

Please quote as: Riedl, C.; Wagner, S.; Leimeister, J. M. & Krcmar, H. (2010): Exploring Large Collections of Ideas in Collaborative Settings through Visualization. In: 20. Annual Workshop on Information Technologies and Systems (WITS) 2010, St. Louis, MO, USA.

Exploring Large Collections of Ideas in Collaborative Settings through Visualization

Christoph Riedl,¹ Steffen Wagner,¹ Jan Marco Leimeister,² Helmut Krcmar¹

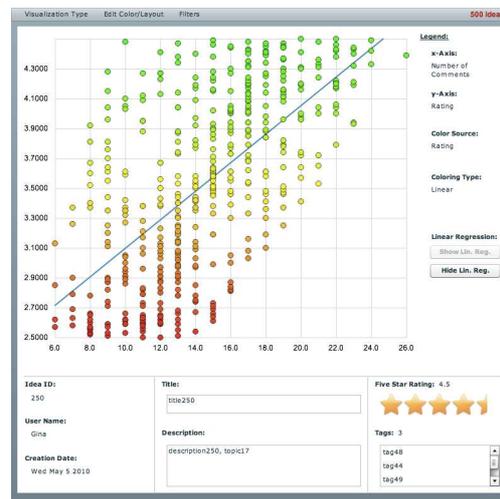


Figure A IdeaViz: Visualizing large collections of ideas in open innovation communities.

Abstract – As innovation becomes more and more important companies increasingly turn to their customers, suppliers, and other external actors to collect ideas for new products and services. Following such an open innovation approach organizations have launched online innovation communities where users collaboratively generate and refine ideas. These communities can generate large amounts of ideas in relatively short time periods. For example, the Starbucks innovation community *mystarbucksidea.com* contains over 88,000 ideas after just two years. To harness the value created through this collaboration and make resulting data usable for later stages of the innovation process we developed a visualization to explore large pools of user-generated ideas. We empirically collected requirements through expert interviews with innovation managers which we then translated into a concrete implementation of a Web-based visualization tool. The visualization serves two purposes. First, it offers a way to explore large idea pools resulting from collaborative idea generation activities. Second, it can be used to generate new input to further develop ideas through collaboration, e.g., by combining similar ideas.

Keywords: Open innovation, innovation community, collective intelligence, collaboration, visualization, human-computer interaction.

Acknowledgement: This research received funding from the German Federal Ministry of Economics and Technology (BMWi) under grant code 10MQ07024. The responsibility for the content of this publication lies with the authors.

¹ Technische Universität München, Boltzmannstr. 3, 85748 Garching, Germany,
{riedlc | wagners | krcmar @in.tum.de}

² Universität Kassel, Nora-Platiel-Str. 4, 34127 Kassel, Germany, leimeister@uni-kassel.de

Exploring Large Collections of Ideas in Collaborative Settings through Visualization

Christoph Riedl, Steffen Wagner, Jan Marco Leimeister, Helmut Krcmar

Introduction

Innovation is considered a key success criterion for companies. As competition increases and innovation cycles get shorter due to rapid technological developments in a globalizing world, companies increasingly focus on their innovation activities. Most notably, organizations are opening up their innovation processes to allow an inflow of innovations from outside the organization (Chesbrough, 2006). Following this open innovation paradigm companies increasingly turn to their customers, suppliers, and other external actors like specialized R&D labs to collect ideas for new products and services (Chesbrough, 2006).

Using Web 2.0 technologies, large-scale collaboration has been enabled which makes the integration of a large group of actors into an organization's innovation process possible. Several organizations have launched online innovation communities where users collaboratively generate and refine ideas. These communities can generate large amounts of ideas in relatively short time periods which can thus result in large data pools. For example, the Starbucks innovation community *My Starbucks Idea*³ contains over 88,000 customer-generated idea submissions. Obviously, generating ideas easily is a big advantage of open innovation processes but it is also a huge challenge with regard to the necessary analysis of this data. Current innovation communities employ simple list-based representations to explore these idea submissions which have severe limitations as the number of data entries increases. To harness the value created through collaboration in online innovation communities and make the resulting data usable for later stages of the innovation process, means to analyze, filter, and refine this data are necessary. Due to their abstract nature, ideas have to be analyzed by humans which limits the applicability of automated information retrieval approaches. Visualization helps humans to gain an overview of large amounts of data to develop insights into the data.

This research addresses the research question how large idea pools resulting from collaborative innovation settings can be visualized in order to gain overview and insights into an idea pool, with the ultimate goal of extracting the most promising ideas for subsequent steps of the innovation process. Our research employs information visualization techniques in a real world setting that suffers from large pools of unstructured data that need to be processed in order to facilitate further use. Our work contributes to collaboration research, in particular collaboration in open innovation communities, by presenting an approach how collaboration results can be visualized in order to facilitate subsequent steps of collaboration.

We chose information visualization as a fruitful approach to address the challenge of exploring the large amounts of data collected through online innovation communities as interactive, graphical visualization allow to gain better understanding of data (Ware, 2004). In presenting large amounts of data in a human understandable way users of a computer visualization can

³ <http://mystarbucksidea.com/> last accessed 2010-10-10.

better comprehend this data, gain an overview of present data, and discover emergent properties and unknown relationships (Ware, 2004).

The visualization serves two purposes. First, it offers a way to explore large data pools resulting from collaborative idea generation activities. Second, using the insights gained through the visualization new input can be generated to further develop ideas through collaboration. In particular, relationships discovered in the data can be used to combine related ideas which can serve as new input for later collaboration activities.

Empirical State of the Art Analysis

The visualization of large idea pools has not been addressed systematically in scientific research and no specialized domain visualizations have been reported so far. We therefore analyzed existing idea portals regarding the display modes of ideas that currently form the dominant design. We performed detail analysis of five portals which contain at least 175⁴ ideas: (1) SAPIens,⁵ (2) My Starbucks Idea,³ (3) IdeaStorm,³ (4) Ubuntu Brainstorm,³ and (5) IdeaExchange.³ The current dominant design of idea representation in online innovation portals is to display ideas in a list. The individual list items are ideas itself, usually displayed using a title, a short description, and additional meta-data such as an author name and publication date. All analyzed portals use this mode of presentation. Long lists are broken into sub-pages following the *pagination* design pattern with 5 to 20 ideas per page (Tidwell, 2006). For the smallest portal, SAPIens with 175 ideas, this results in 30 sub-pages, for the largest, My Starbucks Idea with over 88,000 ideas, this results in currently 475 sub-pages on the main category. Individual pages can be accessed by next/previews buttons as well as the page numbers. Most portals organize ideas into categories but ideas within categories follow the same list-based view. Additional interaction includes the sorting of lists (e.g, by popularity, by user rating, by date) and a keyword search. To display aggregated information SAPIens uses a tag cloud. The other portals offer no aggregated view.

The list-based visualization currently used by innovation portals offers extensibility and the benefit of, in theory, displaying large amounts of data. Shortcomings of the list-based view are space requirements which mandate that long lists be broken up into sub-pages. This forces a tradeoff on the visualization design between the number of items on a single page and the number of resulting sub-pages. As the example of My Starbucks Idea shows, this can result in the need to navigate many sub-pages to gain an overview of the idea pool. Furthermore, as lists emphasize symbolic information, logical relationships between elements are difficult to display as this would require emphasize on spatial information (Vessey, 1991). Accordingly, patterns in data are usually not discovered and non-related items cannot easily be compared. A general shortcoming of the list-based view is that data is not aggregated above the level of an idea which fails to provide the necessary abstraction to make large idea pools navigable. Consequently, the theory of “cognitive fit” developed by Vessey serves as the basis for this research.

⁴ All numbers as of 2010-10-10.

⁵ <http://www.sapiens.info/>, <http://mystarbucksidea.com/>, <http://www.ideastorm.com/>, <http://brainstorm.ubuntu.com/>, <http://sites.force.com/ideaexchange> last accessed 2010-10-10.

IdeaViz – A Visualization for Large Idea Pools

We developed *IdeaViz*,⁶ a Web-based visualization for the exploration and analysis of idea pools, more precisely user-generated new product or service ideas, in online innovation communities. The focus of this research was on the domain problem characterization rather than on the specific algorithm design (Munzner, 2009). Consequently, we focused on requirements elicitation and definition to gain the necessary domain insights. Based on the findings, the visualization has been implemented in Flash using the open-source visualization library Flare.⁷

Design Requirements

When designing IdeaViz we were aware that our proposed user group and target domain had several unique properties. To collect domain requirements we performed three interviews with key users experienced in innovation management during January 2010. Each interview was between 30 and 40 minutes in length. We asked key users to describe tasks that they regularly perform on the idea pool they are responsible for. We also asked which insights into the idea pool they are most interested in and answering which questions is most valuable to them. The following list summarizes the aggregated lead questions resulting from the interviews: (1) How many ideas are in the idea pool? (2) Who are the most active users? (3) Which ideas have been most commented? (4) Which ideas are similar? (5) Which topics/tags do ideas belong to? (6) Are there groups of related ideas (which could be combined)? (7) Are there relationships between ideas (other than tags)? (8) How have idea submissions/comments developed over time?

Similar lead questions related to innovation management have also been reported by Riedl et al. (2009). Based on these lead questions, generic requirements of information visualization, and the shortcomings we observed in the visualizations currently used, we derived requirements for IdeaViz (Table 1).

Table 1 Summary of requirements based on expert interviews.

<i>Generic / non functional requirements</i>	<i>Domain specific / functional requirements</i>
<ul style="list-style-type: none"> • expressivity • effectiveness • good usage of space • scalability • performance 	R1. display all attributes of an idea R2. sort ideas by different attributes R3. filter by attributes to allow data reduction R4. contrast the relationship of attributes R5. explore development over time R6. explore relationship between ideas R7. edit ideas within the visualization R8. partial loading/displaying of large idea pools (not implemented)

Data and Interface

In the following, we describe the general interface of IdeaViz and what data is available to the user. For the basic architecture of the visualization, we chose the model-view-controller (MVC) pattern which allows separating and isolating the user interface (view) from the data (model) and

⁶ Interactive online demo available at <http://tinyurl.com/3akg92e>

⁷ <http://flare.prefuse.org/> last accessed 2010-10-10.

domain/business logic (controller). As a result, the interface with its views (and the controller component) can be easily connected to different sets of data with only minor changes. To develop and test the visualization we used a sample data set of 500 randomly selected ideas from My Starbucks Ideas which we collected in our own database. To specify the data to be displayed we follow the data schema defined by the Idea Ontology, an ontology for idea representation (Riedl et al., 2009). The Idea Ontology defines a central idea class consisting of attributes such as title, description, date, author, rating, tagging, as well as attached comments or documents.

Implemented Views

The visualization contains four views which can be accessed through the “Visualization Type” menu. Each view has specific functions to answer the different lead questions collected above. Ideas are displayed as nodes in the form of small circles. Based on this general encoding of ideas as nodes two different visualization forms can be distinguished: scatter plots and graphs. In scatter plots, information is encoded in the position and the color of the node. In the graph views, on the other hand, information is encoded in the linking of nodes through edges. In the graph view, the position of a node holds no information. Content properties of an idea such as title or tags are not usually displayed. At the bottom of each screen the complete details of an individually selected idea is visible (e.g., Figure 1). The detail view displays the main content of a single idea (e.g., title, author, and description) and thus addresses R1. In general, users can access this information by hovering over a particular idea with the mouse. The details view can also be used to edit individual ideas within the visualization (R7). Regarding the interaction design it is possible to select individual or groups of ideas with the mouse in order to apply filters in all views (“Filters” menu).

Table View: The table view (Figure 1) shows two data grids. The first is author-oriented and presents data aggregated by idea author; the second is idea-oriented and shows all ideas. In both data grids, the columns can be sorted. In the idea-oriented table, it is also possible to select a particular idea in order to retrieve the content information for that idea.

This view addresses in particular R1 (details) as well as R2 (sort). The main objective of the table view is to allow symbolic tasks such as looking for a particular idea and selecting individual ideas based on sorting (e.g., select the 10 most commented or highest rated ideas).

Scatter Plot View: The scatter plot (Figure 2) is the most flexible visualization type. Ideas are presented as nodes in a diagram with two axes. Each axis can be configured separately and therefore allows the user to analyze the data. In the picture, the axes are set to "Number of Comments" (x-axis) and "Rating" (y-axis). Additionally, a linear color encoder is active which colors the nodes depending on their rating. The color coding is freely configurable by the user. Furthermore, a linear regression line is visible (which is possible since both axes are configured with numeric scales). Within the scatter plot groups of ideas can be selected and then filtered based on a rectangular graphical selection.

In addition to R2 (sort), and R3 (filter), the scatter view allows in particular to contrast two attributes (R4). As shown in Figure 2, this view can be used to discover a relationship between the two attributes “rating” and “number of comments.” Furthermore, it emphasizes spatial information and thus facilitates comparison of related attributes.

User	Ideas	Min. Rating	Max. Rating	Avg. Rating
Eva	57	2.51	4.43	3.61
Gina	55	2.53	4.5	3.48
Alice	54	2.59	4.49	3.59
Frank	54	2.52	4.49	3.57
Dave	53	2.59	4.4	3.55
Carl	51	2.54	4.5	3.62
Bob	49	2.51	4.46	3.53
Kylie	47	2.5	4.46	3.38

Idea ID	User	Creation Date	Rating	Comments
250	Gina	Wed May 5 2010 12:00:00 AM	4.5	22
61	Carl	Thu Aug 14 2008 12:00:00 AM	4.5	20
139	Frank	Sat Aug 22 2009 12:00:00 AM	4.49	15
344	Alice	Sat Jan 26 2008 12:00:00 AM	4.49	24
403	Gina	Thu Nov 4 2010 12:00:00 AM	4.48	17
117	Isabelle	Mon Sep 29 2008 12:00:00 AM	4.48	21
474	Isabelle	Sat Jul 5 2008 12:00:00 AM	4.48	10
145	Frank	Sun Aug 15 2010 12:00:00 AM	4.47	19
499	Frank	Thu Mar 5 2009 12:00:00 AM	4.47	13
492	Gina	Fri Aug 14 2009 12:00:00 AM	4.47	18

Idea ID: 250	Title: title250	Five Star Rating: 4.5 ★★★★★
User Name: Gina	Description: description250, topic17	Tags: 3 tag48 tag44 tag49
Creation Date: Wed May 5 2010		

Figure 1 Table view.

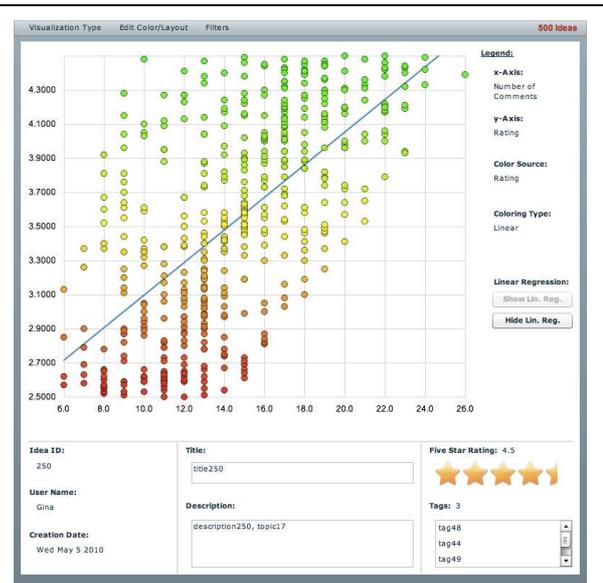


Figure 2 Scatter plot view.

Timeline View: The timeline view (Figure 3) is a specialized version of the scatter plot. It extends the scatter plot with lines (edges) between nodes to create a timeline for a single or multiple author(s) of ideas. In this view, the visualization user can compare multiple timelines and trim them, i.e. filter/remove nodes of the timeline at the beginning or the end, by specifying the start and the end date in the popup shown at the bottom of the picture.

Supplemental to R2 (sort), R3 (filter, in this case by author and time), and R4 (contrast), the timeline view allows to explore development over time (R5).

Graph Plot View: The graph view (Figure 4) displays syntactic relationships between ideas through a graph with a force-directed layout. This axes-free layout allows the user to focus on the relationships (edges) rather than the ideas (nodes) and their particular content. By double-clicking on a node of a cluster, the user gets a more detailed visualization of the cluster (titles of the ideas are visible inside the visualization layout). Depending on the number of directly connected nodes, called degree, the nodes (ideas) are colored categorically and the tooltip shows the maximum degree of the cluster. We use a simple syntax-based word count to determine similarities in idea descriptions. However, tags or other automated text analysis methods could also be used to determine a relationship between ideas.

In addition to R3 (filter, in this case by degree and topic/cluster), the graph view focuses on displaying relationships between ideas (R6) and thus focuses on supporting spatial tasks. The view allows identifying clusters of similar ideas which could be combined. The view also allows the assessment of major topic areas and allows assessing their number and relative size.

Discussion & Future Work

In this paper we presented an overview of the challenges regarding the analysis of large idea pools resulting from online innovation communities. The domain requirements for IdeaViz have been collected through interviews with experienced key users. The interviews stress in particular the need of support for identifying clusters of similar ideas. We also created an interactive

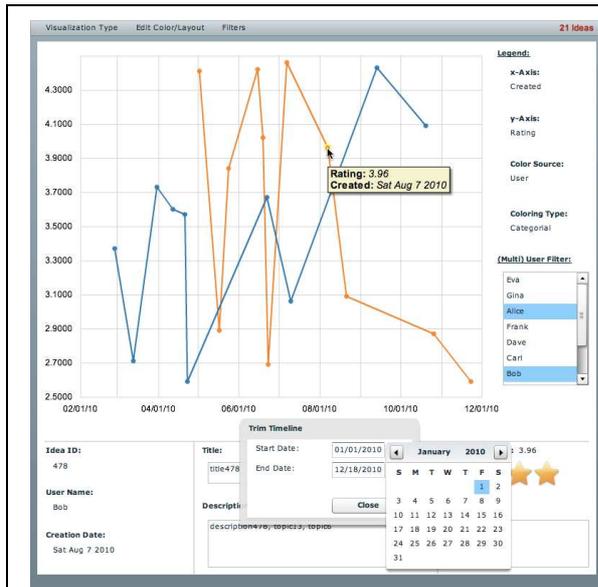


Figure 3 Timeline view.

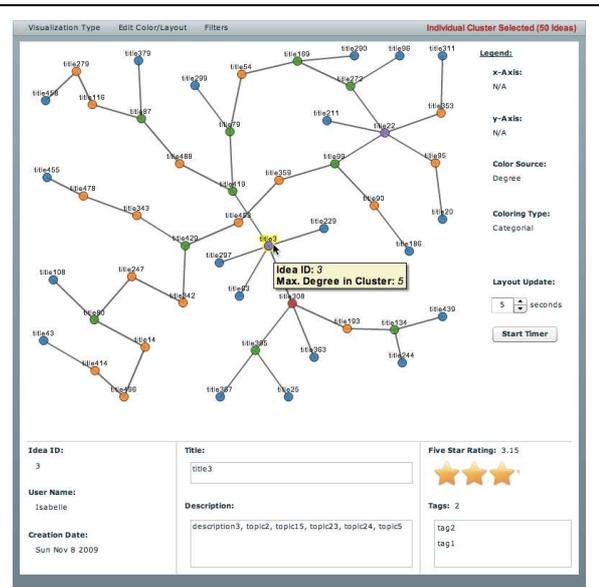


Figure 4 Graph view.

visualization for this domain. The visualization addresses these domain specific requirements (R1-R8). An initial walk-through of the visualization prototype with the key users confirms that the visualization could indeed be used to answer most of the relevant questions, in particular regarding the identification of related ideas (graph view) and the flexible sorting, filtering, and comparison of ideas (scatter plot). This initial walk-through used a data set of 500 ideas, 50 tags (with 2188 idea-to-tag assignments), and 250 edges (i.e., relationships) which also confirms the scalability of our solution in terms of data set size and complexity. As the visualization is Web-based it could easily be integrated into an online innovation portal. Future research should explore a formal evaluation of the visualization and further refine the domain requirements in order to enhance features that are most valuable to innovation managers and key users. Future research should also more closely analyze the “task-visualization”-fit to guide potential user to determine which visualization is most useful to answer a specific type of question.

References

- Chesbrough, H. *Open Innovation: The New Imperative for Creating and Profiting from Technology*, Harvard Business School Press, Boston, MA, USA, 2006.
- Munzner, T. “A Nested Process Model for Visualization Design and Validation,” *IEEE Transactions on Visualization and Computer Graphics* (15:6), 2009, pp. 921-928.
- Riedl, C.; May, N.; Finzen, J.; Stathel, S.; Kaufman, V. & Krcmar, H. “An Idea Ontology for Innovation Management,” *International Journal on Semantic Web and Information Systems* (5:4), 2009, pp. 1-18.
- Tidwell, J. *Designing Interfaces: Patterns for Effective Interaction Design*, O'Reilly, Sebastopol, CA, USA, 2006.
- Vessey, I. “Cognitive Fit: A Theory-Based Analysis of the Graphs Versus Tables Literature,” *Decision Sciences* (22:2), 1991, pp. 219-240.
- Ware, C. *Information Visualization: Perception for Design*, Morgan Kaufmann, San Francisco, CA, USA, 2004.