

# **Knowledge Discovery from Texts with Conceptual Graphs and FCA**

Mikhail Bogatyrev, Kirill Samodurov

Tula State University, Tula, Russia  
okkambo@mail.ru, zmeymc@gmail.com

**Abstract.** Building conceptual lattices from conceptual graphs looks as natural way in Formal Concept Analysis but still is not discovered at length. If conceptual graphs are acquired from natural language texts then they contain specific material for knowledge discovery. Conceptual graphs serve as semantic models of text sentences and the data source for concept lattice. With the use of concept lattice it is possible to extract information which can be treated as facts. Facts can be extracted by using navigation in the lattice and interpretation its concepts and hierarchical links between them. Experimental investigation of this knowledge discovery technique is performed on the annotated textual corpus consisted of descriptions of biotopes of bacteria.

**Keywords:** Knowledge discovery, Conceptual graphs, Formal context, Concept lattice, Bacteria biotopes.

## **1 Introduction**

In the Formal Concept Analysis (FCA) community there is growing interest in the application FCA to textual data. Such interest corresponds to the overall popularity of Text Mining methods due to the prevalence of textual data, especially in the Internet.

There is certain number of works concerned with FCA and Text Mining devoted as to linguistic applications of FCA [1] as to information retrieval with FCA [2, 3].

The actual problem here is the problem of building formal contexts on textual data. If textual data is represented as natural language texts then this problem becomes acute. There are several approaches to the solution of this problem. One, mostly applied variant, is the context in which the objects are text documents and the attributes are the terms from these documents [4]. Another variant is building formal context directly on the texts. On this way, various features of texts have been analyzed and used for constructing formal context. Semantic relations (synonymy, hyponymy, hypernymy) in a set of words are used for semantic matching with FCA in [5], verbobject dependencies from texts are applied in [6] for learning concept hierarchies from text corpora and more general lexico-syntactic features of words are applied in [4].

In addition to the direct use of text for building formal contexts, semantic models of text and textual corpora tagging tools are used. We apply this approach and use conceptual graphs (CGs) for representing semantics of individual sentences of a text.

One of the early mentions of applications of conceptual graphs in FCA can be found in [7]. Modern results concerned with conceptual graphs and FCA are in the work [8].

Although the join of two paradigms of conceptual modeling - conceptual graphs and concept lattices - looks attractive it is still not discovered at length. In this paper we present some results of knowledge discovery which were obtained by using our framework for conceptual modeling on natural language texts [15]. Now due to certain improvements made in the framework it is possible to extract from the texts more information being interpreted as knowledge. Experimental investigation of this knowledge discovery technique is performed by learning of bacteria biotopes [17, 18]. A biotope (also known as habitat) is an area of uniform environmental conditions providing a living place for plants, animals or any living organism.

## 2 FCA and Conceptual Graphs

Briefly recall the main FCA notions and consider some links between concept lattices and conceptual graphs.

### 2.1 Standard Definitions

There are two basic notions FCA deals with: *formal context* and *concept lattice* [9]. Formal context is a triple  $\mathbf{K} = (G, M, I)$ , where  $G$  is a set of objects,  $M$  – set of their attributes,  $I \subseteq G \times M$  – binary relation which represents facts of belonging attributes to objects. The sets  $G$  and  $M$  are partially ordered by relations  $\sqsubseteq$  and  $\sqsubseteq$ , correspondingly:  $G = (G, \sqsubseteq)$ ,  $M = (M, \sqsubseteq)$ . Formal context is represented by  $[0, 1]$  matrix  $\mathbf{K} = \{k_{i,j}\}$  in which units mark correspondence between objects  $g_i \in G$  and attributes  $m_j \in M$ . The concepts in the formal context have been determined by the following way. If for subsets of objects  $A \subseteq G$  and attributes  $B \subseteq M$  there exist mappings (which may be functions also)  $A' : A \rightarrow B$  and  $B' : B \rightarrow A$  with properties of  $A' := \{\exists m \in M | \langle g, m \rangle \in I \forall g \in A\}$  and  $B' := \{\exists g \in G | \langle g, m \rangle \in I \forall m \in B\}$  then the pair  $(A, B)$  that  $A' = B$ ,  $B' = A$  is named as *formal concept*. The sets  $A$  and  $B$  are closed by composition of mappings:  $A'' = A$ ,  $B'' = B$ ;  $A$  and  $B$  are called the *extent* and the *intent* of a formal context  $\mathbf{K} = (G, M, I)$ , respectively.

A conceptual graph is a finite oriented connected bipartite graph [10] which has two different kinds of nodes: *concepts* and *conceptual relations*. Concept nodes may have simple form representing entities and complex form representing entities (named as referents) and their types. A type of entity indicates the class of the element represented by the concept. A referent indicates the specific instance of the class

referred to by the node. For example, the concept  $\langle \text{Human}: \text{John} \rangle$  has complex form where “Human” is type and “John” is referent. Referents may be generic or individual. Relation nodes also have two attributes: *valence* and *type*. Valence indicates the number of the neighbor concepts of the relation, while the type expresses the semantic role of each one.

These two parameters of the CG model – the type of concept and valence of relation – are used in our algorithms of CGs processing. Concept types constitute hierarchically ordered set  $S_t$ , not necessarily a lattice [11]. In the general case the set of relations  $S_r$  is also ordered but relations in CGs acquired from texts represent semantic roles which are not ordered.

Here and henceforth we consider conceptual graphs have been acquired from texts only. Those conceptual graphs become *labeled graphs* when types of concepts are supported. In the previous example “Human” is the label for “John”. In the labeled concept  $(g, l)$  with concept name  $g$  and its label  $l$   $g \in G$  as possible element of the set of objects in the formal context and  $l \in S_t$ . There is the *pattern structure* introduced in [12] for concepts and labels:  $P = (G, (S_t, \wedge), \delta)$ , where  $\delta : G \rightarrow S_t$  is a mapping. In this structure  $S_t$  is a meet-semilattice. It is realized as a thesaurus with hierarchy of its terms. This thesaurus is used for identifying CGs concepts and then for applying them as objects in formal context.

## 2.2 Acquiring and implementing conceptual graphs in FCA

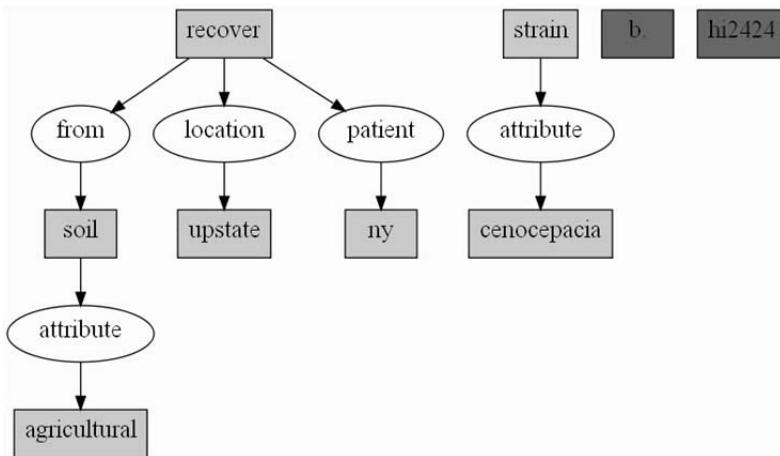
The method of acquiring conceptual graphs from natural language texts is considered in [13]. Some peculiarities of conceptual graphs created with this method are illustrated in [14]. The method has standard phases of lexical, morphological and semantic analysis extended with solution of the problem of Semantic Role Labeling. This problem is non-trivial since semantic roles are not elements in the processed sentence and must be discovered by means of morphological analysis. Solution of the problem of Semantic Role Labeling is essential for building conceptual relations in CGs. As for concepts of CGs, there are several approaches for extracting them from texts. Among these approaches verb-oriented approach [13] has certain advantages. It is based on discovering predicate constructions in the text. Resulting CG has usually central verb as the main concept. A sub graph which has such main concept may be treated as roughly representing semantics of the sentence.

**Filtering sentences.** Except of using standard lexicographic operations on the set of sentences (stemming, part-of-speech tagging, etc.) it is needed to filter sentences according with previously established topics which will be represented in formal context. Filtering means excluding some sentences from the set from which CGs have been acquired. The simplest criterion for filtering is checking existence of key words from the set  $S_t$  in the sentences of processing texts. Filtering is important for the texts having free subject area. Problem-oriented texts usually have high determined topics which makes them free from filtering. But filtering is needed if the number of topics in a text is greater than ones were established for representing in formal contexts.

**Creating formal contexts.** To construct formal context on the set of conceptual graphs it is necessary to select one set of concepts as objects and other set of concepts

as object's attributes. At the first glance, this problem seems simple: those concepts of conceptual graphs which are connected by "attribute" relation have been put into formal context as its objects and attributes. Actually the solution is much more complex.

To illustrate this, consider an example of processing the sentence which is typical in the learning of bacteria biotopes: *B. cenocepacia strain HI2424 was recovered from agricultural soil in upstate NY*. Conceptual graph for this sentence is shown on Fig. 1



**Fig. 1.** Example of conceptual graph with isolated concepts.

The sentence being analyzed is about bacterium named as *Burkholderia cenocepacia*. Its name is used in the text in abbreviated form as *B. cenocepacia* and *HI2424* is the code of its strain. The decision that this sentence is about *Burkholderia cenocepacia* may be found on the stage of analyzing and filtering sentences by learning the algorithm to recognize bacteria names. If the context which has the names of bacteria as an objects is creating, then the sub graph <strain> - (attribute) - <cenocepacia> and two isolated concepts <b.> and <hi2424> do not participate in forming that context.

The verb "recover" is the key word which marks the predicate in conceptual graph to be processed for creating formal context. The main meaningful attribute of *Burkholderia cenocepacia* in the sentence is that it inhabits in the soil, more concrete – in agricultural soil. The key word "soil" must be in the thesaurus containing information about habitats of bacteria for marking the sub graph <soil> - (attribute) - <agricultural> for using in the context. Then although the "attribute" conceptual relation plays significant role in creating formal context it is not always applied in it. Another conceptual relations - "location" and "patient" on Fig. 1 - also belong to the list of important semantic roles. These relations produce possible attributes to formal context ("upstate" and "NY") which are not informative.

Another problem in creating formal contexts is linking in one context objects and attributes from different sentences. Its solution is connected with anaphora resolution and described in [15].

### 3 Learning Bacteria Biotopes with FCA

Bioinformatics is one of the fields where Data Mining and Text Mining applications are growing up rapidly. New term of “Biomedical Natural Language Processing” (BioNLP) has been introduced here. This term is stipulated by huge amount of scientific publications in Bioinformatics and organizing them into textual databases and corpora with access to full texts of articles via such systems as PubMed [16]. There is the innovation devoted to competitive solving BioNLP problems known as BioNLP Shared Task. It started in 2009 and its last issue was in 2016 [22]. One of the BioNLP Shared Tasks is learning of bacteria biotopes (BB-Task).

#### 3.1 Related work

There are several solutions of BB-Task presented mainly in the BioNLP Shared Task workshops; the recent proceedings of such workshops are in [23]. Analyzing them, we can formulate the following general approach to solving the BB-Task.

**BB-Subtasks.** There are three subtasks in which the whole BB-Task is divided.

The first subtask is named as *Bacteria and habitat detection and categorization*. In this subtask biotope entities (names of bacteria) need to be detected in a given biological text and must be mapped onto a given ontology.

The second subtask, *Entity and event extraction* is devoted to event extraction from texts. In this case the term “*event extraction*” corresponds to the Text Mining term “*fact extraction*”. This subtask is focused on the single event “*Lives\_In*” which denotes the fact of living bacteria in certain environment (habitat): water, soil or other organisms.

The third subtask is named as *Knowledge Base extraction*. Here text processing systems are evaluated for their capacity to build a knowledge base from the textual corpus. Actually, as names of bacteria as its relations to the habitat must be detected and enrich given ontology.

All the diversity of BB-Task can be often transformed to two standard problems of Named Entity Recognition (NER) and Relations Extraction (RE) on textual data.

**Information resources.** Textual corpora, databases and ontologies have been applied for storing data in BB-Task. Large ontology of biotopes called OntoBiotope [25] is applied for mapping detected data. From the BioNLP Shared Task 2013 up to now more and more external information systems were used as external program applications in the BB-Task solutions. Among them there are *POS Tagging systems*, *Parsing systems*, *Term Extraction systems*, *Named Entity Recognition systems*. More information about them can be found in [18, 22, 24].

**Methods.** The current trend in solving BB-Task is using methods from Data Mining and Text Mining areas of research. Subtasks of BB-Task is reformulated as the

tasks of data clustering or data classifying for applying appropriate methods of Data Mining. Among these methods there are Support Vector Machine classifier [26], Conditional Random Fields [27] and rule-based and ontology methods from computer linguistics [28].

### 3.2 FCA based solution

Now any new solution of BB-Task may be classified in accordance to considered framework of *BB-Subtasks –Information Resources – Methods* from previous section. Our solution is classified by the following way.

**BB-Subtasks.** Solving the first subtask of BB-Task, we extract the names of bacteria from the textual corpus [17] which contains articles about bacteria. All the texts were preliminary filtered as it is shown in the section 2.2. Extracted names of bacteria are used in formal context and then in concept lattice. Concept lattices serve as known frames of ontologies [9], so the mapping to ontology is presented. Here we solve the NER task and it has direct solution with conceptual graphs. The only problem which is here is anaphora resolution considered in [15].

We formulate the second subtask as Relations Extraction (RE) one. Using conceptual graphs not only “*Lives\_In*” relation but some others may be extracted. We construct three formal contexts of “Entity”, “Areal” and “Pathogenicity”. In the “Areal” context there is “*Lives\_In*” relation linking objects and attributes. In other contexts of “Entity” and “Pathogenicity” the “*Attribute*”, “*Instrument*”, “*Location*”, etc. semantic roles are applied as relations for constructing these contexts – see examples above.

Concept lattices which we create as data storage for our fact extraction system together with the software of this system constitute the basis for constructing knowledge base. So the elements of all three subtasks are presented in the FCA based solution of BB-Task.

**Information resources.** We had selected 130 mostly known bacteria and have processed data from corresponding corpus [17]. Formal contexts of “Entity”, “Areal” and “Pathogenicity” have the names of bacteria as objects and corresponding concepts from conceptual graphs as attributes. Among attributes there are bacteria properties (gram-negative, rod-shaped, etc.) for “Entity” context, mentions of water, soil and other environment parameters for “Areal” context and names and characteristics of diseases for “Pathogenicity” context. Table 1 shows numerical characteristics of created contexts.

Context name	Number of objects	Number of attributes	Number of formal concepts
<b>Entity</b>	130	26	426
<b>Areal</b>	130	18	127
<b>Pathogenicity</b>	130	28	692

**Table 1.** Numerical characteristics of created contexts.

As it is followed from the table there is relatively small number of formal concepts in the contexts. This is due to the sparse form of all contexts generated by conceptual graphs.

**Methods.** One of the problems in learning bacteria biotopes is the problem of bacteria classification: it is needed to classify bacteria according with their properties characterizing them as entities, characterizing their areal and pathogenicity. Various bacteria may have similar properties or may not. It is interesting to find clusters of bacteria containing ones having similar properties. This clustering task may be solved with concept lattice. Every concept in concept lattice being the set of one or several names of bacteria and their properties may be treated as fact. Facts can be extracted by using navigation in the lattice and interpretation its concepts and hierarchical links between them. For extracting facts and clustering we use visualization together with database technique of processing input queries. Special functionality was created in our system to visualize sub lattices of concept lattice to form special *views* consisted of sub lattices corresponding to certain property (intent in the lattice) or entity (extent in the lattice) of bacteria. We applied open source tool [19] which was modified and integrated to our system [15].

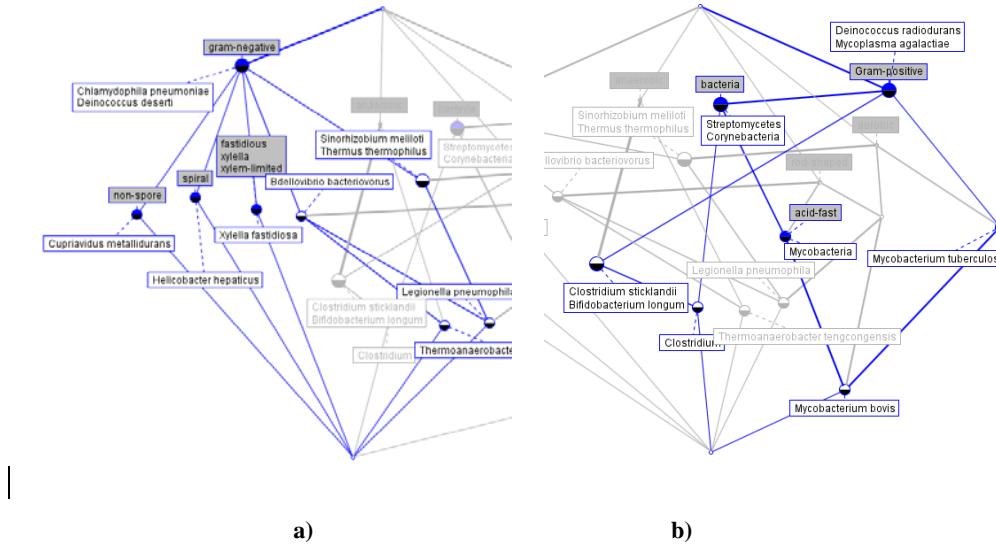
Fig. 2 shows a fragment of the formal context with the attributes related to some properties of bacteria: Gram staining, the property of being aerobic, etc.

	gram-negative	Gram-positive	acid-fast	rod-shaped	spiral	aerobic	anaerobic	non-spore	bacteria
Mycobacterium bovis						X			
Mycoplasma agalactiae		X							
Bifidobacterium longum		X							
Corynebacteria		X					X		
Mycobacteria		X							
Streptomycetes		X							
Mycobacterium tuberculosis		X							
Clostridium		X							
Clostridium sticklandii		X							
Deinococcus radiodurans		X							
Deinococcus deserti		X							
Cupriavidus metallidurans	X								
Thermoanaerobacter tengcongensis	X								
Thermus thermophilus	X								
Sinorhizobium meliloti						X			
Edellobiobacter bacteriovorus						X			
Xylella fastidiosa							X		
Chlamydophila pneumoniae									
Helicobacter hepaticus	X				X				
Legionella pneumophila	X				X				

**Fig. 2.** A fragment of the formal context “Entity” for 20 bacteria.

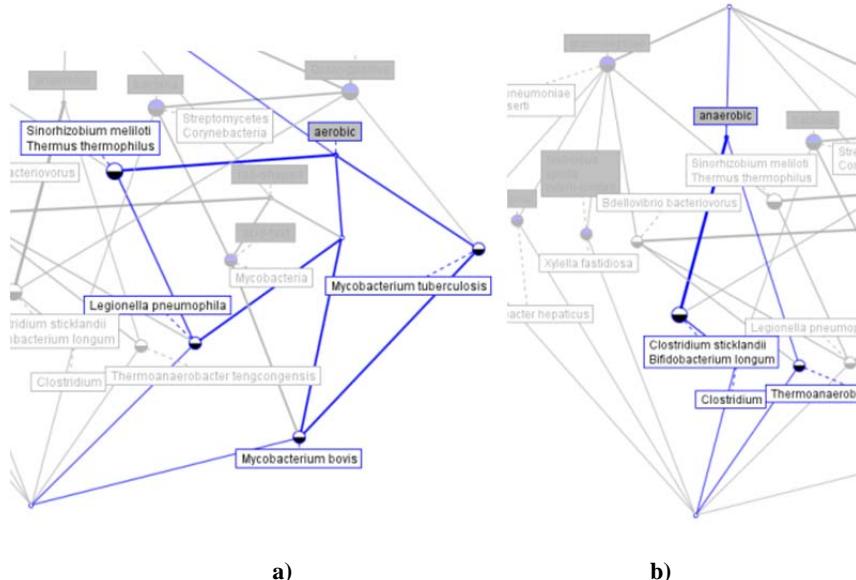
It is evident directly from the context that these 20 bacteria constitute two clusters according to the Gram staining: there is no bacterium which is simultaneously Gram-positive and Gram-negative. Lattice diagrams on the Fig. 3 confirm this fact.

Interpreting views on Fig. 3 we resolve that bacteria are clustered according with their Gram staining because the views on Fig. 3 a) and b) do not intersect.



**Fig. 3.** Views of concept lattice demonstrating Gram staining: a) – Gram-negative property, b) – Gram-positive property.

Clustering bacteria according with the property of being aerobic is not evident from the context on Fig. 2. Lattice diagrams on the Fig. 4 confirm the clustering bacteria according with this property by the same manner as for Fig. 3.



**Fig. 4.** Views of concept lattice demonstrating the property of bacteria to be aerobic.

However, the number of bacteria in Fig. 3 and 4 is not the same: Fig. 3 contains all 20 bacteria (10 in Figure 3-a and 10 in Figure 3-b.) and Fig. 4 - contains only 9 bacteria. This is due to the fact that the relevant texts do not contain information about the property of being aerobic for some bacteria.

### 3.3 Comparing results.

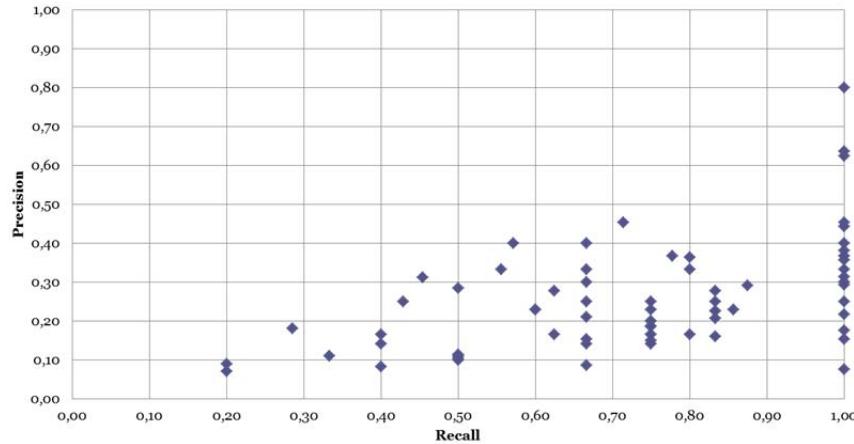
We can compare our results with two known similar solutions related to fact extraction problem. The first solution of extracting events is presented in [20] and is based on using special framework of EventMine [29]. This solution is realized as marking the text by highlighting its lexical elements as elements of event.

The second solution [21] is directly connected with BioNLP. The tasks of NER and RE were solved in [21] with Alvis framework [30] and results of relations extraction are also presented as marked words in the texts. Table 2 shows Recall, Precision and F-score calculated on the results of NER for Alvis framework in [21] and for our system.

	Recall	Precision	<i>F</i> -score
Alvis	0.52	0.46	0.49
Our system	0.42	0.62	0.50

**Table 2.** Recall, Precision and *F*-score for Named Entity Recognition problem.

The Precision / Recall ratio is more informative for evaluating the quality of solution of many problems in Data Mining. On the Fig. 5 it is shown such ratio calculated for 62 bacteria names extracted from texts in one of our experiments.



**Fig. 5.** Precision / Recall ratio for 62 bacteria names extracted from texts.

As it is followed from the Fig. 5, approximately half of the total number of objects has Precision / Recall ratio equal to unity that characterizes our solution as not bad.

Comparing our current results of fact extraction with the known ones we also have to resume that using concept lattice provides principally another variant of solution of fact extraction problem. The main difference of this solution is that it is not realized in the processed text by highlighting its lexical elements but it is realized with new external resource, conceptual model in the form of concept lattice.

#### 4 Conclusions and future work

This paper describes the idea of joining two paradigms of conceptual modeling - conceptual graphs and concept lattices. Current results of realizing this idea on textual data show its good potential for knowledge extraction. Concept lattice may serve as a frame of ontology constructed on texts. Its data which may or may not be interpreted as facts constitutes a knowledge stored in concept lattice being ready to extract.

In spite of certain useful features of presented technology there are some problems which need to be solved for improving the quality of modeling technique.

1. Conceptual graphs acquired from texts contain many noisy elements. Noise is constituted by the text elements that contain no useful information or cannot be interpreted as facts. Noisy elements significantly decrease efficiency of algorithms of fact extraction.
2. The next stage of developing current technology is creating of fledged information system which processes user queries and produces solutions of certain tasks on textual data. Not only visualization but also special user oriented interfaces to concept lattice will be created in this system.

**Acknowledgments.** The paper concerns with the work partially supported by the Russian Foundation for Basic Research, grant № 15-07-05507.

#### References

1. Priss, U.: Linguistic Applications of Formal Concept Analysis. In: Ganter; Stumme; Wille (eds.), Formal Concept Analysis, Foundations and Applications. Springer Verlag. LNAI 3626, p. 149-160 (2005)
2. Carpineto, C., Romano, G.: Using Concept Lattices for Text Retrieval and Mining. In B. Ganter, G. Stumme, and R. Wille (Eds.): Formal Concept Analysis: Foundations and Applications. Lecture Notes in Computer Science 3626, pp. 161-179. Springer-Verlag, Berlin, (2005)
3. Kuznetsov S. O., Strok F. V., Ilvovsky D. A., Galitsky B.: Improving Text Retrieval Efficiency with Pattern Structures on Parse Thickets // Proceedings of the FCAIR. Vol. 977. M.: CEUR Workshop Proceeding, pp. 6-21 (2013)

4. Otero P. G., Lopes G. P., Agustini, A.: Automatic Acquisition of Formal Concepts from Text. *Journal for Language Technology and Computational Linguistics*, vol. 23(1), pp. 59-74 (2008)
5. Meštrović, A.: Semantic Matching Using Concept Lattice. Proc. Concept Discovery in Unstructured Data, (CDUD-2012), pp. 49-58 (2012)
6. Cimiano, P. Hotho, A. Staab, S.: Learning Concept Hierarchies from Text Corpora using Formal Concept Analysis. *Journal of Artificial Intelligence Research*, Volume 24, pp. 305-339 (2005)
7. Wille, R. Conceptual Graphs and Formal Concept Analysis. *Proceedings of the Fifth International Conference on Conceptual Structures: Fulfilling Peirce's Dream*, pp. 290 - 303. Springer-Verlag, London (1997)
8. Galitsky, B., Dobrocsí, G., de la Rosa, J.L. Kuznetsov, S.O. From Generalization of Syntactic Parse Trees to Conceptual Graphs. In: M. Croitoru, S. Ferre, D. Lukose, Eds., *Conceptual Structures: From Information to Intelligence*, Proc. 18th International Conference on Conceptual Structures (ICCS 2010), Lecture Notes in Artificial Intelligence (Springer), vol. 6208, pp. 185-190 (2010)
9. Ganter, B., Stumme, G., Wille, R., eds.: *Formal Concept Analysis: Foundations and Applications*, Lecture Notes in Artificial Intelligence, No. 3626, Springer-Verlag (2005)
10. Sowa, J.F.: *Conceptual Structures: Information Processing in Mind and Machine*. Addison-Wesley, London (1984)
11. Chein, M., Mugnier, M.L.: Conceptual graphs are also graphs. Technical Report RRI-Chein-95, LIRMM (1995)
12. Ganter, B., Kuznetsov, S.O.: Pattern structures and their projections. In Harry S. Delugach and Gerd Stumme, editors, *Concept. Struct. Broadening Base*, volume 2120 of Lecture Notes in Computer Science, pages 129–142. Springer, Berlin, Heidelberg (2001)
13. Bogatyrev M., Tuhtin, V.: Creating conceptual graphs as elements of semantic texts labeling. In: *Computational Linguistics and Intellectual Technologies*. Proc. Int. Conference “Dialogue”. Moscow, pp. 31-37 (in Russian) ( 2009)
14. Mikhail Bogatyrev: Conceptual Modeling with Formal Concept Analysis on Natural Language Texts. *Proceedings of the XVIII International Conference «Data Analytics and Management in Data Intensive Domains»*. CEUR Workshop Proc. Vol.-1752, pp. 16-23 (2016)
15. Mikhail Bogatyrev, Kirill Samodurov: Framework for Conceptual Modeling on Natural Language Texts. Proc. Int. Workshop on Concept Discovery in Unstructured Data (CDUD 2016) at the Thirteenth International Conference on Concept Lattices and Their Applications. Moscow, 2016. CEUR Workshop Proc. Vol.-1625. pp. 13-24 (2016)
16. U.S. National Library of Medicine. <http://www.ncbi.nlm.nih.gov/pubmed>
17. Bossy R, Jourde J, Manine A-P, Veber P, Alphonse E, Van De Guchte M, Bessières P, Nédellec C: BioNLP 2011 Shared Task - The Bacteria Track. *BMC Bioinformatics*, 13: S8, pp. 1-15 (2012)
18. Bossy, R., Golik, W., Ratkovic, Z., Bessi'eres, P., and N'edellec, C.: BioNLP shared Task 2013 – An Overview of the Bacteria Biotope Task. In *Proceedings of the BioNLP Shared Task 2013 Workshop*, pages 161–169, Sofia, Bulgaria. ACL. (2013)
19. ConExp-NG. <https://github.com/fcatools/conexp-ng>
20. Miwa M, Ananiadou S.: Adaptable, high recall, event extraction system with minimal configuration. *BMC Bioinformatics*, 16(10):1-11 (2015)
21. Ratkovic, Z., Golik, W., Warnier, P.: Event extraction of bacteria biotopes: a knowledge-intensive NLP-based approach. - *BMC Bioinformatics*, 13, (Suppl 11): S8, pp. 1-11 (2012)
22. The 4th BioNLP Shared Task. <http://2016.bionlp-st.org>

23. Proceedings of the 4th BioNLP Shared Task Workshop. Berlin, Germany, August 13, 2016. <http://aclweb.org/anthology/W/W16/W16-30.pdf>
24. Pontus Stenetorp, Wiktoria Golik, Thierry Hamon, Donald C. Comeau, Rezarta Islamaj Dogan, Haibin Liu, W. John Wilbur: BioNLP Shared Task 2013: Supporting Resources. In Proceedings of the 3d BioNLP Shared Task Workshop. (2013)
25. OntoBiotope ontology. <http://genome.jouy.inra.fr/bibliome/MEM-OntoBiotope/>
26. Bjorne, J. and Salakoski, T. (2011). Generalizing biomedical event extraction. In Proceedings of BioNLP Shared Task 2011 Workshop. ACL. (2011)
27. Parisa Kordjamshidi, Wouter Massa, Thomas Provoost, Marie-Francine Moens: Machine Reading for Extraction of Bacteria and Habitat Taxonomies. In: Fred A., Gamboa H., Elias D. (eds.) Biomedical Engineering Systems and Technologies. Communications in Computer and Information Science, vol. 574. Springer, pp. 239-255. (2015)
28. Karadeniz, I. and Ozgur, A.: Bacteria biotope detection, ontology-based normalization, and relation extraction using syntactic rules. In Proceedings of the BioNLP Shared Task 2013 Workshop, pages 170–177, Sofia, Bulgaria. ACL. (2013)
29. EventMine framework. <http://www.nactem.ac.uk/EventMine/>
30. Alvis system. [http://www.quaero.org/module\\_technologique/alvis-nlp-alvis-natural-language-processing/](http://www.quaero.org/module_technologique/alvis-nlp-alvis-natural-language-processing/)