Effects of Data Exploration and Use of Data Mining Tools to Extract Knowledge from Databases (KDD) in Early Stages of the Engineering Design Process (EDP)

by

Ma Lorena ESCANDON-QUINTANILLA

MANUSCRIPT-BASED THESIS PRESENTED TO ÉCOLE DE TECHNOLOGIE SUPÉRIEURE IN PARTIAL FULFILLEMENT FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
Ph.D.

MONTREAL, JUNE 5, 2017

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

BOARD OF EXAMINERS

THIS THESIS HAS BEEN EVALUATED

BY THE FOLLOWING BOARD OF EXAMINERS

Mr. Mickaël Gardoni, Thesis Supervisor
Department of Automated Production Engineering, École de technologie supérieure

Mr. Patrick Cohendet, Thesis Supervisor
Department of International Business, HEC Montréal

Mr. Yvan Beauregard, President of the Board of Examiners
Department of Mechanical Engineering, École de technologie supérieure

Mr. Michel Rioux, Member of the Board of Examiners
Department of Automated Production Engineering, École de technologie supérieure

Mr. Vincent Sabourin, External Examiner
Department of Strategy, Social and Environmental Responsibility, UQAM

THIS THESIS WAS PRESENTED AND DEFENDED

IN THE PRESENCE OF A BOARD OF EXAMINERS AND THE PUBLIC

ON JUNE 1ST, 2017

AT ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

ACKNOWLEDGMENTS

The last four years have been some of the most challenging and exciting of my career. For this, I owe credit to my advisors, Mickael Gardoni and Patrick Cohendet. For everything they taught me, their guidance, and the wonderful opportunities they made possible, thank you!

I am also grateful to the board of examiners who took the time to review my thesis, and for their comments to improve the end result: Dr. Yvan Beauregard, Dr. Michel Rioux, and Dr. Vincent Sabourin.

A very special mention to Louis Côté and all the *ÉTS Summer School on innovation and technological design* team, Marjolaine, Hélène, Yung, Jorge, Mario and Samuel, who worked tirelessly to create and manage the course. Thank you for your trust and support.

The continued support by TKM, who provided the data mining software, is also to thank. In particular Mr. Christophe Lecante, Florian and Romain.

I greatly appreciate our research group, Luz Maria J.N., Ahmed C., Mario D., Patrick M., Florent L.N., Arash N., and Thierry Z., for their continuous support throughout my studies. Also, the volunteers at the *24 hours of innovation* organization team, the members of *DécliQ*, and the researchers at the group *Conception Innovante*, especially Laurence and Marine.

Alejandro, my partner in adventures, also deserves great commendation. Thank you for being in my team, for your support, and your confidence in us.

Mom, dad, este reconocimiento va más allá de mis logros académicos, todo lo que soy es gracias a ustedes, me enseñaron a seguir mis sueños y por siempre se los voy a agradecer.

Antero and Gerardo, you inspire me by following your own paths; I admire your strengths and courage.

Laura and Edgar, thank you for the encouragement, the lively discussions, and most of all, your true friendship.

This doctorate was made possible by the financial support from the Mexican National Council for Science and Technology (CONACyT), and the 2014 Complementary scholarship from the Mexican Ministry of Public Education (DGRI-SEP).

EFFETS DE L'EXPLORATION DE DONNÉES ET DE L'UTILISATION D'OUTILS DE *DATA MINING* POUR EXTRAIRE DES CONNAISSANCES À PARTIR DE BASES DE DONNÉES (KDD) DANS LES PREMIÈRES ÉTAPES DU PROCESSUS DE CONCEPTION D'INGÉNIERIE (EDP)

Ma Lorena ESCANDON-QUINTANILLA

RÉSUMÉ

Cette thèse décrit des travaux de recherche originaux dont l'objectif était de fournir aux équipes l'accès aux données et d'observer l'effet de son utilisation aux premières étapes créatives du processus de conception d’ingénierie. À la suite d'une recherche théorique sur l'utilisation des technologies de l'information pour soutenir la génération d'idées et l'utilisation des données comme entrée créatif, une procédure a été conçue suite au processus de découverte de la connaissance des bases de données (KDD) et testée sur plusieurs itérations d'amélioration travaillant avec des équipes créatives dans différents contextes.

Après deux études exploratoires, trois cas ont été réalisés où le chercheur a tenté de mieux appuyer les différentes étapes du EDP par l'application de données provenant de l'exploitation des brevets. Pour observer les différences, nous avons fourni trois niveaux d'accès pour explorer les données dans un outil de *data mining*: bas, intermédiaire et élevé.

- Cas 1 - Les participants à une séance de créativité ont été invités à identifier des besoins ou des problèmes (première étape du processus de conception d’ingénierie). Ils ont eu un accès intermédiaire pour explorer les données dans un outil de *data mining*; ils pourraient explorer, mais pas faire de nouvelles recherches ou ajouter des données. L'analyse des résultats indique que les participants gravitent vers des termes et des mots-clés liés à des idées précédemment générées, de sorte que l'augmentation de la nouveauté est faible. Afin de corriger la question de l'exploration intermédiaire, il a été décidé de former les participants à l'utilisation de l'outil de *data mining* pour les cas suivants; si les équipes ont plus de liberté pour explorer les données, elles peuvent générer des combinaisons plus nouvelles.

- Cas 2 - Les équipes chargées de relever les défis techniques d'un cours ont été formées à l'utilisation de l'outil d'exploration de données. Ils ont ensuite été invités à continuer à utiliser l'outil pour générer de nouvelles idées. Dans ce cas, les équipes avaient un accès élevé à l'outil d'exploration de données; ils ont pu ajouter des données et effectuer des recherches. Les équipes qui ont choisi d'explorer les données pour un soutien créatif ont trouvé des améliorations ou des composants à partir de

solutions existantes pour faire avancer leur propre conception, et ont reçu des évaluations plus positives par un jury d'experts. Cependant, l'objectif d'obtenir des solutions plus diverses ou nouvelles n'a pas été atteint. Une explication possible est que l'utilisation de l'outil peut surcharger les participants avec trop d'options à explorer, menant les équipes à revenir aux solutions connues. Une contre-réaction possible pour résoudre la question de trop d'options est d'avoir un acteur externe (comme un modérateur) extraire des mots-clés à partir des données, et de fournir aux participants ces termes pour combiner dans des idées nouvelles.

- Cas 3 - Les équipes participant à un concours d'innovation ont reçu des mots clés choisis par un expert sur l'outil. Les participants avaient un faible accès à explorer les données dans un outil de *data mining*. Le chercheur a effectué l'analyse des données pour deux défis dans la compétition, et a sélectionné des mots clés pertinents provenant de la base de connaissances du problème. Les résultats montrent que les équipes qui ont choisi les défis supporté par les mots-clés ont généré des idées plus diverses et nouvelles, par rapport aux équipes sans le soutien. En fournissant des mots-clés pertinents, il était possible d'obtenir les avantages du KDD sans les questions de formation des participants sur l'utilisation de l'outil, et les ressources qui les équipes devraient consacrer pour explorer les données.

En conclusion, les données et le KDD peuvent être utilisés comme une entrée créative pour un EDP à différentes étapes. Il est recommandé de déterminer si l'objectif d'inclure des données dans un effort EDP est de générer une idée nouvelle ou de résoudre un problème. Pour générer des idées nouvelles, il semble préférable de fournir des données sous la forme de mots-clés sélectionnés par un acteur externe, pour inciter les combinaisons originales. Si l'équipe recherche des améliorations ou des éléments de solutions existantes, il semble bénéfique d'avoir accès à une base de connaissances à explorer. Il est important de délimiter l'exploration afin de ne pas être étourdis en raison de la quantité d'information disponible.

Pour les trois expériences, le logiciel IPMetrix a été utilisé pour effectuer l'exploration de données. Le processus de sélection, de chargement, de nettoyage et de transformation des données est décrit dans chaque chapitre, en fonction du travail effectué sur les données pour le cas spécifique.

Mots-clés: Processus de conception d'ingénierie, découverte de connaissances, extraction de brevets, exploration de données

EFFECTS OF DATA EXPLORATION AND USE OF DATA MINING TOOLS TO EXTRACT KNOWLEDGE FROM DATABASES (KDD) IN EARLY STAGES OF THE ENGINEERING DESIGN PROCESS (EDP)

Ma Lorena ESCANDON-QUINTANILLA

ABSTRACT

This thesis describes original research work where the objective was to provide teams with access to data, and observe the effect of its use at the early creative stages of the engineering design process. Following a theoretical research on the use of information technologies to support idea generation, and the use of data as creative input, a procedure was designed following the Knowledge Discovery from Databases process, and tried over several iterations of improvement working with creative teams in different contexts.

After two exploratory studies, three cases were performed where the researcher attempted to better support the different stages of the EDP through the application of data from patent mining. To observe the differences, we provided three levels of access to explore data in a data mining tool: low, intermediate and high.

- Case 1 - Participants in a creativity session were asked to identify needs or problems (first stage of the engineering design process). They were given intermediate access to explore data in a data mining tool, meaning they could explore, but not make new searches or add data. The analysis of the results indicates that participants gravitated towards terms and keywords related to previously generated ideas, thus the increase in novelty was low. In order to correct the issue of intermediate exploration, it was decided to train participants in the use of the data mining tool for subsequent cases; if teams have more freedom to explore data, they can potentially generate more novel combinations.

- Case 2 - Teams tasked with engineering challenges in a course were trained in the use of the data exploration tool. They were then invited to continue using the tool to generate new ideas. In this case, teams had high access to the data exploration tool; they were able to add data, and make searches. Teams who chose to explore data for creative support found improvements or components from existing solutions to advance their own design, and received more positive evaluations by a jury of experts. However, the objective of obtaining more diverse or novel solutions was not achieved. A possible explanation is that the use of the tool can overwhelm participants with too many options to explore, leading teams to return to known solutions. A possible counteraction to resolve the issue of too many options is to have

an external actor (such as a moderator) extract keywords from the data, and provide participants with these terms to combine into novel ideas.

- Case 3 - Teams participating in an innovation contest were given keywords selected by an expert on the tool. In other words, participants had low access to explore data in a data mining tool. The researcher performed the data analysis for two challenges in the competition, and selected keywords relevant to the knowledge base of the problem. The results show that teams who selected the keyword supported challenges generated more diverse and novel ideas, compared to teams without the support. By providing relevant keywords, it was possible to obtain the benefits of the KDD without the issues of training participants on the use of the tool, and the resources teams would have to dedicate to explore the data.

It was concluded that data and KDD can be used as a creative input for an EDP at different stages. It is recommended to determine whether the objective of including data in an EDP effort is to generate a novel idea or to solve a problem. To generate novel ideas, it seems preferable to provide data in the form of keywords selected by an external actor, to prompt original combinations. If the team is searching for incremental improvements or elements of existing solutions, then it appears to be beneficial to have access to a knowledge base to explore. It is important to delimit the exploration to avoid becoming stunned because of the amount of available information.

For the three experiences, the software IPMetrix was used to perform the data mining. The process of data selection, loading, cleaning and transformation is described in each chapter, according to the work performed on the data for the specific case.

Keywords: engineering design process, knowledge discovery, patent mining, data mining

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABREVIATIONS

DM	Data mining
ED	Engineering design
EDP	Engineering design process
ICT	Information and communication technologies
IS	Information system
KB	Knowledge base
KD	Knowledge discovery
KDD	Knowledge discovery in databases
KM	Knowledge management
NPD	New product development
SME	Small and medium enterprises

INTRODUCTION

Context

Creative teams in engineering are continuously challenged to design novel solutions to our everyday problems. Engineering teams work to solve technological challenges, relying on all the information they can possibly process and retain (Sim & Duffy, 2003). This brings us to the issue of information: a person can only process and retain so much information, but there is a constant stream of new data being generated. The speed with which a domain of technology advances surpasses human capacity to acquire, process, and retain this information.

Information technologies in data mining have advanced in the processing of large amounts of information, which allow for a quick overview of a domain, while also allowing the user (in this case, engineering teams) to dig deeper into the specifics when needed. Compared to other combinations of technology, for example, a web search and a spreadsheet or text document, data mining tools and techniques enable the inclusion of different data sources, the application of different algorithms to obtain a particular view or analysis, and the ability to relate data automatically (Nielsen, 2012).

Though companies are already using data mining to extract business intelligence, companies able to source and process data more efficiently are winning the market. But the applications and benefits are not restricted to business intelligence (forecasting, marketing, etc.); it can also be applied to the design and development of new products, devices and solutions.

The advancement of information technologies now enables a wealth of information to be digitally documented, exploited and reused for further knowledge creation. The information can be mined using big data techniques to find the most common terms and correlations between ideas (Chen, Li & Hung, 2013).

It is possible to map the connections between concepts in a domain by using big data techniques to analyze the knowledge in patents and scientific articles. The resulting visualizations of information can be used as input to help bolster the creativity of participants generate ideas given that, as described by Hamman (2000), “creative thinking involves a process of iterative activation of ‘cues’”; furthermore, the likelihood of creating new knowledge from recombination is greater as we increment the number of external inspirations (Cohen & Levinthal, 1989, cited in Kabir & Carayannis 2013).

To create something new, it is necessary to combine what we already know (Gilfillan, 1935; Schumpeter, 1939; Nelson & Winder, 1982; Basalla, 1988, and Fleming & Sorenson, 2004 cited in Fleming & Szigety, 2006). Arthur Koestler (1964) coined the term bisociation, a creative act where a situation or idea is perceived in two incompatible frames of reference (can also be defined as associative contexts, types of logic, codes of behavior, universes of discourse, matrices of thought) and the subject is able to meld them together by “thinking aside”.

Research problem

As mentioned before, there is a difficulty to be solved between the large amounts of information available for creative teams, and the time and effort it requires to process. It is estimated that engineering designers can spend up to 30% of their time searching for information relevant to their problem (Sim & Duffy, 2003). Data mining tools can help condense the information and make exploration easier for teams. The problem addressed in this thesis is to determine the stages in the process where this creative support can generate greater benefits, particularly regarding the variety of ideas, and novelty of the final solution.

Objectives

This research aims to propose and observe the use of data mining tools and data as input for the early creative phases of engineering design. The specific objectives are:

1) Review the literature on the use of data as input for creative design
2) Determine the stages of the engineering design where data mining tools and data can be integrated to support the process
3) Propose the use of data mining tools and data to creative engineering teams at different stages of the process
4) Document the results of the use of data mining tools and data in the EDP

Research question

The following research question guided the work during this thesis:

> *How can the use of data mining tools and data support early creative stages of engineering design?*

Scope

The work described in this thesis aims to propose the use of data mining tools, and the resulting information for analysis and insight, to support the early creative stages of engineering design. It does not, however, include the materialization of a concept into a functional product. It also covers the results at a team level, not individual, meaning it does not take into account the creative production of individual team members, but the result of the effort of the team as a whole. It is also worth noting that the application is limited to engineering design.

Structure of the thesis

As this is a thesis by articles, first a theoretical framework is presented, which guided the case studies. Then, chapters 2 to 5 present the case studies and results of the research work, and finally, a conclusion and discussion completes this work:

- Chapter 1 presents the conceptual framework for the work here documented. It discusses the theories enabling the development of the thesis, the methodology, Design Science Research, an introduction to idea generation and bisociation, Data mining, Knowledge discovery from databases (KDD) and Engineering design process (EDP).
- Chapter 2 presents article 1, “Big Data Analytics as Input for Problem Definition and Idea Generation in Technological Design”, in which teams have to identify new issues or problems in a domain. This article was presented in the PLM16 conference in Columbia, SC, USA on July, 2016, and later published as part of the proceedings of said conference.
- Chapter 3 presents article 2, “Improving concept development with data exploration in the context of an innovation and technological design course”, which documents the development of a prototype for the same challenges, where half the teams opted to use the data mining tool to explore data related to the problem domain. This article has been published in the *International Journal on Interactive Design and Manufacturing (IJIDeM)* on February, 2017.
- Chapter 4 presents article 3, “Prompting inventive solution design with keyword cues from patent mining in an innovation competition”, where participants in an innovation competition propose novel solutions to the issues stemming from the session in Chapter 2. This article has been submitted to the Creativity and Innovation Management journal.
- Chapter 5 presents article 4, “Effects of information cues from knowledge discovery in the early creative stages of engineering design”; it describes the findings from the three cases in terms of the support that data exploration and the results from data mining

provided the creative teams at different stages of the engineering design process. This article has been submitted to the Journal of Engineering Design.

- Chapter 6 presents the discussion and conclusion. The chapter includes a summary of the cases, an overview of the articles presented, the limitations of the research, a discussion of the results, implications for the industry and future work, which is divided into future work on engineering design, and on the application of data mining tools for engineering design.

Abstracts and conference communications

Additional communications were presented at conferences during the development of this research. For clarity and length purposes, they are not included in this thesis. The works are described in the following paragraphs.

The first work presented at a conference was an abstract entitled “Opportunities to exploit Big data in idea generation sessions”, presented by co-author Mickaël Gardoni at the International Conference on Industrial Engineering and Operations Management IEOM15 in Dubai, UAE on March 3 - 5, 2015. The abstract was prepared by Ma-Lorena Escandon-Quintanilla and Patrick Cohendet as well. In this paper, the authors identified the stages of the idea generation process where big data tools and techniques could be used to support creative teams. They suggest creative teams can benefit from big data analytics throughout the idea generation process: to identify areas of opportunity (need identification), have information as input for inspiration (information gathering), identify unrelated ideas to promote bisociation (idea generation), and to obtain insight from a large amount of ideas from a crowdsourcing effort (evaluation).

The first conference article, "Strategies to employ social networks in early design phases (idea generation)", was written with Luz-Maria Jimenez-Narvaez, and Professor Mickael

Gardoni for the 20th International Conference on Engineering Design ICED15, which took place in Milan on July 27-30, 2015. It discusses the use of social media as input for creative teams trying to solve a problem, identifying the different issues when adding a new technology to an idea generation session. One key takeaway from this particular paper is the need to facilitate recombination in creativity sessions, a recurring theme in this work.

Two papers were presented by Ma-Lorena Escandon-Quintanilla at the 13th IFIP International Conference on Product Lifecycle Management PLM16, which took place July 11-13 in Columbia, SC, USA. The first paper, penned with Mickaël Gardoni and Patrick Cohendet, titled “Big data analytics as input for problem definition and idea generation in technological design” is presented in Chapter 2 as it is directly linked to the articles derived from the work presented in this thesis, in Chapters 3, 4 and 5.

The second paper presented at PLM16 was composed with colleague and first author Patrick Mbassegue and Professor Mickaël Gardoni. The work, “Knowledge Management and Big Data: opportunities and challenges for small and medium enterprises (SME)” presents a theoretical basis for the opportunities and challenges that can stem from the use of big data tools and techniques in the context of knowledge management for SMEs, considering their particular limitations regarding their financial, human and technological resources.

CHAPTER 1

CONCEPTUAL FRAMEWORK & METHODOLOGY

This chapter presents the conceptual framework and methodology that supported the development of the thesis. First, the theoretical background is described: the first section introduces the concepts of Engineering design process (EDP) and Knowledge discovery from databases (KDD), as well as an overview on the uses of data as an input for creativity. Then, the chapter continues to present the methodological framework applied to guide the definition of goals for each experience and the continuous improvement for future iterations.

1.1 Theoretical background

This first section of this chapter presents the theoretical background that helps frame the cases followed in this thesis. It presents the existing literature on engineering design process (EDP), knowledge discovery from databases (KDD), the use of data as creative input, data mining for creativity, and bisociation. A deeper understanding of creative teams and their process provided the basis for the design of the information support during the three cases.

1.1.1 Engineering design process

Though the concept originated much earlier, the Accreditation Board for Engineering and Technology (ABET), defined Engineering Design in 1996 as “the process of devising a system, component, or process to meet desired needs. It is a decision making process (often iterative), in which the basis sciences, mathematics, and engineering sciences are applied to convert resources optimally to meet a stated objective” (in Ertas & Jones, 1996, page 2).

Engineering design process is the series of steps, stages or activities an engineering team goes through when designing a new solution to an engineering problem (Sim & Duffy, 2003, Atman et al., 2007). The process is usually described as being non-linear and iterative, designers go back and forth between stages or activities when they are faced with an issue, or discover new information about the problem. Shneiderman et al. (2006) proposed the following phases for new product development cycles:

1) Problem definition (need identification)
2) Information gathering
3) Idea generation
4) Modeling (description of potential solutions)
5) Feasibility analysis
6) Evaluation
7) Selection
8) Communication
9) Implementation

Atman et al. (2007) later completed the above process with “Need identification” to adapt the process to the Engineering Design Process (EDP), shown in Table 1.1.

Table 1.1 Engineering design process,
taken from Atman et al. (2007)

Design stages	Design activities
Problem scoping	Need identification Problem definition Information gathering
Development of alternative solutions	Idea generation Modeling Feasibility analysis Evaluation
Project realization	Selection Communication Implementation

The design process is usually performed in the form of work sessions, which set an environment and implement creativity techniques that help participants produce, combine and express ideas.

It has been found that more experienced engineers spend more time in the initial phases of the process (Atman et al., 2007), as they know through experience that more information in the first stages will ultimately save time and iterations later.

1.1.2 Idea generation

Idea generation “is central to engineering design” (Glier et al., 2011), and it is a fundamental step of the innovation process. Ideas are not fully developed solutions that can be patented or launched to market, they are a notion in development, and will need further work. According to Cohendet, Parmentier and Simon (2016), a larger investment of resources is required for an idea to be developed into a concept with value that can be implemented.

Studies show that creativity techniques are useful, and they usually induce participants to explore ideas "outside their normal frame of reference" (Dove & Jones, 2014). Ideas of others sometimes promote the creation of related ideas or new ideas, working as a sort of "intelligent trigger" (Munemori, Yoshino & Yunokuchi 2001) where one piece of information triggers the generation of new ideas. However, more research is needed to find how to support idea generation using ICTs (Ardaiz-Villanueva et al., 2011).

1.1.3 Bisociation

In the middle of last century, Arthur Koestler found that innovative ideas are generated when two fields of knowledge previously considered incompatible are connected in a *bisociation*; the juxtaposition creates a spark of creativity that leads to something completely different from existing solutions to a problem (Koestler, 1964). This belief has been also postulated under different terms, such as conceptual blending or forced relationships.

Koestler suggests that *really* creative combinations "result from a blending of elements drawn from of two previously unrelated frames or matrices of thought into a new matrix of meaning by way of a process involving comparison, abstraction and categorization, analogies and metaphors" (1964). Nielsen echoes the feeling by stating that creative ideas stem from the combination of unrelated ideas (2012).

Bisociation is trying to blend together to domains of knowledge, disciplines or ways of thinking that are seemingly unrelated or incompatible, and coming up with something completely different. Thinking in one single matrix can perform tasks only of a kind already encountered in past experience, this is associative thought; it is not capable of original, creative achievement. Table 1.2 shows the contrast between associative and bisociative thought:

Table 1.2 Comparison between associative and bisociative thought, taken from Koestler (1964)

Associative thought (habit)	**Bisociative thought (originality)**
Association within the confines of a given matrix	Bisociation of independent matrices
Guidance by pre-conscious or extra-conscious processes	Guidance by sub-conscious processes normally under restraint
Dynamic equilibrium	Activation of regenerative potentials
Rigid to flexible variations on a theme	Super-flexibility
Repetitiveness	Novelty
Conservative	Destructive-Constructive

1.1.4 Innovation contests

Two cases presented in this thesis were performed during innovation contests. An innovation contest is defined by the participation of teams usually trying to solve a technological problem in a defined amount of time. Innovation contests can seem similar from afar, but they are all unique, as they have different purposes, durations, and target audience, among other things. Adamczyk, Bullinger and Möslein (2012) made a categorization of the different elements in innovation contests, shown in Table 1.3.

Table 1.3 Innovation contests categorization
taken from Adamczyk, Bullinger & Möslein (2012)

Attraction (marketing / activation)	Online, offline, mixed
Community functionality	Given, not given
Contest period	Very short term, short term, long term, very long term
Contest phases	One, two, more
Degree of elaboration	Idea, sketch, concept, prototype, solution, evolving
Evaluation	Jury evaluation, peer review, self-assessment, mixed
Facilitation	Professional facilitation, peer facilitation, mixed
Media	Online, offline, mixed
Organizer	Company, public organization, non-profit, individual
Participation as	Individual, team, both
Replication	Biannual, annual, less frequent, more frequent
Reward / motivation	Monetary, non-monetary, mixed
Sponsorship / partnership	Family, friends and colleagues, universities, national associations, specific industries, state and local agencies, mixed
Target group	Specified, unspecified
Task / topic specificity	Open task/low, specific task/high

Several cases have been documented where organizations and companies have used innovation contests to obtain ideas and develop new products with actors outside their boundaries. A sample of cases found in the literature can be seen in Table 1.4.

Table 1.4 Sample of innovation contests found in the literature

Initiative	Description
IBM Innovation Jam	IBM used an internal application to bring together employees around to world to generate ideas for new business units. Participants are encouraged to comment on the ideas of others and a jury selects the best ideas to be then implemented in the company. (Bjelland and Wood 2008)
IdeasProject by Nokia	IdeasProject was the “first external idea crowdsourcing” effort by Nokia to obtain ideas from clients, developers and just about anyone in the crowd. They used text-mining, clustering and regression analysis to study the data and made an internal report to use as creative input. (Vuori, 2012)
Innocentive	Open innovation site where individuals or organizations publish challenges and offer a cash prize for the winning participant (Wagner & Jiang, 2012).
Lego Mindstorms	Lego deployed a “virtual product design space” (Majchrzak & Malhotra, 2013) for users to create their own design. Lego selects the winners, awards a prize but keeps all intellectual property.
My Starbucks Idea	Starbucks set up a website as a way for consumers to propose new products, experiences and actions. Just as Lego, they keep the intellectual property, however no rewards are given (Rosen, 2011).
Netflix Prize	Netflix invited teams of programmers to come up with a better recommendation algorithm, the teams could see the leaderboard (but not the actual codes from other teams), and the winning team got $1 million USD (Rosen, 2011).

The cases documented in the literature have gaps in information use that we will attempt to resolve in this thesis through the exploration of data from a KDD process. First, in most cases the participants receive little or no information about the knowledge domain; they rely on participants’ own knowledge and experiences to generate ideas. In this respect, the ideas could fall short in the novelty spectrum, as participants tend to resort to known problems and solutions, and combine common ideas.

A second limitation is that participants are experts in the domain, for example in the Netflix and the Innocentive cases, and can therefore be fixated to domain or industry paradigms. By involving students with technical knowledge, and prompting them to combine it with data from the application domain, the novelty and diversity of ideas can potentially be increased.

1.1.5 Data mining tools and techniques

Data mining is the application of software algorithms to a set of data to find correlations, trends and other patterns in data, such as regressions. Data mining tools are the software applications that enable the processing of data and application of these algorithms. The extraction of these patterns and trends in the data allow us to see new connections, new perspectives, easily re-organize the ideas, assess and preserve them.

1.1.6 Knowledge discovery from databases

It has always been important for companies to extract information from data, be it from within the organization, or outside data (patents, scientific articles, social media posts and content). It is possible to distinguish three levels of information in organizations (Ackoff, 1989):

- The data which represent facts and is often quantitative
- The information as data aggregates. These are built according to rules and require human intermediation (or at least a consensus as to their meaning)
- The knowledge perceived as high-value information and requires human expertise

The purpose of analyzing the data is to have better information that leads to better informed decision making in all aspects of a business. Software tools can apply algorithms to large sets of data to find relevant trends and patterns (Fayyad, Piatetsky-Shapiro & Smyth, 1996).

Knowledge discovery is an interdisciplinary area that focuses on methods and techniques for extracting useful knowledge by analyzing large sets of data (Fayyad, Piatetsky-Shapiro & Smyth, 1996). Its purpose is the conversion of low-level data, which is normally too voluminous to be explored and analyzed manually, into a more compact, abstract or useful format (Fayyad, Piatetsky-Shapiro & Smyth, 1996).

Figure 1.1 shows the steps involved in a KDD process, where it can be observed that data mining is one of the steps of KDD. In the context of a KDD process, data mining is bound to the application of data analysis and discovery algorithms to the data with the objective of extracting a pattern.

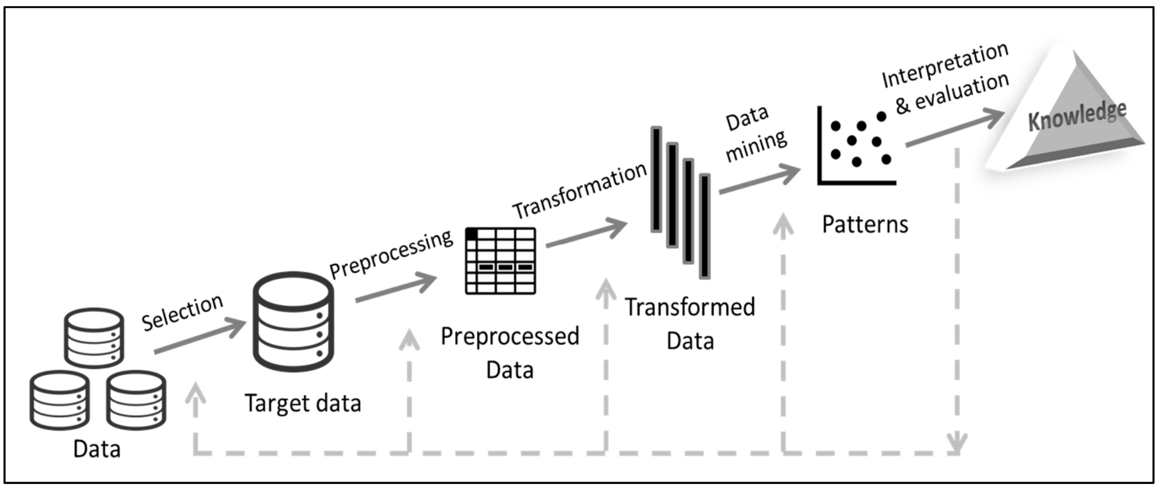


Figure 1.1 Knowledge discovery from databases
taken from Fayyad, Piatetsky-Shapiro & Smyth (1996), Baesens (2014)

Closely related to KDD is the concept of big data. Big data is defined by the availability of large quantities of data where traditional methods and algorithms are not applicable, and new approaches are required to process the data. Some authors refer to the three V's, volume, velocity and variety (Gartner's Laney, 2001, in Kabir & Carayannis 2013) as the criteria for big data.

On the other hand, Howkins (2002) explains that the criteria for big data is not only the size, but the variety in the data, the potential relationships between the data and the need for new tools to be able to exploit the data (see also Maniyka et al., 2001 cited in Kabir & Carayannis 2013).

1.1.6.1 TKM's IPMetrix (data mining tool)

For the three cases presented in this thesis, the author used the software IPMetrix by French company TKM to perform the KDD steps. TKM is a consulting and information services company; their expertise is the exploitation of scientific data sources such as patents, scientific publications and project reports to map collaborations, patent filings and patent evolutions. It is because of the automated analysis of scientific documents and visualizations that this software was chosen to be used for the cases presented here. The company was not involved in the preparation of the data, or the cases themselves.

For the semantic analysis visualization on the IPMetrix tool, TKM uses TF-IDF as a base to determine the most important terms in documents uploaded to the database (meeting with Florian Carichon-TKM, March 2017). TF-IDF stands for term frequency-inverse document frequency. TF is the number of occurrences of a term in a document (Manning et al., 2009), and IDF is a measure used to minimize the effects of the terms that occur frequently in a collection of documents, but do not add value in determining relevance (Manning et al., 2009). The combination of TF and IDF result in a total weight for each term in every document, the equation for the TF-IDF is shown below.

$$TF\text{-}IDF_{t,d} = TF_{t,d} \times IDF_t \tag{2.1}$$

The formula for IDF is shown below, where N is the total number of documents in a collection, and document frequency (df) is the number of documents in the collection that contain a term t.

$$IDF_t = log\frac{N}{df_t} \tag{2.2}$$

According to Manning et al. (2009), TF-IDF has the following characteristics:

- A term obtains high value when it occurs multiple times in a small group of documents.
- A term obtains lower value when it appears a lower amount of times in a document or it appears in many documents.
- A term obtains the lowest value when it appears in all documents constantly.

Figure 1.2, Figure 1.3 and Figure 1.4 present different visualizations that can be extracted from the IPMetrix tool to have a general overview or explore data in detail in a particular domain of knowledge.

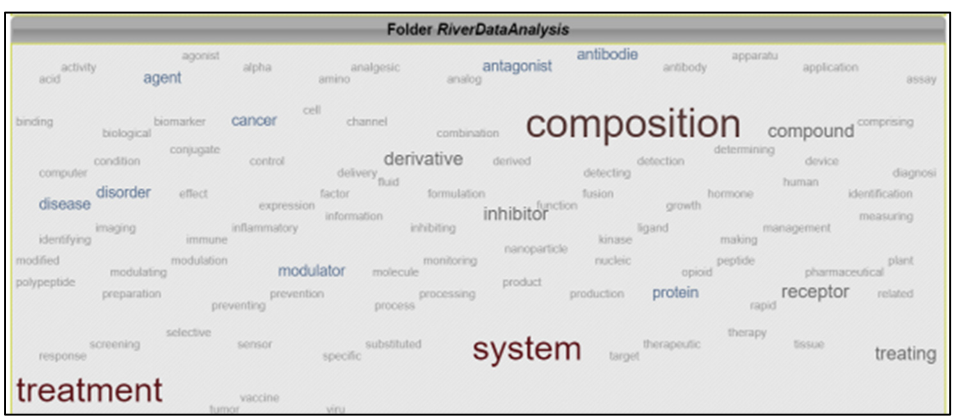
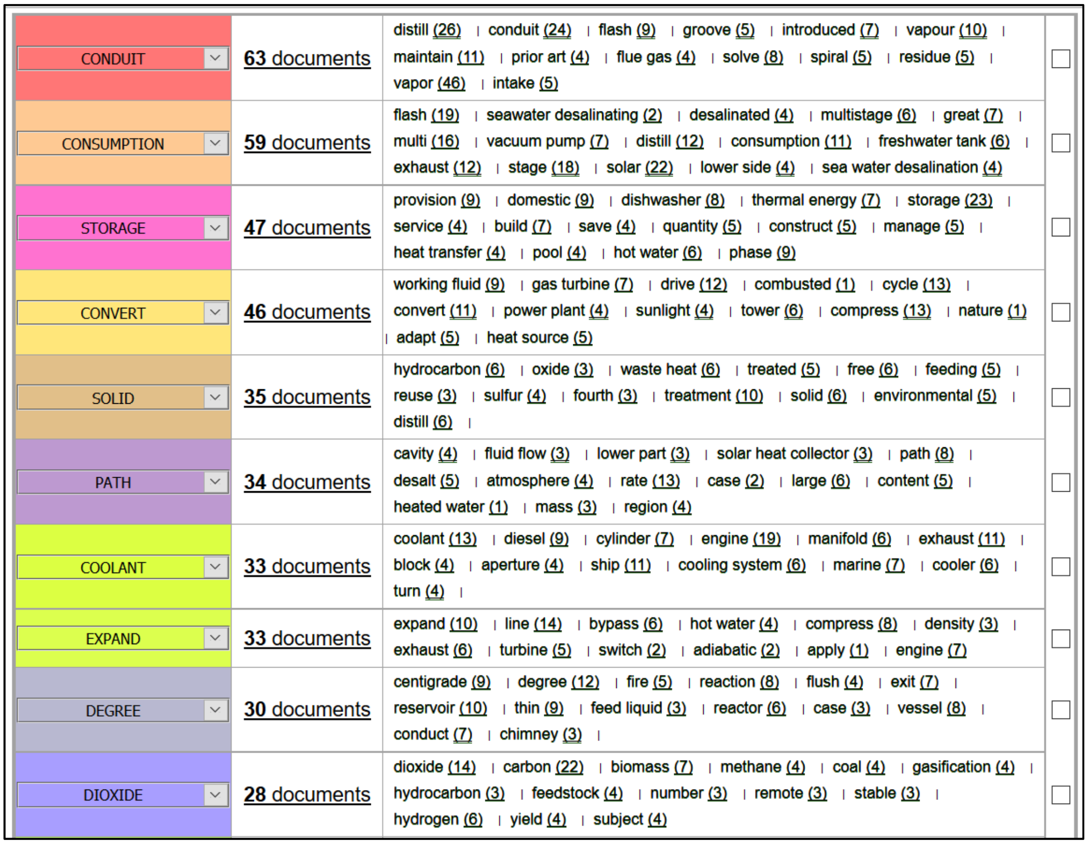


Figure 1.2 Word cloud visualization in IPMetrix



CONDUIT	63 documents	distill (26) | conduit (24) | flash (9) | groove (5) | introduced (7) | vapour (10) | maintain (11) | prior art (4) | flue gas (4) | solve (8) | spiral (5) | residue (5) | vapor (46) | intake (5)
CONSUMPTION	59 documents	flash (19) | seawater desalinating (2) | desalinated (4) | multistage (6) | great (7) | multi (16) | vacuum pump (7) | distill (12) | consumption (11) | freshwater tank (6) | exhaust (12) | stage (18) | solar (22) | lower side (4) | sea water desalination (4)
STORAGE	47 documents	provision (9) | domestic (9) | dishwasher (8) | thermal energy (7) | storage (23) | service (4) | build (7) | save (4) | quantity (5) | construct (5) | manage (5) | heat transfer (4) | pool (4) | hot water (6) | phase (9)
CONVERT	46 documents	working fluid (9) | gas turbine (7) | drive (12) | combusted (1) | cycle (13) | convert (11) | power plant (4) | sunlight (4) | tower (6) | compress (13) | nature (1) | adapt (5) | heat source (5)
SOLID	35 documents	hydrocarbon (6) | oxide (3) | waste heat (6) | treated (5) | free (6) | feeding (5) | reuse (3) | sulfur (4) | fourth (3) | treatment (10) | solid (6) | environmental (5) | distill (6)
PATH	34 documents	cavity (4) | fluid flow (3) | lower part (3) | solar heat collector (3) | path (8) | desalt (5) | atmosphere (4) | rate (13) | case (2) | large (6) | content (5) | heated water (1) | mass (3) | region (4)
COOLANT	33 documents	coolant (13) | diesel (9) | cylinder (7) | engine (19) | manifold (6) | exhaust (11) | block (4) | aperture (4) | ship (11) | cooling system (6) | marine (7) | cooler (6) | turn (4)
EXPAND	33 documents	expand (10) | line (14) | bypass (6) | hot water (4) | compress (8) | density (3) | exhaust (6) | turbine (5) | switch (2) | adiabatic (2) | apply (1) | engine (7)
DEGREE	30 documents	centigrade (9) | degree (12) | fire (5) | reaction (8) | flush (4) | exit (7) | reservoir (10) | thin (9) | feed liquid (3) | reactor (6) | case (3) | vessel (8) | conduct (7) | chimney (3)
DIOXIDE	28 documents	dioxide (14) | carbon (22) | biomass (7) | methane (4) | coal (4) | gasification (4) | hydrocarbon (3) | feedstock (4) | number (3) | remote (3) | stable (3) | hydrogen (6) | yield (4) | subject (4)

Figure 1.3 Semantic analysis from the Voronoi diagram visualization in IPMetrix

To generate a Voronoi diagram, which is a form of visualization where term clusters are arranged in partitions in a plane, the terms in the documents are mapped into vectors. The vectors are then clustered together according to similarity. IPMetrix follows these steps for semantic clustering:

1) Extraction of words and expressions from all the documents uploaded to the database.
2) Lemmatization and stemming of the different keywords and expressions to group words into lexical families.
3) Assembly of a "stop words" list to filter common keywords and expressions.
4) Creation of vectors to represent the different documents.

5) Application of K-means method to create clusters.
6) Valuation of TF-IDF of clusters to determine the representative keywords/expressions.
7) Display the 20 keywords or expressions with the largest TF-IDF ratio.

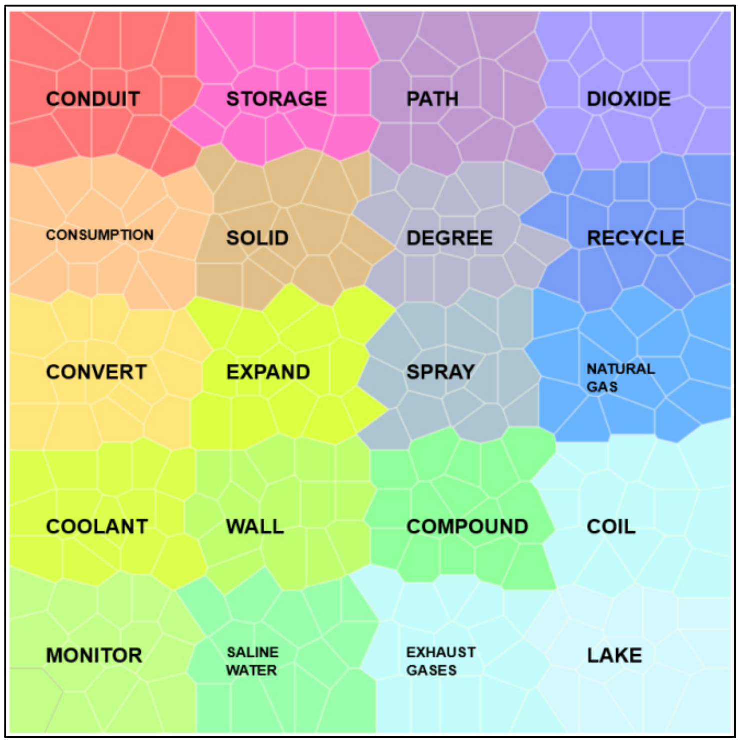


Figure 1.4 Example of the Voronoi diagram in IPMetrix

1.1.7 Use of data mining for idea generation

Whenever a new technology is made available, it is worth finding possible uses in different domains. Howkins (2002) urges companies to use data in a more creative fashion because it is a resource that can be reused and analyzed in different ways to find new insights. While the task of mapping a domain of knowledge and visualize connections between concepts has been facilitated by technologies of data mining and visualization, we have yet to develop technologies that generate new and innovative ideas. Some authors have already attempted to use data as an input for idea generation; Table 1.5 shows a sample of interesting previous studies where data is used to support idea generation and creativity:

Table 1.5 Previous studies where data is used to support idea generation and creativity

Authors	Type of study	Creativity support	Objective	Findings
Maccrimmon & Wagner, 1991	IS design	Data prompting	Support the generation of alternative ideas using data	Free-form techniques help generate ideas, but users need more stimulating techniques to continue being productive.
Hamman, 2000	IS design	Cues and suggestions	Propose the use of an IS to support the creativity of music composers	Algorithms can be used to propose combinations to inspire composers. Software and visualizations can support creation.
Müller et al. 2012	IS design	Data attributes and visualizations	Create a software to support identification of unexplored biomedicine research areas	The software can help researchers look at data in new ways to help generate hypotheses.
Chen, Li & Hung, 2013	Case study	Data correlation report for interest analysis	Use of big data techniques to analyze the results of crowdsourced idea generation	Organizers of large scale idea generation can benefit from DM to assess results.
Shan, Zhu & Zhao, 2013	IS design	Idea network exploration	Support brainstorming by recommending computer generated ideas using idea networks	Participants perceived that image suggestions were useful to generate ideas, but encyclopedia entries were not as useful.
Dove & Jones, 2014	Case study	Data visualization	Explore the use of aggregated data to support idea generation in workshops	Use of data supported collaboration and engagement, helped participants build upon their knowledge. Idea novelty not as expected.

Data mining technologies are about finding similarities, trends and correlations, and it is up to people to evaluate the results and gain insights. However, if we follow Koestler's theory that there is value in connecting what is separated or incompatible, we must then search for the disconnections.

Shan, Zhu and Zhao (2013) believe that using data in the idea generation process can help participants get "unstuck", and that by using data and information exploration there is enormous potential for insight discovery. Dove and Jones (2014) propose that data can be useful to aid in the idea generation process, especially when there is no predefined outcome. From the process by Shneiderman et al. (2006) and Atman et al. (2007), we believe data can be used in four moments of the collaborative idea generation process for engineering design:

- **Need identification / problem definition.** Data mining permits the analysis of data in a way that was not possible before, by bringing together different sources of information and finding trends that are only visible with large amounts of data. This will make it easier to visualize the gaps in a domain (Müller et al., 2012).
- **Information gathering.** Data from different sources can be mined and used as input for information gathering, increasing the external stimuli for teams developing an engineering solution (Dove & Jones, 2014).
- **Idea generation.** Data mining tools and techniques can be used to identify which ideas are not being connected, but are already in the knowledge base of participants or in the domain. The purpose would be to enable bisociation, to connect two frames of reference previously considered to be incompatible (Koestler, 1964; Nielsen, 2012).
- **Evaluation.** A wealth of information is generated throughout the development of a concept or solution; unfortunately, only a few of the ideas are developed, and the rest are discarded. Data can help identify interesting concepts and keep the data for future developments (Chen, Li & Hung, 2013).

1.1.7.1 Bisociative networks

Some authors have already undertaken the task to automate bisociation in what they call "bisociative networks". Proponents of bisociative networks have suggested three types of networks can support bisociation (Dubitzky et al., 2012):

- Bridging concepts (one concept links two graphs or clusters)
- Bridging graphs (a graph links two other graphs or clusters)
- Structural similarities (two graphs have the same shape)

However, in their proposition, the links they find in the networks are between already connected elements. To inspire participants in an idea generation session to combine elements that are distant or disconnected, we need to find them.

1.2 Methodology

The second section of this chapter is concerned with the methodological framework guiding this work. An iterative methodology was selected, as it facilitates increasing the understanding of the issues and a cyclical improvement of the use of data mining tools and data as a creative input.

Design Science Research was selected because of the practical approach of analyzing the current state of the application domain, the iterative process to improve the design of a device or process, and finally the grounding of the findings into new applicable knowledge.

1.2.1 Design Science Research framework

The researcher selected the Design Science Research (Hevner, 2007) approach because it enables a research based on the study of the current environment. This will provide an understanding of how the engineering design process functions today and where in the process the participants are able to benefit more from the input of information.

Once the process has been designed, it is implemented and evaluated, giving basis for further improvements, and serving as foundation to build a knowledge base for new theories and methods. The research methodology allows iterating solutions progressively, improving the process designed at each stage of the research (Figure 1.5).

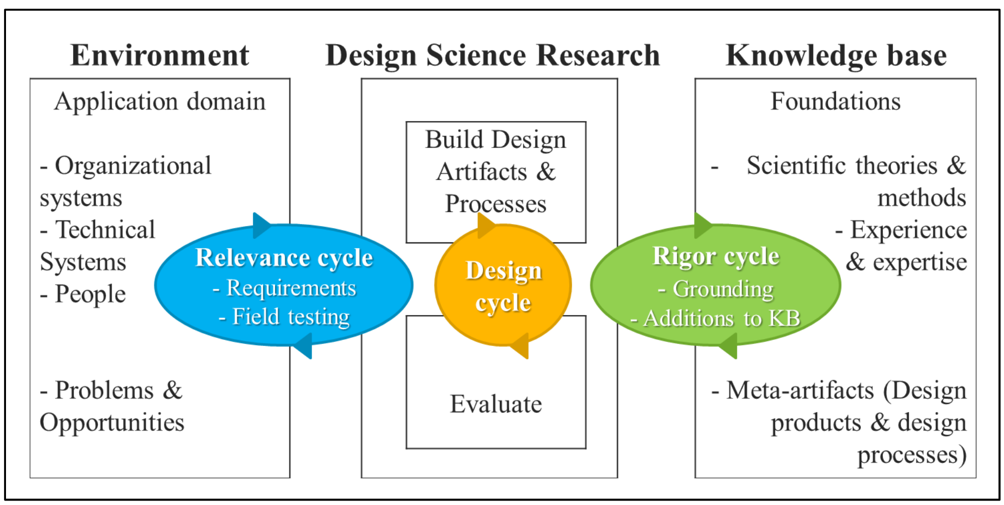


Figure 1.5 A Three Cycle View of Design Science Research
Taken from Hevner (2007)

Following the Design Science Research methodology, each chapter will present an evaluation of the results and the implications for the research, in the final section called “Design cycle evaluation”.

1.2.2 Evaluation

The creative process is measured by different metrics depending on the authors; they propose to evaluate idea generation by the process or by the outcome, using the criteria of quality, quantity, variety and novelty. Table 1.6 summarizes the metrics found in the literature along with a brief definition.

Table 1.6 Summary of metrics to evaluate the creative process

Metric	Definition
Applicability	Measures how much the concept complies with the pre-defined needs, and includes the viability of the concept in the context. (Ardaiz-Villanueva et al., 2011)
Conclusion characters	The amount of characters in the developed concept. (Munemori & Nagasawa, 1996)
Chats	Number of communications between participants in a session or project. (Munemori & Nagasawa, 1996, Yuizono et al., 2005)
Comments	Number of feedbacks received by an idea shared by a participant. (Ardaiz-Villanueva et al., 2011)
Complexity	Refers to the participants taking initiative and dividing the problem into sub-systems for further development. (Ardaiz-Villanueva et al., 2011)
Ideas evaluated	Number of ideas evaluated by other participants, indicates external interest on the idea. (Ardaiz-Villanueva et al., 2011)
Ideas shared	Number of ideas shared by the participants with others. (Graetz et al., 1997)
Level of detail	Level of detail for the concept provided by the participant. (Wodehouse & Ion, 2012)
Novelty	Degree inventiveness, measured by the principles used in the solution. (Ardaiz-Villanueva et al., 2011, Glier et al., 2011, Wodehouse & Ion, 2012)
Number of ideas	Amount of ideas produced by the participants for a session or project. (Ardaiz-Villanueva et al., 2011, Glier et al., 2011, Graetz et al., 1997, Jung, Schneider & Valacich, 2010, Munemori & Nagasawa, 1996, Parjanen, Hennala & Konsti-Laakso, 2012, Wang & Ohsawa, 2013, Wodehouse & Ion, 2012)
Participants	Number of participants involved in the creative process. (Yuizono et al., 2005)

Table 1.6 Summary of metrics to evaluate the creative process (continued)

Metric	Definition
Perceived team cohesiveness / effort	Indicates how much participants felt integrated as a team, and how much effort they perceived from their teammates. (Graetz et al., 1997)
Quality of concepts / ideas accepted	Degree to which the concept responds to the needs of the problem or established filters. (Glier et al., 2011, Jung, Schneider & Valacich, 2010, Wang & Ohsawa, 2013, Wodehouse & Ion, 2012) *Note: The author believes this is actually 'Applicability'
Cards / Notes	Measure of the process by the amount of individual contributions. It is not the same as an "idea", as one record can contain multiple ideas, one concept formed by several ideas, or just a principle with no grounded idea. (Gumienny et al., 2013, Yuizono et al., 2005)
Time	Different authors measure time according to their particular focus, for example: time to reach a conclusion, time to generate ideas, time to develop ideas, time to make a decision. (Graetz et al., 1997, Gumienny et al., 2013, Munemori & Nagasawa, 1996, Yuizono et al., 2005)
Variety	If the ideas produced are clustered by principle, variety measures the different categories. (Glier et al., 2011, Wodehouse & Ion, 2012)
Whiteboard events	Number of times the participants went into the system to collaborate or provide ideas. (Gumienny et al., 2013)

To the knowledge of the author, there is currently no method to objectively measure the value of an idea, therefore the results cannot be evaluated based on the result of the EDP. It is also assumed that the concepts will be applicable to the problem at hand. Consequently, the focus will be on the four metrics in the present work:

- Number of ideas (productivity)
- Complexity (sophistication)
- Variety of ideas
- Novelty

1.2.3 Expected results

Harnessing data at different stages of the EDP, idea generation could see an improvement in the complexity and variety of the resulting concepts. Complexity refers to the level of subsystems considered in the study of the problem (Ardaiz-Villanueva et al., 2011), while the variety measures the number of categories in which the solutions could be divided (Glier et al. 2011, Wodehouse & Ion 2012).

An increase in both this metrics would suggest that the participants were able to look for different types of solutions and did not gravitate towards known solutions. It is expected that the number of ideas will not increase significantly, or will even decrease, given that the creativity exercise is not to diverge in a brainstorm but to try to find a new way to connect the disconnected elements. However, we expect that sessions where participants are given concepts to blend will result in a greater variety of technical solutions and with increased novelty compared to teams who do not have this support.

1.2.4 Stages

To be able to achieve the necessary knowledge to design the process and evaluation procedure to include the use of data from KDD in the context of an engineering design, the following stages are necessary:

1) Literature review
2) Develop protocol for data mining tool use in EDP
3) Design evaluation tool to specific conditions of EDP
4) Gather data for incumbent domains (KDD)
5) Implement protocol in EDP with students

6) Analyze results

1.3 Exploratory studies

To determine the environment and current issues with the EDP followed by engineering teams, two exploratory cases were performed to observe areas of opportunity to support teams in the early stages, particularly for idea generation. The cases documented in this provided a base to theorize the proposed process to follow in subsequent cases.

1.3.1 Outdoor lighting company

The first experience using big data analytics as input for creativity was during a 6 hour ideation session with 2 teams of 6 participants in an outdoor lighting company on January 23, 2015. The purpose was to generate ideas for new research projects.

The data mined were concepts from global design contest. Teams were asked to make bisociations with concepts to refine the proposed ideas. The process for the session was the following:

- Preparation - Concepts mined from global design contest (before session)
- Divergence
- Convergence
- Teams were asked to make bisociations with the concepts to refine the proposed ideas.
- Participants selected the concepts to bisociate

The result observed was that participants selected concepts already in their ideas (for an example, see Figure 1.6).

Figure 1.6 Photograph of work sheets from teams

Following the Design Science Research approach, the researcher learned that for the following idea generation sessions, it is necessary to have a moderator or guide in each group to motivate participants to make unexpected combinations, and not select the concepts which only serve to reinforce their current propositions.

1.3.2 Summer school on innovation and technological design 2015

An opportunity to explore the capabilities of data mining and exploration in the context of idea generation was implemented during the 2015 ETS International Summer School on innovation and technological design. The challenges were initially presented in the innovation contest “Les 24 heures de l’innovation”, and were then retaken by the Summer school participants. The researcher used a data mining tool (not TKM’s IPMetrix) to provide students with access to data related to their projects.

In this preliminary case, 21 students worked on three challenges, one related to solving a parking issue, a second on the development of an automated garbage collection vehicle, and a third one on the automation of a warehouse with robots.

Two efforts were made in using data mining tools to support the engineering design process of teams, as they had the mandate to develop a prototype. For the first attempt, the researched utilized the software tool from a company based in Montreal, which required manual tagging of patents to identify relevant information. It also did not have automatic document import directly from patent database providers, thus the researcher had to download patents one by one, a time-consuming process.

The second attempt was the use of data from social media. Another local company from Montreal that specializes in the analysis of social media posts offered to make an extraction of publications related to the issues being tackled. Each team received a report to be analyzed by themselves.

The first tool proved too inefficient for the researcher to upload large amounts of documents and perform the manual tagging, and complicated for participants to explore the data. With the second tool, the data was only useful to support their arguments in presenting their solutions.

A second issue, unrelated to the data, was the ownership of the problems being solved. The students in this case had clients who dictated the expected outcome, and were bound by those constraints. For the subsequent cases, several desirable conditions were defined:

- Access to a data mining tool with more upload capabilities
- Access to a data mining tool with better ease of use for final users
- Open problems with no external clients

1.4 Design cycle evaluation

The exploratory studies presented at the end of this chapter were a building stone for the subsequent cases. After the execution of these cases, the author was able to determine the characteristics needed for the cases to be able to better implement the utilization of the KDD in an ED process.

The opportunity for such cases was discovered with the AquaHacking competition, by the *de Gaspé Beaubien* Foundation. This competition, open to everyone, aims for teams to propose innovative solutions for the conservation of bodies of water, and awards the best initiatives to support the development of start-ups.

The following three chapters (Chapters 3, 4 and 5) each present a case where the EDP was supported by the use of KDD. Each case presents an increase in the scope of the EDP and the involvement of participants in the KDD.

CHAPTER 2

BIG DATA ANALYTICS AS INPUT FOR PROBLEM DEFINITION AND IDEA GENERATION IN TECHNOLOGICAL DESIGN

Ma-Lorena Escandón-Quintanilla[1], Mickaël Gardoni[1,2], and Patrick Cohendet[3]

[1] École de Technologie Supérieure, Montréal, Canada

[2] Institut National des Sciences Appliquées, Strasbourg, France

[3] HEC Montreal, Montreal, Canada

Published in: PLM 2016, IFIP AICT 492 proceedings

This article documents the first case where a design session was held to identify problems related to the theme of a start-up competition. Participants were self-selected, as they responded to an open call published on the school weekly newsletter. To generate ideas, teams first worked by themselves, and after were provided with access to a data exploration tool to find hints for new areas of opportunity. The paper discusses the observed results and future work.

This paper was presented at the IFIP International Conference on Product Lifecycle Management PLM16 in Columbia, South Carolina, USA on July, 2016.

Abstract

Big data analytics enables organizations to process massive amounts of data in shorter amounts of time and with more understanding than ever before. Many uses have been found to take advantage of this tools and techniques, especially for decision making. However, little applications have been found in the first stages of innovation, namely problem definition and idea generation. This paper discusses how big data analytics can be utilized in those stages. It

includes an example of application in problem definition and proposes a case study implementation in a higher education setting for idea generation.

2.1 Introduction

The current economy's fast-paced product development cycle has lead companies to decrease the time in all stages of new product development. Even before this change, companies spent proportionally little time in the idea generation process, compared to the time spent in technical development and testing. Little by little, companies are realizing the need for and the power of good ideas, thus requesting employees to dedicate more time and resources to the first stages of the new product development process, namely the identification of the opportunity or problem statement, information gathering, and the idea generation.

To create new ideas, the individual must form new combinations of knowledge he or she already possesses (Fleming & Szigety, 2006, Koestler, 1964). However, it has been found that that participants will gravitate towards known solutions (Howard, Culley & Dekoninck, 2006) and that popular ideas are constantly recombined (Fleming & Szigety, 2006). To produce a radical result, the ideator needs to make highly varying ("wild") combinations (Fleming & Szigety, 2006). It is necessary to find ways to promote wild combinations.

In previous literature, authors have discussed options to manage ideas in a product development process, designing collaboration platforms and software to facilitate the documentation and exchange of ideas. But with new information technologies, it is possible to benefit from the wealth of data we are able to collect and process. Data can enable organizations to find insights related to their processes, clients and market.

This article discusses the use of big data analytics for problem definition and idea generation. It includes a case where big data analytics was used to identify problems and a proposed use of readily available analytics tools to facilitate idea generation.

2.2 Idea generation sessions

Idea generation is the fundamental step of the innovation process and, more importantly, it "is central to engineering design" (Glier et al. 2011). Participants from different domains or areas of expertise can work together during idea generation (ideation) sessions, to exchange and create knowledge, usually for a specific aim.

The purpose of ideation sessions is to set an environment and implement creativity techniques that will help participants produce, express and combine ideas. Another advantage of idea generation sessions is that the ideas of others sometimes trigger the creation of related or new ideas (Munemori, Yoshino & Yunokuchi 2001).

Ideation sessions are an interesting example to explore creativity support systems because of their unique characteristics: a defined purpose, limited time, multidisciplinary teams and willingness to create knowledge (Jiménez-Narvaez, Desrosiers & Gardoni, 2011).

While there is not one generally agreed process for idea generation sessions, Shneiderman et al. (2006) propose the following phases, found in recent literature and commonly accepted for new product development cycles (Figure 2.1)

There are many areas of opportunity to improve for the process of idea generation: sharing more ideas, providing feedback and decreasing the time it takes for the team to develop ideas into concepts.

Based on the process for idea generation sessions by Shneiderman et al. (2006) and the examples found on extant literature, we categorized the use of information and identified how big data analytics can be used tool to help teams. It can be used in four phases of the process: to identify areas of opportunity (need identification), as input for inspiration (information gathering), to identify unrelated ideas to combine in new concepts, and to obtain insight from a large amount of ideas from a crowdsourcing effort (evaluation). For this work, the focus lays on the first stages, highlighted in Figure 2.1.

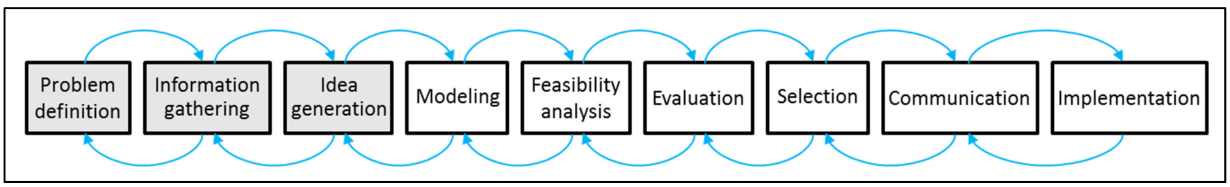


Figure 2.1 Process for idea generation sessions, based on Shneiderman et al. (2006)

2.3 Big data analytics

People collaborate in many different ways, by sending emails with attachments, by sharing documents on the cloud, talking over the phone, exchanging messages. Information systems allow for those communications to occur, and to document the exchanges. All the data generated and collected in an organization is a source of untapped knowledge that can lead to inventive designs of new products and services if analyzed using powerful tools.

Big data is characterized not only by the speed of generation (velocity), but also the different types of data that must be analyzed (variety) and the massive amount of data being collected (volume) (Gartner's Laney, 2001, in Kabir & Carayannis, 2013). To those characteristics, more recent authors have appended the dimensions of veracity (Koutroumpis & Leiponen, 2013), meaning how reliable information is, and value (Koutroumpis & Leiponen, 2013), which considers the impact the data can have on the organization when analyzed.

Big data analytics enables organizations to analyze their data in a way that was not possible before, by bringing together different sources of information and finding trends that are only visible with large amounts of data. This will make it easier to visualize the gaps in a domain (Müller et al. 2012). The use of big data analytics will depend on the availability of the tools required to perform the analysis, and the characteristics (e.g. duration, number of participants, access to external sources of data) and aim of the idea generation session.

2.3.1 Problem definition / need identification

Müller et al. (2012) created a software to support the identification of unexplored research areas through data attributes and visualizations. They propose that information (data) can be used to guide researchers to new unexplored paths. They theorize that data can be examined iteratively for "divergent and convergent thinking" to generate new hypotheses (Müller et al. 2012).

In this same spirit, data from various sources can be collected and exploited to find areas of opportunity for an organization. It is possible to find new applications or markets for the products and services, or even expertise already possessed. For researchers, it can signal new areas to explore. For artists and creators, it can find previously unthinkable combinations. Figure 2.2 depicts the flow of information to use big data analytics for problem or need identification.

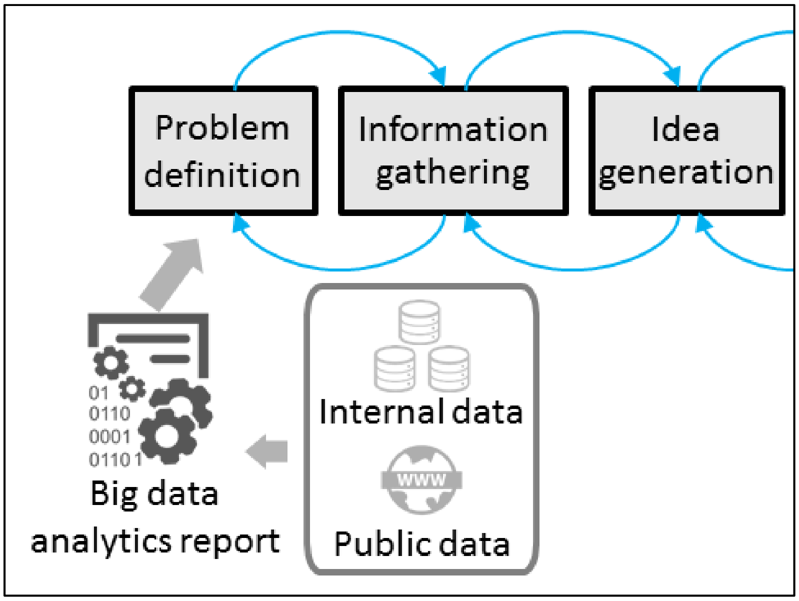


Figure 2.2 Flow of information to use big data analytics
for problem or need identification

2.3.2 Idea generation

Information inputs can help bolster the creativity of participants to generate ideas given that "creative thinking involves a process of iterative activation of 'cues'" (Hamman, 2000); furthermore, the likelihood of creating new knowledge from recombination is greater as we increment the number of external inspirations (Cohen & Levinthal, 1989, in Kabir & Carayannis, 2013). Several works discuss the use of information as input for creativity:

- In (Hamman, 2000) to support music composers through cues and suggestions.
- In (Shan, Zhu & Zhao, 2013) to support brainstorming by recommending computer generated "ideas" (extractions from a three databases).
- In (Maccrimmon & Wagner, 1991) to support the generation of alternative ideas using data prompting.

- In (Dove and Jones, 2014) to complement the idea generation process by using aggregated data.

The examples listed demonstrate that there is an interest to enhance idea generation through the use of information. However the risk is that the material selected to form the knowledge base will already be biased towards a known solution. By using big data analytics, the information will reveal trends and connections that were previously unseen. This effect can potentially be amplified when extracting date from unrelated or complementary knowledge domains to promote new combinations.

Figure 2.3 depicts the flow of information to use big data analytics for problem or need identification.

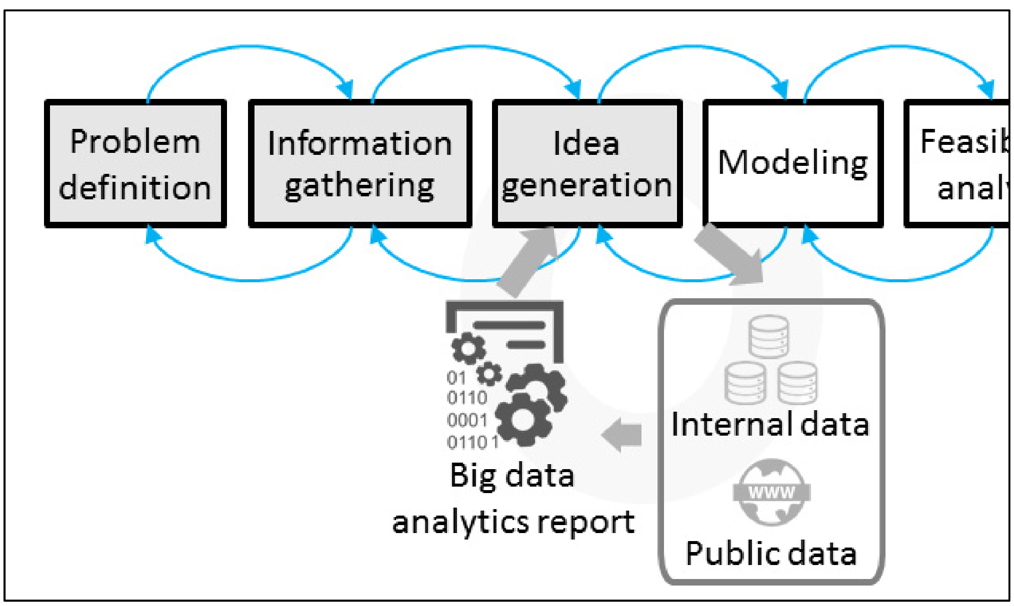


Figure 2.3 Flow of information to use big data analytics as input for idea generation

2.4 Application in a higher education setting

Big data analytics in the context of a complete new product development process can be used as a support for participants to identify problems and generate ideas. To test this hypothesis, the authors designed three case studies to be performed sequentially. This will enable to study the impact of the use throughout the whole creative process. The three case studies will take place in several higher education environments:

- **Problem definition:** big data analytics will be used to define the challenges to be solved in subsequent activities. This case study has been completed and is presented in section 2.4.2.
- **Information gathering:** the authors will build a knowledge base to be provided to participants of an innovation competition. This proposed case study is presented in section 2.4.3.
- **Idea generation:** during a month-long intensive master course on innovation, the authors will provide students with access to big data analytics tools. This proposed case study is presented in section 2.4.4.

2.4.1 Evaluation criteria

The creative process is measured by different metrics depending on the authors, for example:

- Applicability of concepts (Ardaiz-Villanueva et al., 2011)
- Complexity level of concepts (Ardaiz-Villanueva et al., 2011)
- Detail of concepts (Wodehouse & Ion, 2012)
- Novelty of concepts (Ardaiz-Villanueva et al., 2011, Glier et al., 2011, Wodehouse & Ion, 2012)

- Number of characters of a conclusion (Munemori & Nagasawa, 1996)
- Number of chats (Munemori & Nagasawa, 1996, Yuizono et al., 2005)
- Number of comments (Ardaiz-Villanueva et al., 2011)
- Number of ideas (Ardaiz-Villanueva et al., 2011, Glier et al., 2011, Graetz et al., 1997, Jung, Schneider & Valacich, 2010, Munemori & Nagasawa, 1996, Parjanen, Hennala & Konsti-Laakso, 2012, Wang & Ohsawa, 2013, Wodehouse & Ion, 2012)
- Number of ideas evaluated (Ardaiz-Villanueva et al., 2011)
- Number of ideas shared (Graetz et al., 1997)
- Number of participants (Yuizono et al., 2005)
- Number of record cards / sticky notes (Gumienny et al., 2013, Yuizono et al., 2005)
- Number of whiteboard events (Gumienny et al., 2013)
- Perceived team cohesiveness / effort (Graetz et al., 1997)
- Quality of concepts / Ideas accepted (Glier et al., 2011, Jung, Schneider & Valacich, 2010, Wang & Ohsawa, 2013, Wodehouse & Ion, 2012)
- Time (Graetz et al., 1997, Gumienny et al., 2013, Munemori & Nagasawa, 1996, Yuizono et al., 2005)
- Variety of concepts (Glier et al., 2011, Wodehouse & Ion, 2012)

Given that there is currently no method to objectively measure the quality of an idea, this criterion will not be considered. Other metrics, such as the number of characters in a description, do not seem relevant to assess the impact of big data analytics for problem definition, information gathering or idea generation. It is also assumed that the concepts will be applicable to the problem at hand. Consequently, the focus will be on these four metrics: comments (feedback from the participants), complexity, ideas shared, and variety of ideas (to be assessed by domain experts).

We believe that the use of big data analytics as input for creativity will provide participants with hints to novel associations that may result in ideas with greater complexity and varied from current or competing solutions. We expect to obtain positive feedback from participants regarding the use of data as input for the session and as a support to merge concepts and find innovative solutions.

The following sections describe the application performed to support one of the organizations in finding the challenges to propose to an innovation competition, a proposal to apply big data analytics for information gathering for participants of the innovation competition, and the proposed approach to support idea generation for solution design during a summer school.

2.4.2 Problem definition / need identification

In order to apply big data analytics for problem definition, the researchers worked with one of the organizations that will propose challenges for both the competition and the summer school. Their objective is to work with challenges related to river water quality and conservation. Since the problems to be solved were not defined, a creativity session was held to identify areas of opportunity. The session took place on the 30th of March, 2016 at the *École de technologie supérieure* in Montreal. All the community was invited to participate through the weekly bulletin board, 18 participants were registered, and 15 attended the session.

2.4.2.1 Input data

As discussed before, there is an enormous wealth of external and publicly available data that can be utilized. However, there is a difficulty in selecting relevant and valuable data, and cleaning the information to make it usable for the purpose. In this case, because the aim was to identify problems related to rivers, freshwater and water conservation which can potentially be solved by a technological solution or a data analysis solution, the data selected to be used as input are patents. Patents offer the advantage of having pre-defined sections, describing a problem and the solution. To perform the data analysis, patents from Patbase which include keywords such as "freshwater" and "data analysis" + "river" were extracted.

2.4.2.2 Work session

It is important to set objectives and to provide participants with a sense of progress. To ensure the achievement of the purpose of the work session, a series of activities were planned for participants to follow (Table 2.1).

Table 2.1 Activities followed during the problem definition session

Activity	Time allocated
Welcome / Introduction to the topic	20 minutes
Group formation / Personal introduction	10 minutes
Identification of elements in the problem	30 minutes
Identify connections and relationships between elements	30 minutes
Identify key issues	30 minutes
Use of big data analytics (identification of new issues)	30 minutes
Presentation of issues identified	30 minutes

After a brief welcome and explanation of the purpose of the session, participants were first asked to identify all the elements of the problem (stakeholders, inputs, outputs). The second step was to relate all the elements and identify which cause problems. The boards pictured in Figure 2.4 demonstrate the different approaches of teams to identify the elements of river issues. Each group of participants proceeded to identify key issues (examples can be seen in Figure 2.5). The purpose of this activity was to clearly state the key issues.

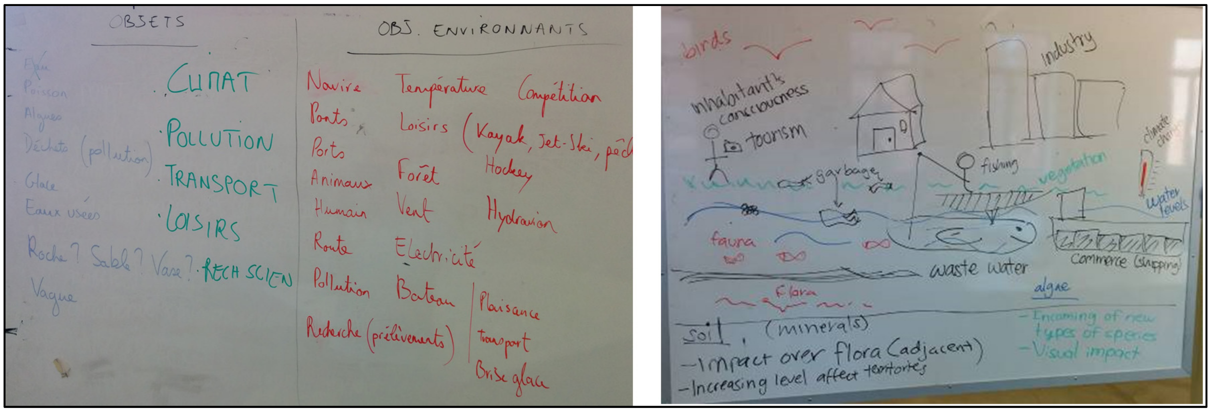


Figure 2.4 Teams list the elements of the issue

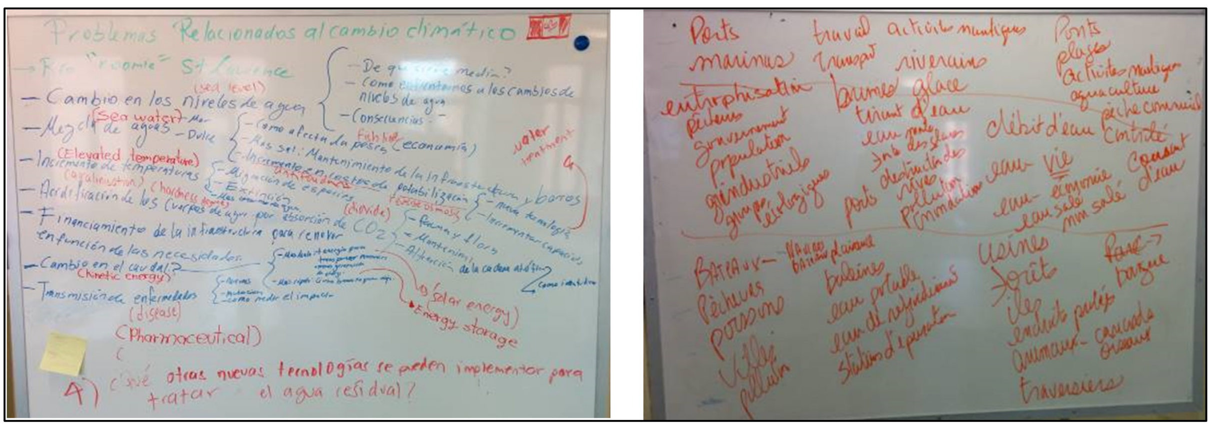


Figure 2.5 Teams identify key issues

For the following phase of the work session, the teams were provided with access to a big data analytics analysis tools pre-loaded with freshwater and river related patents. The software used to analyze the data is IPMetrix (IP Metrix Solution - TKM, 2016), from the company TecKnowMetrix, which provides semantic analysis and cartographies of the information. In this session, the purpose was to use big data analytics as an information input to trigger new relations. Participants had time to explore the different concepts in the visualizations and selected various concepts to combine with their previously identified issues. The objective was to provide participants with new concepts that could work as prompts to open new fields of possibility, by considering the materials, measures, technologies or concepts in the mapped domain. Table 2.2 is a comparison of the results from each group of participants before and after the access to external data:

Table 2.2 Results of issues identified per group

Group	Number of participants	Issues identified originally	Issues identified with support	Total
1	5	5	3	8
2	6	3	1	4
3	4	5	2	7

Participants mentioned that they were able to identify links because of their previous knowledge, reinforcing the notion that the use of data as input can trigger the exploration of different directions.

2.4.3 Information gathering

For the competition, a world-wide event called "*Les 24 heures de l'innovation*", organizations propose a challenge, and participants have 24 hours to work on a solution. At the end of the 24 hours, the best solutions are awarded a prize. The competition takes place in over 40 sites in 20 countries around the world, at the same time. All students will be given access to information gathered to give them insights to the challenges proposed. The main site of the competition is the *École de technologie supérieure* in Montreal.

2.4.3.1 Measuring the impact

The objective of providing participants with data is to improve the novelty and originality of the solutions proposed. To measure the effect, a comparison will be made in the evaluation grid scores for innovativeness given to the winning solutions from this edition (compared to previous editions).

2.4.4 Idea generation

The next ground for experimentation will take place in the month of July, during the “ÉTS Internationals Summer School on innovation and technological design”. In total, fifty engineering students will take part in the course, where they learn about the innovation process, creativity techniques, and work on a team project solving one of the challenges. The objective is to arrive to a functional prototype. Students will be placed in one of the 6 project teams. The teams will select one challenge to solve during the course, and will be guided by professors in the technical side and the creativity and innovation approach.

The students will have the possibility to implement different idea generation tools and techniques. For each, they will have a workshop where they will use the tool or technique and apply it to the problem they are trying to solve. The ultimate goal of providing participants with creative tools and techniques is to arrive to an original solution for the challenge (problem) to be solved.

Additional to the aforementioned tools and techniques, one course will be taught where they will learn to explore data to find hints for solutions.

2.4.4.1 Tracking the results

Because students will employ different tools and techniques, we need to compare the results of the application of each. To do so, students will be required to carry an “idea journal”. In this journal, each group must document the ideas generated during each workshop. Ideas can be documented using brain-maps, lists, drawings, sketches or photographs to represent the work achieved with the tool / technique.

2.5 Discussion and conclusion

Solving problems and creating good ideas for new products, services and technologies are too important to rely only on human capacity to create and collaborate. Great inventions are built in the vast knowledge that was created before us. However, we live in an age where there is too much information for humans to absorb. There is a latent need to make sense of all the data generated every day. In this data therein lay clues for exciting combinations, hints to better solutions.

The purpose of using big data analytics in an idea generation context is precisely that of taking advantage of the wealth of knowledge available through the application of information technologies. The data by itself does not generate value, it is the participants making sense of it and making new connections which can create value.

This paper presented an example of how including data in a problem-identification process can spark new combinations to explore different directions. The next work is expected to provide insights into the use of big data analytics for idea generations with the purpose of designing novel solutions.

2.6 Design cycle evaluation

NOTE: This section does not appear in the article, it is meant as a conclusion and transition for the next chapter of the thesis.

In this case, it was found that teams needed training in the use of the data mining tool to be able to generate different visualizations; it also appears that more time to explore the data would be beneficial to find interesting insights in the data. These findings support the notion found in the literature that the use of information from external sources can provoke new connections in participants in a design effort.

However, direct access to more data for exploration did not prevent participants from selecting keywords that supported previously generated ideas, and novelty was not increased as expected. For the following case, it is proposed to change the timing of the access to the data, to an earlier moment of the EDP.

CHAPTER 3

IMPROVING CONCEPT DEVELOPMENT WITH DATA EXPLORATION IN THE CONTEXT OF AN INNOVATION AND TECHNOLOGICAL DESIGN COURSE

Ma-Lorena Escandón-Quintanilla[1], Mickaël Gardoni[1,2], and Patrick Cohendet[3]

[1] École de Technologie Supérieure, Montréal, Canada

[2] Institut National des Sciences Appliquées, Strasbourg, France

[3] HEC Montreal, Montreal, Canada

International Journal on Interactive Design and Manufacturing (IJIDeM) February, 2017

DOI: 10.1007/s12008-017-0380-5

In previous cases, teams had very little time to execute the KDD process, and explore data to find hints for new solutions in the defined time-frame. In this chapter, we present Case 2, where teams would be trained on the use of the data exploration tool, and would be allotted time to search and upload additional data to improve their knowledge base.

In this case, we followed 8 teams who had the choice to utilize data exploration tools as a support for their idea generation and concept development to solve an engineering challenge.

This article was published in the *International Journal on Interactive Design and Manufacturing (IJIDeM)* on February, 2017.

Abstract

Innovation is about continuously pursuing better, more efficient solutions, and organizations allocate vast amounts of resources to achieve this goal. One challenge is the access to and exploitation of information, as teams attempt to harness existing knowledge to design

solutions efficiently. This article is concerned with two of the earliest stages of the concept development process, information gathering and idea generation. Information gathering and idea generation can be enhanced to find hints for more innovative or diverse concepts for engineering solutions by the use of data mining tools and techniques to exploit patent data. A case is presented where teams of engineering students, in the context of a higher education course for innovation and technological design, explore data from domain specific patents to develop innovative solutions. The findings indicate that the use of data can be advantageous for team creativity, as it helps identify potential solution elements, materials and current technologies.

3.1 Introduction

The job of engineers is to solve complex problems and design new solutions or products which are usually constrained by the possibilities of available materials, budgets, and time. They attempt to foresee market needs, while also following new technologies in other fields which can impact their own industry.

A common challenge for teams responsible of creating new products, services and solutions trying to solve problems when it comes to innovation is to see beyond existing solutions (Agogué et al., 2014), based on their own knowledge and the knowledge that exists within the company (Dodgson, Gann & Salter, 2006). As a wealth of information is generated everyday by companies, users, organizations, and increasingly machines (Dove & Jones, 2014), it becomes more difficult to keep up to date and process all this information within a company. By integrating the use of new data mining technologies, which make it possible to process and visualize data more efficiently, teams would be able to tap into the knowledge from their domain, and adjacent domains to find hints for new solutions (Dubitzky et al., 2012) through interactive exploration of data.

Data mining technologies allow for a more efficient analysis of massive amounts of data, and quick visualization for exploration and interpretation. The most significant benefit from these technologies is the ability to obtain weak signals, or newly developing trends, to be explored. It also enables teams to quickly get an overview of what is happening in a domain, through which they can identify potential new connections to be explored and maybe materialize into a new solution (Dubitzky et al., 2012).

As noted by Siau (2000), there is a delicate balance of domain knowledge needed in order to be able to identify novel combinations. If the participants possess too little domain knowledge related to the problem or situation, it will be very challenging to perceive the potential links; however, participants with too much domain knowledge can be fixated to "correct" answers and not be open to new possibilities. Undergraduate and graduate students who are almost at the end of their studies have sufficient domain knowledge to understand the field, and because they have not been embedded in the ambient too long, are usually not yet entirely fixated on "correct" answers (Burkus, 2013).

In the world of academia, researchers and students increase knowledge by building upon the work of others, expanding theories by proving or disproving what has been hypothesized before. However, there is a trend towards interdisciplinary collaboration and the combination of domains, which leads to finding unexpected solutions. Industry lines are also blurring, the amalgamation of knowledge areas (mechanical engineering + electronic engineering = mechatronics, biology + medicine + engineering = biomedical engineering, etc.) result in innovations for the market. Universities are increasingly creating programs that enable students to take part in projects across field boundaries, providing an opportunity to study how these groups identify research projects that will include the skills and knowledge that each individual can bring with their background.

In this paper, we present a case study where eight teams of students from different engineering fields search for innovative solutions linked to water access and conservation in the context of an intensive multidisciplinary summer course on innovation and technological design. This study aims to understand the use of data mining software to exploit data from patents as input for technological design, where the data mining tool allows for an exploration of the current solution space to find hints for novel solutions.

It is posited that teams taking advantage of the data mining technology to evaluate current solutions and finding incumbent technologies and materials will be able to produce solutions which are deemed more innovative by a panel of experts. It is also expected that the level of sophistication of solutions will be increased, compared to teams not using the tool for exploration of the solution space.

3.2 Background

It has been proposed that to produce a really novel solution, the individual or team would have more success by making highly varying (wild), unexpected combinations (Fleming & Szigety, 2006). Popular ideas and known solutions get combined constantly since it they are known to work (Agogué et al., 2014), as Abraham Maslow's saying goes "if all you have is a hammer, everything looks like a nail". When faced with a new challenge, the individual or team will naturally gravitate towards solutions that are familiar or are proven in order to save time, or because they are "good enough".

If creative thinking is an iterative process of activating cues (Hamman, 2000), then external information inputs can help counter this fixation effect by providing hints to explore and open the possibilities to other types of solutions (Shan, Zhu & Zhao, 2013). Some authors have found that giving examples for expansion can actually increase originality (Agogué et al., 2014); furthermore, the likelihood of creating new knowledge from recombination is greater as the number of external inspirations is increased (Cohen & Levinthal, 1989, cited in

Kabir & Carayannis, 2013), especially when there is a large base of data and information to explore (Dove & Jones, 2014).

3.2.1 Data mining for new concept development

Data mining technologies are software programs that make it easier to process data and present it in formats such as reports and visualizations for users to interpret; it is the application of algorithms to extract meaningful associations in data (Siau, 2000) that can help retrieve correlations and trends.

Large companies call these “innovation technologies”, to refer to software that facilitates access to large amounts of data, by making it easier for employees to navigate through the information at hand (Dodgson, Gann & Salter, 2006). Innovation technologies enable companies to analyze internal and external data (e.g. websites, patents) with the purpose of finding what competitors are doing, potential new technologies, clients and collaborators, and recently, to find gaps in an existing domain or the appearance of new domains (Rhéaume & Gardoni, 2016). Some authors propose that to generate radical innovations, it is necessary to combine already existing knowledge but in an unexpected fashion (Fleming & Szigety, 2006). However, as people are trained to respond with known solutions, it is no easy task to try and find diverse elements to combine, particularly to identify elements that solve part of a problem if they come from a different domain; interactive data exploration can help by providing new pieces of data.

Discovering knowledge from data is an interactive and iterative process, wherein the user makes many decisions regarding the objective, the data sources, the processing and interpretation (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). Authors in the computer aided creativity domain have proposed several software applications and algorithms that can help in finding hints for solutions. Siau (2000) proposes Knowledge discovery to support organizational creativity, describing the process and the various techniques, while presenting the challenges at each stage, for example, setting the objectives, accessing the right data, and selecting the right processing for the objective. Trappey, Trappey and Wu (2009) highlight the importance of having a patent database in a new product development effort, to share knowledge within the team and analyze existing knowledge more efficiently.

3.2.2 Patent mining for creativity

Patents are a way to share technical knowledge in a detailed fashion that allows the extraction of important keywords related to materials, processes, functions, parameters, considerations and constraints of the artifact. Having all this information enables the mapping that can, to a certain point, describe the outlook of a domain, in the past and in the present. Though technology and information can help in the processing of data, human analysis is ultimately still required to make sense of the information and envision the novel connection.

Patent mining is used for new product or solution development to avoid overlapping with existing solutions that can lead to patent infringement, as well as to aid in the creation of new concepts (Trappey, Trappey & Wu, 2009). The use of patents for the early design stages can help decrease the time it takes to analyze a knowledge domain (Ríos-Zapata, Duarte, Pailhès, Mejía-Gutiérrez, & Mesnard, 2016) and benchmark current solutions. As mentioned in (Holzinger & Jurisica, 2014), there is a need for methods to facilitate the discovery of knowledge through interactive systems such as the one presented here, where a data mining tool enables the visual analysis and exploration of patterns in data.

Already several authors have documented the use of technologies for exploration of a knowledge domain with the intent to take advantage of the data for solution design. Dodgson, Gann and Salter (2006) document Procter and Gamble's use of "innovation technologies" to mine data from internal and external sources, namely patents and other documents, to identify potential new products. Verhaegen et al. (2011) document a software program that typifies the elements of a solution, they randomly selected patents with no specific target domain, and the algorithm classifies the solutions grouped by the purpose (function of the object), it then proposes concepts which achieve the same purpose to spark ideas for analogy design.

3.3 Hypothesis

Exploitation of patent data through the use of data mining tools can potentially be used by a team during the idea generation phase, given that the analysis of information for this purpose is time consuming. Also, the team can benchmark current solutions either to combine parts of different solutions into a new solution, map the domain to take expansive examples to generate new alternatives, or by exploring other domains through the search of keywords to find solutions that perform the same function for inspiration.

In this case, participants were introduced to a data mining tool, and were assigned a period of time in the context of a summer course to learn how to use the tool, explore the data, and use the data in their idea generation process. This allows teams to explore the data, decide on search terms and queries, and generate the visualizations which can best support their process (among others, the tool used provides semantic analysis, mapping of incumbent organizations, evolution of terms, and clustering).

By giving participants training in the data mining tool and access to data, patents in this case, we expect the teams using patent data to complement their information research in the idea generation phase to increase the level of sophistication in the solution proposed, compared to a previous iteration of the concept presented.

> ***Hypothesis 1:*** *Teams using patent data as input for idea generation will increase the sophistication of their proposed solution*

A second hypothesis is related to the performance of the ideas when compared to other solutions. We will be able to review the results of the teams' performances, based on the evaluation by a jury of experts who followed three presentations in a one month course where the teams of students worked to propose a novel solution to a challenge. It is expected that teams using patent data as an additional support for idea generation will be able to design technological solutions that will not only aptly resolve the challenge, but will be better ranked by the panel of experts in the final presentation of the course.

> ***Hypothesis 2:*** *Teams using patent data as input for idea generation will be graded more favorably in the expert evaluations*

3.3.1 Evaluation of results

Many different metrics exist to measure quantity and quality of the creative outputs. Other metrics are concerned with measuring the process itself or team dynamics. Quantity metrics measure number of ideas generated in total, of ideas evaluated, of ideas shared, or number of characters in a description. Quality metrics measure the quality of concepts or ideas accepted, applicability, detail of concepts, complexity level, novelty, or variety. Metrics focused on team dynamics measure number of participants, perceived team cohesiveness / effort; meanwhile process-based metrics count number of chats, comments, record cards / sticky notes, whiteboard events, and duration of the creative session.

The interest of this research is concerned with the quality of the results; it is assumed that the solutions will be applicable to the problem, and there will likely not be many teams working on the same issue as to evaluate the diversity. Therefore the metrics selected to assess the results are complexity and novelty of concepts. The novelty of concepts is the degree of inventiveness, measured by the principles used in the solution (Ardaiz-Villanueva et al., 2011, Glier et al., 2011, Wodehouse & Ion, 2012). The complexity level of concepts refers to the participants taking initiative and dividing the problem into sub-systems for further development (Ardaiz-Villanueva et al., 2011). Given the negative connotations of the term, and because in other domains it has the implication of interplay between domains of knowledge, it was decided to use the term “sophistication” instead, as this term indicates that the concept was refined and further developed to consider subsystems of the solution.

The sophistication level will provide an indication of whether by having access to patent data enabled participants in the teams to better define the different components or sub-systems of their solution. Having a well-defined solution will impact their scores in each presentation in front of a jury, as they will have more answers as to what their solution does and how. The novelty of concepts will also be reflected in the jury evaluation, as part of the evaluation includes having identified current solutions, and proposing a different and novel concept.

3.4 Case study

To observe how the use of a data mining tool to facilitate the exploitation of patent data impacts the information gathering and idea generation process in a creative team, it was decided to work with a group of students in an innovation and technological design course who had to solve a technical challenge in a period of one month.

One member of the research team gave a lecture to the group of engineers on the use of data for creativity, and an introduction to a specific software with pre-loaded data related to the proposed challenges. The tool would allow teams to interactively explore the knowledge for current solutions, by identifying main trends and elements present in published patents.

The utilization of the tool was optional, meaning it had no impact on their evaluation, and the students were aware of this fact. The research team only found out after the final evaluation which teams had used the tool based on the activity journal the students were asked to keep throughout the course (this will be discussed in detail in section 3.4.2).

3.4.1 ETS Summer School

The ETS International Summer School on innovation and technological design is an intensive one month program, where students are trained to prototype iteratively to develop a technological solution. It is aimed at master students, as it provides three master course credits, but it is open to last year bachelor students and PhD students.

The course took place during the month of July, 2016. The group was formed by 48 students coming from 11 different countries (Brazil, Canada, Colombia, Czech Republic, Germany, Hong Kong, Iran, Italy, Mexico, South Korea, Singapore), two in the PhD level, 15 in the master level and 31 bachelor level students, spanning twenty-four different engineering specializations.

The purpose of the Summer School is to train engineering students on the innovation process, creativity tools and techniques, decision making tools for innovation projects, prototyping and presenting (pitching) a project. At the end of each week during the course, starting on week 2, students had to present an advance on their project: first, the problem definition; second, the initial concept, and third, the final concept. The presentations were evaluated by a group of professors who are experts in different engineering areas, who were available throughout the course to respond to technical questions by the students, but were not involved in the pedagogical planning or the research presented in this article. Table 1 is an overview of the academic program with the different concepts and theories the students were exposed to during the Summer School:

Table 3.1 Overview of the Summer School pedagogical program

<table>
<tr><th colspan="3">Week 1</th><th colspan="2">Week 2</th><th>Week 3</th><th>Week 4</th></tr>
<tr><td colspan="2">Project</td><td>2-hour challenge</td><td>2-day challenge</td><td colspan="3">Project: 2-week challenge</td></tr>
<tr><td colspan="3">Evaluation</td><td colspan="2">Problem definition</td><td>Initial concept</td><td>Final concept</td></tr>
<tr><td>Courses</td><td colspan="2">1. The innovation process
2. Creative processes
3. Teamwork
4. Basic prototyping
5. Problem definition
6. Making your pitch
7. Creativity tools & techniques</td><td colspan="2">8. Design thinking
9. Big data for innovation</td><td>10. Prototyping
11. Business aspects
12. Concept selection</td><td>Team coaching</td></tr>
</table>

3.4.2 The activity journal

In the course of the Summer School, students were trained in different creativity tools and techniques. After a theoretical introduction of each tool or technique, they were allocated time to apply the tool or technique to the problem they were trying to solve. The ultimate goal of providing participants with creative tools and techniques is to arrive to a novel solution for the challenges (problems) proposed during the course.

Because students employed different tools and techniques, we needed to compare the results of the application of each. To do so, the teams were required to carry an “activity journal”. In this journal, each group chronicled the ideas generated during each workshop and the activities performed during the day. Ideas could be documented using brain-maps, lists, drawings, sketches or photographs to represent the work achieved with the tool / technique.

3.4.3 Team composition

To form the teams, we first allowed students to vote for the "team leaders", based on their experience during the first week of the program, where all activities were executed in varying teams. The group selected 6 leaders, thus the professors were tasked with selecting two more, also based on observing participants for leadership potential. Team leaders were responsible of selecting the challenge to work on for the team; they had 16 hours to consider their selection. The available challenges were:

1) Taking water samples from the river: How to take regular water samples on several strategic points along the St. Lawrence River using autonomous technology powered by a (or various) renewable energy source(s)?
2) Notifying the population about the quality of the river's water: How do you inform population in real time about the water quality of the St. Lawrence River and the risks of contamination due to sewer overflow?
3) Examine, clean and repair damaged sewer pipes: Develop a non-polluting, "intelligent", autonomous system that examines, cleans and repairs the cracks on the concrete pipes that carry sewage to water treatment plants.
4) Recognize and list the river's rare or invasive species: How can we allow citizens to recognize and geo-index rare or invasive species (plant or animal) from the St. Lawrence River's ecosystem in order to protect or control them?
5) Reducing erosion along the river banks: How do you prevent damage caused to the St. Lawrence River's banks by waves from passing ships or wind, without harming the ecosystem or hindering access to the river - all while recovering lost energy?
6) Removing solid waste that pollutes rivers: How do you remove solid residue and pathogenic microorganisms from sewer overflow before they pollute the St. Lawrence River, using an autonomous process powered only by renewable energy sources?

After each team leader communicated their project selection, the 8 team leaders then selected the members they deemed necessary for their project. To avoid having teams based on social compatibility, and to promote a more objective team formation, team leaders selected their team in a draft akin to sports players, based on statistics and profile.

Student profiles were anonymized by abstracting the level and field of studies, and removing gender and university of origin. Each team leader was able to select one profile at a time in several rounds of selection, until all profiles were assigned to a team. Figure 3.1 shows an example of a "player card" with the abstracted information of students. Table 3.2 is a summary of the resulting teams.

Graduate student in Environmental engineering
Background
Process engineering
Experience
Petroleum refining
Personal skills
Cooperative, attentive, determined and ambitious
Software / Programming languages
Fortran, HYSYS, SuperProDesigner, Stella, Solidworks
Player #
P26

Figure 3.1 Example of player cards used to form teams

Table 3.2 Overview of teams

Team	Bac.	MS.	PhD	Males	Females	Fields
1	3	3		5	1	Architecture, Civil Eng., Environmental Eng. (2), Industrial Eng., Mechanical Eng.
2	4	2		5	1	Automated production Eng., Environmental Eng., Industrial Design, Industrial Eng., Innovation management, Sustainable Development Eng.
3	3	3		4	2	Aerospace Eng., Business Administration, Energy Eng., Health Risk and Occupational Safety, Mining Eng., Total Quality Eng.
4	3	2	1	4	2	Aerospace Eng., Chemical and Biomolecular Eng., Electrical Eng., Industrial Eng. (2), Mechanical Eng.
5	5	1		4	2	Architecture, Bioeng., Civil Eng., Energy Eng., Mechanical Eng., Surveying
6	5	1		2	4	Architecture, Energy Eng., Industrial Eng., Mechanical Eng., Surveying, Sustainable Development Eng.
7	3	2	1	3	3	Environmental Eng., Industrial Eng., Logistics operations, Materials Eng. (2), Total Quality Eng.
8	5	1		4	2	Energy Eng., Environmental Eng., Informatics, Mechanical Eng., Nuclear and Risk Eng., Software Eng.

From the moment the teams were formed, they began to work on their selected challenge. The first evaluation required each team to define the problem they would focus on, based on the challenge, to determine a scope. Table 3.3 shows the problem statement to which each team arrived.

Table 3.3 Problem statement defined by the teams

Team	Challenge	Problem statement
1	6	Remove solid residue from sewer overflow to prevent contamination of the St. Lawrence River.
2	6	Prevent solid waste of the sewer system from reaching the St. Lawrence river during overflow by using autonomous solution powered by a renewable energy.
3	1	How to develop an efficient system to support water quality monitoring in the St. Lawrence river?
4	1	Sampling, measuring and transporting St. Lawrence river water for human activity and environment quality.
5	5	Dissipate energy along the riverbank.
6	2	Inform and educate people about the water quality of the St. Lawrence river
7	2	To inform public in real time about the quality and risks of the St. Lawrence river caused by sewer overflow.
8	4	Motivate citizens to be interested in endangered species from the St. Lawrence river ecosystem.

3.5 Big data for creativity

To introduce students to the use of big data and data mining tools for the purpose of creativity, one of the researchers presented a one-hour course that introduced the following concepts:

- Difference between data, information and knowledge (Ackoff, 1989)
- Definition of big data (Howkins, 2002; Shan, et al. 2013)
- Five key characteristics of big data: volume, velocity, veracity and value (Laney, 2001)
- The data analytics process: identification of data sources, selection of data, data cleaning and processing, analytics, and finally interpretation (Baesens, 2014)
- Phases of the creative process (adapted from Shneiderman et al., 2006) where data mining tools and big data can be included: problem definition, information gathering, idea generation, and idea selection

- Overview of IPMetrix tool: patent search, data upload, generation of reports and visualizations.

There are many software tools available in the market to mine data, however, the market for data mining tools for big data is not as comprehensive, and many of these require specialized professionals to implement and manage the software. The authors of this research selected the software IPMetrix, by French company TKM to perform the data mining due to its ease of use and accessibility to the end user. It requires minimal training for end users, and it offers direct connection to patent databases such as PatBase and the European Patent Office.

As all challenges were concerned with water conservation and access, and due to the loading times necessary to build a knowledge base in the software, a pre-load was performed before the students were given access to the tool. Challenges 1, 2, 4, 5 and 6 were selected by the teams, thus were considered for the data pre-load. The pre-load procedure followed is described next:

1) Identify relevant expressions in the challenge to build an initial knowledge base (first column in Table 3.4.
2) Conduct a search for patents accessible from the data mining tool for each challenge. The following were found for each challenge (columns “Keywords” and “Records found” in Table 3.4.
3) Upload resulting patents to the data mining tool.
4) Cleaning the results - The IPMetrix tool automatically cleans the data for inconsistencies, normalizes the content of patents to facilitate processing.

All teams obtained one access to the IPMetrix tool with previously loaded data, but the use of the tool was optional, as was the additional load of information. The researchers were not made aware of which teams opted to use the tool, and only found out by going over the activity journals which of the teams used the tool.

Table 3.4 Queries performed for each challenge for the data pre-load

Challenge	Keywords	Records found
1. How to take regular water samples on several strategic points along the St. Lawrence River using autonomous technology powered by a (or various) renewable energy source(s)?	water sampling + techniques	293
	water sampling + analysis	560
	water sampling + collection	463
	water sampling + procedures	195
	autonomous water drone	296
	river water quality sensor	803
2. How do you inform population in real time about the water quality of the St. Lawrence River and the risks of contamination due to sewer overflow?	real time + notifications + mass communication	42
	real time public mobile communication	661
	water sampling + techniques	293
	water sampling + analysis	560
	water sampling + collection	463
	water sampling + procedures	195
	autonomous water drone	296
	river water quality sensor	803
4. How can we allow citizens to recognize and geo-index rare or invasive species (plant or animal) from the St. Lawrence River's ecosystem in order to protect or control them?	image recognition + vegetation	98
	image recognition species fish	80
	real time + notifications + mass communication	42
	real time public mobile communication	661
5. How do you prevent damage (erosion) caused to the St. Lawrence River's banks by waves from passing ships or wind, without harming the ecosystem or hindering access to the river - all while recovering lost energy?	riverbank + erosion	99
	riverbank + management	32
	wave energy converter terminator	37
	oscillating water column	115
	overtopping device + wave energy	10
	wave energy + attenuator	876
	wave energy + point absorber	138
6. How do you remove solid residue (waste) and pathogenic microorganisms from sewer overflow before they pollute the St. Lawrence River, using an autonomous process powered only by renewable energy sources?	solid waste removal rivers	13
	solid waste removal water	66
	solid waste removal ocean	9
	solid waste removal sea	12
	sewage overflow	221

3.6 Results

In their activity journal, four teams were identified as having used the data mining software to complement their information research to further develop their solutions:

- Team 3: One team member was in charge of identifying competitors from the data in the tool. The data was then used to generate ideas.
- Team 4: From the data available, they selected technologies to combine with previously generated ideas into a new solution.
- Team 5: The team organized the data in the tool, listed and discussed current solutions to generate ideas.
- Team 7: Combined ideas from the data available with other market trends.

To verify the extent to which Hypothesis 1, "Teams using patent data as input for idea generation will increase the sophistication of their proposed solution" occurred, we compared the proposed concepts before and after the introduction to the data mining tool.

The initial concepts proposed by the teams in the first iteration of the solutions, before having access to the data mining tool and being exposed to the concept of using big data as an input for idea generation, are listed in Table 3.5. Table 3.5 also shows the solutions presented by teams after being offered the course and the access to the data mining tool.

Table 3.5 Initial vs. final concepts (before the data mining / big data for creativity lecture)

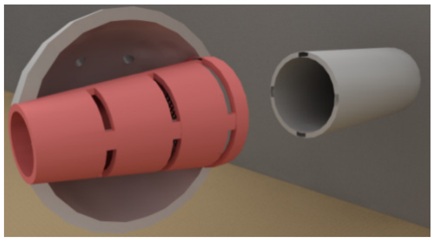
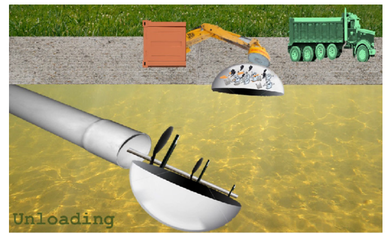
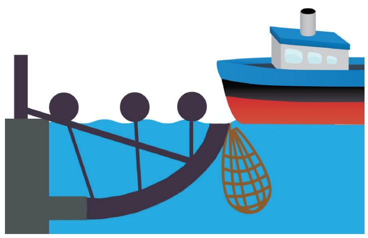
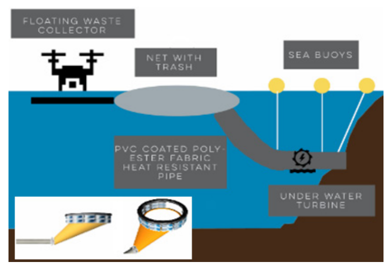
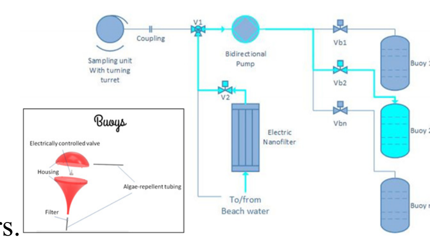
Team	Initial concept	Final concept
1	Cone shaped filter/turbine which gradually removes garbage of different sizes and stores them in a removable container. 	Cone shaped filter/turbine which gradually removes garbage of different sizes and stores them in a removable container, paired with container changing station in the shore. 
2	Floating waste collector net that traps garbage while water continues to flow through (attached to buoys to float). 	Floating waste collector cone that traps garbage while water continues to flow through (attached to buoys to float). A turbine in the pipes keeps the flow of water and waste. 
3	1: System attached to boat 2: On demand sample and storage 3: Crowdsourcing measurements 4: Onsite bacteria analysis and warning (no image of solution was provided)	Solar-powered set of interconnected buoys which use a Peristaltic pump to send off-shore samples to the shore for collection in QR coded containers. 

Table 3.5 Initial vs. final concepts (before the data mining / big data for creativity lecture) (continued)

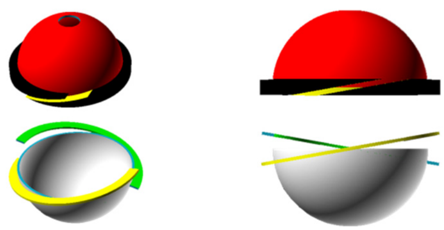
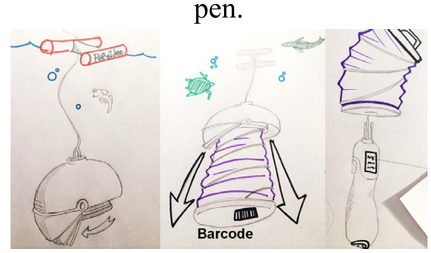
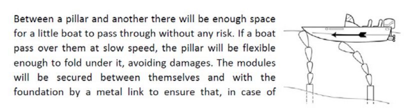
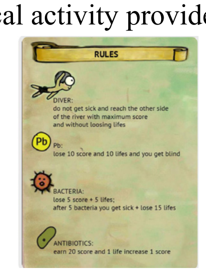
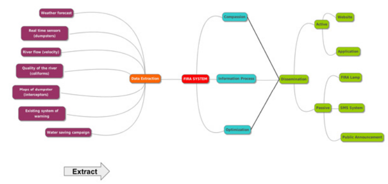
Team	Initial concept	Final concept
4	Autonomous ball-like device that measures in situ parameters and collect water samples. 	Ball-like submersible device with barcoded expansible sample container and portable testing pen. 
5	1: Floating sidewalk 2: Pillar system 3: Buoyant system (no image of solution was provided)	System of wave-breaking pillars that bend under boats and generate electricity from the movement with piezoelectric generators.  
6	Game in a mobile application to educate users about the river water quality. (no image of solution was provided)	Game in a mobile application to educate users about the river water quality. The business model includes collaboration with team 7 to obtain real-time water quality information and advertisement of local activity providers.  

Table 3.5 Initial vs. final concepts (before the data mining / big data for creativity lecture) (continued)

Team	Initial concept	Final concept
7	Mobile application to inform users about water quality, information is extracted from different public sources (no image of solution was provided)	Website and mobile application to inform users about water quality, information is extracted from different public sources and processed for forecasting. Solution is paired with solutions for non-smartphone users (SMS alert) and even non-mobile users (on-site light indicators) 
8	Mobile application to allow users to take pictures and upload them to a database where experts can collaborate with information. Users are engaged with gamification techniques. 	Mobile application to collect pictures of flora and fauna species found along the river shore. It allows users to share, get points and request information from experts. 

From the solutions presented after the big data for creativity lecture, changes can be observed for all teams; however, some teams presented only aesthetic improvements, while other teams developed complete systems around the initial solutions. A summary of the changes are listed in Table 3.6.

Table 3.6 Analysis of changes to team solutions

Team	Use of data mining tool	Observable changes	Sophistication increase
1	No	Added automation of the garbage pick-up, using a static pole in the shore	Intermediate
2	No	Changed "net" (flexible) to "cone" (rigid)	Low
3	Yes	Designed complete system based on buoys	High
4	Yes	Improved sample taking device, designed complete system (management platform, transportation device, test kit)	High
5	Yes	Designed flexible and energy-capturing (subsystems) wave-breaking pillars (system)	High
6	No	Developed a business model based on local advertisement	Low
7	Yes	Identified user segments, designed physical solution to go along digital	Intermediate
8	No	Aesthetic improvements to design	Low

To test Hypothesis 2, "Teams using patent data as input for idea generation will be graded more favorably in the expert evaluations", we referred to the three presentations and the resulting evaluations by the panel of technical expert professors, who are external to the teaching and research staff. The first evaluation was concerned with the problem definition; in the second presentation, teams offered an initial concept of the solution; while the final concepts were revealed during the third and final presentation. The juries were provided with an evaluation grid for each presentation, the aspects reviewed in each presentation are summarized in Table 3.7. The panel is listed on Table 3.8.

Table 3.7 Summary of aspects for evaluation in
Presentations 2 (initial concept) and 3 (final concept)

Criteria	Initial concept	Final concept
Team presentation		3
Clearly stated problem		10
Time management	20	2
Benchmarking		10
Target market		5
Concept solves the problem	40	15
Creativity		10
Sustainability		5
Competitive advantage		10
Prototype	20	20
Evolution of concept / prototype	20	10
Total points	100	100

Table 3.8 Panel of experts

	Degree	Areas of specialization
1	PhD, Hydrogeology MSc, Integrated water resources management	Water Analysis, Conservation, Quality, Balance, Chemistry, Engineering, Sampling, Sediment Pollution. Environment, Hydrogeochemistry, Glaciology, Contaminant Transport Hydrology, Field Sampling, Heavy Metal Pollution,
2	PhD, Nondestructive Testing MASc, Mechanical Eng.	Ultrasound, Guides waves, Structural health monitoring, Biomedical imaging, Doppler ultrasound, Dynamic elastography, Transducer development
3	PhD, Water Resources MSc, Geology	Hydrology, Water resources, Impact of climate change on water resources, Hydrologic modelling, Hydrometeorology, Hydraulics
4	PhD, Computer Eng. BS, Computer Science	Software design for stability, Software verification and validation
5	PhD, MScA, Civil Eng.	Environmental chemistry and engineering, processes for treating wastewater, Nanotechnologies for environmental protection, 4R-VD for waste management
6	PhD, Master, Electrical Eng.	Nonlinear control and optimization applied to robotics, flight control systems and multizone power network control.
7	PhD, MEng, Electrical and computer Eng.	Medical image analysis, Computer vision, Machine learning, Image processing, Information theory, Local invariant features, Computer assisted diagnosis

Table 3.9 shows the results from the evaluations by the experts for the Initial and Final concepts.

Table 3.9 Results from expert evaluations

Team	Initial concept	Final concept
1	80.5	84.6
2	77.8	83.9
3	89.8	**87.8**
4	84.5	**87.6**
5	92.2	**93.7**
6	90.3	73.8
7	89.9	**83.9**
8	84.8	82.6

Figure 3.2 compares the results of teams using the data mining tool, versus the teams not using the tool:

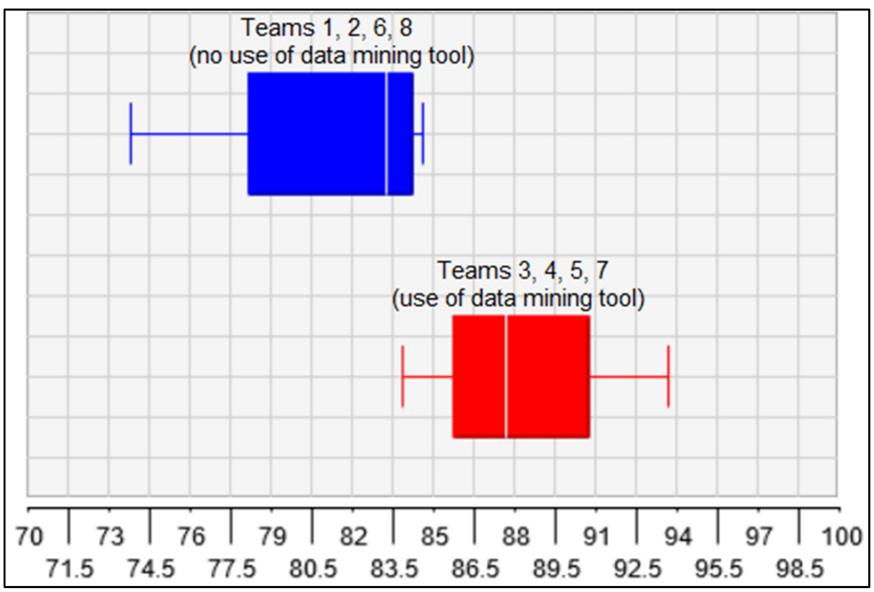


Figure 3.2 Results from teams using the data mining tool vs. teams not using the data mining tool

3.7 Discussion and conclusion

By comparing the concepts presented by the teams before and after the lecture on the use of data mining and big data as tools for creativity, the teams that indicated the use of the data mining tool in the activity journal show an increase in the sophistication of the proposed concepts; they were able to identify components to the solution they had not considered before, making a more complete offer for the potential users and clients. These teams designed more components for their solution; by contrast, teams not using the data mining tool only changed their solution with incremental improvements to their original proposition. The use of the data mining tool to explore the data allowed the teams to interact differently with data; it enabled the exploration of a broader knowledge base, not restrained to the knowledge within the team.

The expert evaluations also graded the teams using the data mining tool with higher marks, which indicates that the concepts were, among other things, deemed more creative while solving the problem at hand. It is worth noting that the experts were unaware of the difference, as they were not familiar with the pedagogical plan of the course and were not informed on the differences in the use of the available tools and techniques by the teams.

Another possible explanation as to why the teams using the data mining tool to explore patents received better evaluations from the jury could stem from the fact that the participants in the team were more motivated to use all the resources available to them to improve their idea generation and the iterations for their concept. Meaning it is possible that they not only used the data mining tool more than other teams, but also the additional set of creativity techniques and methods in their unsupervised work time, while the teams choosing not to use the data mining tool might have also avoided or underused the techniques and relied solely on improvisation or commonly used techniques such as brainstorming and trend identification.

It can be argued that the volume of data was not enough to be considered big data, and that teams could have benefitted from a larger dataset. Nonetheless, one important characteristic of data mining techniques for big datasets is the speed with which the user can analyze information when needed. A team of 6 members would not be able to analyze all patents found in the domain in time to come up with a novel concept in the required timeframe.

The observations made during this case signal a positive effect stemming from the use of patent data for idea generation when comparing teams using the information against teams not using the information. It provides an advantage to benchmark a solution and increase the sophistication in the proposed concepts. Engineering teams would benefit from having access to a data mining tool during a new concept design process to quickly get an overview of the domain and the current technologies and solutions.

However, as is the case with many creativity tools, methods and techniques, it cannot be determined that the use of the tool is by itself the panacea for creative teams. The data mining tool is but one tool that can be used along other tools and techniques to provide an additional advantage in terms of efficiency in the search for potential technologies, materials, collaborators and even users, but one must be careful not to expect that information by itself will provide an answer.

3.8 Future work

The scope of this project does not allow finding out whether the participants were able to generate diverse solutions because they were inspired by the data to combine existing solutions, or whether they kept trying to come up with a different solution to what has been done, in an effort to differentiate their solution. It would be interesting for future research in this area to record the work sessions and analyze the decision making in the teams at the moment of the data exploration, to identify the attitudes towards the information, and the subsequent actions.

The interdisciplinarity in this exercise was also limited, as most participants except 2 have an engineering background. A more diverse group could potentially increase the levels of originality in the ideas generated, as their knowledge of correct answers in engineering solutions would be lower. It would also be interesting to pair the data exploration with different creativity methods with groups working on the same project, as to assess the differences on the use of data mining tools at the different stages of product or service development.

Finally, the use of data mining tools to explore solution spaces interactively can also lead to other types of interaction, within the team by suggesting new data connections, and with potential users to design prototypes. It is worth exploring the different interactions made possible by the inclusion of data mining tools, and their impact on idea generation, and new solution design.

3.9 Design cycle evaluation

NOTE: This section does not appear in the article, it is meant as a conclusion and transition for the next chapter of the thesis.

The objective for this case was to provide participants with training on the data exploration tool, and more time to perform the exploration, as a way to support idea generation to develop a novel concept for engineering.

It was found that half of the teams elected to take advantage of the exploration tool, and only one team added information to the database. This suggests that even with training, the use of the tool is not as evident to non-experienced users, which can discourage the adoption of this type of tools in design teams.

A second relevant conclusion from this case is that teams were not able to generate radically novel ideas. It is possible that access to a large database of information with no limit can overwhelm participants, who search only for improvement to their ideas and not radical innovations.

For the next case, it is proposed to provide teams with pre-selected keywords from the KDD process by an external actor, in this case the researcher, who will artificially delimit the combination possibilities, and assess whether this reduction of the exploration space is useful for teams in the early stages of the EDP.

CHAPTER 4

PROMPTING INVENTIVE SOLUTION DESIGN WITH KEYWORD CUES FROM PATENT MINING IN AN INNOVATION COMPETITION

Ma-Lorena Escandón-Quintanilla[1], Mickaël Gardoni[1,2], and Patrick Cohendet[3]

[1] École de Technologie Supérieure, Montréal, Canada

[2] Institut National des Sciences Appliquées, Strasbourg, France

[3] HEC Montreal, Montreal, Canada

This article was submitted to the Creativity and Innovation Management journal on December, 2016

Following the findings from Case 2, presented in the previous chapter, it was decided to provide participants with a selection of keywords extracted from a KDD process performed by the researcher. The data was provided early in the EDP, in order to support participants in the early stages, particularly idea generation.

By providing participants with relevant keywords, they will be able to combine previous knowledge with hints of solutions of the problem space, obtained from the application of KDD in a pool of patents related to the problem.

Abstract

Innovation contests are an opportunity for companies and organizations to obtain ideas from participants with diverse backgrounds who are not fixed with the bias of the industry. However, the novelty of proposed ideas is hampered by an inability of teams to create original ideas in time constrained competitions. The knowledge base of participants can be incomplete, and the search for data can be inefficient with traditional web searches. Data on a knowledge domain can help teams participating in an innovation contest by providing clues

to propose novel ideas. Data mining technologies make it easier to identify trends and interesting relationships in data which are otherwise not straightforwardly observed, or cannot be detected unless massive amounts of data are being analyzed. The utilization of data mining technologies is not yet accessible to everyone; trained professionals are needed to take advantage of potential insights. This article presents the use of the product of a data analysis made available to teams during an innovation contest to facilitate the generation of novel ideas. Teams with access to keyword cues showed increased variety in the ideas proposed, and in the elements in the solutions to respond to defined constraints.

4.1 Introduction

An innovation contest is a limited-time competition, usually based on information technologies, where the organizers call on the public or targeted groups to propose innovative solutions to problems (Adamczyk, Bullinger & Möslein, 2012). Because participants are asked to propose a novel solution to an issue, it would be disadvantageous to rely solely on the knowledge within the team, and not the large amounts of data available a few clicks away. However, when pressed for time, the majority of teams will usually rely on knowledge possessed by team members, and solutions that are known to work (Howard, Culley & Dekoninck, 2006).

Though the internet provides access to information and is widely used, it is not as useful in the context of idea generation, particularly when time is limited. There are many reasons for this: the information is not entirely reliable, it can be contradictory, there is no rigor or limitation to publish. It is also laborious to go through thousands of websites and synthesize what has been found, and it is difficult to compare results from one search to another, whereas data mining tools make it possible to compare a set of results to another, provide an overview of the knowledge domain (players, main technologies, applications) and provide visualizations to facilitate analysis.

Exposing participants to expansive examples can actually increase the originality of ideas produced (Agogué et al., 2014), and while the analysis of data mining results could provide examples of solutions in a knowledge domain, the use of the tools is not evident to non-experts; there is a need to make data mining more accessible to participants (Dove and Jones, 2014). To facilitate the use of data as an input for idea generation, a process was devised to extract knowledge from a domain and present it to time-pressed participants. This paper aims at responding how and whether data can be used for time-constrained idea generation, and if the resulting solutions are more varied and novel, compared to solutions where the participants did not have additional input.

The bounty of data presents some disadvantages, for instance, it has become increasingly difficult to process without the correct tools, and requires more financial, technological and human resources to manage. This issue of too much data and the difficulty in analyzing it is all the more evident at short idea generation sessions or innovation contests where participants are time-constrained.

Being able to access data from the problem domain can support idea generation by providing hints to trends and potential correlations. Creators can identify potential combinations and gaps in the knowledge which otherwise would not be recognized (Müller et al., 2012) by drawing from a large pool of data through the use of data mining tools. The task of mapping a domain of knowledge has been facilitated by data mining and visualization technologies, but we have yet to develop technologies that can generate novel and valuable ideas by themselves (Boden, 2004).

4.2 Background

Much research in creativity has been dedicated to finding ways to support teams with information or cues during idea generation efforts; we can find an area of information systems design focused on the application of information technologies for this purpose. A few examples include Maccrimmon & Wagner (1991) who support the generation of alternative ideas using data in short sessions, and found that free-form techniques can help generate ideas, but users need more stimulating techniques to continue being productive. Hamman (2000) proposed the use of information systems to support the creativity of music composers, concluding that algorithms can be used to propose combinations to inspire composers. Müller et al. (2012) created a software program to support the identification of unexplored research areas in biomedicine that will help researchers look at data in new ways to help generate new hypotheses. Shan, Zhu & Zhao (2013) support brainstorming by recommending computer generated ideas using idea networks. And Dove & Jones (2014) explore the use of aggregated data to support idea generation in workshops and found that it supported collaboration and participant engagement, however, ideas generated were not as novel as expected.

These examples show that data can be an input to boost creativity, given that creative thinking is an iterative process of activating cues (Hamman, 2000). Using data exploration in the idea generation process has enormous potential for insight discovery, and can also help participants get "unstuck" (Shan, Zhu, & Zhao, 2013, Agogué et al., 2014). Furthermore, the likelihood of creating new knowledge from recombination is greater as we increment the number of external stimuli (Cohen & Levinthal, 1989, cited in Kabir & Carayannis 2013). It is here where access to data through data mining technologies can actually be of help for idea generation: by providing clues and stimuli.

Data mining technologies make it easier to process large amounts of data, and visualize it in more accessible ways. Data mining is the application of algorithms to extract meaningful associations in data (Siau, 2000), meaningful correlations and trends that are otherwise not easy to observe. Data mining technologies help find similarities, trends and correlations, and it is up to people to evaluate the results and gain insights. Many applications exist for these technologies, and while the software goes a long way in processing the data, the last step always involves a person or group of people to interpret the results (Fayyad, Piatetsky-Shapiro, & Smyth, 1996), giving meaning to them and transforming the resulting information into knowledge (Ackoff, 1989).

At innovation contests, the challenges or questions proposed to participants come from organizations or companies facing those issues, however, they are usually open questions, meaning there is no pre-defined outcome and participants are free to propose solutions with no restrictions on budget, functionality, materials, target market, etc. Participating teams are usually multidisciplinary, working towards a defined objective, with limited time, and are willing to collaborate and create knowledge (Jiménez-Narvaez, Desrosiers & Gardoni, 2011). The confluence of diversity in team members, and time and task restriction, make up an interesting context to explore how the use of keywords cues extracted from patent mining using data mining tools can support creativity in the generation of novel solutions.

There have also been documented cases of innovation competitions where companies make open calls for the public to submit ideas or solutions for a prize. For example, in the IBM Innovation Jam, IBM used an internal application to bring together employees around to world to generate ideas for new business units. Participants are encouraged to comment on the ideas of others and a jury selects the best ideas to be then implemented in the company (Bjelland and Wood 2008). Lego deployed a "virtual product design space" (Majchrzak & Malhotra, 2013) for users to create their own design in Lego Mindstorms. Lego selects the winners, and awards a prize, but keeps all intellectual property. Netflix also decided to invite teams of programmers to come up with a better recommendation algorithm, during the contest the teams could see the leaderboard and their results to try and surpass that number (Rosen, 2011). IdeasProject was the "first external idea crowdsourcing" effort by Nokia to obtain ideas from just about anyone. They used text-mining, clustering and regression analysis to study the data and made an internal report to use as creative input (Vuori, 2012).

A commonality to these efforts in extracting innovation from participants in a contest is that data has been used as the object of the invention, and not as an input or inspiration for novel ideas, as was the case in the examples where information cues have been a source of information to prompt creativity.

There is an interest to enhance idea generation through the use of data (Maccrimmon & Wagner, 1991, Hamman, 2000, Shan, Zhu, & Zhao, 2013, Dove & Jones, 2014). In a notable case, Dove and Jones (2014) attempted to complement the idea generation process by using data as a support for creativity in workshops. They conclude that data can be useful to aid in the idea generation process especially when there is no predefined outcome (open questions), but it is necessary to make data more accessible to participants. Human interpretation plays a key role, as one must be careful not to get lost in the data, or delay decision making infinitely due to new data constantly arriving, or being uncertain of the veracity and usefulness of available data. Dove and Jones (2004) propose the use of moderators to surpass the difficulties of using data mining tools to exploit data.

There is an enormous wealth of data that can potentially be exploited. However, there is a difficulty for novice or non-expert users to select relevant and valuable data, “cleaning” the data and enriching it with meta-data created by calculation to make it usable for the intended purpose.

Patents offer the advantage of having pre-defined sections, describing a problem and the solution. It is possible to map the connections, to some extent, between keywords in a domain by using data mining techniques to analyze the knowledge in patents and scientific articles. Analyzing patents can help paint a picture of how a domain has evolved (George, Osinga, Lavie, & Scott, 2016), buzzwords or popular technologies over time, organizations with most expertise, and even the locations where this domain is of most interest. There can also be potential downsides to using patents for idea generation, for instance: inventions or solutions can exist outside the domain of knowledge, whereas the user is imposing a bias merely by defining the search keywords; some inventions or solutions can be protected by other types of intellectual property (industrial secret, copyright, etc.).

In this paper, we present a case where participants in an innovation contest were presented with keywords from a domain of knowledge to spark the bisociation of existing knowledge in a new problem or challenge, as it has been hypothesized that to create a new idea, the person applies existing knowledge to a new context (bisociation). Bisociation is defined by Koestler (1964) as the ability to merge two incompatible frames of reference, when an individual can consider a solution from one domain being applied in another. We can find this notion throughout the creativity literature also as “conceptual blending” (Fauconnier & Turner, 1998).

4.3 Hypothesis

If we follow Koestler’s theory that there is value in linking what is separated or incompatible, we must then search for the disconnections. Harnessing data in the proposed manner, the idea generation phase could potentially be improved in terms of the variety of resulting ideas. The variety measures the number of categories in which the solutions could be divided (Ardaiz-Villanueva et al., 2011, Glier et al., 2011, Wodehouse & Ion, 2012). An increase in this metric would suggest that participants did not gravitate towards known solutions, and were able to look for different types of solutions.

As mentioned before, the use of data mining to exploit data can provide links and correlations, and it’s up to the person or team to interpret and make meaning out of the results. It has been found that during idea generation tasks, especially if time is restricted, participants will try and tackle the idea generation portion by relying on tools and techniques they are familiar with. If a new tool or technique is made available, they will not take the time to learn to use it, even if the new tool promises better results. Because the data mining tool requires training to be able to use proficiently, it was decided that the data analysis should be done beforehand, providing participants with keywords to promote bisociation to find novel solutions.

The purpose of this research is to test whether participants are able to use a previously analyzed set of data in the form of a keyword input during an innovation contest, and if this input promoted the generation of more novel solutions. We expect that participants who are given keywords to blend will show a greater variety of technical solutions compared to participants who do not have this support.

Hypothesis: Teams with access to additional input from data analysis will have an increased variety in the proposed solutions

4.3.1 Procedure

To test the hypothesis, it is first necessary to prepare the input (keywords), provide access to participants in the competition, and evaluate the results (solutions proposed by teams). Overall, this is the process followed for the case study:

- Before the competition
    - Identification of comparable challenges (similar constraints)
    - Identification of keywords to conduct data analysis
    - Search for data
    - Upload data to data mining software
    - Analyze data
    - Select keywords

- During competition
  - Publish keywords for selected challenges
- After competition
  - Transcribe descriptions of solutions verbatim
  - Abstract elements of each solution
  - Group elements into the similar constraints identified for the challenges
  - Analyze results

4.3.2 Limitations

One restriction is the uncertainty of participation, as some participants who register before the event don't complete the challenge, while other participants don't register until it is time to submit the end result, and some local sites perform local registrations and do not share this information with the main site. Given that searching for keywords, patents and analyzing the data takes preparation and time, giving access to participants would have given them hints as to the topic of the challenges, which are unveiled on the launch of the competition. Thus, providing access to pre-registered participants would give them an unfair advantage over participants that register locally only or until the end of the contest.

4.4 Case study

We proposed the use of data analytics in the context of an innovation contest to facilitate idea generation for innovative concept development. During the innovation contest, 22 challenges were proposed to over 2,000 participants in 195 teams, distributed all over the world. Each team selected a challenge to work on, and proposed a solution by the end of the contest.

4.4.1 The 24 hours of innovation

The "24 hours of innovation" is a non-profit innovation contest that occurs every May, organized by the ÉTS (*École de technologie supérieure*) in Montreal, Canada. The competition aims to promote collaboration between students, researchers, experts, businesses and the general public, to propose innovative solutions. There are no restrictions for participants regarding location, age, gender, education, occupation, background, etc. Companies and organizations propose challenges they are currently facing, and teams have 24 hours to create a solution and present it in a 2-minute pitch video. Table 4.1 concisely represents the elements of the 24 hours of innovation competition (Adamczyk, Bullinger & Möslein, 2012):

Table 4.1 Characteristics of the 24 hours of innovation competition, compared to the categorization by Adamczyk, Bullinger & Möslein (2012)

Categories	Description	24 hours of innovation
Media	Online, offline, mixed	Mixed (online, offline)
Organizer	Company, public organization, non-profit, individual	University (nonprofit)
Task / topic specificity	Open task/low, specific task/high	Open task / low specificity
Degree of elaboration	Idea, sketch, concept, prototype, solution, evolving	Concept
Target group	Specified, unspecified	Aimed at university students
Participation as	Individual, team, both	Teams
Contest period	Very short term, short term, long term, very long term	Very short (24 hours)
Reward / motivation	Monetary, non-monetary, mixed	Monetary
Community functionality	Given, not given	Given: social media channels before, during and after competition

Table 4.1 Characteristics of the 24 hours of innovation competition, compared to the categorization by Adamczyk, Bullinger & Möslein (2012) (continued)

Categories	Description	24 hours of innovation
Evaluation	Jury evaluation, peer review, self-assessment, mixed	Jury evaluation (local and international) Peer review (for public's favorite)
Attraction (marketing / activation)	Online, offline, mixed	Mixed (varies according to local site)
Facilitation	Professional facilitation, peer facilitation, mixed	Mixed (depends on local site)
Sponsorship / partnership	Family, friends and colleagues, universities, national associations, specific industries, state and local agencies, mixed	Mixed (universities, private organizations, public entities, municipalities)
Contest phases (rounds)	One, two, more	One
Replication	Biannual, annual, less frequent, more frequent	Annual

It is important to distinguish between the concept of innovation contest, and an idea market: innovation contests are usually a limited-time competition, based on information technologies, where organizers call the public or specific groups for innovative solutions to problems (Adamczyk, Bullinger & Möslein, 2012), while idea markets are virtual marketplaces where an idea "provider", who can be an individual or organization, can sell a solution to a "buyer", a company searching for an innovative solution (Natalicchio, Messeni Petruzzelli, & Garavelli, 2014).

In the case presented here, the competition is not an idea market, as the sponsor does not buy the solution from the teams, but rather proposes an "open problem", and the winner is selected by a panel of international experts based on Innovation and creativity, Analysis of scientific and technical information, Quality of presentation and Eco-responsibility, and compete with solutions for all challenges, not only the challenge by the same sponsor. Furthermore, the organizer of the contest does not keep any intellectual property of ideas proposed by participants, nor does it obtain monetary gain from the contest, as it is a non-profit event organized by an educational institution.

The 24 hours of innovation competition is based on collaboration principles. Teams are encouraged to collaborate with other teams, discuss ideas and potential issues, and even ask for help outside the contest. The organizers also invite experts who can solve technical questions regarding the solutions. The companies, organizations, researchers and experts can provide additional information through email, social media or through the live web conference feed if participants have questions about the technical aspects of the challenge; organizers usually only answer questions regarding contest rules.

In the weeks leading up to the event, articles are published in the three main languages of partner organizers, English, Spanish and French, providing participants with recommendations on team formation[1], creativity guides[2], recommendations for efficient time management[3], etc. The purpose of the publications is to aid teams during idea generation, selection, concept development and pitch making.

[1] Dubois, M. (2015,02) The Importance of Team Preparation for an Ideation Session, http://substance-en.etsmtl.ca/team-preparation-ideation-importance/ accessed October 20, 2016

[2] Dubois, M. (2015,02) A Creativity Guide for Your Ideation Sessions http://substance-en.etsmtl.ca/creativity-guide-ideation-sessions-2/ accessed October 20, 2016

[3] Dubois, M. (2015, 03) The 24 hours of innovation: Montreal’s secret recipes to win! http://substance-en.etsmtl.ca/24-hours-innovation-official-guide-international-competition/ accessed October 20, 2016

Different to other innovation contests with longer durations, the teams participating in the challenge have a pressing time constraint. In only 24 hours, the teams are tasked with the following:

- Read and understand all proposed challenges
- Select a challenge based on team interests and competences
- Clearly define the problem they will solve from the challenge selected (scope)
- Generate ideas to solve the problem
- Develop a concept solution
- Try to make the concept solution more sustainable
- Analyze the feasibility of the solution and benchmark to solutions in the market
- Prototype (the prototypes range from basic sketches to 3D printed models)
- Create a 2-minute pitch video
- Complete team registration and submit video

At the end of the 24 hours, the videos are evaluated by a local jury in the sites where there is a partner organizer (participants on site). The main site at the ÉTS campus holds a special jury to select a winner among all other participants (online participants). Each local site winner participates in the international jury evaluation, where three international and five regional prizes are awarded. The local and international juries use the same evaluation grid to assess the videos, grading innovation and creativity, analysis of scientific and technical information, quality of presentation and eco-responsibility (shown in Table 4.2).

Table 4.2 Evaluation grid used by local and international juries at the 24 hours of innovation competition

Innovation and creativity	Analysis of scientific and technical information
The concept is innovative	The concept is applicable and relevant
The team shows creativity	Process accurate and logical design
The team used avant-garde technologies	Solid theoretical basis and documented
The concept has a positive impact on society	The concept is financially feasible
The concept is aesthetically pleasing	The team focused its target
The concept differs from existing products	The concept is technically feasible
Quality of presentation	**Eco-responsibility**
The presentation demonstrated structure and organization	The concept presented meets the sustainable development
Presenters they captured the attention of the jury	The concept uses minimal hardware resources
The hypothesis of the problem is formulated	The concept is energy efficient
The illustrations are clear and relevant	The concept has a defined life cycle and sustainability in mind

All solutions uploaded to YouTube[4] for anyone to see, so everyone can benefit from the ideas generated. Companies and organizations proposing the challenges are encouraged to get in touch with the teams to develop ideas further, while some teams have formed startups from the resulting concept.

4.4.2 Challenge selection

For this research, the competition of 24 hours of innovation in its May, 2016th edition was selected to test the hypothesis of keyword input from data to support bisociation. Because

[4] Official "24 hours of innovation" YouTube channel https://www.youtube.com/user/24hinno accessed October 20, 2016

challenges vary in complexity and constraints, previous to the competition we located four challenges with similar objectives to be able to prepare the data and compare the results:

- Type A: Challenge 1 has similar complexity and constraints as challenge 3, as the objective is to conceive a device that performs a specific task, with constraints related to communication, use of renewable energies and mobility.
    - Challenge 1 - Taking water samples from the river: How to take regular water samples on several strategic points along the St. Lawrence River using autonomous technology powered by a (or various) renewable energy source(s)?
    - Challenge 3 - Examine, clean and repair damaged sewer pipes: Develop a non-polluting, "intelligent", autonomous system that examines, cleans and repairs the cracks on the concrete pipes that carry sewage to water treatment plants.
- Type B: Challenge 2 and 4 can also be compared, as they are both related to a form communication with the public and continuous information updates.
    - Challenge 2 - Notifying the population about the quality of the river's water: How do you inform population in real time about the water quality of the St. Lawrence River and the risks of contamination due to sewer overflow?
    - Challenge 4 - Recognize and list the river's rare or invasive species: How can we allow citizens to recognize and geo-index rare or invasive species (plant or animal) from the St. Lawrence River's ecosystem in order to protect or control them?

4.4.3 Data preparation

To demonstrate the usefulness of providing participants with keywords to spark new solutions, a process was followed to determine the relevant keywords. For challenges 1 and 2, one of each type, a data analysis was performed, after which keywords were selected to promote bisociation. Challenges 3 and 4 were left as the rest of the challenges.

The software used to analyze the data is IPMetrix[5], which provides semantic analysis and information cartographies (IP Metrix Solution - TKM, 2016). The software was selected after comparing with other patent mining and data analysis tools. It was found that IPMetrix was easier to navigate to search for patents and scientific literature to build a map of a given domain of knowledge, and its semantic analysis tool facilitated the identification of relevant clusters and keywords (for example, other tools required manual tagging to identify clusters). IPMetrix is developed by TecKnowMetrix, French company founded in 2004 as a spinoff from research group in innovation economics at the University of Grenoble.

Using this specialized data analytics software, we extracted patents from PatBase, a world-wide patent information provider, to identify the current and past solutions for specific issues in the domain. The data analysis process was performed as follows:

1) Identify keywords from the challenges - To build a domain specific knowledge base, we first identified the relevant expressions in the challenge:
   a) Challenge 1 - Taking water samples from the river: How to take regular water samples on several strategic points along the St. Lawrence River using autonomous technology powered by a (or various) renewable energy source(s)?
   b) Challenge 2 - Notifying the population about the quality of the river's water: How do you inform population in real time about the water quality of the St. Lawrence River and the risks of contamination due to sewer overflow?
2) Search for patents and scientific articles accessible from the data mining tool - For each challenge, a keyword search was performed in the IPMetrix software. The patents resulting from the search were then uploaded to the tool. For example, for challenge 1,

[5] IPMetrix presentation on TKM website http://tkm.fr/en/ip-metrix-solution.php accessed October 20, 2016

the following keywords were selected: water sampling, water analysis, water collection, water sampling procedures, autonomous collection, water quality sensor. Figure 4.1 is a screenshot of the patent search results from the IPMetrix PatBase API.

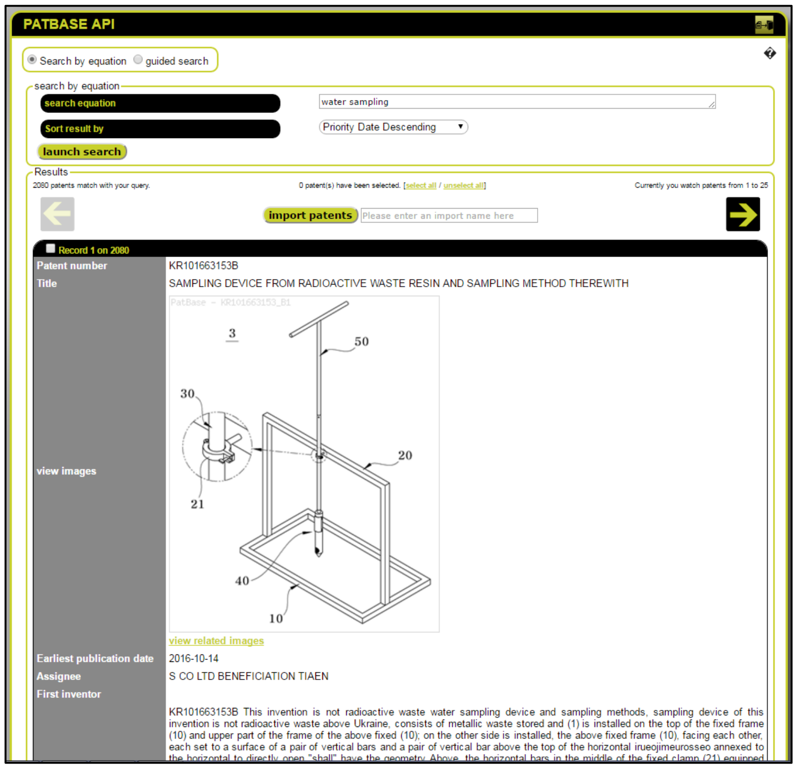


Figure 4.1 Screenshot of IPMetrix patent search results

As the first round of search provided information in the semantic analysis as to other terms that might be incumbent to the search, it was decided to search iteratively to include the new terms by applying common Boolean operators (“and”, “or”, “not” or “and not”). A total of 4303 patent families were found for challenge 1, and 1504 for challenge 2. Figure 4.2 depicts the semantic analysis used to identify the common expressions in the patents.

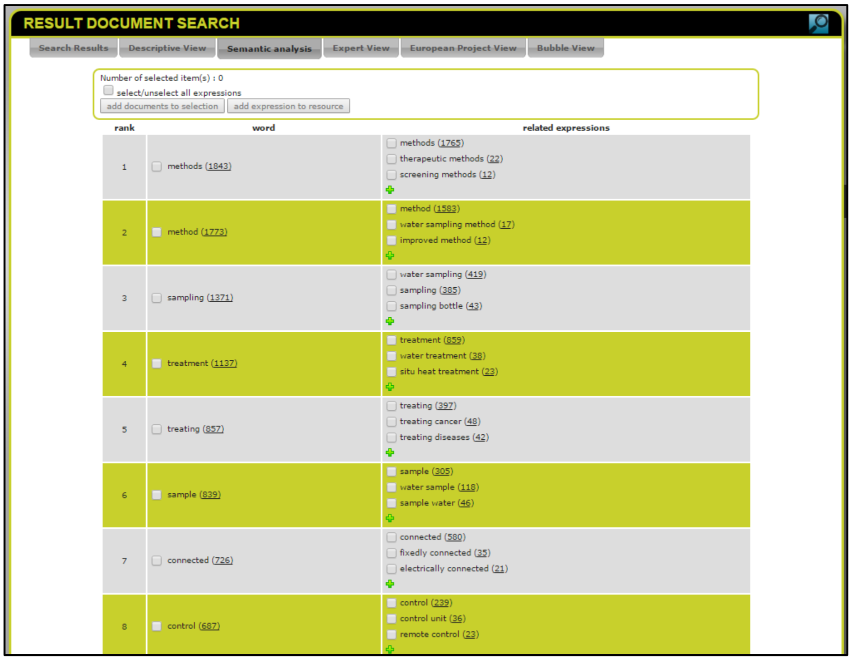


Figure 4.2 IPMetrix semantic analysis

3) Cleaning the results - The IPMetrix tool automatically cleans the data for inconsistencies, normalizes the content of patents to facilitate processing.
4) Data visualizations to identify clusters - The Voronoi diagram visualization in IPMetrix allowed us to identify the clusters in patents. Voronoi diagrams help visualize information by mapping nearest points in a plane (Aurenhammer, 1991). Given that to bisociate it is necessary to combine or place new knowledge in a different context, this visualization can help identify far away clusters in a domain. Each cluster groups related patents with keywords, which help identify different technologies, materials and actions in the domain. An example for the Voronoi diagram for challenge 1 can be seen in Figure 4.3.

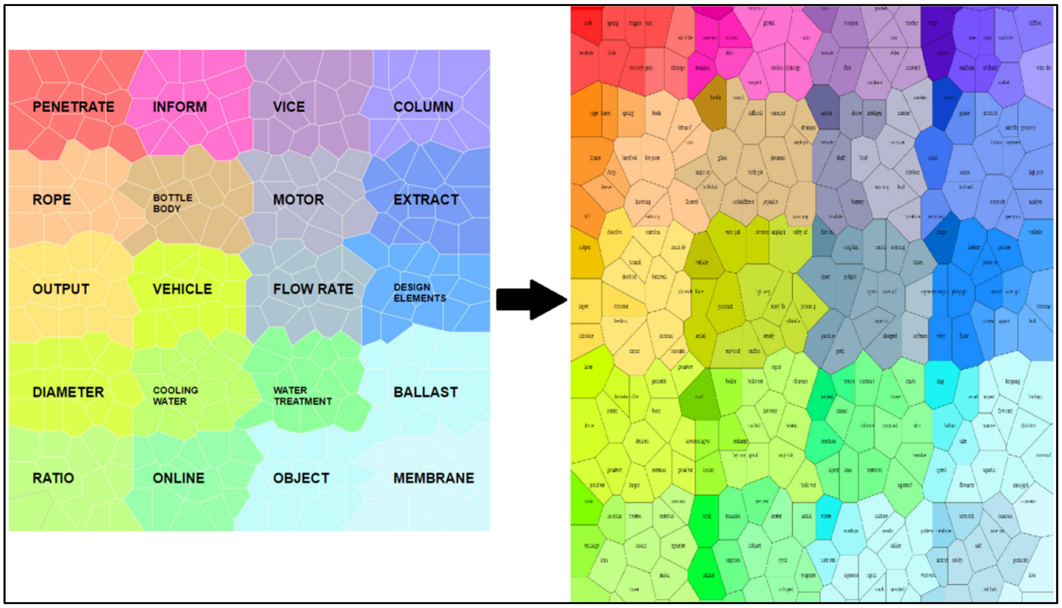


Figure 4.3 Voronoi diagram for challenge 1 knowledge domain

5) Identify prominent keywords for each cluster - To avoid overwhelming participants, the data visualizations for each challenge were studied, and keywords deemed "disconnected" were selected to be published along with the challenges so as to promote bisociation. Twenty-five keywords were selected by the researchers for challenge 1; this number was eleven for challenge 2. The keywords reflected different types of technologies and functions which exist in the domain of knowledge analyzed, and were identified as the most prominent in the clusters, which signals interest and needs.

Because of the limited amount of time participants have to generate a new idea, the results from this data analysis were published along with the descriptions of the selected challenges on the day of the contest to support inventive solution design and avoid issues with learning to use the tool and lack of time to make sense of the results. All participants had access to a public folder with the description of the question, along with the email of the key contact person in the organization proposing the challenge.

4.4.4 Analysis of team solutions

Once the contest was finished, we were able to tally the number of teams who selected each of the challenges compared in this case (challenges 1, 2, 3 and 4). Over 212 participants composed the 34 teams working with the four challenges: challenge 1 was selected by 15 teams, 10 teams selected challenge 3; challenge 2 was selected by 5 teams, and challenge 4 was selected by 4 teams. A summary of the teams and number of participants can be seen in Table 3.

Table 4.3 Summary of teams, participants for each challenge

Challenge	Keyword cues	Number of teams	Number of participants
Type A - Challenge 1. Taking water samples from the river	**Yes**	15	+93 1 team did not specify
Type A - Challenge 3. Examine, clean and repair damaged sewer pipes	No	10	+55 2 teams did not specify
Type B - Challenge 2. Notifying the population about the river's water quality	**Yes**	5	36
Type B - Challenge 4. Recognize and list the river's rare or invasive species	No	4	28

We first measured variety as a score counting the different types of solutions proposed that fulfill the objective of each challenge. For this case, we identified two levels where the variety can be measured, first, for the type of solution, and second for the elements that resolve the constraint proposed in the challenge. Figures 4.4, 4.5, 4.6 and 4.7 show the screenshots depicting the types of proposed solutions to each challenge.

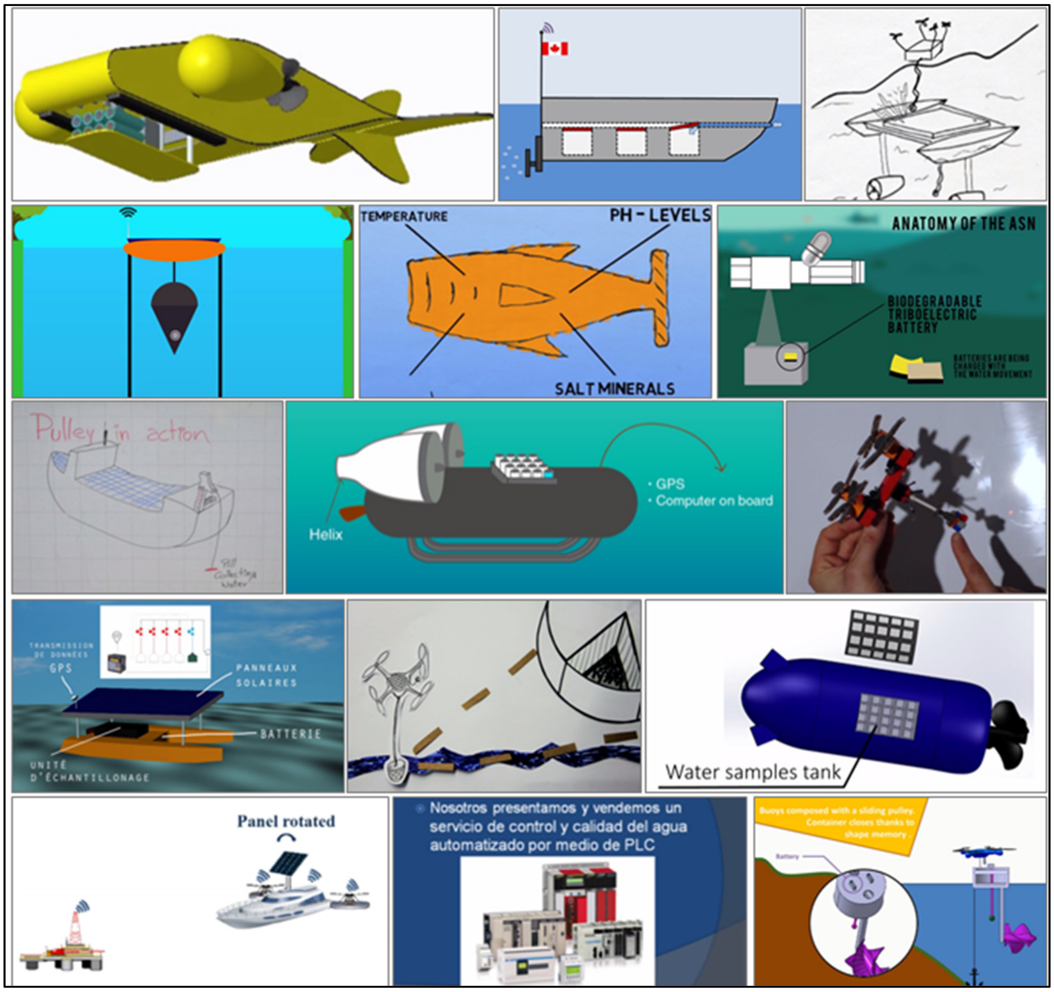


Figure 4.4 Screenshots for Challenge 1 (Type A) - Seven solution types

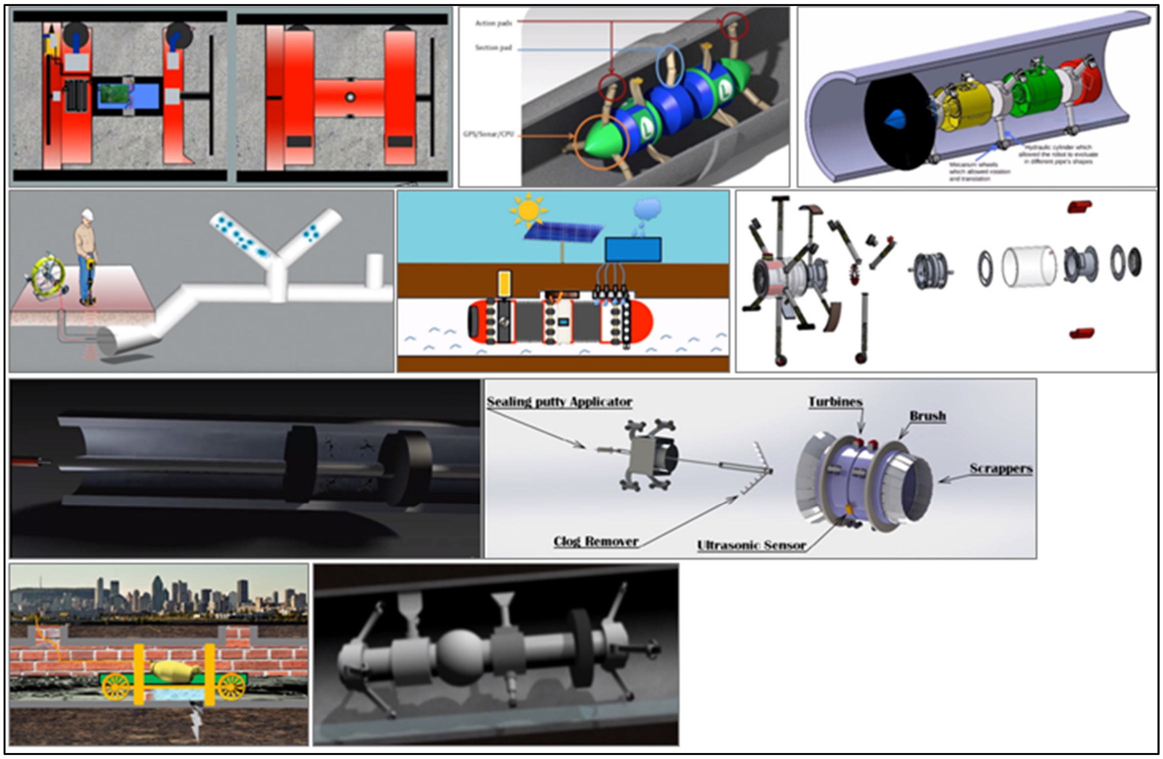


Figure 4.5 Screenshots for Challenge 3 (Type A) - Four solution types

Figure 4.6 Screenshots for Challenge 2 (Type B) - Five solution types

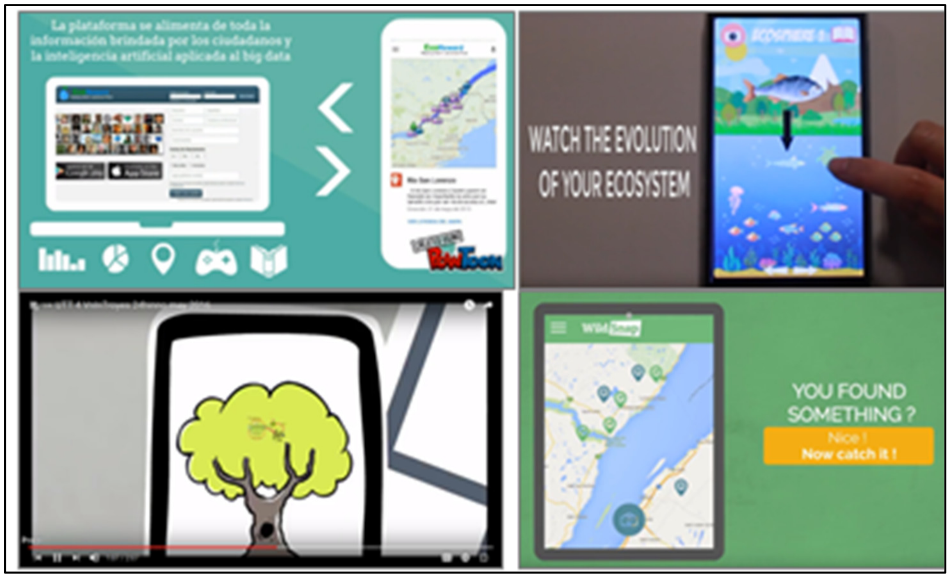


Figure 4.7 Screenshots for Challenge 4 (Type B) - Two solution types

Table 4.4 summarizes the types of solution found for each challenge, based on the description:

Table 4.4 Summary of types of solution for each challenge

Challenge	Keyword cues	Solution types	Total solution types
Type A Challenge 1	Yes	floating, flying, hybrid floating-flying, submersible, hybrid submersible-vehicle, fixed stations, service	7
Type A Challenge 3	No	propeller, wheels, turbine, probe	4
Type B Challenge 2	Yes	buoys, water filter, water dispenser, interactive mirror, information panels	5
Type B Challenge 4	No	game, simulation	2

To better understand the impact in the comparable constraints, the elements from each solution were abstracted from the verbatim description, Tables 4.5, 4.6, 4.7 and 4.8 contain the team name, video identifier (to search for the YouTube video, add the identifier after https://www.youtube.com/watch?v=, for example, to watch the first video on the list, the complete address would be https://www.youtube.com/watch?v=5fU6xjrGD50), location of team, number of team members, number of elements and list of elements in solutions.

Table 4.5 Team solutions for Type A - Challenge 1

Team Name	YouTube video identifier	University / Country	Team members	Number of elements	Elements
Le Don de Dieu	5fU6xjrGD50	ÉTS CAN	5	6	Autonomous boat + renewable energy + GPS + water sampling compartments + robotic collection station + data transmission
Alpha	D89Fbb6XC1A	UTPL ECU	NA	5	Autonomous boat + renewable energy + GPS + water sampling capsules + collection & recharging station
Stream team	DNyZFI20dWU	ESIPE FRA	6	5	Submarine + water sampling compartments + collection & recharging station + analysis in station + data transmission
CariOutan	FZ5bnLq_S5M	NTU SGP	5	5	Autonomous boat + renewable energy + GPS + water sampling compartments + collection & recharging station
Fishalyzer	30fQ1cGY-qc	RWTH Aachen DEU	6	4	Autonomous mini submarines + renewable energy + sensors + collection hubs
Paos de Queijo	A6dvbhe5P3o	USP BRA	4	4	Drone + water sampling attachment + image processing
Food 4 Thought	zAVMhJJ4rrw	Antel URY	7	4	Buoys + renewable energy + submersible water sampler + data transmission
FSI Team	bGXHQgnMSBE	UMBB DZA	8	4	Autonomous boat + renewable energy + GPS + water sampling detachable drones
Grupo110	s2mATN7fSQg	UNS ARG	6	4	Autonomous amphibian submarine
Jane's Addiction	MVwXRvm_pQY	UNICEN ARG	10	3	PLC + renewable energy + service
RiverRider	4FH2sj-eIcQ	UTT FRA	7	3	Submarine drone + GPS + water sampling compartments

Table 4.5 Team solutions for Type A - Challenge 1 (continued)

Team Name	YouTube video identifier	University / Country	Team members	Number of elements	Elements
Solucion	G5P7fRslUVs	UNS ARG	8	3	Fixed underwater sampling stations + hydrophones + sonar communication system
Vik team	bTKMXJs6rj4	UTBM FRA	9	3	Fixed buoys + drones + collection station
NTU Innovators	r3RCPBijrEY	NTU SGP	6	3	Autonomous vehicle + smart container + collection & recharging station
Samply	b0CebD7V3lc	RWTH Aachen DEU	6	3	Floating collection & recharging platform + renewable energy + drones

Table 4.6 Team solutions for Type A - Challenge 3

Team Name	YouTube video identifier	University / Country	Team members	Number of elements	Elements
AllNightLong	7MeqhvQlY_E	UTT FRA	6	6	Rotating brushes + laser + suction cup + gum applicator + propeller + wifi
Esifourmis	MHq-zn7MI14	UPEM FRA	NA	5	Brush + scrappers + ultrasonic sensors + turbines + sealing applicator
Italian Plumbers	I8HQe160ll4	UNIPD ITA	9	5	ultrasound sensor + data analysis + repair unit + sealing diaphragm + propulsion system
The Plumbers Society	s77LgIft9Es	UTBM FRA	8	5	Robot + Ultrasound sensor + suction cups + rotating brushes + communication
A.R.S	9DpNNT9loOg	ÉTS CAN	5	4	RF comm module + micro controller + propeller + alternator
BetesMobiles	l3sPIzPcIak	ÉTS CAN	6	4	Sonar exploration ball + mobile reparation unit
The D'EauCtors	vzC8IXlOL7s	ESIPE FRA	5	4	Rotating brushes + scanning laser + repairing bacteria + application

Table 4.6 Team solutions for Type A - Challenge 3 (continued)

Team Name	YouTube video identifier	University / Country	Team members	Number of elements	Elements
INO-STARS	KnGKg2A8MU U	UMBB DZA	9	3	Sonic waves + sealer + cement
Los Calapatitas	JDskcwvTFnM	PUCP PER	7	3	Water pressure + image analysis + resin applicator
Solución Alcantarillado	ujC1PBY50OE	UTP PAN	NA	2	Sonar + calcite creating bacteria

Table 4.7 Team solutions for Type B - Challenge 2

Team Name	YouTube video identifier	University / Country	Team members	Number of elements	Elements
APPANKO	f1-FiR52l5w	UV MEX	9	5	Buoy + analyzing module+ led lights + application + website
Los Tlacuaches	0corD-2ZbYY	ITESM MEX	8	5	Floating platform + sensors + radio transmission + application + physical notification
FUN	2RVPFupEpdk	PUCP PER	6	4	Sensors + display + water dispenser + awareness campaign
Kofola Team	8USVLhhPq_Y	ČVUT CZE	7	2	Connected indicator for water tap
CreativeMin ds	8h2aguYroSo	UNFV PER	6	2	Informative videos + displays

Table 4.8 Team solutions for Type B - Challenge 4

Team Name	YouTube video identifier	University / Country	Team members	Number of elements	Elements
Les pommes croustillantes	_aC64UuLSk0	UPEM FRA	6	5	Application + database + image recognition + crowdsourcing + gamification
VnInTroyes	74x-QuQGZ3g	UTT FRA	8	5	Application + database + image recognition + crowdsourcing + gamification
F10	3m20wV2HjaE	Antel URY	3	4	Big data + crowdsourcing + artificial intelligence+ gamification
ECOSIM	ymsY1y6PJlo	City U HKG	11	3	Simulation game + shape recognition + gamification

The solutions were analyzed to assess the variety of the solutions. To arrive at the variety of solutions, we consider the objective of each challenge type: Type A challenges had constraints related to communication, energy source and mobility, while Type B Challenges had constraints related to communication with the public and continuous information updates. Each element in the solution was classified into the branch related to the challenge constraint, for example, in challenges 1 and 3, the device was required to move; all elements that enabled the mobility function were grouped under that branch.

All elements were mapped into a solution tree, which enabled a side by side comparison the variety of solutions by comparing the number of different elements for each constraint. Figure 4.8 maps the elements of the solutions into a tree of comparable constraints for Type A challenges, the branches where teams proposed more diverse elements are marked with a gray box. Figure 4.9 maps the elements of solutions for Type B challenges, the branches where teams proposed more diverse elements are also marked with a gray box.

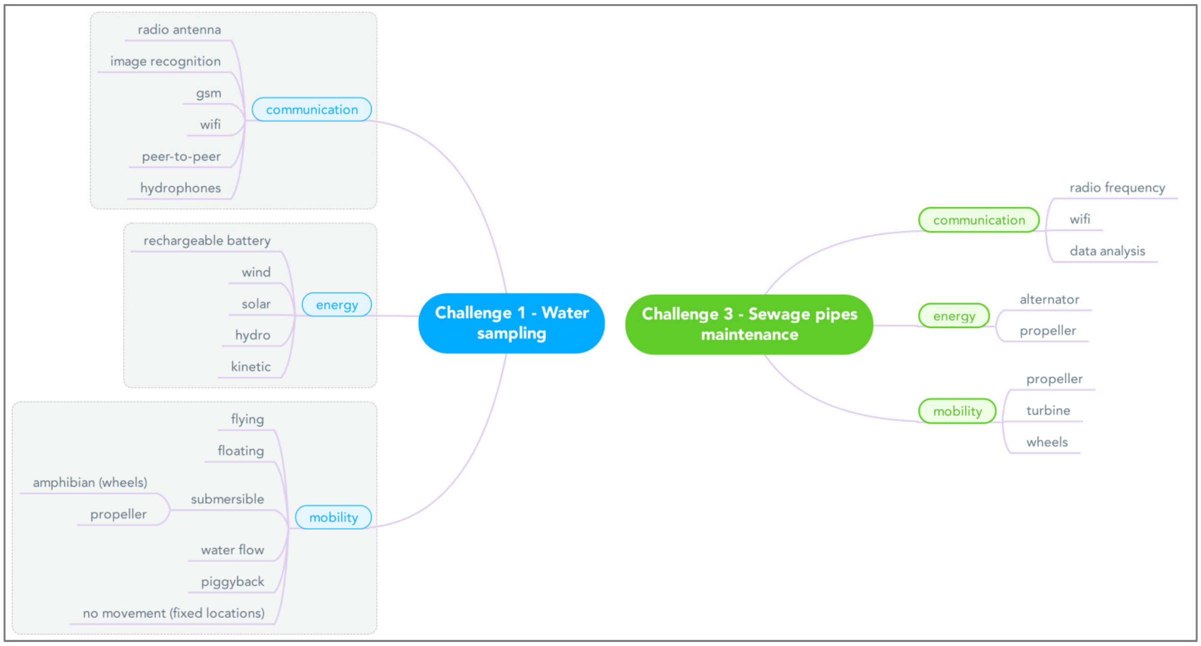


Figure 4.8 Tree comparison for solutions to Type A challenges (1 in blue, 3 in green)

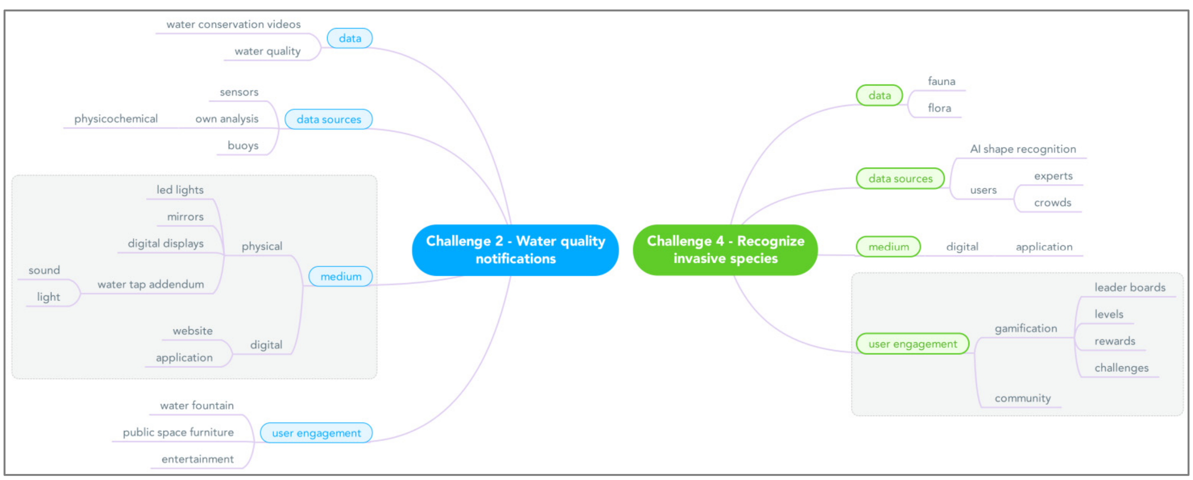


Figure 4.9 Tree comparison for solutions to Type B challenges (2 in blue, 4 in green)

4.5 Results

The variety in the solution for challenges where an input from data mining was provided was increased by 75% for challenge 1 (7 types of solutions v. 4 types of solutions) and a staggering 150% for challenge 2 (5 types of solutions v. 2 types of solutions), as can be seen in Figure 4.10. The assessment of the types of solution for each challenge indicates that teams with access to selected keywords from the data analysis were able to provide more varied solutions overall, as indicated in Table 4.4.

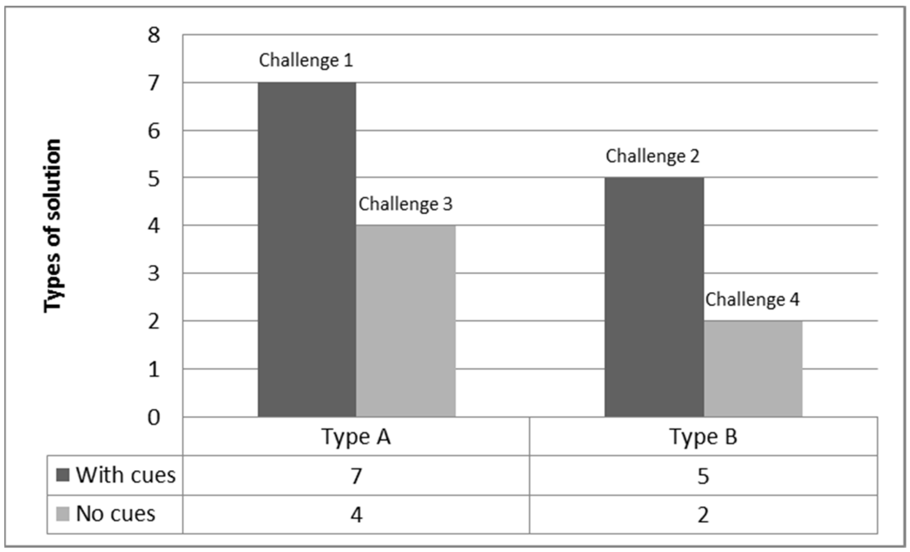


Figure 4.10 Type of solutions for Type A and Type B challenges

Following the classification of the elements for comparable constraints, it was clear to see that teams with access to keywords to support their idea generation were able to generate more diversity in the elements, as shown in Figures 4.11 and 4.12. To obtain these numbers, we first typed a complete verbatim description of the concept solution proposed by each team from the videos submitted.

We then mapped the types of solution and compared the variety of elements per branch shown in the maps, Figures 4.8 and 4.9. Each branch was expanded with the elements proposed in each solution. This process was described more in detail in the “Analysis of team solutions” section of this article. For type A challenges, all branches of the comparable constraints show more diversity in the elements for teams using a keyword input, by more than double. For type B challenges, only one branch has quantifiably more elements than teams with no keyword input support, and one branch for the teams with input has fewer elements. Nonetheless, it could be argued that the branch with more elements is defining for the solution, thus is a better indicator for this exercise.

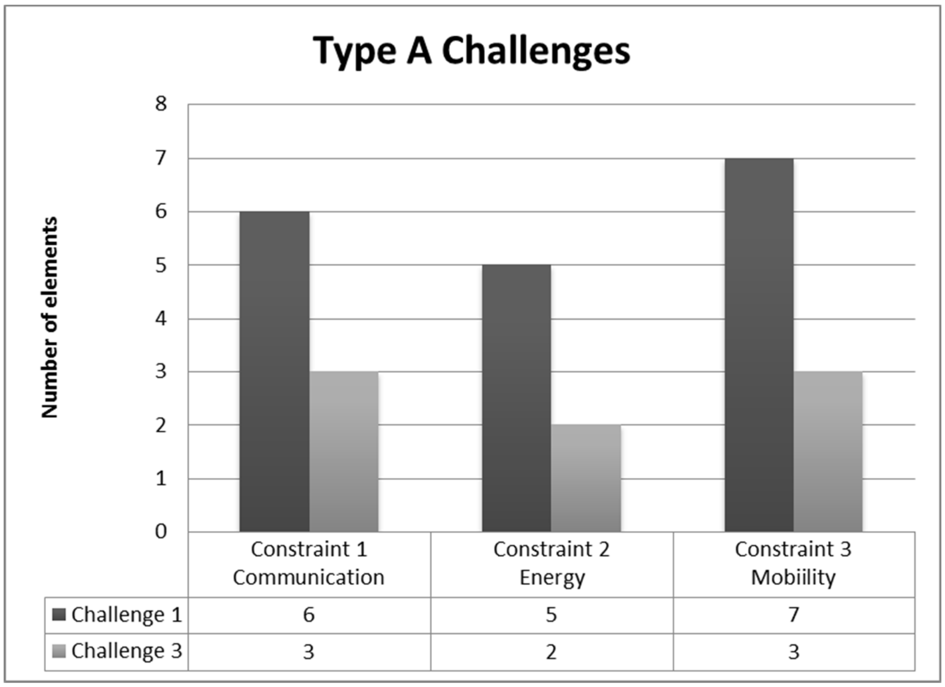
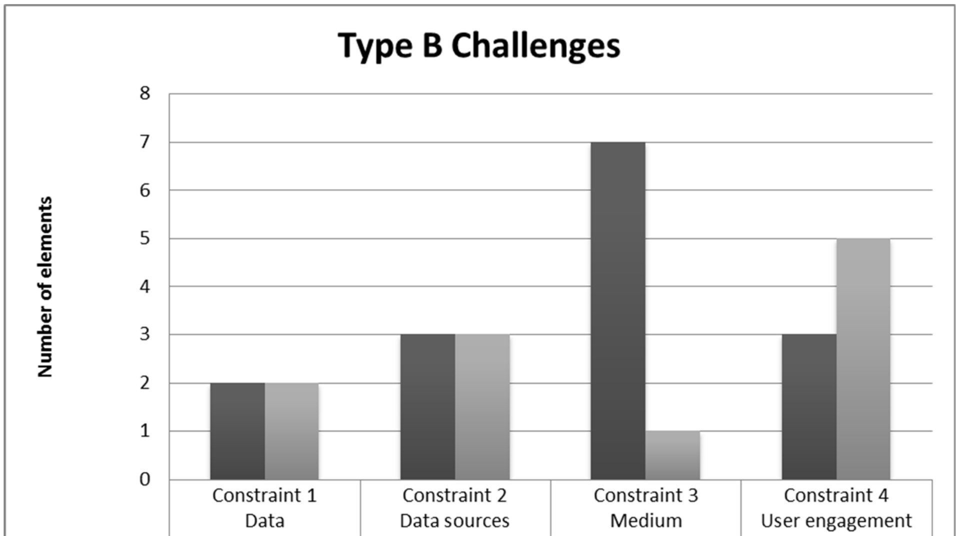


Figure 4.11 Elements per comparable constraint in Type A challenges



	Constraint 1 Data	Constraint 2 Data sources	Constraint 3 Medium	Constraint 4 User engagement
Challenge 2	2	3	7	3
Challenge 4	2	3	1	5

Figure 4.12 Elements per comparable constraint in Type B challenges

One unexpected indicator to the usefulness of having additional input is the fact that more teams selected the challenges with the keywords to bisociate, a 30% increase for challenge 1, and a 20% increase for challenge 2. While selecting the challenge, teams usually perform an idea generation session for a pre-selection of challenges to better explore their potential inventiveness in the subject. Considering that more teams selected the challenges with additional input from data mining can indicate that the cues were useful from the first stages of the competition; it could also indicate that giving participants additional input makes the decision easier, as they have more data to work with.

Finally, an even more encouraging indicator was the fact that the grand winner of the competition was one team working with challenge 1, and the second place was a team working with challenge 2, meaning they both had access to the keywords input. This in turn can denote that the solutions presented by the teams were deemed more novel than the counterparts by a panel of external experts. The winning team obtained 436 points out of 600 (100 points per judge, 6 judges), 17 more points than the next team, which in turn obtained 8 more points than the third place. Because the evaluation used by judges emphasizes creativity and feasibility, the results suggest that the use of the keywords coming from the analysis of the domain of knowledge not only provided hints for creative solutions, but also more grounded on feasible solutions.

4.6 Discussion and conclusion

We attempted to provide the benefits of additional input extracted from data analysis during an innovation contest to stimulate bisociation, the creative combination of existing knowledge within the participants and in the domain of the problem, into a novel solution. The results suggest that having additional data in a time-pressed environment can be useful for participants from the moment of selection, as more teams were drawn to those challenges were the keyword input was provided. It could be argued that there are other explanations for teams to select the challenges with additional input, be it because it provides them with a sense of certainty, or save time in their search for information.

As can be seen, the variety of the concepts proposed by the teams indicates that the keyword cues provided were useful for idea generation. In the comparable constraints, the teams who elected to work with the keywords created more diverse solutions by an important margin. It supports the notion that an input extracted from data can have a positive impact for creative teams in the context of time-constrained idea generation sessions.

More work is needed in order to adapt this process for enterprises and organizations looking to enhance their creative processes through the use of data. It can potentially be used as an ongoing support for product and service design, by inspiring teams with the use of keywords as an input. This can be challenging for organizations with no access to a specialist who can perform the tasks of identifying the right data sources, cleaning and selecting data.

4.7 Design cycle evaluation

NOTE: This section does not appear in the article, it is meant as a conclusion and transition for the next chapter of the thesis.

The observations and analysis of results in this experience, Case 3, show that teams involved in an EDP benefited from being exposed to keywords extracted from the KDD process performed by an external actor. Teams were able to combine (bisociate) their existing knowledge with data from the domain knowledge base. This is demonstrated by the diversity in solutions and elements which make up the proposed concepts.

Teams working without access to pre-selected keywords from the knowledge domain resort to common solutions, as the exploration space for new solutions can be infinite, and thus unmanageable in a short-time contest.

For future cases with similar time constraints, it is recommended to perform a KDD process in advance, so teams can benefit from the condensed information.

CHAPTER 5

EFFECTS OF INFORMATION CUES FROM KNOWLEDGE DISCOVERY IN THE EARLY CREATIVE STAGES OF ENGINEERING DESIGN

Ma-Lorena Escandón-Quintanilla[1], Mickaël Gardoni[1,2], and Patrick Cohendet[3]

[1] École de Technologie Supérieure, Montréal, Canada

[2] Institut National des Sciences Appliquées, Strasbourg, France

[3] HEC Montreal, Montreal, Canada

This article was submitted to the Journal of Engineering Design on March, 2017

Chapters 3, 4 and 5 presented three cases where KDD was applied to support the EDP of teams in various contexts, with different scopes and involvement of participants. While the observations and results of each case are interesting in a self-contained manner, a meta-analysis of the results and the implications of the use of data as a creative input can provide more understanding.

This article attempts to conciliate the findings from all three cases, and explain how the different teams used the data and the analytics tool for idea generation in the various stages and contexts of an EDP, and what we can learn from this.

Abstract

Engineering teams tasked with finding a novel solution will usually search for information from external sources to complement knowledge within the team. This process is time and resource consuming, and overwhelmed teams tend to go back to improving existing designs with known solutions. This paper presents the analysis of three cases, totaling 45 teams and over 275 participants, where knowledge discovery from databases (KDD) is applied in the early creative stages of the engineering design process (EDP), in the context of higher

education. Teams were given varying levels of access to a data mining tool to explore data from patents. By integrating the exploration of data, teams can discover information, and combine it with existing knowledge to increasing the novelty of ideas generated in early stages of the design process. The use of data was found to be useful in all three cases, it appears to help teams find new possibilities, generate more diverse ideas, and develop the design. The findings suggest that teams with restricted access to data performed better in terms of novelty, but data is also useful at other stages to generate incremental improvements.

5.1 Introduction

Teams of engineers are often tasked with finding solutions to diverse problems, with the added difficulty that the solution should be novel and unlike current solutions. Perhaps unwillingly, the natural path for a team will be to look for the current solutions and try to combine and improve known solutions (Fleming & Szigety, 2006).

Experienced engineers know to search for information early in the process to save time later (Atman et al., 2007). However, time-constrained teams don't have the time or resources, and often resort to known solutions. In order to promote novel combinations for original ideas, it is important to provide teams with external information to complement their knowledge (Karlsson & Torlind, 2014).

It has been reported that teams in time-constrained engineering challenges research the internet to find additional information (Jiménez-Narvaez, Dalkir & Gardoni, 2012). And while there is indeed a great amount of information available on the internet, there are many downsides to using it as a source for idea generation: it is effort and time-consuming to condense and find the relationships between data, there is no standard of validity, as anyone can publish, many sites do not require sources or proof of veracity, and there is a continuous stream of new information, the relentless pace and volume of which can overwhelm users.

The aim of this study is to observe the effects of using a data mining tool and the data itself for knowledge discovery (KDD) throughout the engineering design process (EDP), and how this impacts the early creative stages, particularly idea generation. In the first case, the KDD aimed to support participants in the identification of problems, in the second case, the purpose was to support participants in the creation of an innovative idea, and in the third case, the purpose was to support participants in the development of a concept. Each case provided learnings that were integrated to the subsequent cases.

This article is structured as follows: the first section presents a theoretical background to introduce the concepts of knowledge discovery from databases (KDD) and engineering Design process (EDP). After, we present the three cases where engineering groups in diverse contexts were tasked with proposing problems, ideas and concepts. The article ends with a discussion of the findings, and the implications for future research in the area.

5.2 Theoretical background

Faced with ever-increasing amounts of data, it has become a challenge for most teams, individuals and even organizations to be able to manage the information, and to obtain a useful outcome (Fayyad, Piatetsky-Shapiro & Smyth, 1996, George, Osinga, Lavie, & Scott, 2016). Fortunately, data mining tools and techniques have evolved to help cope with the rising needs for data analysis. Data mining tools and techniques have allowed for faster processing and analysis of large amounts of data; they automate the processing, and leave the interpretation to the user (Fayyad, Piatetsky-Shapiro & Smyth, 1996, Baesens, 2014).

5.2.1 The knowledge discovery from databases (KDD) process

The process of collecting, processing and analyzing data to uncover knowledge and obtain insights is known as Knowledge discovery from databases (KDD), a term coined by Piatetsky-Shapiro (1991) to indicate that the desirable result of the data analysis is knowledge; the process can be seen in Figure 5.1. Here, the difference between data, information and knowledge is worth noting: data is a piece of information, which is not useful by itself without context; information is aggregated data, which can be useful with context; and finally, knowledge is usually the interpretation of information to provide a high-value insight (Ackoff, 1989).

Before the process begins, the user or users must attempt to define the objective as best as they can, to select the data sources and algorithms that will provide the results that fit the problem. The KDD process is iterative, so the users can go back to previous stages if necessary. The process then includes the following stages (Fayyad, Piatetsky-Shapiro & Smyth, 1996, Siau, 2000, Baesens, 2004):

- Data selection, which refers to the actions of defining the information sources to be included in the database.
- Data preprocessing is cleaning the data to remove noise, such as repeated, incomplete or inaccurate data, and defining a strategy for inconsistencies.
- Data transformation, in which the content in the database is converted so it can be mined.
- Data mining is the application of statistical methods and algorithms to find patterns in the data. Data mining methods are: classification, regression, clustering, summarization, dependency modeling, and outlier and change detection.

- Interpretation and evaluation: there, the user can interpret the reports and visualizations resulting from the data mining, and decide if the results can help attain the objective.
- Application: the final step always involves the users applying the newly acquired knowledge.

KDD is applied in business to assess efficiency (find bottlenecks, shortcuts, improve planning), identify market segments and design marketing strategies (i.e. with sentiment analysis), and gain competitive advantage (Fayyad, Piatetsky-Shapiro & Smyth, 1996). Some experiences have been documented previously that attempt to use data to support idea generation sessions, to infer the usages of electricity from aggregated consumption data (Dove & Jones, 2014), to help researchers keep up to date in incumbent or relevant technologies in their working domain (Müller et al. 2012), or even suggesting random Wikipedia pages to promote inventiveness (Shan, Zhu & Zhao, 2013).

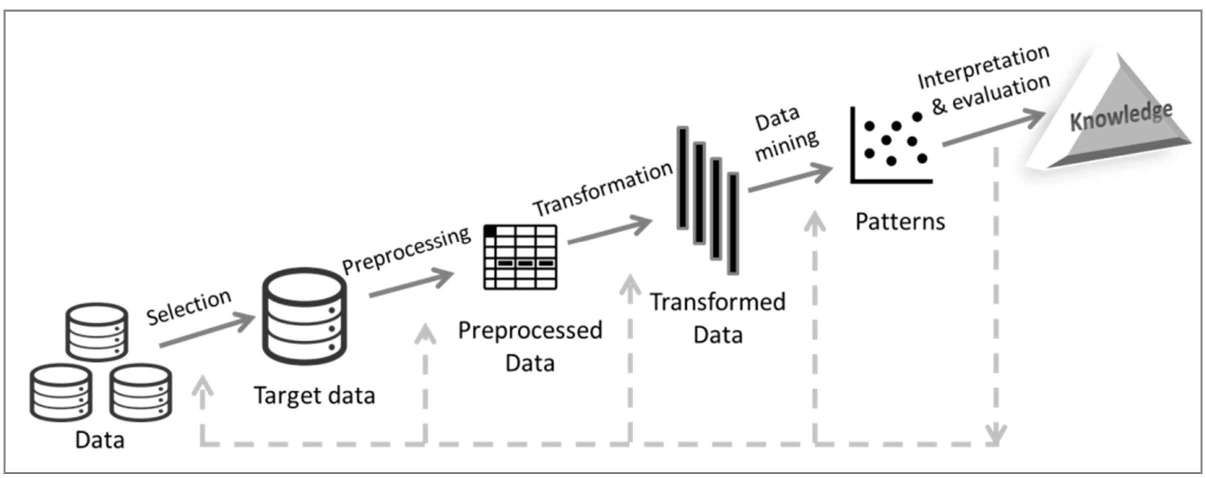


Figure 5.1 Overview of the KDD process,
taken from Fayyad, Piatetsky-Shapiro & Smyth (1996)

5.2.2 The engineering design process

The engineering design process is the process by which engineering teams develop the process, mechanism or device to solve the problem; the process is typically non-linear and iterative, as designers move between stages or activities when they are faced with an issue, or discover new information about the problem (Erta & Jones, 1996, Sim & Duffy, 2003, Atman et al., 2007).

To tackle the design of a new solution in engineering, practitioners in these fields usually follow a process that begins with the reception of a request, or the realization of a need. The design process followed by engineers resembles processes in other domains. Schneiderman et al. (2006), propose a general process of nine steps for idea generation efforts: Problem definition, information gathering, idea generation, modeling, feasibility analysis, evaluation, selection, communication and implementation.

Later, Atman and colleagues (2007) proposed the same activities for engineering design process, adding a “need identification” activity, and classifying the activities into three stages: Problem scoping, development of alternative solutions, and project realization. This process is shown in Figure 5.2. The scope of this article pertains to the first two stages of the engineering design process:

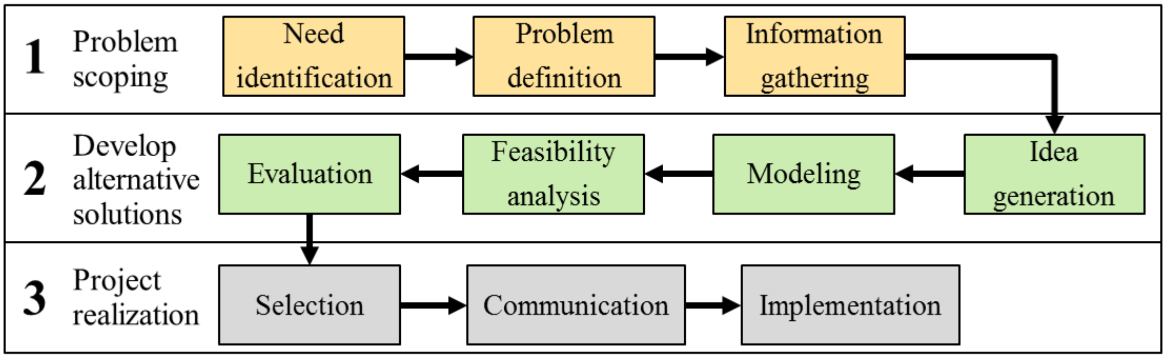


Figure 5.2 Engineering design process,
based on Schneiderman et al. (2006), and Atman et al. (2007)

5.2.3 Measuring the creative process

Judging an idea can be indeed difficult. An idea is not a complete solution, it is a notion that is usually not patentable; a complete solution is most likely a combination of ideas that have been advanced. There are many steps between the generation of the idea and the actual implementation which can make the idea evolve and change (Karlsson & Torlind, 2014), rendering the assessment of value of an idea problematic. Therefore, many authors opt instead to assess the creative process, which can be measured quantitatively or qualitatively, and can measure diverse aspects (the session itself, the perception of participants, the productivity, etc.). Creative sessions have been measured differently in previous literature; the following examples include a measure for idea quality, which is deemed subjective by the authors of the present work:

Table 5.1 Measures used to evaluate the results of a creativity session

Measure	Category
Quantitative	
- Number of characters of a conclusion (Munemori & Nagasawa, 1996) - Number of ideas (Ardaiz-Villanueva et al., 2011, Glier et al., 2011, Graetz et al., 1997, Jung, Schneider & Valacich, 2010, Munemori & Nagasawa, 1996, Parjanen, Hennala & Konsti-Laakso, 2012, Wang & Ohsawa, 2013, Wodehouse & Ion, 2012) - Number of record cards / sticky notes (Gumienny et al., 2013, Yuizono et al., 2005)	Productivity
- Number of ideas shared (Graetz et al., 1997) - Number of comments (Ardaiz-Villanueva et al., 2011) - Number of chats (Munemori & Nagasawa, 1996, Yuizono et al., 2005)	Interaction
- Number of ideas evaluated (Ardaiz-Villanueva et al., 2011)	Evaluation
- Time (Graetz et al., 1997, Gumienny et al., 2013, Munemori & Nagasawa, 1996, Yuizono et al., 2005) - Number of participants (Yuizono et al., 2005)	Work session
Qualitative	
- Applicability of concepts (Ardaiz-Villanueva et al., 2011) - Novelty of concepts (Ardaiz-Villanueva et al., 2011, Glier et al., 2011, Wodehouse & Ion, 2012) - Quality of concepts / Ideas accepted (Glier et al., 2011, Jung, Schneider & Valacich, 2010, Wang & Ohsawa, 2013, Wodehouse & Ion, 2012)	Concepts
- Perceived team cohesiveness / effort (Graetz et al., 1997)	Teams
Hybrid	
- Complexity level of concepts (Ardaiz-Villanueva et al., 2011) - Detail of concepts (Wodehouse & Ion, 2012) - Variety of concepts (Glier et al., 2011, Wodehouse & Ion, 2012)	

The cases in this article were measured differently due to the scope of the EDP process, the amount of teams available to compare the results, and the availability of data by an external jury. Case 1 measures Number of ideas, Case 2 measures Complexity, Novelty, Number of participants and Variety, and Case 3 measures Complexity, Novelty and Applicability.

5.3 Study cases

This section presents the three cases where KDD was applied for different stages of the EDP, and explains how the data was selected, pre-processed, transformed and mined. The different tasks to select, preprocess, transform and analyze the data in each case were performed by either one of the researchers of this study, or the participants in the session. In this article, we will refer to the researcher who performed the KDD and EDP activities as a 'moderator', as she is serving as a buffer between the data mining tool or the data, and the teams, while not directly involved in the EDP.

Each case subsection discusses the context, the participants, the session proceedings, and the use of information. All three cases took place within a higher education context, and while the participants differed, the challenges are linked to the same topic and feed the subsequent cases. The findings at the end of each case were integrated for the next case, providing grounds for improvement regarding the use of the tool and the data as creative input. Table 5.2 presents a summary of the objective, session duration, teams and participants of the three cases.

Table 5.2 Overview of the three cases: Objective, session duration, teams and participants

Case	Objective	Duration	Participants	Number of participants	Number of teams
1	Need identification	4 hours	Members of the ETS community	15	3
2	Benchmark Novel concept Prototype	2.5 weeks (≈100 hours)	Students at the ETS Summer School on innovation and technological design	49	8
3	Novel concepts	24 hours	Participants of the 24h of innovation competition	≈212	34
			Total	≈276	45

To put the three cases into perspective, different elements of each session are distinguished, according to Adamczyk, Bullinger & Möslein's (2012) categorization of innovation contests. In table 5.3, the categories are arranged from most similar (element coincides in all three cases), to least similar (element differs in all three cases). This will enable to have a discussion on which elements might have also affected the results of the teams.

Table 5.3 Innovation contest categorization of the three sessions
(categories taken from Adamczyk, Bullinger & Möslein, 2012)

Categories (Adamczyk, Bullinger & Möslein, 2012)	**Case 1**	**Case 2**	**Case 3**
Organizer Company, public organization, non-profit…	Non-profit	Non-profit	Non-profit
Facilitation Professional, peer, mixed	Mixed	Mixed	Mixed
Participation as Individual, team, both	Team	Team	Team
Target group Specified, unspecified	Specified	Specified	Specified
Task / topic specificity Open task / low specificity, specific task / high specificity	Open task	Open task	Open task
Sponsorship / partnership Family, friends, universities, associations, industries, agencies, mixed…	University	University	Mixed
Community functionality Given, not given	Not given	Not given	Given
Replication Biannual, annual, less or more frequent	No replication	Annual	Annual
Media Online, offline, mixed	Offline	Offline	Mixed

Table 5.3 Innovation contest categorization of the three sessions
(categories taken from Adamczyk, Bullinger & Möslein, 2012) (continued)

Categories (Adamczyk, Bullinger & Möslein, 2012)	**Case 1**	**Case 2**	**Case 3**
Degree of elaboration Idea, sketch, concept, prototype, solution…	Idea	Prototype	Concept
Contest period Very short term … very long term	Very short	Medium	Short
Reward / motivation Monetary, non-monetary, mixed	No reward	Non-monetary	Monetary
Evaluation Jury, peer, self-assessment, mixed	No evaluation	Jury evaluation	Jury evaluation Peer review
Attraction (marketing / activation) Online, offline, mixed	Online	No advertising	Mixed
Contest phases (rounds) One, two, more	No contest	Two rounds	One

5.3.1 Data selection

For the three case studies, patents were selected as data sources because of their availability, the richness in the data, and the structure of the documents (authors, claims, etc.), which simplifies data pre-processing. Compared to other open sources of data, patents have the advantage of having a complete description of a solution to a problem. Authors of a patent must describe in detail what the invention does, and how it is composed. Access to patent information can help access existing knowledge in a domain, and communicate that knowledge within the design team (Trappey, Trappey & Wu, 2009).

5.3.2 Data pre-processing, transformation and mining

Because of the time constraints of each case, the participants were given access to varying levels of access to data and a data mining tool. These differences in the KDD involvement of participants were due to the different duration, scope, and time available to train users in the tool.

The software IPMetrix from French company TKM was used to gather, pre-process, transform and exploit the data to deliver semantic analysis and information cartographies. This software was selected because it specializes in the exploitation of scientific data (patents, scientific articles and technical reports) to produce visualizations and reports that help users have a condensed overview of a given technology or domain, as opposed to a web search.

5.3.3 Case 1 - Co-located teams, brief session for problem identification

The first case had the goal to find problems in the area of river conservation (climate change), access to rivers, spill-related issues, and other issues related to the welfare of rivers, particularly the St. Laurence River in Canada. The issues found during this session would then be proposed at the AquaHacking competition website for participating teams to develop solutions to a challenge of their choice.

5.3.3.1 Participants

Participants were self-selected. A call for participation was published in the school electronic bulletin; all members of the community were welcome to participate. No rewards were offered to participants for their work. From the 18 volunteers registered for the session, fifteen actually participated in the session. Three teams were formed, and were animated by volunteer Master and PhD students from a student club which regularly organizes creative sessions and creativity tools workshops at the *École de technologie supérieure* in Montreal, Canada.

5.3.3.2 Session proceedings

The activity sequence followed during this session was:

1) Welcome
2) Introduction to the topic
3) Group formation
4) Identification of elements in the problem
5) Identify connections and relationships between elements
6) Identify key issues
7) Use of data visualizations to identify new issues
8) Presentation of issues identified

5.3.3.3 Information application

Teams were provided with access to a data mining tool pre-loaded with freshwater and river related patents. In this session, the purpose was to use the results of data mining for knowledge discovery as an information input to trigger new relations. Participants had time to explore the different keywords and relationships in the visualizations and selected various keywords to combine with their previously identified issues.

5.3.3.4 Results

In the first iteration, the three teams identified 5, 3 and 5 problems respectively. After they were provided access to the data mining tool visualizations of the knowledge base created before the session, they were able to find 3, 1 and 2 additional problems. Considering the time constraints, the results are considered to be mostly positive, as the teams were able to discover further problems just by accessing a new source of information. However, participants seem to have selected keywords which supported their existing knowledge. Thus, while teams were able to find new ideas, the novelty of ideas was deemed as low. It is possible that having access to a more ample source of information earlier in the process would have enabled the teams to explore more alternatives.

The teams were not motivated by a reward, had little time, and the use of the tool was late into the process, as the teams had already been working on the task before they were presented with the data visualization. These conditions would be corrected on future sessions. The main takeaway from this experience, for the process design, is that the keywords for blending should be provided at an earlier stage, before the participants get fixated on existing knowledge.

5.3.4 Case 2 - Co-located teams, short project for concept development

The second case took place in the context of an intensive summer course on innovation, where teams of engineers selected one challenge identified in case 1 and were tasked with finding a solution within three weeks. Given that the participants had more time to explore the data by themselves, it was decided to give them an introduction workshop to learn to use the data mining tool. After the workshop, the teams had the option to continue using the tool, but it was not mandatory. The purpose was to observe how many teams would in fact utilize the tool and the data for their idea generation process, and the impact of the data exploration on the novelty and complexity of their solutions.

5.3.4.1 Participants

Participants were students in the ÉTS International Summer School on innovation and technological design. The 48 students came from 10 countries, over 20 different engineering fields, most of them at a master level of studies; only a handful of them had work experience. After the Summer school was over, it was found that half of the teams made use of the tool.

5.3.4.2 Session proceedings

The timeline for this course was the following:

Table 5.4 Timeline of course followed in case 2

Week	Lectures / workshops	Objective for participants
1	Introduction to creativity and innovation techniques	Team formation
2	Introduction to use of data for creativity, and data mining tool	Problem definition
3	Introduction to prototyping	Initial concept
4	Team coaching	Final concept

5.3.4.3 Information application

As teams had more time to explore the information by themselves, and were able to evolve the problem definition, it was decided to provide them with a training session to use the IPMetrix software. The researcher pre-loaded the database with relevant patents, according to the domain of knowledge of each challenge, and participants were free to upload additional data from diverse sources. There was no incentive to use the tool that would affect their grade; teams were able to decide whether to allocate time and resources to exploring the data or to other activities.

5.3.4.4 Data collection

Teams participating in the Summer school were asked to carry an activity journal (diary), to report the techniques and tools applied every day, along with the evolution of their project, notes and insights. The moderator did not have access to the journal until the course was finished. It was identified that half of the teams made use of the tool to explore patent data related to their problem. Teams using the data received better evaluations from the expert jury panel, comprised of university professors who mentored all teams and were not part of the teaching staff or the research team. Teams exploring the data were also able to evolve their concept, adding elements to their proposed solution. However, the novelty expected from the added data exploration did not occur.

5.3.4.5 Results

Half of the teams decided to take advantage of the tool for data exploration; this can indicate that the tool might still be complex to use, or that not all teams see the benefit of data as a creativity support. Teams using the data mining tool were able to identify more potential solutions, and were better graded by a panel of experts. However, the observed improvements could be categorized as ‘incremental innovations’. The moment in the EDP at which the students received the training and access to patent data might have influenced the fixation level of teams. And though the teams were able to make incremental innovations, the novelty was deemed low; this finding is further debated in the Discussion section. For the next case, it is proposed that the teams are provided data without access to the tool, to see if the data by itself can help the teams in the idea generation phase. This would enable teams to benefit from the data, without the need to learn to use a tool.

5.3.5 Case 3 - Distributed teams, very short project for idea generation

The international innovation competition Les 24 heures de l'innovation provided the opportunity to test the process to use data mined from patents to support creative teams with the development of an innovative solution, while correcting for the circumstances observed in Cases 1 and 2: data should be provided early in the process, teams have no time to perform the search in time-constrained projects, and the data exploration tool can be too complex to use by non-experts. For this case, the moderator found two airs of challenges also within the water conservation challenges proposed in Case 1, with similar scope and constraints to provide data to the participants of two challenges, and compare them to teams working with the similar challenges with no data support.

The challenges are unveiled at the launch of the competition with a brief description, after which the teams have 24 hours to dissect the problem, find a solution, and make a video to present their concept to the jury. The keywords from the data analysis were published along the description of the problem, thus participants had the opportunity to work with them since the beginning. In this case, no team had access to the data exploration tool.

5.3.5.1 Participants

Over two thousand participants in 195 teams took part in the competition; however, we observed the results of only the teams who worked on the selected challenges. Thirty-four teams selected to work with one of the four challenges considered for this case. The number of participants in the teams is estimated to be over 212, based on the format completed by teams at the end of the competition and average number of participants in teams in the competition, but cannot be specified as some teams only list the team leader, while other teams have intermittent participants or external collaborators at different points of the competition.

5.3.5.2 Session proceedings

The following schedule is suggested to teams taking part of the competition, but it is not mandatory:

1) Read and understand all proposed challenges
2) Select a challenge based on team interests and competences
3) Clearly define the problem they will solve from the challenge selected (scope)
4) Generate ideas to solve the problem
5) Develop a concept solution
6) Try to make the concept solution more sustainable
7) Analyze the feasibility of the solution and benchmark to solutions in the market
8) Prototype (the prototypes range from basic sketches to 3D printed models)
9) Create a 2-minute pitch video
10) Complete team registration and submit video

To be considered in the competition, teams must make sure to send their video before the event's deadline.

5.3.5.3 Information application

The objective of the database was defined according to each challenge, based on the problem statement. Two of the challenges were selected to build the knowledge base, and the results of the analysis by the moderator were published along the problem statement in the form of keywords for participants to use as input for their EDP. To be used as a baseline for the novelty, complexity and variety in the results, the two other challenges did not receive additional information from the moderator.

5.3.5.4 Data collection

All teams participating in the 24 hours of innovation competition must upload a pitch video presenting their concept solution for the selected challenge; all videos are publicly available in the YouTube page of the competition. To analyze the results, the descriptions of all solutions were transcribed verbatim, and then tagged to identify the elements of the solutions. By mapping the results of each challenge, it was observed that the solutions by teams selecting the keyword supported challenges went for more diverse solutions, compared to baseline challenges.

5.3.5.5 Results

The results of this third case were more positive than cases 1 and 2. Teams using the keywords from the data analysis as input for idea generation were able to generate more diverse ideas, more complex (included more elements in the solution) and more varied, compared to teams with no input.

5.4 Evaluation of results

As mentioned before, at each case the researchers aimed at correcting the process of the use of the tool and the information produced by the tool as input for the engineering design process. Figure 5.3 shows which actors performed the different KDD steps during the three cases.

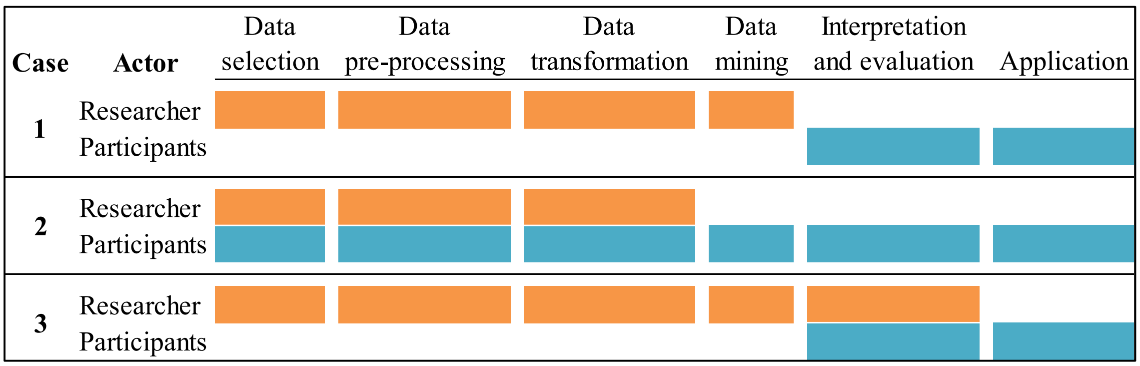


Figure 5.3 KDD steps executed by the researcher and participants in the three cases

Figure 5.4 shows the EDP activities in the three cases, and identifies the actors who performed them. An actor called 'external' was added to signal that the need was suggested by an external entity.

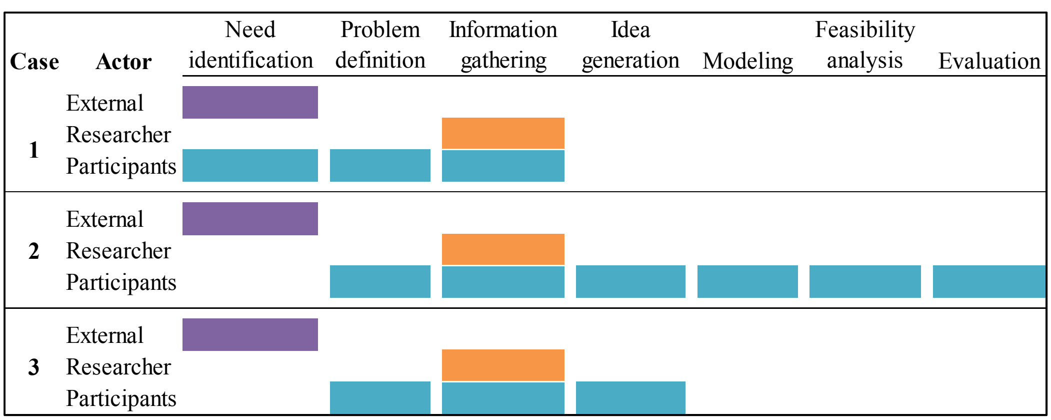


Figure 5.4 EDP activities by the researcher and participants in the three cases

Each case had different scope, involvement by researchers and participants regarding the extraction of information from the data mining tool, context and duration. For this reason, it is not possible to apply the same evaluation in all three cases (a summary of the metrics used in each case is shown in Table 5.5).

Table 5.5 Metrics applied to assess the output of the sessions

Case	Type of measure	Measure	Performed by
1	Quantitative - Productivity	Number of ideas	Researcher
2	Hybrid	Complexity	Researcher
	Hybrid	Novelty	Researcher
	Qualitative - Concepts	Applicability	External panel of experts
3	Hybrid	Complexity	Researcher
	Hybrid	Novelty	Researcher
	Hybrid	Variety	Researcher
	Quantitative - Work session	Number of participants	Researcher

For the first case, the researchers evaluated only the number of ideas generated. It was found that the use of data as an input did increase the number of ideas generated, but teams seemed to select keywords that confirmed previously generated ideas, or generated similar ideas.

In the second case, the researchers succeeded in training a group of 48 students on the use of data mining tools for information exploration, and giving access to the tool pre-loaded with relevant data. Half of the teams chose to use the data mining tool to explore the data available, and only one team actually uploaded additional information to the tool. The results are again positive, as teams with the data support were able to advance in their concept development and were better assessed by a jury of experts, compared to teams not using the tool. However, the teams did not succeed in generating novel solutions as expected.

For the third case, aiming to correct the fixation of teams on related keywords that occurred in Case 1, and the low novelty in Case 2, teams were given pre-selected the keywords by finding interesting but distant terms in the reports produced by the data mining tool for each challenge.

The results from this session showed several benefits for teams using the keywords as input for their idea generation: it is theorized that having the keywords from the beginning attracted more participants to this challenge, as they had more 'support' of information to begin; teams were able to design more diverse ideas, compared to teams without the support, and finally, the elements within the ideas were also more diverse than teams without the keywords, which resulted in more complex solutions. Given the positive results from this session, the researchers hypothesized that providing the teams with direct access to the data mining tool would allow them to further explore the information if given training on selecting distant terms to combine for novel solutions.

Table 5.6 shows a summary of the three cases, noting the use of data by the team to support creativity, and the observed results:

Table 5.6 Summary of cases and findings

	Case 1	**Case 2**	**Case 3**
Objective	Data exploration as creative input for problem definition	Data exploration as creative input for concept development	Data as creative input for idea generation
Analysis level	Teams	Teams	Teams
Measure	Number of ideas	Complexity Novelty Applicability	Complexity Novelty Variety Number of participants
Degree of data exploration freedom	Medium	High	Low
Increase in number of ideas	Weak	DNA	DNA
Increase in novelty	DNA	Weak	High
Increase in complexity	DNA	Medium	Medium
Increase in diversity	No	Weak	High

5.5 Discussion

There is an opportunity to include data in the engineering design process, and search for ways to make information a part of the design process to increase diversity, originality, and complexity in the solutions proposed by teams. This is a unique study of cases, as all three cases are related to the same issues, and the variations in each case make for an interesting comparison as to what works, that should be kept for further sessions, and what can be improved to increase creativity for engineering design.

The use of data mining tools is beneficial at different stages of the engineering design process. All teams using either the information stemming from the analysis by the researcher, or the tool itself, benefitted from the input of data, whereas we can say that teams using other information sources or no information sources had a weaker performance (fewer ideas, less diversity, low solution development).

Each case had a different scope in the EDP, which let us observe whether the inclusion of data mining, or data extracted from it, would have a positive impact in the creative results of teams. The observation for the three cases was that it is indeed useful. It helps teams with the condensation of information, compared to having an open research on the internet which requires manually creating links and selecting relevant information. However, more work is needed in the future to improve the access to this information, and to test the best moment to provide the access and/or data.

We expected the access to information to help increase novelty and diversity, but the improvements were only incremental. However, in both cases were the tool was directly accessible by the participants, the results in the novelty were not as expected. There can be various explanations for this: first, it is possible that, not having an expert user in the data mining tool, the team can only use limited functionalities, thus not exploiting the information resulting from the tool more effectively. By limiting the combination and exploration capacity in Case 3, teams had more diverse and novel ideas, even compared with Cases 1 and 2, were teams were given broader access to information.

Another possible explanation to the differences between cases 2 and 3 is the reduction of the exploration space. It is possible that the teams in case 3, where they had a limited set of distant keywords, were able to make combinations more quickly, as they did not have the burden of exploring and deciding which information is useful and which is not, they worked with what they had, and combined it with their technical knowledge. In case 3, it appears as if students steered away from expected or known solutions, but perhaps it was due to a potential monetary reward, and the fact that they were not required to build a prototype, as participants in Case 2 were.

Teams not working with information from the data mining tool have an unlimited search space, which can be infinite. The prospect of an infinite search space can motivate teams to go back to known solutions, especially when time is limited. This fixation effect can be countered by artificially delimiting a search space. This finding is congruent with the need for moderators to help teams get un-stuck.

The proposition of keywords can artificially create a search space, and provide guidance for a design strategy. This is because participants in teams have technical knowledge, but not domain-specific knowledge, and therefore usually go for known solutions that are good but not novel. The combination of their existing technical knowledge and limited domain-specific knowledge provokes more original and diverse ideas, which confirms there is creative value in the combination of KDD with teams' existing knowledge.

5.6 Design cycle evaluation

NOTE: This section does not appear in the article, it is meant as a conclusion and transition for the next chapter of the thesis.

The use of data extracted from a KDD process has demonstrated to be of value at the different stages of an EDP. However, the main conclusion from this study of cases is that the timing and format of access to data is extremely important if the objective is to generate novel ideas.

To yield more innovative solutions, it is preferable to create a knowledge base related to the domain of the problem before the EDP begins, as to begin with novel combinations of keywords with knowledge within the team.

The use of data in later stages will aid the teams in developing the maturity of the concepts, to support the development of sub-systems and constituting elements of the solution based on existing technologies that are proven to work. In other words, the data helps to solve problems and to expand the search for components.

It is also recommended that the selection of relevant keywords should be made by an external actor, who can assess the distance of the keywords in the visualizations, and does not have a fixation on known keywords in the problem domain.

CHAPTER 6

DISCUSSION AND CONCLUSION

6.1 Summary of cases

The author of this thesis set out to study the effects of data exploration in the different early stages of the engineering design process, particularly on the novelty, diversity and complexity (sophistication) of ideas.

Three cases were performed with varying scopes of the EDP: the first case called for the identification of problems (Chapter 2), the second case (Chapter 3) went as far as prototyping the proposed solution, and the third case (Chapter 4) had the purpose to propose novel concepts. Table 6.1 presents a summary of the three cases presented in the thesis, noting the EDP scope, the KDD steps performer by the researcher and the participants, the total number of teams and participants.

Table 6.1 Summary of the three cases presented in this thesis

Case	Objective	Duration	Number of teams	Total participants
1	Need identification	4 hours	3	15
2	Benchmark Novel concept Prototype	2.5 Weeks (≈100 hours)	8	49
3	Novel concepts	24 hours	34	≈212

Chapters 2, 3 and 4 present the communications detailing the cases, while the article in chapter 5 attempts to conciliate the findings from the three cases under a theory on how the input from data obtained by mining patents impacted the results of teams working on an engineering challenge. All of these communications were shared with the scientific community in the form of conference or journal articles. A summary of the articles is presented in Table 6.2

Table 6.2 Summary of articles presented in the thesis

	Article 1	**Article 2**	**Article 3**	**Article 4**
Thesis chapter	**2**	**3**	**4**	**5**
Case	**1**	**2**	**3**	
Problem	Data exploration as creative input for problem definition	Data exploration as creative input for concept development	Data as creative input for idea generation	Data as creative input for different stages of the EDP
Research question	What is the impact of data exploration on problem definition?	What is the impact of data exploration for concept development?	What is the impact of keywords as creative input for idea generation?	How does KDD impact the different stages of the EDP?
Analysis level	Teams	Teams	Teams	Case studies
Approach	Deductive	Deductive	Deductive	Inductive
Method	Case study	Case study	Case study	Study of cases
Research type	Qualitative	Qualitative / Quantitative	Qualitative / Quantitative	Qualitative
Data collection	Primary	Primary	Primary	Primary
Article type	Empirical	Empirical	Empirical	Conceptual
Publication	PLM16 Conference	IJIDeM journal	Submitted to the Creativity and Innovation Management Journal	Submitted to the Journal of Engineering Design
Keywords	problem definition, idea generation, big data analytics, innovation	idea generation, data mining, patent mining, innovation, solution design	idea generation, innovation contests, bisociation, patent mining	idea generation, engineering design, solution design, knowledge discovery

6.2 Limitations of the research

As with all scientific work, this research was bound by certain limitations that might have influenced the outcome. We would like to address those concerns, as it is important to discuss potential improvements for future work.

First, the data selected to be included in the knowledge base was selected by the researcher, who is not an expert in the application area. This can have two potential implications: either the data extracted is not enough and could have been expanded, thus having a greater potential for innovation, or the information was enough for teams to design novel solutions, while remaining applicable solutions that could be accepted by experts in the domain. To test this, future design sessions can propose random words or keywords from different knowledge domains, to verify if novelty can be increased and the solution would still be applicable.

The result evaluations for Cases 1 and 3 was performed by the researcher, who might be a victim of confirmation bias, as she was expecting the data from data mining to help teams in their EDP. For the first case, it was attempted to control by having a quantitative measure of the idea production. For the second case, the researcher tried her best to quantify as well the elements in the solutions and the diversity, as it is shown in the resulting article, presented in Chapter 5. This can be solved by having other experts make an assessment of the results, in which case, data can be obtained from the YouTube page of the *Les 24 heures de l'innovation* for the challenges listed.

6.3 Results

During the Exploratory study documented in section 1.3.1, the teams selected pieces of information that supported their previous knowledge, and therefore the ideas already generated were reinforced. For the second exploratory study, documented in section 2.3.2, teams had access to two reports from data mining; one related to patents in the domain of their problem, the second related to social media postings about the issue. However, they were not able to propose novel solutions because they were bound by the clients to specific solution constrains. After these two cases, the objective was to apply the KDD in EDP efforts where the team had more inference on the problem, and therefore the solution. It was also important to observe the level of freedom to explore the data, and how that impacted the novelty of the solutions.

In Case 1, teams had an intermediate level of freedom to explore data: they were free to explore, but the database was already uploaded. Teams selected keywords that only slightly varied their previously generated ideas. One explanation for this would be the timing of the access to the tool. They had already been working on defining the problems, and time was limited to make a new iteration. It was then proposed that future teams with access to data should have more time to upload additional data and explore by themselves. It is likely that the timing of the access to the data was too late in the process, a fault that was corrected for Case 2.

An opportunity arose in Case 2 to test this notion, teams were given a workshop to learn to use the tool, but the use was optional. In this case, the freedom to explore the tool and the information was unrestrained. Half of the teams decided to make use of the tool, and though they had the opportunity to propose a radically novel concept, teams opted instead to improve their concepts.

In the case of teams with direct access to the data mining tool, which occurred in Cases 1 and 2, the results support the case for having an external moderator or actor selecting the data as creative input for the idea generation phases.

In Case 3, teams were provided with the pre-selected keywords at the very beginning of the process, just after learning what the challenge was. The freedom to explore the tool and the data was limited, as we provided only the result of the data analysis. We believe that by providing the keywords, the teams created their own exploration space which was broader than the knowledge contained within the team, but not endless, as a web search would be.

6.4 Discussion of results

Design problems, as the ones presented in the three cases here described, are complex because there is more than one possible answer, they are usually unstructured, and the outcome is not clear from the start. For example, if the problem is to design a new building, the team knows in advance that the end result is a building, there is a process to follow that is known.

In the case of the issues in the three cases, there is no known outcome or expected answer. If the team decides to work on 'how to communicate the quality of the river water', there are many possible directions: What is 'quality'? Is the communication happening on-site or distant? What type of communication can include more users? What is the frequency required? What would be the business model? and so on. The problem is there, but the objective of it is not defined.

To structure the problem, designers will tap into their knowledge base, and try to gather as much information as possible from external sources, such as clients, the design brief, etc. In the three cases, we consider that the input from the data mining exercise that was provided to the teams was useful in the definition of the problem. The data can serve as an indication to interesting concepts in the domain.

This can actually be because equipping the team with external data can help expand the exploration search, all the while delimiting it. In the end, the information available to the team determines the solution produced.

6.5 Implications for the industry

Creativity is one of the most important traits of this generation, companies that do not encourage creativity cannot innovate, and if a company does not innovate, it will be obsolete. The work presented here aims to shed a light on the inclusion of data from the analysis of large bodies of data in the engineering design process. It was found that its use, at different stages, can be helpful for teams.

However, its use and application will depend on several conditions. The first is the data included in the database to be mined. It is important, as previous authors have mentioned, to determine the objective of the data, and select the appropriate sources and algorithms to be applied to the data. The second is the timing of the access to the data, as we concluded, in order to have more novelty, the access should be provided early in the process, before the teams get fixated on ideas already proposed. The third is the use of external moderators or actors that select 'relevant' or 'interesting' information, to prevent teams from selecting information that confirms or supports previous knowledge or solutions.

It is certain that companies looking to innovate would benefit from keeping a knowledge base available to make decisions not only regarding the managerial activities, but also as a source for new product or solution design.

6.6 Future work

6.6.1 Engineering design process

This research focused on teams of engineering students on a deadline and with limited time to complete their objective. A future line of research would be to test whether the results are similar in an industrial context with experienced engineers and with more time available.

The students who participated in the cases possess knowledge of an engineering area but do not have many years of experience in the industry, are not fixed or stuck on known solution or viewpoints. Experienced engineers may have more rigid views that can lead them to use patent visualizations and reports differently.

Another area that can be further investigated is the processes used by teams to incorporate information visualization into the design process. One difficulty in performing this type of research is the acquisition of data on the process. It may be necessary to record video or audio throughout all the work sessions.

6.6.2 Use of machine learning and artificial intelligence

The field of data mining is continuously evolving, and the new wave of data analysis tools that derive from artificial intelligence and machine learning algorithms, can also benefit the engineering design process.

This research was performed using a tool based on frequency of terms to extract important terms from patents. Although it is a strategy that is commonly used for the extraction of information, there are advances in the area of machine learning have opened the possibility of detecting more sophisticated patterns in the data.

For example, through the application of neural network models for learning patterns in unstructured or raw text, as is the case for some sections of the patents, it is possible to capture semantic and syntactic information of words in vectors. You can use these results to generate new types of visualizations and new systems with the ability to explore relationships between words or concepts, which is a step further than the analysis used for the Voronoi diagram in the IPMetrix tool discussed in section 1.1.5.

As a future line of research, it is possible to test new reports and visualizations generated from the application of machine learning algorithms to determine if more advanced reports have a different effect on the solutions proposed by participants.

6.7 Conclusion

We were able to compare results between cases where the data was pre-selected for the teams, which was the case for the exploratory study and for Case 2, and cases where the teams had direct access to the data mining tool and selected by themselves the data to consider.

What appears to work best when tackling an engineering problem, based on the three cases here presented, is the combination of human and machine, where the previously possessed knowledge is expanded by providing domain-specific data extracted from a large database, specifically selected to provoke novel combinations. This is demonstrated by the fact that, under the same conditions, teams with no keyword support propose solutions with significantly lower variety in the theme and the elements within the solution. Ergo, the exploration space in teams with keyword support appears to have expanded.

It was found that in all three cases, teams using data as input performed better compared to teams not using it. For cases where the teams had access to the tool, the teams were able to improve their diversity of ideas was improved for teams using data. There is value in the combination of existing knowledge and knowledge from mining databases.

BIBLIOGRAPHICAL REFERENCES

Ackoff, R.L. (1989). From data to wisdom. *Journal of applied systems analysis,* 16 (1), 3-9.

Adamczyk, S., Bullinger, A. C., & Möslein, K. M. (2012). Innovation contests: A review, classification and outlook. *Creativity and Innovation Management*, 21(4), 335-360.

Agogué, M., Kazakçi, A., Hatchuel, A., Masson, P., Weil, B., Poirel, N., & Cassotti, M. (2014). The impact of type of examples on originality: Explaining fixation and stimulation effects. *The Journal of Creative Behavior*, 48(1), 1-12.

Ardaiz-Villanueva, O., Nicuesa-Chacón, X., Brene-Artazcoz, O., Sanz De Acedo Lizarraga, M. L., & Sanz De Acedo Baquedano, M. T. (2011). Evaluation of computer tools for idea generation and team formation in project-based learning. *Computers and Education,* 56(3), 700-711.

Atman, C. J., Adams, R. S., Cardella, M. E., Turns, J., Mosborg, S., & Saleem, J. (2007). Engineering design processes: A comparison of students and expert practitioners. *Journal of engineering education*, 96(4), 359-379.

Aurenhammer, F. (1991). Voronoi diagrams—a survey of a fundamental geometric data structure. *ACM Computing Surveys* (CSUR), 23(3), 345-405.

Baesens, B. (2014). *Analytics in a big data world: The essential guide to data science and its applications*. Hoboken, NJ. John Wiley & Sons.

Bjelland, O. M. & Wood, R. C. (2008). An inside view of IBM's 'Innovation Jam'. *MIT Sloan Management Review,* 50(1), 32-40.

Boden, M. A. (2004). *The creative mind: Myths and mechanisms*. Psychology Press.

Burkus, D. (2013). *The Myths of Creativity: The Truth About How Innovative Companies and People Generate Great Ideas* (1st ed.). John Wiley & Sons.

Chen, P. L., Li, S. C., & Hung, M. T. (2013, July). Co-occurrence analysis in innovation management: Data processing of an online brainstorming platform. In *Technology Management in the IT-Driven Services (PICMET)*, 2013 Proceedings of PICMET'13: (pp. 688-694). IEEE.

Cohendet, P., Parmentier, G., & Simon, L. (2016) Managing knowledge, creativity, and innovation. Work in Progress, version: July 25, 2016. HEC Montreal / Université Grenoble Alpes.

Dodgson, M., Gann, D., & Salter, A. (2006). The role of technology in the shift towards open innovation: the case of Procter & Gamble. *R&D Management*, 36(3), 333-346.

Dove, G. & Jones, S. (2014). Using Information Visualization to Support Creativity in Service Design Workshops. *Paper presented at the ServDes*. Lancaster, UK.

Dubitzky, W., Kötter, T., Schmidt, O., & Berthold, M. R. (2012). Towards creative information exploration based on Koestler's concept of bisociation. In *Bisociative Knowledge Discovery* (pp. 11-32). Springer Berlin Heidelberg.

Ertas, A., & Jones, J. C. (1996). *The engineering design process* (2nd ed. ed.). New York, N.Y.: J. Wiley and Sons.

Fauconnier, G., & Turner, M. (1998). Conceptual integration networks. *Cognitive science*, 22(2), 133-187.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17(3), 37

Fleming, L., & Szigety, M. (2006). Exploring the tail of creativity: An evolutionary model of breakthrough invention. *Advances in Strategic Management*, 23, 335-359.

George, G., Osinga, E. C., Lavie, D., & Scott, B. A. (2016). Big data and data science methods for management research. *Academy of Management Journal, 59*(5), 1493-1507.

Glier, M. W., Schmidt, S. R., Linsey, J. S. & McAdams, D. A. (2011). Distributed ideation: Idea generation in distributed capstone engineering design teams. *International Journal of Engineering Education*, 27(6), 1281-1294.

Graetz, K. A., Proulx, N., Barlow, C. B., & Pape, L. J. (1997, November). Facilitating idea generation in computer-supported teleconferences. In *Proceedings of the international ACM SIGGROUP conference on Supporting group work: the integration challenge* (pp. 317-324). ACM.

Gumienny, R., Gericke, L., Wenzel, M., & Meinel, C. (2013, February). Supporting creative collaboration in globally distributed companies. In *Proceedings of the 2013 conference on Computer supported cooperative work* (pp. 995-1007). ACM.

Hamman, M. (2000). Priming computer-assisted music composition through design of human/computer interaction. *Urbana*, *51*, 61801.

Hevner, A. R. (2007). A Three Cycle View of Design Science Research. *Scandinavian Journal of Information Systems*, 19(2), 87-92.

Hevner, A. R., March, S. T., Park, J. & Ram, S. (2004). Design science in information systems research. *MIS quarterly*, 28(1), 75-105.

Holzinger, A., & Jurisica, I. (2014). Knowledge discovery and data mining in biomedical informatics: The future is in integrative, interactive machine learning solutions. In *Interact knowledge discovery and data mining in biomedical informatics*, 1-18, Springer Berlin Heidelberg.

Howard, T. J., Culley, S. J., & Dekoninck, E. (2006). Information as an input into the creative process. In *DS 36: Proceedings DESIGN 2006, the 9th International Design Conference*, Dubrovnik, Croatia.

Howkins, J. (2002). *The creative economy: How people make money from ideas*. United Kingdom: Penguin.

Jiménez, L. M., Desrosiers, S. & Gardoni, M. (2011). Creative teamwork in quick and long term project development, 24 hours of innovation. *Symposium on Models and Modeling Methodologies in Science and Engineering* (MMMse 2011). Orlando, USA.

Jiménez-Narvaez, L. M., Dalkir, K., & Gardoni, M. (2012) Managing knowledge needs during new product lifecycle design on Quick-term Project Development QPD: Case study of 24 hours of innovation - ETS Montreal. *Vol. 388 AICT* (pp. 25-34).

Jung, J. H., Schneider, C. & Valacich, J. (2010). Enhancing the motivational affordance of information systems: The effects of real-time performance feedback and goal setting in group collaboration environments. *Management Science*, 56(4), 724-742.

Kabir, N. & Carayannis, E. (2013). Big Data, Tacit Knowledge and Organizational Competitiveness. *Journal of Intelligence Studies in Business*, 3(3).

Karlsson, A., & Törlind, P. (2016). Mitigating lack of knowledge: a study of ideas in innovative projects. *International Journal of Design Creativity and Innovation, 4*(3-4), 144-161. doi: 10.1080/21650349.2014.961553

Koestler, A. (1964). *The act of creation* (6th ed.). London: Hutchinson.

Koutroumpis, P., & Leiponen, A. (2013). Understanding the value of (big) data. In *Proceedings of 2013 IEEE international conference on big data* (pp. 38-42).

Maccrimmon, K. R. & Wagner, C. (1991). The architecture of an information system for the support of alternative generation. *Journal of Management Information Systems* 8(3), 49-67.

Majchrzak, A. & Malhotra, A. (2013). Towards an information systems perspective and research agenda on crowdsourcing for innovation. *The Journal of Strategic Information Systems,* 22(4), 257-268.

Manning, C., Raghavan, P., & Schütze, H. (2009). *Introduction to information retrieval* (1st ed.). Cambridge: Cambridge University Press.

Müller, L., Wetzel, T., Hobohm, H. C., & Schrader, T. (2012, November). Creativity Support Tools for Data Triggered Hypothesis Generation. In *Knowledge, Information and Creativity Support Systems,* 24-27.

Munemori, J. & Nagasawa, Y. (1996). GUNGEN: Groupware for a new idea generation support system. *Information and Software Technology*, 38(3 SPEC. ISS.), 213-220.

Munemori, J., Yoshino T. & Yunokuchi, K. (2001). A spiral-type idea generation method support system for sharing and reusing ideas among a group. *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*.

Natalicchio, A., Messeni Petruzzelli, A., & Garavelli, A. C. (2014). A literature review on markets for ideas: Emerging characteristics and unanswered questions. *Technovation, 34*(2), 65-76. doi: http://dx.doi.org/10.1016/j.technovation.2013.11.005

Nielsen, M. (2012). *Reinventing discovery*. Princeton, N.J.: Princeton University Press.

Parjanen, S., Hennala L. & Konsti-Laakso, S. (2012). Brokerage functions in a virtual idea generation platform: Possibilities for collective creativity? *Innovation: Management, Policy and Practice,* 14(3), 363-374.

Piatetsky-Shapiro, G. (1991). Knowledge Discovery in Real Databases: A Report on the IJCAI-89 Workshop. *AI Magazine* 11(5): 68–70.

Rhéaume, L., & Gardoni, M. (2016). Strategy-making for innovation management and the development of corporate universities. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 10(1), 73-84.

Ríos-Zapata, D., Duarte, R., Pailhès, J., Mejía-Gutiérrez, R., & Mesnard, M. (2016). Patent-based creativity method for early design stages: case study in locking systems for medical applications. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 1-13.

Rosen, P. A. (2011). Decision making in a Web 2.0 environment: Crowdsourcing lessons for organizations. *Journal of Decision Systems,* 20(3), 309-324.

Shan, Z., Zhu, Y. & Zhao, T. (2013). Kaleidoscope: Computer-Assisted Ideation through Idea Network Exploration. CS 224W 2013 Project Final Report. Department of Computer Science, Stanford University.

Shneiderman, B., Fischer, G., Czerwinski, M., Resnick, M., Myers, B., Candy, L., Edmonds, E., Eisenberg, M., Giaccardi, E., Hewett, T., Jennings, P., Kules, B., Nakakoji, K., Nunamaker, J., Pausch, R., Selker, T., Sylvan E., & Terry, M. (2006). Creativity Support Tools: Report From a U.S. National Science Foundation Sponsored Workshop. *International Journal of Human-Computer Interaction,* 20(2), 61-77.

Siau, K. (2000). Knowledge discovery as an aid to organizational creativity. *The Journal of Creative Behavior*, 34(4), 248-258.

Sim, S. K., & Duffy, A. H. (2003). Towards an ontology of generic engineering design activities. *Research in Engineering Design*, *14*(4), 200-223.

Trappey, A. J., Trappey, C. V., & Wu, C. Y. (2009). Automatic patent document summarization for collaborative knowledge systems and services. *Journal of Systems Science and Systems Engineerin,g* 18(1), 71-94

Verhaegen, P. A., D'hondt, J., Vandevenne, D., Dewulf, S., & Duflou, J. R. (2011). Identifying candidates for design-by-analogy. *Computers in Industry,* 62(4), 446-459

Vuori, M. (2012). Exploring uses of social media in a global corporation. *Journal of Systems and Information Technology,* 14(2), 155-170.

Wagner, C. & Jiang, L. (2012). Harnessing the power of social media for creativity support: A three-pronged approach. *Knowledge Management and E-Learning,* 4(2), 174-194.

Wang, H., & Ohsawa, Y. (2013). Data-Driven Innovation Technologies for Smarter Business from Innovators' Market Game to iChance Creativity Support System. In *Advances in Chance Discovery,* (pp. 107-125). Springer Berlin Heidelberg.

Wodehouse, A. & Ion, W. (2012). Augmenting the 6-3-5 method with design information. *Research in Engineering Design,* 23(1), 5-15.

Yuizono, T., Kayano, A., Shigenobu, T., Yoshino, T., & Munemori, J. (2005, September). Groupware for a new idea generation with the semantic chat conversation data. In *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems,* (pp. 1044-1050). Springer Berlin Heidelberg.

Websites and articles

Competition AquaHacking (2016). Fondation de Gaspé-Beaubien. Retrieved March 18, 2016, from http://2016.aquahacking.com/en/challenge/

Dubois, M. (2016). A Creativity Guide for Your Ideation Sessions. Retrieved October 20, 2016 from http://substance-en.etsmtl.ca/creativity-guide-ideation-sessions-2/

Dubois, M. (2016). The 24 hours of innovation: Montreal’s secret recipes to win! Retrieved October 20, 2016 from http://substance-en.etsmtl.ca/24-hours-innovation-official-guide-international-competition/

Dubois, M. (2016). The Importance of Team Preparation for an Ideation Session, Retrieved October 20, 2016 from http://substance-en.etsmtl.ca/team-preparation-ideation-importance/

IP Metrix Solution - TKM. (2016). Tkm.fr. Retrieved June 10, 2016, from http://www.tkm.fr/en/ip-metrix-solution.php

Laney, D. (2001). 3D Data Management: Controlling Data volume, Velocity and variety. Retrieved March 29, 2016 from https://blogs.gartner.com/doug-laney/files/

Les 24 heures de l’innovation (2016). Les 24h de l'innovation. Retrieved May 24, 2016, from http://24h-innovation.org/en/

Les 24 heures de l’innovation YouTube page (2016). Les 24h de l'innovation. Retrieved May 26, 2016, from https://www.youtube.com/user/24hinno