

# Detecting Features from Confusion Matrices using Generalized Formal Concept Analysis



Carmen Peláez-Moreno and Francisco J. Valverde-Albacete \*

Dpto. de Teoría de la Señal y de las Comunicaciones.  
Universidad Carlos III de Madrid  
Avda. de la Universidad, 30. Leganés 28911. Spain  
`carmen,fva@tsc.uc3m.es`

**Abstract.** We claim that the confusion matrices of multiclass problems can be analyzed by means of a generalization of Formal Concept Analysis to obtain symbolic information about the feature sets of the underlying classification task. We prove our claims by analyzing the confusion matrices of human speech perception experiments and comparing our results to those elicited by experts.

## 1 Motivation

For  $n, p \in \mathbb{N}$ , let  $G = \{g_i\}_{i=1}^n$  be a set of input labels or *stimuli* and  $M = \{m_j\}_{j=1}^p$  a set of output labels or *responses* for a multiclass classifier task embodied in a human or artificial agent. Consider the joint event “presenting a stimulus  $g_i$  to a classifier and obtaining response  $m_j$ ,” ( $\mathcal{G} = g_i, \mathcal{M} = m_j$ ). A *contingency table* or *confusion matrix* (CM) for the classifier  $C \in \mathbb{N}^{n \times p}$  is a record of the decisions of  $N$  repetitions of such an experiment<sup>1</sup>.

Confusion matrices are rich summaries of how the classifier performed in a *test set*. This is usually transformed into an aggregate figure of merit, like accuracy, or a visual depiction, like a multi-class ROC, thereby losing information about the particular errors the classifier may commit.

We contend that some information about the underlying task can be obtained from the numerical data in the confusion matrix via a special type of biclustering scheme, a *concept lattice*, from Formal Concept Analysis (FCA) [1]. Furthermore, concept lattices allow us both to observe the global behavior of classifiers and to analyze their confusions in detail.

FCA, unfortunately, cannot deal in an automatic way with non-binary incidences, but generalizations of it to cater for the notion of *degree of incidence* have been developed [2,3,4,5,6].

In this paper, we use  $\mathcal{K}$ -Formal Concept Analysis (kFCA) [5,7], which enables the analysis of practical real-valued CM by embedding them into an idempotent

---

\* This work has been supported by Spanish Government-Comisión Interministerial de Ciencia y Tecnología projects TEC2008-02473/TEC and TEC2008-06382/TEC.

<sup>1</sup> We consider here the general case where the labels used in the training speech samples differ from those considered by the recognizer.

semifield  $\mathcal{K}$ —actually a bounded lattice-ordered group [8]—, to try and prove that a concept lattice can elicit a symbolic description of the features being used in the classification process and how they are misused by the classifier.

## 2 Generalized Formal Concept Analysis of Confusion Matrices

**From count matrix to  $\varphi$ -confusion lattice.** To illustrate  $\mathcal{K}$ -Formal Concept Analysis of confusion matrices, consider that of Fig. 1(a). The first design choice is to find an adequate domain to express the strength of confusions. From a count matrix  $N_{GM}$  we may obtain an estimate of the mutual information distribution for the events  $\hat{C}_{GM}$ , like that of Fig. 1(b). A proper choice for the semiring in  $\mathcal{K}$ -Formal Concept Analysis is  $\mathbb{R}_{\max,+}$  (read “completed max-plus”). This is the completed set of reals with the “max” operation used as addition and normal addition as multiplication.

$N$	p	m	t	f	th	k	s
p	150	0	38	7	13	88	0
m	0	201	0	0	0	0	0
t	30	0	193	1	0	28	0
f	4	1	3	199	46	5	4
th	11	0	6	85	114	4	10
k	86	0	45	4	1	138	0
s	0	0	2	5	38	1	170

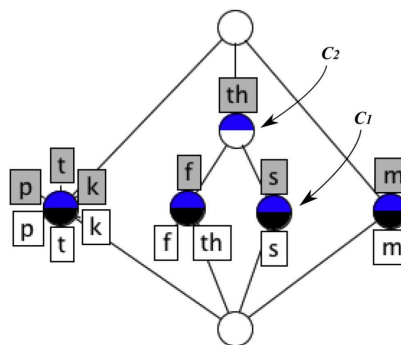
(a)  $N_{GM}$

$\hat{C}$	p	m	t	f	th	k	s
p	2.851	$-\infty$	0.824	-1.717	-0.305	2.155	$-\infty$
m	$-\infty$	4.202	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
t	0.761	$-\infty$	3.401	-4.292	$-\infty$	0.735	$-\infty$
f	-2.213	-3.793	-2.674	3.277	1.683	-1.817	-1.626
th	-0.567	$-\infty$	-1.487	2.236	3.179	-1.953	-0.117
k	2.149	$-\infty$	1.169	-2.424	-3.904	2.905	$-\infty$
s	$-\infty$	$-\infty$	-3.047	-1.826	1.619	-3.928	3.995

(b)  $\hat{C}_{GM}$

	p	m	t	f	th	k	s
p	x	x			x		
m		x					
t	x		x			x	
f				x x			
th				x x			
k	x	x			x	x	
s					x	x	x

(c)  $(G, M, I_{C,\varphi}^+)_{\mathbb{R}_{\max,+}}$



(d)  $\mathfrak{B}(G, M, I_{C,\varphi}^+)_{\mathbb{R}_{\max,+}}$

Fig. 1: example analysis using kFCA: (a) count confusion matrix, obtained from the Miller and Nicely experiments [9] for SNR = 0dB—only phonemes  $G = M = \{/m/, /p/, /t/, /k/, /f/, /s/, /th/\}$  have been retained as both stimuli (left) and responses (above); (b) its mutual information distribution; (c) structural matrix and (d) structural lattice for  $\varphi = 0.056585$ .

For  $n, p \in \mathbb{N}$ , given two sets of stimuli  $G = \{g_i\}_{i=1}^n$ , and responses  $M = \{m_j\}_{j=1}^p$ , and a  $(\overline{\mathbb{R}}_{\max,+})$ -valued matrix  $C \in \overline{\mathbb{R}}_{\max,+}^{n \times p}$ , the triple  $(G, M, C)_{\overline{\mathbb{R}}_{\max,+}}$  is called a  $(\overline{\mathbb{R}}_{\max,+})$ -valued *formal context*, where  $C(i, j) = \lambda$  reads as “stimulus  $g_i$  is confused with response  $m_j$  to degree  $\lambda$ ” and dually “response  $m_j$  is evoked by stimulus  $m_i$  in degree  $\lambda$ ”.

We may associated multi-valued sets of stimuli  $A$  and responses  $B$  by means of a pair of functions  $(\cdot)_{C,\varphi}^+ : \overline{\mathbb{R}}_{\max,+}^n \rightarrow \overline{\mathbb{R}}_{\max,+}^p$  and  ${}_{C,\varphi}^+(\cdot) : \overline{\mathbb{R}}_{\max,+}^p \rightarrow \overline{\mathbb{R}}_{\max,+}^n$  forming a Galois connection [1,7] as follows: define  $\varphi$ -concepts as pairs  $(A, B)_\varphi$  such that  $(A)_{C,\varphi}^+ = B \iff A = {}_{C,\varphi}^+(B)$ . The Basic Theorem of  $\mathcal{K}$ -Formal Concept Analysis asserts that the set of formal  $\varphi$ -concepts is a complete lattice  $\underline{\mathfrak{B}}^\varphi(G, M, C)_{\overline{\mathbb{R}}_{\max,+}}$  (see [5,7] for details). The parameter  $\varphi \in \mathbb{R}$  is called the *threshold of existence* and it describes a *minimum degree of confusion* required for concepts to be considered members of the  $\underline{\mathfrak{B}}^\varphi(G, M, C)_{\overline{\mathbb{R}}_{\max,+}}$ .

**Structural Confusion Lattices.** The  $\varphi$ -concept lattice  $\underline{\mathfrak{B}}^\varphi(G, M, C)_{\overline{\mathbb{R}}_{\max,+}}$  has a huge number of concepts (infinite, in the typical case) and is hard to visualize. Therefore, for each choice of  $\varphi$  deemed interesting, we introduce its *structural (confusion) lattice*  $\underline{\mathfrak{B}}(G, M, I_{C,\varphi}^+)$ , the (standard) confusion lattice of the binary incidence,  $I_{C,\varphi}^+$ , depicting only those concepts above a fixed threshold of existence  $\varphi$ . The following *lattice exploration algorithm* must be carried out once for each choice of  $\varphi$ <sup>2</sup>:

1. Work out the concepts  $\gamma(g_i)_{C,\varphi}^+$  and  $\mu(m_j)_{C,\varphi}^+$  associated to singleton stimuli and responses, respectively.
2. Build a binary incidence  $I_{C,\varphi}^+$  associated to those concepts by adequately comparing them to create the binary context  $(G, M, I_{C,\varphi}^+)$  with the binary incidence  $g_i \langle I_{C,\varphi}^+ \rangle m_i \iff \gamma(g_i)_{C,\varphi}^+ \leq \mu(m_j)_{C,\varphi}^+$ .
3. Use a standard tool for Formal Concept Analysis, called CONEXP [11], to build and visualize the *structural concept lattice at  $\varphi$* ,  $\underline{\mathfrak{B}}(G, M, I_{C,\varphi}^+)$ .

**Structural confusion lattice interpretation.** For a *boolean* confusion matrix  $I$ —such as that of Fig. 1(c)— the triple  $(G, M, I)$  is called a *formal context*, and assumed to encode all information pertaining to the phenomenon being analyzed. Pairs of a particular set of stimuli that are *all* confused with a particular set of responses, and vice versa, are called *formal concepts*. For instance,  $c_1 = (\{\mathbf{/s/}\}, \{\mathbf{/s/}, \mathbf{/th/}\})$  is one such pairs for the context above, and  $c_2 = (\{\mathbf{/s/}, \mathbf{/f/}, \mathbf{/th/}\}, \{\mathbf{/th/}\})$  another. The set of stimuli in a concept is called the *extent* and the set of responses is the *intent* of the concept:  $\{\mathbf{/s/}\}$  and  $\{\mathbf{/s/}, \mathbf{/th/}\}$ , are the extent and intent, respectively of  $c_1$ , meaning stimulus  $\mathbf{/s/}$  is confused with responses  $\mathbf{/s/}$  and  $\mathbf{/th/}$ . To distinguish between stimuli and responses, boldface characters will be used for the former throughout the text.

Concepts are partially ordered by inclusion of extents, or, equivalently, reverse inclusion of intents: if  $(A_1, B_1) \leq (A_2, B_2) \iff A_1 \subseteq A_2 \iff B_1 \supseteq B_2$  we

<sup>2</sup> An on-line demonstration of this can be accessed in [10].

say that the first concept is *more specific* (*less general*) than the second. For instance,  $c_1$  is more specific than  $c_2$ . The Basic Theorem of Formal Concept Analysis asserts that the set of formal concepts of a formal context, as related by this order relation, is a complete lattice called the *concept lattice*  $\mathfrak{B}(G, M, I)$ .

In the Hasse diagram of a confusion lattice, stimulus labels appear *in white boxes just below* the corresponding concept and response labels usually appear *in gray boxes just above*. To diminish visual clutter, instead of completely labeling each node with all labels of either sort *we put the label of each response only in the highest—most abstract—concept it appears, and the label of each stimulus only in the lowest—most specific—concept it appears*. This is the *reduced labeling* shown in Fig. 1(d).

In this labeling scheme, concepts capture the confusions between more phones than those that actually appear attached to the concept. To recover the *confusion extent*, the set of stimuli being confused at a particular concept, we take the union of all *stimulus labels found from the node downwards* in the lattice. Similarly, to build the *confusion intent*, we take the union of all *response labels found from the node upwards* in the lattice. In the example, if we go from  $c_1$  downwards in the lattice collecting stimulus labels (below the nodes) we obtain its extent  $\{/s/\}$ , and if we go upwards we find the labels in its intent,  $/s/$  (above  $c_1$  itself) and  $/th/$  (above  $c_2$ ).

There are two types of complementary, domain-specific information that can be gleaned from a lattice: *specific concept information* and *overall lattice information*. As to the first, the most interesting concepts are the *join-irreducible concepts* (bottom half-filled in black in Fig. 1(d)), and *meet-irreducible concepts* (top half-filled in gray, blue online). Call the rest of the concepts in the example lattice  $c_{ptk} = (\{/p/, /t/, /k/\}, \{/p/, /t/, /k/\})$ ,  $c_m = (\{/m/\}, \{/m/\})$ , and  $c_{fth} = (\{/f/, /th/\}, \{/f/, /th/\})$ . The set of join-irreducibles is  $\mathcal{J} = \{c_{ptk}, c_{fth}, c_1, c_m\}$ , and the set of meet-irreducibles is  $\mathcal{M} = \{c_{ptk}, c_{fth}, c_1, c_2, c_m\}$ . In confusion lattices, join-irreducibles, always annotated with a stimulus label, are the concepts to peruse in order to know what responses each individual stimulus invokes. And likewise, meet-irreducibles, annotated with response labels, show what set of stimuli evokes a particular response.

Regarding overall information about the matrix, consider the three separate sublattices of Fig. 1(d) including, the first, concepts top, bottom and  $c_{ptk}$ , to the left; the second, concepts top, bottom and  $c_m$ , to the right; and the third, concepts top, bottom,  $c_1$ ,  $c_2$  and  $c_{fth}$ , at the center. Concepts in different sublattices are incomparable except for the top and bottom. We will say that such sublattices are *adjoined factor sublattices* of the confusion lattice. Notice that *stimuli and responses that lie in adjoined factor sublattices are never confused*, hence the presence of some adjoined sublattices in the confusion lattice is essentially the lattice-theoretic manifestation of as many different *virtual channels* in the classifier system. By this we mean that the classifier succeeds in conveying definite information from input to output without error. In the example, the channels for  $\{/m/\}$ ,  $\{/f/, /s/, /th/\}$  and  $\{/p/, /t/, /k/\}$  seem evident.

### 3 The elicitation of symbolic knowledge from phonetic confusion matrices

Confusion matrices became a key tool for the analysis of human speech perception since the Miller & Nicely experiments [9]. After a thorough analysis their major conclusion was that phone recognition is grounded on *hierarchical categorical discrimination*, that is, English consonant sounds form groups identified in terms of hierarchical clusters of *articulatory features*. They introduced the notion of *virtual articulatory communication channels*, according to such clusterings, and posited that the channels were characterized by five distinctive acoustic-articulatory features, namely, *voicing*, *nasality*, *affrication*, *duration* and *place*. In the following we will try to reproduce these results using  $\mathcal{K}$ -Formal Concept Analysis of perceptual confusion matrices.

To assess the complexity of the structural confusion lattices, we have worked out the *concept counts* defined at different thresholds  $\varphi$ . The concept count represents the number of nodes present in the corresponding structural lattice, and provides a rough measure of the complexity of the resulting representation. Small values of the threshold  $\varphi$  bring into the picture non-systematic, difficult to explain, confusions evident in the analysis of structural lattices with a high number of nodes. On the contrary, if a larger value of  $\varphi$  is chosen, the number of concepts will be reduced offering a much simpler structural lattice showing the most prominent confusions.

The different plots of Fig. 2 represent the evolution of the number of concepts for several Signal to Noise Ratios (SNR) and a full 200 – 6 500Hz band for the Miller and Nicely experiments. We can clearly notice how the maximum number of concepts attained by each plot is inversely related to the SNR of the emitted syllables. Therefore, the confusion lattice analysis is capturing the complexity of the CM that corresponds to each SNR: as the speech signal quality gets better the errors become more systematic or *structured* and therefore the number of concepts decreases.

The evolution of the number of concepts suggests a method for describing the information in structural lattices:

1. Begin by observing the most salient properties of the system, that is, those lattices obtained with higher values of the threshold  $\varphi$ .
2. Subsequently, try to bring more detail into the picture by sweeping from higher to lower values of  $\varphi$  (from right to left in figures 2).

We thus obtain a sequence of structural lattices starting from the least complex—with the least number of concepts—and gradually increasing the complexity as new concepts appear.

Figure 3 is a typical structural lattice for the Miller & Nicely experiments at 0dB and a particular  $\varphi$  where six adjoined factor sublattices can be observed. To the left the *voiced* phonemes with */m/* and */n/*, the *nasals*, even represented in two separate sublattices. To the right, three sublattices representing *unvoiced* phonemes: the *(oral) stops* */p-t-k/*, *fricative* */sh/* and the *rest of fricatives*.

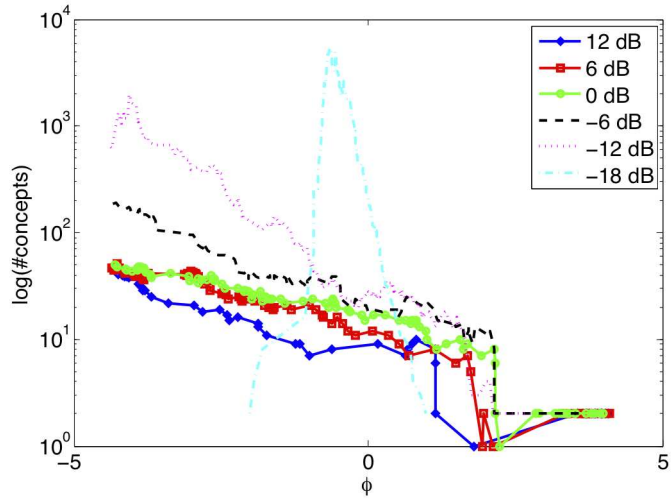


Fig. 2: (color online) number of concepts vs.  $\varphi$  for HSR confusion matrices (data from [9]). The maximum number of concepts attained by each plot is inversely related to the SNR of the emitted syllables.

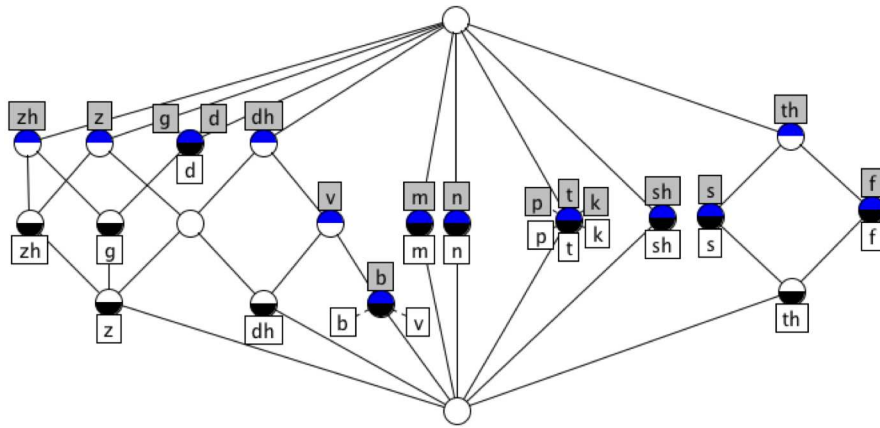


Fig. 3: phonetic confusion lattice at  $\varphi = 0.11716$  and SNR = 0dB (data from [9]).

Hence our hypothesis is that *adjoined factor sublattices in a structural confusion lattice reflect virtual feature transmission channels*. Since this has to be contrasted to the Miller and Nicely findings, a direct method to elicit what phonetic knowledge the sublattices reflect would be to show the stimuli and responses in each lattice. This would demand, afterwards, the concurrence of a phonetic expert to elicit the features.

However, a clustering of phonemes in terms of their *voicing*, *manner* and *place of articulation* can also be cast into an Formal Concept Analysis concept lattice as shown at the top of Fig. 4(a)—showing two phonemes for each feature that correspond to *unvoiced* (on the left) and *voiced* sounds (on the right). We may use this knowledge to *label* the structural lattices automatically by selecting the feature label adequate for each phonemic concept extent.

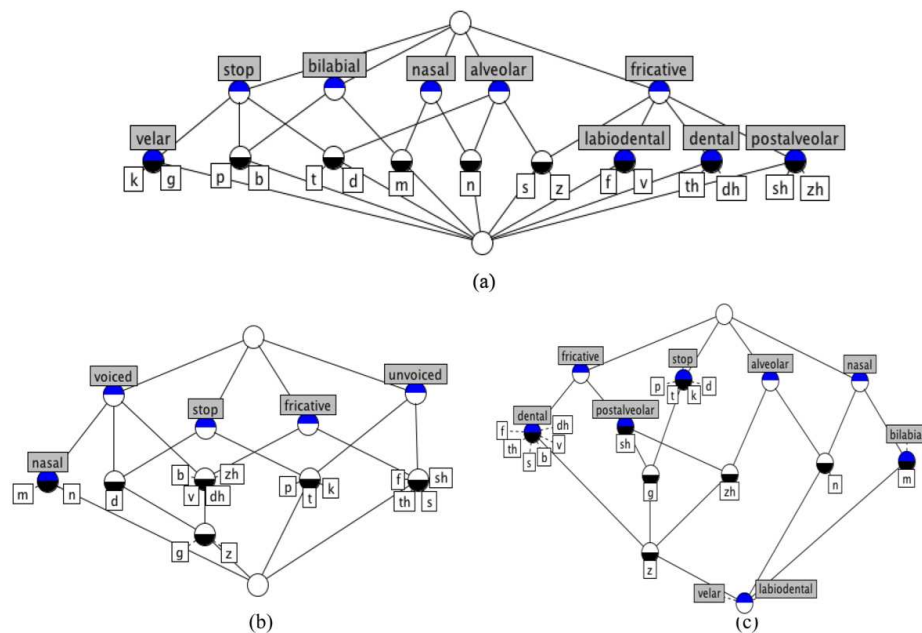


Fig. 4: phonemes vs. articulatory features concept lattices: (a) canonical clustering with unvoiced sound concepts on the left and voiced ones on the right; (b) clustering elicited from the confusion lattice of Fig. 3; (c) *id.* including the *place* feature.

The lattices at the bottom of figure 4 demonstrate which part of the clustering can be actually elicited from the confusion matrix in Fig. 3. *Voicing*, *manner of articulation*—stop, nasal, fricative—can be obtained almost without error as shown in Fig. 4(b), although clear mismatches between the canonical and empirically induced representations can be observed: /b/ and /g/ are perceived as *fricative*, /z/ as *stop*. But *place of articulation* is hopeless as Fig. 4(c) shows. In fact, *labiodental* and *velar* can not be defined at all. This agrees in all with

the Miller & Nicely conclusions, except for result in *place* of articulation, which has often been disputed.

## 4 Conclusions

We have provided evidence that  $(\overline{\mathbb{R}}_{\max,+})$ -Formal Concept Analysis of confusion data for a multiple-classification task can identify features present in the classification act. Since our generalization considers non-binary matrices in the analysis, it is ideally suited to the analysis of count confusion matrices.

After a preprocessing stage which amounts to considering the confusion matrix as a joint-distribution of input stimuli and output responses, we are able to pinpoint adjoined sublattices in the concept lattice which we take as evidence that some definite feature is being transmitted.

For assessment purposes, we also elicited these features using conventional articulatory acoustic knowledge. Our results agree with expert-drawn conclusions in all but the most contested ones, what we take to reflect the robustness of the elicitation process.

## References

1. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations. Springer, Berlin, Heidelberg (1999)
2. Burusco, A., Fuentes-González, R.: The study of the L-fuzzy Concept Lattice. *Mathware and Soft Computing* **1** (1994) 209–218
3. Bělohávek, R.: Lattice generated by binary fuzzy relations. *Tatra Mt. Mathematical Publications* **16** (1999) 11–19
4. Krajci, S.: A generalized concept lattice. *Logic Journal of IGPL* **13** (2005) 543
5. Valverde-Albacete, F.J., Peláez-Moreno, C.: Towards a generalisation of Formal Concept Analysis for data mining purposes. In Missaoui, R., Schmid, J., eds.: *Concept Lattices. Proceedings of the International Conference on Formal Concept Analysis, ICFCA06, (Dresden, Germany)*. Volume 3874 of LNAI., Berlin, Heidelberg, Springer (2006) 161–176
6. Medina, J., Ojeda-Aciego, M., no, J.R.C.: Formal concept analysis via multi-adjoint concept lattices. *Fuzzy Sets and Systems* **160** (2009) 130–144
7. Valverde-Albacete, F.J., Peláez-Moreno, C.: Galois connections between semimodules and applications in data mining. In Kusnetzov, S., Schmidt, S., eds.: *Formal Concept Analysis. Proceedings of the 5th International Conference on Formal Concept Analysis, ICFCA 2007, Clermont-Ferrand, France*. Volume 4390 of LNAI., Springer (2007) 181–196
8. Cuninghame-Green, R.: *Minimax Algebra*. Number 166 in *Lecture notes in Economics and Mathematical Systems*. Springer (1979)
9. Miller, G.A., Nicely, P.E.: An analysis of perceptual confusions among some English consonants. *Journal of the Acoustic Society of America* **27** (1955) 338–352
10. Esteban-Alonso, V., Valverde-Albacete, F.J., Peláez-Moreno, C.: Generalised Formal Concept Analysis demo. On-line (date last viewed 28/02/2010) (2008)
11. Yevtushenko, S.A.: System of data analysis "Concept Explorer". In: *Proceedings of the 7th national conference on Artificial Intelligence KII-2000*. (2000) 127–134 (In Russian) <http://sourceforge.net/projects/conexp>.