R. Bělohlávek, M. Košťák, and P. Osička: Reconstruction of belemnite evolution using FCA

Department of Computer Science, Palacky University, Olomouc (17. listopadu 12, CZ–77146 Olomouc, Czech Republic)

Introduction

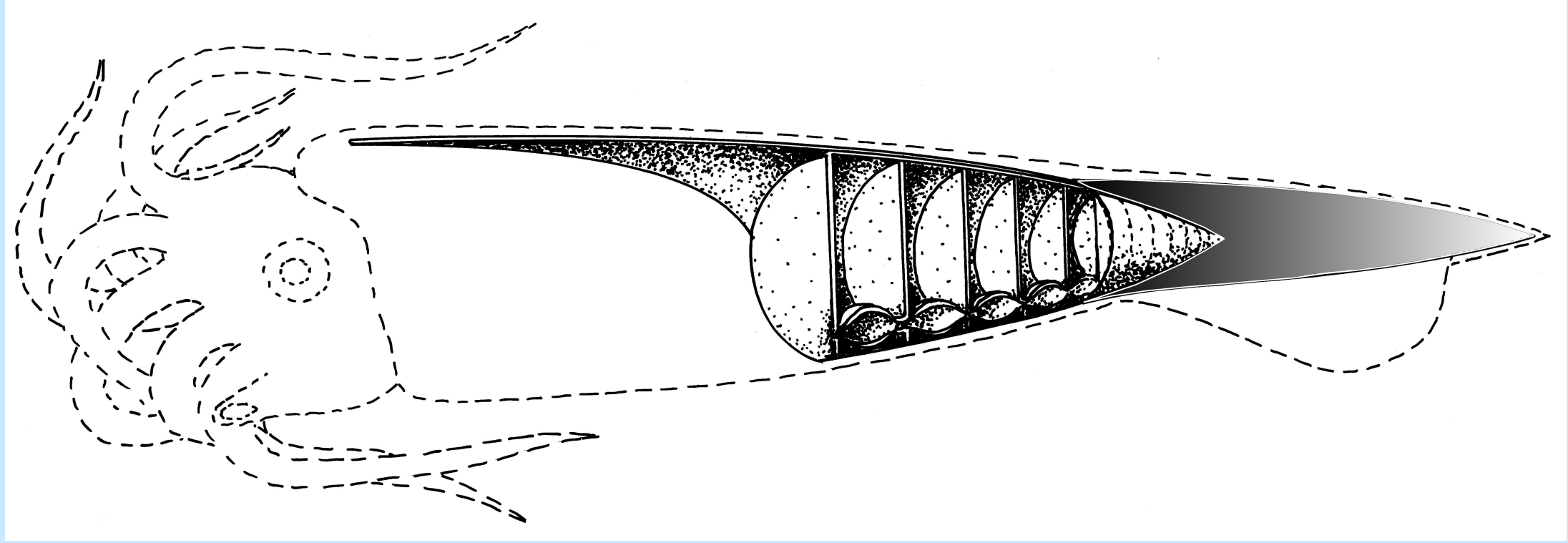
We present a case study in identification of taxa in paleobiological data. Our approach utilizes formal concept analysis and is based on conceiving a taxon as a group of individuals sharing a collection of attributes. In addition to the incidence relation between individuals and their attributes, the method uses expert background knowledge regarding importance of attributes which helps filter out correctly formed but paleobiologically irrelevant taxa. We present results of experiments performed with belemnites—a group of extinct cephalophods which seems particularly suitable for such a purpose. We demonstrate that the methods are capable of revealing taxa and relationships among them that are relevant from a paleobiological point of view.

Taxonomy

- 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 method
- 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).

Belemnites

- Extinct cephalopod group with no descendents (their habitus partly resembles some Recent squids)
- Their systematic, palaeoecology, palaeobiogeography and stratigraphy is based on palaeontological research.



Preliminaries

Formal Concept Analysis

- Relational object-attribute input data in the form of a table describing objects (rows), their attributes/features (columns), and their "yes/no" relationship.Data table is formalized by formal context: $\langle X, Y, I\rangle$, where $I \subseteq X \times Y$.
- Concept-forming operators:

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

- Primary output: a hierarchically ordered collection of concept-clusters. Formal concept is a pair $\langle A, B\rangle$ such that $A^{\uparrow} = B$ and $B^{\downarrow} = A$. Formal concepts form a complete lattice w.r.t to the order $\langle A_1, B_1\rangle \leq \langle A_2, B_2\rangle$ iff $A_1 \subseteq A_2$ (iff $B_2 \subseteq B_1$)
- Ganter, Wille: *Formal Concept Analysis: Mathematical Foundations.* Springer-Verlag, Berlin, 1998

Attribute dependencies

- Expert may find some formal concepts interesting and relevant while others not. Possible reason: the expert may have further knowledge regarding the objects and attribute
- We utilize the idea that attributes may not be equally important in concept formation. Formalization: An AD-formula over a set Y of attributes is an expression $D_1 \sqsubseteq D_2$ where $D_1, D_2 \subseteq Y$. Meaning: attributes of D_2 are more important than attributes of D_1. A formal concept $\langle A, B\rangle$ satisfies the AD-formula $D_1 \sqsubseteq D_2$, if $D_1 \cap B \neq \emptyset$ implies $D_2 \cap B \neq \emptyset$.
- A formal concept satisfies a set T of AD-formulas if $\langle A, B\rangle$ satisfies each AD-formula $D_1 \sqsubseteq D_2$ from T.
- Example: consider the following attributes of books: hardbound, paperback, engineering, science, philosophy. We naturally consider the attributes engineering, science, and philosophy more important than hardbound or paperback. This corresponds to AD-formula

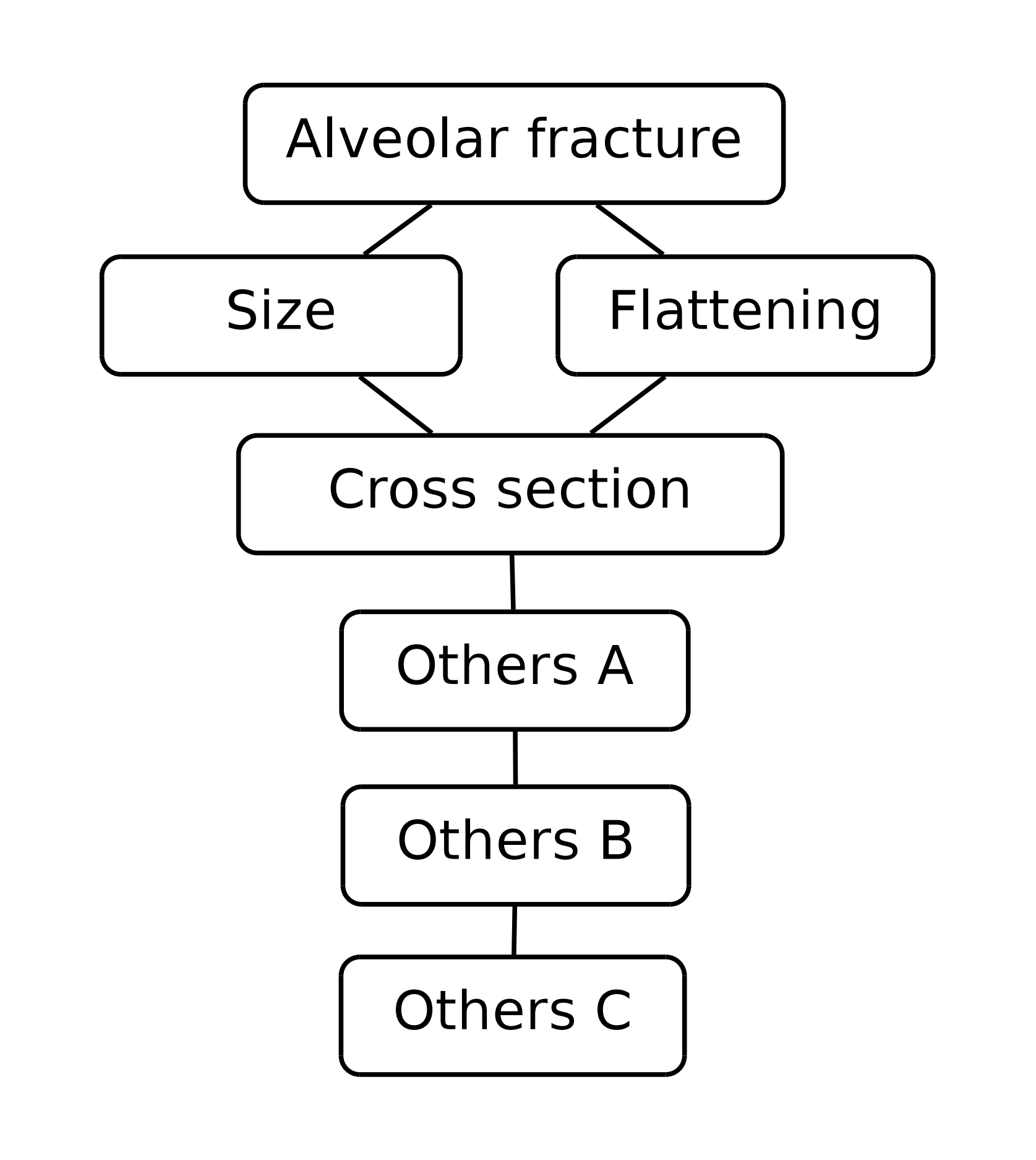
$$\textit{hbound} \sqcup \textit{pback} \sqsubseteq \textit{engineering} \sqcup \textit{science} \sqcup \textit{philosophy},$$

- formal concept characterized by the attribute hardbound is not natural. The one characterized by engineering, and the one characterized by engineering and paperback would be considered natural
- Belohlavek, Vychodil: *Formal concept analysis with background knowledge: attribute priorities* IEEE Transactions on Systems, Man, and Cybernetics, Part C, 39(4)(2009), 399 – 409.

Analysis and Results

The dataset

- The data (i.e., a formal context $\langle X, Y, I\rangle$) used in our experiment consists of 26 belemnite species (objects of X) 36 rostrum characteristics (attributes of Y).
- As no soft body parts are preserved for taxonomic distinction, the attributes are selected based on the following morphological characteristics of the belemnite rostrum (grey part in the picture):
  - structure of the alveolar end,
  - shape and size of the rostrum,
  - internal structures at alveolar end,
  - external characteristics of the rostrum,
  - structure of apex.
- 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 and we suppose this to be suitable for our purpose.
- Based on expert opinion, we partitioned the set Y of attributes into the seven groups and partially ordered them.



Alveolar fracture

Size

Flattening

Cross section

Others A

Others B

Others C

- Then, for every two groups $D_1, D_2 \subseteq Y$ of the seven groups we add to the set T of our AD-formulas the AD-formula $D_1 \sqsubseteq D_2$.

Analysis and Results

Results

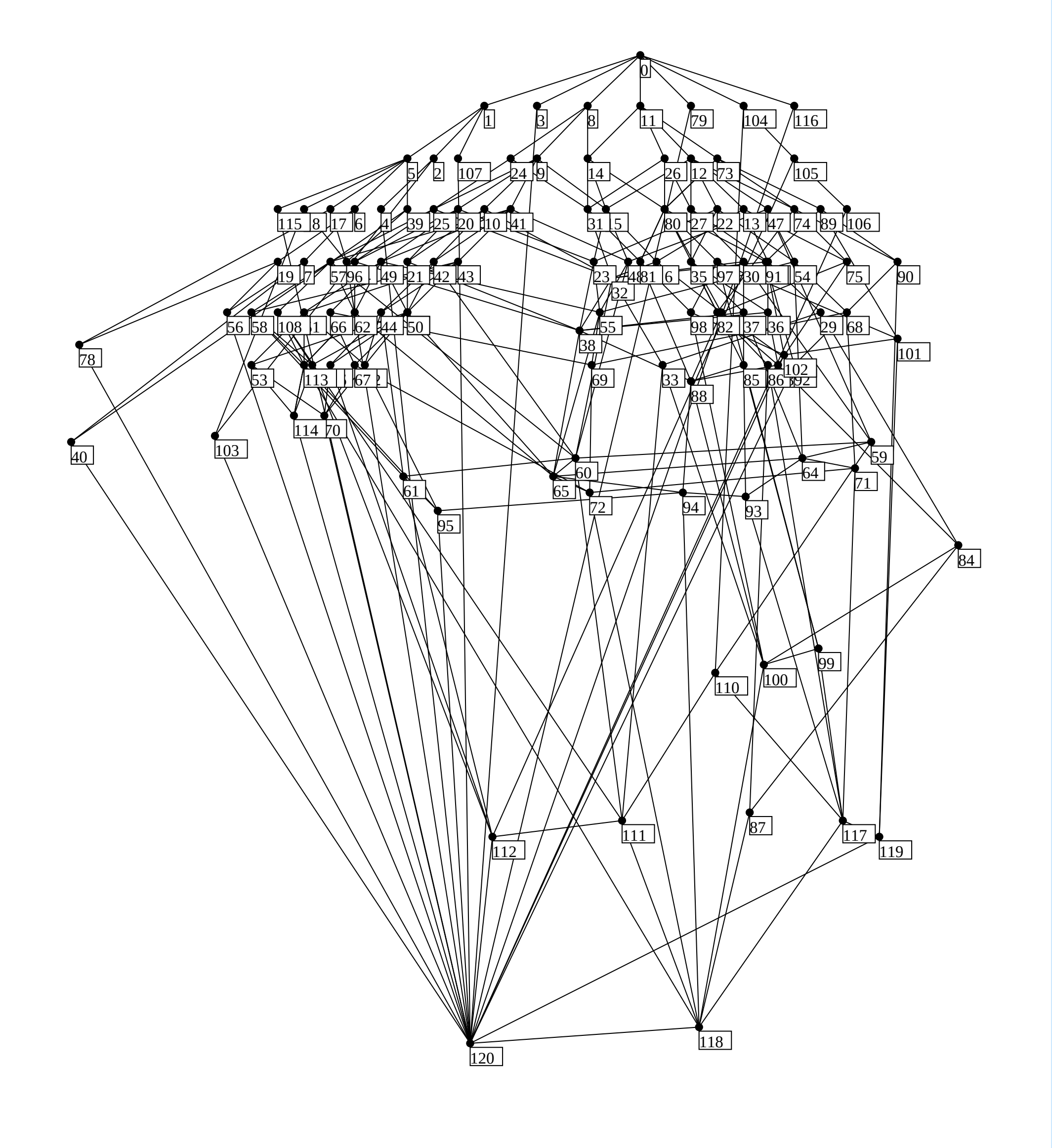


Figure 1. Constrained concept lattice produced by the method. Its formal concepts represent relevant paleobiological taxa.

Interpretation of results

From a paleobiological viewpoint, the major conclusions of the present analysis may be summarized as follows:

- The analysis suggests that the belemnite taxonomic group studied (family Belemnitellidae) is polyphyletic This suggestion contradicts the present opinion according to which this family is monophyletic.
- The analysis indicates the existence of three morphologic trends going to independent, but morphologically similar species in one belemnite genus (so called "parallel evolution"). Up to now, only two parallel morphological lineages were considered inside this genus.
- The analysis revealed the genus origin and derivation of endemic taxa.
- Challenges and extends current knowledge. Stimulates new paleontological research.

International Center for Information and Uncertainty
SSIU 2013: International Summer School "Information and Uncertainty"
http://mcin.upol.cz/SSWIU-2013/

iciu

european social fund in the czech republic

EUROPEAN UNION

MINISTRY OF EDUCATION, YOUTH AND SPORTS

OP Education for Competitiveness

INVESTMENTS IN EDUCATION DEVELOPMENT