FCA-Based Models and a Prototype Data Analysis System for Crowdsourcing Platforms

Dmitry I. Ignatov[1], Alexandra Yu. Kaminskaya[1,2], Anastasya A. Bezzubtseva[1,2], Andrey V. Konstantinov[1], and Jonas Poelmans[1,3]

[1] National Research University Higher School of Economics, Russia, 101000, Moscow, Myasnitskaya str., 20
dignatov@hse.ru
http://www.hse.ru
[2] Witology
http://www.witology.com
[3] KU Leuven, Belgium

Abstract. This paper considers a data analysis system for collaborative platforms which was developed by the joint research team of the National Research University Higher School of Economics and the Witology company. Our focus is on describing the methodology and results of the first experiments. The developed system is based on several modern models and methods for analysing of object-attribute and unstructured data (texts) such as Formal Concept Analysis, multimodal clustering, association rule mining, and keyword and collocation extraction from texts.

Keywords: collaborative and crowdsourcing platforms, Data Mining, Formal Concept Analysis, multimodal clustering.

1 Introduction and Related Work

The success of modern collaborative technologies is marked by the appearance of many novel platforms for holding distributed brainstorming or carrying out so called "public examination". There are a lot of such crowdsourcing companies in the USA (Spigit [1], BrightIdea [2], InnoCentive [3] etc.) and Europe (Imaginatik [4]). There is also the Kaggle platform [5] which is most beneficial for data practitioners and companies that want to select the best solutions for their data mining problems. In 2011 Russian companies launched business in that area as well. The two most representative examples of such Russian companies are Witology [6] and Wikivote [7]. The reality as yet is far away from technological breakthrough, though some all-Russian projects have already been finished successfully (for example, Sberbank-21, National Entrepreneurial Initiative-2012 [8] etc.). The core of such crowdsourcing systems is a socio-semantic network [9,10,11,12], which data requires new approaches to analyze. This paper is devoted to the new methodological base for the analysis of data generated by collaborative systems, which uses modern data mining and artificial intelligence models and methods. As a rule, while participating in a project, users of such

H.D. Pfeiffer et al. (Eds.): ICCS 2013, LNAI 7735, pp. 173–192, 2013.

crowdsourcing platforms [13] discuss and solve one common problem, propose their ideas and evaluate ideas of each other as experts. Finally, as a result of the discussion and ranking of users and their ideas we get the best ideas and users (their generators). For deeper understanding of users's behavior, developing adequate ranking criteria and performing complex dynamic and statistic analyses, special means are needed. Traditional methods of clustering, community detection and text mining need to be adapted or even fully redesigned. Moreover, these methods require ingenuity for their effective and efficient use (finding non-trivial results). We briefly describe models of data used in crowdsourcing projects in terms of Formal Concept Analysis (FCA) [14]. Furthermore, we present the collaborative platform data analysis system CrowDM (Crowd Data Mining), its architecture and methods underlying the key steps of data analysis.

The remainder of the paper is organized as follows. Section 2 contains descriptions of the Witology crowdsourcing methodology and Sberbank-21 project. In section 3 we describe some key notions from FCA, our data and methods. In section 4 we discuss the analysis scheme of the developed system. In section 5 we present the results of our first experiments with the Sberbank-21 data. Section 6 concludes our paper and describes some possible directions for future research.

2 Witology Crowdsourcing Methodology and Projects

One proverb says "Two heads are better than one", but crowdsourcing projects may take several thousands of heads. The term "crowdsourcing" is a portmanteau of "crowd" and "outsourcing", coined by Jeff Howe in 2006 [13]. There is no general definition of crowdsourcing, but it takes some specific features. Crowdsourcing is a process, both online and offline, that includes task solving by a distributed and large group of people who are usually from different organisations, and not necessarily paid by money for their work.

We shortly describe the methodology of the Witology crowdsourcing company, Witodology, considering as an example its Sberbank-21 project. Note that the company clearly says that Witodology is based on the notion of socio-sematic networks [11,12]. In 2011, from October till November, Witology and Sberbank launched one of the first successful crowdsourcing projects in Russia: Sberbank-21. The Russian company Sberbank is the largest and oldest Russian bank which history started in 1841; its name can be formally translated into English as "Savings Bank of the Russian Federation". The project was devoted to the theme "Office of Sberbank in 2012 (Office SB-2012)". The main project topics include "Office of SB-21 for private clients", "Office SB-21 for individual entrepreneurs", "Office SB-21 for small businesses", "Internal filling of "physical" SB-21 office", and "Internal filling of "virtual" SB-21 office". The goal was formulated as "a selection of the best well-founded and innovative solutions for the formulated tasks, which include format and content of the proposed solutions as well as reasons for their appearance and adoption". During the preliminary test, 450 experts were selected out of 5198 people. Amongst them 33% were women and 67% were man, 21% were Sberbank employees and 79% of them were either clients or other interested

persons. The main stages of the project were "Solution's generation", "Selection of similar solutions", "Generation of counter-solutions", "Total voting", "Solution's improvements", "Solution's stock", "Final improvements" and finally "Solution's review".

In total, 222 experts proposed 1581 solutions for 15 tasks which were grouped into 5 topics. After "Selection of similar solutions" by participants, 24 574 analytical operations including comparison, clustering, and filtering of ideas were performed. As a result of this stage, 589 solutions were selected. The stage "Total voting" resulted in the selection of 182 solutions. After the stage "Solution's stock" 75 solutions were left. From 15 remaining solutions after "Solution's review" 3 solutions were nominated as the best ones.

The first stage "Solution's generation" is performed individually by each user. A key difference between traditional brainstorming and the "Solution's generation" stage is that nobody can see or listen to ideas of other participants. The main similarity is the absence of criticism which was moved to later stages. In the "Selection of similar solutions" phase participants are selecting similar ideas (solutions) and their aggregated opinions are transformed to clusters of similar ideas.

For Sberbank-21 projects all the proposed solutions were divided into 15 clusters (tasks), three per topic: (Sberbank and private client: interface in 2021?, Sberbank-21 service for every 21 years old person?, Unique service of 2021 for private clients?), (Sberbank and entrepreneur: interface in 2021?, Service in 2021 for startupers?, Unique service in 2021 for entrepreneurs?), (Sberbank and small businesses: interface in 2021?, Service in 2021 for new businesses?, Unique service for small businesses?), (What will disappear in the "physical" office of SB-21?, What will change in the "physical" office SB-21?, What will appear in the "physical" office of SB-21?) (What will disappear in the online office of SB-21?, What will change in the online office of SB-21?, What will appear in the online office of SB-21?).

Counter-solutions generation includes criticism (pros and cons) and evaluation of proposed ideas by communication between an author and experts. During this stage an idea's author can invite other experts to his team taking into account their contribution to discussion and criticism. Total voting is performed by evaluation of each proposed idea by all users in terms of their attitude and quality levels of the solution (marks are integers between -3 and 3). Two stages, i.e. "Solution's improvements" and "Final improvements", involve active collaboration by experts and authors who improve their solutions together.

The system calculates 10 user's ratings based on their activity; among them are "Popularity","Social capital", "Performance", "Gamer", "Actor", "Judge", "Commenter", "Importance", "Influence", and "Reputation". For texts the company uses the following rates: "Significance", "Influence", "Popularity", "Quality", "Attitude" and also "Reputation".

Solution's stock is one of the most interesting game stages of the project when all participants with a positive reputation rate accumulated on the previous stages take money in internal currency "wito" and can perform stock trade.

The solutions with the highest price become winners. Finally, during the review of solutions, experts with high reputation make their final evaluations based on several criteria in -3 to 3 scale: "Solution efficiency", "Solution originality", "Solution performability", and "Return on investment".

We have to tell the reader about some best solutions but cannot go into details because of non-disclosure requirements. For example, for the task "What will disappear in the "physical" SB-21 office?" the best solution said that "You shouldn't to fill in the same documents several times". For the task "What will change in the SB-21 "physical" office" the solution was "Changing the access mode for a safe deposit box to biometric data for corporate clients and Near Based Communication (NFC) chips for private clients". And the answer for the question "What will appear in the SB-21 "physical" office?" was "Videowall", a sort of interface for communication with a distant operator and making regular financial transactions.

3 Mathematical Models and Methods

At the initial stage of collaborative platform data analysis two data types were identified: data without using keywords (links, evaluations, user actions) and data with keywords (all user-generated content). These two data types totally correspond with two components of a socio-semantic network. For the analysis of the 1st type of data (with keywords) we suggest to apply Social Network Analysis (SNA) methods, clustering (biclustering and triclustering [15,16,17], spectral clustering), FCA (concept lattices, implications, association rules) and its extensions for multimodal data, triadic, for instance [18]; recommender systems [19,20,21,22] and statistical methods of data analysis [23] (the analysis of distributions and average values).

3.1 Formal Concept Analysis and OA-biclustering

Methods described in this paper are mainly from the multimodal clustering block at the analysis scheme (see fig. 2). The protagonists of crowdsourcing projects (and corresponding collaborative platforms) are platform users (project participants). We consider them as *objects* for analysis. More than that, each object can (or cannot) possess a certain set of *attributes*. The user's attributes can be: topics which the user discussed, ideas which he generated or voted for, or even other users. The main instrument for analysis of such object-attribute data is FCA [14]. Let us give formal definitions. *The formal context* in FCA is a triple $\mathbb{K} = (G, M, I)$, where G is a *set of objects*, M is a *set of attributes*, and the relation $I \subseteq G \times M$ shows which object possesses which attribute. For any $A \subseteq G$ and $B \subseteq M$ one can define *Galois operators*:

$$\begin{aligned} A' &= \{m \in M \mid gIm \text{ for all } g \in A\}, \\ B' &= \{g \in G \mid gIm \text{ for all } m \in B\}. \end{aligned} \tag{1}$$

The operator $''$ (applying the operator $'$ twice) is a *closure operator*: it is idempotent ($A'''' = A''$), monotonous ($A \subseteq B$ implies $A'' \subseteq B''$) and extensive ($A \subseteq A''$). The set of objects $A \subseteq G$ such that $A'' = A$ is called closed. The same properties hold for closed attribute sets, i.e. subsets of the set M. A couple (A, B) such that $A \subset G$, $B \subset M$, $A' = B$ and $B' = A$, is called *formal concept* of a context K. The sets A and B are closed and called *extent* and *intent* of a formal concept (A, B) respectively. For the set of objects A the set of their common attributes A' describes the similarity of objects of the set A, and the closed set A'' is a cluster of similar objects (with the set of common attributes A'). The relation "to be more general concept" is defined as follows: $(A, B) \geq (C, D)$ iff $A \subseteq C$. We denote by $\mathfrak{B}(G, M, I)$ the set of all concepts of a formal context $\mathbb{K} = (G, M, I)$. The concepts of a formal context $\mathbb{K} = (G, M, I)$ ordered by extensions inclusion form a lattice, which is called a *concept lattice*. For its visualization a *line diagram* (Hasse diagram) can be used, i.e. the cover graph of the relation "to be a more general concept".

To represent datasets with numerical (e.g., age, word frequency, number of comments) and categorical (e.g., gender, job) attributes there are many-valued contexts. *A many-valued context* (G, M, W, I) consists of sets G, M and W and a ternary relation $I \subseteq G \times M \times W$ for which it holds that

$$(g, m, w) \in I \text{ and } (g, m, v) \in I \Rightarrow w = v.$$

The elements of G are still called objects, those of M (many-valued) attributes and the elements of W attribute values. Sometimes we write $m(g) = w$ to show that the object g has the value w of the attribute m.

We can transform the many-valued context into a one-valued one by means of **conceptual scaling** [14].

In the worst case (Boolean lattice) the number of concepts is equal to $2^{\{\min|G|,|M|\}}$, thus, for large contexts, FCA can be used only if the data is sparse. Moreover, one can use different ways of reducing the number of formal concepts (choosing concepts by stability [24] index or extent size). The alternative approach is a relaxation of the definition of formal concept as maximal rectangle in object-attribute matrix which elements belong to the incidence relation. One of such relaxations is the notion of object-attribute bicluster [16]. If $(g, m) \in I$, then (m', g') is called *object-attribute bicluster* with the density $\rho(m', g') = |I \cap (m' \times g')|/(|m'| \cdot |g'|)$.

The main features of OA-biclusters are listed below:

1. For any bicluster $(A, B) \subseteq 2^G \times 2^M$ it is true that $0 \leq \rho(A, B) \leq 1$.
2. OA-bicluster (m', g') is a formal concept iff $\rho = 1$.
3. If (m', g') is a bicluster, then $(g'', g') \leq (m', m'')$.

Let $(A, B) \subseteq 2^G \times 2^M$ be a bicluster and ρ_{min} be a non-negative real number such that $0 \leq \rho_{\min} \leq 1$, then (A, B) is called *dense*, if it fits the constraint $\rho(A, B) \geq \rho_{\min}$. The above mentioned properties show that OA-biclusters differ from formal concepts since unit density is not required. Graphically it means that not all the cells of a bicluster must be filled by a cross (see fig. 1). Besides

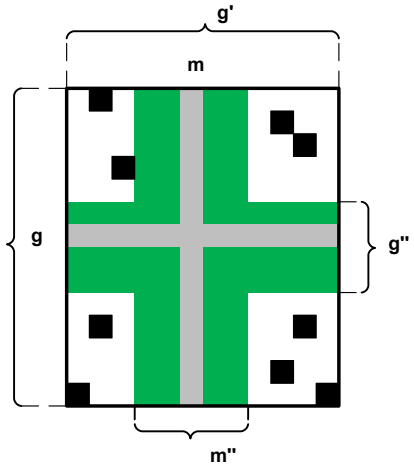

Fig. 1. OA-bicluster

formal lattice construction and visualization by means of Hasse diagrams one can use implications and association rules for detecting attribute dependencies in data. Then, using the obtained results, it is easy to form recommendations (for example, offering users the most interesting discussions for them). Furthermore, structural analysis can be performed and then used for finding communities. Statistical methods are helpful for frequency analysis of the different users' activities. Almost all of the above mentioned methods can be applied to data containing users' keywords (in this case they become attributes of a user).

3.2 Triadic FCA and OAC-triclustering

To deal with three-way data within FCA, an extension to Triadic Concept Analysis (TCA) was proposed by Lehman and Wille [25,26]. In [18] the author introduced the TRIAS algorithm for mining all frequent triconcepts from 3-dimensional data and applied it to the popular Bibsonomy (users-tags-papers) dataset. Voutsadakis [27] extended triadic concept analysis to n-dimensional contexts.

There exist some known difficulties in mining binary data, such as a lack of fault tolerance, an explosion of the number of patterns leading to large computational complexity and to many small patterns that appear to be false positive observations. In triadic or n-ary contexts these problems are seriously aggravated. To cope with these issues, several techniques have been introduced for faster selection of interesting patterns. For example, there is an extended box clustering approach [28] and triadic concept factors [29]. Another approach, called constraint-based mining, also scales up to n-ary relations and is discussed in [30] and [31]. In [17] we also proposed a new triclustering approach for mining so-called (dense) OAC-triclusters, where OAC stands for Object Attribute

Condition. This algorithm has a better theoretical time complexity than existing exact algorithms like TRIAS and is therefore better suited for very large datasets. Moreover, during experimentations with the bibsonomy dataset, we found the number of triclusters generated by our algorithm to be significantly lower than the number of triconcepts extracted by TRIAS. Manual validation of the extracted tricommunities revealed that the majority of them was meaningful.

A triadic context $\mathbb{K} = (G, M, B, Y)$ consists of sets G (objects), M (attributes), and B (conditions), and ternary relation $Y \subseteq G \times M \times B$. An incidence $(g, m, b) \in Y$ shows that the object g has the attribute m under condition b.

For convenience, a triadic context is denoted by (X_1, X_2, X_3, Y). A triadic context $\mathbb{K} = (X_1, X_2, X_3, Y)$ gives rise to the following diadic contexts

$$\begin{aligned} \mathbb{K}^{(1)} &= (X_1, X_2 \times X_3, Y^{(1)}), \\ \mathbb{K}^{(2)} &= (X_2, X_1 \times X_3, Y^{(2)}), \\ \mathbb{K}^{(3)} &= (X_3, X_1 \times X_2, Y^{(3)}), \end{aligned} \tag{2}$$

where $gY^{(1)}(m, b) :\Leftrightarrow mY^{(1)}(g, b) :\Leftrightarrow bY^{(1)}(g, m) :\Leftrightarrow (g, m, b) \in Y$. The derivation operators (primes or concept-forming operators) induced by $\mathbb{K}^{(i)}$ are denoted by $(.)^{(i)}$. For each induced dyadic context we have two kinds of derivation operators. That is, for $\{i, j, k\} = \{1, 2, 3\}$ with $j < k$ and for $Z \subseteq X_i$ and $W \subseteq X_j \times X_k$, the (i)-derivation operators are defined by:

$$\begin{aligned} Z \mapsto Z^{(i)} &= \{(x_j, x_k) \in X_j \times X_k | x_i, x_j, x_k \text{ are related by Y for all } x_i \in Z\} \\ W \mapsto W^{(i)} &= \{x_i \in X_i | x_i, x_j, x_k \text{ are related by Y for all } (x_j, x_k) \in W\}. \end{aligned} \tag{3}$$

Formally, a triadic concept of a triadic context $\mathbb{K} = (X_1, X_2, X_3, Y)$ is a triple (A_1, A_2, A_3) of $A_1 \subseteq X_1, A_2 \subseteq X_2, A_3 \subseteq X_3$, such that for every $\{i, j, k\} = \{1, 2, 3\}$ with $j < k$ we have $A_i^{(i)} = (A_j \times A_k)$. For a certain triadic concept (A_1, A_2, A_3), the components A_1, A_2, and A_3 are called the extent, the intent, and the modus of (A_1, A_2, A_3). It is important to note that for interpretation of $\mathbb{K} = (X_1, X_2, X_3, Y)$ as a three-dimensional cross table, according to our definition, under suitable permutations of rows, columns, and layers of the cross table, the triadic concept (A_1, A_2, A_3) is interpreted as a maximal cuboid full of crosses. The set of all triadic concepts of $\mathbb{K} = (X_1, X_2, X_3, Y)$ is called the concept trilattice and is denoted by $\mathfrak{T}(X_1, X_2, X_3, Y)$.

To simplify notation, we denote by $(.)'$ all prime operators, as it is usually done in FCA. For our purposes consider a triadic context $\mathbb{K} = (G, M, B, Y)$ and introduce primes, double primes and box operators for particular elements of G, M, B, respectively. In what follows, we write g' instead of $\{g\}'$ for 1-set $g \in G$ and similarly for $m \in M$ and $b \in B$: m' and b'.

We do not use double primes, because of their rigid structure; they do not tolerate exceptions like some amount of missing pairs. To allow missing pairs in the operators results we introduce box operators:

$$\begin{aligned} g^{\square} &= \{\ g_i \mid (g_i, b_i) \in m' \textit{ or } (g_i, m_i) \in b'\} \\ m^{\square} &= \{\ m_i \mid (m_i, b_i) \in g' \textit{ or } (g_i, m_i) \in b'\} \\ b^{\square} &= \{\ b_i \mid (g_i, b_i) \in m' \textit{ or } (m_i, b_i) \in g'\}. \end{aligned} \tag{4}$$

Table 1. Prime and double prime operators of 1-sets

Prime operators of 1-sets	Their double prime counterparts
$m' = \{ (g,b) \mid (g,m,b) \in Y\}$	$m'' = \{ \tilde{m} \mid (g,b) \in m' \quad and \quad (g,\tilde{m},b) \in Y\}$
$g' = \{ (m,b) \mid (g,m,b) \in Y\}$	$g'' = \{ \tilde{g} \mid (m,b) \in g' \quad and \quad (\tilde{g},m,b) \in Y\}$
$b' = \{ (g,m) \mid (g,m,b) \in Y\}$	$b'' = \{ \widetilde{b} \mid (g,m) \in b' \quad and \quad (g,m,\tilde{b}) \in Y\}$

Let $\mathbb{K} = (G, M, B, Y)$ be a triadic context. For a certain triple $(g, m, b) \in Y$, the triple $T = (g^{\square}, m^{\square}, b^{\square})$ is called a *OAC-tricluster* based on box operators.

The density of a certain tricluster (A, B, C) of a triadic context $\mathbb{K} = (G, M, B, Y)$ is given by the fraction of all triples of Y in the tricluster, that is $\rho(A, B, C) = \frac{|I \cap A \times B \times C|}{|A||B||C|}$.

The tricluster $T = (A, B, C)$ is called *dense* if its density is greater than a predefined minimal threshold, i.e. $\rho(T) \geq \rho_{min}$. For a given triadic context $\mathbb{K} = (G, M, B, Y)$ we denote by $\mathbf{T}(G, M, B, Y)$ the set of all its (dense) triclusters.

The main features of OAC-triclusters are listed below:

1. For every triconcept (A, B, C) of a triadic context $\mathbb{K} = (G, M, B, Y)$ with nonempty sets A, B, and C we have $\rho(A, B, C) = 1$.
2. For every triclucter (A, B, C) of a triadic context $\mathbb{K} = (G, M, B, Y)$ with nonempty sets A, B, and C we have $0 \leq \rho(A, B, C) \leq 1$.

Proposition 1. *Let $\mathbb{K} = (G, M, B, Y)$ be a triadic context and $\rho_{min} = 0$. For every $T_c = (A_c, B_c, C_c) \in \mathfrak{T}(G, M, B, Y)$ there exists a tricluster $T = (A, B, C) \in \mathbf{T}(G, M, B, Y)$ such that $A_c \subseteq A, B_c \subseteq B, C_c \subseteq C$.*

In the table 2 we have $3^3 = 27$ formal triconcepts, 24 with $\rho = 1$ and 3 void triconcepts with $\rho = 0$ (they have either emptyset of users or ideas or tags). Although the data is small, we have 27 patterns to analyze (maximal number of triconcepts for the context size $3 \times 3 \times 3$); this is due to the data being the power set triadic context. We can conclude that users u_1, u_2, and u_3 share almost the same sets of tags and resources. So, they are very similar in terms of $(term, idea)$ shared pairs and it is convenient to reduce the number of patterns describing this data from 27 to 1. The tricluster $T = (\{u_1, u_2, u_3\}, \{t_1, t_2, t_3\}, \{i_1, i_2, i_3\})$ with $\rho = 0.89$ is exactly such a reduced pattern, but its density is slightly less than 1. Each of the triconcepts from $\mathfrak{T} = \{(\emptyset, \{t_1, t_2, t_3\}, \{i_1, i_2, i_3\}), (\{u_1\}, \{t_2, t_3\}, \{i_1, i_2, i_3\}), \ldots (\{u_1, u_2, u_3\}, \{t_1, t_2\}, \{i_3\})\}$ is contained, w.r.t. component-wise set inclusion, in T.

Table 2. A toy example with Witology data for users $\{u_1, u_2, u_3\}$ describing ideas $\{i_1, i_2, i_3\}$ by terms $\{t_1, t_2, t_3\}$

	t_1	t_2	t_3
u_1		×	×
u_2	×	×	×
u_3	×	×	×

i_1

	t_1	t_2	t_3
u_1	×	×	×
u_2	×		×
u_3	×	×	×

i_2

	i_1	i_2	i_3
u_1	×	×	×
u_2	×	×	×
u_3	×	×	

i_3

3.3 Socio-semantic Networks for Crowdsourcing

One of the possible models for crowdsourcing platforms is the so called *socio-semantic network* [12]. *A social network* is usually modeled as a weighted multi-graph

$$G = \{V, E_1, \ldots, E_k; \pi, \delta_1, \ldots, \delta_k\},$$

where

- V represents members of the network or crowdsourcing platform,
- $E_1, \ldots, E_k \subset V \times V$ denote different relations between the members, e.g. being a friend, follower, relative, co-worker etc.
- $\pi : V \to \Pi$ is a *user profile* function, which stores personal information about the network members.
- $\delta_i : E_i \to \Delta_i$ $(i \in \{1, \ldots, k\})$ keeps parameters and details of the corresponding relation.

The model of the content has a very similar definition. It is a multi-graph

$$C = \{T, R_1, \ldots, R_m; \theta, \gamma_1, \ldots, \gamma_m\},$$

where

- T stands for the set of all elements of the generated content, e.g. posts, comments, evaluations, tags etc.
- $R_1, \ldots, R_m \subset T \times T$ denote different relations on the content, e.g. being a reply on, have the same subject, etc.
- $\theta : T \to \Theta$ stores parameters of the content;
- $\gamma_i : R_i \to \Gamma_i$ $(i \in \{1, \ldots, k\})$, similarly, keeps parameters and details of the corresponding relation.

The basic connections between the social graph and the content are defined by the authorship relation $A \subset V \times T$.

One can also consider other kinds of connections between users and generated content items, but usually all of them could be modeled via introducing a new type of content. For example, the relation *John is interested in post "Announcement"* could be modeled by introducing a new content node *interest evidence*, which points to "Announcement" (use the corresponding relation R_i here) and is authored by John.

Since we deal with binary relations between users and users, users and items, and items and items, it is easy to turn the socio-semantic based models to FCA-language and possibly, by so doing, obtain some benefits for finding communities, groups of interests and making recommendations. For visualising socio-sematic networks refer to [32].

3.4 FCA-Based Models for Crowdsourcing Data

From socio-semantic networks we move on to formal concept analysis. It is easy to show that all key crowdsourcing platform data can be described in FCA terms by means of formal contexts (single-valued, multi-valued or triadic).

1. The data below are described by a single-valued formal context $\mathbb{K} = (G, M, J)$.

Let $\mathbb{K}_P = (U, I, P)$ be a formal context, where U is a set of users, I is a set of ideas, and $P \subseteq U \times I$ shows which user proposed which idea. Two other contexts, $\mathbb{K}_C = (U, I, C)$ and $\mathbb{K}_E = (U, I, E)$, describe binary relations of idea commenting and idea evaluation respectively.

The user-to-user relationships can also be represented by means of a single-valued formal context $\mathbb{K} = (U, U, J \subseteq U \times U)$, where $u_1 J u_2$ can designate, for example, that user u_1 commented some idea proposed by u_2. Relationships between content items can be modelled in the same way, e.g. $\mathbb{K} = (T, T, J \subseteq T \times T)$, where $t_1 J t_2$ shows that t_1 and t_2 occurred together is some text (idea or comment).

2. A multi-valued context $\mathbb{K}^W = (G, M, W, J)$ can be useful for representing data with numeric attributes.

Let $\mathbb{K}^F = (U, K, F, J)$ be a multi-valued context, where U is a set of users, K is a set of keywords, F is a set of keyword frequency values, $J \subseteq U \times K \times F$ shows how many times a particular user u applied a keyword k in an idea description or while discussing some ideas. The context $\mathbb{K}^F$ can be reduced to a plain context by means of (plain) scaling.

The commenting and evaluation relations can be described through multi-valued contexts in case we count each comment or evaluation for a certain topic. E.g, the multi-valued context $\mathbb{K}^V = (U, I, V = \{-3, -2, -1, 0, 1, 2, 3\}, J)$ describes which mark a particular user u assign to an idea i, where V contains values of possible marks; it can be written as $u(i) = w$, where $v \in V$.

3. A triadic context $\mathbb{K}_B = (G, M, B, Y \subseteq G \times M \times B)$ can be used for data containing tags as descriptors.

Consider the formal context $\mathbb{K}_T = (U, I, T, Y)$, where U is a set if users, I is a set of ideas, T is a set of tags (e.g. keywords and keyphrases), Y shows that a particular user u used keyword t in the description of an idea i.

It is worth to mention that all considered data can be sorted out into two groups: 1) data with keywords $\mathbb{K}^F, \mathbb{K}_T$ and 2) data without them $\mathbb{K}_P, \mathbb{K}_I, \mathbb{K}_E$.

The main advantage of such a representation is that FCA can be applied to community detection which is the main part of social network analysis. Social network analysis is a popular research field in which methods are developed for analysing 1-mode networks, like friend-to-friend, 2-mode [33,34,35], 3-mode

[18,36,17] and even multimodal dynamic networks [9,11,12]. Here we focused on the subfield of bicommunity and tricommunity identification.

As it was shown above, crowdsourcing data can be represented as bipartite or tripartite graphs. Standard techniques like "maximal bicliques search" return a huge number of patterns (in the worst case exponential w.r.t. the input size). Therefore we need some relaxation of the biclique notion and good interestingness measures for mining biclique communities.

It is widely known in the social network analysis community (see, e.g. [37,38,39,40,41]) that the notion of formal concept is (almost) the same thing as a biclique.

A concept-based bicluster (OA-bicluster) [16] is a scalable approximation of a formal concept (biclique). The advantages of concept-based biclustering are:

1. Less number of patterns to analyze;
2. Less computational time (polynomial vs exponential);
3. Manual tuning of bicluster (community) density threshold;
4. Tolerance to missing (object, attribute) pairs.

For analyzing three-mode network data like folksonomies [42] we also proposed a triclustering technique [17]. The reader can refer to [43] to see how that approach was empirically validated on real online social network data.

Thus every formal concept or OA-bicluster of contexts from paragraph 1 or 2 (after scaling) can be considered as a bicommunity of users sharing similar interests or behaving similarly and every triconcept or tricluster of contexts from paragraph 3 can be interpreted as a tricommunity of users, their ideas and keywords they used. These patterns are crucial for team building and recommendation of relevant topics and persons for discussions. According to practitioners in the field, exploiting these patterns can make crowdsourcing work more comfortable and increase user's activity.

3.5 FCA-Based Recommender Model

Two kinds of recommendations seem to be potentially useful for crowdsourcing. The first one is a recommendation of like-minded persons to a particular user, and the second one is able to find antagonists, users which discussed the same topics as a target one, but with opposite marks.

1. **Recommendations of like-minded persons and interesting ideas**
Let $\mathbb{K}_P = (U, I, P)$ be a context which describes idea proposals. Consider a target user $u_0 \in U$, then every formal concept $(A, B) \in \mathfrak{B}_P(U, I, P)$ containing u_0 in its extent provides potentially interesting ideas to the target user in its intent and prospective like-minded persons in $A \setminus \{u_0\}$.

Consider the set $\mathfrak{R}(u_0) = \{(A, B) | (A, B) \in \mathfrak{B}_P(U, I, P)\, and\, u_0 \in A\}$ of all concepts containing a target user u_0. Then the score of each idea or user to recommend to u_0 can be calculated as follows $score(i, u_0) = \frac{|\{u | u \in A,\ (A,B) \in \mathfrak{R}(\mathfrak{u}_\circ)\ and\ i \in u'\}|}{|\{u | u \in A\ and\ (A,B) \in \mathfrak{R}(\mathfrak{u}_\circ)\}|}$ or $score(u, u_0) = \frac{|\{A | u \in A\ and\ (A,B) \in \mathfrak{R}(\mathfrak{u}_\circ)\}|}{|\mathfrak{R}(\mathfrak{u}_\circ)|}$ respectively. As a result we have a set of

ranked recommendations $R(u_0) = \{(i, score(i))|i \in B \; and \; (A, B) \in \mathfrak{R}\}$. One can select the topmost N of recommendations from R ordered by their score.

2. **Recommendations of antagonists**
Consider two evaluation contexts: the multi-valued context $\mathbb{K}^W = (U, I, W = \{-3, -2, -1, 0, 1, 2, 3\}, J)$ and binary context $\mathbb{K}_E = (U, I, E)$. Then consider (X, Y) from $\mathfrak{R}(u_0) = \{(A, B)|(A, B) \in \mathfrak{B}_P(U, I, P) \; and \; u_0 \in A\}$. Set X contains people that evaluated the same set of topics Y, but we cannot say that all of them are like-minded persons w.r.t relation E. However, we can introduce a distance measure, which shows for every pair of users from X how distant they are in marks of ideas evaluation:

$$d_{(X,Y)}(u_1, u_2) = \sum_{\substack{u_1, u_2 \in X \\ i \in Y}} |i(u_1) - i(u_2)|. \tag{5}$$

As a result we again have a set of ranked recommendations $R_{(X,Y)}(u_0) = \{(u, d(i))|u \in U \; and \; (A, B) \in \mathfrak{R}\}$. The topmost pairs from $R_d(u_0)$ with the highest distance contain antagonists, that is persons with the opposite views on most of the topics which u_0 evaluated. To aggregate $R_{(X,Y)}(u_0)$ for different (X, Y) from $\mathfrak{R}(u_0)$ into a final ranking we can calculate

$$d_{(u_0,u)} = \max\{d_{(X,Y)}(u_0, u)|(X, Y) \in \mathfrak{R}(u_0) \; and \; u_0, u \in X\}. \tag{6}$$

The proposed models need to be tuned and validated, and also assume several variations such as using biclusters instead of formal concepts and other ways of final distance calculation. An additional possible recommender model can exploit triadic data structures for more diverse recommendations from the different sets of a triadic context.

3.6 Keywords and Keyphrases Extraction

We consider *Keywords (keyphrases)* as a set of the most significant words (phrases) in a text document that can provide a compact description for the content and style of this document. In the remainder of this paper we do not always differentiate between keywords and keyphrases, assuming that a keyword is a particular case of a keyphrase. In our project two similar problems of keyword and keyphrase extraction arise:

1. Keywords and keyphrases of the whole Witology forum;
2. Keywords and keyphrases of one user, topic etc.

In the first case we concentrate on finding syntactically well associated keywords (keyphrases). In the second case specific words and phrases of a certain user or topic are the subject of interest. Hence, we have to use two different methods for each keyword (keyphras) extraction problem. The first one is solved by using any statistical measure of association, such as Pointwise Mutual Information (PMI), T-Score or Chi-Square [44]. To solve the second problem we may use TF-IDF

or Mutual Information (MI) measures that reflect how important the word or phrase is for the given subset of texts. All the above mentioned measures define the weight of a specific word or phrase in the text. The words and phrases of the highest weight then can be considered as keywords and keyphrases. We are more interested in the quality of extracted keywords and keyphrases than in the way we obtain them. To tokenize texts we use a basic principle of word separation: there should be either a space or a punctuation mark between two words. A hyphen between two sequences of symbols makes them one word. To lemmatize words we use the Russian AOT lemmatizer [45], which is far from being ideal, but it is the only freely available one (even for commercial usage) for processing Russian texts. To normalize bi- and tri-grams we use one of our Python scripts that normalizes phrases according to their formal grammatical patterns. We are going to use formal contexts based on sets of extracted keyphrases and people who use them, the occurrence of keyphrases in texts and so on. By analogy, keyphrases, texts and users all together form a tricontext for further analysis. Moreover, keyphrases are an essential part of a socio-semantic network model, where they are used for semantic representation of the network's nodes.

4 Analysis Scheme

The data analysis scheme of CrowDM, which is developed now by the project and educational team of Witology and NRU HSE is presented in figure 2. As it was mentioned before, after downloading data from a platform database, we obtain formal contexts and text collections. In turn, the latter become formal contexts as well after keyword extraction. After that, the resulting contexts are analyzed. The FCA and multimodal clustering blocks of CrowDM were implemented by N. Romashkin and K. Blinkin in Python for the project.

5 First Experiments Results

We performed different experiments with the following methods: formal concepts, iceberg-lattices and stability indices, biclustering, triclustering, implications, association rules, power law analysis, and SNA methods.

For carrying out experiments we constructed formal concepts where objects are users of the platform and attributes are ideas which users proposed within one of 5 project topics ("Sberbank and private client"). We selected only the ideas that reached the end or almost the end of the project. An object "user" has an attribute "idea" if this user somehow contributed to the discussion of this idea, i.e. he is an author of the idea, commented on the idea and evaluated the idea or comments which were added to the idea. Thus, the extracted formal concepts (U, I), where U is a set of users, I is a set of ideas, correspond to so called epistemic communities (communities of interests), i.e. the set of users U who are interested in the ideas of I. Figure 3 displays the diagram of the obtained upper part of a certain concept lattice.

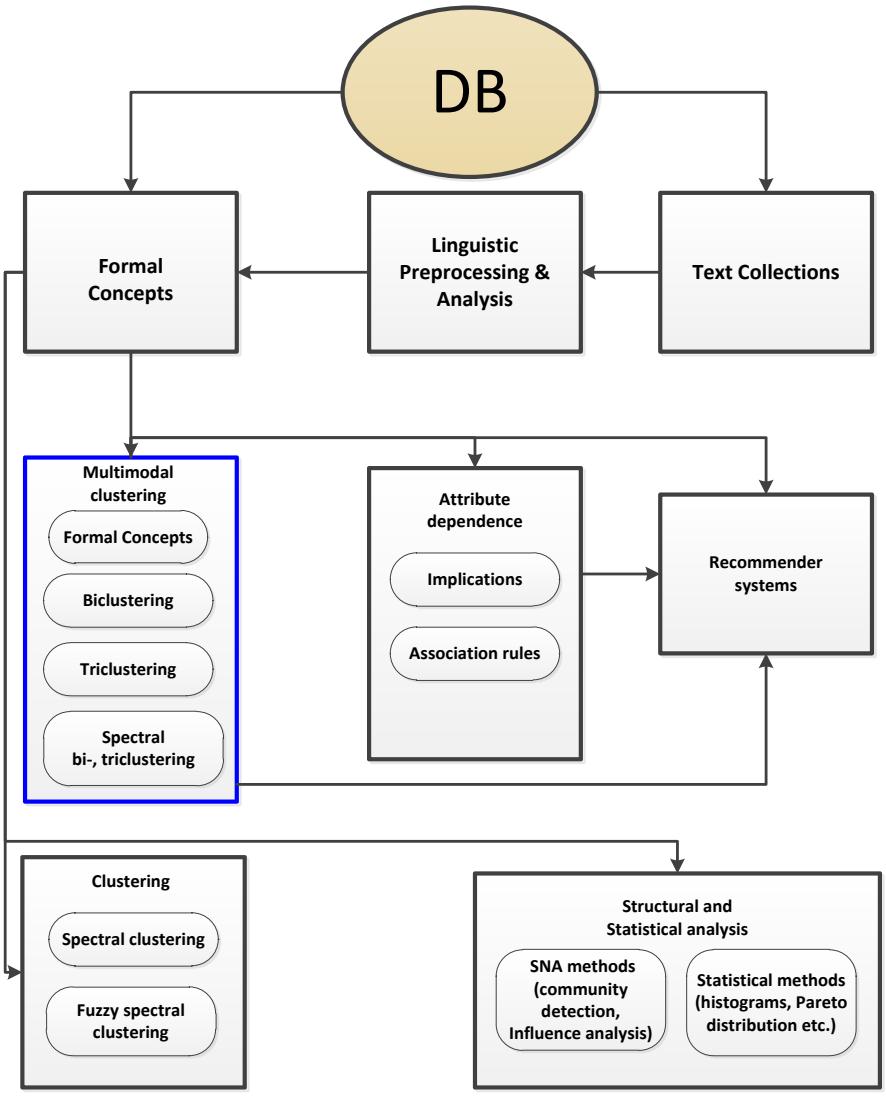

Fig. 2. The data analysis scheme of CrowDM

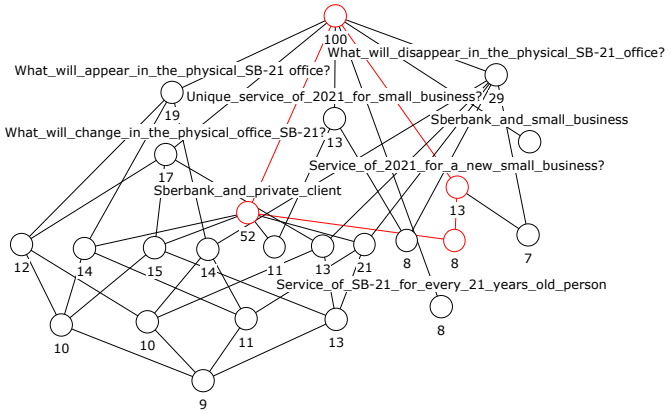

Fig. 3. The order filter (iceberg-lattice) diagram of the concept lattice for a certain users-tasks context with 24 concepts and $minsupp = 7$. The diagram is obtained by CrowDM system.

Each node of the diagram coincides with one formal concept (in total the lattice contains 198 concepts). A node cam be marked by the label of an object (or the count of objects in a formal concept extent) or an attribute if this object (moving bottom-up by diagram) or attribute (moving top-down) first appeared in this node. It is obvious that the obtained diagram is too awkward to be analyzed as a static image. Usually in such cases one can use order filters or diagrams of the sets of stable concepts or iceberg-lattices for visualization. We will showcase how to read a concept lattice using the lattice fragment in figure 4. Some first experiments were carried out using the program Concept Explorer (ConExp) which was developed for applying FCA algorithms to object-attribute data [46]. Later we applied our own data analysis system CrowDM. The system is able to build the formal concepts and biclusters for a given context. Clicking on a lattice node, one can see the objects and attributes corresponding to the concept which this node represents. Objects are accumulated from below (in the given example the set of objects contains User45 and User22), attributes come from above (we have only one attribute, "Microcredits from 1000 to 5000"). This means that User45 and User22 together took part in the discussion of the given idea and nobody else discussed it.

We demonstrate the results of applying biclustering algorithms on the same data below.

Let us explain the figure 5. During experiments we used the system for gene expression data analysis BicAT [15]. Rows correspond to users, columns are ideas of a given topic ("Sberbank and private client"), in the discussion of which users participated. The color of the cell of the corresponding row and column intersection depicts the contribution intensity of a given user to a given idea. The contribution is a weighted sum of the number of comments and evaluations to that idea and takes into account the fact whether this user is an author of this

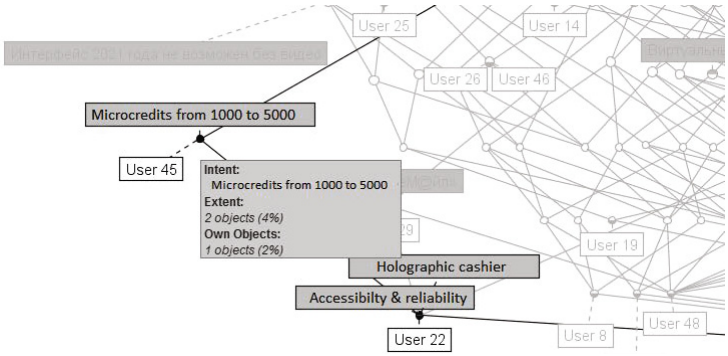

Fig. 4. Fragment of concept lattice diagram obtained by ConExp

idea. The lightest cells coincide with zero contribution, the brightest ones (fig. 5, top left cell) show the maximum contribution. After data discretization (0 – zero contribution, 1 – otherwise) we applied the BiMax algorithm which found some biclusters (see fig. 5 for example). Since one of the important crowdsourcing project problems is the search for people with similar ideas, the presented bicluster with 6 users is most interesting. The majority of the other found biclusters contained less than 4-5 users (we constrained the number of ideas in a bicluster to be strictly greater than 2).

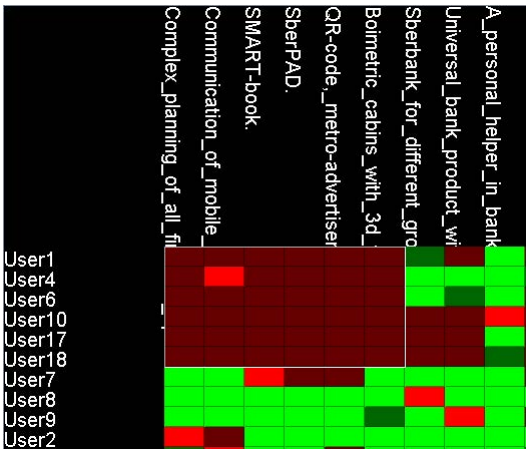

Fig. 5. Bicluster with a large number of users

Then, to gain a better understanding of the evaluation process in the project, the evaluation distribution was plotted in several ways. One of them is presented in fig. 6; it shows the cumulative number of users, who made more than a certain amount of evaluations during the entire project.

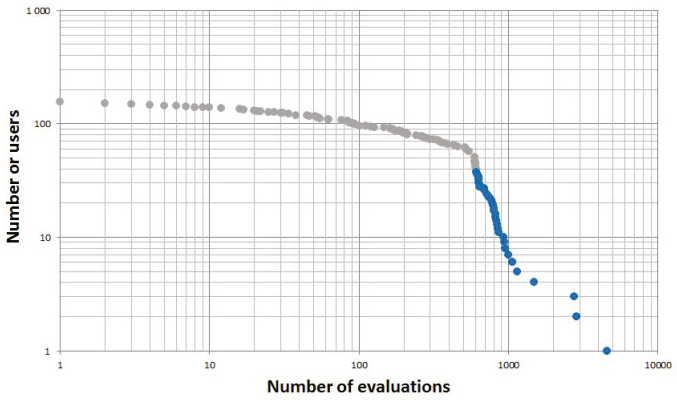

Fig. 6. Evaluation distribution

The horizontal axis displays the amount of submitted evaluations. The vertical axis represents the number of users, who made more than a fixed amount of evaluations. For instance, there is only one participant, who produced more than 5000 evaluations, and one more person, who made more than 3000 but less than 5000 evaluations. Thus, the rightmost dot on the X-axis shows the first participant (the y-coordinate is 1), and the next dot shows both of them (the y-coordinate is 2). The total number of users, who have once evaluated something, is 167. The set of graph points is explicitly split into two parts: the long gentle line (from $x = 0$ to 544 inclusive) and the steep tail. The fact, that both lines seem almost straight in logarithmic scales, indicates that the evaluation activity on the project might follow a Pareto distribution. It is reasonable to seek the individual distribution functions for the main and the tail parts of the sample, as testing the whole sample for goodness of fit to a Pareto distribution results in strong rejection of the null hypothesis ($H0$: "The sample follows a Pareto distribution").

We perform further analysis with two subsamples of the initial sample by means of Matlab tools from [23]. This analysis implies useful consequences according to the well-known "80:20" rule:

$$W = P^{(\alpha-2)/(\alpha-1)},$$

which means that the fraction W of the wealth is in the hands of the richest P of the population. In our case, 70% of users make 80% of all evaluations ($\alpha = 3.48$, p-value$= 0.78$ for $x \in [614, 5020]$). Thus, the situation is quite wealthy and there is no serious disproportions in evaluations. But if one finds strong disproportion in evaluation or commenting activity like 20% of users make 80% of actions, it implies that facilitators, who is responsible for monitoring and control in the crowdsourcing system, should involve more inactive users into the process. We also built a typology of Witology platform users [47].

6 Conclusion

The results of our first experiments suggest that the developed methodology will be useful for data analysis of crowdsourcing systems. The most important directions for future work include the analysis of textual information generated by users, applying multimodal clustering methods and using them for developing recommender systems. Development and experimental validation of recommender models, including FCA-based, are in R&D plan of the company. Some interesting results concern regression between different user's ratings and various measures of actor importance in SNA were obtained; for example, one of the Witology ratings was well-described by such SNA measures as user's centrality and degree. Thus SNA can be a good tool for developing and redesigning the Witology ranking methods.

Acknowledgments. The main part of this work was performed by the project and educational group "Algorithms of Data Mining for Innovative Projects Internet Forum". Further work was supported by the Basic Research Program at the National Research University Higher School of Economics in 2012 and performed in the Laboratory of Intelligent Systems and Structural Analysis. We also would like to thank Oleg Anshakov, Sergei Kuznetsov and Rostislav Yavorsky for their valuable comments and the rest of our group: Nikita Romashkin, Konstantin Blinkin, Ekaterina Chernyak, Olga Chugunova, Daria Goncharova, Daniil Nedumov, Fedor Strok.

References

1. Spigit company, `http://spigit.com/`
2. Brightidea company, `http://www.brightidea.com/`
3. Innocentive comp., `http://www.innocentive.com/`
4. Imaginatik company, `http://www.imaginatik.com/`
5. Kaggle, `http://www.kaggle.com`
6. Witology company, `http://witology.com/`
7. Wikivote company, `http://www.wikivote.ru/`
8. Sberbank-21, national entrepreneurial initiative-2012, `http://sberbank21.ru/`
9. Roth, C.: Generalized preferential attachment: Towards realistic socio-semantic network models. In: ISWC 4th Intl Semantic Web Conference, Workshop on Semantic Network Analysis. CEUR-WS Series, Galway, Ireland, vol. 171, pp. 29–42 (2005) ISSN 1613-0073
10. Cointet, J.-P., Roth, C.: Socio-semantic dynamics in a blog network. In: CSE (4), pp. 114–121. IEEE Computer Society (2009)
11. Roth, C., Cointet, J.P.: Social and semantic coevolution in knowledge networks. Social Networks 32, 16–29 (2010)
12. Yavorsky, R.: Research Challenges of Dynamic Socio-Semantic Networks. In: Ignatov, D., Poelmans, J., Kuznetsov, S. (eds.) CDUD 2011 - Concept Discovery in Unstructured Data, CEUR Workshop Proceedings, vol. 757, pp. 119–122 (2011)
13. Howe, J.: The rise of crowdsourcing. Wired (2006)
14. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations, 1st edn. Springer-Verlag New York, Inc., Secaucus (1999)

15. Barkow, S., Bleuler, S., Prelic, A., Zimmermann, P., Zitzler, E.: Bicat: a biclustering analysis toolbox. Bioinformatics 22(10), 1282–1283 (2006)
16. Ignatov, D.I., Kaminskaya, A.Y., Kuznetsov, S., Magizov, R.A.: Method of Biclusterzation Based on Object and Attribute Closures. In: Proc. of 8th International Conference on Intellectualization of Information Processing (IIP 2011), Cyprus, Paphos, October 17-24, pp. 140–143. MAKS Press (2010) (in Russian)
17. Ignatov, D.I., Kuznetsov, S.O., Magizov, R.A., Zhukov, L.E.: From Triconcepts to Triclusters. In: Kuznetsov, S.O., Ślęzak, D., Hepting, D.H., Mirkin, B.G. (eds.) RSFDGrC 2011. LNCS, vol. 6743, pp. 257–264. Springer, Heidelberg (2011)
18. Jäschke, R., Hotho, A., Schmitz, C., Ganter, B., Stumme, G.: TRIAS–An Algorithm for Mining Iceberg Tri-Lattices. In: Proceedings of the Sixth International Conference on Data Mining, ICDM 2006, pp. 907–911. IEEE Computer Society, Washington, DC (2006)
19. Ignatov, D.I., Kuznetsov, S.O.: Concept-based Recommendations for Internet Advertisement. In: Belohlavek, R., Kuznetsov, S.O. (eds.) Proc. CLA 2008. CEUR WS, vol. 433, pp. 157–166. Palacký University, Olomouc (2008)
20. Ignatov, D., Poelmans, J., Zaharchuk, V.: Recommender System Based on Algorithm of Bicluster Analysis RecBi. In: Ignatov, D., Poelmans, J., Kuznetsov, S. (eds.) CDUD 2011 - Concept Discovery in Unstructured Data. CEUR Workshop Proceedings, pp. 122–126 (2011)
21. Ignatov, D.I., Poelmans, J., Dedene, G., Viaene, S.: A New Cross-Validation Technique to Evaluate Quality of Recommender Systems. In: Kundu, M.K., Mitra, S., Mazumdar, D., Pal, S.K. (eds.) PerMIn 2012. LNCS, vol. 7143, pp. 195–202. Springer, Heidelberg (2012)
22. Ignatov, D.I., Konstantinov, A.V., Nikolenko, S.I., Poelmans, J., Zaharchuk, V.: Online recommender system for radio station hosting. In: [48], pp. 1–12
23. Clauset, A., Shalizi, C.R., Newman, M.E.J.: Power-law distributions in empirical data. SIAM Rev. 51(4), 661–703 (2009)
24. Kuznetsov, S.O.: On stability of a formal concept. Ann. Math. Artif. Intell. 49(1-4), 101–115 (2007)
25. Lehmann, F., Wille, R.: A Triadic Approach to Formal Concept Analysis. In: Ellis, G., Rich, W., Levinson, R., Sowa, J.F. (eds.) ICCS 1995. LNCS, vol. 954, pp. 32–43. Springer, Heidelberg (1995)
26. Wille, R.: The basic theorem of triadic concept analysis. Order 12, 149–158 (1995)
27. Voutsadakis, G.: Polyadic concept analysis. Order 19(3), 295–304 (2002)
28. Mirkin, B.G., Kramarenko, A.V.: Approximate Bicluster and Tricluster Boxes in the Analysis of Binary Data. In: Kuznetsov, S.O., Ślęzak, D., Hepting, D.H., Mirkin, B.G. (eds.) RSFDGrC 2011. LNCS, vol. 6743, pp. 248–256. Springer, Heidelberg (2011)
29. Belohlavek, R., Vychodil, V.: Factorizing Three-Way Binary Data with Triadic Formal Concepts. In: Setchi, R., Jordanov, I., Howlett, R.J., Jain, L.C. (eds.) KES 2010, Part I. LNCS, vol. 6276, pp. 471–480. Springer, Heidelberg (2010)
30. Cerf, L., Besson, J., Robardet, C., Boulicaut, J.F.: Data peeler: Contraint-based closed pattern mining in n-ary relations. In: SDM, pp. 37–48. SIAM (2008)
31. Cerf, L., Besson, J., Robardet, C., Boulicaut, J.F.: Closed patterns meet n-ary relations. ACM Trans. Knowl. Discov. Data 3, 3:1–3:36 (2009)
32. Drutsa, A., Yavorskiy, K.: Socio-semantic network data visualization. In: Tagiew, R., Ignatov, D.I., Neznanov, A.A., Poelmans, J. (eds.) EEML 2012 - Experimental Economics and Machine Learning. CEUR Workshop Proceedings, vol. 757 (2012)
33. Latapy, M., Magnien, C., Vecchio, N.D.: Basic notions for the analysis of large two-mode networks. Social Networks 30(1), 31–48 (2008)

34. Liu, X., Murata, T.: Evaluating community structure in bipartite networks. In: Elmagarmid, A.K., Agrawal, D. (eds.) SocialCom/PASSAT, pp. 576–581. IEEE Computer Society (2010)
35. Opsahl, T.: Triadic closure in two-mode networks: Redefining the global and local clustering coefficients. Social Networks 34 (2011) (in press)
36. Murata, T.: Detecting communities from tripartite networks. In: Rappa, M., Jones, P., Freire, J., Chakrabarti, S. (eds.) WWW, pp. 1159–1160. ACM (2010)
37. Freeman, L.C., White, D.R.: Using galois lattices to represent network data. Sociological Methodology 23, 127–146 (1993)
38. Freeman, L.C.: Cliques, galois lattices, and the structure of human social groups. Social Networks 18, 173–187 (1996)
39. Duquenne, V.: Lattice analysis and the representation of handicap associations. Social Networks 18(3), 217–230 (1996)
40. White, D.R.: Statistical entailments and the galois lattice. Social Networks 18(3), 201–215 (1996)
41. Roth, C., Obiedkov, S., Kourie, D.: Towards Concise Representation for Taxonomies of Epistemic Communities. In: Yahia, S.B., Nguifo, E.M., Belohlavek, R. (eds.) CLA 2006. LNCS (LNAI), vol. 4923, pp. 240–255. Springer, Heidelberg (2008)
42. Vander Wal, T.: Folksonomy Coinage and Definition (2007), `http://vanderwal.net/folksonomy.html` (accessed on March 12, 2012)
43. Gnatyshak, D., Ignatov, D.I., Semenov, A., Poelmans, J.: Gaining insight in social networks with biclustering and triclustering. In: [48], pp. 162–171
44. Manning, C.D., Schütze, H.: Foundations of statistical natural language processing. MIT Press, Cambridge (1999)
45. Russian project on automatic text processing, `http://www.aot.ru`
46. Grigoriev, P.A., Yevtushenko, S.A.: Elements of an Agile Discovery Environment. In: Grieser, G., Tanaka, Y., Yamamoto, A. (eds.) DS 2003. LNCS (LNAI), vol. 2843, pp. 311–319. Springer, Heidelberg (2003)
47. Bezzubtseva, A., Ignatov, D.I.: A New Typology of Collaboration Platform Users. In: Tagiew, R., Ignatov, D.I., Neznanov, A.A., Poelmans, J. (eds.) EEML 2012 - Experimental Economics and Machine Learning. CEUR Workshop Proceedings, vol. 757, pp. 9–19 (2012)
48. Aseeva, N., Babkin, E., Kozyrev, O. (eds.): BIR 2012. LNBIP, vol. 128. Springer, Heidelberg (2012)