Reconstruction of Belemnite Evolution Using Formal Concept Analysis

Radim Belohlavek,[1,2]* Martin Košťák[3]†, and Petr Osicka[2]
[1]State University of New York at Binghamton, Binghamton, NY
rbelohla@binghamton.edu
[2]Palacky University, Olomouc, Czech Republic
[3]Charles University, Prague, Czech Republic

Abstract

The paper presents results on using formal concept analysis in the problem of identification of taxa and reconstructing evolution from paleontological data. We present results of experiments performed with belemnites—a group of extinct marine cephalophods which seems particularly suitable for such a purpose. We demonstrate that the methods of formal concept analysis are capable of revealing taxa and relationships among them which are relevant from a paleobiological point of view.

1 Introduction and Problem Setting

An important system-theoretic concept is that of a taxonomy, i.e. a classification scheme arranged in a hierarchical structure. Biological taxonomies and the methods of devising them are perhaps the best known and most widely studied. There exist several approaches to biological classification, with phylogenetics (cladistics) and phenetics (numerical taxonomy) being perhaps the two most important methods [Dunn and Everitt, 2004; Jardin and Sibson, 1971; Kitching et al., 1998; Sneath and Sokal, 1973]. The aim of this paper is to explore the idea of utilizing formal concept analysis in identification of taxa and devising taxonomies in paleobiology. The basic idea is to identify taxa with formal concepts which are particular groupings of objects characterized by sharing certain properties (attributes).

The overall goal of our research is to find out whether and how formal concept analysis may help in identification of taxa and reconstructing evolution from paleobiological data. In the paper, we report some of our first results which we obtained from data about a particular group of extinct cephalophods called belemnites, which are similar to modern squid.

*The research of R. Belohlavek was supported by grant no. MSM 6198959214 and by grant no. P103/10/1056 of the Czech Science Foundation.

†The research of M. Košťák was supported by grant no. 205/07/1365 of the Czech Science Foundation and by grant no. MSM 0021620855.

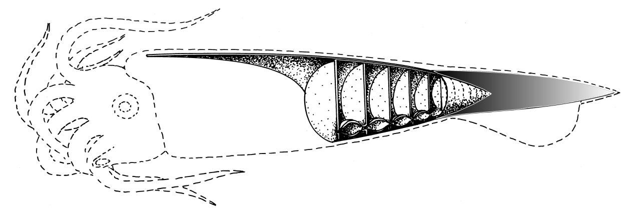

Figure 1: Schematic sketch of belemnite body with position of the internal shell. Apical grey part represents the rostrum.

Coleoid cephalopods played an important role in the Cretaceous ecosystem in both the Northern and Southern hemispheres. Inside this diverse group, especially belemnites were a common part of nectonic assemblages in marine shallower water ecosystems. They belong to coleoids with internal shell. The external morphology of belemnites resembles partly Recent squids and their behavior—larger sets of specimens concentrated probably during spawning acts. However, the internal shell characteristic is strongly different in both, non related cephalopod groups. In belemnites, the internal shell is composed of three major components—the most commonly preserved calcitic rostrum, aragonitic phragmoconus, and dorsal pro-ostracum. The belemnite taxonomy and systematics are based on rostrum characterstics. This hydrodynamic and also hydrostatic organ shows gradual changes in their evolutionary history. A schematic sketch of a belemnite body is shown in Fig. 2. For detailed morphologic features of rostrum see [Christensen, 1997; Košťák et al., 2004].

The most important morphologic features in rostrum morphology are size, shape and especially the part called the alveolar end and/or fracture. From quite a large systematic group (order Belemnitida), we chose a part of family Belemnitellidae—a dominant Upper Cretaceous belemnite group.

As no soft body parts are preserved for taxonomic distinction, belemnite taxonomy is mainly based on the following morphological characteristics of the belemnite rostrum:

- Shape and size of the rostrum.
- Structure of the alveolar end.

- Internal structures at alveolar end.
- External characteristics of the rostrum.
- Structure of apex.

Belemnites are extinct cephalopod group with no descendents. Their systematic, palaeoecology, palaeobiogeography and stratigraphy is based on palaeontological research. However, the simulation of belemnite evolution opens new options for using various mathematical models, for which the presented cephalopod group is particularly suitable.

2 Formal Concept Analysis

Formal concept analysis (FCA) is a method for data analysis with applications in various domains (see [Ganter and Wille, 1999] for foundations and [Carpineto and Romano, 2004] for applications). The input to FCA consists of a data table describing a set $X = \{1, \ldots, n\}$ objects and a set $Y = \{1, \ldots, m\}$ of attributes. The table specifies attribute values of the objects. The main aim in FCA is to reveal from the data a hierarchically organized set of particular clusters, called formal concepts, and a small set of particular attribute dependencies, called attribute implications. FCA aims at formalizing and utilizing a traditional theory of concepts, in which a concept is understood as an entity consisting of its extent (collection of all objects to which the concept applies) and its intent (collection of all attributes to which are characteristic for the concept). For example, the extent of the concept *dog* consists of all dogs while its intent consists of attributes such as *barks*, *has tail*, *has four limbs*, etc. The information extracted from data in FCA is well-comprehensible by users because FCA works with notions which humans are used to reason with in the ordinary life. This is an important feature of FCA which makes it appealing for users.

In the basic setting of FCA, attributes are assumed to be binary, i.e. a given object either has or does not have a given attribute. A data table with binary attributes is represented by a triplet $\langle X, Y, I\rangle$, called a *formal context*, which consists of the above-mentioned sets X and Y of objects and attributes, and a binary relation I between X and Y (incidence relation, to-have relation). Thus, $I \subseteq X \times Y$, $\langle x, y\rangle \in I$ indicates that object x has attribute y, $\langle x, y\rangle \notin I$ indicates that x does not have y. Objects $x \in X$ correspond to table rows, attributes $y \in Y$ correspond to table columns, and I is represented by 0s and 1s in the table entries (i.e., 1 indicates that the corresponding object does have the corresponding attribute and 0 indicates the opposite). A formal concept of $\langle X, Y, I\rangle$ is any pair $\langle A, B\rangle$ of sets $A \subseteq X$ (*extent*, set of objects to which the concept applies) and $B \subseteq Y$ (*intent*, set of attributes characterizing the concept) such that B is just the set of attributes shared by all objects from A, and A is the set of all objects sharing all attributes from B. In symbols, this can be written as $A^{\uparrow} = B$ and $B^{\downarrow} = A$, where

$$A^{\uparrow} = \{y \in Y \mid \text{for each } x \in A\colon \langle x, y\rangle \in I\},$$
$$B^{\downarrow} = \{x \in X \mid \text{for each } y \in B\colon \langle x, y\rangle \in I\}.$$

The thus introduced arrow operators form a Galois connection between X and Y and play an important role in FCA. The set of all formal concepts of $\langle X, Y, I\rangle$ is denoted by $\mathcal{B}(X, Y, I)$. That is,

$$\mathcal{B}(X, Y, I) = \{\langle A, B\rangle \mid A^{\uparrow} = B,\ B^{\downarrow} = A\}.$$

Under a partial order $\leq$, defined for $\langle A_1, B_1\rangle, \langle A_2, B_2\rangle \in \mathcal{B}(X, Y, I)$ by

$$\langle A_1, B_1\rangle \leq \langle A_2, B_2\rangle \text{ iff } A_1 \subseteq A_2 \text{ (iff } B_2 \subseteq B_1),$$

$\mathcal{B}(X, Y, I)$ happens to be a complete lattice, so-called *concept lattice* associated to $\langle X, Y, I\rangle$ [Ganter and Wille, 1999]. The partial order $\leq$ represents the subconcept-superconcept (generalization-specialization) relationship between formal concepts. That is, $\langle A_1, B_1\rangle \leq \langle A_2, B_2\rangle$ means that formal concept $\langle A_2, B_2\rangle$ (representing e.g. *mammal*) is more general than $\langle A_1, B_1\rangle$ (e.g. *dog*). Efficient algorithms for computing $\mathcal{B}(X, Y, I)$ exist.

The concept lattice is visalized by a so-called labelled line diagram (or Hasse diagram) [Ganter and Wille, 1999]. The diagram consists of nodes, some of which are connected by lines. The formal concepts are represented by nodes, the subconcept-superconcept relationship is represented by the lines, and a particular way of labeling the nodes by object and attribute names is used. In particular every node (formal concept) is connected to all of its direct predecessors (formal concepts which are more specific) and direct successors (formal concepts which are more general). The diagram is easily understood by users and provides a hierarchical view on the data.

In several situations, the expert finds certain formal concepts in the concept lattice interesting and relevant while others not. One reason for this is that in addition to the input data (object, attributes, their relationship), the expert may have further knowledge regarding the objects and attributes. According to this knowledge, the expert may regard some formal concepts interesting (relevant) and the others not. In [Belohlavek and Sklenar, 2005; Belohlavek and Vychodil, 2009], we developed an approach to handling one particular type of a background knowledge which we utilize in this paper. The idea is that the attributes may not be equally important to form concepts. To give a simple example, consider the following attributes of books: *hardbound*, *paperback*, *engineering*, *science*, *philosophy*. In a reasonable taxonomy of books, we naturally consider the attributes *engineering*, *science*, and *philosophy* more important than *hardbound* or *paperback*. As a consequence, we would consider a formal concept characterized by (i.e. with its intent consisting of) *hardbound* (all books which are hardbound) not natural (relevant, interesting). One the other hand, the formal concept characterized by *engineering* (books on engineering), the one characterized by *science*, and perhaps also the one characterized by *engineering* and *paperback* (paperback books on engineering) would be considered natural. Such a background knowledge may be represented by so-called *attribute-dependency formulas* (AD-formulas) [Belohlavek and Sklenar, 2005;

Belohlavek and Vychodil, 2009]. In our example, the AD-formula representing our background knowledge is

$$hbound \sqcup pback \sqsubseteq engineering \sqcup science \sqcup philosophy,$$

but in general, there background knowledge is represented by a set of AD-formulas. An AD-formula over a set Y of attributes is an expression $D_1 \sqsubseteq D_2$ where $D_1, D_2 \subseteq Y$. A formal concept $\langle A, B\rangle$ of $\langle X, Y, I\rangle$ is considered compatible with a set T of AD-formulas ($\langle A, B\rangle$ satisfies T) if for each $D_1 \sqsubseteq D_2$ from T, if $D_1 \cap B \neq \emptyset$ then $D_2 \cap B \neq \emptyset$. That is, a formal concept characterized by *hardbound* is not compatible with the above background knowledge because in this case, $B = \{hardbound\}$, and thus the for above AD-formula $D_1 \sqsubseteq D_2$ with $D_1 = \{hbound, pback\}$ and $D_2 = \{engineering, science, philosophy\}$, we have $D_1 \cap B \neq \emptyset$ but $D_2 \cap B = \emptyset$. Intuitively, $\langle A, B\rangle$ contains a less important attribute but does not contain any of the prescribed more important attributes. The set of all formal concepts of $\langle X, Y, I\rangle$ compatible with T is denoted by $\mathcal{B}_T(X, Y, I)$ and is considered the set of all natural (interesting, relevant) formal concept in the data $\langle X, Y, I\rangle$ given the background knowledge T.

3 Belemnite Evolution in Time, Space

The morphologic changes in time and space are well documented especially in early belemnitellids [Košťák, 2004]. Relatively few morphologic features in belemnite rostrum, clear taxonomy, and new systematic and stratigraphic revision (see below) make it possible to use formal methods for identification of taxa.

Belemnites show an interesting model of migration patterns and provincialism in Jurassic and Cretaceous coleoids. We have used the Upper Cretaceous belemnites (Belemnitellidae). Their evolution centre lied in the SE parts of the Russian Platform and mass migrations affected shallow seas in Central and NW Europe, Siberian areas and North America during the Cenomanian through Coniacian. They show marked provincialism in the North European, the North American and the East European Provinces [Christensen, 1997; Košťák et al., 2004]. The last belemnitellids became extinct at the K/T boundary (although, it concerns only last 1-2 species; note: K/T boundary is the Cretaceous/Tertiary boundary with mass extinction 65.5 Ma). The fall in belemnitellid diversity is related to the reduction of shallow sea areas, eustasy, and the radiation of new vertebrate groups (i.e. Neoselachii, Teleostei). The analysis selected results of which are provided below focuses on their evolution, particularly on the frequency of morphologic changes, endemicism, and species and generic relationships.

We chose the Upper Cretaceous belemnites from the Cenomanian–Coniacian interval for the folowing reasons

- Systemetic revision has been finished recently.
- Clear stratigraphic position.
- The evolutionary lineage from one ancestor is well documented.
- Recognized migration patterns and spreading succession in time and space
- Quite simple morphology of preserved parts, with clear changes in time.

Table 1: Objects: selected belemnite species

belemnite species	label
Actinocamax verus antefragilis	0
Praeactinocamax primus	1
P. plenus	2
P. plenus cf. strehlensis	3
P. triangulus	4
P. aff triangulus	5
P. sozhensis	6
P. contractus	7
P. planus	8
P. coronatus	9
P. matesovae	10
P. medwedicicus	11
P. sp. 1	12
P. sp. 2	13
P. strehlensis	14
P. bohemicus	15
P. aff. bohemicus	16
P. cobbani	17
P. manitobensis	18
P. cf. manitobensis	19
P. sternbergi	20
P. walkeri	21
Goniocamax intermedius	22
G. surensis	23
G. christenseni	24
G.lundgreni	25

We have analyzed 26 species belonging to three genera in the Cenomanian–Coniacian interval (i.e., 97–87 Ma). This phase of belemnite evolution shows the highest freqency in morphologic changes. Thus, we suppose this to be a very suitable for our purpose.

4 Experiments

Dataset The data (formal context $\langle X, Y, I\rangle$) used in our experiment consists 26 belemnite species (objects of X) depicted in Table 4 and 38 rostrum characteristics (attributes of Y) depicted in Table 2. Due to lack of space, we do not display the table representing which species have which attributes.

Attribute priorities and AD-formulas The corresponding concept concept lattice $\mathcal{B}(X, Y, I)$ consists of 616 formal concepts is is thus quite large. From the paleontological point of view, it contains both formal formal concepts which seem natural and interesting as well as those which are not natural. To eliminate the latter, we employed AD-formulas (see Section 2). Namely, it is indeed the case that a paleontologist considers some attributes more important than others. The use of AD-formulas to make his background knowledge explicit and to take it into account in the subsequent analysis is therefore a natural solution.

Table 2: Attributes and their groups

Group/name	label
Size	
large guards (more 65mm)	a
medium guards (65-80mm)	b
small guards (less 65mm)	c
cigar shape in dorsovenral view	d
lanceolate in DV view	e
slightly lanceolate in DV view	f
subcylindrical in DV view	g
conical in DV view	h
cigar shape in lateral view	i
lanceolate in L view	j
slightly lanceolate in L view	k
subcylindrical in L view	l
conical in L view	m
Flattening	
laterally flattened	n
dorsally flattened	o
ventrally flattened	p
Alveolar fracture, pseudoalveolus	
high conical alveolar fracture	q
low conical alveolar fracture	r
shallow pseudoalveolus (less 3mm)	s
deep pseudoalveolus (more 3mm)	t
Cross section	
oval cross section of alveolar fracture	u
oval to triangular cross section of AF	v
triangular cross section of AF	w
circular cross section of AF	x
Others A	
bottom of ventral fissure	y
dorsolateral furrows	z
venral furrow	α
Others B	
dorsolateral depressions	β
granulation of the whole guard	γ
striation of the whole guard	δ
mucro	ϵ
vascular imprints frequent	η
Others C	
conellae	θ
granulation of a part of the guard	ι
striation of a part of the guard	κ
vascular imprints rare	λ

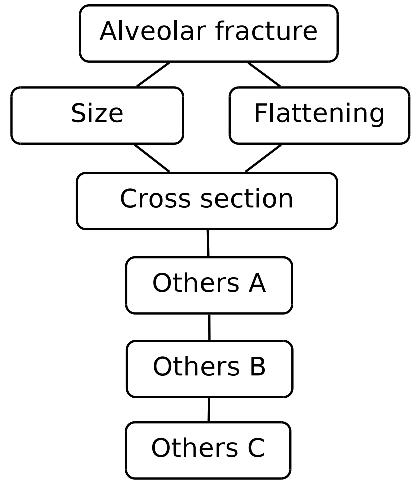

Figure 2: Hasse diagram of the order set of attribute subsets used for the AD formula generation

According to expert opinion, we partitioned the set Y of attributes into the seven groups shown in Table 2 and partially ordered them as depicted in Fig. 4. Then, for every two groups $D_1, D_2 \subseteq Y$ of the seven groups such Y_1 is a direct predecessor of Y_2, we add to the set T of our AD-formulas the AD-formula $D_1 \sqsubseteq D_2$. For example, since the group Others B is a direct predecessor of Others A, we add

$$\beta \sqcup \gamma \sqcup \delta \sqcup \epsilon \sqcup \eta \sqsubseteq y \sqcup z \sqcup \alpha$$

to T.

The concept lattice The constrained concept lattice $\mathcal{B}_T(X, Y, I)$ (containing only the formal concepts which are compatible with the AD-formulas from T, see Section 2) contains 121 formal concepts only and is depicted in Fig. 4. The concepts are labeled by numbers which we use below when interpreting the formal concepts and taxa and their relationships. Due to lack of space, we postpone the description of all the formal concepts to a full version of this paper.

Interpretation The Upper Cretaceous belemnites (family Belemnitellidae = belemnitellids in this paper) are represented by 7 genera: *Actinocamax*, *Praeactinocamax*, *Goniocamax*, *Gonioteuthis*, *Belemnella*, *Belemnellocamax*, *Belemnitella*, and problematic genera *Belemnocamax* and *Fusiteuthis*. While the phylogenic lineage going from *Goniocamax* to younger *Gonioteuthis*, *Belemnella*, and *Belemnitella* is quite clear and is connected with progressive calcification of the aleveolar part [Christensen, 1997], the relationships between earlier genera like the origin of the Late Cretaceous belemnites are still unclear.

The concept analysis strictly derived and separated genus *Actinocamax* from other belemnitellids (concept No. 3). We have used one species (subspecies) of *Actinocamax verus antefragilis* (the stratigraphic range of this species is very long and falls into the period of rapid evolution of another belemnite taxa).

Table 3: Interesting formal concepts

Concept	Extent	Intent
3	0	c d i n q x z
2	22 23	c e f l p t v z β λ ϵ
4	23	c d e f l p s t v z β κ λ ϵ
103	22	b c e f k l p t v z β δ λ ϵ
8	2 3 7 14 15 16 17 18 20 21	b p r v α
79	10	b e l p s u z β δ λ ϵ
11	1 2 3 7 8 14 16 17 20 21	b p r v β κ α
104	1 2 19 20	a f p r u β κ α
116	6	a b c e k l p q u z β δ
17	11 13	b l p t v β α
115	4	a b c e k l p t w z β λ ϵ α
18	11 22	b l p t v z β
78	13	b e l p t u v β κ α
40	9	b g m p t u z α
114	18	a b f k l p r s v w z β γ ι α
70	15	b f g l m p r v w β γ ι κ α
95	17	b e f k l p r u v z β ι κ ϵ α
112	20	a b f k p r u v z β γ ι κ λ ϵ α
117	1 2	a b c e f g k l p r u z β δ κ α
119	1	a b c d e f g j k l p r u z β δ κ α
118	2	a b c e f g k l p r u v z β δ κ λ ϵ α
87	2 8	b e k l p r u z β δ λ ϵ α
84	2 7 8	b e k l p r u z β λ ϵ α
102	14	b c e j k p r u v α
101	1 14	b c e j k p r u α
90	1 3	b e j p r u β α
5	4 9 11 12 13 22 24 25	b p t
56	24 25	b f l o p t u θ y z β δ λ ϵ α

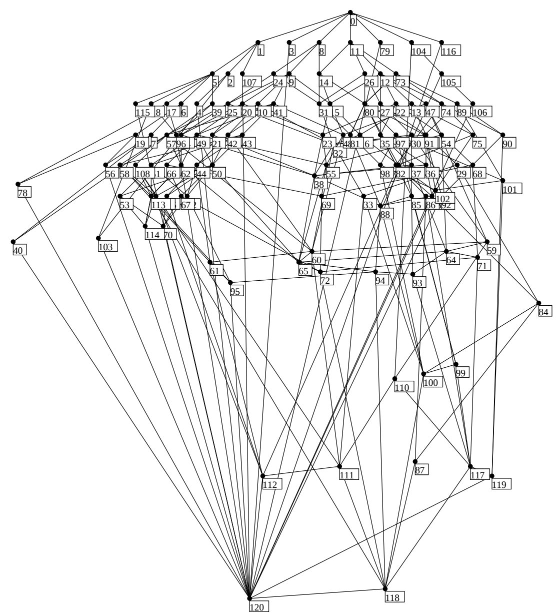

Figure 3: Hasse diagram of ordered set of 121 conceps resulting from the size reduction via AD formulas.

This and later species of *Actinocamax* are so similar each other that their distinguishing is possible just at subspecies level. The concept analysis of this genus showed no relationship to other belemnite taxa. This should be interpreted like an extreme derivation from belemnitellids (size, shape of the rostrum and alveolar fracture) and/or by presence of another belemnoid ancestors. In this respect, the *Actinocamax* evolutionary lineage should be excluded from belemnitellids as their ancestor probably does not belong to taxa related to earlier genus *Praeactinocamax*. Species of *Praeactinocamax* were commonly distributed in Euroasia and North America from the Cenomanian through Coniacian (Santonian in Greenland). This genus occurred a few milions years before the earliest *Actinocamax* and the first species of this genus—*P. primus* is generaly considered [Naidin, 1964; Christensen, 1997] to be also the earliest species of belemnitellids. If this opinion is true, *P. primus* and its stright descendent *P. plenus* (*primus/plenus* group) are initial taxa for all younger belemnitellids (i.e. monophyletic group) and this must be clearly detected in concept analysis. If we look at concept No. 11, we can see attributes common to another taxa of this genus. Original morphologic features in *primus/plenus* group (also concepts No. 117,118, etc.) are present especially in the East European species.

Of a high importance is concept No. 8. showing attributes common for both conservative morphological features in the East European species (group No. 14 (related to No. 11)) but also to relatively distinct Turonian North American (*P. manitobensis*, concept No. 114, *P. cobbani* and *P. sternbergi*—concepts No. 61, 95, 112) and the Late Turonian species from Central and NW Europe (*P. bohemicus*, concept No. 70). In this respect, we observe at least two different evolutionary lineage, i.e. the expression of allopatric speciation in different geographic areas.

High endemic species with unknown or poorly documented phylogenetic relationship are *P. sozhensis* (concept No. 116) and *P. matesovae* (concept No. 79).

Another belemnite lineage is evolved also from the earliest species of *Praeactinocamax* by forming a deeper pseudoalveolus (a space surrounding the phragmoconus). This evolutionary innovation is partly observable already in some specimens of *P. plenus* population. So, the origin of this lineage took place probably in the Late Cenomanian. A typical morphologic features inside this morphologic lineage are summarized in the concept No. 1. (see attributes in concepts No. 1, 5). This evolutionary lineage shows markedly higher number of rare and endemic species (*P. medwedicicus*, *P. coronatus*, *P. aff. triangulus*, *P.* sp. 1 and 2, etc.). The iteration could be understood also as an expression of allopatric speciation in the East European Cretaceous sea. This lineage is also important by the origin of the earliest species of genus *Goniocamax* (*G. intermedius* and *G. surensis*—concepts No. 2,4,103) and advanced *G. christenseni* and *G. lundgreni* (concept No. 56). The last mentioned species couple is also considered to be an ancestor of younger Belemnitella stock (including also genus *Gonioteuthis*). The transition between *Goniocamax* and *Belemnitella* is gradual and relatively clear.

5 Conclusions

The results demonstrate that natural taxa and interesting relationships have been revealed using this approach. In particular, the analysis confirmed: separated Actinocamax lineage which is probably derived from another ancestor than in genus Praeactinocamax; three different morphologic trends in genus Praeactinocamax, explained by allopatric speciation; clearly separated North American and rare NW and Central European species from the East Europaen Praeactinocamax taxa; iteration with endemic taxa; the genus origin and its derivation (Goniocamax from Praeactinocamax). In future research, we plan to focus on the following topics: further analyses of the present data (employ further methods of FCA); application of the method to further fossil groups; a long-term goal is to study how expert paleontologist form taxa and taxonomies and possible formalization of this process.

References

[Belohlavek and Sklenar, 2005] R. Belohlavek and V. Sklenar. Formal concept analysis constrained by attribute-dependency formulas. LNCS (Proc. ICFCA 2005), **3403**(2005), pp. 176-191.

[Belohlavek and Vychodil, 2009] R. Belohlavek and V. Vychodil. Formal Concept Analysis With Background Knowledge: Attribute Priorities. *IEEE Transactions on Systems, Man, and Cybernetics, Part C.* **39**(4)(2009), 399–409.

[Carpineto and Romano, 2004] C. Carpineto and G. Romano. *Concept Data Analysis. Theory and Applications.* J. Wiley, 2004.

[Christensen, 1997] W. K. Christensen. The Late Cretaceous belemnite family Belemnitellidae: Taxonomy and evolutionary history. *Bulletin of the Geological Society of Denmark.* **44**(1997), 59–88.

[Dunn and Everitt, 2004] G. Dunn and B. S. Everitt. *An Introduction to Mathematical Taxonomy.* Dover, 2004.

[Ganter and Wille, 1999] B. Ganter and R. Wille. *Formal Concept Analysis. Mathematical Foundations.* Springer, Berlin, 1999.

[Jardin and Sibson, 1971] N. Jardin and R. Sibson. *Mathematical Taxonomy (Probability & Mathematical Statistics).* J. Wiley, 1971.

[Kitching et al., 1998] I. Kitching et al. *Cladistics: Theory and Practice of Parsimony Analysis.* Oxford University Press, 1998.

[Košťák, 2004] M. Košťák. Cenomanian through the lowermost Coniacian Belemnitellidae Pavlow (Belemnitida, Coleoidea) of the East European Province. *Geolines.* **18**(2004), 59–109.

[Košťák et al., 2004] M. Košťák, S. Čech, B. Ekrt, M. Mazuch, F. Wiese, S. Voigt, C. J. Wood. Belemnites of the Bohemian Cretaceous Basin in a global context. *Acta Geologica Polonica.* **54**, 4(2004), 511–533.

[Naidin, 1964] D. P. Naidin. *Upper Cretaceous belemnites from the Russian Platform and adjacent areas. Actinocamax, Gonioteuthis, Belemnellocamax* (In Russian). Izdatelstvo Moskovskogo Gosudarstvennogo Universiteta, Moskva, 1964, 190 pp.

[Priss, 2003] U. Priss. Formalizing Botanical Taxonomies. LNAI (Proc. ICCS 2003) **2746**(2003), pp. 309–322.

[Sneath and Sokal, 1973] P. H. A. Sneath, R. R. Sokal. *Numerical Taxonomy: The Principles and Practice of Numerical Classification.* W. H. Freeman & Co., 1973.