifier of the XML document is retrieved, It is verified if the database table already contains the XML file. If the document was already added, the record is updated, otherwise a new record is created. The timestamp of the file is retrieved and stored, in the database. The object node list is retrieved from the XML file, the element <object> is retrieved, the value of the attribute "type" is retrieved, the value of the object is retrieved. A new record is created in the data source table. The following attributes of the record are inserted: datasource_id, unique identifier UUID, filename, document path, document url, XML document. Then a new Lucene document is created.

Lucene, Luke:

Lucene index is used to index and store the documents. The field "content" will contain the unstructured text of the XML input file. The Lucene index stores for each term where it appears in which documents. Lucene allows to quickly search and find out in how many for example "Person" fields or "content" fields the name "Jan Janssen" appears. When filling the database, the documents are also stored in the Lucene index. Lucene has many interesting options for storing documents, options should be compared for optimizing performance. Timestamps and data can be stored in Lucene as a number. This number stored in the field "date" can be used for defining temporal rules by using the [... To ...] operator-which returns all documents between two dates.

PostgreSQL domain ontology:

The database not only contains an ontology defining the data architecture but also a domain ontology defined by the user containing terms, term clusters, temporal rules, compound attributes, etc. Text mining attributes, temporal attributes and compound attributes can be added and removed using the respective ontology maintenance modules.

Text mining attributes

A text mining attribute can be a term or a term cluster. A term is an array of search terms. A term cluster is a list of terms. For example the term cluster "family" consists of the terms "mother", "father", "uncle", etc. The term "father" consists of the search terms "my father" "my dad", "my daddy", etc.

Temporal attributes

A temporal attribute consists of a name and an XML rule that uses timestamps available in the data. For example to find all criminals that were seen 4 times or more by the police between january 2009 and june 2009." This rule can be used to find unknown repeat offenders. It uses the timestamps of the police reports. A list of examples will be given in the temporal attributes section. An XML syntax should be introduced for defining these rules. It is possible to use the date field in Lucene. A temporal rule language should be defined for working with these dates. Complex rules should be transformed to operations on these date fields.

Compound attributes

Compound attributes have a name and XML rule. This XML rule uses temporal and text mining attributes. Again an XML syntax should be defined.

Language module:

Different languages including English, Dutch and Russian should be supported. The user must be able to choose between these languages. Lucene indexer of documents has a Russian Analyzer,

Dutch Analyzer, etc.

I/O module

To load an ontology into the database or write it to a file from the database. A parser should be developed for reading and writing.

Object-cluster rules

Object-cluster rules are used to group individual documents or records based on a common characteristic they share. For police investigations, documents can be grouped based on the person who committed the crime or the date at which events occurred for example. A new object of type object-cluster is created and these objects are used for creating the artefact input files instead of the individual documents.

Data segmentation module

The goal of segmentation rules is to distill a segment from the data based on the occurrence or absence of attributes or attribute combinations in the documents. An XML syntax should be defined for defining these rules. On implementation level, the Lucene index should be defined as an object according to OO principles. When a selection is made from this index, this new index object is created in the RAM memory. To each segmentation rule is a Lucene index stored in the RAM related. This is done as follows: the subset of documents in the Lucene index that fulfils the requirements of the query is indexed in the RAM memory and the tool further uses this index as long as the segmentation rule is selected. The student can do this by creating a new RAM Directory and give it as parameter to IndexWriter method of Lucene API.

Data segmentation rule example:

1. Select all patients who had a length of stay in the hospital larger than 6 days and received mastectomy surgery.

- This rule uses the temporal attribute that indicates if the timestamp of the hospitalization record of the patient subtracted from the timestamp of the discharge record of the patients is larger than 6 days. The rule also uses the predefined field in the XML input file that indicates the type of surgery. The relevant objects are selected by applying the rule on the Lucene index.

Concept Explorer

Concept Explorer is a tool for lattice creation and visualization. It takes a csv file as input in which the rows contain the objects, for example a police report and the columns the attributes. For each object, a line is stored containing the document identifier, and a 0 or 1 indicating that the report does not or does have the attribute in the corresponding column. From this data table a concept lattice is derived. We want to provide an extra data column with the csv file containing the document url. The user can then choose to visualize the object labels as URLs on which the user can click. A separate text highlighting module then displays the text and marks the found thesaurus elements. This facilitates browsing the data. Also, a deselect all option should be added for the attributes and objects. The lattice should initially be created with all attributes from the table deselected.

Databionics ESOM

This package for creating a special class of topographic maps is written in java and takes a lrn and cls file as input. With each neuron in the map, objects are associated, these objects can be individual documents or clusters of documents, e.g. a police report or a person who has multiple police reports. If a user clicks on a neuron, the documents appear in the lower pane of the tool. Also here a clickable label should be used which opens a separate window with the document

and the highlighter module enabled. If the neuron contains document clusters, e.g. a person, a separate TCA lattice showing all the documents of this cluster and the concepts they contain should be opened in a separate window.

TCA lattice

If the user chose to apply an object-cluster rule, documents are clustered based on a common characteristic, for example the person who committed the crime. If these persons, i.e. document clusters, are used as objects for the FCA or ESOM artifact creation, it should be possible to easily explore the conceptual structure of these persons by generating a profile of the suspect selected for investigation. If the user clicks on the name of a person in the FCA lattice, a TCA profile should be created, showing the lattice with the documents of these person as objects and their attributes. The documents are moreover chronologically ordered in the visualization based on their timestamps. This allows to investigate appearance and change of concepts over time and to monitor the suspect's evolution.

Hidden Markov Model

For a list of events, e.g. a list of activities performed to patients, the time stamps and activities performed are transformed in event sequences. HMM is created from event sequences.

Highlighter module

When a user selects a document and clicks on it in the FCA lattice or ESOM map, a separate window opens in which the text in the document is shown and attributes from the text mining module are marked in color. Terms and phrases in this document that are selected by the user are automatically filled in, in the fields provided by the user interface of the thesaurus maintenance module that automatically opens up.

4. User interface

1. First main window, $K \rightarrow C$ step, selecting the existing K artefacts for analysis:
 - Loading data source into tool
 - Create or load ontology containing terms, term clusters, temporal attributes, compound attributes
 - Start ontology maintenance module containing:
 - i. Text mining attribute maintenance module
 - ii. Temporal attribute maintenance module
 - iii. Compound attribute maintenance module.
2. Second main window, $C \rightarrow C$ step, composing and creating artifacts
 - Compose artefact input module containing:
 - i. Text mining attribute selection: select and deselect all, select individual, etc. from list
 - ii. Temporal attribute selection from list
 - iii. Segmentation rule selection
 - iv. Object-cluster rule selection
 - v. Compound attribute selection from list
 - Choose artefact to be created
 - Create artefact
3. Third main window, $C \rightarrow K$ step, browsing artifacts, selecting new text mining attributes, annotating documents
 - Depending on the artefact choice:
 - i. Artefacts should be made browsable and object labels clickable, such that the document opens and is displayed in a separate window

- A term highlighter should be used that shows the found terms and term clusters from the ontology in the document.
 - An intuitive to use display and method should be given to select a term and add it to a thesaurus.
4. Fourth main window, $K \rightarrow K$ step, reporting to organization and updating knowledge base
- Show list of case numbers with notes made during previous step
 - Create and store classification rules
 - Jump back to first screen button
 - Create summary report containing:
 - i. Annotated case numbers and thesaurus elements added during $C \rightarrow K$ which also indicates the police reports from which they were selected
 - ii. List of classification rules added during $K \rightarrow K$ step

5. Data analysis scenarios

This section describes some interesting analysis scenarios and how they can be performed with CORDIET.

1. The user is interested in finding all persons who were spotted more than 4 times in a prostitution area by the police in a period of 4 months, defined to be January 2010 until May 2010.
 - A data segmentation rule that uses the temporal attribute for the given period is defined. The temporal attribute says `timestamp_doc > 01012010` and `timestamp_doc < 01052010`. A data segmentation rule is defined that selects all police reports that have this temporal attribute.

- We are however only interested in persons not in individual documents. All suspect names known to the police can be retrieved from a police database and were stored in the Lucene index. The Lucene index indicates in which documents the names appear, using an object-cluster rule we can cluster the segmented set of reports based on the known persons they contain. These known persons are a text mining attribute.
- We then select the attributes to be used for building the lattice or ESOM map. The FCA lattice can be used to investigate the properties of these suspects and find potential relations between them. The ESOM map can reveal suspects with a similar profile. For the HMM model, the documents and indicators observed for the suspect are turned into an event sequence that may give insight into the criminal career of the suspect.

6. Appendix

Example input XML file

```
<?xml version="1.0" encoding="UTF-8" ?>
<datasource uuid="1212121">

  <filename>thisfile.xml</filename>

  <filepath>d:\data\thisfile.xml</filepath>

  <fileurl>pages/showxml.jsp?id=1212121</fileurl>

  <datetime>2010-09-16T12:00:00</datetime>

  <objects>

    <object type="person" id="1">Jansen, Jan 1966-12-01</object>

    <object type="location" id="2">Amsterdam, Jameswattstraat 84
      1097DM</object>

  </objects>

  <content>[[CDATA this the content ]</content>

  <binarycontent />

</datasource>
```

7. Publications

Journal papers published:

1. Poelmans, J., Elzinga, P., Viaene, S., Van Hulle, M. & Dedene G. (2009). Gaining insight in domestic violence with emergent self organizing maps, Expert systems with applications, 36, (9), 11864 – 11874. [SCI=2.596]
2. Poelmans, J., Elzinga, P., Viaene, S., Dedene, G. (2010) Formally Analyzing the Concepts of Domestic Violence, accepted for Expert Systems with Applications.
3. Vuylsteke A., Baesens B., Poelmans J. (2010). Consumers' search for information on the internet: how and why China differs from Western Europe, Accepted for Journal of interactive marketing.

4. Poelmans, J., Elzinga, P., Viaene, S., Dedene, G., Van Hulle, M. (2009). Analyzing domestic violence with topographic maps: a comparative study, Lecture Notes in Computer Science, 5629, 246 – 254, Advances in Self-organizing Maps, 7th International Workshop on Self-Organizing Maps (WSOM). St. Augustine, Florida (USA), 8-10 June 2009, Springer. [SCI=0.295]
5. Poelmans, J., Elzinga, P., Viaene, S., Dedene, G. (2009). A case of using formal concept analysis in combination with emergent self organizing maps for detecting domestic violence, Lecture Notes in Computer Science, 5633, 247 – 260, Advances in Data Mining. Applications and Theoretical Aspects, 9th Industrial Conference (ICDM), Leipzig, Germany, July 20-22, 2009, Springer. [SCI=0.295]
6. Poelmans, J., Elzinga, P., Viaene, S., Dedene, G. (2008). An exploration into the power of formal concept analysis for domestic violence analysis, Lecture Notes in Computer Science, 5077, 404 – 416, Advances in Data Mining. Applications and Theoretical Aspects, 8th Industrial Conference (ICDM), Leipzig, Germany, July 16-18, 2008, Springer. [SCI=0.295]
7. Poelmans, J., Elzinga, P., Viaene, S., Dedene, G. (2010), Formal Concept Analysis in knowledge discovery: a survey. Lecture Notes in Computer Science, 6208, 139-153, 18th international conference on conceptual structures (ICCS): from information to intelligence. 26 - 30 July, Kuching, Sarawak, Malaysia. Springer.
8. Manyakov, N., Poelmans, J., Vogels, R., Van Hulle, M. (2010), Combining ESOMs trained on hierarchy of feature subsets for single-trial decoding of LFP responses in monkey area V4. Lecture Notes in Artificial Intelligence, 6114, 548-555, 10th International Conference on Artificial Intelligence and Soft Computing (ICAISC). June 13-17, Zakopane, Poland. Springer
9. Poelmans, J., Dedene, G., Verheyden, G., Van der Mussele, H., Viaene, S., Peters, E. (2010). Combining business process and data discovery techniques for analyzing and improving integrated care pathways. Lecture Notes in Computer Science, Advances in Data Mining. Applications and Theoretical Aspects, 10th Industrial Conference (ICDM), Leipzig, Germany, July 12-14, 2010. Springer

Conference proceedings published:

10. Poelmans, J., Dedene, G., Snoeck, M. Viaene, S. (2010). Using Formal Concept Analysis for the Verification of Process-Data matrices in Conceptual Domain Models, Proc. IASTED International Conference on Software Engineering (SE 2010), Feb 16 - 18, Innsbruck, Austria. Acta Press, pp..

11. Poelmans, J., Elzinga, P., Viaene, S., Dedene, G. (2010). A method based on Temporal Concept Analysis for detecting and profiling human trafficking suspects. Proc. IASTED International Conference on Artificial Intelligence (AIA 2010). Innsbruck, Austria, 15-17 february. Acta Press ISBN 978-0788986-817-5, pp. 330-338.
12. Elzinga, P., Poelmans, J., Viaene, S., Dedene, G. (2009), Detecting domestic violence – Showcasing a Knowledge Browser based on Formal Concept Analysis and Emergent Self Organizing Maps, Proc. 11th International Conference on Enterprise Information Systems ICEIS, Volume AIDSS, pp. 11 – 18, Milan, Italy, May 6-10, 2009.
13. Poelmans, J, Elzinga, P., Van Hulle, M., Viaene, S., and Dedene, G. (2009). How Emergent Self Organizing Maps can help counter domestic violence, World Congress on Computer Science and Information Engineering (CSIE 2009), Los Angeles (USA), Vol. 4, IEEE Computer Society Press ISBN 978-0-7695-3507-4, 126 – 136.
14. Vuylsteke, A., Wen, Z., Baesens, B. and Poelmans J. (2009). Consumers Online Information Search: A Cross-Cultural Study between China and Western Europe. Paper presented at Academic And Business Research Institute Conference 2009, Orlando, USA, available at <http://www.aabri.com/OC09manuscripts/OC09043.pdf>
15. Elzinga, P., Poelmans, J., Viaene, S., Dedene, G., Morsing, S. (2010) 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, 77-82.
16. Poelmans, J., Elzinga, P., Viaene, S., Dedene, G. (2010). Concept Discovery Innovations in Law Enforcement: a Perspective, accepted for Computational Intelligence in Networks and Systems workshop (INCos 2010), Thessaloniki, Greece.
17. Dejaeger, K., Hamers, B., Poelmans, J., Baesens, B. (2010) A novel approach to the evaluation and improvement of data quality in the financial sector, accepted for 15th International Conference on Information Quality (ICIQ 2010) UALR, Little Rock, Arkansas USA, November 12-14.

Journal papers submitted:

18. Verheyden, G., Poelmans, J., Viaene, S., Van der Musselle, H., Dedene, G., van Dam, P. (2010). Key Success Factors for significantly improving Patient Satisfaction on Breast Cancer care: a Case Study, submitted for The Breast

19. Poelmans, J., Elzinga, P., Viaene, S., Dedene, G. (2010) Curbing domestic violence: Instantiating C-K theory with Formal Concept Analysis and Emergent Self Organizing Maps, submitted for IEEE transactions on knowledge and data engineering.
20. Poelmans, J., Elzinga, P., Viaene, S., Van Hulle, M. & Dedene G. (2010) Text Mining with Emergent Self Organizing Maps and Multi-Dimensional Scaling: A comparative study on domestic violence, submitted for Applied Soft Computing.
21. Poelmans, J., Elzinga, P., Viaene, S., Dedene, G. (2010) Formal Concept Analysis in Information Engineering: a Survey, submitted for ACM Computing Surveys.
22. Poelmans, J., Elzinga, P., Viaene, S., Dedene, G. (2010) Informatiegestuurd politiewerk: een slimme en nieuwe kijk op de data, submitted for Informatie.
23. Elzinga, P., Poelmans, J., Viaene, S., Dedene, G. (2010) Formele concept analyse: een nieuwe dimensie voor intelligence, submitted for Blauw.

Book chapter:

24. Poelmans J, Van Hulle M, Elzinga P, Viaene S, Dedene G (2008) Topographic maps for domestic violence analysis. Self-organizing maps and the related tools, pp. 136 - 145.

Dutch Publications:

25. Vuylsteke A., Poelmans J., Baesens B. (2009) Online zoekgedrag van consumenten: China vs West-Europa, Business In-Zicht, December, 2-3.
26. Dejaeger K., Ruelens J., Van Gestel T., Jacobs J., Baesens B., Poelmans J., Hamers B. (2009) Evaluatie en verbetering van de datakwaliteit. Informatie, November, Jaargang 51/9, 8-15.

Awards:

Nominated for best paper award at 8th Industrial Conference on Data Mining (ICDM), Leipzig, Germany, July 16-18, 2008

Winner of young professionals best paper award at 9th Industrial Conference on Data Mining (ICDM), Leipzig, Germany, July 20-22, 2009

Winner of best paper award at 10th Industrial Conference on Data Mining (ICDM), Berlin, Germany, July 12-14, 2010