

# Classifying and comparing community innovation in Idea Management Systems

Adam Westerski<sup>a,\*</sup>, Theodore Dalamagas<sup>b</sup>, Carlos A. Iglesias<sup>a</sup>

<sup>a</sup>*Universidad Politécnica de Madrid, Escuela Técnica Superior de Ingenieros de Telecomunicación, Avenida Complutense 30, Ciudad Universitaria, 28040 Madrid, Spain*

<sup>b</sup>*IMIS Institute, "Athena" Research Center, Artemidos 6 & Epidavrou, 15125, Marousi, Greece*

---

## Abstract

The Idea Management Systems are a tool for collecting ideas for innovation from large communities. One of the problems of those systems is the difficulty to accurately depict the distinctive features of ideas in a rapid manner and use them for judgement of proposed innovations. Our research aims to solve this problem by introducing annotation of ideas with a domain independent taxonomy that describes various characteristics of ideas. The findings of our study show that such annotations can be successfully transformed into new metrics that allow the comparison of ideas with similar successfulness as the metrics already used in Idea Management Systems but greater detail. The presented results are based on experiments with over 50000 ideas gathered from case studies of four different organisations: Dell, Starbucks, Cisco and Canonical.

*Keywords:* idea management system, metrics, annotation, taxonomy, classification, innovation management, product development, process improvement

---

---

\*Corresponding author. Tel.: +34 91 549 57 00, Extension: 3061; Fax: +34 91 336 7333; email: westerski@dit.upm.es

## 1. Introduction

In the era of globalization the markets become more competitive and the organisations seek new ways of innovating. Among those attempts, are Idea Management Systems that employ Information Technology and crowd-sourcing principals to support innovation processes in the organizations. In particular, the notion behind those systems originates from simple suggestion boxes but is transformed into a more sophisticated process (Turrell, 2002). During the last decade of their evolution Idea Management Systems have extended their coverage from collecting ideas from large communities via computer networks to collaborative improvement of those ideas, the assessment of ideas and idea management in synergy with other enterprise processes (Westerski et al., 2011).

Currently, Idea Management Systems are considered a very promising branch of computer software market (Fenn and LeHong, 2011) and various analyses of the vendor landscape (Rozwell et al., 2010; Brown et al., 2009) show rapid adoption growth in many enterprises in recent years. Nevertheless, current state of the art Idea Management Systems still face key problems related to the large amount of human effort needed during the idea management process. Based on the testimonials of Idea Management Systems vendors (Baumgartne, 2008) and case studies of various companies (Jouret, 2009; Belecheanu, 2009), the main origins of those problems are: large volume of submitted ideas, sudden peaks of submissions, redundancy of ideas, and large quantities of trivial ideas.

In our research we relate the above issues to the idea assessment phase and focus on challenges that arise when trying to quantify the value of information contained in ideas and its impact on innovation in the enterprise. According to the study of contemporary solutions by Harstinski et al. (Hrastinski et al., 2010), the problems of idea assessment are approached by: 1) the use of a handful of automatically generated yet very simple community statistics; and 2) expert reviews that require a considerable amount of knowledge and impose serious time constraints thus increase the costs of the entire idea management process.

In this article we present a solution for idea assessment that combines the advantages of those two cases mentioned by Harstinski: rapid generation of metrics that require little expert knowledge yet offer more diversity and versatility than the current community metrics. In particular, we deliver a methodology for obtaining the metrics via analysis of idea annotations made with a domain independent taxonomy that expresses idea characteristics. The focus of the following article is to show that the proposed set of metrics can be applied to Idea Management Systems in a meaningful way that would allow to capture the distinctive features of ideas and compare entire idea datasets.

The article is structured as follows: firstly we summarize the past research achievements in terms of metric generation for Idea Management Systems as well as other kinds of computer-supported cooperative work systems (see Sec. 2). Additionally, in the same section, we discuss research on capturing the meaning of innovation in general and show how it influenced our work. Afterwards, we introduce our contribution in a form of a taxonomy for describing idea characteristics and present in more detail the theoretical grounding by referring to particular innovation models (see Sec. 3). Finally, we show how to utilize the proposed taxonomy in practice of Idea Management Systems by transforming the idea annotations into metrics that characterise entire systems (see Sec. 4). At the end, we present the results of our experiments that test the usage of the taxonomy for annotation (see Sec. 5.1) as well as verify the performance of metrics generated from those annotations in relation to the contemporary parameters of Idea Management Systems (see Sec. 5.2).

## **2. Related Work**

Having a significant presence in the industry, Idea Management Systems have also been investigated by the academia in search of problems and patterns that emerge when using this class of systems in an organization (e.g. (Bailey and Horvitz, 2010)). In our case, the investigative work on idea assessment is of

special interest. Within this area, Hrastinski et al. (Hrastinski et al., 2010) surveyed a number of selected products and pointed out that the current commercial systems employ rather simple idea evaluation methods most often being analysis of community statistics (number of ideas per user, community voting results, number of idea comments etc.) or internal business metrics that are delivered by designated experts (e.g. return of investment, market value etc.). Both of those approaches have been evaluated by Gangi et al. (Gangi and Wasko, 2009) and compared to conclude that in practice none of current methods have a significant impact on which ideas are being implemented by the organizations. Following those conclusions, there have been various approaches that attempted to find a solution to time efficient and effective automatic idea assessment problem e.g. with prediction markets (Bothos et al., 2008), by applying problem solving algorithms (Adamides and Karacapilidis, 2006), calculating metrics for the quality of management (Conn et al., 2009) or using data from other enterprise systems to automatically assess ideas (Ning et al., 2006; Westerski and Iglesias, 2011). All those solutions are based on the notion of reusing existing data whereas the approach proposed by us claims that there is a necessity to attach some additional data to ideas in order to improve assessment and selection phases of the idea life cycle.

Apart of Idea Management Systems domain research, there has been a huge number of works that attempt to analyse characteristics of discussions or content created by communities in a collaborative way e.g. (Stromer-Galley, 2007; Alexandru Spataru and Bendixen, 2004; Nisbet, 2004). Among those, Perey (Perey, 2008) describes a necessity to go beyond simple metrics that count number of interactions with the system in time. However, in contrast to us, in his work Perey focused only on measuring characteristic features of users and their interactions with each other rather than metrics on content that those users create. Klein (Klein, 2012) notices similar problems with regard to difficulties in assessment and browsing community submissions but he attempts to find a remedy through experimenting with novel system interaction methods, in particular

argumentation tools (Kirschner et al., 2002). While this approach is different to ours it shows an interesting alternative not only for generation of new metrics but altering the entire philosophy of Idea Management System front-end which in turn can create new opportunities in the back-end.

Outside of the aforementioned areas of computer science, there has been a large number of works that investigate ways of categorising innovation and attempt to quantify it. While preparing for the creation of the taxonomy and validating it afterwards we analysed those models as a reference. We started from the very origins of Schumpeter’s innovation theories (Schumpeter, 1934) and finished with the contemporary work on the topic. The selection of models that we have analysed as related was based on studies from a number of works that attempt to revise the state of the art on innovation models (Eris and Saatcioglu, 2006; Popadiuka and Choo, 2006; Damanpour and Gopalakrishnan, 1998; Chuang et al., 2010; Garcia and Calantone, 2002). During our work, we prepared a taxonomy model that included the various perceptions of innovation from those models. The preliminary experiments with this taxonomy version have shown that most reviewers did not know how to apply the terms. Consequently we made a choice to propose the taxonomy, as described in the next section, only based on the analysis of idea content from Idea Management instances that we gathered.

### **3. A Domain Independent Taxonomy for Idea Annotation**

In this Section, we introduce a taxonomy that captures the characteristics of ideas published in an Idea Management System. In our methodology, the taxonomy is used to annotate ideas with terms that later serve as a base for calculating metrics. The choice of terms that establish the taxonomy is based on our experience with different kinds of Idea Management datasets gathered during the course of Gi2MO project (Westerski et al., 2010). This initiative aimed to enrich contemporary Idea Management Systems with an extensive use of metadata

according to the Semantic Web principles. During the project we gathered various datasets ranging from ideas for technology to products for mass consumer (see Sec. 5.2 for detailed dataset description). Based on the analysis of those datasets, we enumerated the key characteristics of ideas that could be inferred from the idea text and organized them into a hierarchy. The taxonomy model that we propose can be summarized by the following hypothesis:

”Every idea that was **proposed** has been **triggered** by a particular experience and describes a certain **innovation** put in context of application in a given **object**.”

”Proposed”, ”triggered”, ”innovation” and ”object” represent the four main characteristics of an idea that we established as the root for further taxonomy terms which detail a particular aspect of the idea characteristics (see Fig. 1).

The **trigger** branch details aspects related to experiences that influenced creation of the idea. While analysing the ideas gathered in different Idea Management Systems, we noticed that users often tend to mention how they came up with a particular innovation in order to justify their claims. Similarly, innovation models of Kelly and Kranzberg (Kelly et al., 1975), Usher (Usher, 1954), Myers (Myers and Marquis, 1969), Hughes (Hughes, 1975) as well as contemporary research (Narasimhalu, 2005) notice the existence of various causes that lead to idea generation. In particular, innovation is described as being a result of recognition of a problem, need for changes or recognition of technical feasibility or demand. Those different types of triggering experiences are referred by us in the trigger branch as **Observation Types**. Additionally, Usher (Usher, 1954) has shown that innovation is not only triggered by experiences related to a personal observation but also events that influence the innovator and lead to an act of insight. We relate to this by characterising the type of event that led to the idea with **Creativity Origin** classification and by identifying the connection between the triggering experience and the object that is innovated (**Associated Object**).



Figure 1: Gi2MO Types taxonomy as proposed in Section 3.

The **innovation** branch relates the idea proposal to the reality of the enterprise and the state of the Idea Management facility. As such, the assessment made by annotators that use this taxonomy branch goes to the origins of the very understanding of innovation in enterprise discussed for tens of years since the original contributions by Schumpeter (Schumpeter, 1934). The verity of models proposed since then show that interpretation of innovation can be extended in many different directions depending on the context and goals. In our work, we

took into account the previous models (see Sec. 2), however we narrowed down the amount of terms based on experiences with idea datasets and inferences that could be made based on idea text. As a result, we noticed three key aspects that were mentioned by innovators and reviewers in Idea Management Systems: relations to other ideas previously posted in the system or innovations introduced by the organisation (**Dependence**), descriptions of usefulness of the idea for a particular group (**Target Audience**), references to idea originality with respect to current state of organisation or entire market (**Originality**).

The **object** branch focuses on describing the entity that is being innovated and the changes proposed in relation to the original. Apart of the classical distinction between products and processes (Abernathy and Utterback, 1978; Utterback and Abernathy, 1975; Adner and Levinthal, 2001), we also recognize service innovation as it has been advocated by a number of researchers that studied innovation past the time when manufacturing was the dominant element of economies (Susman et al., 2006; Drejer, 2004; Rubalcaba et al., 2010). With regard to classifying how those entities are transformed by ideas, Gilfillan (Gilfillan, 1935) noticed that innovation is often a chain of small improvements, modifications and additions rather than a single act of brilliance of one innovator. We relate to this observation by classifying the type of changes proposed for an object (**Structure**), as well as recognizing if the introduced change is a reoccurring innovation from some past iteration or a completely new proposal (**History Relationship**). Additionally, following the observations of Abernathy and Clark on firm competence (Abernathy and Clark, 1985), we noticed that proposed changes in the analysed ideas may have a different impact on the current design of the product as well as associated product knowledge. Some ideas propose adding or removing elements in an existing design while others introduce a totally new product. Those kind of differences are classified in the **Offering Placement** sub-tree. Finally, following research in the engineering design (Jarratt et al., 2010) we notice that the proposed changes in the object and their implementation may affect existing related products in a different way.



A report by AberdeenGroup (Brown, 2006) shows that analysis of those kind of changes is of crucial importance for organizations when making decisions on adopting certain innovations or not.

The final **proposal type** branch is connected to the way the text of an idea has been written. The analysis of idea datasets has shown that not all users express their requests for innovation in the same way. Some of the ideas differ on the level of completeness of the description, while others vary in the way the entire idea has been formulated. We perceive those differences as lack or presence in description of selected innovation process stages such as problem definition or solution (as defined in many innovation management models, e.g. (Kelly et al., 1975; Baker and Freeland, 1972; Usher, 1954)). The goal of the Proposal Type taxonomy branch is to capture those differences and later allow the idea reviewers or moderators of idea contests to filter out certain proposal types that are not wanted at all.

#### 4. Calculation of Metrics Based on Idea Annotations

The taxonomy presented in the previous section enables to identify the characteristics for individual ideas. Nevertheless, in big datasets the amount of idea annotations made using the taxonomy terms can be overwhelming and therefore difficult to analyse and interpret for a practitioner. Since our goal is to facilitate idea dataset comparison, we propose to summarize the annotations and describe their meaning for the entire dataset.

In particular, in the next step of our methodology, we propose to utilize the taxonomy described in the previous section to annotate ideas and afterwards produce metrics based on the quantitative analysis of the annotations.

The said methodology for generating metrics includes:

- assuming a certain interpretation of terms in the taxonomy and assigning

a metric to each taxonomy sub-tree

- calculating the metric value for every idea individually based on idea annotations with the taxonomy terms
- calculating the metric value for entire dataset as a median of metric value of all ideas from the dataset
- supplementing every calculated dataset metric with the diversity measure of annotations per each taxonomy sub-tree (using information entropy)

In the first step, we define 14 metrics (see Table 1) corresponding to different branches of taxonomy tree and an additional single metric (Idea Completeness) that measures how many branches of the taxonomy are used for describing an idea. The metrics that relate to particular sub-trees have an ordinal scale based on the particular interpretation of term order in the respective taxonomy sub-tree. The explanation of the approach taken for each sub-tree can be observed in table 1, while the results of applying the median for calculations of metrics for entire datasets can be seen in the next section when we report on evaluation results (see Sec. 5.2).

The aforementioned metrics summarise the information expressed by the annotations and transform it by providing a certain interpretation. However, the problem that arises is that some of the information is lost in comparison to term frequency analysis. In particular, one cannot say what is the diversity of terms just looking at the metric (e.g. Idea Originality is 0.5 if half of the ideas are *New* and half have *no innovation* but also when all ideas are tagged as *Incremental*). Therefore, apart of the metrics based on interpretation of the taxonomy terms we also propose to measure the diversity of terms in the annotated datasets. Whereas the first set of metrics flattens the perception of terms to a common level, in the second case we sought for a solution that will enable the reviewer of the idea contests to assess the diversity of terms and judge how much ideas are similar to each other under certain criteria. As a result we have chosen

Table 1: List of metrics based on Gi2MO Types taxonomy.

<b>Taxonomy Branch</b>	<b>Metric</b>	<b>Explanation</b>	<b>Scale</b>
Trigger/ Observation Type	Trigger Experience Completeness	How complete was the experience with the triggering object	1 = Potential Opportunity 0.5 = Faulty Experience 0 = Lack Of Feature
Trigger/ Creativity Origin	Trigger Situational Dependence	How much is the trigger dependence on occurrence of some particular event	1 = Event 0 = Object Interaction
Trigger/ Associated Object	Trigger Relatedness	How closely related are the object of innovation and the object that triggered the idea	1 = Object Of Innovation 0 = Other Object
Innovation/ Dependence	Idea Dependability	How much is the idea connected to other ideas (how much does it change influence other)	1 = (Proceeds, Follows, Duplicates, Is Part Of) 0.5 = (Encapsulates, Excludes) 0 = None
Innovation/ Target Audience	Idea Adaptiveness	How much are the ideas meant for new markets and how much for existing ones	1 = Existing Audience 0 = New Audience
Innovation/ Originality	Idea Originality	How original is the idea	1 = New 0.5 = Incremental 0 = None
Innovation/ Related to	Idea Originality Scope	How far does the originality of the idea reach. Is it only the particular element of the organisation, entire organisation or the entire market ?	1 = Market 0.5 = Organization 0 = Innovation Proposals
Proposal Type	Community Cooperativeness	How well do users formulate and communicate their ideas and the implementation of ideas.	1 = Solution 0.75 = Suggestion 0.25 = Request 0 = Issue Report
Object/ History Relationship	Implementation Freshness	How much are the new ideas related to former products	1 = Evolutionary 0 = Regressive
Object/ Type	Implementation Integrity	How tangible is the object of innovation	1 = Product 0.5 = Service 0 = Process
Object/ Type/ Product	Implementation Applicability Scope	How broad is the application of the idea (measured by product offering scope)	1 = Product Type 0.5 = Product Line 0 = Specific Product
Object/ Offering Placement	Implementation Constructiveness	How much ideas are creating new products or just improving the old products	1 = New 0 = Existing
Object/ Structure	Implementation Scope	How much the ideas want to modify of the current state	1 = Complete 0.5 = Element 0 = Characteristic
Object/ Relationships	Implementation Dependability	How much does the introduction of innovation impact other products	1 = (Part Of, Composed Of) 0 = Complementary With

information entropy as a statistic that would fulfil this need (see Eq. 1).

$$E(tb_x) = - \sum_{i=1}^n p(i) \log_2 p(i) \quad (1)$$

When applied to our case,  $E(tb_x)$  is the diversity for the  $tb_x$  taxonomy branch;  $p(i)$  is the frequency of annotations with the certain ( $i$ ) term combination; while  $n$  is the number of all such combinations in a given taxonomy branch (we assume that all combinations are possible, e.g. an idea can describe product and service innovations). Based on the above, we propose to calculate term diversity understood in such way for every taxonomy branch and for each dataset. As an outcome, our hypothesis is that the entropy should allow to decide how similar to each other are ideas of different datasets. The results of experiments that evaluate this hypothesis in practice and calculate entropy for particular datasets are presented in next section.

## 5. Evaluation and Discussion of Results

In order to test our hypothesis about the taxonomy and the formulated metrics, we performed a series of experiments to cover the entire presented methodology. Firstly, we studied how does the taxonomy perform when annotations are applied by groups of people of different sizes and different expertise levels, as well as how does manual annotation compare to the automated approach (see Sec. 5.1). Further, having obtained satisfactory results with the annotation experiments, we evaluated the second step of the methodology that delivers the actual metrics. The performed experiments aimed to evaluate the feasibility to use the metrics for comparison of datasets (see Sec. 5.2) as well as usage of metrics for selection process of best ideas by measuring their correlation with some of the currently used statistics (see Sec. 5.2.1).

### *5.1. Annotation of Data in Idea Management Systems*

Our ultimate desire was to construct a taxonomy that could be complex enough to cover all the idea characteristics but at the same time suitable for usage by non-experts or with automatic annotation algorithms. During the experiments we realised that this might be a difficult task to achieve due to some characteristics being very detached from the sole idea text. Therefore, we downsized the taxonomy in different ways to find the set of its elements that would fit the desired goals best. We present the results of a number of experiments that compare performance of annotators when using the full taxonomy as well as certain parts of taxonomy for: manual annotation of ideas (see Sec. 5.1.1) and automatic annotation (see Sec. 5.1.2).

#### *5.1.1. Manual Annotation*

In case of manual annotation, we measured the differences in annotations proposed by different people as well as the differences in annotations of the same annotator repeated in certain time intervals. For the first experiment, we arranged for 10 people to individually annotate the same set of 10 ideas. All participants of the experiment were computer scientists, aged 25-30 and working in the academia; none of them have had any previous contact with innovation theory or our taxonomy in specific. We did not limit the annotators in any way with regard to annotation rules (e.g. annotators could apply many different terms of the same branch to a single idea). Following the experiment, we measured the agreement of annotators as a percent of cases in which they either agreed to put the same annotation or agreed on not putting a certain annotation at all. As a result, we discovered the differences in decisions were quite considerable with only 34% of cases where annotators fully agreed and 5% of cases where no agreement could be reached at all (half of the annotators put an annotation and the other half did not).

Pursuing the same line of inquiry, we repeated the experiment inviting 5 in-

novation theory experts to provide the annotations for the same 10 ideas. In comparison with the first experiment, the innovation experts reached a consensus in 2% more cases than non-experts. For full results and comparison please see figure 2.

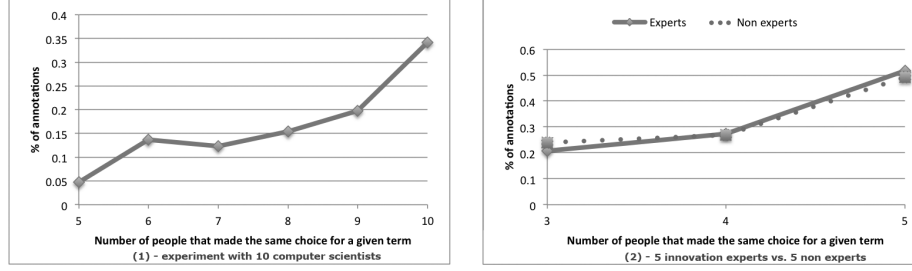


Figure 2: Evaluation results for manual annotation of 10 random ideas taken from IdeaStorm dataset.

In the second experiment referring to manual annotation, we asked the same person to annotate the same set of 100 ideas twice but in time distance of 3 months. The differences in that case were smaller than in the first experiment with 10 different annotators - 70% of annotations turned out the same in second annotation round as in the first. The worst result was noted for the Trigger/CreativityOrigin branch (only 48% of the same annotations and the only branch below 50%) due to the annotator categorising specifically types of triggering events very differently in consequent iterations of the experiment. On the other hand, the best results were achieved for Object type and Originality branches (82 % and 79 % respectably). Additionally, if we include in our calculations the cases of agreement on not marking a certain annotation, the final result for single annotator agreement rises to 90 %.

The presented manual annotation experiments show that the characterisation of innovation can be very subjective and relies in a great manner on the understanding of the topic by the annotator. This is in line with statement made by Garcia et al. (Garcia and Calantone, 2002) who presented a number of examples where the same innovation was labelled as radical or incremental depending on

very small differences in understanding of those terms. In the case of annotation of ideas in Idea Management System the experiments showed that this is quite a valid problem if the annotations are made by a collective of reviewers (regardless if they are innovation experts or not). However, in the case of a single person doing all the work, the annotations are quite coherent, especially if limited to certain taxonomy branches.

### *5.1.2. Automatic Annotation*

Regarding the automatic annotation, we tested whether it is possible to automatically extract features of the ideas that would suggest certain annotations without the need of pointing to keywords or using any additional knowledge base. Therefore, we experimented with a machine learning approach that was based on comparing similarity of idea texts. In particular, we used a supervised machine learning approach and the weighted k-nearest neighbour (kNN) algorithm (Dasarathy, 1990). Our evaluation was done using a tool called GoNTogle that was previously proven to successfully work for automatic annotation of documents (Bikakis et al., 2010). In the implementation of GoNTogle the nearest neighbours are selected based on text similarity calculated by the document similarity algorithm of Lucene library (Lucene, 2012).

During our experiments, we used the annotated data corpus of 400 ideas from the previous manual annotation tests: 200 ideas were used as a training set and 200 for evaluation of the accuracy of the automatic annotation proposals. Taking into account the results of the manual annotation experiment, both of the datasets used during the automatic annotation were prepared by the same single annotator. In our first approach, we merged idea description with idea title into a single block of text and treated it as a document. For the analysis of results we used the typical measures for judgement of information retrieval effectiveness: precision (Eq. 2), recall (Eq. 3) and their harmonic mean, i.e. F-measure (Eq. 4).

$$precision = \frac{|correct\ automatic\ annotations|}{|all\ automatic\ annotations|} \quad (2)$$

$$recall = \frac{|correct\ automatic\ annotations|}{|all\ evaluation\ set\ idea\ annotations|} \quad (3)$$

$$F = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (4)$$

In the described case, we allowed annotation with all taxonomy terms and as a result the average F-measure was 0.46. To investigate further reasons for such performance, we analysed the results for particular elements of the taxonomy to discover which branches of the taxonomy could be fit for use with automatic annotation algorithms (see Fig. 3). In particular, we found that most promising elements of the taxonomy are located in the Trigger sub-tree.

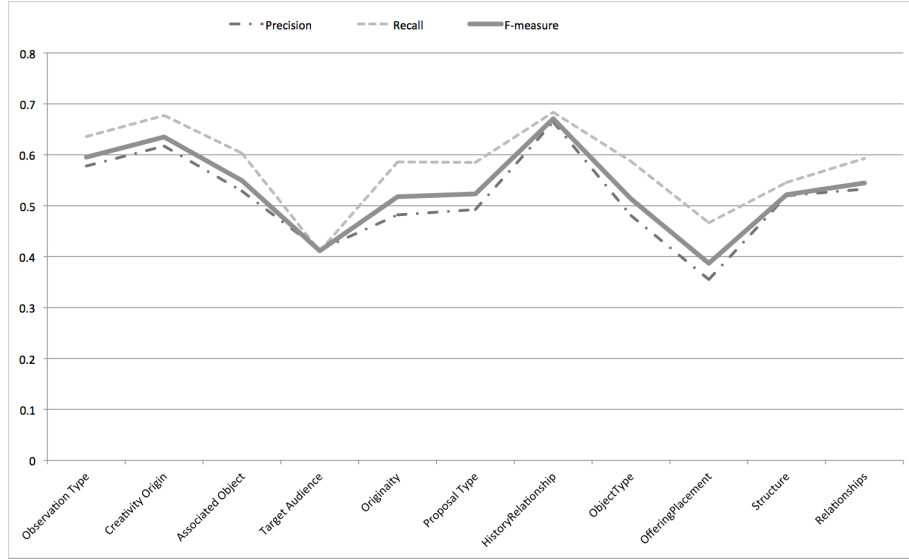


Figure 3: Evaluation results of automatic annotation split per taxonomy sub-tree (IdeaStorm dataset).



In addition, some elements of the Object branch also gave interesting results but we detected that in some cases the high variance of learning set had a big impact on those results (which was not the case for the Trigger branch as shown on Figure 4).

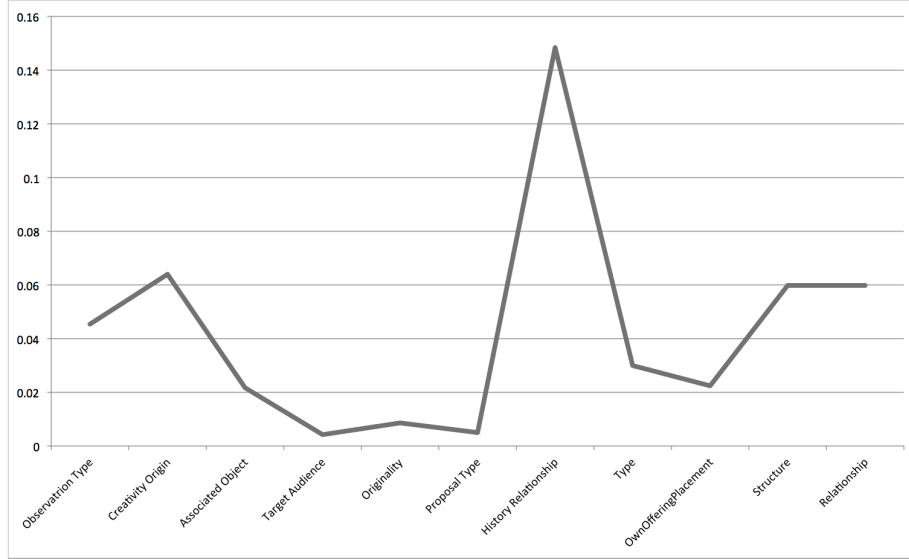


Figure 4: Variance of the dataset manual annotations split per taxonomy sub-tree (IdeaStorm dataset).

In an attempt to search for different options and improve the automatic annotation results, we took a few paths to change our process, most notably: split the ideas into paragraphs and treat them as separate documents during annotation time, increase the learning set size, add additional rules for annotation (e.g. define terms that exclude each other etc.). In the first case, splitting of ideas into paragraphs brought a quite substantial improvement and interesting observations. Taking into account different taxonomy branches, on average the F-measure increased by 24% (with best case of 83% F-measure for Creativity Origin branch). In addition, we noticed that the amount of annotations per idea shrank because individual paragraphs did not hold enough information to assign terms from certain taxonomy branches (see Fig. 5).

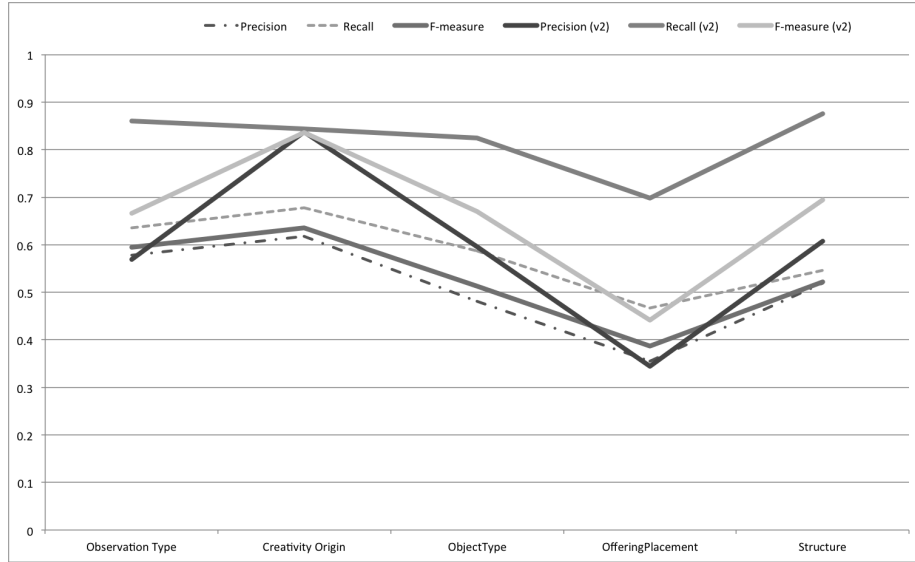


Figure 5: Differences in precision, recall and F-measure between annotation of IdeaStorm dataset when using full idea text and text split into paragraphs (v2).

In case of increasing the learning set size by 50% (up to 300 ideas) we got a 2% F-measure improvement in case of taking the full taxonomy into account and 5% F-measure improvement if analysing only the term branches filtered out earlier during the paragraph experiment. Finally, by adding some additional rules on top of the regular algorithm we did not get any improvement at all.

Concluding the experiment, we noticed that utilising the full taxonomy as originally proposed is very challenging if we desire to obtain all annotations in an automatic manner. Nevertheless, by measuring the performance of particular branches of the taxonomy we got interesting insight into the elements of the taxonomy that are already fit to be used with automatic annotation and which should be left for manual process. In addition, our results have shown that splitting idea text into paragraphs proved to work best for the type of textual submissions provided by innovators in Idea Management Systems.

### 5.2. *Testing metrics with datasets*

After achieving satisfactory results with taxonomy annotation tests, we proceeded with experiments to evaluate the metrics that can be generated after the annotations are delivered. The goal of the following tests was to verify if the metrics diversify enough between different datasets to be able to observe distinctive features of selected Idea Management instances and make assumptions about the types of communities engaged in the innovation process. Furthermore, we compared the proposed metrics with the currently available and measured if they have any correlation with the successfulness of ideas or each other.

We analysed a total of 4 datasets (see Table 2), from each we extracted and manually annotated 200 ideas: 120 random selected ideas, 40 ideas that have been implemented, 10 top rated ideas, 10 lowest rated ideas, 10 top commented ideas, 10 least commented ideas. The ideas were selected based on the analysis of the entire lifetime of the respectable instances since their start until the time our experiment was conducted (February 2011).

Two of the chosen instances are based on the same SaleForce Idea Management System. Both are administered in a similar manner as indefinite idea competitions: Dell IdeaStorm exists since February 2007, while the myStarbucks system is running since March 2008. In both cases, the organizations that own the systems are large multinational corporations with huge user base (e.g. Dell sold 44 million PC units just in 2009 (Dell, 2011), while Starbucks claimed to serve 60 million customers weekly in 2011 (Starbucks, 2011)). Up until the time of our experiment both instances presented similar user interface and workflow for the innovators as well as participants of the community. We have chosen those two instances to see if systems deployed in the same way from the perspective of infrastructure as well as idea management practices would diversify due to the fact that ideas are collected for different kinds of products (see Table 2).

The third instance included in our tests was Canonical’s Ubuntu Brainstorm that was opened in February 2008 and is based on an open-source IdeaTorrent

platform. In comparison to the previous instances, the idea submission rules are different and force innovators to deliver solutions for their ideas. Another major difference is that Canonical user base is smaller in comparison to Dell or Starbucks (20 million users total as estimated by Canonical (Ubuntu, 2011)) but also very collaborative (Lakhani and von Hippel, 2003; Feldstein, 2007) and only focused on a single type of an open-source product. The implementation process of ideas is significantly more transparent due to the fact that Ubuntu is an open-source project and all its production infrastructures are available to public and linked to Brainstorm. We have chosen to analyse this instance to see if the computer technology literate audience of Canonical that is used to giving contributions for free would propose ideas that differ in comparison to mass consumer customer base of Dell and Starbucks.

The final dataset that we analysed came from an instance called i-Prize, operated by multinational corporation called Cisco. The instance started running in February 2010 and was only open for three months. Apart of setting a limited time frame for the collection of ideas, Cisco also offered considerable money incentives for the winners that proposed the best ideas. In contrast other instances do not have any incentives apart of public mentions of the winning ideas. Additionally, the goal of i-Prize contest was to collect ideas for a new major future Cisco business while in all three other instances there were no precise goals other than gathering feedback from clients on current products and services.

Taking into account the described differences between the datasets we applied the previously introduced metrics to see if those four different datasets would indeed differ as expected when measured with our metrics. The process of applying the metrics to a dataset included: calculation of metric value per every idea individually, calculating the median value out of all 200 ideas annotated. We followed this methodology with all metrics and for all data samples from every dataset. When visualised on a chart (see Fig. 6) we were able to observe the differences between the datasets and interpret them. As hypothesised before, the biggest similarities can be observed with Starbucks and IdeaStorm

Table 2: List of datasets used for the experiments.

System name	#Ideas/ #Comments/ #Users	Area	Case Characteristic
Dell IdeaStorm	15.000/ 90.000/ 2.000	Computers, telecommunication devices and related services.	Focused on collecting ideas for existing products over indefinite amount of time with periodically organized focus sessions
myStarbucks Ideas	8.000/ 80.000/ 3.000	Coffee and related products sold in a coffeehouse chain.	Focused on collecting ideas for existing products and changes in services over indefinite amount of time
Cisco i-Prize	1.000/ 4.000/ 1.000	Computer, networking and communications equipment.	Viewable only after registration and available only during a set amount of time. Focused on collecting very abstract ideas for new area of activity. Introduces considerable money incentives for best inventors.
Ubuntu Brainstorm	27.000/ 90.000/ 2.000	Open-source operating system and related software.	Very collaborative, computer literate community gathered around open-source software products. Apart of ideas system enables submission of proposed implementation methods for ideas.

instances which gather ideas in competently different areas but are run by the same operator (11 out of 14 metrics had the same values). The most standing out difference between these two datasets can be observed with regard to Innovation Freshness: IdeaStorm ideas in majority were never implemented before while most of Starbucks ideas are reoccurring requests to bring back old innovations.

Based on the metrics calculations, the Cisco i-Prize is the instance that exceeds

others by a large margin in many areas but also has most contrasts (5 top scores and 4 lowest scores). Most notably this instance is characteristic for remarkably high Implementation Constructiveness and Scope, which could be attributed to i-Prize contest explicitly asking for ideas in new areas covering very board scope. This assumption is also confirmed by high Idea Originality Scope that shows that most proposed ideas are original with regard to a very broad scope of markets.

Lastly, the ideas originating from Ubuntu open-source community stand out most in two areas: Implementation Dependability and Applicability Scope. The first metrics shows that Ubuntu users most often propose changes in key ele-

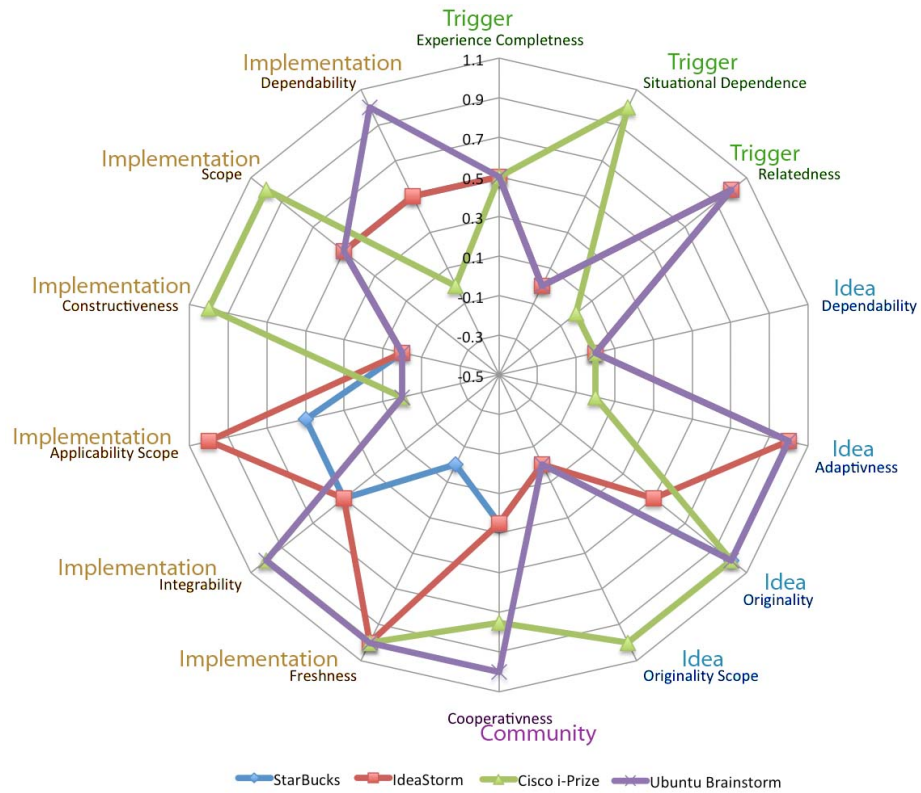


Figure 6: Comparison of metrics based on interpretation of taxonomy terms.

ments of offering that have impact on many software modules. The second metric shows that the proposed ideas are very specific and aimed only for particular products from Canonical offering.

Concluding the above analysis we observed that the metrics enabled to verify judgement about certain instances and deliver proof to how certain communities exceed others. In addition to such interpretation we evaluated the diversity of datasets measured with entropy. When visualised on a radar chart (see Fig. 7) the area taken by dataset determined how similar to each other are ideas of different datasets. In this particular case our experiment has shown that overall ideas posted in IdeaStorm and myStarbucks instances were most diversified while Cisco and Ubuntu Brainstorm least.

#### *5.2.1. Metrics Correlation Analysis*

The analysis presented in the previous section has evaluated the metrics when used for judgement of entire datasets or groups of ideas gathered in idea contests. However, Idea Management Systems suffer not only from lack of tools for assessment of entire instances but assessment of individual ideas in particular. Therefore, in the final experiment, we compared the currently used metrics for idea assessment and idea selection processes (Jouret, 2009) with the metrics proposed by us. Our goal was to check: 1) if there would be a meaningful change in the proposed metrics values or correlations when calculated for particular idea subsets (e.g. top commented ideas or top rated ideas); and 2) if the relationships between our metrics and idea adoption would be similar to the impact of legacy metrics on idea adoption.

To achieve the stated goals, we related our metrics with the following legacy metrics used in contemporary Idea Management Systems: idea rating value, number of comments for idea and idea age (amount of days until idea gets implemented; for not implemented ideas days until the date of newest idea in the test dataset). In particular, we measured the bivariate correlations between

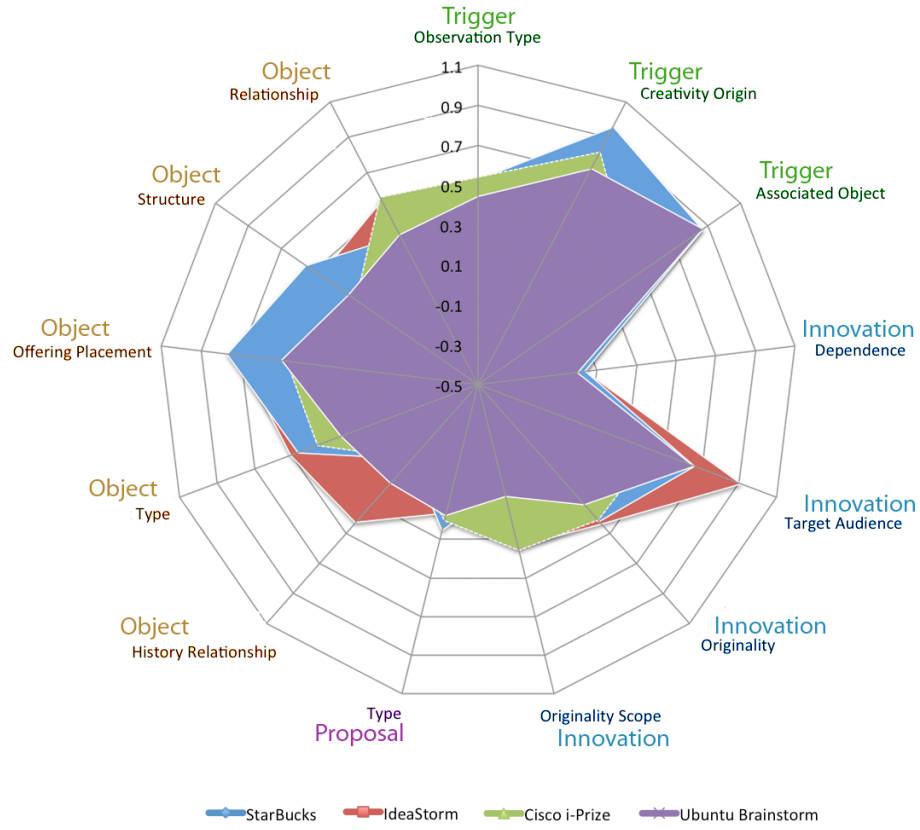


Figure 7: Comparison of diversity of datasets with respect to different taxonomy branches based on entropy measure.

our metrics and the legacy metrics (see example of Dell IdeaStorm in Table 3). The correlation between all variables turned out small (according to Cohen scale (Cohen, 1988)) which suggests that there is little point for analysis of our metrics in border line conditions of community metrics typically used in Idea Management Systems.

To assure that those results were not only the case of a single dataset we measured the aforementioned correlations for all other test datasets and observed the differences between the correlations of the same metrics. While in most cases the correlations remained small as in IdeaStorm, the standard deviations were



Table 3: Bivariate correlations between the proposed metrics and legacy metrics (Dell IdeaS-term dataset)

Metric	# Comments	Rating	Idea Age
Completeness	0.04	0.11	0.01
Experience Completeness	-0.15	0.03	-0.15
Situational Dependence	0.17	0.28	0.13
Relatedness	0.04	-0.04	-0.13
Dependability	-0.04	-0.03	-0.1
Adaptiveness	-0.04	-0.19	-0.16
Originality	-0.17	-0.13	-0.11
Originality Scope	-0.01	0.03	-0.01
Cooperativeness	-0.08	-0.14	-0.07
Freshness	-0.03	0.08	0.07
Integrability	-0.17	-0.22	-0.18
Applicability Scope	0.11	0.07	0.09
Constructiveness	-0.09	-0.04	-0.049
Scope	-0.10	-0.15	-0.22
Dependability	0.22	0.15	0.46

quite significant in comparison to the mean correlation value of all datasets (see Table 4). This could lead to a conclusion that the behaviour of idea characteristics (expressed with our metrics) in relation to community activity (measured with legacy metrics) is individual for every particular environment and setting of idea campaign.

As a follow up, we also measured correlations exclusively between the metrics proposed in this article. Yet again, the results were very different depending on the dataset, however some metrics within the scope of a single dataset were strongly correlated allowing to make interesting observations about the communities:

- **Starbucks:** ideas for products impose more modifications in existing offering than ideas for services or processes (strong correlation of Implementation Dependability and Integrity)

---

<sup>1</sup>correlation undefined for one of the datasets

Table 4: Mean and standard deviation of bivariate correlations between all datasets for Gi2MO  
Types metrics and legacy metrics

Metric	# Comments		Rating		Idea Age		Idea Adoption	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Completeness	0.06	0.02	0.08	0.04	-0.02	0.04	0.12	0.05
Experience Completeness	0.02	0.13	-0.01	0.05	-0.01	0.13	0.00	0.14
Situational Dependence	0.08	0.07	0.11	0.11	0.10	0.10	0.10	0.13
Relatedness	-0.01	0.10	-0.01	0.08	-0.10	0.12	0.01	0.12
Dependability	n/a <sup>1</sup>	n/a <sup>1</sup>	n/a <sup>1</sup>	n/a <sup>1</sup>	n/a <sup>1</sup>	n/a <sup>1</sup>	n/a <sup>1</sup>	n/a <sup>1</sup>
Adaptiveness	0.01	0.06	-0.05	0.13	-0.07	0.06	-0.01	0.10
Originality	0.00	0.12	0.01	0.10	-0.08	0.16	-0.01	0.18
Originality Scope	0.02	0.05	-0.01	0.03	0.03	0.14	-0.01	0.16
Cooperativeness	0.07	0.13	0.02	0.14	-0.07	0.16	0.06	0.10
Freshness	n/a <sup>1</sup>	n/a <sup>1</sup>	n/a <sup>1</sup>	n/a <sup>1</sup>	n/a <sup>1</sup>	n/a <sup>1</sup>	n/a <sup>1</sup>	n/a <sup>1</sup>
Integrability	-0.10	0.09	-0.11	0.11	-0.08	0.08	-0.06	0.06
Applicability Scope	-0.04	0.12	0.01	0.06	0.02	0.08	-0.09	0.06
Constructiveness	0.04	0.16	0.04	0.10	-0.02	0.03	-0.03	0.08
Scope	0.04	0.13	-0.02	0.13	-0.06	0.12	-0.03	0.08
Dependability	0.08	0.11	0.06	0.12	0.14	0.22	0.11	0.13

- **Cisco:** the only instance where inventors connect new products with gain of new type of customers (strong negative correlation between Constructiveness and Adaptiveness); generic ideas for product types are related to modifications in existing offering, while specific and detailed ideas are more typical for completely new items (strong negative correlation between Constructiveness and Applicability Scope)
- **Ubuntu:** very original ideas are also the ones that deliver most complete description (strong correlation of Originality and Idea Completeness); similar as in StarBucks product ideas impose more modifications in related items of offering (strong correlation of Implementation Dependability and Integrity)
- **IdeaStream:** none of the metrics had a strong correlation

Except for the IdeaStorm instance, all other datasets had one single standing

out similarity: ideas that proposed a complete structure change of products often referred to creating new products rather than redesigning old ones (very strong correlation between Implementation Scope and Constructiveness).

Finally, in addition to correlations between metrics, we measured and compared the correlations of all metrics to idea adoption (determines if an idea was implemented or not) to see if our metrics do better or worse as a tool for detecting good ideas. In case of IdeaStorm (see Table 5), in majority of cases, our metrics had a better correlation with idea adoption than the legacy metrics. The most standing out results were achieved by Innovation and Object metrics. Nevertheless, according to the Cohen scale the impact in best cases can be described as medium. After repeating the experiment for the 3 other datasets (myStarBucks, Cisco and Ubuntu), the final conclusions were similar.

Table 5: Bivariate correlations between metrics and idea adoption (Dell IdeaStorm dataset)

<b>Metric</b>	<b>Idea Adoption</b>	<b>Metric</b>	<b>Idea Adoption</b>
Completeness	0.16	Freshness	0.05
Experience Completeness	-0.03	Integrability	-0.1
Situational Dependence	-0.07	Applicability Scope	-0.09
Relatedness	0.02	Constructiveness	-0.15
Dependability	-0.07	Scope	0.14
Adaptiveness	0.05	Dependability	0.23
Originality	-0.27	# Comments	-0.04
Originality Scope	-0.2	Rating	-0.04
Cooperativeness	0.01	Idea Age	-0.06

Concluding all experiments with the correlation measure, the proposed metrics provide a small improvement over the legacy metrics in terms of picking the winning ideas. Our results show that Idea Originality as well as Object Dependability are better indicators than any other. Additionally, the correlation analysis delivered another proof that our metrics can be used for comparison of different environments and discovering characteristics of the communities.

## 6. Conclusions

We have proposed a set of new automatically generated metrics to aid the decision making process during the assessment of ideas in Idea Management Systems. Our hypothesis was that these metrics could be derived from annotations made with a specially crafted taxonomy, and used to characterise community generated innovation in a sufficient way to compare the gathered data. This hypothesis has been confirmed with a number of experiments that used the taxonomy as a tool to discover differences and similarities of various case studies.

Furthermore, we presented an evaluation of all the steps underlying the generation of metrics and obtained valuable insight into conditions under which our methodology performs best. We determined that out of four proposed taxonomy sub-trees substantial parts of two (Trigger and Object) can be applied automatically with satisfactory results, while the characteristics represented in the two remaining sub-trees (Innovation and Proposal Type) should be analysed and applied by a human. Furthermore, we have shown that the manual annotation delivers significantly better results when done by a single annotator rather than a group (regardless of the level of expertise with innovation theory).

Finally, we evaluated the use of metrics not only for comparison of entire datasets but also for decision making process of selecting the individual ideas for implementation. We determined that the borderline cases of community activity that are currently used for filtering ideas (vote count, comment count etc.) do not influence the values of metrics proposed by us (e.g. more original ideas are not more commented or voted on). In addition, the obtained results have shown that our metrics deliver slightly better results to predict winning ideas in comparison with the contemporary used community metrics. Most notably, our results show best performance for Idea Originality and Object Dependability as best measures of idea adoption, standing out in comparison to any other metric.

In terms of future work we envision to peruse a fully automated approach by

putting more impact on analysis of different automatic annotation methods and attempting to simplify the taxonomy without much sacrifice on the level of knowledge that it carries. Furthermore, the introduction of the taxonomy opens a range of new possibilities for clustering and ranking ideas that could be a significant step toward bringing better organisation to Idea Management.

### **Acknowledgements**

This research has been partly funded by the Spanish Ministry of Industry, Tourism and Trade through the project RESULTA (TSI-020301-2009-31) and Spanish CENIT project THOFU. In addition we would like to express our gratitude to: Nikos Bikakis, Giorgos Giannopoulos, Tadhg Nagle and George Anadiotis for their help and input during various stages of our research.

### **References**

- Abernathy, W., Utterback, J., 1978. Patterns of innovation in technology. *Technology Review* 80 (7), 40– 47.
- Abernathy, W. J., Clark, K. B., February 1985. Innovation: Mapping the winds of creative destruction. *Research Policy* 14 (1), 3– 22.
- Adamides, E. D., Karacapilidis, N., 2006. Information technology support for the knowledge and social processes of innovation management. *Technovation* 26 (1), 50–59.
- Adner, R., Levinthal, D., 2001. Demand heterogeneity and technology evolution: implications for product and process innovation. *Management Science* 47, 611–628.
- Alexandru Spatariu, K. H., Bendixen, L. D., 2004. "defining and measuring quality in online discussions. *Journal of Interactive Online Learning* 2 (4).

- Bailey, B. P., Horvitz, E., 2010. Whats your idea? a case study of a grassroots innovation pipeline within a large software company. In: CHI '10 Proceedings of the 28th international conference on Human factors in computing systems.
- Baker, N. R., Freeland, J. R., 1972. Structuring information flow to enhance innovation. *Management Science* 19, 105–116.
- Baumgartne, J., 2008. An introduction to idea management. Tech. rep., Bwiti bvba.
- Belecianu, R., 2009. D6.3.1 sap living lab environment. Tech. rep., Laboranova Collaboration Environment for Strategic Innovation.
- Bikakis, N., Giannopoulos, G., Dalamagas, T., Sellis, T., 2010. Integrating keywords and semantics on document annotation and search. In: The 9th International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE 2010).
- Bothos, E., Apostolou, D., Mentzas, G., 2008. Idem: A prediction market for idea management. designing e-business systems. markets, services, and networks. In: 7th Workshop on E-Business, WeB 2008. Paris, France.
- Brown, J., 2006. Managing product relationships: Enabling iteration and innovation in design. Tech. rep., AberdeenGroup.
- Brown, M., Powers, S., Nicolson, N., March 2009. Vendor landscape: Innovation management software. Tech. rep., Forrester.
- Chuang, L.-M., Liu, C.-C., Tsai, W.-C., Huang, C.-M., May 2010. Towards an analytical framework of organizational innovation in the service industry. *African Journal of Business Management* 4 (5), 790– 799.
- Cohen, J., 1988. Statistical power analysis for the behavioral sciences. Lawrence Erlbaum.

- Conn, S. J., Torkkeli, M. T., Bitran, I., 2009. Assessing the management of innovation with software tools: an application of innovationenterprizer. *International Journal of Technology Management* 45 (3/4), 323–336.
- Damanpour, F., Gopalakrishnan, S., March 1998. Theories of organizational structure and innovation adoption: the role of environmental change. *Journal of Engineering and Technology Management* 15 (1), 1– 24.
- Dasarathy, B. V., 1990. Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques. IEEE Computer Society.
- Dell, 2011. Form 10-k for fiscal 2009. Tech. rep., Dell Inc.
- Drejer, I., April 2004. Identifying innovation in surveys of services: a schumpeterian perspective. *Research Policy* 33 (3), 551– 562.
- Eris, E. D., Saatcioglu, O. Y., July 2006. A system look for technological innovation: Firm based perspective. In: *European and Mediterranean Conference on Information Systems (EMCIS)*. Costa Blance, Alicante, Spain.
- Feldstein, A. P., 2007. Brand communities in a world of knowledge-basedproducts and common property. In: *MIT5: creativity, ownership and collaboration in digital age*.
- Fenn, J., LeHong, H., July 2011. Hype cycle for emerging technologies, 2011. Tech. rep., Gartner.
- Gangi, P. M. D., Wasko, M., 2009. Steal my idea! organizational adoption of user innovations from a user innovation community: A case study of dell ideastorm. *Decision Support Systems* 48, 303– 312.
- Garcia, R., Calantone, R., 2002. A critical look at technological innovation typology and innovativeness terminology: a literature review. *The Journal of Product Innovation Management* 19, 110—132.
- Gilfillan, S. C., 1935. *The Sociology of Invention*. Follett Publishing Company.

- Hrastinski, S., Kviselius, N. Z., Edenius, M., 2010. A review of technologies for open innovaton: Characteristics and future trends. In: Proceedings of the 43rd Hawaii International Conference on System Sceiences.
- Hughes, T. P., 1975. Innovators: the problems they choose, the ideas they have, and the innovations they make. In: Kelly, P., Kranzberg, M., Rossini, F. A., Baker, N. R., Tarpley, F. A., Mitzner, M. (Eds.), Technological innovation: A critical review of current knowledge. Advanced Technology and Science Studies Group, Georgia Tech.
- Jarratt, T. A. W., Eckert, C. M., Caldwell, N. H. M., Clarkson, P. J., 2010. Engineering change: An overview and perspective on the literature. *Research in Engineering Design* 22 (2), 103–124.
- Jouret, G., September 2009. Inside ciscos search for the next big idea. *Harvard Business Review*.
- Kelly, P., Kranzberg, M., Rossini, F. A., Baker, N. R., Tarpley, F. A., Mitzner, M., 1975. Ecology of innovation. In: Kelly, P., Kranzberg, M., Rossini, F. A., Baker, N. R., Tarpley, F. A., Mitzner, M. (Eds.), Technological innovation: A critical review of current knowledge. Advanced Technology and Science Studies Group, Georgia Tech.
- Kirschner, P. A., Buckingham-Shum, S. J., Carr, C. S., 2002. Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making. Springer.
- Klein, M., 2012. Enabling large-scale deliberation using attention-mediation metrics. *Computer Supported Cooperative Work* 21 (4-5), 449–473.
- Lakhani, K. R., von Hippel, E., 2003. How opensource software works: free user-to-user assistance. *Research Policy* 32 (6), 923943.
- Lucene, 2012. Lucene implementation of knn classifier.  
URL [http://lucene.apache.org/core/old\\_versioned\\_docs/versions/3\\_0\\_1/api/core/index.html](http://lucene.apache.org/core/old_versioned_docs/versions/3_0_1/api/core/index.html)



- Myers, S., Marquis, D. G., 1969. Successful Commercial Innovation,. National Science Foundation.
- Narasimhalu, A., 2005. Innovation cube: Triggers, drivers and enablers for successful innovations. In: ISPIM.
- Ning, K., O’Sullivan, D., Zhu, Q., Decker, S., 2006. Semantic innovation management across the extended enterprise. *International Journal on Industrial and Systems Engineering* 1.
- Nisbet, D., 2004. Measuring the quantity and quality of online discussion group interaction. *Journal of eLiteracy* 1.
- Perey, C., 2008. Beyond eyeballs: Improving social networking metrics. In: W3C Workshop on the Future of Social Networking.
- Popadiuka, S., Choo, C. W., 2006. Innovation and knowledge creation: How are these concepts related? *International Journal of Information Management* 26 (4), 302– 312.
- Rozwell, C., Harris, K., Mesaglio, M., August 2010. Who’s who in innovation management technology. Tech. rep., Gartner.
- Rubalcaba, L., Gallego, J., Gago, D., 2010. On the differences between goods and services innovation. *Journal of Innovation Economics* 30 (4), 611–628.
- Schumpeter, J., 1934. *The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest and the Business Cycle*. Harvard University Press, Cambridge, MA.
- Starbucks, 2011. Starbucks corporation fiscal 2011 annual report. Tech. rep., Starbucks Corporation.
- Stromer-Galley, J., 2007. Measuring deliberation’s content: a coding scheme. *Journal of Public Deliberation* 3 (1).

- Susman, G., Warren, A., Ding, M., 2006. Product and service innovation in small and medium-sized enterprises. Tech. rep., Smeal College of Business. The Pennsylvania State University.
- Turrell, M., August 2002. Idea management and the suggestion box. Tech. rep., Imaginatik.
- Ubuntu, 2011. Ubuntu homepage.  
URL <http://www.ubuntu.com/>
- Usher, A. P., 1954. A History of Mechanical Invention. Harvard University Press.
- Utterback, J. M., Abernathy, W. J., 1975. A dynamic model of product and process innovation. *Omega* 3 (6), 639–656.
- Westerski, A., Iglesias, C. A., 2011. Exploiting structured linked data in enterprise knowledge management systems: An idea management case study. In: Enterprise Distributed Object Computing Conference Workshops (EDOCW), 2011 15th IEEE International.
- Westerski, A., Iglesias, C. A., Nagle, T., 2011. The road from community ideas to organisational innovation: A life cycle survey of idea management systems. A Special Issue of the Journal Web-Based Communities. Community-based Innovation: Designing Shared Spaces for Collaborative Creativity.
- Westerski, A., Iglesias, C. A., Rico, F. T., 2010. A model for integration and interlinking of idea management. In: Metadata and Semantic Research: 4th International Conference, MTSR 2010. Alcalá de Henares, Spain.