The CUBIST Project: Combining and Uniting Business Intelligence with Semantic Technologies

ANDREWS, Simon <http://orcid.org/0000-0003-2094-7456>

The CUBIST Project: Combining and Uniting Business Intelligence with Semantic Technologies

Simon Andrews
Conceptual Structures Research Group, Sheffield Hallam University, UK

ABSTRACT

As a preface to this Special 'CUBIST' Edition of IJIIT, this article describes the European Framework Seven CUBIST project, which ran from October 2010 to September 2013. The project aimed to combine the best elements of traditional BI with the newer, semantic, technologies of the Sematic Web, in the form of the Resource Description Framework (RDF), and Formal Concept Analysis (FCA). CUBIST's purpose was to provide end-users with "conceptually relevant and user friendly visual analytics" to allow them to explore their data in new ways, discovering hidden meaning and solving hitherto difficult problems. To this end, three of the partners in CUBIST were use-cases: recruitment consultancy, computational biology and the space industry. Each use-case providing their own requirements and problems that were finally addressed by the prototype CUBIST visual-analytics developed in the project.

Keywords: Semantic Technology; Business Intelligence; Formal Concept Analysis; FCA; Resource Description Framework; RDF; Semantic Web; CUBIST Project; Gene Expression; Satellite Telemetry; Recruitment Consultancy.

INTRODUCTION

There have been three conference workshops dedicated to the CUBIST project (Andrews, S., Dau,F., 2012; Andrews, S., Dau,F., 2013; F. Dau, 2011). Each has involved a number of papers produced by members of the project consortium and by people external to CUBIST with an interest in its aims and approaches. In this special edition of IJIIT, selected extended versions of some of these papers are presented.

In *Browsing Large Concept Lattices through Tree Extraction and Reduction Methods*, Melo, Le-Grand and Aufare describe an approach used in CUBIST to simplify concept lattices by extracting and visualising trees derived from them. Browsing concept lattices from Formal Concept Analysis becomes a problem as the number of concepts can grow significantly with the number of objects and attributes. Alternative, browse-able trees, combined with reduction methods such as fault-tolerance and concept clustering provide a much more manageable approach.

M^{c}Leod, Iskandar and Burger take the semantic approach to exploring biological images in *Towards the Semantic Representation of Biological Images: From Pixels To Regions*.

Biomedical images and models contain vast amounts of information, only accessible by domain experts. Although a semantic representation in which every image pixel is featured is expensive, more abstract renditions, such as those made possible through Region Connection Calculus and the W3C Geospatial Vocabulary, provide a basic description that enables a non-expert end-user to perform a number of queries.

In *Gene Co-Expression in Mouse Embryo Tissues*, Andrews and M^{c}Leod develop some existing ideas in Formal Concept Analysis to provide an analysis of a large data set of gene expressions in mouse embryo tissues. In biology, there is a mapping between a protein and the gene that helped create it. Working together in groups, genes are responsible for the production of tissues and organs, thus the identification of such groups is of interest to biologists exploring the development and specialisation of tissues. This paper describes a new technique for managing complexity based on 'fault tolerance' and the identification of disjoint sets in data. Using this technique, distinct groups of co-expressed genes are identified.

Polovina explores the possibilities of using the semantics of transactions to give enterprises new insight into their business processes. He describes *A Transaction-Oriented Architecture for Enterprise Systems*. Many enterprises risk business transactions based on information systems that are incomplete or misleading, given that 80-85% of all corporate information remains outside of the processing scope of such systems. Enterprise architectures are needed that captures more of this 'soft' information. Such architectures can be achieved through modelling more holistically the transactions that collectively describe the business and its processes. Such an architecture captures the real-world meaning of the transactions, something not possible using traditional, data driven, IT approaches.

In this 'extended preface' to the special edition, the CUBIST project itself is described, covering its background, the real-world use-cases involved in the project, the semantic technologies developed and employed and some of the results obtained. Finally, some key areas of further work and exploitation are described with a conclusion to summarise the overall success of the project.

It is anticipated that the reader of this CUBIST edition of IJIIT will be motivated to explore semantic technologies, and, in particular, the new possibilities of applying Formal Concept Analysis (FCA) and the Resource Description Framework (RDF) to their data, perhaps initially by exploring the contents of the original CUBIST conference workshops. Creating an ontology of their information, storing their data as RDF triples and employing new visual analytics may give the reader new ways to be productive with their own data and find hidden insights into their area of business or research domain.

THE CUBIST PROJECT - BACKGROUND AND OVERVIEW

Constantly growing amounts of data, complicated and rapidly changing economic interactions, and an emerging trend of incorporating unstructured data into analytics, is bringing new challenges to Business Intelligence (BI). Contemporary solutions involve BI users dealing with increasingly complex analyses. According to a 2008 study by Information Week, the complexity of BI tools and their interfaces becomes the biggest barrier for success of these systems. Moreover, classical BI solutions have, so far, neglected the meaning of data, which can limit the

completeness of analysis and make it difficult, for example, to remove redundant data from federated sources.

Semantic Technologies (ST), however, deal with the meaning of data and are capable of dealing with both unstructured and structured data. By having the meaning of data (and a reasoning mechanism) in place, users can be guided during his work with data. For example, a piece of information can be explained or a new fact brought to the user's attention. In particular, we employ Formal Concept Analysis (FCA) (B. Ganter, Stumme, & Wille, 2005; B. Ganter & Wille, 1998), which is a well-known semantic technique, as a key element in a new hybrid BI system. FCA can be used to guide a user in discovering new facts, which are not explicitly modelled in the data warehouse schema. However, semantic technologies have, traditionally, operated on data sets a magnitude smaller than classical BI solutions. They also lack standard BI functionalities such as Online Analytical Processing (OLAP) queries, making it difficult to perform analysis over semantic data. On the other hand, 'understanding data' could improve classical methods in BI, such as data reduction and duplicate detection. Figure 1 compares the classical and sematic approaches side-by-side.

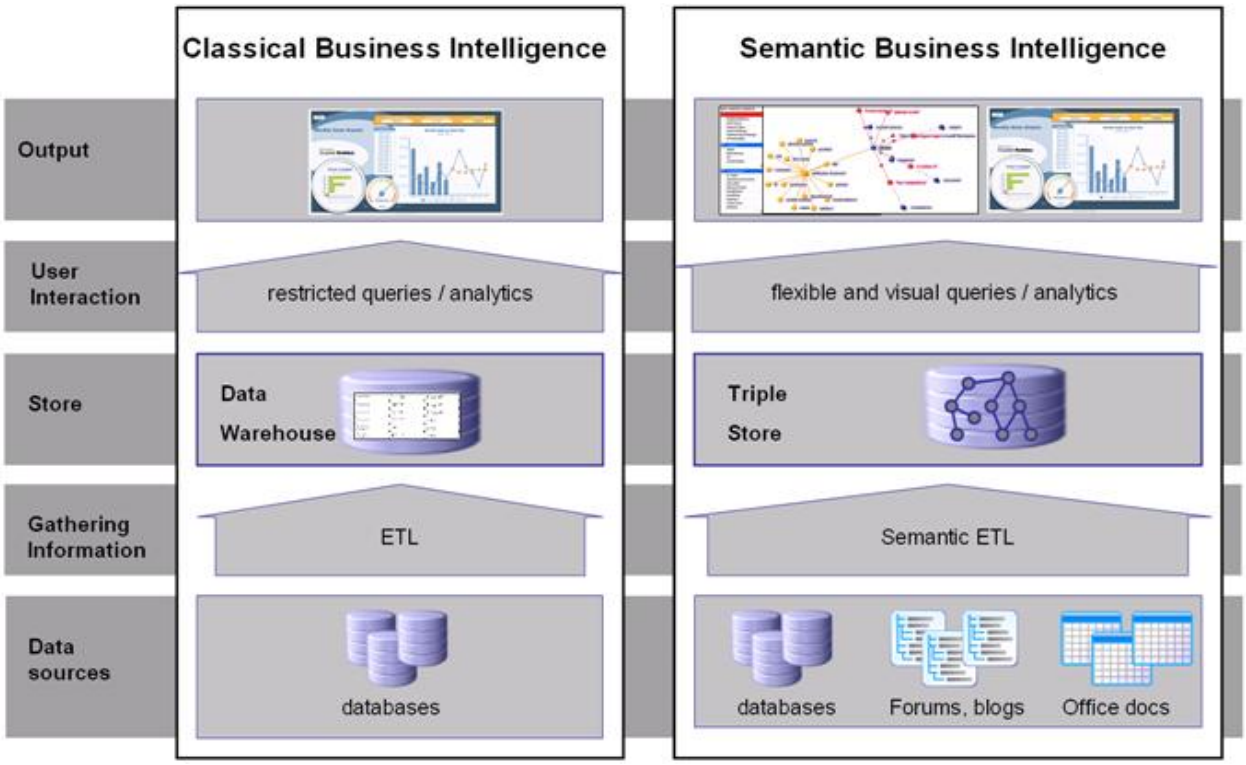

Figure 1. From Classical to Semantic BI

To address these issues, the CUBIST project developed methodologies and a platform, which combines essential features of Semantic Technologies and Business Intelligence. The CUBIST project has developed a system with the following core features:

- It supports federation of data from a variety of unstructured and structured sources;
- The persistency layer is a Semantic Data Warehouse; a hybrid approach based on mature database technology and a BI enabled triple store;

- Semantic information is used to improve BI best practices in, for example, data reduction and pre-processing;
- CUBIST enables a user to perform BI operations over semantic data;
- The Semantic Data Warehouse is used to realize advance mining techniques known from, in particular FCA (Formal Concept Analysis). FCA guides the user in performing BI and helps the user discover facts not expressed explicitly by the warehouse model;
- Novel ways of applying visual analytics in which meaningful diagrammatic representations of the data are used for depicting the data, navigating through the data and for visually querying the data.

Technology and academic partners provided the expertise for creating the system platform and software: the Semantic Data Warehouse, data extraction, transformation and loading, FCA tools and visual analytics.

CUBIST has demonstrated the resulting technology stack in the fields of recruitment consultancy, computational biology and the space industry. These use-cases were fundamentally involved in the developed CUBIST system, from functionality to data representation, to user interfaces, to visualisation of results. Each use-case is described in the next section.

USE-CASES

Recruitment Consultancy

The use-case was a leading technology company currently making a big impact in the UK HR and staffing sectors. It provides sophisticated access to information that enables the UK staffing sector to find and better understand companies who advertise online. The company's proprietary software tracks live job vacancies advertised on the UK's leading job boards as well as employers' own websites. Using a powerful OLAP reporting suite, clients can research the market in seconds and talk with authority about their industry; highlighting recruitment trends within it and revealing information about their biggest competitor. The main source of information is job postings gathered from job boards and employer websites. With analytics over these unstructured sources provided by CUBIST, the company wanted to significantly improve the market intelligence and competitive intelligence for their clients.

The company's goal in CUBIST was to extend its services in two new directions: Market Intelligence – with extended analysis of recruitment behaviour to answer who, where, when and how questions about recruitment strategies, trends of jobs and specific sectors in the market, and Competitive Intelligence – with provision of data back to UK employers to help track and better understand the recruitment activity of their competitors.

Space Industry

The use-case was a leading provider of system and operations engineering as well as software engineering expertise in the field of space and aerospace applying these capabilities to industrial applications including:

- Space system engineering, specification, operations engineering, training and software development from the earliest phases of spacecraft and mission concept definition to on-orbit operations;
- Software Engineering: Design and development of monitoring and control systems, distributed control for fixed and mobile robots in structured and unstructured environments;
- Research and Development: Establishment of methods and processes for collaborative multi modal human-computer interaction; development of knowledge management systems.

The company deals with large quantities of mission data, of structured and unstructured types, accumulated over a long period of time. The company aimed to deploy CUBIST solutions that will increase the cost-efficient utilisation of the vast quantity of structured and unstructured data present within the development and operations lifecycle of their systems. They wanted particularly to develop CUBIST methods of identifying and solving operational problems related to data consistency, completeness and determinism, and for the automatic generation of new content and documents from existing data.

Computational Biology

The use-case was a Biomedical Informatics Systems Engineering Laboratory focussing on issues that arise from bringing to bear the latest computer science developments in the context of biomedical research. The work concentrates on the next generation of distributed informatics systems and includes research on interoperability of biomedical atlases, multi-agent systems in biomedical informatics, augmentation systems and AI planning for distributed task composition. Work on ontologies, includes cross-species anatomy ontology integration and 2D and 3D visualization of anatomy ontologies. The group has extensive experience in using semantic web technologies in the context of biomedical atlases and wanted to investigate the possibilities of new visual analytics for the Edinburgh Mouse Atlas of Gene Expression (EMAGE) (Richardson et al., 2010), a database of gene expression assays conducted on mouse embryo tissue.

SEMANTIC DATA WAREHOUSE

CUBIST employs an RDF triple store and ontology as the backbone for the information warehouse, to improve performance and reduce the complexity of the integration of heterogeneous data sources. ST enriches BI by enabling the discovery of new implicit information through logical reasoning. The standard RDF query language SPARQL is used as the query interface between the triple store and the new, FCA-based, visual analytics. The information warehouse uses advanced indexing and materialisation techniques known from state-of-the-art data warehousing to improve the performance of the RDF triple store.

Facilitating RDF as the uniform data representation model within an enterprise reduces the complexity of integrating many heterogeneous data sources (structured or unstructured) during the Extract, Transform, Load (ETL) phase. Additionally, RDF makes it possible for sound logical reasoning rules to be defined in order to infer new facts from the original data acquired during the ETL process. The RDF conceptual model also introduces unmatched flexibility and agility to the process of modelling the data warehouse (as compared to traditional ‘star schema’

approaches). And, since the semantics and relationships between various facts in the warehouse are explicit, it is also possible to perform advanced semantic search and exploration that goes beyond the traditional faceted search in a BI context. Whilst relational databases go some way to provide semantics, ontologies more easily and naturally capture extended semantics, such as inference, hierarchies and object-relations.

In CUBIST, each of the use-case data was modelled with an agreed ontology before a semi-automated ETL process converted the data into RDF and loaded it into the triple store. Figure 2 shows the ontology for the EMAGE data set, showing the relations between genes, strength of expression, tissues, stage of embryo development (Theiler Stage) and experiment. The taxonomy of anatomy of mouse embryo is captured by the "is part of" hierarchical relation of tissues. A natural way of analysing and visualising such relations is provided by FCA, as described in the next section.

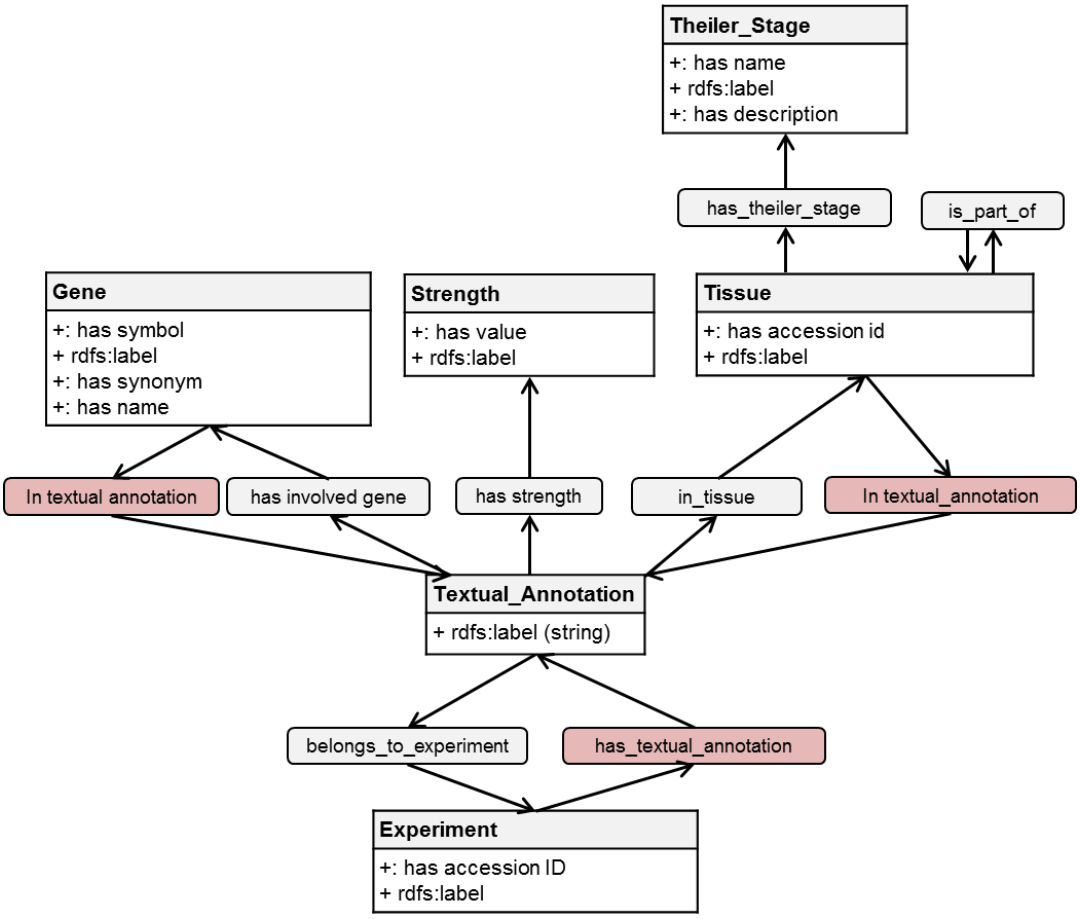

Figure 2. EMAGE Ontology

FORMAL CONCEPT ANALYSIS

Although the Semantic Web and RDF are reasonably well-known, FCA is less so. Thus it is sensible to provide a short introduction to it here. Formal Concept Analysis is a term that was

introduced by Rudolf Wille in 1984 and builds on "applied lattice and order theory that was developed by Birkhoff and others in the 1930's" (Wille, 2005). It was initially developed as a subsection of Applied Mathematics based on the mathematisation of *concepts* and concepts hierarchy. A *formal context* is a triple (*G*, *M*, *I*) consisting of a set of *objects*, *G*, a set of object *attributes*, *M*, and a relation *I* between *G* and *M*. A *formal concept* (*A*,*B*) is a set of objects, $A \subseteq G$, and a set of attributes, $B \subseteq M$, such that all objects in *A* have all the attributes in *B*, and there are no other objects in *A* that have all attributes in *B* and no other attributes in *B* that are shared by all objects in *A*. More formally, if a pair of closure operators are defined by:

$$A^{\uparrow} = \{m \in M \mid \forall g \in \mathrm{A} \rightarrow (g, m) \in I\}$$
$$B^{\uparrow} = \{g \in G \mid \forall m \in B \rightarrow (g, m) \in I\}$$

Then a formal concept, (*A*, *B*), is defined as $A = B^{\uparrow}$ and $B = A^{\uparrow}$. *B* is called the *intent* of the concept and *A* is called the *extent*. A *concept lattice* visualisation (also known as a *Hasse diagram*) arises from the hierarchical relation between concepts (Priss, 2008): when an attribute is added to intent the corresponding extent becomes smaller (more specialised) and *vice-versa*.

Figure 3 shows a concept lattice derived from the EMAGE data in the CUBIST triple store. Each node is a concept, where the intent contains all attributes at and above the node (genes in this case) and the extent contains all the objects at and below the node. For example, taking the third node down on the left side of the lattice, the intent is {*Crkl*, *Shh*, *Igdcc3*, *Ndufb11*} and the extent is {"*>=11 times strongly det. in TS17*", "*>=10 times strongly det. in TS17*", "*>=8 times strongly det. in TS17*", "*>=7 times strongly det. in TS17*", "*>=6 times strongly det. in TS17*", "*>=5 times strongly det. in TS17*"}. The concept could be called "The genes strongly detected 11 times or more in Theiler Stage 17".

Because FCA is predicated on binary relations (an object either has an attribute or not), traditional, many valued, data have to be converted in CUBIST to this Boolean form, a process in FCA known as *scaling* (Wolff, 1993)(S. Andrews & Orphanides, 2010). A data item with n possible values becomes n formal attributes. An original attribute for *Gender*, for example, with the possible values *Male* and *Female*, is scaled in FCA as two formal attributes, *Gender-Male* and *Gender-Female*. Each gene in Figure 3, for example, is a formal attribute. A continuous, numerical, item can be discretised by using numerical ranges. In CUBIST, an existing tool called *FcaBedrock* (S. Andrews & Orphanides, 2010) was developed, adapted and integrated into the system for this purpose.

Applying FCA to data sets is an intensive process, computationally speaking. Computing formal concepts is an exponential task, with often hundreds of thousands, if not millions, of concepts being present in a data set. To deal with this, an existing high-performance concept-mining algorithm called In-Close (S. Andrews, 2009) was implemented in the CUBIST system. When the concept miner is combined with a query (that produces a sub-set of the data), CUBIST is an efficient system with fast response times.

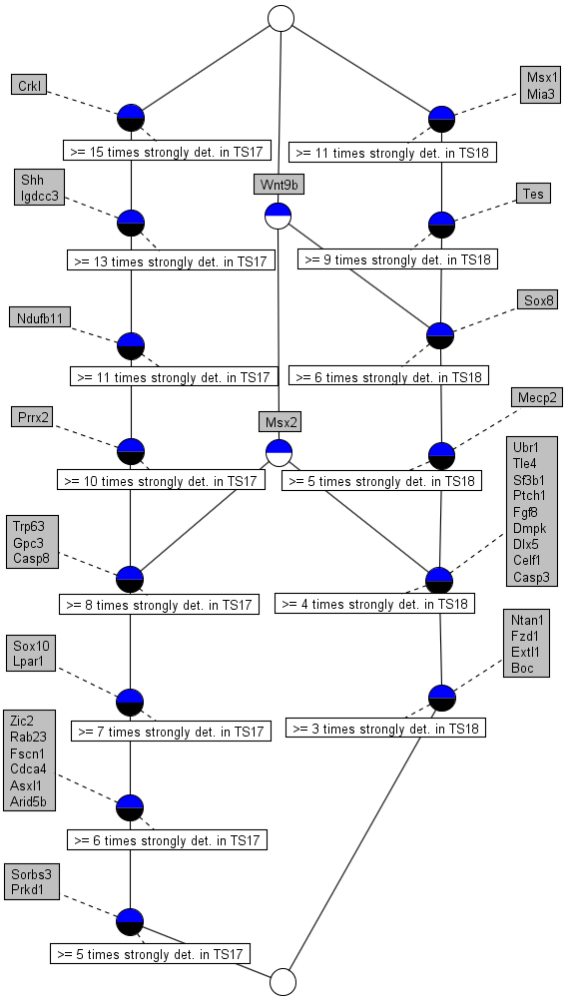

Figure 3. Concept lattice of genes strongly detected in Theiler Stages 17 and 18. (F. Dau, 2013)

RESULTS AND USE-CASE EVALUATION

The formal evaluation of CUBIST is on-going at the time of writing, so presented here are a number of CUBIST results with some initial feedback from the end-users.

Recruitment Consultancy

This data base contains a mixture of structured and unstructured data: Structured data such as Job Title, Salary, Posting Date, Job reference, and unstructured data present within the job description, such as skills or experience requirements. An ontology has been developed for the information and the data has undergone ETL into the CUBIST Semantic Data Warehouse.

Data browsing in the CUBIST system enables the user to navigate through the use case's data set using the explicit links within the ontology. If the user starts with the vacancies, they can view all details for each vacancy, move for a selected vacancy to its advertiser (which shows other vacancies of that advertiser) and move from there to the contact details of the advertiser.

Graph exploration is similar to data browsing, but this time the user navigates through the data visually. This makes the RDF graph explicit – see Figure 4.

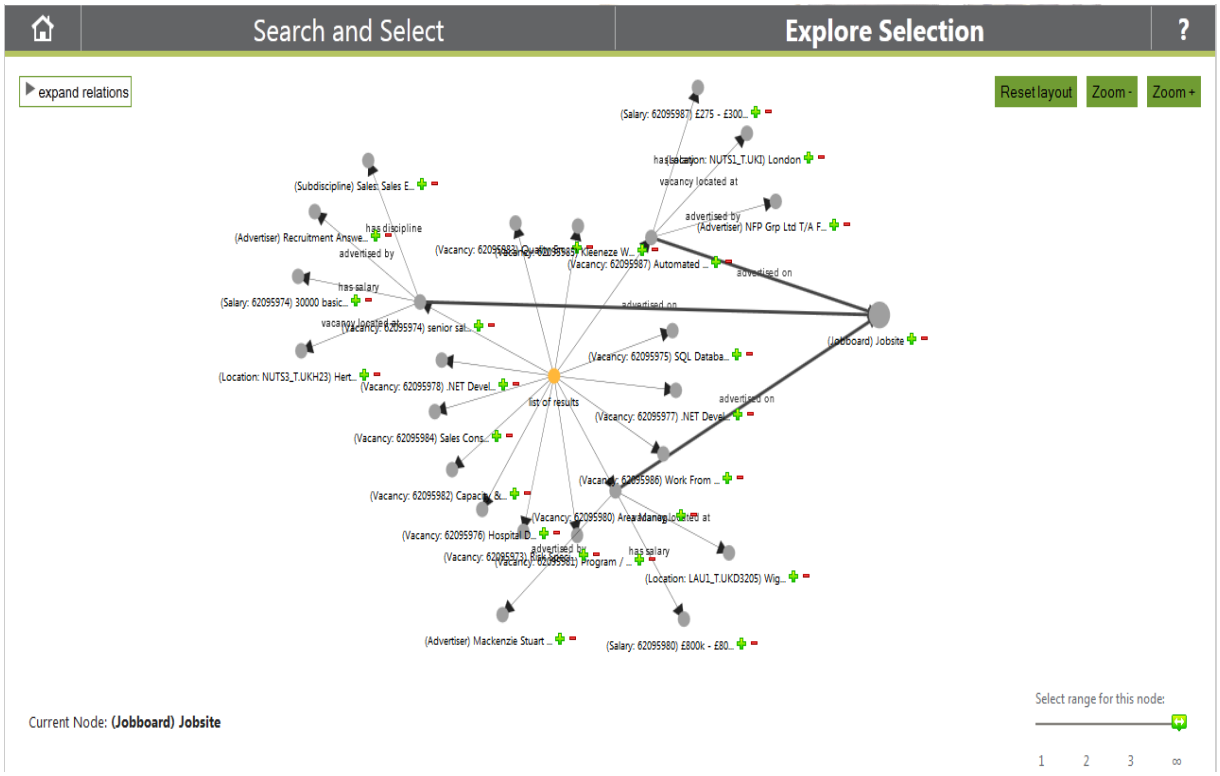

Figure 4. Browsing vacancies with the graph exploration view.

Multiple facets represent classes of data in the repository, such as vacancies or advertisers. Facets can be used to enable quick data filtering, which is useful for both navigating the data set and creating queries for analysis. The use-case has eight different classes of data, thus eight facets. Moreover, queries can be developed for different types in a class, such as of types of vacancies (seeking developers for four different programming languages, for example).

Visual analytics, achieved via conceptual scaling, are provided in CUBIST by a graphical work-bench called CUBIX (Melo, C., Mikheev, A., Le-Grand, B., Aufaure,M.-A., 2012). It offers a range of visualisations including Hasse, Sunburst, Icicle, Bar Charts and Sankey. For each selected attribute in a query, it is also possible to drill down in the data via filtering out values of that attribute. For example, it is possible to select two subsets of attribute-values and show with a bar-chart for each pair of values (one value per attribute set) how many objects have both values. In Figure 5, all job-titles are chosen as one set and the minimum salaries as the other set, which leads to a stacked bar-chart, showing for each job-type the distribution of salaries. The analysis suggests, for example, that C++ developers are least searched for, but best paid.

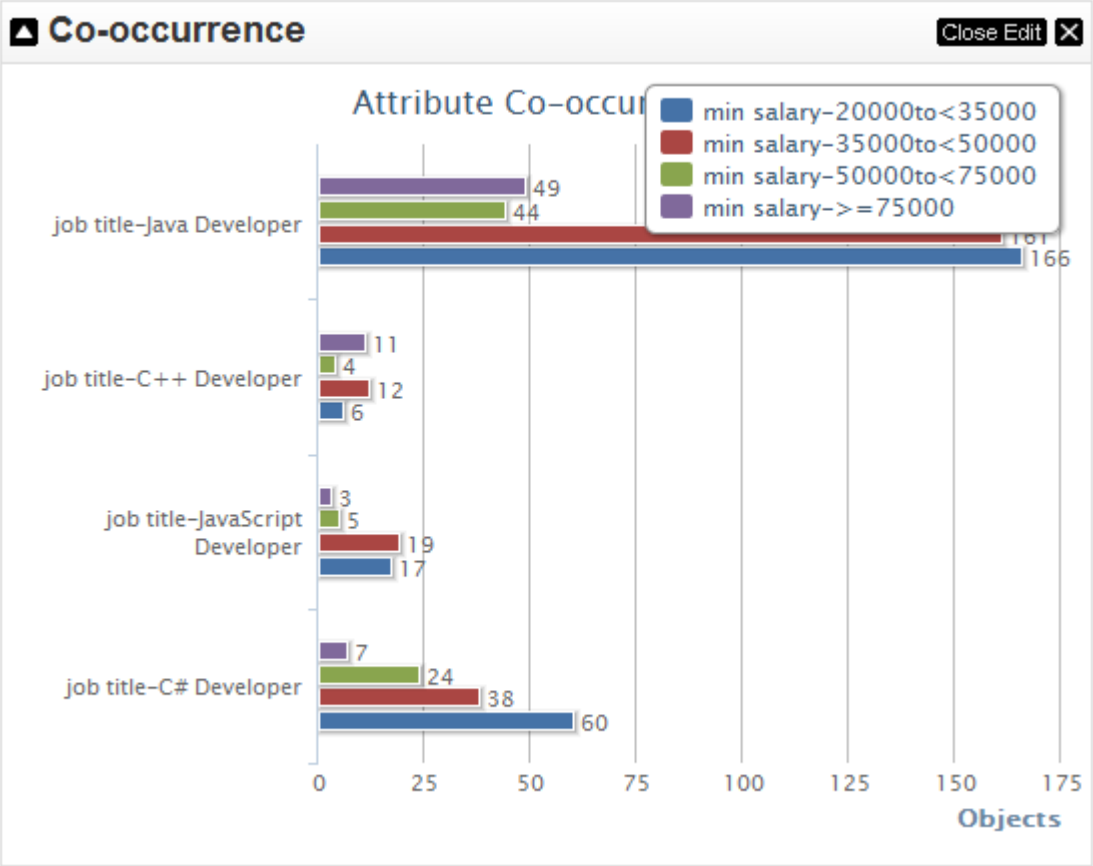

Figure 5. Co-occurrence of Job Title and Minimum Salaries.

Space Industry

The use case is based on a space mission of a science observatory on the Columbus Laboratory, which is part of the International Space Station. The data is from space science instruments providing detailed measurements of the Sun's spectral irradiance. The use-case was particularly interested with respect to CUBIST capabilities for flexible data retrieval and fast data analysis. A single ontology was developed, reflecting the set of parameters and their ranges for the telemetry data resulting in about 2.5 million entities in the triple store, each represented by about 200 different parameters.

Data browsing enables the user to select the relevant parameters very easily, thanks to its AutoComplete functionality, enabling the user to pick a few parameters out of more than 200. For example, if the user starts with the word 'time', she can view every parameter related to time.

Scaling is a critical aspect for this use-case, before proceeding with visual analytics. In order to render the next step of analysis and visualization, the continuous data are scaled in CUBIST to create a discrete form, grouped into bins, so that they become manageable and understandable as visualisations.

Visual analytics provided by CUBIX offered a range of visualisations useful to the space industry use-case. Being able to combine Hasse and co-occurrence matrix visualisations (Figure 6) has provided a particularly valuable tool for error analysis by the expert operators.

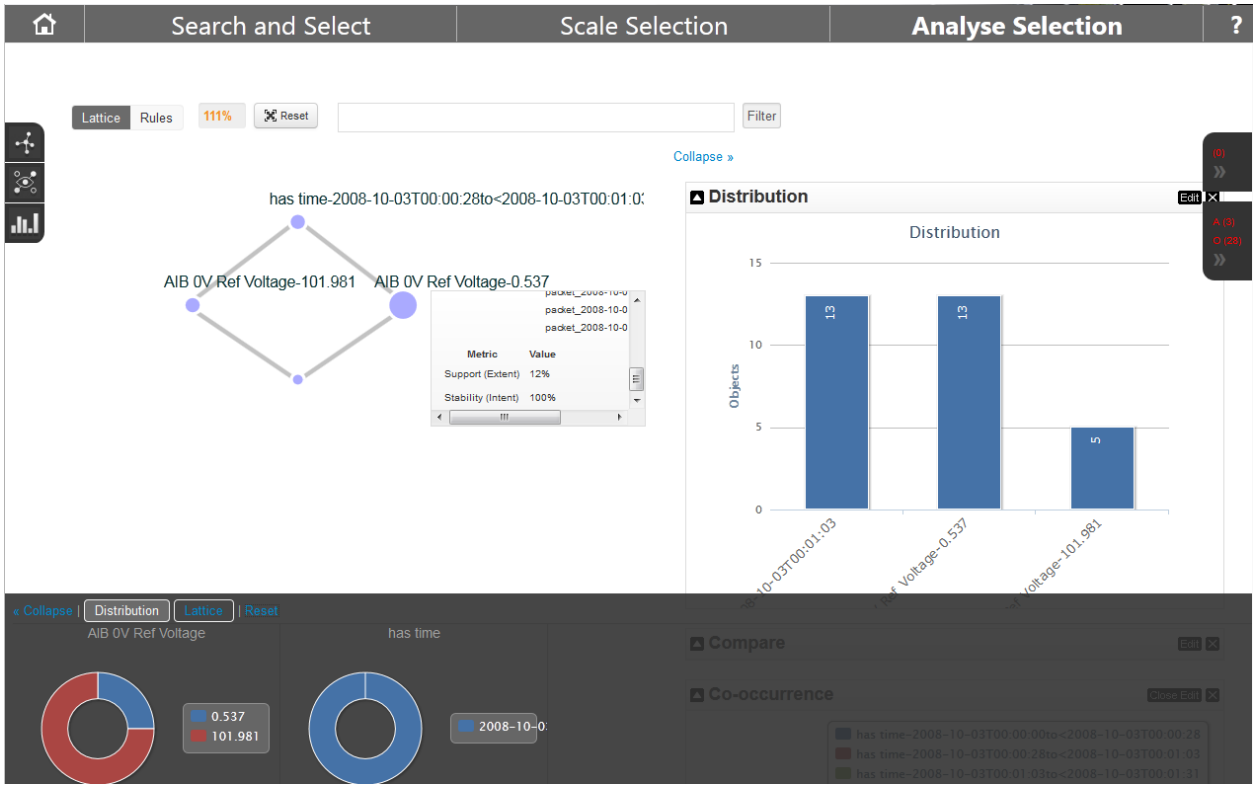

Figure 6. Hasse and Co-Occurrence Visualisations of Satellite Telemetry Data.

Computational Biology

An ontology has been developed (see Figure 2) of the gene expression information from the EMAGE database, and a subset of this data, covering the textual annotations (tissues), has been provided to CUBIST and federated into the Semantic Data Warehouse.

Data browsing in CUBIST enables the user to navigate through the EMAGE data set using the explicit links within the ontology. For example, if the user starts with the gene Bmp4, they can view every textual annotation for that gene. One of those annotations features the tissue Mesoderm TS11, thus they can navigate to that tissue and view it. If they desire they can then view all of its textual annotations.

Graph exploration is similar to data browsing, but this time the user navigates through the data visually. This makes the RDF graph explicit, allowing the user to visually navigate from entity to entity in the ontology, at each stage listing the instances associated with a selected value.

Multiple facets: A facet represents a class of data in the repository, such as genes, strength of expression, textual annotation and Theiler Stage. Facets can be used to enable quick data filtering, which is useful for both navigating the data set and creating queries for analysis.

Visual analytics in CUBIST include extra mechanisms in the CUBIX tool are provided in order to give alternative data exploration viewpoints. For example, the Attribute Implication view provides a simple way of determining the relationships within the data. It is clear from Figure 7, for example, that the gene *Bmp4* is expressed in most tissues. However, both genes are expressed in "apical ectodermal ridge TS17" and "mesenchyme TS17" (this is an example of gene co-expression, a kind of information crucial for this use case and not possible to detect with traditional BI-tools)

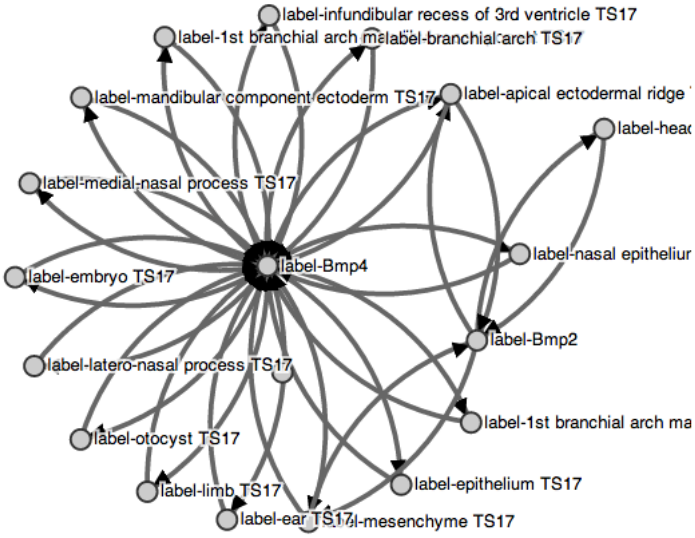

Figure 7. Attribute Implication for Genes Bmp2 and Bmp4 in TS17

The "is part of" relation between tissues in the ontology allows gene expression results from an experiment to be propagated. If a gene is detected in a digit it is thus in the paw, arm and body (positive propagation). If a gene is not detected in the body it is not in any part of the body (negative propagation). Inference rules in RDF allow these 'extra' data to be computed. In CUBIX a Sunburst diagram captures this with colour coding (see Figure 8) [REF]. The coloured nodes in the sunburst indicate where at least one of the genes is expressed. If the mouse is moved over one of the coloured nodes the box in the top right corner is updated to show the tissue name and the list of genes expressed there.

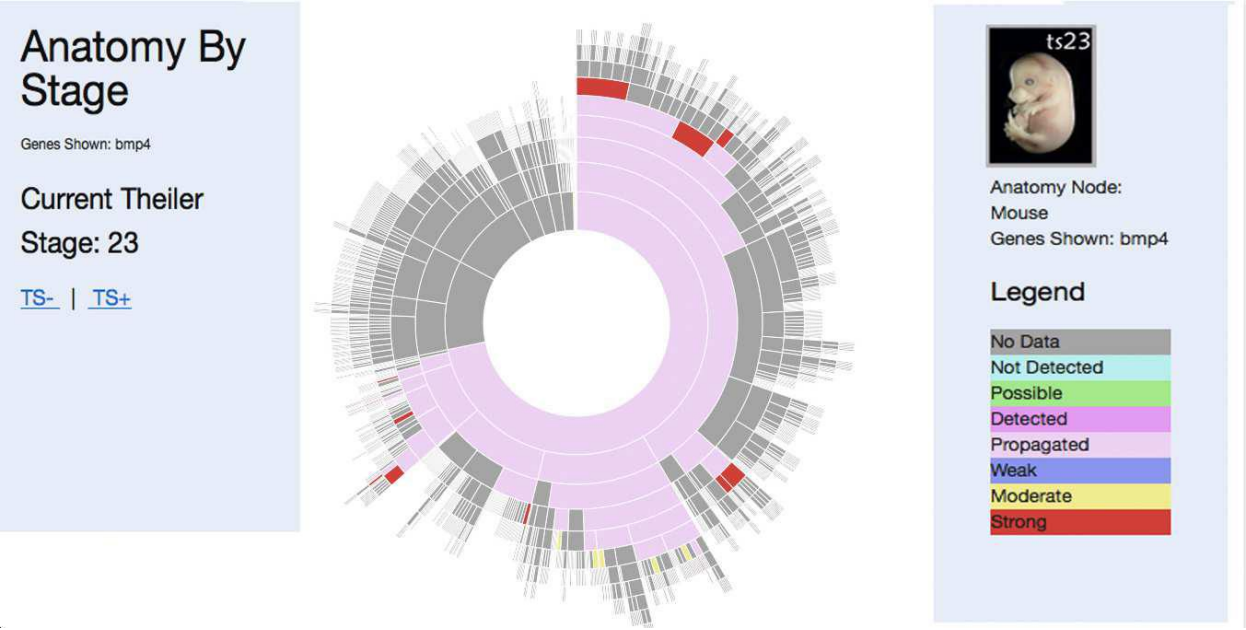

Figure 8. Expression profile of Gene Bmp4 in TS23 (Taylor, A., McLeod, K., Burger,A., 2013).

FURTHER WORK AND EXPLOITATION

Results from CUBIST are to be developed and exploited in a number of ways by the technological and academic project partners. New and existing projects will use, apply and adapt CUBIST analytics and tools. The use-case partners also have plans to employ and exploit CUBIST in their businesses and research areas. A number of key (but by no means exhaustive) developments are listed below.

It is planned to take over some of the CUBIST-results into a project called *ProteomicsDB*, a large scale study of proteins using semantic technology. The overall goal, context and even the used data structures of this project are closely related to the biological use-case in CUBIST (e.g. both for genes and proteins, biologists investigate where and how strongly they are expressed –in tissues for genes and in cells for proteins). It is currently being decided which CUBIST results are the most important ones and how the corresponding software components of CUBIST have to be adapted in order to reuse them in ProteomicsDB.

The 'fault tolerance' approach to FCA, where concepts are approximated by adding 'missing' relations (Pensa & Boulicaut, 2005)will be developed further. The potential is not only for approximation of concepts to manage scale and complexity of results, but also in predictive analysis. As shown in CUBIST work (S. Andrews & McLeod, 2011), it is possible to predict the results of experiments where there are a small number of relations missing from an otherwise compete concept. If results can be predicted with a high degree of confidence then substantial savings of time and money are possible. The opportunity to develop this approach further for large scale data analytics will be taken in the European ATHENA project (Andrews, S., Yates, S., Akhgar, B., Fortune,D., 2013). In ATHENA, Formal Concept Analysis will play an important role in the analysis of social media during crisis situations. It will provide better situational awareness of Member States' Emergency Services. In such situations, where information is likely

to be partial and inaccurate, a fault tolerance approach to situational awareness will facilitate faster and more effective responses from Emergency Services.

The development of the high-performance concept mining algorithm In-Close in CUBIST gives an opportunity to bring new tools to the well-known problem of classification in Machine Learning. The European FP7 ePOOLICE project aims to develop a system to scan the internet to detect Organised Crime (OC). For this a high-performance classification system based on indicators of OC is being developed that will harness In-Close to quickly mine associations and dependencies. Because OC indicators are of several types (categorical, continuous, geospatial, temporal, etc.), the FCA scaling component of CUBIST, FcaBedrock, will be adapted for the ePOOLICE system.

Some of the visualisations created for CUBIST are proving to be useful in other domains, such as Complex System Design. A new method for dealing with continuous attributes in FCA (e.g. temperature) and corresponding visualisations is being proposed. The approach takes advantage of the use of Similarity-based Formal Concept Analysis (SFCA) to classify, visualize, and explore simulation data in order to help system designers to identify relevant design choices. In contrast with traditional FCA which takes as input a binary table of objects and attributes, SFCA uses a similarity measure to group multi-valued attributes in their corresponding concepts. The approach was tested on an aircraft cabin design case study, which concerns the simulation of different configurations of the ventilation system to study the passengers' comfort in the cabin.

Arising from the space industry use-case, a novel distributed approach for mining formal concepts over data streams is being developed to predict anomalies from real-time satellite telemetry data. The proposed platform computes and maintains closed itemsets incrementally and can return the current computation in real time on user's request. It is comprised of several components that carry out the computation of concepts from a basic transaction, to filter and transform data, store data and provide analytic features to visually explore data. Currently, tests are being carried out with different prediction methods based on formal concepts. The idea is to provide a warning interface able to detect anomalies before they happen.

Following interest from users of the biomedical resource (EMAGE), the semantic visualisation of gene expression query results is being further explored. In particular, two existing CUBIX visualisations (sunburst and icicle) have been specialised and deployed within a standalone prototype. The aim of this prototype was to learn how valuable such visualisations were and how they could be implemented for maximum usability and usefulness. Evaluation results from within the EMAGE community were positive. Consequently, this work is being tested with other biomedical resources (for example the GenitoUrinary development resource GUDMAP).

CONCLUSION

Although formal evaluation is on-going, it is clear that the CUBIST project has been successful in its main aim: Combining and Uniting Business Intelligence with Semantic Technologies. Two important semantic technologies: FCA and RDF have been combined with traditional BI in a number of ways; from using concept lattices to visualise numerical data to using traditional BI

visualisation to display the results of RDF queries. Each use-case has been able to exploit these combined approaches to provide new insight into their data and more effective ways to carry out their business. The results of CUBIST show that there is great potential in providing BI-type visual analytics to end-users that use, as their data structure, the richer expressivity of an ontology compared to traditional database structures. A number of important new projects and further work have already been identified and started that will take CUBIST ideas, tools and techniques into new areas, solving new problems and providing new ways for end-users to harness the otherwise hidden semantic in their information systems.

ACKNOWLEDGEMENT

This work is part of the CUBIST project ("Combining and Uniting Business Intelligence with Semantic Technologies"), funded by the European Commission's 7th Framework Programme of ICT, under topic 4.3: Intelligent Information Management.

REFERENCES

Andrews, S., Dau, F. (Ed.). (2012). *2nd CUBIST workshop at ICFCA, leuven, belgium* KULeuven.

Andrews, S., Dau, F. (Ed.). (2013). *3rd CUBIST workshop at ICCS, derby, UK.* CEUR-WS.

Andrews, S., Yates, S., Akhgar, B., Fortune, D. (2013). The ATHENA project: Using formal concept analysis to facilitate the actions of responders in a crisis situation. (pp. 167-180) Elsevier: Butterworth-Heinemann.

Andrews, S. (2009). In-close, a fast algorithm for computing formal concepts. *ICCS 2009,*

Andrews, S. & McLeod, K. (2011). Gene co-expression in mouse embryo tissues.

Andrews, S. & Orphanides, C. (2010). FcaBedrock, a formal context creator. *ICCS 2010,*

Dau, F. (2013). Towards scalingless generation of formal contexts from an ontology in a triple store *International Journal of Conceptual Stuctures and Smart Applications, 1*(1), 18-37. doi:10.4018/ijcssa.2013010102

Dau, F. (Ed.). (2011). *1st CUBIST workshop, at ICCS 2011, derby, UK.* CEUR-WS.

Ganter, B., Stumme, G. & Wille, R. (Eds.). (2005). *Formal concept analysis: Foundations and applications* Springer.

Ganter, B. & Wille, R. (1998). *Formal concept analysis: Mathematical foundations* Springer-Verlag.

Melo, C., Mikheev, A., Le-Grand, B., Aufaure, M.-A. (2012). Cubix: A visual analytics tool for conceptual and semantic data. *Proceedings of the 12th International Conference on Data Mining Workshops,* 894-897. doi:10.1109/ICDMW.2012.41

Pensa, R. G. & Boulicaut, J. (2005). Towards fault-tolerant formal concept analysis. *Advances in Artificial Intelligence, 9th Congress of the Italian Association for Artificial Intelligence,*

Priss, U. (2008). Formal concept analysis in information science. *Annual Review of Information Science and Technology (ASIST), 40*

Richardson, L., Venkataraman, S., Stevenson, P., Yang, Y. a. B.,N., Rao, J., Fisher, M., . . . Christiansen, J. H. (2010). EMAGE mouse embryo spatial gene expression database: 2010 update. *Nucleic Acids Research, 38*(Database issue), D703-D709.

Taylor, A., McLeod, K., Burger, A. (2013). Semantic visualisation of gene expression information. *Proceedings of the 3rd CUBIST Workshop,*

Wille, R. (2005). Formal concept analysis as mathematical theory of concepts and concept hierarchies. In B. Ganter, G. Stumme & R. Wille (Eds.), *Formal concept analysis: Foundations and applications* (pp. 1-33) Springer.

Wolff, K. E. (1993). A first course in formal concept analysis: How to understand line diagrams. *Advances in Statistical Software, 4*, 429-438.