Where are They?

Using Text Analytics to Select the Right Idea

Andera Gadeib
CEO
Dialego AG
Karmeliterstr. 10
52064 Aachen
andera@gadeib.de

Abstract: Text analytics reduces complex information vital to the innovation process. Delivery is fast and objective. Text analytics identifies potential target areas for innovation. These targets are extracted automatically from a plethora of ideas, hidden in unstructured text. The paper discusses empirical findings from three exemplary innovation projects in the field of consumer goods.

1 A Typical Dilemma in Innovation

Open Innovation, Crowd Sourcing and Co-Creation are much discussed topics currently. [Ch03]. From marketeers in the packaged goods industry to agencies are consulting with their clients on what idea should be launched next in the market. One major step is to include the consumer in the innovation process [Gr09]. It becomes a game changing new rule - that the consumer him/herself should be generating the idea.

Open innovation describes the process of generally opening up the boundaries of developing the next generation of products inside a business. It includes experts from the outside, i.e., lead users [Hi88] or consumers to find a solution to a well-defined problem, such as, “How do we create a chewing gum that cleans teeth as good as brushing the teeth”. Crowd Sourcing gives the task to a number of consumers in cyberspace, setting up an idea contest or something similar. Co-creation would enable consumers to describe the next generation of chewing gum, adding benefits and features they would love to have in the market.

erschienen im Tagungsband der INFORMATIK 2011
Lecture Notes in Informatics, Band P192
ISBN 978-3-88579-286-4

1.1 Empirical Study

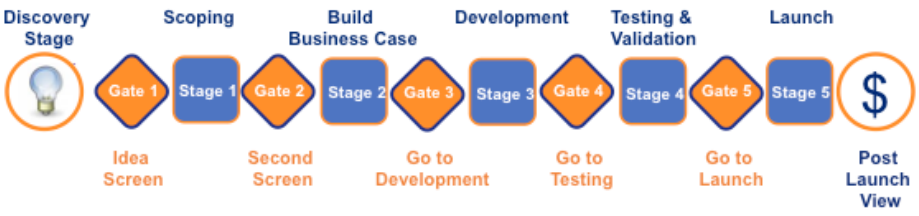
The author, of this Paper, is a practitioner in the field of consumer research especially in innovation processes utilizing different web-based co-creation and creativity methods. This paper intends to demonstrate the proposed methodology in three exemplary cases. The new methodology will help to evaluate the quality and quantity of ideas generated in the early stage of an innovation process. Innovative methodology will be used for evaluation, namely, text mining technologies which have been developed and tested on more than twenty successful projects by the author so far. Companies for which this methodology has been utilized, include Gruner + Jahr, Danone Waters and Melitta.

The cases explored in this paper follow:

- the development of new products for oral care for children and for canines in the UK
- the generation of new concepts for a novel, multi-functional device for runners.

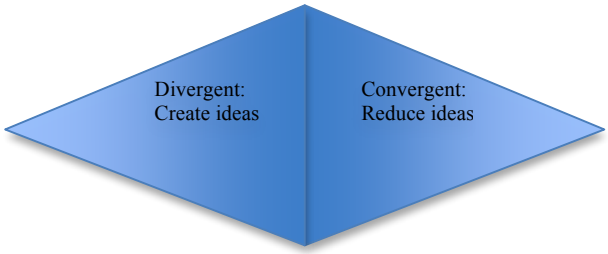
1.2 Typical Innovation Processes in Practice

In practice Cooper´s Stage Gate Process [Co11] is a decision-making process which is widely used for organizing innovation within a company. It is organized in 5 stages. The formalized process starts with discovery and runs from scoping via business cases to development, testing and finally: Launch (Picture 1). Gates are defined as decision points to making the decision whether the product development, with regard to the specific innovation topic, shall be continued or dismissed.



Picture 1 - Stage Gate Process by Cooper

In the early stage of any innovation process, the focus is on Ideation: discovering new ideas and scoping for the right area or consumer need to innovate. This first step is often described as creative problem solving (CPS) whereas the problem is to find a new product within a certain category or market, e.g., oral care for dogs or a new runners' device. Creative problem solving processes use divergent ideation and convergent judgment [Br 00]. While the divergent ideation process starts at the single point and leads to many ideas, the convergent step of judgment aims to reduce the various ideas to the one or few most successful product ideas (Picture 2).



Picture 2 - Divergence and Convergence in Ideation Processes

Opinions differ on how to best succeed in ideation processes. *"The best way to get a good idea is to get a lot of ideas.",* Linus Pauling, American Nobel Prize winner for Chemistry and Peace [Ke01].

Often ideation projects are stuck because they include too few participants in the process, such as a workshop with just about 8 or 10 people. They generate too few and too generic ideas to select from. Some approaches try to meet the need for good ideas by screening the most creative people to take part in such a workshop [Ke06]. Although empirical findings support doubts on the reliability of creativity tests [Ki06], they are still widely used. So far, practicable alternatives are missing.

Most innovation processes make use of different creativity techniques such as more-or-less structured brainstorming [Ke01]. Some creativity processes are already offered online, either to ideate in collaborative sessions or individually. [Fo10]. In all cases, the outcome is a plethora of unstructured text leading to the challenge of organizing all these ideas in the next, convergent step.

The convergent phase is to evaluate the ideas and bring them down to the "right" ideas. Most models suggest the same procedure: either participants of the ideation process or experts evaluate the ideas for selection [Fo10]. This can be done by scoring, e.g., identifying uniqueness and feasibility or judges commenting on the ideas.

ISBN 978-3-88579-286-4

Both methods are limited to a manageable number of ideas, and thereby are in conflict with the "wish for many ideas" sought to maximize the effectiveness of ideation processes. Furthermore any judgment is subjective and therefore tends to dismiss unusual ideas.

2 New Approach Leading to Strong Divergence

In practice, innovation managers would rather request the generation of many ideas, following Linus Pauling´s thought: 'There is a better chance to find a good idea in many ideas.' Until today, divergence has been limited by the mere need to have humans evaluate the results of ideation sessions. As soon as a creativity workshop delivers several pages of ideas, human brains are not able to manage the selection process of the right idea.

2.1 The New Approach in Detail

The author has developed in her company, Dialego, a web-based creativity process supporting innovation processes from the early stage of ideation, called InnovationLounge. InnovationLounge has been inspired by Cooper´s Stage Gate Process and has already been used in several innovation projects for different fast moving consumer goods companies. It engages the consumer in creativity sessions *online* to create ideas and passes the ideas through several milestones of filtering and selecting and finally to Launch (Picture 3). This paper focuses on the first two steps of the described process: the divergent phase of idea generation and the convergent idea filter stage. Similar procedures have been discussed for the early stage of idea generation and idea filter by Forster outlining computer based collaborative creativity processes [Fo10].

Picture 3 - InnovationLounge Process Involving the Consumer

For InnovationLounge projects, participants are recruited from an online access panel. The sample of at least 100 consumers is well-spread regarding gender and age. For idea generation, several creativity techniques are implemented in a web-based questionnaire and forum. The process focuses on intuitive-creative techniques rather than systematic-analytic methods [Fo10]. First 4 iterative steps of different creativity techniques are run through one-by-one before participants receive the ideas of other participants as stimulus for their inspiration.

The creativity techniques used are taken from De Bono´s methodology of lateral thinking [DE73, DE07]. These are

1. Provocative Operation: Challenge (take away a usual attribute from a product and try to find an alternative)

2. Random Visual (a photograph of an object is shown, solutions for the new product will include the object. In this proposed methodology, the photograph is randomly selected for each respondent out of a database of 60 different photographs, e.g., spider, firemen, flower)

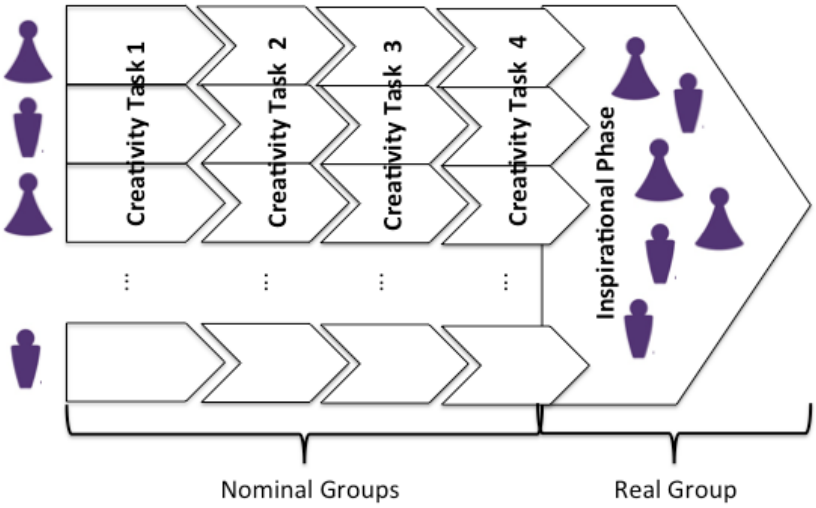
3. Random Words Method (five random words are shown to each individual. About 120 words are in a database and each participant sees different 5 words randomly selected upon questionnaire request)

4. Change Perspective (What would a different person, e.g. Santa Claus or a Business Manager, think about the new product?)

These four tasks are completed by the participants as nominal groups, which means that everyone sees only his or her ideas. In the next Step ideas from all other participants are shown on a „virtual board“ within the questionnaire and respondents are asked to reflect the ideas and use them for inspiration to create more new product ideas. Research by Taylor et. al early in 1958 revealed that real groups (members actually work together as a group) do not necessarily create more or better ideas [TBB 58]. Nominal groups, in fact, created more ideas which has been evidenced in different experiments that followed Taylor et. al.

The innovation process described here uses both nominal and real groups. First many ideas are generated in four subsequent nominal creativity groups. These are followed by two creativity rounds in a real group comparable to the brainstorming technique with the aim to stimulate more in-depth details and additional ideas, when reflecting the ideas generated individually (Picture 4).

ISBN 978-3-88579-286-4



Picture 4 - Idea Generation Process

2.1 Findings from Case Studies

Experience from real cases proves that the process delivers a great quantity of ideas and that, in fact, the two different techniques of nominal and real groups deliver different quantity and quality. Two exemplary projects shall be used to present the results (please refer to Table 1). Project Kids included 510 participants (parents of a child), Project Dogs had 423 dog owners. While nominal groups created 2,267 and 1,745 ideas respectively, the real groups only added 867 and 797 ideas.

Applying SemanticMining techniques to evaluate the content, it reveals that the stimulation and communication of ideas from others does improve the richness and details of ideas. The content indicator shown in table 1 statistically describes the entropy within text. While nominal groups show an average content indicator of 1.2 and a maximum of 4.1 and 4.3, respectively. The real groups deliver more diversity when explaining their ideas. They range at an average of 2.0 and 2.2 and reach higher maximal values of 5.4 and 5.7 respectively. The content does not necessarily deliver a score for the quality or even success of the ideas in the market, but reveals how detailed and different ideas are described when compared with each other.

	Number of Participants (n=)	Number of Ideas Nominal Groups	Average Number of Ideas per Person	Content Indicator	Number of Ideas Real Groups	Average Number of Ideas per Person	Content Indicator
Project Kids	510	2267	4,4	Average 1.2, Max 4.1	867	1,7	Average 2.0, Max. 5.4
Project Dogs	423	1745	4,1	Average 1.2, Max 4.3	797	1,9	Average 2.2, Max 5.7

Table 1 - Quality and Quantity of Ideas from InnovationLounge Process

The two cases clearly show that an abundance of ideas is being generated. The methodology works well regarding divergence but places challenges for the subsequent phase of convergence. Judging almost 2,000 ideas from nominal group stage or almost 900 ideas from real group stage over- taxes human capabilities.

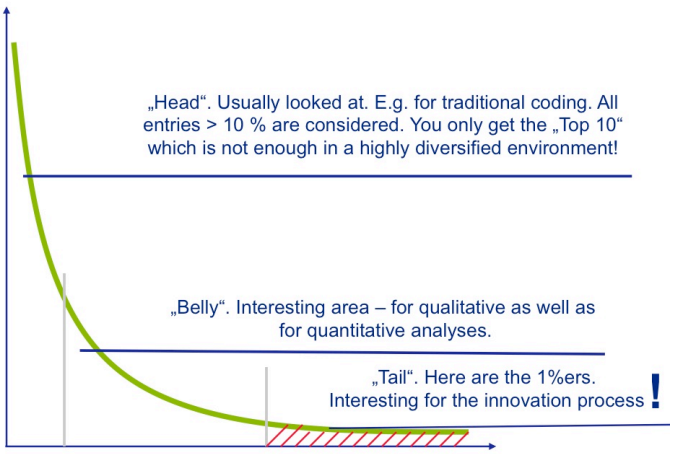
A solution for the next convergent step is, again, the use of text mining techniques.

3 Convergence: Reducing Complexity Using Text Mining Techniques

3.1 The Long Tail of Ideas

The main reason to use text mining technologies is to objectively reduce complexity and to give access to all ideas, from the more mainstream ones to the brilliant, sometimes crazy 1% of ideas that deserve to be considered and might be a starting point for revolutionary new products. These ideas often are kicked out in an early stage due to doubts regarding their feasibility.

We call them "1 %er" following the idea of the Long Tail [An06]. In talking with responsible managers about their satisfaction with innovation processes they often complain that most of the output (sometimes even all of the output) was known before. They probably look at the "Head" of the Long Tail where all the common thoughts about a category are summed up. Asking a consumer what he wants looking at new car, he´ll come up with 4 wheels and a roof. Being a little more creative, it might become "bucket seats" or a roof that can slide open. The "crazy" ideas (and often starting point for revolutionary innovations) imagine engines running on water for fuel.



Picture 5 - The Long Tail of Innovation

If brainstorming sessions deliver too many great ideas that no human can easily digest and structure them, the question becomes, " How can a machine do this work for us?"

Engaging in text mining techniques, over a period of time, enabled us to structure open-ended answers from consumer surveys. We started to experiment with the technology with the aim to solve the issue of breaking down all the ideas to manageable areas of interest. After a several successful text mining projects, we were convinced it would continue to work, and continue to work, very well.

3.2 Reducing Complexity

Bringing the plethora of ideas down to an assessable size is a first important target. Also it is very important for the target group of practitioners and especially marketers, to have an engaging visualization to see results. It is important to communicate the output as effectively as possible. The following examples of the established process try to meet both needs, utilizing text analytics including empirically developed and proven algorithms to segment the text information. This process delivers the critical information in a very visual and easy to understand way.

As in most text analytics processes, pre-processing is being applied before any analysis starts. This includes the removal of symbols, figures and stop words and the stemming of all concepts received.

First of all, a simple visualization of the input is being delivered in the form of word clouds. With these simple means, clients get the opportunity to dive into the sphere of concepts that have been generated in the innovation project. The Word Cloud (Picture 2) enables a fast overview and is often the starting point for input from further brainstorming within the company.

The example shown in Picture 2, is taken from an innovation project for the next generation of running devices; we named it "Runner's Aid". It´s easily recognizable that consumers wish for a MP3 player, pedometer and sat nav. Some of which they are already using. The larger texts are more often mentioned and demonstrate the head of the Long Tail. Once you dive into the smaller concepts, the 1%er ideas show up, such as climate information or joining partners when they are approaching. It´s more or less a rough overview, showing even just the major concepts (i.e. top 150 words out of 1,000 words collected). For brainstorming sessions, we would use a large poster print out of the full Cloud showing every single word used to reflect the importance and input from the 1% of very special ideas (not even seen in the example below).

ISBN 978-3-88579-286-4

Picture 6 - Word Cloud, a Simple but Effective Communication Mean

3.3 Telling the Story

In the next step, the measure of betweenness, widely known from network analysis, can be used to tell the story of all ideas (Picture 3). This measure will deliver the most central topics in all entries. These are the topics that should not be left out from the consumers' viewpoint today. But be aware: in an innovation process, some of these topics might be common sense already. But still it is worth looking at what is at the core of a product or market in the eyes of the consumer.

In the example in Picture 3, obviously voice (control) and mobile phone have been mentioned both 4 times, whereas voice is the more relevant topic. It shows a much higher score in what we call the storyteller score.

Top 15

	STORY TELLER	FREQUENCY
control	1.00000000	5
phone	0.88382351	7
radio	0.84705883	5
voice	0.62058824	4
monitor	0.41470587	2
internet	0.38235295	5
record	0.25588235	2
coach	0.17941177	2
rate	0.17941177	2
directions	0.17941177	3
connection	0.15882353	3
mobile	0.15882353	4
pressure	0.09117647	1
fi	0.09117647	1
light	0.09117647	2

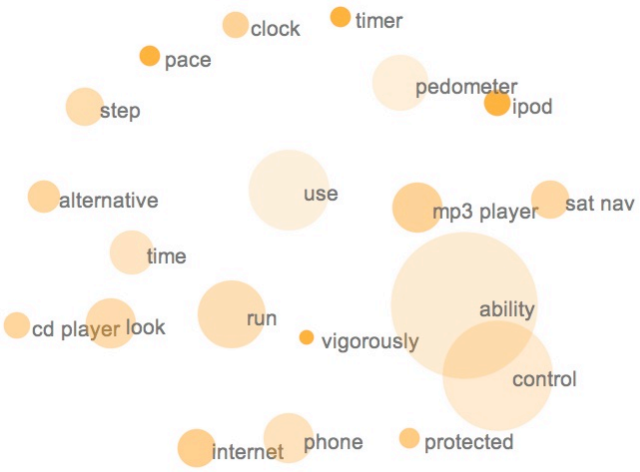
.

Picture 7 - Most Relevant Topics for the Runner´s Aid

3.4 Mapping Idea Clusters

Finally and most relevant, the means of the term frequency – inverse document frequency (tf-idf) weight has been used as a basis to generate clusters. The weight describes how important a concept is within the collection of all ideas. The cluster solution shows which topics belong together and how distant or close different topics are to each other. In the example shown in Picture 4, twenty clusters have been generated to separate twenty different areas of revolutionary or evolutionary product development.

ISBN 978-3-88579-286-4



Picture 8 - Innovation Landscape: Most Relevant Topics for the Runner´s Aid

The Internet and phone are close to each other on the issue of data privacy in terms of protection and control of what is being measured and transmitted, i.e., online. Satisfying our runners' desire for this new Aid and successfully offering the new product appears to be the type of major issue today that is worth exploring. Company survival depends upon discovering new ways to target the consumer, and opening up new markets, which might be found in this "landscape of ideas".

All the above clusters are composed of the initial text input and clustered by their difference or similarity. As with numbers, each cluster demonstrates homogenous mentions within each cluster and heterogeneity towards the other clusters. Managers running the innovation process can click on each cluster to read through the original ideas of the participants and thereby get an insightful understanding of their needs.

This approach is made to enable customers to visit the consumers' voice but in a way that the input is automatically pre-structured. Leaving the tedious task of sorting out the topics that belong together to the machine, is a valuable time-saver. And, this way, managers can make use of *all* ideas, and not leave out any very good idea that they might have discarded subjectively earlier in the process.

4 Outlook

This paper has drafted the process of reducing complexity within innovation and brainstorming processes by integrating text analytics techniques. These techniques enable the user to work with the unabridged quantity and quality of rich text datasets, delivering the basis for groundbreaking new ideas. Ideas are easily extracted and brought to the next gate and phase within innovation processes, namely concept development and testing.

The text analytics techniques is a unique business solution which provides an insightful journey, into new ideas that yield new marketable products. These techniques have outperformed many traditional methodologies. In fact, many more marketable ideas are being generated through these techniques. Also. the 1%ers are considered by analysis to have more of the important revolutionary, new ideas. Continued refinement of this process and continued practical applications, offer the user a "leg up" in ground-breaking information to develop innovative "products-by-demand" and establish new markets, in the coming months and years.

Bibliography

[An06] Anderson, C.: The Long Tail. Hyperion, New York, 2006.
[Br00] Brophy, D. R.: Comparing the Attributes, Activities, and Performance of Divergent, Convergent, and Combination Thinkers. In: Creativity Research Journal, Vol. 13, Nos. 3 & 4, 439–455, 2000–2001.
[Ch03] Chesbrough, H. W.: Open Innovation: The New Imperative for Creating and Profiting from Technology. Harvard Business School Press, Boston, 2003.
[Co11] Cooper, R. G.: Winning at New Products: Accelerating the Process from Idea to Launch. Perseus Books, New York, 2001.
[De73] De Bono, E.: Lateral Thinking: Creativity Step by Step. Harper & Row, New York, 1973.
[De07] De Bono, E.: How to Have Creative Ideas. Vermillion, London, 2007.
[Fo10] Forster, F.: Computerunterstützung von kollaborativen Kreativitätsprozessen. Südwestdeutscher Verlag für Hochschulschriften, Saarbrücken, 2010.
[Gr09] Grant Thornton: Innovation: the key to future success? Grant Thornton Global Innovation Report. Chicago, 2009.
[Hi88] Hippel, E.v.: The Sources of Innovation, Oxford University Press, Oxford, 1988.
[Ke06] Kearon, J.: Creative Consumers: Adding Inspiration to Innovation. ESOMAR Publications. Research Paper Innovate: 1-17, 2006.
[Ke01] Kelley, T.: Art of Innovation: Success Through Innovation the IDEO Way. Profile Books, London, 2001.
[Ki06] Kim, H. K.: Can We Trust Creativity Tests? A Review of the Torrance Tests of Creative Thinking (TTCT). In: Creativity Research Journal, Vol. 18, No. 1, 3–14, 2006.

ISBN 978-3-88579-286-4

[TBB58] Taylor, D. W., Berry, P. C., Block, C. H.: Does Group Participation When Using Brainstorming Facilitate or Inhibit Creative Thinking? In: Administrative Research Quarterly, Vol. 3, No. 1, Jun., 1958.

[Va92] VanGundy, A. B.: Idea Power: Techniques and Resources to Unleash the Creativity in Your Organization. AMACOM Books, New York, 1992.

erschienen im Tagungsband der INFORMATIK 2011
Lecture Notes in Informatics, Band P192
ISBN 978-3-88579-286-4