Building up Shared Knowledge with Logical Information Systems

Mireille Ducassé, Sébastien Ferré, Peggy Cellier

Building up Shared Knowledge with Logical Information Systems

Mireille Ducassé[1], Sébastien Ferré[2], and Peggy Cellier[1]

[1] IRISA-INSA de Rennes, France,
{ducasse, cellier}@irisa.fr
[2] IRISA-University of Rennes 1, France
ferre@irisa.fr

Abstract. Logical Information Systems (LIS) are based on Logical Concept Analysis, an extension of Formal Concept Analysis. This paper describes an application of LIS to support group decision. A case study gathered a research team. The objective was to decide on a set of potential conferences on which to send submissions. People individually used Abilis, a LIS web server, to preselect a set of conferences. Starting from 1041 call for papers, the individual participants preselected 63 conferences. They met and collectively used Abilis to select a shared set of 42 target conferences. The team could then sketch a publication planning. The case study provides evidence that LIS cover at least three of the collaboration patterns identified by Kolfschoten, de Vreede and Briggs. Abilis helped the team to build a more complete and relevant set of information (Generate/Gathering pattern); to build a shared understanding of the relevant information (Clarify/Building Shared Understanding); and to quickly reduce the number of target conferences (Reduce/Filtering pattern).

1 Introduction

Group work represents a large amount of time in professional life while many people feel that much of that time is wasted. Lewis [13] argues that this amount of time is even going to increase because problems are becoming more complex and are meant to be solved in a distributed way. Each involved person has a local and partial view of the problem, no one embraces the whole required knowledge. Lewis also emphasizes that it is common that *"groups fail to adequately define a problem before rushing to judgment"*. Building up shared knowledge in order to gather relevant distributed knowledge of a problem is therefore a crucial issue.

Logical Information Systems (LIS) are based on Logical Concept Analysis (LCA), an extension of Formal Concept Analysis (FCA). In a previous work [5], Camelis, a single-user logical information system, has been shown useful to support serene and fair meetings. This paper shows how Abilis, a LIS web server that implements OnLine Analytical Processing (OLAP [3]) features, can be applied to help build shared knowledge among a group of skilled users.

The presented case study gathered a research team to decide on a publication strategy. Starting from 1041 call for papers, each team member on his own

In Proceedings of teh 8th International Conference on Concept lattices and Their Applications, Amedeo Napoli and Vilem Vychodil eds. 17-20 October 2011, Nancy, France. ISBN 978-2-905267-78-8

preselected a set of conferences matching his own focus of interest. The union of individual preselections still contained 63 conferences. Then, participants met for an hour and a half and collectively built a shared set of 42 target conferences. For each conference, the team shared a deep understanding of why it was relevant. The team could sketch a publication planning in a non-conflictual way.

Kolfschoten, de Vreede and Briggs have classified collaboration tasks into 16 collaboration patterns [12]. The contribution of this paper is to give evidences that LIS can significantly support three of these patterns which are important aspects of decision making, namely *Generate/Gathering*, *Clarify/Building Shared Understanding* and *Reduce/Filtering.* Firstly, the navigation and filtering capabilities of LIS were helpful to detect inconsistencies and missing knowledge. The updating capabilities of LIS enabled participants to add objects, features and links between them on the fly. As a result the group had a more complete and relevant set of information (Generate/Gathering pattern). Secondly, the compact views provided by LIS and the OLAP features helped participants embrace the whole required knowledge. The group could therefore build a shared understanding of the relevant information which was previously distributed amongst the participants (Clarify/Building Shared Understanding pattern). Thirdly, the navigation and filtering capabilities of LIS were relevant to quickly converge on a reduced number of target conferences (Reduce/Filtering pattern).

In the following, Section 2 briefly introduces logical information systems. Section 3 describes the case study. Section 4 gives detailed arguments to support the claim that logical information systems help build up shared knowledge. Section 5 discusses related work.

2 Logical Information Systems

Logical Information Systems (LIS) [7] belong to a paradigm of information retrieval that combines querying and navigation. They are formally based on a logical generalization of Formal Concept Analysis (FCA) [8], namely Logical Concept Analysis (LCA) [6]. In LCA, logical formulas are used instead of sets of attributes to describe objects. LCA and LIS are generic in that the logic is not fixed, but is a parameter of those formalisms. Logical formulas are also used to represent queries and navigation links. The concept lattice serves as the navigation structure: every query leads to a concept, and every navigation link leads from one concept to another. The query can be modified in three ways: by formula edition, by navigation (selecting features in the index in order to modify the query) or by examples. Annotations can be performed in the same interface. Camelis[3] has been developed since 2002; a web interface, Abilis [4], has recently been added. It incorporates display paradigms based on On-Line Analytical Processing (OLAP). Instead of being presented as a list of objects, an extent can be partitioned as an OLAP cube, namely a multi-dimensional array [1].

[3] see `http://www.irisa.fr/LIS/ferre/camelis/`

[4] http://ledenez.insa-rennes.fr/abilis/

3 The Case Study

The reported case study gathered 6 participants, including the 3 authors, 4 academics and 2 PhD students. All participants were familiar with LIS, 4 of them had not previously used a LIS tool as a decision support system. The objective was to identify the publishing strategy of the team: in which conferences to submit and why. This has not been a conflictual decision, the group admitted very early that the set of selected conferences could be rather large provided that there was a good reason to keep each of them.

One person, the facilitator, spent an afternoon organizing the meeting and preparing the raw data as well as a logical context according to the objective. She collected data about conference call for papers of about a year, related to themes corresponding to the potential area of the team, from WikiCFP, a semantic wiki for Calls For Papers in science and technology fields [5]. There were 1041 events: conferences, symposiums, workshops but also special issues of journals.

Then every participant, on its own, spent between half an hour to two hours to browse the context, update it if necessary and preselect a number of conferences (Section 3.1). The group met for one hour and a half. It collaborately explored the data and selected a restricted set of conferences (Section 3.2). After the meeting, every participant filled a questionnaire. The context used for the case study can be freely accessed [6].

3.1 Distributed Individual Preselection and Update

When the context was ready, every participant was asked to preselect a set of conferences that could be possible submission targets. The instruction was to be as liberal as wanted and in case of doubt to label the conference as a possible target.

During this phase, each of the academics preselected 20 to 30 conferences and each of the PhD students preselected around 10 conferences. Each participant had his own "basket". There were overlappings, altogether 63 conferences were preselected. Participants also introduced new conferences and new features, for example, the ranking of the Australian CORE association [7] (Ranking), and the person expected to be a possible first author for the target conference (Main author).

Figure 1 shows a state of Abilis during the preselection phase. LIS user interfaces give a *local view* of the concept lattice, centered on a focus concept. The local view is made of three parts: (1) the query (top left), (2) the extent (bottom right), and (3) the index (bottom left). The query is a logical formula that typically combines attributes (e.g., `Name`), patterns (e.g., `contains "conference"`), and Boolean connectors (`and`, `or`, `not`). The extent is the set of objects that are matched by the query, according to logical subsumption. The extent identifies

[5] http://www.wikicfp.com/cfp/

[6] http://ledenez.insa-rennes.fr/abilis/, connect as guest, load `Call for papers`.

[7] http://core.edu.au/index.php/categories/conference rankings

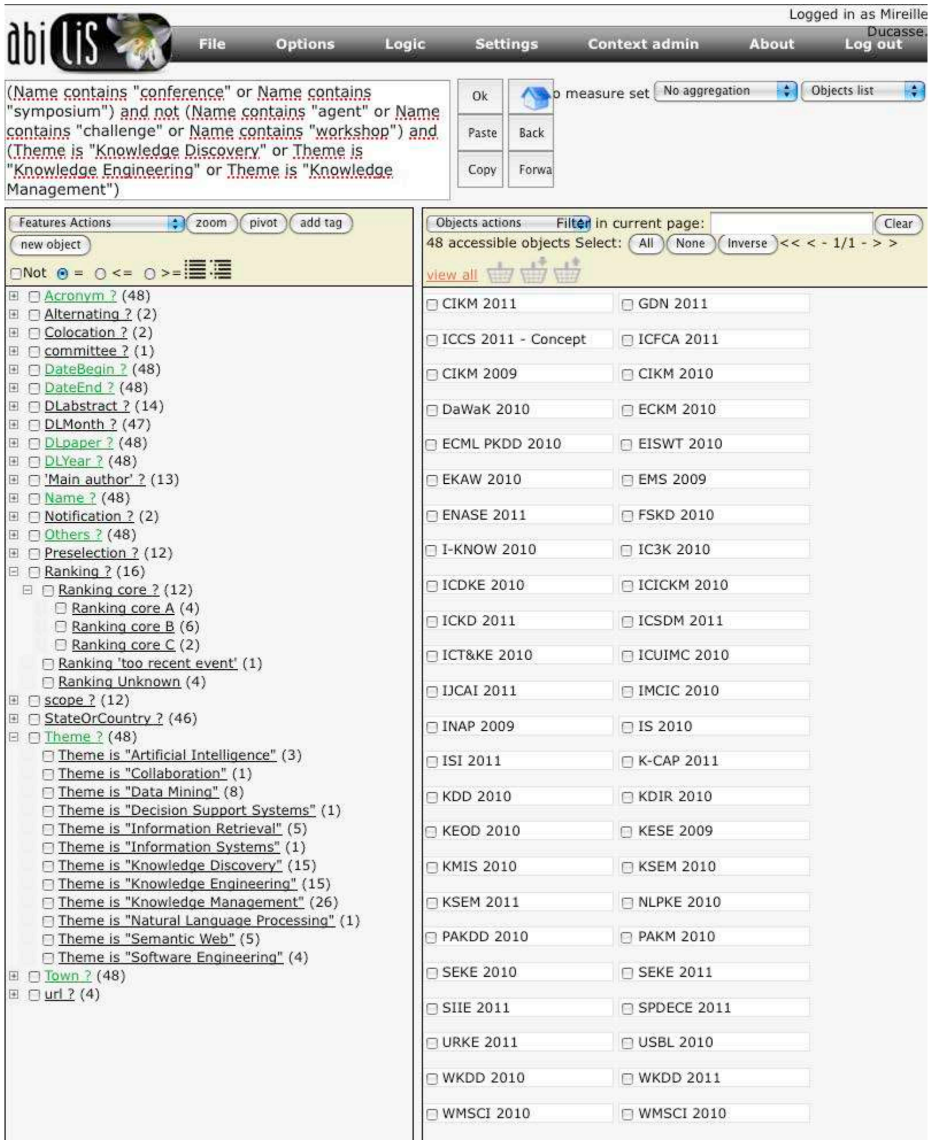

Fig. 1. Snapshot of Abilis during preselection: a powerful query

the focus concept. Finally, the index is a set of features, taken from a finite subset of the logic, and is restricted to features associated to at least one object in the extent. The index plays the role of a summary or inventory of the extent, showing which kinds of objects there are, and how many of each kind there are (e.g., in Figure 1, 8 objects in the extent have `data mining` as a theme). In the index, features are organized as a taxonomy according to logical subsumption.

The query area (top left) shows the current selection criteria: `(Name contains "conference" or Name contains "symposium") and not (Name contains "agent" or Name contains "challenge" or Name contains "workshop") and (Theme is "Knowledge Discovery" or Theme is "Knowledge Engineering" or Theme is "Knowledge Management")`. Note that the query had been obtained solely by clicking on features of the index (bottom left). Let us describe how it had been produced. Initially there were 1041 objects. Firstly, opening the `Name ?` feature, the participant had noticed that names could contain "conference" or "symposium" but also other keywords such as "special issue". He decided to concentrate on conferences and symposiums by clicking on the two features and then on the `zoom` button. The resulting query was `(Name contains "conference" or Name contains "symposium")` and there were 495 objects in the extent. However, the displayed features under `Name ?` showed that there were still objects whose name in addition to "conference" or "symposium" also contained "agent", "challenge" or "workshop". He decided to filter them out by clicking on the three features then on the `Not` button then on the `zoom` button. The resulting query was `(Name contains "conference" or Name contains "symposium") and not (Name contains "agent" or Name contains "challenge" or Name contains "workshop")` and there were 475 objects in the extent. He opened the `Theme ?` feature, clicked on the three subfeatures containing "Knowledge", then on the `zoom` button. The resulting query is the one displayed on Figure 1 and there are 48 objects in the displayed extent.

In the extent area (bottom right), the 48 acronyms of the selected conferences are displayed. In the index area, one can see which of the features are filled for these objects. The displayed features have at least 1 object attached to them. The number of objects actually attached to them is shown in parentheses. For example, only 14 of the preselected conferences have an abstract deadline. All of them have an acronym, a date of beginning, a date of end, a date for the paper deadline, a name, some other (not very relevant) information, as well as at least a theme and a town. The features shared by all selected objects have that number in parentheses (48 in this case). For the readers who have a color printout, these features are in green. The other features are attached to only some of the objects. For example, only 16 objects have a ranking attached to them: 4 core A, 6 core B, 2 core C, 1 'too recent event', 4 unknown (to the Core ranking).

One way to pursue the navigation could be, for example, to click on `Ranking ?` to select the conferences for which the information is filled. Alternatively, one could concentrate on the ones for which the ranking is not filled, for example to fill in this information on the fly for the conferences which are considered interesting.

Another way to pursue the navigation could be, for example, to notice that under the `Theme ?` feature, there are more than the selected themes. One can see that among the selected conferences, one conference is also relevant to the `Decision Support Systems` theme. One could zoom into it, this would add `and Theme is "Decision Support Systems"` to the query ; the object area would then display the relevant conference (namely GDN2011).

3.2 Collaborative Data Exploration, Update and Selection

The group eventually had a physical meeting where the current state of the context was constantly displayed on a screen.

Using the navigation facilities of Abilis, the conferences were examined by decreasing ranking. Initially, the group put in the selection all the A and A+ preselected conferences. After some discussions, it had, however, been decided that Human Computer Interaction (HCI) was too far away from the core of the team's research. Subsequently, the HCI conferences already selected were removed from the selection. For the conferences of rank B, the team decided that most of them were pretty good and deserved to be kept in the selection. For the others, the group investigated first the conferences without ranking and very recent, trying to identify the ones with high selection rate or other good reasons to put them in the selection. Some of the arguments have been added into the context. Some others were taken into account on the fly to select some conferences but they seemed so obvious at the time of the discussion that they were not added in the context.

Figure 2 shows the selection made by the group at a given point. In the extent area, on the right hand side, the selected objects are partitioned according to the deadline month and the anticipated main author thanks to the OLAP like facilities of Abilis [1]. Instead of being presented as a list of objects, an extent can be partitioned as an OLAP cube, namely a multi-dimensional array. Assuming object features are valued attributes, each attribute can play the role of a dimension, whose values play the role of indices along this dimension. Users can freely interleave changes to the query and changes to the extent view.

The query is `SelectionLIS02Dec and scope international`. Note that the partition display is consistent with the query. When the group added `and scope international` to the query, the national conferences disappeared from the array.

Some conferences, absent from WikiCFP have been entered on the fly at that stage (for example, ICCS 2011 - Concept). Not all the features had been entered for all of them. In particular, one can see in the feature area that only 28 out of 29 had been preselected. Nevertheless, the group judged that the deadline month, the potential main author and the ranking were crucial for the decision process and added them systematically. It is easy to find which objects do not have a feature using `Not` and `zoom`, and then to attach features to them.

Fig. 2. Snapshot of Abilis during collaborative data exploration: a partition deadline month/mainAuthor

One can see that there are enough opportunities for each participant to publish round the year. One can also see at a glance where compromises and decisions will have to be made. For example, PC will probably not be in a position to publish at IDA, ISSTA, KDD and ICCS the same year. Thanks to this global view PC can discuss with potential co-authors what the best strategy could be.

A follow up to the meeting was that participants made a personal publication planning, knowing that their target conferences were approved by the group.

4 Discussion

In this section, we discuss how the reported case study provides evidences that LIS help keep the group focused (Section 4.1) and that LIS also help build up shared knowledge (Section 4.2). As already mentioned, participants filled up a questionnaire after the meeting. In the following, for each item, we introduce the arguments, we present a summary of relevant parts of participant feedbacks, followed by an analysis of the features of LIS that are crucial for the arguments.

4.1 Logical Information Systems Help Keep the Group Focused

It is recognized that an expert facilitator can significantly increase the efficiency of a meeting (see for example [2]). A study made by den Hengst and Adkins [10] investigated which facilitation functions were found the most challenging by facilitators around the world. It provides evidences that facilitators find that *"the most difficult facilitation function in meeting procedures is keeping the group outcome focused."*

In our case study, all participants reported that they could very easily stay focused on the point currently discussed thanks to the query and the consistency between the three views.

As the objective was to construct a selection explicitly identified in Abilis by a feature, the objective of the meeting was always present to everybody and straightforward to bring back in case of digression. Furthermore, even if the context contained over a thousand conferences, thanks to the navigation facilities of LIS, only relevant information was displayed at a given time. Therefore, there was no "noise" and no dispersion of attention, the displayed information was always closely connected to the focus of the discussion.

4.2 Logical Information Systems Help Build Up Shared Knowledge

Kolfschoten, de Vreede and Briggs have identified 6 collaboration patterns: *Generate, Reduce, Clarify, Organize, Evaluate, and Consensus Building* [12]. We discuss in the following three of their 16 sub-patterns for which all participants agreed that they are supported by Abilis in its current stage. For the other sub-patterns, the situation did not demand much with respect to them. For example, the decision to make was not conflictual, the set of selected conferences could be rather large, there was, therefore, not much to experiment about "consensus building." The descriptions of the patterns in italic are from Kolfschoten, de Vreede and Briggs.

Generate/Gathering: move from having fewer to having more complete and relevant information shared by the group.

Before and during the meeting, information has been added to the shared knowledge repository of the group, namely the logical context. A new theme, important for the team and missing from WikiCFP, has been added: Decision Support Systems. New conferences have been added into the context either by individual participants in the preselection phase or by the group during the selection phase. New features were added. For example, it soon appeared that some sort of conference rankings was necessary. The group added by hand, for the conferences that were selected, the ranking of the Australian Core association. Some conferences were added subsequently, sometimes the ranking was not added at once.

All participants acknowledged that the tool helped the group to set up a set of features which was relevant and reflecting the group's point of view.

The crucial characteristics of LIS for this aspect are those which enable integrated navigation and update.

Firstly, the possibility to update the context while navigating in it enables participants to enhance it on the fly adding small pieces of relevant information at a time. Secondly, for each feature, Abilis displays the number of objects which have it. It is therefore immediate to detect when a feature is not systematically filled. The query `Not <feature>` selects the objects that do not have the feature. Users can then decide if they want to update them. Thanks to the query, as soon as an object is updated, it disappears from the extent. Users can immediately see what remains to be updated. Thirdly, updating the context does not divert from the initial objective. Indeed, the `Back` button allows users to go back to previous queries. Fourthly, the three views (query, features, objects) are always consistent and provide a "global" understanding of the relevant objects. Lastly, in the shared web server, participants can see what information the others had entered. Hence each participant can inspire the others.

For the last aspect, the facilitator inputs were decisive. Participants reported that they did not invent much, they imitated and adapted from what the facilitator had initiated. This is consistent with the literature on group decision and negotiation which emphasizes the key role of facilitators [2].

Clarify/Building Shared Understanding: Move from having less to more shared understanding of the concepts shared by the group and the words and phrases used to express them.

Participants, even senior ones, discovered new conferences. Some were surprised by the ranking of conferences that they had previously overlooked. Participants had a much clearer idea of who was interested in what.

All participants found that the tool helped them understand the points of view of the others.

The crucial characteristics of LIS for this aspect are those which enable to grasp a global understanding at a glance. Firstly, the query, as discussed earlier, helps keep the group focused. Secondly, the consistency between the 3 views helps participants to grasp the situation. Thirdly, irrelevant features are not in the index, the features in the index thus reflect the current state of the group decision. Fourthly, the partitions *à la* OLAP sort the information according to the criteria under investigation. Lastly, the shared web server enables participants to know before the meeting what the others have entered.

Reduce/Filtering: move from having many concepts to fewer concepts that meet specific criteria according to the group members.

Both at preselection time and during the meeting, participants could quickly strip down the set of conferences of interest according to the most relevant criteria.

All participants said that the filtering criteria were relevant and reflecting the group's point of view. They also all thought that the group was satisfied with the selected set of conferences.

The crucial characteristics of LIS for this aspect are those of the *navigation core* of LIS. Firstly, the features of the index propose filtering criteria. They are dynamically computed and they are relevant for the current selection of objects. Secondly, the query with its powerful logic capabilities enables participants to express sophisticated selections. Thirdly, the navigation facilities enable participants to build powerful queries, even without knowing anything about the syntax. Lastly, users do not have to worry about the consistency of the set of selected objects. The view consistency of Abilis guaranties that all conferences fulfilling the expressed query are indeed present.

This aspect is especially important. As claimed by Davis et al. [4], convergence in meetings is a slow and painful process for groups. Vogel and Coombes [16] present an experiment that supports the hypothesis that *groups selecting ideas from a multicriteria task formulation will converge better than groups working on a single criteria formulation,* where convergence is defined as *moving from many ideas to a focus on a few ideas that are worthy of further attention.* Convergence is very close to the Reduce/Filtering collaboration pattern. They also underline that *people try to minimize the effects of information overload by employing conscious or even unconscious strategies of heuristics in order to reduce information load,* where information overload is defined as *having too many things to do at once.*

With their powerful navigation facilities, LIS enable to address a large number of criteria and objects with a limited information overload. Indeed, one can concentrate on local aspects. The global consistency is maintained automatically by the concept lattice.

5 Related work

Abilis in its current stage does not pretend to match up to operational group support systems (GSS) which have a much broader scope. LIS, however, could be integrated in some of the modules of GSS. For example, $Meetingworks^{TM}$ [13], one of the most established GSS, is a modular toolkit that can be configured to support a wide variety of group tasks. Its "Organize" module proposes a tree structure to help analyze and sort ideas. That structure looks much like the index of LIS. It can be edited by hand and some limited selection is possible. The navigation capabilities of LIS based on the concept lattice are, however, more powerful.

Concept analysis has been applied to numerous social contexts, such as social networks [15], computer-mediated communication [9] and domestic violence detection [14]. Most of those applications are intended to be applied *a posteriori*, in order to get some understanding of the studied social phenomena. On the contrary, we propose to use Logical Concept Analysis in the course and as a support of the social phenomena itself. In our case, the purpose is to support a collaborative decision process. Our approach is to other social applications, what information retrieval is to data mining. Whereas data mining automatically computes a global and static view on *a posteriori* data, information retrieval (i.e. navigation in and update of the concept lattice) presents the user with a local and dynamic view on *live* data, and only guides users in their choice.

A specificity of LIS is the use of logics. This has consequences both on the queries that can be expressed, and on the feature taxonomy. The use of logics allows to express inequalities on numerical attributes, disjunctions and negations in queries. In pure FCA, only conjunctions of Boolean attributes can be expressed. Previous sections have shown how disjunction and negation are important to express selection criteria. In the taxonomy, criteria are organized according to the logical subsumption relation between them in pure FCA, criteria would be presented as a long flat list. Logics help to make the taxonomy more concise and readable by grouping and hierarchizing together similar criteria. The taxonomy can be dynamically updated by end-users.

6 Conclusion

In this paper we have shown that a Logical Information System web server could be used to support a group decision process consisting of 1) data preparation 2) distributed individual preselection and update and 3) collaborative data exploration, update and selection. We have presented evidences that the navigation and filtering capabilities of LIS were relevant to quickly reduce the number of target conferences. Secondly, the same capabilities were also helpful to detect inconsistencies and missing knowledge. The updating capabilities of LIS enabled participants to add objects, features and links between them on the fly. As a result the group had a more complete and relevant set of information. Thirdly, the group had built a shared understanding of the relevant information.

Acknowledgments The authors thank Pierre Allard and Benjamin Sigonneau for the development and maintenance of Abilis. They thank Pierre Allard, Annie Foret and Alice Hermann for attending the experiment and giving many insightful feedbacks.

References

1. Allard, P., Ferré, S., Ridoux, O.: Discovering functional dependencies and association rules by navigating in a lattice of OLAP views. In: Kryszkiewicz, M., Obiedkov, S. (eds.) Concept Lattices and Their Applications. pp. 199–210. CEUR-WS (2010)
2. Briggs, R.O., Kolfschoten, G.L., de Vreede, G.J., Albrecht, C.C., Lukosch, S.G.: Facilitator in a box: Computer assisted collaboration engineering and process support systems for rapid development of collaborative applications for high-value tasks. In: HICSS. pp. 1–10. IEEE Computer Society (2010)
3. Codd, E., Codd, S., Salley, C.: Providing OLAP (On-line Analytical Processing) to User-Analysts: An IT Mandate. Codd & Date, Inc, San Jose (1993)
4. Davis, A., de Vreede, G.J., Briggs, R.: Designing thinklets for convergence. In: AMCIS 2007 Proceedings (2007), http://aisel.aisnet.org/amcis2007/358
5. Ducassé, M., Ferré, S.: Fair(er) and (almost) serene committee meetings with logical and formal concept analysis. In: Eklund, P., Haemmerlé, O. (eds.) Proceedings of the International Conference on Conceptual Structures. Springer-Verlag (July 2008), lecture Notes in Artificial Intelligence 5113
6. Ferré, S., Ridoux, O.: A logical generalization of formal concept analysis. In: Mineau, G., Ganter, B. (eds.) International Conference on Conceptual Structures. pp. 371–384. No. 1867 in Lecture Notes in Computer Science, Springer (Aug 2000)
7. Ferré, S., Ridoux, O.: An introduction to logical information systems. Information Processing & Management 40(3), 383–419 (2004)
8. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations. Springer, Heidelberg (1999)
9. Hara, N.: Analysis of computer-mediated communication: Using formal concept analysis as a visualizing methodology. Journal of Educational Computing Research 26(1), 25–49 (2002)
10. den Hengst, M., Adkins, M.: Which collaboration patterns are most challenging: A global survey of facilitators. In: HICSS. p. 17. IEEE Computer Society (2007)
11. Kilgour, D.M., Eden, C.: Handbook of Group Decision and Negotiation, Advances in Group Decision and Negotiation, vol. 4. Springer Netherlands (2010)
12. Kolfschoten, G.L., de Vreede, G.J., Briggs, R.O.: Collaboration engineering. In: Kilgour and Eden [11], chap. 20, pp. 339–357
13. Lewis, L.F.: Group support systems: Overview and guided tour. In: Kilgour and Eden [11], chap. 14, pp. 249–268
14. Poelmans, J., Elzinga, P., Viaene, S., Dedene, G.: A case of using formal concept analysis in combination with emergent self organizing maps for detecting domestic violence. In: Perner, P. (ed.) Advances in Data Mining. Applications and Theoretical Aspects, Lecture Notes in Computer Science, vol. 5633, pp. 247–260. Springer Berlin / Heidelberg (2009), http://dx.doi.org/10.1007/978-3-642-03067-3_20, 10.1007/978-3-642-03067-3_20
15. Roth, C., Bourgine, P.: Lattice-based dynamic and overlapping taxonomies: The case of epistemic communities. Scientometrics 69, 429–447 (2006), http://dx.doi.org/10.1007/s11192-006-0161-6, 10.1007/s11192-006-0161-6
16. Vogel, D., Coombes, J.: The effect of structure on convergence activities using group support systems. In: Kilgour and Eden [11], chap. 17, pp. 301–311