Chances and Marketing: On-line Conversation Analysis for Creative Scenario Discussion

Xavier Llorà, **David E. Goldberg**,
Yukio Ohsawa, **Kei Ohnishi**, **Hiroshi Tamura**,
Yuichi Washida, and **Masataka Yoshikawa**

IlliGAL Report No. 2004025
May, 2004

Illinois Genetic Algorithms Laboratory
University of Illinois at Urbana-Champaign
117 Transportation Building
104 S. Mathews Avenue Urbana, IL 61801
Office: (217) 333-2346
Fax: (217) 244-5705

Chances and Marketing: On-line Conversation Analysis for Creative Scenario Discussion

Xavier Llorà[1], **David E. Goldberg**[1], **Yukio Ohsawa**[2], **Kei Ohnishi**[1], **Hiroshi Tamura**[3], **Yuichi Washida**[3], and **Masataka Yoshikawa**[3]

[1]Illinois Genetic Algorithms Laboratory (IlliGAL),
National Center for Supercomputing Applications,
University of Illinois at Urbana-Champaign, Urbana, IL, 61801
{xllora,deg,kei}@illigal.ge.uiuc.edu

[2]Chance Discovery Consortium, University of Tsukuba &
University of Tokyo, Tokyo, Japan, 112-0012
osawa@gssm.otsuka.tsukuba.ac.jp

[3]Research and Development Division, Hakuhodo Inc., Tokyo, Japan, 108-8088
{yuichi.washida,hiroshi.tamura,masataka.yoshikawa}@hakuhodo.co.jp

Abstract

Creativity protocols and methodologies tend to be time consuming if applied manually. This paper presents how innovation support technologies can be applied to collaborative scenario creation and discussion. The fusion of change discovery, genetics algorithms, and computer-supported collaborative tools, allow a group of participants in a creative processes to have a pervasive access to reflection feedback about the scenario to be analyzed in real time. The work summarized in this paper introduces such innovation support technologies, as well as it summarizes a first marketing research workshop powered by such technologies.

1 Introduction

Since the appearances of *Applied Imagination* by Osborn in 1953 (Osborn, 1953), the need for methodological approaches to innovation and creativity has become a key element to success in fields such as decision making, problem solving, or total quality management—to mention a few. Usually, such methodologies require the usage of manual protocols, becoming tedious and counterproductive if not performed correctly. Moreover, techniques such as brainstorming face handicaps arised by the skills of the session leader and the users' fatigue that bound the duration of the manual protocols of such creative sessions.

In the last decade, information technologies have widely spread, beyond business and scientific applications. In early 90's Weiser (Weiser, 1991) envision how information technologies will eventually penetrate most of the facets of human activities. In early 1996, Holt explored the connection between traditional creativity and innovation protocols and the new counterparts supported by information technologies (Holt, 1996) .

This paper presents how innovation support technologies can be applied to collaborative scenario creation and discussion. Innovation technologies, as a combination of change discovery and genetics algorithms, allow a group of participants in a creative processes to have a pervasive access to reflection feedback about the scenario to be analyzed in real time. Such feedback is crucial for providing a visual road-map, by means of KeyGraphs (Ohsawa, Benson, & Yachida, 1998), of the current creative effort, helping the users to navigate through the concepts and chances of the scenario.

The rest of the paper is structured as follows. Section 2 introduces KeyGraphs, the change discovery tool used in our work to provide the visual map of the scenario discussions. We also briefly introduce the connection between genetic algorithms and human innovation in section 3. Section 4 overviews the infrastructure technology for the on-line scenario creation, analysis, and discussion. Such infrastructure was used for real-world scenario discussions. The results are summarized in section 5. Finally, section 6 presents a brief discussion and conclusions of the work presented in this paper.

2 Chance discovery, visual scenario maps, and marketing

The discovery of new product values is always a key element on marketing activities. Usually, for such purposes large data bases containing information gathered in surveys are available. However, even with the usage of data mining techniques for analyzing big volumes of data, there is a remaining need. The marketing groups involved in the creation—or identification—of new product values need to discus different scenarios where such products may become competitive. Moreover, such discussions are potential sources for understanding the different parallel scenario, identifying bridges between them. The rest of this section briefly reviews related concerns of chance discovery (Ohsawa & McBurney, 2003), as well as introduce the main computation embodiment, KeyGraphs (Ohsawa, 2003), used throughout the work presented in this paper.

2.1 On-line scenario analysis and reflection

Participants of a working group proposing and discussing scenarios that describe customers' life, may create valuable information of customers behaviors. Suppose we have two scenarios of customers behaviors, a new scenario may emerge as the result of the discussion among the participants in a creative workshop. If such chance could be identified during the discussion, the scenario understanding and creation would be greatly speedup. Innovation support technologies (Goldberg, Welge, & Llorà, 2003; Llorà, Ohnishi, Chen, Goldberg, & Welge, 2004) focus on providing automated mechanism for such collaborative endeavor. Such technologies help automatize the manual protocols that constraint the duration of such workshops, as well as help to easy users fatigue. The steps in the scenario discussion and creation can be summarized as follows:

1. Data gathering.
2. Data processing and mining.
3. Initial scenario creation.
4. Free discussion and reflection using the initial scenario.
5. Creation of new scenarios.
6. Analysis of relevant data and knowledge available.
7. Free discussion and reflection on the new scenarios.

8. Go back to point 5 until agreement is reached, or time runs out.

Such process, a simplified view of the the so-called double helix model (Ohsawa & McBurney, 2003) for chance discovery, modeled using human-based genetic algorithms (Kosorukoff & Goldberg, 2002), may take advantage of innovation technologies. Computer-supported collaborative tools can archive all the required information. Besides automatizing the data processing and mining, such technologies can track and log the interaction and discussions among participant. If such information is available, chance discovery computational embodiments can produce real-time feedback during the creative workshop. Such immediate feedback help stimulate the creativity of the participants, as we will explain later in this paper.

2.2 KeyGraphs and visual discussion maps on on-line environments

KeyGraphs (Ohsawa, Benson, & Yachida, 1998; Ohsawa & McBurney, 2003) can be applied to documents in order top get a visual map out of its contents. The work presented in this paper used them to analyze the on-line conversations of a group collaborating in a creativity workshop. DISCUS can apply such method, among others, to archived documents, message boards, or chat room logs. Here we assume that a document D is composed of sentences and each sentence is composed of words. The main steps of the KeyGraph algorithm—for a detailed description please see Ohsawa & McBurney (2003) (Ohsawa & McBurney, 2003)—can be outlined as follows.

Document preprocessing. Consists of two tasks: (1) *document compactation* and (2) *phrase construction*. The first one consists on stop-word removal and word stemming using the Porter algorithm (Porter, 1980). The second tasks takes a subset of ℓ_{phrase} words and all the possible combinations out of those words are constructed, retaining only the high frequency ones appearing in the document. Thus, the document D is reduced to a document D' which contains unique terms $w_1, w_2, \ldots, w_\ell$, where w_i refers to either a world or a phrase.

Extracting high-frequency terms. Terms in D' are sorted by their fitness. N_{hf} denotes the set of the top n_{nodes} high-frequency terms, being represented as nodes in a graph G.

Extracting links. Links represent *co-occurrence*-term-pairs that often occur in the same sentence. A measure for co-occurrence of terms w_i and w_j is defined as

$$assoc(w_i, w_j) = \sum_{s \in D'} \min\left(|w_i|_s, |w_j|_s\right), \tag{1}$$

where w_i and w_j are element of the N_{hf}, and $|w_i|_s$ the number of times a term w_i occurs in a sentence s. The *assoc* values are computed for all term in N_{hf}. The term-pairs are sorted by their *assoc* values and the top n_{links} term-pairs are represented by edges in G.

Extracting key terms. Key terms are terms that connect clusters of high-frequency terms together. To measure the tightness with which a term w connects a cluster, the following function is defined:

$$key(w) = 1 - \prod_{g \subset G} \left[1 - \frac{based(w, g)}{neighbors(g)}\right] \tag{2}$$

where g is a cluster, and

$$based(w, g) = \sum_{s \in D'} |w|_s, |g - w|_s, \tag{3}$$

$$neighbors(w) = \sum_{s \in D'} \sum_{w \in s} |w|_s, |g - w|_s, \quad (4)$$

where $|g - w|_s = |g|_s - |w|_s$ if $w \in g$, and $|g|_s$ otherwise, being $|g|_s$ the number of times a cluster g occurs in a sentence s. The *key* values are computed for all the terms in D' and sorted accordingly. The top n_{key} terms—the K_{hk} set—are added as nodes to G if they were not present previously.

Extracting key links. For each high-frequency term $w_i \in N_{hf}$, and each key term $w_j \in K_{hk}$, $assoc(w_i, w_j)$ is calculated. Links touching w_j are sorted by their *assoc* values for each key term $w_j \in K_{hk}$. A link with highest *assoc* values connecting w_j to two or more clusters is chosen as a key link. Key links are represented by edges—if they are not already present—in G.

Extracting Keywords. Nodes in G are sorted by the sum of *assoc* values associated with the key links touching them. Terms represented by nodes of higher values of these sums than a certain threshold are extracted as keywords for the document D.

3 The innovation intuition and creative processes

Genetic algorithms are a core technology for the *innovation technology* endeavor. Starting in 1983 (Goldberg, 1983), Goldberg (Goldberg, 2002b) developed the so called *fundamental intuition of genetic algorithms*, or the *innovation intuition*. Specifically, the innovation intuition of GAs is about the work together of: (1) selection and mutation, and (2) selection and recombination. Moreover, the innovation intuition of GAs provide a facet-wise modeling of human innovation. This approach models two orthogonal *facets of human innovation*.

Selection + mutation = Continual improvement. Selection and mutation working together are a form of hill-climbing mechanism. Mutation suggests variants in the neighborhood of the current solutions; selection acts as the decision process which accepts improving changes with a high probability. This simple model describes one of the facets of human innovation, the so called *continual improvement* in total quality management literature, or as Japanese call it, *kaizen*.

Selection + crossover = innovation. Another facet of human innovation is the so called *cross-fertilizing innovation*. People usually grasp a set of good solution features in one context, and a notion in another context and juxtaposing them, thereby speculating that the combination might be better than either notion taken individually. Taking together selection and crossover, GAs are a computation model of *cross-fertilizing innovation*.

GAs also are main role players for the *innovation technology* revolution. As early mentioned, humans are to become the main measure of such a technology. Pervasive GA-guided interaction between human and computers opens a new research path to creativity- and innovation-support. Two well-known models of such support are interactive GAs, and human-based GAs. Interactive GAs (iGAs) replace the computer computation of the relative fitness of solutions and the selection process by the judgment of a human evaluation. More detailed information about the progress of interactive GAs and interactive evolutionary computation (iEC) are presented in a review by Takagi (Takagi, 2001). Whereas iGAs replace the evaluation and selection by the human judgment, human-based GAs (HBGAs) (Kosorukoff & Goldberg, 2002) move one step further and permit evaluation, selection, and variation to be performed by a human. For such reasons, the previous facets of GAs may be regarded as a first order model of human innovation.

4 DISCUS, an integration framework

The use of modern tools from artificial and computational intelligence to model and lead the interaction between participants may be regarded as a kind of human-machine collaboration. To achieve effective and scalable collaboration in difficult problem domains, the DISCUS project [1] (*Distributed Innovation and Scalable Collaboration in Uncertain Settings*) envisions the fusion of human-human collaboration (HHC) and human-machine collaboration (HMC)—as primary ingredients for the creative innovation, collaboration, and decision support management. To achieve these goals, DISCUS fuses a number of key elements: (1) interactive genetic algorithms (iGAs) (Takagi, 2001), (2) human-based genetic algorithms (HBGAs) (Kosorukoff & Goldberg, 2002), (3) scalable genetic algorithms (machine based) (Goldberg, 2002a), (4) flexible data and text mining (D2K/T2K) (Welge, Auvil, Shirk, Bushell, Bajcsy, Cai, Redman, Clutter, Aydt, & Tcheng, 2003), and (5) chance discovery using KeyGraphs (Ohsawa, Benson, & Yachida, 1998). Chance discovery—when used together with genetic algorithms—provides a natural mechanism to identify salient fortuitous events, as Goldberg, Sastry, & Ohsawa described (Goldberg, Sastry, & Ohsawa, 2003). A detailed description of DISCUS is beyond the scope of this paper. We refer the interested reader to Goldberg, Welge, & Llorà (2003) (Goldberg, Welge, & Llorà, 2003), and Llorà et al. (2004) (Llorà, Ohnishi, Chen, Goldberg, & Welge, 2004). Some of the DISCUS features that supported the work presented in this paper may be summarized as follows.

Data mining for marketing surveys. Using D2K (Welge, Auvil, Shirk, Bushell, Bajcsy, Cai, Redman, Clutter, Aydt, & Tcheng, 2003), DISCUS is able to deal with large volumes of information. Such capabilities were used for extracting association rules out of the questionnaires available. Using such knowledge about customer behavior data can be then filtered and selected, becoming the base of the first scenarios creation.

KeyGraph on filtered data. Once the filtered data is available, several initial KeyGraphs can be generated using DISCUS. If the data represents the world of daily life of people behavior, an event may represent an answer to a question in a questionnaire about daily lifestyle. By visualizing the map where answers appear in a graph, as figure 1 shows, one can see the overview of the behaviors of survey subjects (Ohsawa, 2003).

On-line collaboration tools and archives. Once the visual maps provided by KeyGraphs are available, the participants on the workshop discus about these initial scenarios using DISCUS collaboration technologies. The participants have access to enhanced collaboration tools including message boards, chat rooms, and video-conferencing capabilities—to mention a few. The particular layout during the work summarized here involved that the participants only interact with each other using a message board, as shown in figure 2. This setup archives all the information related with the discussion of the scenarios.

Real-time scenario discussion using KeyGraphs. DISCUS archival capabilities open another dimension on innovation support. The archived logs of the discussions can be analyzed while the discussion is active. KeyGraphs of the conversation logs (Llorà, Ohnishi, Chen, Goldberg, & Welge, 2004) present a visual map of the discussion sparked by the original KeyGraphs obtained from the survey data. The goal of displaying such visual maps during a discussion is twofold. First, the maps summarize and focus the discussion held so far. Second, the maps identify relevant scenarios and bridges between them, helping the participants to push the creative process toward new innovative trends.

[1] `http://www-discus.ge.uiuc.edu`

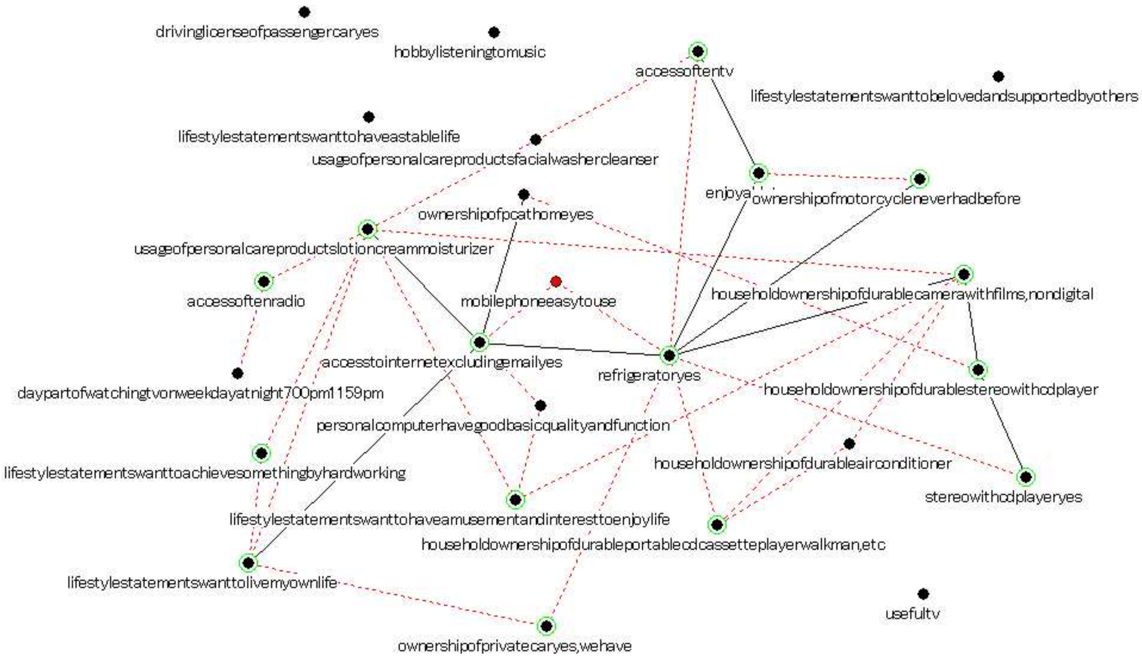

Figure 1: KeyGraph generated out of the filtered marketing data.

5 Marketing, innovation, and on-line creativity

The approach presented above was tested on a real marketing research workshop. Four researchers from the Chance Discovery Consortium and Hakuhodo Inc. used the DISCUS platform to perform a creativity workshop. Five more researchers from the Illinois Genetic Algorithms Laboratory also joined the workshop. Together, they engaged the eight step process introduced in section 2.1. The goal was to discuss marketing needs for the cell phone market in the States.

Before engaging any discussion, the researchers in the workshop used DISCUS capabilities to explore the available data from marketing surveys. Such data base was mined extracting association rules. Such rules helped to characterize and identify the target profile of the workshop: *single young women*. Once the researchers agreed on such target, DISCUS tools extracted all the available information that matched such profile. Such extraction process—combined with the appropriate data transformations—provide the input for the creation of the initial scenarios. Visual maps provided by the KeyGraph analysis of such data are the initialization seed for the discussion powered by DISCUS. Figure 1 presents one of the KeyGraphs used to start the discussion. It presents the visual map of the questions and answers data, as introduced in section 4.

Once the KeyGraphs were ready, the participants joined the discussion about how mobile phones may be used by young single women. Such discussion took place using the computer-supported collaborative tools integrated in DISCUS. All the interactions among participants were conducted through DISCUS, archiving all the relevant information. The participants had access to collaboration tools, search engines, intelligent document retrieval utilities, and chance discovery-guided

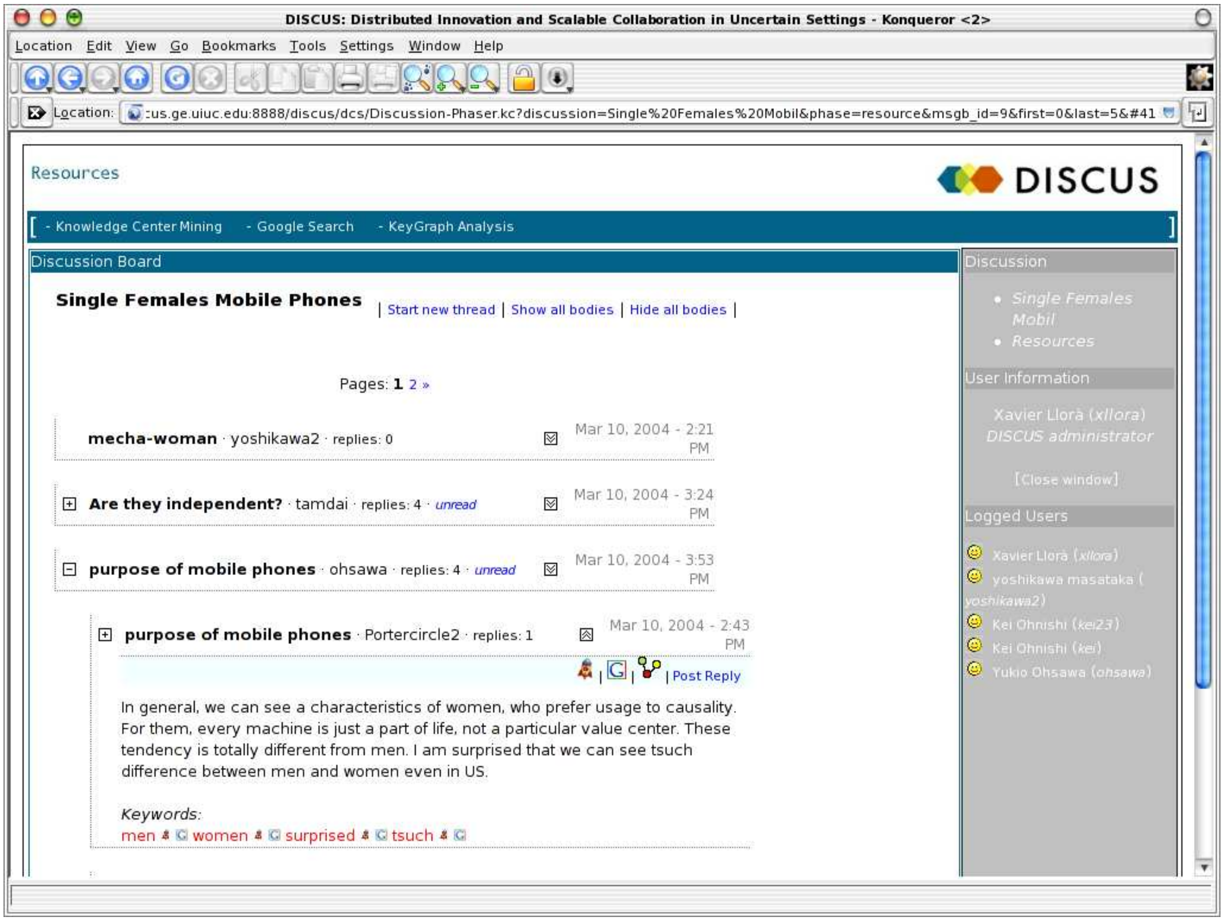

Figure 2: An scenario discussion using DISCUS on-line collaboration tools.

searches—to mention a few. DISCUS integrates all this elements through an intuitive, user-oriented, web interface. Figure 2 presents a screenshot of DISCUS platform during the discussion.

The participants started discussing using the initial scenarios. At the early states, participants were already using the available tools, guided by the innovation support approach of DISCUS technology. The first visible impact of such support was the dynamics of the creative process. In traditional approaches to creativity for groups, the divergence and the convergence phases for idea creation and evaluation usually appear interleaved, having a few alternances before users get fatigued. Such approach require long times elapsed in each step, as the result of the constrained stages where all the participant focus on the same activity. Often, this made the total time too long for a realistic business applications. However, DISCUS participants experimented fast overlapping stages of divergent and convergent thinking. Such speedup was mainly provided by the fast feedback that the DISCUS platform provides. Real time analysis of the discussions using KeyGraphs helped users to have a fast externalization of the ideas involved on the discussion. Figure 3 illustrates how, with a simple mouse click, the participants can analyze a given message, or the whole discussion.

The participants also used the KeyGraphs built using the discussion information in a particular manner. Participants found this discussion like a game, where a success in connecting distant parts of the KeyGraph structure was regarded as a win at each time in the discussion. This also means that participants sometimes felt like connecting the superficial outlook of KeyGraph structure, a kind of convergence thinking connecting different concepts. Either way, the scenarios proposed

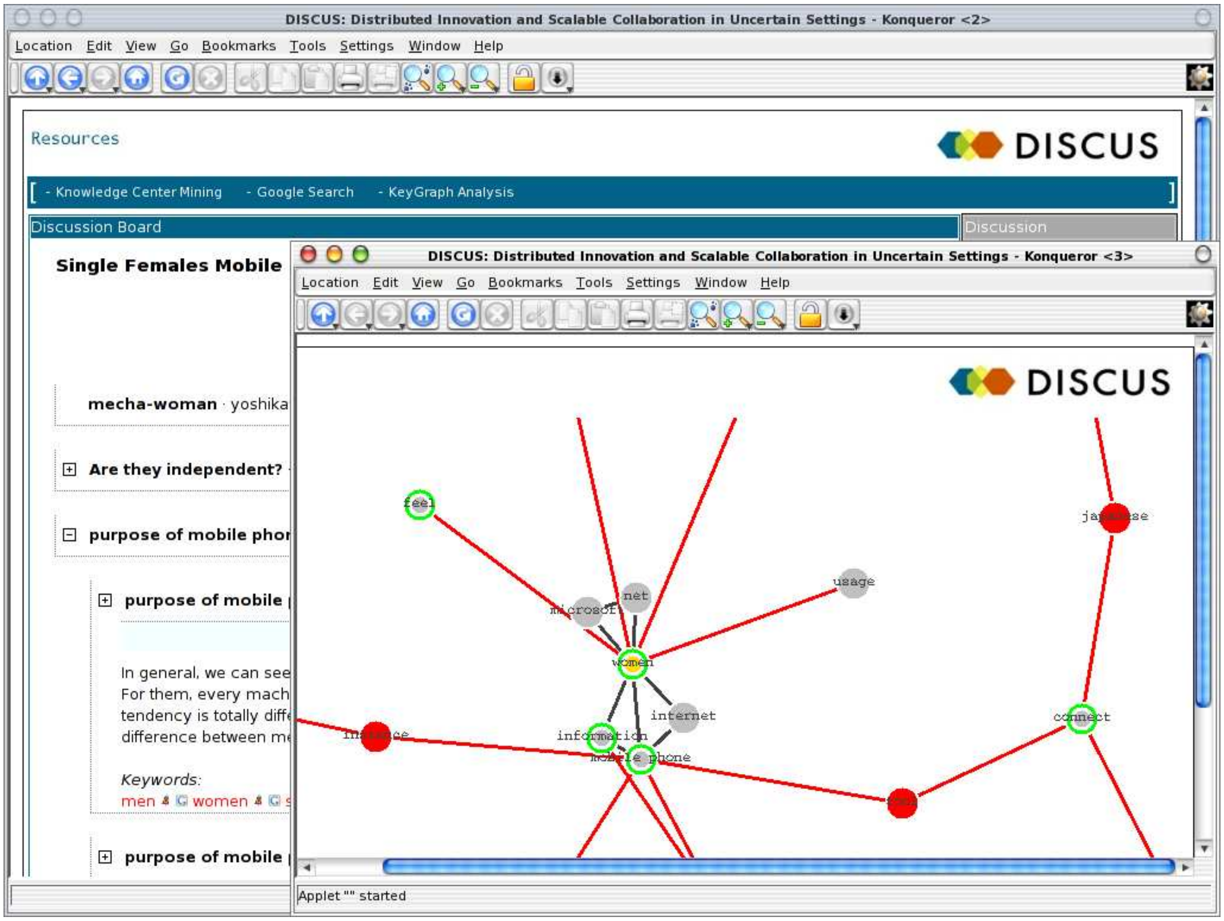

Figure 3: An scenario discussion using DISCUS tools enriched with KeyGraph analysis.

never went beyond the context of the discussion. Sometimes, a new message failed to connect the islands in a KeyGraph against the will of its sender. Other times, a posted message appeared isolated from the previous existing islands in the KeyGraph. In such a case, the discussion went in a divergent direction, introducing new topics and concepts. However, due to the intention of the sender to connect the KeyGraph structure, the contextual distance of the new message was within the reach of other participants' later discussions. This made the overall KeyGraph become connected afterward—convergence thinking again.

This effect of quick interleaving of divergences and convergences accelerated the crossovers of scenarios, and participants obtained many clues to propose new ways to look at the market of mobile phones. More significantly, the feeling of playing a game entertained the participants, thus none of them dropped out of the communication for more than two hours of high quality discussions. Around 37 messages were posted; almost one every five minutes. The half a day workshop powered by DISCUS produced high quality new marketing research scenarios in a record time. The whole process described in section 2.1, including the data processing, mining, and initial scenario creation, was done in a single day. The same kind of workshop following manual protocols may have taken almost three times more the time spent using DISCUS framework. Moreover, key elements as the on-line real time feedback may not be possible without the innovation support technologies of DISCUS.

6 Conclusions

This paper has presented how innovation support technologies can be applied to collaborative scenario creation and discussion. The paper has also introduced how innovation support technologies can help integrate in a single platform—the so-called DISCUS project—all the required steps for scenario discussion on a creativity-oriented marketing workshop. Special attention in this paper was paid to the innovation technology. On-line collaboration tools were enhanced with real-time scenario analysis by means of KeyGraph. Thus, the participants were able to have an instant insight and reflection about the discussion concepts and trends. Finally the paper has briefly reviewed a real workshop on marketing research powered by DISCUS.

Using DISCUS, the participants of the workshop were able to explore and analyze the available data base obtained from customer's surveys at Hakuhodo. The filtered data was used to create the initial scenarios for the discussion. Such process was totally driven by the marketing researchers using the DISCUS tools. Later, the researchers engage the scenario discussion using DISCUS. The interactive real-time feedback provided by DISCUS, turn the creative workshop into a game-flavored activity. Such approach helped to fight the participants fatigue on such kind of workshops, as well as it greatly stimulate their creativity, as the high-quality outcome of the marketing research discussion showed.

Acknowledgments

This work was sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant F49620-03-1-0129, and by the Technology Research, Education, and Commercialization Center (TRECC), at University of Illinois at Urbana-Champaign, administered by the National Center for Supercomputing Applications (NCSA) and funded by the Office of Naval Research under grant N00014-01-1-0175. The US Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Office of Scientific Research, the Technology Research, Education, and Commercialization Center, the Office of Naval Research, or the U.S. Government.

References

Goldberg, D. E. (1983). Computer-aided gas pipeline operation using genetic algorithms and rule learning. *Dissertation Abstracts International*, *44*(10), 3174B. Doctoral dissertation, University of Michigan.

Goldberg, D. E. (2002a). *The design of innovation: Lessons from and for competent genetic algorithms*. Norwell, MA: Kluwer Academic Publishers.

Goldberg, D. E. (2002b). *The design of innovation: Lessons from and for competent genetic algorithms*. Kluwer Academic Publisher.

Goldberg, D. E., Sastry, K., & Ohsawa, Y. (2003). Discovering deep building blocks for competent genetic algorithms using chance discovery. In *Ohsawa, Y. & McBurney, P. (eds.), Chance Discovery* pp. 276–301.

Goldberg, D. E., Welge, M., & Llorà, X. (2003). *DISCUS: Distributed Innovation and Scalable Collaboration In Uncertain Settings* (IlliGAL Report No. 2003017). Urbana, IL: University of

Illinois at Urbana-Champaign, Illinois Genetic Algorithms Laboratory.

Holt, K. (1996). Brainstorming-from classics to electronics. *Engineering Design*, *7*(1), 77–82.

Kosorukoff, A., & Goldberg, D. E. (2002). Evolutionary computation as a form of organization. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2002)* pp. 965–972. Morgan Kaufmann.

Llorà, X., Ohnishi, K., Chen, Y.-P., Goldberg, D. E., & Welge, M. (2004). *Enhanced Innovation: A Fusion of Chance Discovery and Evolutionary Computation to Foster Creative Processes and Decision Making* (IlliGAL Report No. 2004012). Urbana, IL: University of Illinois at Urbana-Champaign, Illinois Genetic Algorithms Laboratory.

Ohsawa, Y. (2003). *Chance discovery* (Chapter KeyGraph: Visualized Structures Among Event Clusters. Springer.

Ohsawa, Y., Benson, N. E., & Yachida, M. (1998). KeyGraph: Automatic indexing by co-occurence graph based on building construction metaphor. In *Proceedings of Advances in Digital Libraries* pp. 12–18.

Ohsawa, Y., & McBurney, P. (2003). *Chance discovery*. Springer.

Osborn, A. F. (1953). *Applied imagination*. Scribners, New York.

Porter, M. (1980). An algorithm for suffix stripping. *Automated Library and Information Systems*, *14*(3), 130–137.

Takagi, H. (2001). Interactive evolutionary computation: Fusion of the capabilities of EC optimization and human evaluation. *Proceedings of the IEEE*, *89*(9), 1275–1296.

Weiser, M. (1991). The computer for the twenty-first century. *Scientific American*, *September*, 94–100.

Welge, M., Auvil, L., Shirk, A., Bushell, C., Bajcsy, P., Cai, D., Redman, T., Clutter, D., Aydt, R., & Tcheng, D. (2003). *Data to Knowledge (D2K)* (Automated Learning Group Technical Report). Urbana, IL: National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign.