MINING CONSTRUCTION SAFETY DOCUMENTS FOR SAFETY CONCEPT STRUCTURE DISCOVERY USING FORMAL CONCEPT ANALYSIS

*H.H. Wang
National Central University
No.300, Jhongda Rd., Jhongli City
Taoyuan County, Taiwan 320
*(*Corresponding author: hhwang@ncu.edu.tw)*

MINING CONSTRUCTION SAFETY DOCUMENTS FOR SAFETY CONCEPT STRUCTURE DISCOVERY USING FORMAL CONCEPT ANALYSIS

ABSTRACT

Construction safety documents regulate significant safety actions and requirements by which construction workers or employees should abide in order to secure them from occupational hazard events. Therefore, facilitating faster identification of applicable safety requirements from the documents has become an important topic in the construction safety domain. To address the need in this regard, tools and techniques have been developed and research efforts have been made. One of the research efforts is to utilize ontology, a knowledge representation and reasoning approach, as the methodology to achieve the goal of identifying applicable safety requirements. In such ontology-based researches, the development and construction of the safety concept ontology is an essential task as the concepts and the relationships between the concepts both representing key safety knowledge need to be carefully identified. In this paper, the author focuses on the safety documents without predefined concept structure and aims to address the concept and relationship identification difficulties. Specifically, the author leverages the Formal Concept Analysis (FCA) methodology, a theory of data analysis that identifies conceptual structures among data sets, to mine and analyze the documents in order to discover safety concept structure and to further assist the development of the concept ontology for the construction safety documents. The author expects the application of FCA to construction safety domain can eventually benefit the ontology-based approaches for identification of applicable safety requirements.

KEYWORDS

Construction safety, Construction safety document, Ontology, Formal Concept Analysis

INTRODUCTION

Safety is one of the important issues in construction industry, especially in the construction management area. To promote construction safety and prevent occupational hazards, numerous research efforts have been made either to understand/identify the root causes of construction accidents or to analyze construction activities or activity-related factors that may result in potential hazards, such as equipment or space requirements. No matter what kind of safety-promoting approaches are taken, to strictly comply with construction safety requirements is always the most fundamental and core task.

Construction safety requirements, usually regulated in construction safety documents, specify necessary actions that construction workers or employees must take in order to secure them from occupational hazard events. Benner (1983) pointed out that the most explicit technique to achieve construction safety is to abide by safety requirements, which aim to improve safety by clearly regulating what an employer or employee should or should not do without requiring personal judgment. Therefore, facilitating automatic or semi-automatic identification of applicable safety requirements from the safety documents to raise workers' awareness of safety, which can help achieve construction safety, has become an important topic in the construction safety domain.

To address the need of identifying applicable safety requirements, various tools and techniques have been developed and research efforts have been made. For example, Wang and Boukamp (2011) utilized ontology as the methodology to identify applicable safety requirements based on the available contextual information of activity, job steps and/or potential hazards. In addition, some safety specification databases, such as the Canadian enviroOSH Legislation plus Standards (Canadian Centre for Occupational Health and Safety 2012), exist and can be used for searching safety requirements based on keywords.

In the aforementioned ontology-based research, the fundamental task is to develop a safety ontology. Ontology is defined as “an explicit and formal specification of a conceptualization” by Gruber (1993). An ontology can model a set of concepts and relationships among these concepts within a knowledge domain. The concept mentioned herein can be defined as a unique term abstracting a domain phenomenon. In the context of construction safety documents, terms abstracting jobs performed, components built, actions taken or resources used are safety concepts, which together with their relationships form safety concept knowledge. Therefore, a safety ontology in the construction safety domain aims to model the safety concept knowledge in order to leverage this knowledge for later reasoning work.

To develop a safety ontology, safety concepts and their relationships need to be carefully identified from construction safety documents. That is, terms which are used in the safety documents and fit in with the definition of concept become the sources of concepts and should be extracted. If safety concepts have already been systematically represented in the documents, identifying the concepts and their relationships would be straightforward. For example, job hazard analysis documents explicitly structure safety concepts in three categories: activity, job step and potential hazard (Roughton and Crutchfield, 2007). The relationship, for example, between an activity and a job step concepts can be defined as *hasStep* (Wang and Boukamp, 2011). If the documents, however, do not have a predefined structure for the concepts, it would be time-consuming and cumbersome to identify the concepts and their relationships, further making the construction of safety concept ontology much more difficult.

In this paper, the author aims to address the difficulty of identifying relationships between concepts, or concept structure more specifically, from construction safety documents without predefined concept structure. The safety document used for analysis and discussion is the Occupational Safety and Health (OSH) Standards for the Construction Industry, 29 CFR 1926, published by Occupational Safety and Health Administration (OSHA) (U.S. Department of Labor, 2013a). The author focuses on the OSHA standard instead of others such as state regulations for two main reasons. First, OSHA is responsible for handling the enforcement and administration of OSH standards in states under federal jurisdiction; hence, OSHA standards can be viewed as the baseline for safety requirements that other organizations have to abide by. Second, safety practitioners recommend referring to OSHA standards when performing job hazard analysis to further facilitate construction safety planning (Wang and Boukamp, 2011). In addition, Formal Concept Analysis (FCA), which is introduced in the following section, is adopted to mine and analyze the 29 CFR 1926 Standard in order to discover potential safety concept structure and to facilitate the development of the safety concept ontology. The reason why FCA is leveraged is because it is an approach suitable for exploring implicit structure and relationships among different explicit terms of a document, which is the issue encountered when analyzing safety documents without predefined concept structure.

This paper first provides a brief introduction to FCA. Then, an overview of the 29 CFR 1926 Standard is presented to demonstrate what information from the Standard is needed for the analysis. Following that is the application of FCA to the proposed research problem. The final section elaborates on the findings of this study and future research directions.

FORMAL CONCEPT ANALYSIS

Formal Concept Analysis (FCA) was developed and introduced by Rudolf Wille in Germany (Wille, 1982). Invented as a mathematical theory, FCA is a method for data analysis, knowledge representation and information management (Priss, 2007). FCA has been applied in many different disciplines, such as medicine, psychology, sociology, biology, library and information science, linguistics, software engineering and industrial engineering (Wolff, 1993). To the author’s knowledge, FCA was rarely applied in the construction engineering and management domain.

FCA can be understood from three aspects: formal context, formal concept, and concept lattice and line diagram, which are introduced respectively in the following sections. It should be noted that the notion of the "concept" described in FCA is not the same as that described in the ontology development process discussed in the previous section.

Formal context

FCA starts with a cross table, including a set of formal objects (denoted by *G*, represented in rows at the heading column of the cross table); a set of formal attributes (denoted by *M*, represented in columns at the heading row of the cross table) and relations between the objects and attributes (denoted by $I \subseteq G \times M$, represented as crosses in the cross table mapping to an object-attribute pair). Therefore, the relation that the object *g* has to the attribute *m* is represented as *gIm*. Figure 1 depicts a cross table example of different types of building. In this example, building types (RC, SRC, SC and SS) act as formal objects whereas materials (concrete, reinforcing bar and steel) attributes. The crosses indicate the relations between the objects and attributes (whether a building type is composed of the materials).

	Concrete	Reinforcing bar	Steel
RC Building	X	X	
SRC Building	X	X	X
SC Building	X		X
SS Building			X

Figure 1 – Cross table of the different building type example

The sets of formal objects, formal attributes and their relations form a formal context, denoted by the three sets (*G*, *M*, *I*); that is, a cross table stands for a formal context. It is noted that the use of "object" and "attribute" is just indicative as different applications may choose object-like items and their features as formal objects and attributes respectively (Priss, 2007). For example, scientific papers are related to a number of terms. Hence, these papers can be viewed as objects and terms attributes (Poelmans et al., 2010).

Formal concept

If a set of formal objects have a set of formal attributes in common which are not shared by the other formal objects, then the pair of this set of formal objects and formal attributes is called a formal concept. For example, the objects "RC Building" and "SRC Building" in Figure 1 both have the attributes "Concrete" and "Reinforcing bar" and the other two objects do not share the same attributes. Thus, the "RC Building, SRC Building" (objects) and "Concrete, Reinforcing bar" (attributes) form a formal concept. The rest of the formal concepts of the example in Figure 1 are (represented in the form of "objects" and "attributes"):

- "RC Building, SRC Building, SC Building" and "Concrete"
- "SRC Building, SC Building, SS Building" and "Steel"
- "SRC Building, SC Building" and "Concrete, Steel"
- "SRC Building" and "Concrete, Reinforcing bar, Steel"
- "RC Building , SRC Building, SC Building, SS Building" and "" (empty intent)

The set of formal objects of a formal concept is called its "extension" whereas the set of formal attributes its "intension". The extension covers all the objects belonging to this formal concept and the intension consists of all attributes valid for all these objects (Poelmans et al., 2009). The formal concepts, their extensions and intensions of a given formal context are uniquely defined and fixed (Priss, 2007).

Concept lattice and line diagram

There is a natural hierarchical order between the concepts of a given context that is called the subconcept-superconcept relation. This relation is defined as follows: the concept (A_1, B_1) is a subconcept of the concept (A_2, B_2), denoted by $(A_1, B_1) \leq (A_2, B_2)$, if and only if $A_1 \subseteq A_2$ or $B_2 \subseteq B_1$, where A_1 and A_2 are sets of formal objects, and B_1 and B_2 are sets of formal attributes valid for A_1 and A_2 respectively (Wille, 1992). In other words, a concept is a subconcept of another concept if and only if its extension is contained in the extension of the other concept. For example, two formal concepts of Figure 1 are (1) "RC Building, SRC Building" and "Concrete, Reinforcing bar" and (2) "RC Building, SRC Building, SC Building" and "Concrete". Since (1)'s objects "RC Building, SRC Building" is a subset of (2)'s objects "RC Building, SRC Building, SC Building", it can be concluded that (2) is a superconcept of (1) and that (2)'s attribute "Concrete" is a subset of (1)'s attributes "Concrete, Reinforcing bar".

The set of all concepts of a formal context together with the subconcept-superconcept relations between these concepts constitutes a complete lattice, called the concept lattice of the formal context. A concept lattice can be represented visually through a line diagram as shown in Figure 2 for the building type example. The nodes in the line diagram represent formal concepts; formal objects are labeled below the nodes while formal attributes above. To read the information contained in a line diagram, the following rule applies: an object *g* has an attribute *m* if and only if there is a path leading upwards from the node named by "*g*" to the node named by "*m*". Then the extension and intension of a formal concept (node) can be readily obtained from a line diagram by respectively collecting all objects below and all attributes above the node of the given concept (Wolff, 1993). Using Figure 2 as an example, for the rightmost concept labeled by the attribute "Reinforcing bar" and object "RC Building", one can conclude that the object "RC Building" has two attributes "Reinforcing bar" and "Concrete" (by following the upwards leading path). One can also conclude that the same concept has extension (objects) "RC Building" and "SRC Building" and intension (attributes) "Reinforcing bar" and "Concrete".

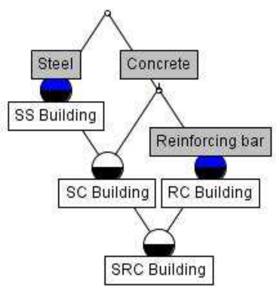


Figure 2 – Line diagram (concept lattice) of the building type example

Implications

The subconcept-superconcept relation suggests that, given two formal concepts (A_1, B_1) and (A_2, B_2), $A_1 \subseteq A_2$ or $B_2 \subseteq B_1$ if and only if $(A_1, B_1) \leq (A_2, B_2)$. If $(A_1, B_1) \leq (A_2, B_2)$ holds, the notion described by the attributes of B_1 can be viewed as a specialization of that of B_2 as it contains more attributes than B_2 does (i.e. B_2 is a subset of B_1). Therefore, the relation implies a potential hierarchical structure among attributes: the attributes of B_2 may be super-attribute of those of B_1. Furthermore, if A_1 and A_2 are partially subsumed by each other (i.e. A_1 and A_2 share just some objects), the attributes of B_1 would be associated with those of B_2. If, on the other hand, A_1 and A_2 do not have objects in common, then the attributes of B_1 *and* B_2 would be deemed independent with each other. When representing formal attributes as concepts in an ontology, the subconcept-superconcept relations implies subclass-superclass relationships between attributes of B_1 and B_2 whereas the partial subsumption and independence relations respectively suggest the use and non-use of association relationships. Therefore, the relation implications of FCA are significant to the ontology development as they provide a way for ontology developers to understand the ontological concept structure and relationships between the concepts.

OVERVIEW OF THE 29 CFR 1926 STANDARD

Safety and health standards of Occupational Safety and Health Administration (OSHA) are contained in Title 29 of the Code of Federal Regulations (29 CFR). Several different parts are in the OSHA

CFR, specifying standards applicable to various types of work environments and workspaces. Standards for construction industry are regulated in Part 1926 (U.S. Department of Labor, 2013a). Figure 3 depicts a snippet of the Subpart M of 29 CFR 1926 Standard. As can be seen from Figure 3, an ordered hierarchical structure exists in 29 CFR 1926: the Part contains several subparts; a subpart consists of several standards; and a standard is composed of several sections. Each section can have a three-level, nesting structure: the first-level section consists of a description and, if any, a second-level section; the second-level section also consists of a description and, if any, a third-level section; and the third-level one only has a description. The first-level section is numbered by the Standard number attached with an lowercase English letter (e.g. 1926.502(d)) while the second- and third-level sections are numbered by the number of its upper level section plus an Arabic number (e.g. 1926.502(d)(6)) and a lowercase Roman number (e.g. 1926.502(d)(6)(i)), respectively.

Part 1926
Safety and Health Regulations for Construction
……
Subpart M
Fall Protection
……
Standard 1926.502
Fall protection systems criteria and practices
……
1926.502(d)
"Personal fall arrest systems." Personal fall arrest systems and their use shall comply with the provisions set forth below. Effective January 1, 1998, body belts are not acceptable as part of a personal fall arrest system. Note: The use of a body belt in a positioning device system is acceptable and is regulated under paragraph (e) of this section.
……
1926.502(d)(6)
Unless the snaphook is a locking type and designed for the following connections, snaphooks shall not be engaged:
1926.502(d)(6)(i)
directly to webbing, rope or wire rope;
1926.502(d)(6)(ii)
to each other;
1926.502(d)(6)(iii)
to a dee-ring to which another snaphook or other connector is attached;
1926.502(d)(6)(iv)
to a horizontal lifeline; or
1926.502(d)(6)(v)
to any object which is incompatibly shaped or dimensioned in relation to the snaphook such that unintentional disengagement could occur by the connected object being able to depress the snaphook keeper and release itself.

Figure 3 – A snippet of OSHA's 29 CFR 1926 Subpart M Standard

To analyze the 29 CFR 1926 Standard, this study focuses on the section level and determines a unit for analysis according to the following principles in this study:

(1) If the description of a section contains multiple conceptual terms, then the description itself will be regarded as a unit for analysis regardless of whether the section has sub-sections. For example, the section 1926.502(d) in Figure 3 describes conceptual terms, such as "personal fall arrest system", "body belt" and "positioning device system"; hence, this description alone represents a unit for analysis.
(2) If the description of a section does not contain conceptual terms and the section has sub-sections, then the sub-sections are evaluated using the previous principle.
(3) If the description of a section contains neither conceptual terms nor sub-sections, then this section is skipped and the next parallel section is evaluated using the first principle.
(4) Only first- and second-level sections are entered into evaluation using the above three principles.

Due to the huge amount of standards in 29 CFR 1926, this study focuses on only two standards, namely 1926.501 and 1926.502, for the convenience of demonstration. The two standards are chosen for they are two of the most frequently cited standards by OSHA during the period October 2011 through September 2012 (U.S. Department of Labor, 2013b). By deploying these principles, the chosen standards are organized and respectively labeled, and ready to be analyzed using FCA. For example, the description of 1926.502(d) is labeled as "2d" while that of 1926.502(d)(6) is labeled as "2d6". Totally 119 sections are identified and labeled for the following FCA use.

APPLICATION OF FCA TO ANALYZING 29 CFR 1926

Preparation

To apply FCA to the chosen standards, this study views the sections of the standards as "formal objects" and the conceptual terms which describe the applicability of the sections as "formal attributes". The conceptual terms are either referred to from the glossary of 29 CFR 1926 (i.e. the Definitions section in 1926.500(b)) or determined from the text of the standards through manual identification. There are 119 formal objects (sections) and 68 formal attributes (conceptual terms) in total used for FCA. In addition, a FCA software, Concept Explorer (ConExp) (Yevtushenko, 2000), is used to develop the formal context and concept lattices/line diagrams in this study.

Cross table and formal context

Figure 4 shows a partial cross table for the formal context of the 29 CFR 1926 example, where the labels for different sections are in the heading column and conceptual terms are in the heading row, respectively representing formal objects and formal attributes. 213 formal concepts in total are obtained from the formal context and in view of such many formal concepts, this study selects some of them for discussion as follows.

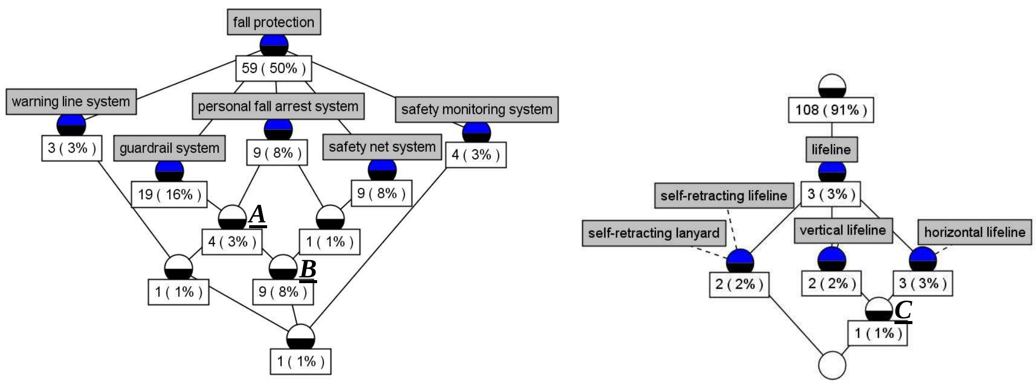
	anchorage	body belt	body harne...	connector	controlled access zone	dangerous equipment	failure	free fall	guardrail system	hole	lifeline
1b1									X		
1b2					X				X		
1b3									X		
1b4									X	X	
1b5											
1b6									X		
1b7									X		
1b8						X			X		
1b9					X				X		
1b10									X		
1b11									X		
1b12									X		
1b13									X		
1b14									X		
1c											
1c1									X		
2b2									X		
2b3							X		X		
2b4									X		
2b5							X				
2b7											
2b10									X		
2b11									X	X	
2b12									X	X	
2b13	X								X	X	
2b14									X		

Figure 4 – Partial cross table of the 29 CFR 1926 example

Demonstration using concept lattices and line diagrams

Figure 5(a) shows a concept lattice of 11 formal concepts, and six formal attributes are included in the concept lattice. Each node has a box below that contains the number of sections belonging to the formal concept and the percentage for which the number of sections accounts in the formal context. Below the topmost formal concept labeled by formal attribute "fall protection" are formal concepts labeled by the rest five attributes "warning line system", "guardrail system", "personal fall arrest system", "safety net system" and "safety monitoring system". That is, the five attributes can be regarded as sub-attributes (or sub-concepts in an ontological term) of the attribute "fall protection". In addition, the five attributes can exist in 29 CFR 1926 sections collectively. For example, formal concept A that owns four sections contains both formal attributes "guardrail system" and "personal fall arrest system" (as shown on the node A below the two nodes with "guardrail system" and "personal fall arrest system" labels) while formal concept B contains formal attributes "guardrail system", "personal fall arrest system" and "safety net system". This indicates that there are some mutual formal objects in formal concept A's two super-concepts, which implies an association relationship exists between formal attributes "guardrail system" and "personal fall arrest system" (e.g. "guardrail system" can be substituted for "personal fall arrest system").

Figure 5(b) shows another concept lattice formed by seven formal concepts, in which five formal attributes are included: "lifeline", "vertical lifeline", "horizontal lifeline", "self-retracting lifeline" and "self-retracting lanyard". One can tell from Figure 5(b) that "vertical lifeline", "horizontal lifeline" and “self-retracting lifeline” can be referred to as sub-attributes of “lifeline”. Although the formal concept having attribute “self-retracting lanyard” is a sub-concept of the formal concept having attribute “lifeline”, self-discretion is required here to decide whether “self-retracting lanyard” should be a sub-attribute of “lifeline” based on the context. Furthermore, one formal concept contains “vertical lifeline” and “horizontal lifeline” (as node C shows) while no formal concepts have “‘vertical lifeline’, ‘horizontal lifeline’” and “‘self-retracting lifeline, ‘self-retracting lanyard’” in common (i.e. the bottommost node is an empty node, indicating no section shares all the formal attributes of the concept lattice). This suggests that the former two formal attributes are independent from the latter two in the selected 29 CFR 1926 standards; when developing an ontology, one has no need to consider association relationships between them. Last, the leftmost node contains both “self-retracting lifeline” and “self-retracting lanyard”, which denotes a tight relation between them and hence, an association relationship is required to represent such a relation.



(a) (b)

Figure 5 – Selected formal concepts of the 29 CFR 1926 example

DISCUSSION

Using FCA provides an alternative way for supporting ontology modellers to determine significant relationships between conceptual terms and hence, the ontology development process can be facilitated. There is a limitation observed during the application: user’s self-discretion of categorizing formal attributes is needed in some circumstance to ascertain the correct representation of these attributes, as demonstrated in the previous section. In other words, the relationships identified through applying FCA might not perfectly represent the actual relationships between conceptual terms in some cases. Nevertheless, this limitation does not cause any harm since modellers still have to review all the possible identified relationships and conceptual terms which are incorrectly represented can then be ignored.

CONCLUSIONS

This study is one of the first attempts to apply FCA in the research domain of architectural, engineering and construction industry, especially to help the ontology development. This study is part of an ongoing research that focuses on using ontology to facilitate the identification of applicable safety requirements from a safety specification. This study tests the applicability of FCA on a small scale while the abovementioned limitation of applying FCA exists. The author expects to deploy FCA in a broad sense in the ongoing research and also address the limitation. In addition, aiming at facilitating the ontology development process, to incorporate natural language processing techniques into the ongoing research in order for automatically identifying the conceptual terms for FCA would be one of the future research tasks.

ACKNOWLEDGEMENT

The authors would like to thank the National Science Council of Taiwan (Contract No. NSC101-2218-E-008 -017) for financially supporting this research.

REFERENCES

Benner, L. J. (1983). What is this thing called a safety regulation?, *Journal of Safety Research*, 14, 139-143. doi: 10.1016/0022-4375(83)90041-5

Canadian Centre for Occupational Health and Safety. (2012). *Web Information Service: Canadian enviroOSH Legislation plus Standards*. Retrieved from http://www.ccohs.ca/products/legislation/legislation.html.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2), 199–220. doi: 10.1006/knac.1993.1008

Poelmans, J., Elzinga, P., Viaene, S. and Dedene, G. (2009, July). *A case of using formal concept analysis in combination with emergent self-organizing maps for detecting domestic violence*. Proceedings of the 9th Industrial Conference on Advances in Data Mining, Applications and Theoretical Aspects, Miami, Leipzig, Germany.

Poelmans, J., Elzinga, P., Viaene, S. and Dedene, G. (2010, July). *Formal concept analysis in knowledge discovery: a survey*. Proceedings of the 18th International Conference on Conceptual Structures: From Information To Intelligence, Kuching, Sarawak, Malaysia.

Priss, U. (2007). Formal concept analysis in information science. *Annual Review of Information Science and Technology*, 40(1), 521-543. doi: 10.1002/aris.1440400120

Roughton, J. E., and Crutchfield, N. (2007). *Job hazard analysis: A guide for voluntary compliance and beyond*, Butterworth-Heinemann, Oxford, UK.

U.S. Department of Labor. (2013a). "Occupational Safety and Health Standards for the Construction Industry - 29 CFR 1926." Retrieved from Occupational Safety and Health Administration website: http://www.osha.gov/pls/oshaweb/owasrch.search_form?p_doc_type=STANDARDS&p_toc_level=1&p_keyvalue=Construction

U.S. Department of Labor. (2013b). "Frequently cited standards – Major Group 15: Building Construction General Contractors And Operative Builders." Retrieved from Occupational Safety and Health Administration website: http://www.osha.gov/pls/imis/citedstandard.sic?p_sic=15&p_esize=&p_state=FEFederal

Wang, H.H. and Boukamp, F. (2011). An ontology-based representation and reasoning framework for supporting job hazard analysis. *Journal of Computing in Civil Engineering*, 25(6), 442-456. doi: 10.1061/(ASCE)CP.1943-5487.0000125.

Wille, R. (1982). Restructuring lattice theory: an approach based on hierarchies of concepts. In I. Rival (Ed.), *Ordered Sets* (pp. 445-470). Dordrecht-Boston, Reidel.

Wille, R. (1992). Concept lattices and conceptual knowledge systems, *Computers & Mathematics with Applications*, 23(6-9), 493-515. doi: 10.1016/0898-1221(92)90120-7.

Wolff, K.E. (1994). A first course in formal concept analysis: how to understand line diagrams. In. F. Faulbaum (Ed.), *Proceedings SoftStat'93* (pp. 429-438). Gustav Fischer Verlag.

Yevtushenko, S. A. (2000). *System of data analysis "Concept Explorer". (In Russian)*. Proceedings of the 7th national conference on Artificial Intelligence KII-2000, Russia. Retrieved from http://conexp.sourceforge.net/index.html