Conceptual Clustering of Patents

Enhancing insight for the decision making process in Strategic Technology Planning

Kevin Kruijthoff

Technische Universiteit Delft

TU Delft Delft University of Technology

Challenge the future

CONCEPTUAL CLUSTERING OF PATENTS

ENHANCING INSIGHT FOR THE DECISION MAKING PROCESS IN STRATEGIC TECHNOLOGY PLANNING

by

Kevin Kruijthoff

in partial fulfillment of the requirements for the degree of

Master of Science
in Engineering & Policy Analysis

at the Delft University of Technology,
to be defended publicly on the 17th of August 2015

Student number	4326415	
Supervisor:	Dr. S. Cunningham	TU Delft
Thesis Committee:	Prof. dr. M. van Geenhuizen	TU Delft
	Dr. S. van Splunter	TU Delft
	Dr. A. Sanz	SKF

TU Delft

—'Geen tijd', is liever wat anders doen.
Opa

Acknowledgments

First of all I would like to thank the graduation committee for their support during this thesis. I would like to thank Prof. dr. Marina van Geenhuizen for chairing my committee, enabling me to perform my research, as well as for her insightful comments on the thesis report and the usability of Decision Support Systems. Dr. Sander van Splunter for his insightful comments on the design elements and the structuring of the thesis, as well as for his shared enthusiasm for the topic. Dr. Alejandro Sanz for his great insights into the patent analysis theory and its practical use, the providing of relevant data needed for my research, as well as taking the effort of traveling to Delft and the providing of insightful feedback. I want to specially thank Dr. Scott Cunningham for supervising me throughout this period. Our initial exchanging of ideas and the introduction to Alejandro, resulted in me working on exactly the type of topic I was searching for. I very much enjoyed our discussions on the topic as well as on adjacent topics covering our shared passion for data analysis and its practical use.

Furthermore, special thanks to the group of EPA friends that I was privileged to interact with over the last two years. The exchanging of ideas over coffee, the hours spend on (modelling) projects, and the wonderful international dinners at 'Tariq's party place' have enriched my life.

Daarnaast wil ik mijn familie en schoonfamilie bedanken voor de steun en gezelligheid, gedurende mijn studie. 'Geen tijd, is liever wat anders doen' zou opa zeggen, en de afgelopen tijd, besteed aan mijn studie, kan ik dan ook niet anders bestempelen dan precies dat wat ik leuk vind. Afsluitend, wil ik mijn vriendin Jobien bedanken voor haar steun en geduld gedurende deze afgelopen jaren van projecten, tentamens en deadlines.

Delft, August 2015
Kevin Kruijthoff

Summary

Within technological R&D companies, the decisions made within their Strategic Technology Planning (STP) process are of great importance due to their financial, legal and competitive repercussions. For this reason their R&D managers want to be as widely informed as possible about the risks and benefits of developing a certain technology. As the subjects for which these decisions are made, are too complex to comprehend, there is exploitation of so-called Decision Support Systems. This research focusses on those Decision Support Systems in the STP process which are based on patent analysis. Patent analysis is relevant to apply in this process as the patent database is considered to be a comprehensive collection of technological concepts. This is due to the fact that companies voluntarily deliver the information, and benefit from describing their technological innovation as complete as possible. Throughout the STP process multiple needs for insight to make informed decisions can be detected. However, as every applied form of analysis will present a representation of the selected subject, which' outcome is based on assumptions made as well as on the underlying method applied. The user cannot be sure that this given outcome is always correctly representing reality. Hence, they are applying multiple representation methods on the same subject, a process which is also referred to as multiplism. Within this process they are searching for robust outcomes that hold true under all the different representations. Due to this process, there is an ever present interest in new methods of representation and analysis, to extend the R&D division's decision support toolkit.

Due to the ever present need for more insight, alongside a methodological interest, this research focusses on the evaluation of the applicability of conceptual clustering to patents, as well as its potential usability in the STP process. Conceptual clustering is a machine learning technique developed in the 1980's, which automatically categorizes the entered information, resulting in a tree shaped hierarchy. The conceptual clustering technique hasn't been found to be applied to patents before and this research will therefore also focus on identifying its potential use within the STP process.

To gain insight into the different needs for patent analysis within the STP process, as well as to provide a source for evaluation, a design has been made for the decision support system by applying the Axiomatic Design method. The inputs for this design are customer needs which have been based on a literature study and expert input on the different applications for patent analysis in the STP process. To assess its potential applicability, a prototype Decision Support System, constructed around the method of conceptual clustering applied to patents, has been developed. For which the implementation of the conceptual clustering functionality is done based on the COBWEB algorithm as proposed by Fisher. The prototype entails two software products written in Python which are respectively a text-mining and a conceptual clustering program, combined with a visualization tool based on the D3 JavaScript library, for which an online environment has been developed.

To illustrate the evaluation, a complete selection of patents related to additive manufacturing (3D printing) retrieved from Thomson Reuters' Derwent patent database is used. This patent set (containing 9360 patents) has been parsed through-, and visualized by-, the developed prototype, showing a representation of this technological field's inter-patent relational structure. Through evaluation based on the design, the prototype has been proven to be usable within the STP process for Exploration analysis, Competitor analysis and Portfolio analysis. The highest added value to the R&D division's toolkit is perceived to lay in the explainability of the outcome, the accessibility through the online visual representation and the possibility of using the decision support system in an interactive way. Allowing the technology scouts and managers to interact with the data, discussing and continuing exploring based on newly gathered insights. Furthermore implementable solutions are presented, allowing to extend the developed prototype's applicability for use to the inclusion of Freedom to operate analysis capabilities, trend analysis and to a limited extend inventive problem solving analysis capabilities. Further evaluation of the prototype based on a comparative analysis of the case with the output created with a self-organizing map technique, leads to further conclusions on the conceptual clustering technique. Firstly, the same high level clusters can be detected, however not all inter-relations of these clusters match. Forcing the user to further investigate, increasing robustness of knowledge obtained. Secondly, the conceptual clustering shows more application domains in addition to the technical concepts. Thirdly, multiple representations can be made of the same set based on user needs. Fourthly, the influence of the attribute number, for which it is suggested to perform a high level search first. Then select a subset of patents for reprocessing and visualization, in order to obtain more insight into sub-branching. Finally, the combinational use of the two methods enhances the insight, due to the explanatory value of the nodes in the conceptual clustering output.

Contents

—Somewhere, something incredible is waiting to be known.
Carl Sagan

1 Introduction

For Research and Development (R&D) sections of technological organizations, Strategic Technology Planning (STP), a process guiding the development of new technologies, is of great importance. This is due to the financial, legal and competitive impact that is the result of this process. Hence, their managers aim to provide the best possible advice to their stakeholders in what to be developed technology to put in the company's resources and on how to further develop these technologies. Strategic Technology Planning incorporates multiple aspects of exploration and is a continuous process aiming to guide and monitor the development process. Ideally the managers would base their advice rationally on full information, meaning that they would know all their decisions' repercussions. However in reality this is impossible as they know that they are limited by bounded rationality[1]. This is due to lack of insight in e.g. the technological field they are operating in or what actions their competitors will take in the future. To reduce this lack of understanding, thus making better informed decisions, Strategic Technology Planning exploits the usage of different Decision Support Systems (DSS).

This research focuses on those Strategic Technology Planning Decision Support Systems that are based on data extracted from patent databases, such as those of the European Patent Office or the United States Patent and Trademark Office. These databases provide an overview of the issued and pending patent applications. Which are due to their organized nature and consistent classification of data, a highly interesting source when analyzing a technological field. It is however important to mention that to get a complete insight into a technological field as a whole, other sources such as scientific publications or international projects, should also be analyzed. To be able to analyze the information that is embedded in the patent databases, it needs to be made interpretable, which is done by extracting, structuring and visualizing it first. These are steps that, due to the large number of patents, are too comprehensive to perform manually. Thus rises the need for automation of this process.

This chapter will introduce the reason for the suggested addition of a structuring and visualization technique to the automation of this process, which is that of conceptual clustering. This will be done by first introducing the problem statement based on the currently available patent analysis tools within the Strategic Technology Planning process in Section 1.1. Secondly, the method of conceptual clustering and the here upon based research question and sub-questions are introduced in Section 1.2. Finally Section 1.3 will explain the research methodology that has been applied during this research.

1.1. The Strategic Technology Planning process

Figure 1.1 (which will be elaborated in Chapter 4) presents an implementation method of the Strategic Technology Planning process, in which can be seen that the needs for representation of the patent field are dependent on the step within the process as well as on the specific actor's role. The process starts off with a call

[1] Bounded rationality: Rationality that is limited due to lack of information, cognitive capability, and/or time (Simon, 1982)

for research and ends with a pilot phase to develop a concept which is from then on being continuously reassessed. In blue, a selection of the possible uses of patent analyses are shown. Each of the individual patent analyses provides 'a' view of the aspects of the patent field under examination. However, given that each analysis will be subject to assumptions and design choices, one cannot guarantee that the presented representation will be just under all circumstances. Adding to this is the fact that the derived results will always be in light of a certain judgment made by the analyst. Therefore managers are continuously searching to extend their toolkit with different tools to gain even more versatile insights on the same subject within these different analysis steps. In which the aim is to compare the multiple analyses to search for robust outcomes, which hold true under the tools different given assumptions and designs, a process also referred to as multiplism. Additionally, as the decisions made based on the outputs of these analyses are of great importance, their trustworthiness is essential. In which trustworthiness refers to the level of confidence one has in the results, to draw conclusions based on this output.

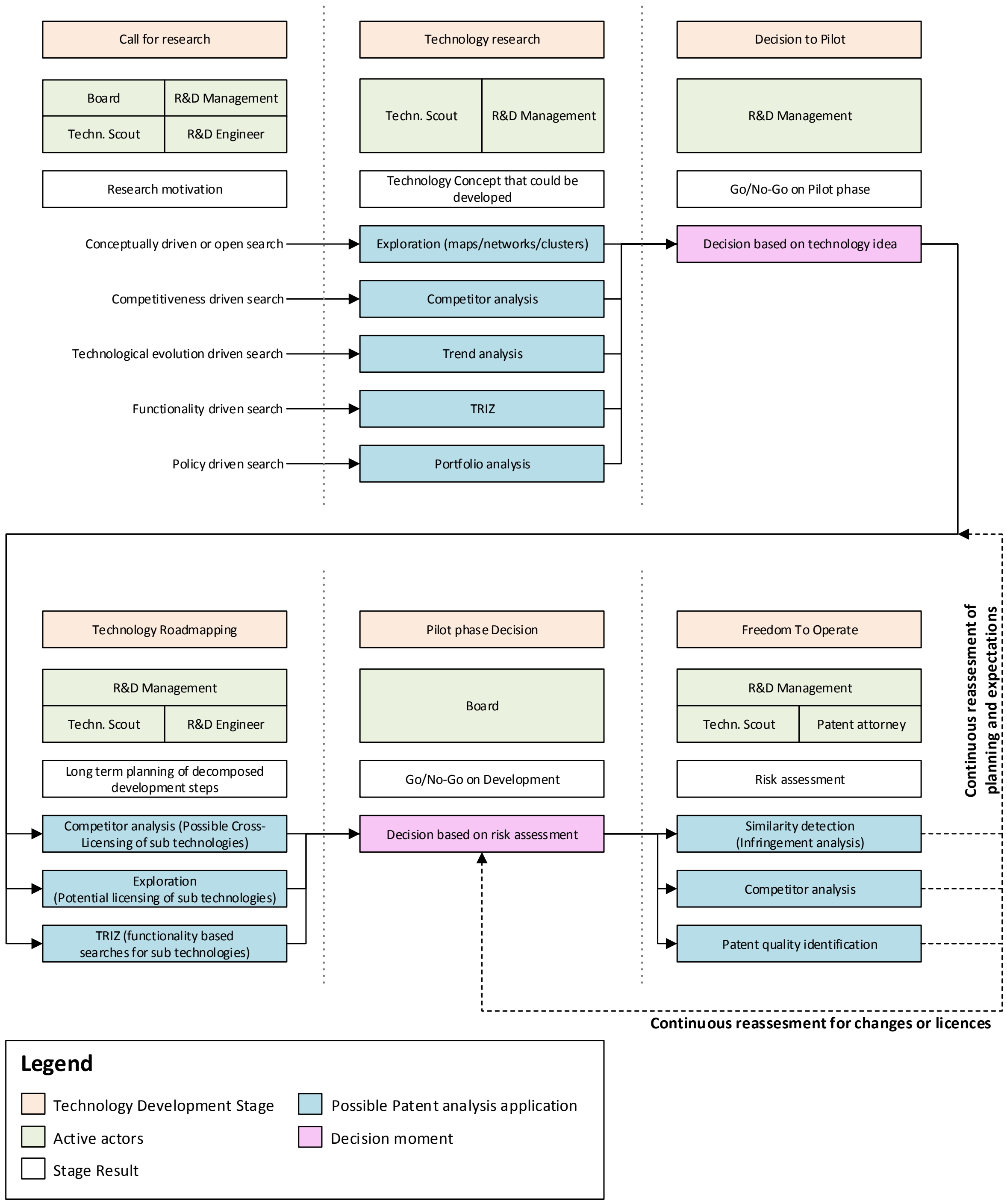


Figure 1.1 – The possible positioning of patent analyses in Strategic Technology Planning

The representation of the technological fields is of relevance in many of the possible applications of patent analysis. In a recent literature study on the state of the art in patent analysis by Abbas, Zhang and Khan the visualization techniques are categorized in: Patent networks, Patent maps and Data clusters (Abbas, Zhang, & Khan, 2014). Of which Figure 1.2 shows a selection of examples, in which the wide range of variety in representation methods can be seen. These are respectively (a) Self organizing claim point maps (B.-U. Yoon, Yoon, & Park, 2002), (b) Keyword-based patent maps (S. Lee, Yoon, & Park, 2009), (c) Patent topic maps (Kasravi & Risov, 2009), (d) Technology networks (B. Yoon, 2010), (e) Patent claim maps (Shin & Park, 2005), (f) Semantic patent maps (Bergmann et al., 2008), (g) Social network analysis applied to patent citations (Sternitzke, Bartkowski, & Schramm, 2008) and (h) Self organizing maps (Thomson Reuters, 2008).

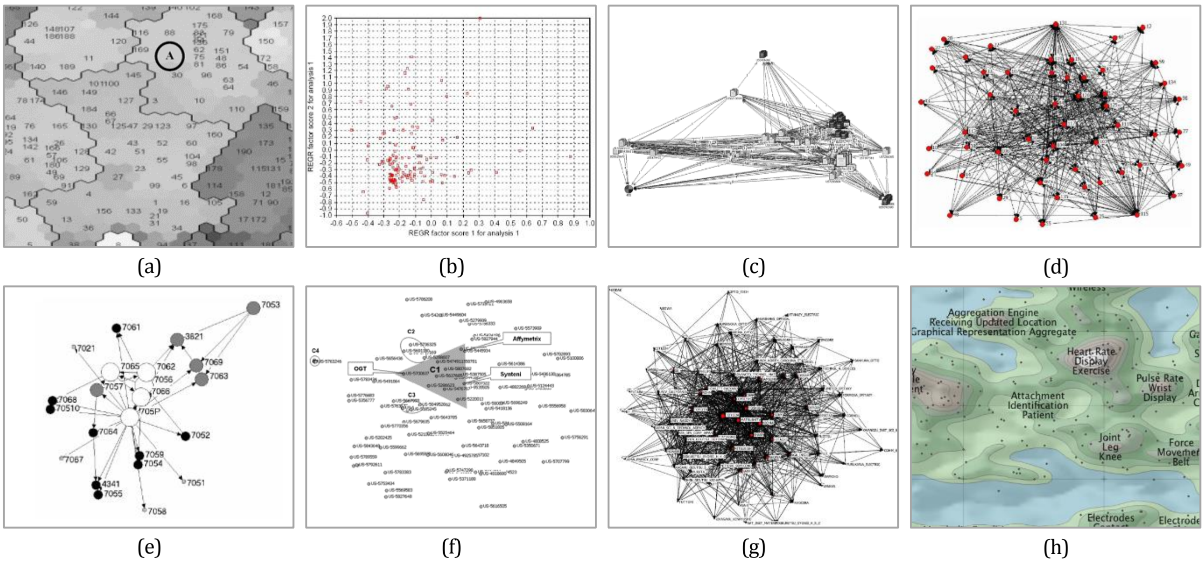


Figure 1.2 – Examples of currently applied patent analysis representations

These different methods of representing aspects of the patent field vary in purpose and usability. Contemplative, the three mayor detectable flaws or shortcomings when examining the currently available methods are:

1. In most of these methods there is a lack of interpretable explainability of the patent locations. Meaning that by looking at these maps, one cannot deduct why a certain patent is located at its given position or is linked to another patent. The interpretation of the links between patents and their positioning is claimed to be crucial in understanding a representation (Rafols, Porter, & Leydesdorff, 2010). This results in potential misinterpretations by their users as well as a negative influence on the trustworthiness.
2. They do not capture the nature of technology supported by the view of data science, which sees the technological space as to be a hierarchical one. This results in a representational problem which may prohibit the analysts from correctly interpreting the technology field.
3. Not all methods are reproducible ones, meaning that based on the same input and settings, different outputs are generated, which in itself is worrisome when regarding the trustworthiness aspect. These representations all show their local optima, so on their own they all have some validity, but one will need to construct and interpret multiple to get a more reliable impression of the actual patent space.

Additionally, for multiple patent analysis applications the possibility to add non patent literature data as an additional data layer can assist in enhancing the insight the tool can provide. For example the addition of business relevant information, such as the likelihood that a certain company will sue when using similarity detection analysis, or the projection of opportunities based on media analysis onto a trend analysis representation. As the decisions that are based on the interpretations of these analyses are of great financial, legal and competitive importance, there is a continuous search for more functionality and reliability from these decision support systems.

1.2. Conceptual clustering as a potential addition to the STP toolkit

Given the ever present need for more insight and the reduction of risks, this research will focus on the evaluation of a new application, which is that of conceptual clustering of the patent field. Conceptual clustering is an existing machine learning algorithm for hierarchical classification, developed in the 1980's, and is considered to fall under the categorization of clustering techniques. Clustering can be described as the process of intelligently grouping items into categories that are meaningful to its user (Michalski, 1980). Conceptual clustering is first defined in 1980 by Michalski as a machine learning method. In which machine learning refers to a method that does not require the user's input on the pre-definition of classes. Rather, it is based on an evaluation function that creates classes which fit the entities best, a method which is described as 'learning by observation' (Fisher, 1987). To study the applicability of a decision support system based on conceptual clustering to patents, the COBWEB algorithm as defined by Fisher (Fisher, 1987) has formed the basis of a prototype. For which the applicability as to being used in the STP process is evaluated. The COBWEB algorithm has been chosen as it is a clearly structured algorithm that is based on four main operators. This algorithm's output is inherently reproducible and by default a hierarchical tree shape which is expected to aid the user in interpreting the data.

There has no literature been found covering the application of conceptual clustering applied to the patent field. Which was in itself unexpected, given that this technique is a relatively old one, the hierarchical nature of technology as well as the hierarchical structure of the patent field. This could be because of the lock-in effect of the currently applied techniques. In which the available tools have been proven to be of use and thus are continuously used and improved. Resulting in no effort being put into the examination of other techniques applicability's. Another reason could be that the conceptual clustering technique is from a different school of thoughts as the currently applied methods, and therefore not having been explored. The conceptual clustering technique originates from the artificial intelligence and machine learning field, where most of the current available methods seem to be originating from the field of statistics.

The aim of this research is firstly to determine whether it is possible to apply conceptual clustering to the patent field. And secondly, to evaluate the suitability of conceptual clustering to enhance (not replace) the current patent analysis applications within the STP process. As it is stated that "combined models are often more effective than their separate parts" (Porter & Cunningham, 2004). To perform this research the research question is therefore stated as:

"For which aspect and in what manner could conceptual clustering enhance the current patent analysis process for Strategic Technology Planning in technical R&D companies?"

This question will be answered by focusing on the following sub questions;

1. What is the role of Decision Support Systems within the STP process of R&D companies?
2. How is technology structured and what does this imply for patent analysis?
3. For what purposes within the STP process is patent analysis used?
4. How can a tool be designed to apply conceptual clustering to patents?
5. For what aspect in the STP process can the designed tool be used to enhance the currently available patent analysis tools?
6. Does conceptual clustering applied to patents prove to be a worthy addition to the currently available patent analysis tools?

The scope of the outcomes will be in light of the strategic technology planning decision making process and its users' needs. This entails its use in those positions where patent analysis can be applied. This excludes technologies that cannot be patented, strategic non-patenting of technologies, as well as those technologies that have not been granted a patent yet. This is further elaborated in Chapter 3.

1.3. Research methodology

Figure 1.3 shows the applied research structure presented in this thesis in order to answer the presented research question. As to analyze for which aspects conceptual clustering applied to patents could enhance the STP process, a design type research is applied in which three main sections can be defined:

1. The design of the Decision Support System's functional needs and constraints based on the STP process needs by applying the Axiomatic Design method
2. The development of a decision support system prototype build around the COBWEB algorithm applied to patents and its visual representation in an online environment
3. The evaluation of the conceptual clustering prototype based on the designed functional needs and constraints, using a patent set representing the additive manufacturing field

The first section is focusing on the design and thus inherently the determination of evaluation criteria, and presents an analysis of the functional needs for Decision Support Systems in the STP process. This is done by first elaborating on the role of Decision Support Systems in multi-actor decision making settings in Chapter 2. Which is followed by relevant theoretical background information in Chapter 3, presenting the notion of technology and its relation to patenting. The first section is concluded by Chapter 4, which provides an overview of potential patent analyses applications within the STP process and their functional needs for the Decision Support Systems. These functional needs and their constraints are derived via a design process called Axiomatic Design as defined by Suh (Suh, 2001). Which is a widely applied design method, guaranteeing a complete design of the relevant aspects.

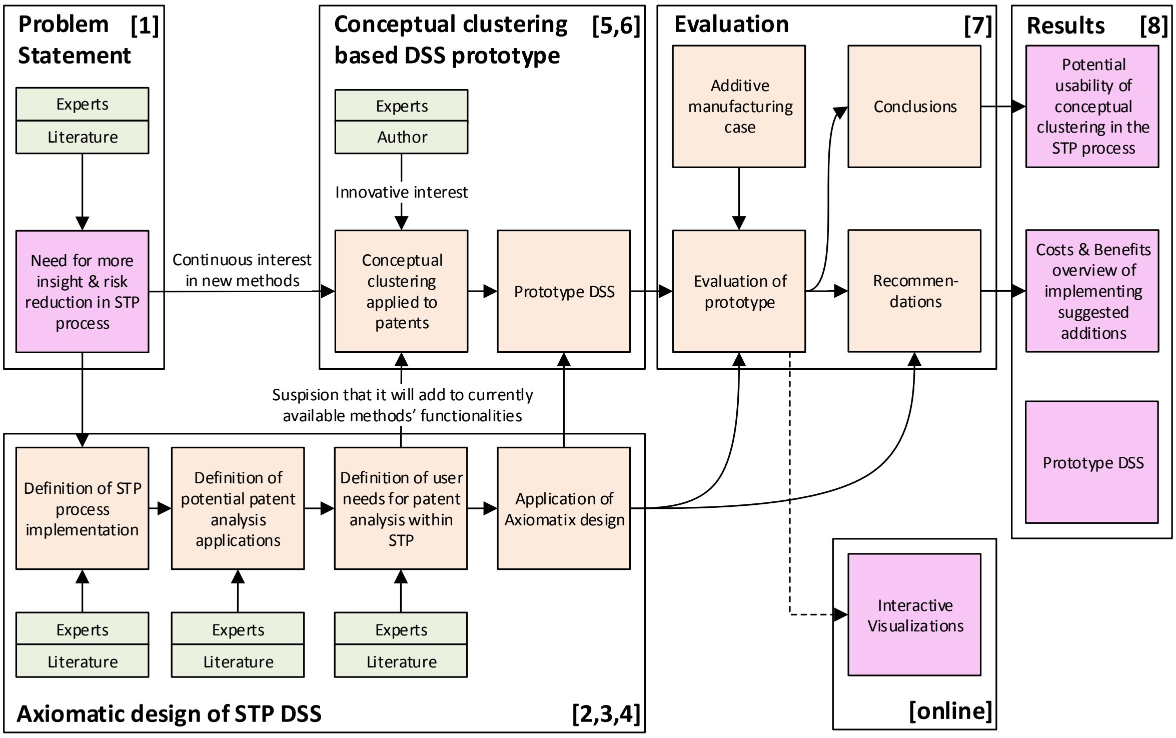


Figure 1.3 – Research flow

The second section covers the development of the new application of Conceptual clustering to the patent field. Which will start with the explanation of the conceptual clustering concept, the COBWEB algorithm and what the expected added values of this technique are when applied to the patent field, in Chapter 5. Followed by describing the actual development of a prototype applying conceptual clustering to patents, the design choices and its structural validation, in Chapter 6.

The third section aims to evaluate the conceptual clustering technique applied to patents as well as for what purpose this new application could enhance the Strategic Technology Planning process. This will be done in Chapter 7, in which the evaluation of the potential enhancement of the STP process is presented based on the designed functional needs and constraints, as defined in the first section, for those elements of the design that are present in the prototype. For those elements that not have been included, implementable solutions are presented together with an indication of implementation costs. Based on this evaluation, conclusions are drawn describing for which aspect of the STP process, conceptual clustering could potentially be used. Secondly a representation, created using the developed prototype, of the technological field of additive manufacturing (3D-printing) is presented. Which is used to evaluate to a certain extend the conceptual clustering technique when compared with the technique of self-organizing maps. Finally, the research' conclusions and recommendations for further work, as well as a reflection, are presented in Chapter 8.

Additionally, in order for the reader that is interested to interact with- or explore the created and described visualizations in this thesis, an online environment has been created to facilitate this. To view and interact with these visualizations, please go to http://www.kevinkruijthoff.nl.

—Research is what I'm doing when I don't know what I'm doing.
Wernher von Braun

2

Decision Support Systems in multi-actor settings

To gain insight into the role of Decision Support Systems in technological organizations. This chapter will introduce the different applied implementations of Decision Support Systems and briefly discus the different theories on which they are built in Section 2.1. Subsequently, Section 2.2 provides an overview of evaluation methods that have been developed specifically for Decision Support Systems.

2.1. THE ROLE OF DECISION SUPPORT SYSTEMS

Throughout decision making in multi-actor networks[2], which can be in intra as well as in inter organizational settings, there is usage of different Decision Support Systems (DSS). The development of DSSs is done to improve the efficiency of the decision making as well as to improve the resulting decision's effectiveness (Pearson & Shim, 1995). As, aside from any political reasons, the more knowledge one possesses on the topic, the less bounded his rationality is. In which knowledge is defined by Dunn as information obtained and interpreted by a policymaker enabling him or her to pursue goals in a changing environment (Dunn, 1994). The (generally computer-based) DSSs allow decision makers through interacting with data and accompanying models to better confront ill-structured problems (McNurlin, Sprague, & Bui, 1989). Thereby enabling its users to process large volumes of quantitative data into an interpretable view. This is why DSSs are valuable tools in complex situations where multiple sources of data need to be analyzed (Martinsons & Davison, 2007).

In 1971 Gorry and Morton defined a DSS as a human-machine problem-solving system, in which a computer system could deal with a structured part of the problem and the decision-maker's judgement would bear the unstructured part of the problem (Gorry & Morton, 1971). This definition clearly shows the DSSs in light of helping the decision-maker through automation, as the structured information is presented in a form which is better interpretable for decision-making. DSS tools are classically comprised of (i) a database management system, (ii) a modeling functions management system, and (iii) a user interface which enables the user to enter its queries and present him or her with the interpretable results (Shim et al., 2002). The model management systems as well as knowledge-based DSSs where primarily based on techniques of artificial intelligence and expert systems (Bonczek, Holsapple, & Whinston, 2014; Courtney & Paradice, 1993). Knowledge based DSSs later evolved into organizational knowledge management (Paradice & Courtney, 1989).

The models on which the DSSs are created can be qualitative or quantitative methods, and some decision support systems even include models based on the theory of organizational decision making of (March & Simon, 1958) and (Cyert & March, 1963). Keen concluded that DSS practitioners should understand how decisions are made and they should be able to relate to the managers perspective (Keen, 1987). In addition Mitroff and

[2] The definition of network that will be used is that defined by de Bruijn and ten Heuvelhof as "a number of actors with different goals and interests and different resources, who depend on each other for the realization of their goals" (de Bruijn & ten Heuvelhof, 2008).

Linstone describe that the stakeholder's perspective based mental models are key to the decision process. They describe these perspectives to be developed from the organizational, personal and technical aspects, as well as to consider less tangible aspects as aesthetics and ethical aspects (Mitroff & Linstone, 1993). As the success of the DSS is related to the integration of relevant perspectives into a model that will create an interpretable result. Model-based decision support development has been divided by Shim into the areas: formulation, solution, and analysis. Referring to the structuring of the problem, applying the derived model, and delivering the model outcome in a usable form (Shim et al., 2002).

Over the last decades, organizational decision making has changed from mainly individual decisions on relatively small issues towards complex decisions needing to be tackled by networks. This has given rise to DSS being adapted for being group support systems (GSS), which allow multiple people within the network to collaboratively make decisions. Using GSS type DSSs allows teams to communicate time and place independently and to make use of group task supporting technologies (Shim et al., 2002). These network type decision making settings have been proven to have more effective interaction with higher effectiveness of decision-making in distributed organizations when making use of a GSS based DSS (Warkentin, Sayeed, & Hightower, 1997). As one of the keys to a successful network focused DSS is the accessibility for all the users within the network, model-driven DSS can clearly benefit from the internet. Web-based DSS reduce the need for specific installment of tools and make it cheaper and faster to make relevant information available in distributed organizations (Shim et al., 2002).

Even though GSS based DSSs have the potential to increase the efficiency in decision making, the actors may disagree on what the problem is and/or not accepts the same forms of evidence as a base for the decision at hand (Enserink et al., 2010). As the individual's objectives are determined by their interest and perceptions, multi-actor complexity arises through the diversity herein (van de Riet, 2003). Thus, the first important decision within the decision making process in networks arises, namely "How do we decide what methods we should use in order to make key decisions?" (Mitroff, 2008). In addition, most DSSs are aimed to provide aid in selecting the optimal solution, which could however be far from doing so as simulations require models, which are by default incomplete and determined by assumptions. When confronted with incompleteness and uncertainties, the search should not be for an optimal but rather for a robust decision with a high probability of success (Chandrasekaran, 2005). When multiple perceptions of reality exist among actors or robustness is searched for, critical multiplism can be applied. Shadish has explained critical multiplism, a technique strongly influenced by the multiple operationalism as defined by Campbell and Fiske (D. T. Campbell & Fiske, 1959), in light of the combination of the two. Where, Multiplism deals with the issue that there are multiple methods to research a certain issue but none of these can be determined to be uniformly best. In these cases multiplism suggest the usage of multiple methods, each subjected to their own biases, to research the given issue. The term critical is used to refer to the effort of identifying the assumptions and biases present in the methods used. The advantage of critical multiplism is that when multiple methods with their own biases produce similar results, the confidence in this outcome increases (Shadish, 1993).

Within the Strategic Technology Planning process, the need for multiplism doesn't necessarily come from multi-actor complexity but rather from the fact that the representation of the technological field is perceived to be represented in different forms. It should therefore be used to determine robust outcomes that hold true under the given uncertainties. Furthermore due to the immense amount of data, it is easy to see why DSSs are being used to deal with the structured part of the problem, making the interpretation for decision making as efficient as possible. Within the STP process there are multiple actors that have different needs at different phases. Given the importance of the perception, needs and abilities of these actors within these phases, it is of great importance to thoroughly research this when designing a DSS for the STP process. Which is why this research will apply the Axiomatic Design method, covered in Chapter 4, which focusses on designing a system by focusing on complying with customer needs throughout the design process.

2.2. THE EVALUATION OF DECISION SUPPORT SYSTEMS

The decision support systems being a human-machine problem-solving system, results in the need for an evaluation process that consist of multiple elements. This is due to the fact that it consists of both an analytical component in the 'machine' part, as well as an application and interpretation component in the 'human' part (Lamy, Ellini, Nobécourt, Venot, & Zucker, 2010). The overall goals of the DSS evaluation are to see whether the system is able to perform its tasks, whether the user needs are met, and whether its users performance improves when using the system (Kirakowski & Corbett, 1990). In their book ‘Handbook on Decision Support Systems 2’, Rhee and Rao summarize applicable evaluation processes for Decision Support Systems, that refrain from purely focusing on empirical examination (Rhee & Rao, 2008). These are (1) the three faceted approach, (2) the sequential approach, and (3) the general approach to DSS evaluation.

As “different needs require different evaluation modes” (Maynard, Burstein, & Arnott, 2001). The three faceted approach deals with the need for multiple types of evaluation based on the DSS element. It makes use of the categorization that Adelman created based on the DSS evaluation needs which are: technical, empirical and subjective facets (Adelman, 1992). The technical facet focusses on the assessment of dataflow, algorithms, logic and the technical testing of the DSS. The empirical facet focusses on the effectiveness of the created DSS as well as on the further improvement of it based on the findings of using it. Finally, the subjective facet, covers the evaluation of those aspects that are not quantifiable, such as the inter-relational effectiveness between the human and machine, the user interface, and the ease of use. Table 2.1 shows the adapted table in which evaluation aspects are shown in the three faceted approach framework.

Table 2.1 – Framework for evaluation objects and criteria, adapted from: (Adelman, 1992; Khazanchi, 1991; Rhee & Rao, 2008)

		Evaluation objects		
		Technical Facet	Empirical Facet	Subjective Facet
Objectivity of criteria	Objective ↑	• Data flow • Application control • Functional operation	• Cost benefit analysis • Utilization information economics	
	↕		• Decision makers’ confidence • Time taken	
	↓ Subjective			• Ease of use • User interface • Understanding

The sequential approach to DSS evaluation has three main development stage dependent evaluations that can be defined. Which are the identification of evaluation criteria, formative evaluation and summative evaluation. Table 2.2 presents an overview of the DSS evaluation process in respect to the DSS development lifecycle, the human decision making process as well as to relevant steps in prototyping design. It aims to provide guidance into when to perform which evaluation, but doesn’t prescribe a specific evaluation method. The formative evaluation should continuously be performed during the design and development testing phase. This is done to detect and eradicate weak points as well as to determine to what extend the current implementation is meeting its goals. Including technical evaluations, expert opinions, and assuring the objectivity during the formative evaluation phase ensures the system output’s reliability (Rhee & Rao, 2008). The summative evaluation aims at evaluating the effectiveness of the final delivered DSS in achieving the goals it was set out to do (Gediga, Hamborg, & Düntsch, 1999; Kirakowski & Corbett, 1990).

Table 2.2 – The DSS evaluation process, adapted from (Rhee & Rao, 2008)

	Sequence ⟶						
Human Decision making process	Intelligence		Design		Choice	Implementation	
DSS Development life cycle	Project Assessment	Problem Analysis	Design	Development Testing		Implementation	Maintenance
DSS evaluation process	**Identification of criteria**		**Formative evaluation**		**Evaluation of system outcome**	**Summative evaluation**	
Relevant steps in prototyping design	Requirement analysis, model analysis		Method selection, Software selection & Design, Transformation			System evaluation, feedback	

In the general approach to DSS evaluation, the key aspect is the determination of which elements to evaluate, and is focusing on the evaluation of the fulfilment of the actual user need. With the quality of the decision outcome, the overall efficiency of the process, and the satisfaction of its users as measurement variables. As the focus lies on the evaluation of the DSS for the specific domain or technology, it uses the so-called spread score, as defined by Adam, Fahy and Murphy (Adam, Fahy, & Murphy, 1998). This is used to determine the extent to which the DSS's usage spread within an organization as well as the complexity of its tasks.

As the general approach to DSS evaluation focusses mainly on the user side, and the main focus of this research lies with the initial exploration of the technique's applicability, this evaluation method will not be applied in this research. This is however advised to be carried out when the prototype is developed into a full-fledged DSS to be integrated in the organizational policy. From the three faceted approach, the applicable evaluation aspects for the process at hand, lay mainly in the technical facets. Regarding the sequential approach, the evaluation in this research will cover the identification of criteria, the formative evaluation and partly the evaluation of the system outcome. The summative evaluation step cannot be performed as this research doesn't fully cover the integration of the DSS into an organization. The evaluation steps that are performed on the DSS prototype developed during this research can be found in Chapter 7.

—An answer looking for a question.
Karl Kruszelnicki

3

Technology and the role of patenting

Before studying the field of patent analysis, it is important to have an understanding of the perceived structure of technology and its relation to patenting as well as of the structure of the patent documents. This chapter will first introduce the different takes on the structural representation of technology in Section 3.1. Secondly the concept of patenting, its relation to technology as well as the patent structure is explained in Section 3.2. In the latter, the advantages of using patents for exploratory analysis is described, and finally the explanation of the patent classification system and its applicability for exploratory analysis is being reviewed.

3.1. The structure of technology

Throughout the literature, multiple widely separated definitions are given to describe technology. In (White & Bruton, 2007) an overview of the multiple definitions of technology is presented: "The processes used to change inputs into outputs", "The application of knowledge to perform work", "The theoretical and practical knowledge, skills, and artifacts that can be used to develop products as well as their production and delivery system", "The technical means people use to improve their surroundings" and "The application of science, especially to industrial or commercial objectives". These definitions all describe technology as involving a systematic process that performs a change to its input, resulting in a rather isolated situation. Contrasting are definitions of technology as a multi-layer dynamic system incorporating economic, social, and geographical circumstances (Watanabe, Zhu, & Miyazawa, 2001) in which technology refers to the whole of isolated functionalities that share attributes, which are described as trends that come and go over different time spans.

In his article "The structuring of invention" (Arthur, 2007), Arthur defines a technology as a means to fulfill a human purpose, and a distinction is made between parts and sub-parts. At the lowest level a technology is described as a functionality being built around a reliable exploitation of some phenomenon. A technology is seen as having a specific purpose, which he names the base concept or base principle, its inputs and outputs. The author uses the term recursiveness to describe the pattern of technologies consisting of multiple other technologies as building blocks to create different functionalities. This multi-level embedding of individual technologies can be seen as a form of hierarchy that by the way they are combined become technologies on their own, this is what the author describes as a working architecture. This view is shared by other researchers that talk about a nested complex system perspective and modularity in product architectures (Murmann & Frenken, 2006). The base principle view on technology will further be used in this document when referring to technology.

3.2. Technology structure and Patenting

A patent is a form of intellectual property, as can be seen in Figure 3.1. However categorized as such it is not actually a property, but rather a negative right, as it blocks all but the one granted this right from using the patented technology. It is an industrial right which protects a technical invention for a limited period of time (in the geographical areas for which the rights are paid). The idea is described as a "package deal", meaning that the patent owner is granted temporary monopoly rights but in return has to disclose the details of his technical invention (Jaffe & Trajtenberg, 2002). The word originates from the term 'litterae patentes' which means 'letters laid open'. This refers to the fact that patent information and the technical knowledge it contains are made available to the public. It is therefore important that the one applying for said patent knows that it will have the exclusive right to commercially exploit the invention (United States Patent and Trademark Office, 2014). That is because otherwise no company would take the risk of putting its resources in the research and development of new products or techniques and have competitors freely benefit. Because of the 'open' character of the patent documents, patent filers often prevent the possibility of reproduction of the invention. This can be done by the usage of legal terms instead of technical terms or by disclosing the essential elements needed for reproduction. As both patent applications as well as patents are published, the moment at which a patent is filed can be of strategic importance for the inventor. The patent information contains both technical as well as legal information. And usually consists of a title page (which contains the bibliographic data), description, claims and drawings. Based on the guidelines defined by the World Intellectual Property Organization (WIPO), individual patent offices apply fixed patent structures (Bodenhausen, 1986). To ensure the ease for patent users of finding relevant fields in the patents over different languages, multiple fields are labeled with INID (Internationally agreed Numbers for the Identification of bibliographic Data) codes, which are linked to specific fields, e.g. 54 always refers to the title of the patent.

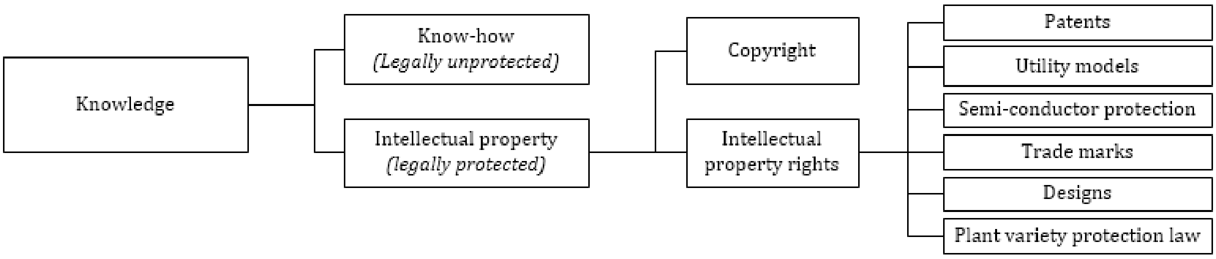


Figure 3.1 – The position of patents in knowledge space (European Patent Office, 2015)

For a technology to be able to be granted a patent, under the law of the European Patent Convention it needs to fulfill multiple criteria. It needs to be a new invention, that involves an inventive step and which can be industrially applicable. It can therefore, before the filing or priority date, not be previously known to the public or be obvious to a person skilled in the relevant field. It is therefore important that before filing for a patent, one does an extensive search into existing inventions. Next to patent literature, it is also important to perform a market research and go through non patent literature such as academic papers, journals etc. to make sure that the invention is indeed novel. As can be distilled from the description, not everything can be patented. At the moment e.g. discoveries, mathematical methods, computer programs and business methods are not seen as inventions. Also, procedures as surgical and therapeutic ones, along with diagnostic methods applied to human or animal bodies are excluded from patentability. Finally, commercial exploitation that is in contradiction with ethical principles are not recognized as inventions (Schröder, 2013).

Regarding the structural components of a patent, the most important bibliographic data are firstly the publication number. This is a unique number which consists of a country code, a series of numbers and a code for the publication status which is called a Kind code. There are multiple Kind codes but common ones are 'A' for a patent application, 'B' for a granted patent and 'E' for reissued patents (United States Patent and Trademark Office, 2013). For example, 'US8905739B2' has the country code 'US' representing the United States, and as can be seen from the 'B2' Kind code it is a granted patent (in this case having a previously published pre-grant publication and available March 2001). Secondly, the data which are related to the application itself. This contains the applicant information, the specific patent office that published the document, the date on which the document was filed at the given office and the application number given to the document. Third, due to the notion of priority, meaning that the same application can be claimed in other states that comply with the Paris Convention within 12 months of filing the original patent (Bodenhausen, 1986). The priority data contains the priority number, the country of priority and when the invention has already been filed before a date of priority. Fourth, the patent classification which is assigned based on the nature of the technological invention. This classification's structure and use will be elaborated in Section 3.2.2. And finally, the inventor's information such as name and geographical location.

When considering the description element of a patent, it should be seen as the subject matter of the invention. It has to describe in detail all aspects as such that it can be carried out by a person skilled in the relevant field. This is imposed to be done so by law. Therefore their content structure is a uniform one and contains the following elements. Firstly, the current state of the art in the field related to the invention is described. Secondly, the gaps such as shortcomings or risks in this current state are addressed, followed by a problem definition based on this shortcomings which the invention will address. Thirdly, the explanation of the invention that will solve the presented gap. Finally, examples of implementation can be included. The structure format of the description is as such that it tries to provide the patent owner with a solid explanation of the technology in case of disputes, as well as to make the new found technology openly available in a way that is understandable for those in the relevant field (European Patent Office, 2015).

The claims section is the legal core of the document and is therefore often considered as the most important part of a patent. The reason for this is that it defines which specific features of an invention are protected by this patent. To be able to have as much protection as possible, these claims are often vague and extensive abstract descriptions for the technical aspects of the claims, when being filed by the applicants. A single claim is a one sentence one, which is divided in what are called the classifying and the characterizing parts. The classifying part describes that part that is not covered by the patent and the characterizing part is the invention applied to the given situation or product (European Patent Office, 2015). The formulation is of great importance when regarding which part and under what condition is protected by the patent, therefore much time is spent on the formulation by the applicant or patent attorneys. Often drawings are used to enhance the understanding of the invention, these are usually black and white ones.

This structured representation that is the patent database can be used to gain insight into the current state of the technology space as it entails a large collection of the currently available technologies. In section 3.2.1 some of the main features of the patent database that make it suitable for the exploration of the technology space are highlighted. An obvious first step when exploring the patent space would be to look at the classification code which is assigned to each individual patent, section 3.2.2 will cover the structure of this code and explain its possible usage in exploring the patent space.

3.2.1. PATENTS AS A TOOL: ADVANTAGES OF PATENT USE IN EXPLORATORY ANALYSIS

As the nature of the patents is that of a structured document within a structured database, which works with global standards, there is a large uniform body of knowledge which can be exploited. This body of knowledge dates back to 1885 and is increasingly expending, as over the period between 1995 and 2013 the number of applications worldwide has significantly risen. This increase has been a 2.5 fold one from just over 1 million in 1995 to 2.57 million applications in the year 2013 (World Intellectual Property Organization, 2014b). A significant share of which is due to the rise in patent applications filed in China, as can be seen in Figure 3.2.

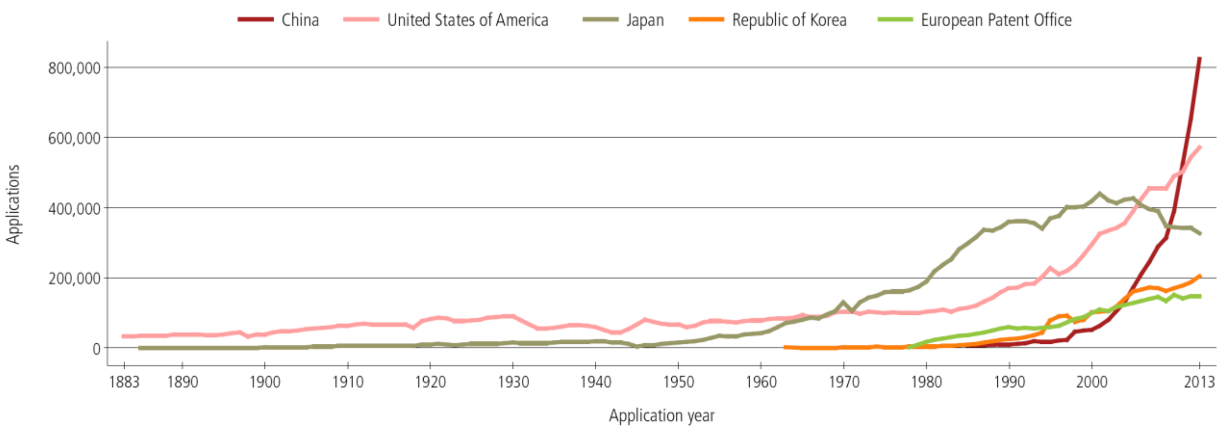


Figure 3.2 – Patent application trend of the top five offices (World Intellectual Property Organization, 2014b)

The ease of large bodies of national and international data being available from centralized electronic systems makes it easy to search for relevant patent documents. As of the legal aspect of the patents, the information provided in the patent is more extensive than in other types of literature sources. Because of this, it also incorporates all relevant aspects of the given technology in detail, which can function as catalysts for other users in the relevant field. However, due to this open character, not all technologies are being patented as a result of strategic decisions, in order to rely on secrecy rather than on the protection of patents.

The fact that inventions are patented in early stages of the discovery of new or improved technologies, together with the open publication of this information makes it an ideal body to detect trends. All applications are published at the latest after 18 months of the priority date (European Patent Office, 2015). There is also no filtering in what is being published, the patent office must publish every patent document, which adds to the completeness of the information body.

An important feature of patents is the reliability of the technical data. For companies interested in the mode of operation of certain technologies, these documents are often the only trustworthy source of information. This has to do with the fact that the content is shared by the inventor in a complete manner and on a voluntary basis (Jaffe & Trajtenberg, 2002). It is important to notice that although the data is easily accessible and complete, that patents are structurally different from other forms of literature. And because of this, specific analysis and classification algorithms are required when applying automated information processing techniques. This is to be able to process the significantly longer texts, different writing styles for the description and claim section, as well as the often present drawings, formulas etc. (Bonino, Ciaramella, & Corno, 2010).

The patent information contains next to the technical and legal information also bibliographical data which is a rich source for market research. For example the inventors and their geographical locations can be used to find out who the key players are in certain fields of technology, or whether certain countries or locations are specialized in the research and development of certain technologies. These insights can for example help governmental bodies such as the European Union in allocation of their R&D budget planning. Patents also include citation information which links them to prior patented technologies or scientific literature. These citations allow for tracing the links between inventions, inventors, firms etc. Based on these citations, researchers make what they call indicators of 'importance' of individual patents and study spillovers of technologies (Jaffe & Trajtenberg, 2002).

Finally, the fact that it consistently exploits the use of a globally applied classification system enables the sorting of the documents according to relevant technologies. How this classification is set up and its applicability for exploratory use will be explained in section 3.2.2.

3.2.2. AN ATTEMPT AT STRUCTURING THE PATENT SPACE: THE IPC CODES

The main form of classification used in patents for obtaining an internationally uniform classification of patent documents is the International Patent Classification (IPC) which has been active since October 7, 1975 and has been structurally reformed for use in the electronic environment in 1999. This classification's main purpose is to serve as a search tool to determine the novelty in patent applications. Other purposes that the IPC serves are e.g. the ordering of documents to facilitate easy access to relevant technical and legal information, the ability of selective spreading of information to relevant users, the definition of the current state of the art in given class, and the creation of a possibility to obtain relevant statistics (World Intellectual Property Organization, 2014a) . As patents often have multiple claims from different technology classes, they are often given multiple classifications.

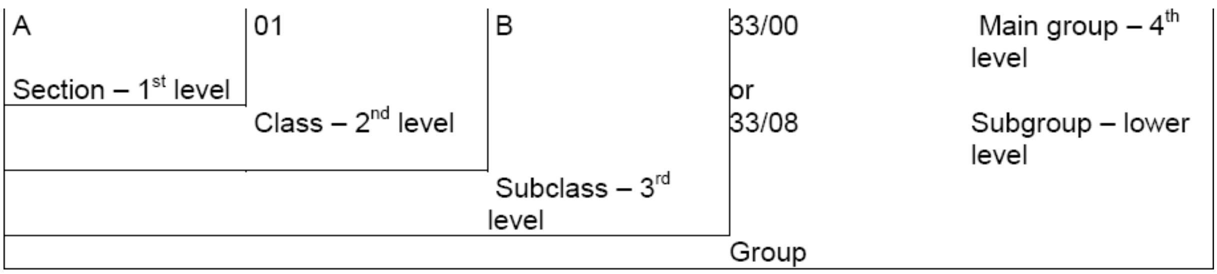


Figure 3.3 – IPC code composition (World Intellectual Property Organization, 2014a)

The classification codes are created based on the merging of the indicators for informational levels. These codes consist of a Section, a Class, a Subclass, and a Group for which their composition into an IPC code can be seen in Figure 3.3. At the top level, the section, there are eight options represented by the capital letters A through H, for which the section titles are shown in Table 3.1. The class level is represented by adding a two digit number to the section code. An item at the subclass level consists of its class symbol (a section code followed by a two digit number) extended by a capital letter. Finally the code can be extended by adding either a main group or subgroups, of which the latter are lower hierarchical levels as compared with the main group. An example of a hierarchical positioning by IPC code is shown in Table 3.2. In which H01S3/14 can be read as "Lasers characterized by the material used as the active medium" as "/14" represents a subgroup of the main group "3/00" that is Lasers. This is possible because the IPC is based on the concept of a hierarchical taxonomy, which means that a patent positioned in a class also belongs to the classes above the assigned class in the hierarchy (Blockeel, Bruynooghe, Dzeroski, Ramon, & Struyf, 2002).

Table 3.1 – IPC Section Codes and Titles

Section symbol	**Section title**
A	Human Necessities
B	Performing Operations; Transporting
C	Chemistry;Metallurgy
D	Textiles; Paper
E	Fixed Constructions
F	Mechanical Engineering; Lighting; Heating;Weapons; Blasting
G	Physics
H	Electricity

Table 3.2 – IPC Code Example

Code symbol	**Section title**
H	Electricity
H01	Basic Electric Elements
H01S	Devices using Stimulated Emission
H01S 3/00	Lasers
H01S 3/14	• characterized by the material used as the activemedium

Regarding these classifications it is important to understand that the determination of the right IPC code by the users is vital for its success as an analysis tool. As ignoring the use of the right subgroups can lead to analysts obtaining information which they assume to be more similar than it truly is. The other way round, when the classification is done into a too deep level, information can be perceived as segregated where in fact there is a degree of similarity (McNamee, 2013). In addition, to be able to add more functionality to their search systems, multiple patent offices have introduced their own classification codes which they include in parallel to the IPC codes in their patent documents and database. This allows them to add additional metadata to the IPC's data enabling even more precise queries. Examples are the ECLA (European Patent Classification System) as used by the European Patent Office (EPO) and the USPC (US Patent Classification System) as used by the United States Patent and Trademark Office (USPTO) (Bonino et al., 2010).

Inherent to hierarchical classifications is the presence of super- and subordinate relations. A superordinate of a given subject represents a higher order or category within the given classification (Oxford University Press, 2015c). This means that those subjects sharing the same superordinate are considered to have a level of compatibility. These relations are embedded within the IPC code structuring e.g. all those patents starting with H01S share, as can be seen in Table 3.2, the fact that they are all technological innovations related to 'Devices using Stimulated Emission'. The classification doesn't however result purely in an is-a-relation hierarchy but rather in an is-related-to hierarchy which contains sub-technologies as well as complete standalone technologies. Thus the fact that two patents are siblings doesn't automatically mean that they are inter-changeable technologies. To be able to detect similar technologies an extra form of analysis is therefore needed. Adding to this is the fact that patents can have multiple IPC codes appointed to them, which results is a lattice like structure instead of a tree like structure, making claims about similarity based on positioning even harder. The common 2D representations of the IPC classification based on the first assigned IPC code might mislead the interpretation by the user due to the simplification of the actual underlying structure. The linguistic equivalent of super- and subordinate relation, is the hypernym and hyponym relation. In which hypernym is a word representing a broader meaning (Oxford University Press, 2015a), e.g. transportation is a hypernym of walking. These hypernym/hyponym relations can be considered as an is-a relationship, meaning “a noun X is a hyponym of a noun Y if X is a subtype or instance of Y” (Snow, Jurafsky, & Ng, 2004). The hypernym aspect becomes relevant when interpreting hierarchies created based on semantic analysis of patents, which is common in many forms of patent analysis representations.

The patent classifications could be used as a tool for exploratory technology analysis as it aims to provide a taxonomy of the patent space, resulting in a topology reflecting the information in the individual classes as well as their relationships to other classes (Schickel-zuber & Faltings, n.d.). This does however limits the user's possibility of exploring to only searches by specific technology applications, and not on other relevant data e.g. the field of application, specific companies, geographical distribution etc. To accommodate these type of explorations, multiple approaches to exploit these other relevant features of the patent documents have been developed and will be elaborated in Chapter 4.

— A good decision is based on knowledge and not on numbers.

Plato

4

Patent Analysis for Strategic Technology Planning

To be able to determine how well the conceptual clustering technique applied to patents will succeed in fulfilling the user's needs, this chapter presents a Decision Support System design. Which is based on the needs for patent analysis within the Strategic Technology Planning (STP) process. As elaborated in Chapter 3, patents are a rich source of technological data combined with relevant bibliographical data. This is why the patent databases can be exploited for many forms of analysis, for which the specific needs will be determined. This Chapter will start in Section 4.1 with an implementation of the STP process which will be used to elaborate on the flow of the process as well as to show the possible positioning of patent analysis within the STP process. Section 4.2 provides per possible patent analysis application, the user needs and constraints which are determined based on literature studies as well as on expert inputs. Section 4.3 presents an overview of common methods for the realization of patent analysis implementations, which will be concluded in section 4.4 . Section 4.5 will provide a brief introduction into the design method, called Axiomatic Design, which is used to design an implementation of patent analysis for the STP process. The actual application of the axiomatic design is presented in Section 4.6, resulting in an implementable design which will also be used in the evaluation of the actually developed prototype.

4.1. The Strategic Technology Planning process

Figure 4.1 shows the Strategic Technology Planning (STP) process, which aims to find novel technology concepts and develop them into full-fledged technologies. It presents 'a' representation of the STP process, for which alternative process implementations also exist. However, the interest for this research lies with the functional needs for the different patent analysis types, which can be considered equal no matter how the STP process is implemented. The presented representation will be used to elaborate on the STP process use as it comprehends the possible different patent analysis steps within STP processes.

Starting from the top left process step, the STP process flow starts with a research motivation in the call for research stage. This can be done for multiple reasons as well as by multiple actors. An open search can be performed to gain insight into the structure of a certain technology field, in order to gain inspiration for new technologies. A research can be done into the company's competitors, to determine who they are, in what fields they are active, and what can be learned from this, or whether they can form a threat. A Trend analysis can be performed to forecast needs based on extrapolation of innovation over time. These technology Forecasting tools are used to gain understanding in the direction, rate, characteristics and effects of technological change (Firat, Woon, & Madnick, 2008). A search can be performed into a certain functionality in other technological fields to research whether these could also apply in the specific technological area of interest by the analysis and categorization of patents into reoccurring trend phases (Verhaegen, D'hondt, Vertommen, Dewulf, & Duflou, 2009). And finally, a policy driven search can be performed based on the company's strategy, to look into the current portfolio and the gaps that it has compared to the company's set technology strategy.

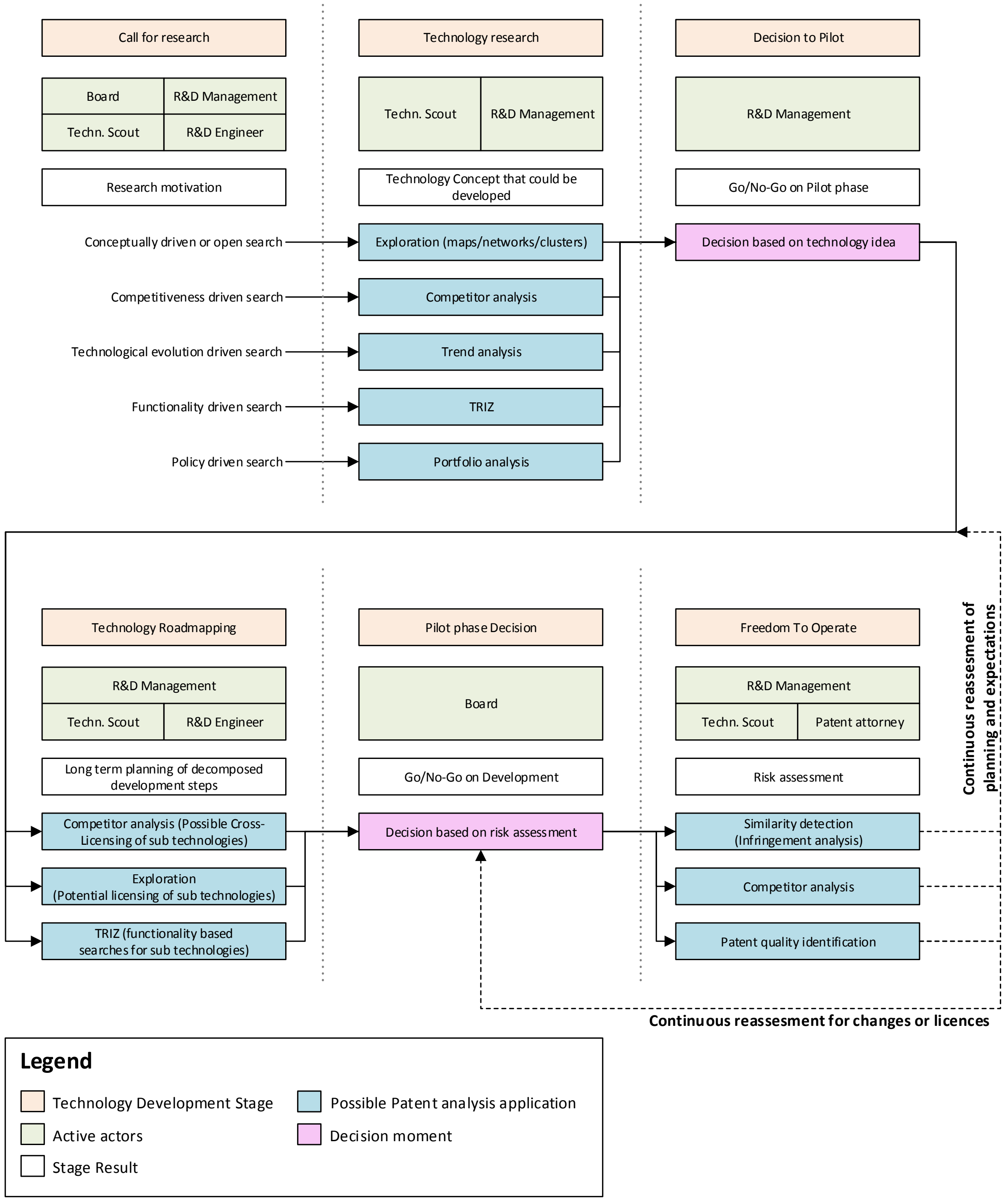


Figure 4.1 – The possible positioning of patent analyses in Strategic Technology Planning

In the second stage the actual technology research is performed by the technology scouts and the R&D managers. This is partly done via multiple patent analysis applications, for which their functional needs will be individually elaborated in Section 4.2. The aim is to produce a technological concept that could potentially be developed. The third stage is the decision by the R&D management whether or not they see the resulting technological concept worthy enough to move onto a pilot or prototype of the concept. Once the proposed technological concept is approved for development into a pilot or prototype, the technology road mapping starts in which the proposed technology is broken down into sub-technologies for which a long term planning for development is made. From this point on decisions are made to further develop the technological concept and research its perceived risks and benefits in the Freedom to Operate stage. In the Freedom to Operate stage, multiple actors work on assessing the risks and benefits of developing the specific proposed technology. Evaluations are done into whether or not the proposed technology is infringing another company's patent. And if so what the likelihood is that the infringed company would sue, or whether an innovation license can be agreed upon. Licenses can be granted for multiple reasons e.g. when the other party is more likely to exploit the innovation, when companies focus on R&D and the licenses are used to generate revenue, used as a strategic tool to stimulate market demand (Shepard, 1987), or to prevent other companies from developing a competitive technology (Gallini, 1984). Interesting to notice here is the notion of cross licensing, a situation in which companies both have patents to block certain of each other's technologies but mutually agree not to (Shapiro, 2001). At the same time Patent Quality Identifying tools can be used to determine the strength of a certain company's own patents. But they can also be used to assist organizations beforehand when deciding on investment options of new products or to maintain legal authority over their existing IP (Trappey, Trappey, Wu, Fan, & Lin, 2013). The result is an overview of the perceived risks as well as the potential licensing options, on which the board can reconsider its decisions on the Go/No-Go of the development of the proposed technological concept. During the development of the technology a continuous reassessment has to take place on the risks as well as on the planning to make sure the technology being developed is still relevant and on track.

4.2. THE FUNCTIONAL NEEDS FOR PATENT ANALYSIS TOOLS WITHIN THE STP PROCESS

This section will provide the functional needs for the independent patent analysis applications within STP. The functional needs and constraints will be elaborated based on an ideal situation, which are not necessarily commercially available at the moment. The described needs are derived from literature studies and have been adjusted based on expert input, for the latter the summary of the discussion can be found in Appendix I. The functional needs will be presented per analysis type in alphabetical order.

COMPETITOR ANALYSIS

In patent analysis for competitor analysis, insight is gathered into the positioning of competitive companies within the technological field of the company. Which is done by exploring potential competitors and estimating their developments, to predict threats or to analyze possible mergers & acquisitions, alliances or cross-licensing (Ernst, 2003). The goal of each competitor analysis is however dependent on the position within the STP process. In the research phase, a search can be done with the goal of finding out **what** competitors are doing. Which is done by determining in which areas of the technological field competitors are developing technologies. It is therefore important to get an overview of the relevant technological field, within which the patents of a certain company are shown. The ability for an additional external data layer to be placed over the technological field is therefore needed. Hereby the individual patents of a certain company can be highlighted, to investigate the actual information of these patents. It would be of assistance to be able to access the individual patent's bibliographical information once they are located, to be able to quickly scan the located technologies. When the patents of the own company are added, a comparative study can be done into the presence in certain areas of the technological field. To be able to draw conclusions, it is important that it is clear how this technological field is structured and why certain patents are located at their specific location. In the Freedom To Operate phase a competitor analysis can be performed to determine **who** is a possible threat. This is done by determining whether a competitive company has a technology in close proximity of the to be developed technology. A reverse search can here also be performed by virtually placing the technology in the patent space and see which patents are closely related. Then determine to which company they belong to, and based on the relationship with this company, determine whether to take any precautions. In the case of non-hostile companies, they can be flagged into possible future competitors. In the Technology Road mapping stage, competitor analysis is done to find out **how** competitors could be used to advantage. Which can be done by determining whether there are potential openings for cross licensing of sub-technologies.

EXPLORATION

When insight into a specific technological field is aimed for or an open search is being performed, exploratory patent analysis can be applied. In this type of analysis, the structural relations of a certain technological domain are presented (B. Yoon, 2010). In this manner, those who have no prior knowledge of the field can obtain knowledge by interpreting the representation (Rafols et al., 2010). It is therefore necessary to create an overview of the relevant technological field, and make clear why certain patents are located at their specific location. As the outputs of this type of research are often used for communicative purposes, the ease of interpretability is an important criterion.

PATENT QUALITY IDENTIFICATION

The quality of the patent determines the protection it provides against competitor infringement as well as to what extend it enhances the invention's value. A patent quality analysis can thus be used to determine which technology is likely to add value and which could pose a threat. To determine the value of a certain patent, it is compared to patents in its field (Trappey, Trappey, Wu, & Lin, 2012). In the Freedom to Operate stage, a patent quality identification can be performed on patents that are in close proximity to the proposed technology on the indicators related to litigation. This can be relevant as it has been noted that high valued patents are more likely to be the start of, and to win, lawsuits (Allison, Lemley, Moore, & Trunkey, 2004). Relevant indicators for the analysis of patents are the number of claims made in the patent (Trappey et al., 2012), (forward and backward) citations (Hegde & Sampat, 2009; Trajtenberg, Henderson, & Jaffe, 1997), and the uniqueness of the assigned IPC (Lanjouw & Schankerman, 2001).

PORTFOLIO ANALYSIS

When a search is performed based on the company policy, a portfolio analysis can be done to see whether the current patent portfolio complies with the strategy set by the company (Ernst, 2003). Therefore a representation of the relevant field is needed, together with the positioning of the patents that are in the current portfolio of the company. To see whether this complies with the company's strategy an overlay of non-patent data is needed.

SIMILARITY DETECTION

Within the Freedom To Operate phase, the similarity detection is one of the most important steps as it presents the patent attorney with the patents that are significantly similar to the to be developed technology. This is of great importance given the possible financial risks of patent litigation (Abbas et al., 2014). The most important feature is therefore to determine the similarity between the concept and the existing individual patents. The found patents must be able for evaluation and therefore show the patents content. An overlay of non-patent data would enhance the process as the relationship to the patent holder could be shown. Which for example can contain the likelihood that the patent owning company will sue or will agree upon a license, thereby increasing the handling speed of the FTO process.

TREND ANALYSIS

Trend Analysis tools can be used to give insight to the development of technologies over time. It can be exploited to perform analytical studies of the emerging, maturing and passing away of technologies. It can thereby indicate the growth pattern of a technology and give insights into occurring shifts within technology (R. S. Campbell, 1983). For policy makers this information is a relevant and important base for their policy decisions, as it has been found that the propensity to patent as well as research productivity are influenced to large extend by the design of policies regarding education, intellectual property and science and technology policies (de Rassenfosse & van Pottelsberghe de la Potterie, 2009). Therefore, the developments within a certain technological field are investigated over time, to determine trends. To be able to do so, a representation of the technological field needs to be shown as it develops over time (in which patents are added to change the existing field according to their date). In which it is necessary to be able to extract how the new entries relate to the existing field. By analyzing how the field changes over the set time span, predictions can be made about future needs.

TRIZ

When a solution is searched to realize or improve a certain functionality with a technology, the TRIZ method could potentially be applied. TRIZ is a Russian acronym that translates into "Theory of Inventive Problem Solving" and is a method developed by Altschuller and his colleagues. The core principle relies on the assumption that problem-solution patterns exist within technology which forms the base for technology transfer. Meaning that the essence of the problem you are facing has most likely been solved before (in a different field), the goal is to find it and adapt that solution to the problem at hand. Both analytical as well as prescriptive methods have been developed to apply TRIZ, of the latter the 40 inventive principles of problem solving is the most accessible concept (Barry, Domb, & Slocum, 2008). As the TRIZ principle focuses on the use of the generalization of a problem, for which a general solution is searched, there is a need to determine the basic functionality of the individual patents. They should be analyzed for their key concepts and the components structural relations, which should then be compared with the needed functionality (Park, Ree, & Kim, 2013). Thus there is a need for grammatical analysis of the patent unstructured text (J. Yoon & Kim, 2011).

4.3. METHODS OF IMPLEMENTATION OF PATENT ANALYSIS TOOLS

The automation of the patent based decision support systems is generally done in two main steps. Firstly, the application of text-mining techniques, which are used to extract the information from the structured and/or unstructured text of the patents. Secondly, the usage of visualization techniques, which are meant to assist the decision maker or technology experts by representing the extracted patent information visually, enabling them to get an interpretable overview of the relevant aspects of the patent space. Important to notice here is that the user's need, based on its role (e.g. legal, technical, managerial) and the positioning within the STP process, defines what the decision support system needs to be able to extract and represent interpretably to its user. This section aims to provide insight into the currently used techniques for patent analysis when considering these two elements.

Due to the rising numbers of patents it becomes hard for experts to be certain that by manually reading up on the patent office notices, they have indeed a clear picture of the relevant area of the technology space. At the same time they need to be able to deliver, with great confidence, outcomes in which they can confirm that all relevant patents have been analyzed for possible infringement. The aim from a managerial perspective in decision-making is to have as less risk as possible regarding competitors and infringement. This is because advises have to be given to the board of directors on Strategic Technology Planning, including areas of technological opportunity and thereby also clearly indicating its perceived risk assessment outcomes, thus are we confident to present and defend these searches and their outcomes e.g. when it comes to court cases? To aid experts in this process, automation can be introduced into certain steps of the process, in which the goal is to create a clear overview of the relevant technology space which can be interpreted by experts and can assist them in their decision-making process. This automation is generally done in two ways; application of text-mining techniques, which are used to extract the information from structured and/or unstructured text, and the use of visualization techniques, which are meant to assist the decision maker or technology experts by representing the extracted patent information visually, enabling them to get a better overview of the patent space.

4.3.1. TEXT-MINING

Text-mining is a term used for tools that analytically extract meaningful information from structured and/or unstructured natural language text and hereby possibly extracts useful patterns (Abbas et al., 2014). The text-mining techniques are categorized by Abbas, Zhang and Khan in the literature study on patent analysis (Abbas et al., 2014) under; Natural Language Processing (NLP) based approaches, Semantic analysis based approaches, rule based approaches, property function based approaches and neural networks based approaches. To perform such text-mining techniques (Trippe, 2003) presents an overview of applicable tools, which also covers techniques for data visualization, in what the author describes as the field of patinformatics. Additionally, the paper (Tseng, Lin, & Lin, 2007) presents text-mining techniques for patent analysis in which the authors provide an overview of the methods which can be used in segmentation of textual data into meaningful structures.

SEMANTIC INFRINGEMENT ANALYSIS

One of the aspects of Strategic Technology Planning is the detection of potential infringing of other companies patents. As concluded in (Bonino et al., 2010) semantic technologies can play a relevant role in patent analysis, by simplifying and improving the patent search and analysis processes. This chapter provides an overview of applications of semantic search algorithms, to detect patents with a significant similarity, which are currently implemented in the field of patent infringement searches.

In (B.-U. Yoon et al., 2002), the authors introduce a self-organizing feature map (SOFM) based patent map (PM) that visualizes the complex relationship among patents and the dynamic pattern of technological advancement which is created by the use of so-called keyword vectors. In the data pre-processing the patent's text is filtered to remove supplementary words, stem words are identified and collected, these are then statistically analyzed and based on the outcome of this analysis a hierarchy is formed. The then created keyword vectors are based on the frequency of keywords co-occurrences and are compared to those of other patents and assigned values. The authors following claim is that through visualization of the valued patents on a patent map it becomes possible to discover areas of patent infringement based on the distance between patents on this map. Downsides to the presented method are due to its nature to focus on individual words. The outcome could be influenced by the writer's style, patents from the same author could be grouped together while covering different contents where conceptually close patents can be distant from each other when different terminology is applied. The method deals with occurrence of keywords and only considers the frequency of these, so the statistics may not reveal the internal structural relationships between patents (Cascini & Zini, 2008).

The Subject–Action–Object (SAO) technique provides a partial answer to the problem of the keyword vector searches lack of identifying the actual concept being patented. An SAO structure explicitly describes the structural relationships among technological components in the patent, and the set of SAO structures is considered to be a detailed picture of the inventor's expertise, which is the specific key-findings in the patent (Park, Yoon, & Kim, 2011). So instead of determining the frequency of occurrence of words and their synonyms only, as is done in the keyword vector technique, the text is analyzed for combinations of means (Subjects) and ends (Actions and Objects) (Bergmann et al., 2008). In this way, an SAO structure explicitly describes the structural relationships among technological components and fully reflects the specific key-findings in the patent, so an SAO based technological similarity is suitable for identifying patent infringement (Park et al., 2011). A downside to the method as presented is that the synonyms which are used for the semantic similarity measurement method are still to be defined by an expert-based analysis which leaves room for humanly induced errors and could be time consuming. Furthermore this method cannot readily be used to create an overview of the entire patent space, which could be useful in the infringement analysis to identify the actual field of use.

In the paper (Bergmann et al., 2008) on applying SAO for patent infringement analysis, which implementation is commonly used in the literature as a base to expand their SAO method upon, the authors implement the usage of the SAO structure analysis in their potential infringement detection process which is a multi-stage analyzing process. The process performs on the selected patents the following general steps: Natural Language Processing using an SAO semantic processor implemented in the software Knowledgist 2.5 (to create the semantic structures), domain specific analysis of speech (the inclusion of synonyms), applying of a patent informational analysis tool (the creation of similarity matrices) and a situation-specific statistics and multivariate process (to create the patent's coordinates) with as an output the patents presented on a patent map. The graphic representation of the patent map is done by means of a multi-dimensional scaling (MDS) and can be used to detect patents for further examination by the expert. Downside to this implementation is that there is still a need for experts in defining the synonyms for the SAO output structures.

Using the SAO methodology output as a base, functional trees are created of the patents which are then compared to other patent's trees to determine the values inferring similarity. This implies that the similarity between two patents is estimated by comparing their components, hierarchical relationships and functions (Cascini & Zini, 2008). The authors suggest to evaluate the similarity between two patents by comparing their functional tree, i.e. the hierarchical architecture of the invention's components and their functional interactions. Working with the functional tree allows to identify conceptual similarities and to limit the influence of the language style (Cascini & Zini, 2008).

In their paper (C. Lee, Song, & Park, 2013) argue that there are problems with the existing automated patent searches which are: Data limitations, Method limitations and Practicality issues. They argue that dependency relationships are not yet investigated and taken into account, however the aforementioned SAO techniques provide a solution for this. Nonetheless do they also provide their own solution to this problem which is a process of data collection and preprocessing (data parsing), transforming patent claims into hierarchical keyword vectors (structural text mining) and examining the possibilities of patent infringement (a tree matching algorithm). In their method the patent claims are transformed into a hierarchical keyword vector after the relations among claim elements are identified by common phrases. These vectors are then used in a tree structure of claims with binary values for the keywords (if the keyword occurs in the claim the value 1 is assigned otherwise 0). The trees of patents are compared for similarity based on three sub-steps: matching claims and adjusting structures, measuring component similarities and deriving final similarity. The tree matching method seems to be similar to that as presented by (Cascini & Zini, 2008) however the base here is a keyword vector which is created by analysis using common phrase structures.

To solve the presented problem of the need for expert-based analysis approach in the methods of (Bergmann et al., 2008) and (Cascini & Zini, 2008) to define synonyms when using the SAO method, (Park et al., 2011) present an implementation which uses WordNet as a replacement for this expert analysis. WordNet is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept (Princeton, 2014). By replacing the step of an expert manually determining synonyms by an automated procedure this can result in an immense time saving. The procedure the authors apply is based on the technological similarity among patents, and considers this as the criterion for judging the possibility of infringement. Their procedure consists of the following steps: extraction of SAO structures from dataset using NLP, measurement of the semantic technological similarities among patents using WordNet, generation of a 2 Dimensional semantic patent map using multidimensional scaling and finally the analysis of the automatically generated clusters to identify possible patent infringement. Downside to the presented methodological solution of implementing WordNet as a replacement for the expert analysis in the step of synonym creation could indeed be time saving, however as the authors also mention it cannot be guaranteed that the WordNet database has sufficient relevant synonyms in specific cases of patent analysis.

Interesting to mention for its use of semantic text-mining techniques applied to patents is the paper by (Fantoni, Apreda, Dell'Orletta, & Monge, 2013) which discusses a method of extracting conceptual information from patents resulting in a fast interpretable outcome, this is however not linked yet to comparison for possible patent infringement analysis but an extension towards this can be easily imagined.

4.3.2. VISUALIZATION

To aid the expert analyst in interpreting structures found by the text-mining techniques to conduct further analysis upon, different visualization techniques can be applied. As presented in (Rafols et al., 2010), there are multiple advantages of using visualizations. First, the facilitation of interpretation for those with no prior knowledge of the field, which however comes with the downside of the possibility of manipulation by strategically structuring the visualization. Secondly, given its representation the information can be packed into a dense and easily view-able structure. And third, the ability of combining multiple types of data into one visualization. In (Abbas et al., 2014) the visualization techniques are categorized in: Patent networks, Patent maps and Data clusters. Visualization methods that are currently used are, among others, Claim point maps (B.-U. Yoon et al., 2002), Keyword-based patent maps (S. Lee et al., 2009), Patent claim maps (Shin & Park, 2005), Semantic patent maps (Bergmann et al., 2008), Patent topic maps (Kasravi & Risov, 2009) and using social network analysis applied to patent citations (Sternitzke et al., 2008). In the review article (B. Yoon, 2010) on patent visualization, the authors conclude that the best way to visualize data for a patent infringement analysis is done via the usage of a Technology map or a Technology network.

PATENT SPACE MAPPING

An often applied method of visually representing technology is what is referred to as science maps, a method developed in 1973 by Small (Small, 1973) and focusing on bibliographic data. In this method the maps are generated based on matrixes that contain similarity measures between documents, which are than visualized on a two or three-dimensional space (Rafols et al., 2010). They are representations of scientific fields or organizations that contain elements which when having similar aspects are placed nearby and otherwise relatively further (Noyons, 2001). In their paper (Rafols et al., 2010) the authors provide an addition to this method by applying science maps in which they add the possibility of overlaying maps adding extra insight into e.g. where companies are positioned. There are however critiques to science maps/(co-)citation mapping as it is claimed that the interpretation of these maps is difficult and that they are showing 'pathways' instead of the structure of technology (Rip, 1997). The latter is due to the fact that it mainly focuses on the representation of bibliographical data and not on the content of the actual documents.

As described in (B. Yoon & Park, 2004), network analysis is a quantitative technique derived from graph theory and is first described by (Knoke & Kuklinski, 1982) as a manner of mapping social networks. It consists of actors and interactions, in which the actors in this case will be represented by patents. The network analysis will structure the relations among actors and their locations based on their behavioral, perceptual, and attitudinal aspects and that of the system as a whole (Marsden & Laumann, 1984). In their article (Chang, Wu, & Leu, 2009), the authors apply a methodology in which they first perform a patent bibliometric analysis which output is then used in a network analysis to identify technological trends in Carbon Nanotube Field Emission Display.

Interesting aspect mentioned is that of a Technology Cycle Time referring to the time interval between two given patents, to give an indication of technological trend development. Presented critiques are that large volumes of patents may cause difficulties when generating the network and that the resulting network might be meaningless when the relation between patents is not clear (Chang et al., 2009).

Self-organizing maps (SOM), based on a theory first presented in (Kohonen, 1982) using neural networks techniques, are used in patent analysis to reduce the dimensions of a dataset as such that it can be visually presented as humanly interpretable data. The current method of use of this method is that of an evolutionary one where data is put in and a map is generated, which is then analyzed with experts and managers of which the resulting conclusions are then used to make a new and improved query, excluding that part which is irrelevant. A recent application of SOM applied to the patent field is that presented in (Segev & Kantola, 2012), in which its implementation is based on patent descriptive vectors of weighted attributes. Followed by a partially automated evaluation of the resulting structure by making use of a Unified Distance Matrix to gain insight into the underlying relations. Downsides of this methods are firstly the fact that the maps created with this technique are of questionable reliability when it comes down to reproducibility. This is due to the fact that in every run they can produce an entirely different output based on the same input. Second, a morphological approach, working from what the parts of a technology are, to make sure that all that is potentially relevant is selected, is something that is not possible using this technique. However, one can easily imagine that it would add value to search for a broad topic, such as electrical vehicle, and automatically be able to find all its relevant sub-systems technologies.

In their paper (Kim, Suh, & Park, 2008), the authors present a method for patent analysis using a k-means clustering algorithm which is described by MacQueen (MacQueen, 1967). The structure resulting from the k-means clustering performed on the attributized patents (based on the bibliographical as well as on the unstructured data) is used to form a so-called semantic network of keywords which is arranged by filling date and patent frequency. Downside in the prescribed method, is that there is a need for expert input prior to the analysis and the setting of semantic clustering groups. Furthermore it is concluded that the implementation could face issues regarding the k-means clustering as sizing and empty clusters (Abbas et al., 2014).

4.4. CONCLUSIONS AND SHORTCOMINGS

Due to the rising amount of patents, automation is applied in the analysis process, which can be separated into two main groups. These groups are text-mining techniques and visualization techniques. Text-mining techniques can be used to attributize the unstructured text to be parseable by visualization techniques, but also for other relevant uses such as semantic infringement detection. Visualization is mainly used to ensure a faster and easier interpretation of complex relations. Relevant to take into account is that the amount of information on a certain patent as well as the nature and the quality of it will change significantly over its lifespan (Bonino et al., 2010). The temporary informative value needs to be taken into account when analyzing any made patent analysis. Furthermore, it must be noticed that the patent applications should also be considered into the patent analysis as their filing date will be the date considered in the case of infringement, here raises the question of completeness as there is a period in which this information is not made public.

Considering the presented visualization techniques and their applicability for Strategic Technology Planning, multiple shortcomings can be detected for one or multiple of the presented patent visualization techniques:

1. First, the non-reproducibility of the outcome which is of great importance from a legal perspective. This goes together with the possibility of traceability of the decisions made, in other words how did it arrive at the given outcome and why is a certain patent at that specific location.
2. Second, the lack of hierarchical relations which are present in technology itself but are not visually present.
3. Third, the possibility to view the visualization evolving over time in order to detect trends or the creation of new technological clusters.
4. Fourth, the ability to guide the search as a user in order to be able to interact with the dataset allowing the possibility to bias or seed the search from the start.
5. Finally, the possibility of customizing the visualization based on the need of its user. In the latter multiple examples can be given e.g. the highlighting of the own patent portfolio to see the positioning of the company within the patent space, the highlighting of companies based on the level of threat they pose onto the company or simply highlight certain words to see in which fields of technology they occur.

As an answer to the shortcomings presented in the currently available analyses and visualizations of the patent space, together with the structural representation of technology and patents, the suggestion is made to investigate the applicability of the conceptual clustering technique to the patent space. This is suggested as it is a reproducible technique that creates a hierarchical structure, of which the output allows for a highly customizable visualization. As one of the most important aspects in knowledge management systems is to extract and put into use the information which their databases contain in an effective manner (Abbas et al., 2014). As well as the impact of the coherence of the different steps within its process. A design for a decision support system for the STP process will be presented in Section 4.6, which is built around the exploitation of the conceptual clustering technique. The concept that is conceptual clustering as well as the specific conceptual clustering algorithm used will be elaborated in Chapter 5.

4.5. A BRIEF INTRODUCTION INTO AXIOMATIC DESIGN

The transformation from user specific needs for patent analysis within the STP process into an applicable design will be done by making use of the axiomatic design method. Axiomatic Design, as described by Suh (Suh, 2001)[3], is a design method that is based on the confirmation with two axioms[4]:

1. **The Independence Axiom**: For a design to be acceptable the adjustment to satisfy one requirement should not affect any other requirement, thereby maintaining independence of requirements.
2. **The Information Axiom**: When multiple solutions are present, choose the one with the maximum probability of success. Minimize the number of elements in the needed solution.

By applying the Independence Axiom, a scientific infrastructure is provided which is able to reduce the time resulting from trial-and-error methods and is therefore becoming more widely applied in the industry and academics (Kulak, Cebi, & Kahraman, 2010). The Information Axiom, maximizing the probability of success and minimizing the number of elements is comparable with the exclusion of waste within lean design (Shirwaiker & Okudan, 2011). To comply with these two axioms, the Axiomatic Design method acknowledges the existence of four domains which are exploited to establish a design. The process starts with user needs and ends with an implementable overview of elements. The design process is done by mapping the consecutive domains, as shown in Figure 4.2, which is done by the use of matrices, each time mapping the adjacent phases. The design takes shape when zigzagging between them while designing, and changing the content in the different domains based on new insights gathered in the adjacent domains. The first domain is the customer domain, in which the Customer needs (CAs) are defined. These are mapped onto the second domain, the Functional domain, which contains the functional requirements (FRs). Third is the Physical domain which contains Design Parameters (DPs). And fourth, the Process domain which contains the process variables (PVs). Next to these domains, Constraints (Cs) are defined to which the design should comply.

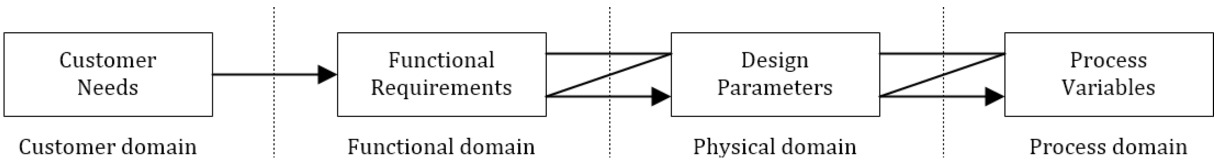


Figure 4.2 – Axiomatic Design domains (Suh, 2001)

The tables presented below show three types of possible mappings between functional requirements and design parameters. Table 4.1 shows what is called a diagonal design, in which for each individual FR there is an individual DP. Table 4.2 shows what is called a bottom triangular design. In both the diagonal as well as the bottom triangular design, the functional requirements can be met independently by the design parameters. Which is considered a good design as it complies with the Independency Axiom, this is referred to as uncoupled design. Table 4.3 however, is neither a diagonal nor a bottom triangular one and is therefore considered as a bad design. It is considered to be a bad design as it doesn't comply with the Independency Axiom, resulting in what

[3] A comprehensive description of axiomatic design including multiple design examples is presented in (Suh, 2001)

[4] Axiomatic: obviously true and therefore not needing to be proved (Cambridge Dictionaries Online, 2015)

is called a coupled design. In coupled designs one cannot comply with a certain requirements setting without changing that of another. A clear example of a coupled design, as presented by Suh, is the traditional water faucet in which the flow and the temperature cannot be separately changed. But rather by changing the temperature, the flow of the water is also changed (Suh, 2001).

Table 4.1 – Diagonal mapping

	DP 1	DP 2	DP 3
FR 1	X		
FR 2		X	
FR 3			X

Table 4.2 – Bottom triangular

	DP 1	DP 2	DP 3
FR 1	X		
FR 2	X	X	
FR 3	X	X	X

Table 4.3 – Bad design

	DP 1	DP 2	DP 3
FR 1	X	X	
FR 2	X	X	
FR 3	X	X	X

In the literature review by Kulak, Cebi and Kahraman covering Axiomatic Design (Kulak et al., 2010), the authors discovered that Axiomatic Design is mostly used in application-based situations and usually focusses on the compliance with the independence axiom. They observe that, next to the usage in the design phase, it is often used as a type of evaluation within product design and that applying it in fuzzy evaluation based research is growingly applied to multi-attribute decision making. Axiomatic Design is said to aim at "making human designers more creative, reduce random search process, minimize iterative trial-and-error process, and determine the best design among those proposed" (Shirwaiker & Okudan, 2011).

In his book, Suh presents a specific application of Axiomatic Design for the development of software. However since decision support systems are a human-machine system, it cannot be considered to be a pure software design. For this reason the use of the axiomatic design in this research will not adopt the suggested approach for software design, but rather the presented more general approach.

4.6. AXIOMATIC DESIGN OF PATENT ANALYSIS FOR STRATEGIC TECHNOLOGY PLANNING

As the application of the conceptual clustering of patents is a novel one and there is thus no reference on its potential applicability within the STP process. For this research a design of a patent analysis tool for the STP process will be created based on the needs and constraints within the different applications of patent analyses, as presented in Section 4.2. This will however be done with one bias from the start, which is that of the application of conceptual clustering for the determination of inter-patent relations. The purpose of the resulting design will be twofold, firstly it will give insight into the specific user needs within the STP process for patent analysis and the functional requirements that these imply. And secondly, it will allow for an evaluation of the implemented prototype of conceptual clustering applied to patents. This evaluation, which can be found in Chapter 7, will be done by a reversed analysis of the Process Variable groups and tracing them back in reversed direction to evaluate which customer needs' it can (potentially) fulfill. The rest of this section will present the final mapping of the design, resulting from the multiple zig-zagging iterations done through the domains. Alongside which explanations about the process and choices are presented. Starting with Table 4.4 which presents the Constraints (Cs) to which the chosen solutions need to comply.

Table 4.4 – Constraints

Constraint	Description
C 1.	Ease of use (Based on user profile)
C 2.	Robustness of outcome
C 3.	Reproducibility of outcome
C 4.	Explainabilty (Interpretability of outcome)
C 5.	Usability as a communication tool

The customer needs (CAs) resulting from the analysis as presented in Section 4.2 are listed in Table 4.5 and are mapped onto the high-level functional requirements (FRs) of the to be developed analysis tool of which the description is presented in Table 4.6.

Table 4.5 – CA to FR mapping

		FR 1.	**FR 2.**	**FR 3.**	**FR 4.**	**FR 5.**	**FR 6.**	**FR 7.**
How is the technological field structured?	**CA 1.**	x						
What is in the portfolio of a specific company?	**CA 2.**	x	x					
Who are competitors in certain technological areas?	**CA 3.**	x	x	x				
What are the risks of developing a certain technology?	**CA 4.**	x	x	x	x			
Is there an applicable solution for a specific problem?	**CA 5.**	x	x	x		x		
Does the portfolio comply with strategy policy?	**CA 6.**	x	x				x	
Where is the technological field expectedly heading?	**CA 7.**	x						x

Table 4.6 – Functional Requirement Descriptions

FR	**Description**
FR 1.	Representation of the technological field
FR 2.	Map patents of a certain company
FR 3.	Present an overview of patents ownership
FR 4.	Facilitate Freedom To Operate process
FR 5.	Present patents that possess technologies providing solutions for comparable problems
FR 6.	Represent the strategical compliance as set by company policy
FR 7.	Represent the technological field evolving over time

In Table 4.5, showing the mapping of CAs to FRs, we can see which functional requirements we need to comply with in order to be able to answer this specific question. This is not always a one to one mapping. To give an answer to how the technological field is structured (CA 1.) a representation of the technological field is needed (FR 1.). But, to answer what the portfolio is of a specific company (CA 2.) a representation is needed of the technological field (FR 1.) as well as the ability to map the patents of a certain company (FR 2.). This shows that to be able to answer a certain question it can be the case that multiple functional requirements must be met. The resulting mapping, shows a (partly) bottom-triangular design, which means that some functionalities overlap in one direction. For example, to comply with CA 3. the tool needs to be able to represent the technological field (FR 1.), have the ability to map patents of a certain company (FR 2.) and to present an overview of who owns the specific patents (FR 3.). This also highlights the existence of difference weights of the FRs, as complying with FR 1. is critical, as without that no CA can be met. In contrast FR 4. to FR 7. are standalone as they are only needed for one specific CA, giving them less weight. Referring to these FRs, the mapping of the FRs to the design parameters (DPs) is presented in Table 4.7, for which the descriptions are provided in Table 4.8.

Table 4.7 – Functional Requirements to Design Parameter mapping

			DP 1.																		DP 4.					DP 5.				DP 6.		DP 7.	
			DP 1.1.						DP 1.2.					DP 1.3.			DP 1.4.				DP 4.1.		DP 4.2.										
			DP 1.1.1.	DP 1.1.2.	DP 1.1.3.	DP 1.1.4.	DP 1.1.5.	DP 1.1.6.	DP 1.2.1.	DP 1.2.2.	DP 1.2.3.	DP 1.2.4.	DP 1.2.5.	DP 1.3.1.	DP 1.3.2.	DP 1.3.3.	DP 1.4.1.	DP 1.4.2.	DP 2.	DP 3.	DP 4.1.1.	DP 4.1.2.	DP 4.2.1.	DP 4.2.2.	DP 4.3.	DP 5.1.	DP 5.2..	DP 5.3.	DP 5.4.	DP 6.1.	DP 6.2.	DP 7.1.	DP 7.2.
FR 1.	FR 1.1.	FR 1.1.1.	x																														
		FR 1.1.2.		x																													
		FR 1.1.3.			x																												
		FR 1.1.4.				x																											
		FR 1.1.5.					x																										
		FR 1.1.6.						x																									
	FR 1.2.	FR 1.2.1.							x																								
		FR 1.2.2.								x																							
		FR 1.2.3.									x																						
		FR 1.2.4.										x																					
		FR 1.2.5.											x																				
	FR 1.3.	FR 1.3.1.												x																			
		FR 1.3.2.													x																		
		FR 1.3.3.														x																	
	FR 1.4.	FR 1.4.1.															x																
		FR 1.4.2.																x															
FR 2.																			x														
FR 3.																				x													
FR 4.	FR 4.1.	FR 4.1.1.																			x												
		FR 4.1.2.																				x											
	FR 4.2.	FR 4.2.1.																					x										
		FR 4.2.2.																						x									
	FR 4.3.																								x								
FR 5.	FR 5.1.																									x							
	FR 5.2.																										x						
	FR 5.3.																											x					
	FR 5.4.																												x				
FR 6.	FR 6.1.																													x			
	FR 6.2.																														x		
FR 7.	FR 7.1.																															x	
	FR 7.2.																																x

Table 4.8 – Functional Requirements and Design Parameter descriptions

FR	Description	DP	Description
FR 1.	Representation of the technological field	DP 1.	Conceptual clustering
FR 1.1.	Attributize individual patents	DP 1.1.	Text-mining
FR 1.1.1.	Ability to specify which patent fields to use	DP 1.1.1.	Select which patent fields to include
FR 1.1.2.	Only include word types that contain information	DP 1.1.2.	Select which word types to include
FR 1.1.3.	Only include English words	DP 1.1.3.	Only attributize English words
FR 1.1.4.	Exclude words that are considered to contain no information	DP 1.1.4.	Select which words to exclude
FR 1.1.5.	Ability to specify the number of attributes	DP 1.1.5.	Select number of attributes
FR 1.1.6.	Keep track of attributized patents	DP 1.1.6.	Store attributes of patents
FR 1.2.	Create hierarchy of patents	DP 1.2.	COBWEB
FR 1.2.1.	Include all entered patents	DP 1.2.1.	Import patents
FR 1.2.2.	Cluster Structuring	DP 1.2.2.	Determine patent position in hierarchy
FR 1.2.3.	Keep track of counts	DP 1.2.3.	Update counts of attributes
FR 1.2.4.	Access clustered structure	DP 1.2.4.	Extract generated structure
FR 1.2.5.	Make structure available for visualization	DP 1.2.5.	Store the node hierarchy and patent locations
FR 1.3.	Show representation	DP 1.3.	Visualization
FR 1.3.1.	Include all patents entered in the hierarchy	DP 1.3.1.	Import the hierarchical structure
FR 1.3.2.	Make the hierarchy structure interpretable	DP 1.3.2.	Visualize the hierarchical structure
FR 1.3.3.	Explain the hierarchy structure	DP 1.3.3.	Visualize structure relevant information
FR 1.4.	Exploration of representation	DP 1.4.	Interactivity
FR 1.4.1.	Assist the user in understanding the high level hierarchies	DP 1.4.1.	Hide children under a given node
FR 1.4.2.	Allow user to learn from the structure	DP 1.4.2.	Explore details of the structure
FR 2.	Map patents of a certain company	DP 2.	Highlight company specific patents
FR 3.	Present an overview of patents ownership	DP 3.	Visualize the individual patent information
FR 4.	Facilitate Freedom To Operate process	DP 4.	FTO analysis
FR 4.1.	Determine potential infringement	DP 4.1.	Similarity analysis
FR 4.1.1.	Determine patent level of similarity	DP 4.1.1.	Determine similarity between siblings
FR 4.1.2.	Insight into patent level of similarity	DP 4.1.2.	Store patent sibling similarity values in nodes
FR 4.2.	Determine patent quality	DP 4.2.	Patent quality analysis
FR 4.2.1.	Insight into patent grounding	DP 4.2.1.	Display patent linking overview
FR 4.2.2.	Insight into patent influence	DP 4.2.2.	Determine patent citations
FR 4.3.	Provide insight into Inter-company relationship	DP 4.3.	Disp. Pat. owner inter-company relation / threat level
FR 5.	Pat. that techn. provide solutions for comparable problems	DP 5.	TRIZ
FR 5.1.	Extract the actions represented within patents	DP 5.1.	Determine SAO structures within patents
FR 5.2.	Make a comparison possible	DP 5.2.	Enter the functionality searched for
FR 5.3.	Determine how similar the pat. actions are compared to the solution searched for	DP 5.3.	Determine SAO functionality similarity
FR 5.4.	Represent the solutions within the technological field	DP 5.4.	Store pat. SAO similarity as compared to the search
FR 6.	Represent the strategical compliance as set by company policy	DP 6.	Strategy / Portfolio analysis
FR 6.1.	Insight into company strategy	DP 6.1.	Highlight company strategy specific technology areas
FR 6.2.	Insight into company strategized technological areas	DP 6.2.	Determine strategy positioning
FR 7.	Represent the technological field evolving over time	DP 7.	Trend analysis
FR 7.1.	Track the changes of the technological field over time	DP 7.1.	Store changes over iterations
FR 7.2.	Show the evolving of the technological field	DP 7.2.	Visualize evolution over time

The resulting matrix as presented in Table 4.7 shows a diagonal design in which the functional independency is guaranteed. The multiple zig-zagging refinements of the design can be recognized in the multi-layered aspect of the design. For example to be able to perform a trend analysis (DP 7.) in order to fulfil the need for a representation of the technological field evolving over time (FR 7.). Here, a refinement in DPs is needed resulting in the conclusion that the functional requirements needed to be further specified. This results in the addition of the need for a storage of how the patent field evolves over time under influence of entering patents (DP 7.1.), and the need for a visualization that can represent the evolution of the patent field over time (DP 7.2.).

Table 4.9 – Design Parameters to Process Variables mapping

	PV 1.	PV 2.	PV 3.	PV 4.	PV 5.	PV 6.	PV 7.	PV 8.	PV 9.	PV 10.	PV 11.	PV 12.	PV 13.	PV 14.	PV 15.	PV 16.	PV 17.	PV 18.	PV 19.	PV 20.	PV 21.	PV 22.	PV 23.	PV 24.	PV 25.	PV 26.	PV 27.	PV 28.	PV 29.	PV 30.	PV 31.		
DP 1.1.1.	X																															A1	A
DP 1.1.2.		X																															
DP 1.1.3.			X																														
DP 1.1.4.				X																													
DP 1.1.5.					X																												
DP 1.1.6.						X																										A3	
DP 5.1.							X																									A2	
DP 5.2.								X																								B2	B
DP 5.3.									X																							B3	
DP 5.4.										X																							
DP 1.2.1.											X																					B1	
DP 1.2.2.												X																				B4	
DP 1.2.3.													X																				
DP 1.2.4.														X																		B7	
DP 1.2.5.															X																		
DP 4.1.1.																X																B5	
DP 4.1.2.																	X																
DP 7.1																		X														B6	
DP 7.2.																			X													C3	C
DP 1.3.1.																				X												C1	
DP 1.3.2.																					X												
DP 1.3.3.																						X										C2	
DP 1.4.1.																							X										D
DP 1.4.2.																								X									
DP 2.																									X								E
DP 3.																										X							
DP 6.1.																											X						
DP 6.2.																												X					F
DP 4.2.1.																													X				G
DP 4.2.2.																														X			
DP 4.3.																															X		H

Table 4.9 is the resulting mapping of the design parameters onto the process variables after multiple zig-zagging iterations, for which the descriptions are shown in Table 4.10. It is ordered to facilitate the grouping of certain process variables for the creation of application implementable blocks (this explains the incoherent ordering of the design parameters). The individual implementable blocks and sub-blocks that are detected are listed in Table 4.11. These are grouped together based on the specific part of the software in which they are implemented. The presented finalized design will partly be implemented in a prototype decision support system as well as be used for the evaluation of this prototype. Finally, it is important to mention that the combined implemented elements will together determine the usefulness of the decision support system.

Table 4.10 – Design Parameter and Process Variables descriptions

DP	Description	PV	Description
DP 1.1.1.	Select which patent fields to include	PV 1.	List of fieldnames to include
DP 1.1.2.	Select which word types to include	PV 2.	List of word types
DP 1.1.3.	Only attributize English words	PV 3.	Exclude non-English words based on existence in dictionary
DP 1.1.4.	Select which words to exclude	PV 4.	Exclude words that occur on Stop word list
DP 1.1.5.	Select number of attributes	PV 5.	input variable of number of attributes
DP 1.1.6.	Store attributes of patents	PV 6.	JSON file containing patents and their attribute counts
DP 5.1.	Determine SAO structures within patents	PV 7.	NLTK SAO analyzer
DP 5.2.	Enter the functionality searched for	PV 8.	input fields for the subject, action and object
DP 5.3.	Determine SAO functionality similarity	PV 9.	Vector based pairwise similarity calculation
DP 5.4.	Store the pat. SAO similarity as compared to the search	PV 10.	Write to similarity measure field of patent
DP 1.2.1.	Import patents	PV 11.	JSON read attributized patents
DP 1.2.2.	Determine patent position in hierarchy	PV 12.	cobweb
DP 1.2.3.	Update counts of attributes	PV 13.	add_attribute_counts
DP 1.2.4.	Extract structure	PV 14.	loop_items
DP 1.2.5.	Store the node hierarchy and patent locations	PV 15.	JSON file containing node hierarchy and patent locations
DP 4.1.1.	Determine similarity between siblings	PV 16.	Calculate Sibling attribute similarity value
DP 4.1.2.	Store patent sibling similarity values in nodes	PV 17.	Store in JSON
DP 7.1.	Store changes over iterations	PV 18.	XML Node change log
DP 7.2.	Visualize evolution over time	PV 19.	JavaScript time slider
DP 1.3.1.	Import the hierarchical structure	PV 20.	JavaScript JSON reader
DP 1.3.2.	Visualize the hierarchical structure	PV 21.	D3.js tree object
DP 1.3.3.	Visualize structure relevant information	PV 22.	Relative important cluster attributes
DP 1.4.1.	Hide children under a given node	PV 23.	JavaScript Onclick close node's children
DP 1.4.2.	Explore structure	PV 24.	JavaScript zoom and drag functionality
DP 2.	Highlight company specific patents	PV 25.	JavaScript Onselect highlight nodes of selected company
DP 3.	Visualize the individual patent information	PV 26.	JavaScript print individual patent information at node
DP 6.1.	Highlight company strategy specific technology areas	PV 27.	JavaScript Onselect highlight nodes of selected strategy
DP 6.2.	Determine strategy positioning	PV 28.	Strategy node selection
DP 4.2.1.	Display patent linking overview	PV 29.	Patent link information from XML interface
DP 4.2.2.	Determine patent citations	PV 30.	Patent citations from XML interface
DP 4.3.	Display pat. owner inter-company relation / threat level	PV 31.	PHP/MySQL connection to database containing threat levels

Table 4.11 – Design Parameter group block Descriptions

Group	Description	Group	Description
A	Text-mining	B-7	Extracting & Exporting
A-1	Setup & attributizing	C	Visualization
A-2	SAO structures	C-1	Import & Structure
A-3	Exporting	C-2	Node/Structural explanation
B	Clustering	C-3	Evolution
B-1	Importing	D	Exploration
B-2	Problem input	E	Analysis
B-3	SAO pairwise comparison	F	Strategy positioning
B-4	Conceptual clustering	G	Quality analysis connection
B-5	Sibling similarity	H	Company database connection
B-6	Change log		

This chapter has shown that based on the user's perspective e.g. legal, technical or managerial as well as on the phase within the STP process, there are different needs when it comes to patent analysis. For which there are currently different implementations present, which present a number of flaws. Conceptual clustering is proposed as a potential solution. For which, based on the STP process a design for a decision support system implementation has been developed using the Axiomatic Design method. The design is realized based on customer and functional needs derived from literature studies and expert input. And will later be used to refer to when evaluating the effectiveness in enhancing the STP process, as presented in Chapter 7.

— Equipped with his five senses, man explores the universe around him and calls the adventure Science.
Plato

5

Conceptual Clustering

One of the methods applied in the representation of patents to gain insight into the, for the user relevant, underlying structures is that of clustering. Clustering can be described as the process of intelligently grouping items into categories that are meaningful to its user (Michalski, 1980). This should be seen as partitioning items into groups as such that the similarity between items intra-groups is high and that between items inter-groups is low. From this principle can be deduced that the type of measure of similarity will thus greatly define the final outcome of the clustering. Many types of similarity measures have been developed over time and the choice of use should depend on the type of measurement and/or the object's representation (Cha, 2007)[5]. When applied to patents, clustering techniques are interesting for managers as they provide an overview of the patent space which can be used in exploring clustered entities to detect similar technologies. This chapter will further focus on one specific family of clustering techniques that is conceptual clustering. This will be done by firstly introducing conceptual clustering and in detail one of its methods, iterative conceptual clustering, in Section 5.1. Secondly the explanation of its measure of similarity is provided in section 5.2. Third, an overview of the operators used in the iterative conceptual clustering are presented. Finally, in Section 5.4 the method is evaluated for use in patent analysis and positioned in the current patent analysis field.

5.1. Conceptual Clustering and the COBWEB algorithm

Conceptual clustering is first defined by Michalski (Michalski, 1980) as a machine learning method. In which machine learning refers to a method that does not require the user's input on the pre-definition of classes. Rather, it is based on an evaluation function that creates classes which fit the entities best, a method which is described as 'learning by observation' (Fisher, 1987). The output of a conceptual clustering algorithm is the structure of a classification tree with its entered entities embedded. The tree structure is formed, based on the evaluation of class qualities. Based on the type of conceptual clustering technique, different approaches to form these trees are applied. There are top-down approaches (splitting nodes), bottom-up approaches (fusing nodes), and combinations of the two applied in searches for the best fitting categorization. The ones that apply a combination have the advantage that they can recover a previous decision in light of newly entered information.

One of the developed methods for conceptual clustering is that of incremental conceptual clustering. Building on the idea of Michalski (Michalski, 1980) and inspired by the systems UNIMEM (Lebowitz, 1987) and CYRUS (Kolodner, 1983), in 1987 Fisher developed a conceptual clustering algorithm named COBWEB (Fisher, 1987). The COBWEB algorithm exploits the use of a bi-directional search possibility to recover previously made decisions as well as allowing for a hill-climbing search which results in an optimization of computational requirements. A fundamental difference when compared to its predecessor, is that this conceptual clustering

[5] An extensive comparative overview of similarity and distance measures is presented in (Cha, 2007)

technique uses an iterative approach. This has multiple advantages, namely that before running the algorithm the complete input does not need to be present or even indexed. Secondly, the bi-directional input allows for different orders of iteration as it compensates when superior categorizations than those present are found. In the COBWEB algorithm concepts are represented as a summary of probabilities of the attributes of the entities within that concept having a certain value. Each concept node is therefore a summary of the objects it contains in its lower hierarchy. Before the COBWEB algorithm can be examined, an understanding of its similarity measure that is the categorical utility is needed, which is therefore presented in Section 5.2.

5.2. CATEGORICAL UTILITY AS A SIMILARITY MEASURE

The similarity measure which determines the choices made within the iterative conceptual clustering algorithm that is COBWEB is the categorical utility. The categorical utility as developed by Corter and Gluck (Corter & Gluck, 1985), is defined by their assumption that the usefulness of concepts is related to the ability of the concept's categories being used in communicating information of the instances' properties. Within the COBWEB algorithm, the categorical utility function is used to determine to which concept a new entry should be added. This is possible as it is a function that suits the needs for the clustering of data as it combines the rewarding of both intra-class similarity ad well as inter-class dissimilarity. Equation 5.3 shows how the categorical utility of adding an entity under a certain concept can be calculated[6].

$$\sum_i \sum_j \mathrm{P}(A_i = V_{ij} | C_k)^2 \qquad \text{Equation 5.1}$$

$$\sum_i \sum_j \mathrm{P}(A_i = V_{ij})^2 \qquad \text{Equation 5.2}$$

$$\mathrm{CU} = \frac{\sum_{k=1}^{n} P(C_k) \left[\sum_i \sum_j \mathrm{P}(A_i = V_{ij} | C_k)^2 - \sum_i \sum_j \mathrm{P}(A_i = V_{ij})^2\right]}{n} \qquad \text{Equation 5.3}$$

This equation is used to calculate the categorical utility *CU* for each of the individual concepts *k* under the root concept. To do so, the newly entered item is virtually placed into these concepts when *CU* is calculated. $P(C_k)$ is equal to the count of items under concept *k* (including the virtually added one) divided by the total number of items under the root concept. Equation 5.1 is the probability of the specific value V_{ij} occurring for attribute A_i in concept C_k, representing the expected number of attribute values that can be correctly guessed for the specific concept. In which the summation is performed over all attributes *i* and all their possible values *j*. Equation 5.2 represents the probability of the specific value for the attribute under the root concept, which is used to account for the effect of those that are correctly guessed with no knowledge. The latter also results in the rewarding of the inter-concept differences. This results in *CU* being equal to sum of the probability of an item being found under a concept multiplied by the difference in correct by concept prediction and correct by chance, which in total is divided by the number of concepts *n* under the root concept.

$$\mathrm{A}_i = \mathrm{V}_{ij} = \alpha + (1 - \alpha)^2 \qquad \text{Equation 5.4}$$

When using Boolean variables (meaning it either does or does not have a certain attribute) this will result in a relatively large influence of the variable when the probability of it occurring in the given concept is 100% and when it is 0%. This is due to the fact that using Equation 5.3 will per attribute result in the influence shown in Equation 5.4. This results in a parabolic shape of the influence of the attribute α as can be seen in Figure 5.1. This should be read as those attributes that are either present or not present in a large share of the concept population having a great explanatory value of the items within the concept. When the probability of an attribute occurring is 0.5 the influence is lowest, this is as expected as it has no explanatory value on the items that the content contains.

[6] The complete derivation of Equation Equation 5.3 can be found in (Fisher, 1987).

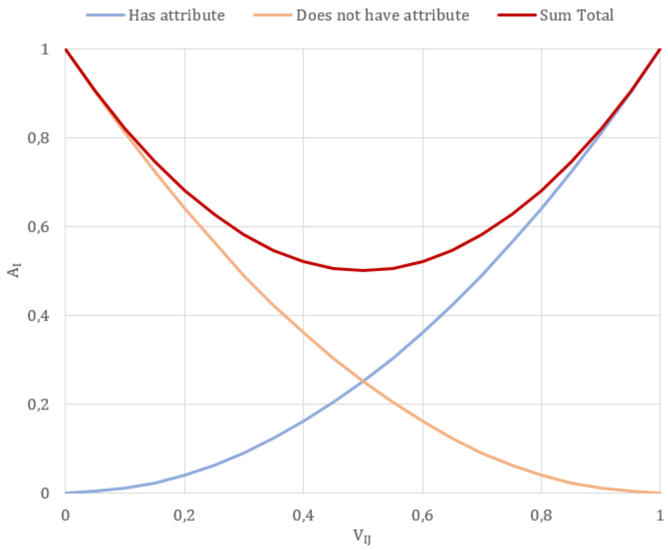


Figure 5.1 – Chart of A and V

When all the entered items have the same value for a certain variable, both the expected predicted probability as well as the probability of a correct guess without knowledge will have the same value. This means that when considering Equation 5.3, this will create the situation in which the two values are canceling each other out, resulting in a value of 0. This means that, as can be expected, this variable has no explanatory value and thus no influence at all in the forming of clusters. Concluding it can be seen that for an attribute to have a relatively high explanatory value for the concept at hand it needs to either have a relatively high probability to occur in the concept and a relatively low probability to occur globally or exactly the opposite. In that way it really says something about what the entries under a concept have in common and sets them apart from the other concepts.

5.3. COBWEB OPERATORS: THE AUTOMATION OF CATEGORIZING

Based on the calculation of the categorical utility, the COBWEB algorithm applies one out of four operations. These possible operations are adding a concept under an existing concept, creating a new concept on the level under examination, splitting a concept by removing it and promoting its children, and finally merging two concepts into and one combining its children. These four options will be explained by the use of example calculations and visualizations which will make use of the data in Table 5.1 in which a value of 1 represents that the given entry possesses this attribute and a 0 that it doesn't. The number of concepts that will be created by the algorithm are not bounded by a system parameter, but rather they emerge based on the operators and the current environment.

Table 5.1 – Example Entries

Entry	A_1	A_2	A_3	A_4	A_5
E_a	1	1	0	1	0
E_b	1	0	1	0	0
E_c	1	1	0	0	0
E_d	1	0	1	1	1
E_e	1	1	0	1	1

5.3.1. ADDING AN ENTRY UNDER AN EXISTING CONCEPT

One of the operators in the COBWEB algorithm is the one of adding an entry by finding the most suitable host concept. This is done by virtually placing it in the individual concepts, calculating each resulting categorical utility, and placing it under the concept which results in the highest possible value. This operator will be elaborated by an example based on Table 5.1.

When considering the situation as presented in Figure 5.2 in which E_a and E_b are present, and E_c will be added. In this situation the categorical utility (CU) of adding the new entry to the individual branches will be calculated

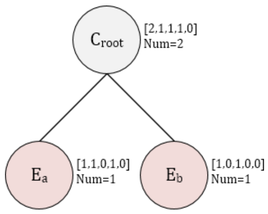


Figure 5.2 – 2 Entries

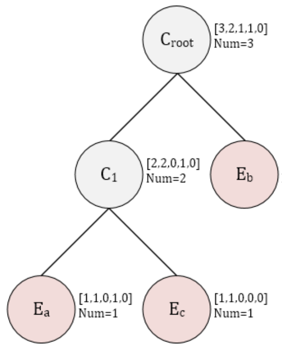


Figure 5.3 – 3 entries

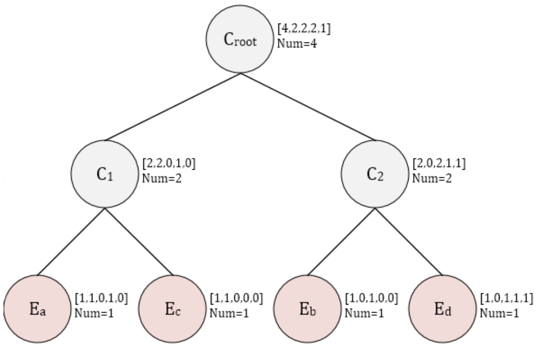


Figure 5.4 – 4 entries

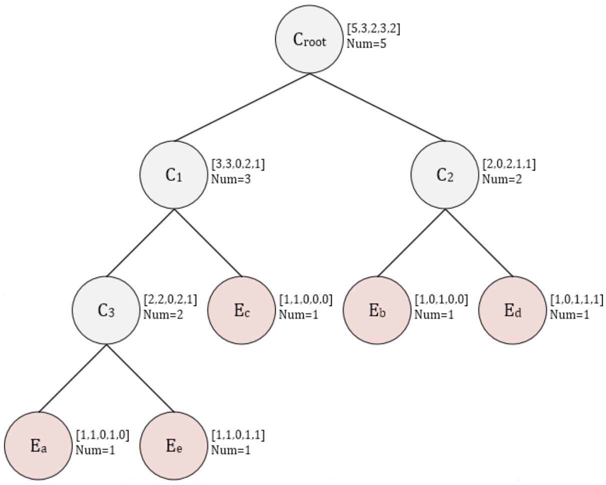


Figure 5.5 – 5 entries

using Equation 5.3. Which results in the CU for E_c to be placed together with E_a is equal to 0.498 and to be placed together with E_b is equal to 0.332. The CU for adding E_c to E_a being the highest and thus the new situation is that which can be seen in Figure 5.3. In this new situation we can see that E_a and E_c are added together under concept C_1. Each concept contains the probability of an attribute being present in the entries it has underneath, which are calculated based on the counts of occurrences and the number of entries it contains. In the case of C_1 these probabilities are $P(A_i \mid C_1)=[1,1,0,0.5,0]$. Adding E_d the same procedure is followed and the CU for both the path towards C_1 as well as towards E_b are calculated given the new situation. Being respectively 0.27 and 0.56, E_d is placed next to E_b in the concept created C_2, of which the resulting situation can be seen in Figure 5.4. Finally the addition of entry E_e is considered. The CU for adding E_e under C_1 equals 0.49 where that of adding it under C_2 equals 0.33, thus the path towards C_1 is followed. From concept C_1 a second set of calculations is performed to determine the best suiting concept. To place E_e together with E_a results in a CU of 0.28 where placing it together with E_c results in a CU of 0.12, thus the final position of E_e is next to E_a under C_3, which can be seen in Figure 5.5. In the adding of E_e the applied strategy in this algorithm that is hill-climbing can be detected.

A hill-climbing strategy is a strategy in which an optimum position for an object is searched for by constantly choosing the best option out of those available at that moment. In the example of adding E_e the first decision was that at the root level of concept C_{root} in which C_1 was the best available option. The second decision was at C_1 at which choosing the option of E_a was the best available option. The main advantage of this way of positioning the new entry is that it is computationally efficient as with each step it eliminates more possibilities which do not have to be considered and therefore not calculated.

Furthermore, considering the situation in Figure 5.5 having the concept probability values, an analysis can be done on the final situation of concept clustering. The first thing that can be said is that A_1 has no influence in the defining of concepts as it is possessed by all the entries and is therefore irrelevant. Second, for C_2 can be seen that A_2 is relevant as this attribute is not present in all entries together with A_3 that is present in all Entries. At the same level, for C_1this is exactly opposite. We can conclude that a first decision here is based on the attributes A_2 and A_3. Following C_1, it can be seen that a sub-cluster had formed in which A_4 is always present where for E_c this attribute is not present. A_5 always seems to be non-influential when considering the clusters formed given these specific entries.

5.3.2. ADDING AN NEW CONCEPT

The second operator in the COBWEB algorithm is the one of adding an entry by the creation of a new concept. This is done by virtually placing it next to the currently present concepts and calculating the resulting CU. The situation for the calculation of E_c being entered and tested for adding a new concept can be seen in Figure 5.6. The CU for this situation calculated by using Equation 5.3 is equal to 0.443, which is in this situation lower than the highest value of adding it to an existing one which was 0.498 with E_a.

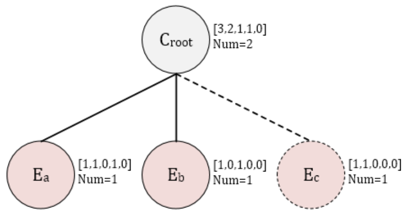


Figure 5.6 – Example of the testing of adding a new concept

5.3.3. SPLITTING A CONCEPT

The third possible operation is that of splitting a concept. Together with the merging two concepts, which will be explained in Section 5.3.4, this operation is present to reduce the effect of the iteration order. They allow the algorithm to correct previous decisions based on the newly available information embedded in the new Entry. A split is considered when a concept is present as a child under the concept location from which the search is performed. The concept selected for splitting is that one that has the highest CU in the determination of the best fitting host concept for this entry, as explained in Section 5.3.1. Figure 5.7 shows the situation in which E_e is added. As concept C_1 is the best suiting concept, this one is considered for splitting. In this situation its children are promoted and these are examined concerning how well they would host E_e. The best hosting concept would be that one combined with E_a which results in a CU of 0.44, however adding it to the existing concept C_1 results in a CU of 0.493 which is thus preferred.

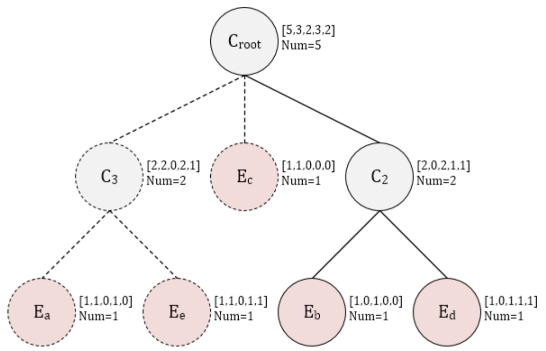


Figure 5.7 – Example of the testing of splitting a concept and adding to its best host

5.3.4. MERGING TWO CONCEPTS

The last operator in the COBWEB algorithm is that of merging 2 existing concepts. This is can be tried when the concept location from which to search has two or more children. If so, the two best hosts as determined by the CU calculations of adding the entry to an existing concept, as explained in Section 5.3.1, are selected to test for a merger. An example of the examination of a merger can be seen in Figure 5.8 in which E_e is added. From the root concept, two children can be seen which are thus considered for a merger in which the children of these two are combined under a new concept together with the new entry. In the given situation the resulting CU is equal to 0, where adding under existing concept C_1 results in 0.493 and is preferred. Relevant to mention is that in the case of only two concepts present as children, a merger will only be a preferred solution when all attributes have the same value as otherwise the resulting CU will always equal 0.

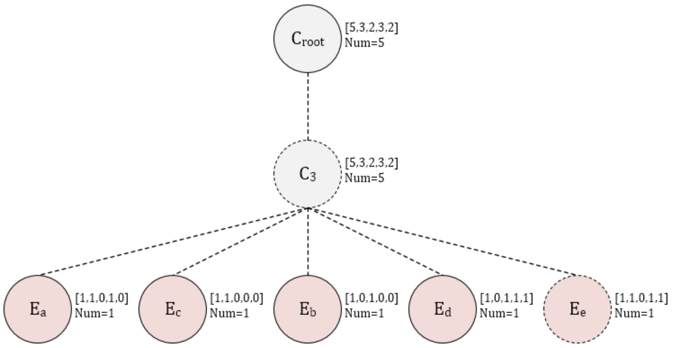


Figure 5.8 – Example of the testing of merging two concepts and adding to the created concept

5.4. CONCEPTUAL CLUSTERING IN THE PATENT ANALYSIS FIELD

The analysis in Chapter 4 has shown that there is not a patent analysis tool available that classifies the patents in a hierarchical manner. It is therefore of interest to see whether conceptual clustering, which' output results in a hierarchical inter-relational structure, can be used in enhancing the insight derived from the patent analysis process. This seems interesting because, as shown in Chapter 3, hierarchical classifications are often used in the structuring of knowledge in the attempt to raise the level of understanding of the data at hand. In this research a prototype will be created to determine what form of technological taxonomy will be represented in the hierarchy created by the COBWEB algorithm. Which could in follow-up research be used to see whether the classification of the IPC system correlates with the resulting classification, for example to make recommendations about determination of these codes to new patent applications. Next to the hierarchical representation, the auditability of this technique is high, as it is a mutually exclusive technique, so one knows how the tree is constructed and one can also look back and see what it would take to remove a certain part, which will add to the reliability of the analysis performed.

The categorization of dimensional (spatial) models of data as presented by Porter and Cunningham, entails three main types of dimensional models. Which are: (1) mapping techniques, (2) Clustering techniques, and (3) Tree techniques (Porter & Cunningham, 2004). These are all described as to be deductive approaches, in which the analysis is started with an already-accepted data structure, meaning that the data is fit into a model with defined operation of structural creation decisions. This is done as it is assumed that the data is of a certain structure, which the selected technique is expected to be able to visually represent. The mapping techniques aim by the use of the simplification of focus to present patterns in the data. They can be recognized by their use of space and location as a method of communicating the findings. The selection of the metrics defining the distance and/or similarity will impact the model outputs in this technique. The clustering techniques are described to be focusing on the discovering of so-called natural clusters. Their representation is easily interpretable as the notion of groups feels natural to users. The downside of using this techniques can be that more complex inter-relations cannot be captured in such a representation. The models based on tree techniques present the data in a top-down manner. Meaning that the top level nodes present aggregates of the data, and the deeper down the tree, the more specific the description of the data gets. This is done with the goal to find the inter-relations of the data, in which the main difference with the clustering techniques is that the key differences between the branches are determinable. Considering the conceptual clustering technique; given its resulting hierarchical structure and it focusing on both similarities as well as differences. The conceptual clustering technique should, despite what its name might suggest, be positioned as a tree type dimensional modeling technique, in the categorization as defined by Porter and Cunningham.

When looking at the taxonomy of patent analysis as presented by Abbas, Zhang and Khan (Abbas et al., 2014) which is represented in Figure 5.9. Given the fact that the conceptual clustering technique itself makes use of attributized data, and its output will generate a structure which can be visualized into a tree like structure, which can be interpreted by an expert. Taking into account the previously introduced categorization of Porter and Cunningham (Porter & Cunningham, 2004), it can be seen that under patent analysis visualization techniques there is currently no such categorization of tree type techniques. Which highlights again the novelty of the application of conceptual clustering applied to patents. Conceptual clustering should therefore be positioned in a new category under the visualization techniques, when applying the taxonomy of patent analysis as defined by Abbas, Zhang and Khan.

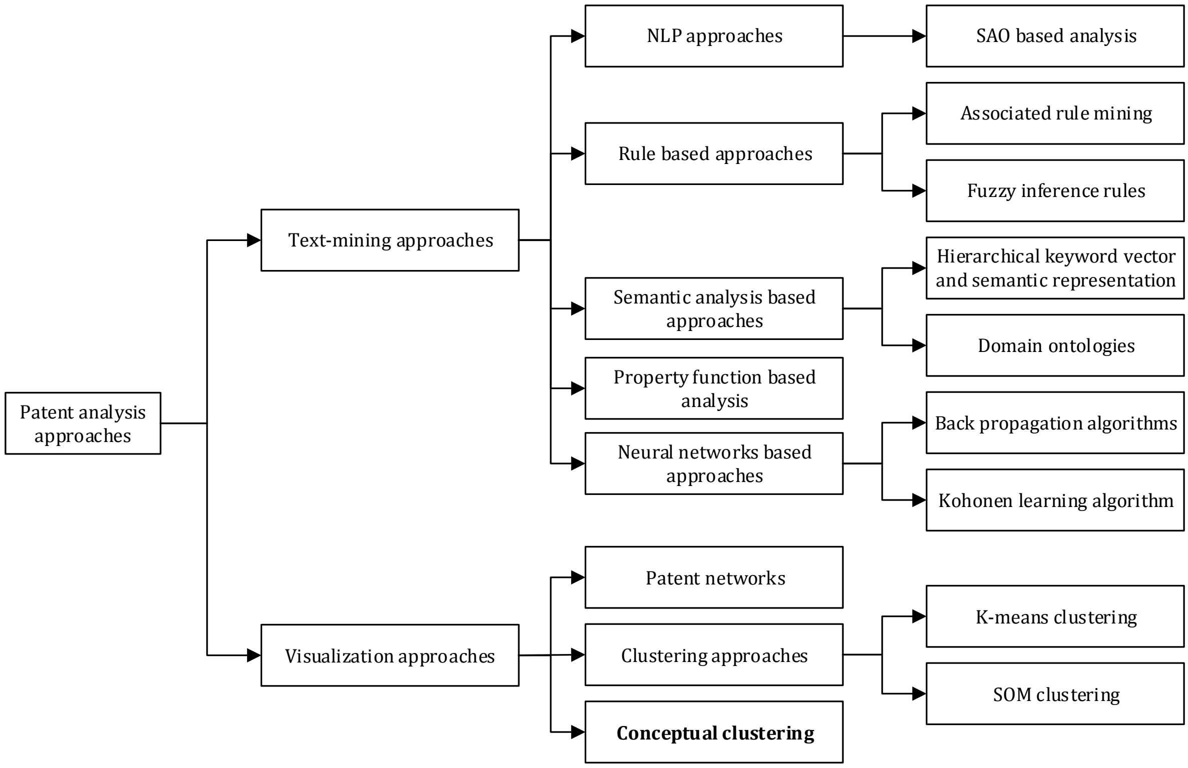


Figure 5.9 – Conceptual Clustering positioned in the Taxonomy of techniques for patent analysis (Abbas et al., 2014)

To be able to assess the usefulness of the conceptual clustering concept applied to the field of patent analysis for the decision making process in Strategic Technology Planning, a prototype applying the COBWEB algorithm to patent documents has been developed which will be introduced in Chapter 6.

—Do, or do not, there is no try.
Master Yoda

6

Implementation: Conceptual Clustering applied to patents

This chapter describes the implementation of a conceptual clustering applied to patents based decision support system. The initial implementation making use of the COBWEB algorithm is done as a three-step process which is visualized in Figure 6.1. Showing the extraction from the database, the parsing and indexing in the text-mining software, storing of the indexed patent entries, the conceptual clustering software, storing the hierarchical relations and finally the visualization. This chapter will highlight the key elements of this implementation, for an overview of the complete conceptual clustering software documentation please refer to Appendix V. The implementation of the Text-Mining software as well as that of the COBWEB algorithm based software is done, as requested, in Python. Python is a high-level general-purpose programming language which is used in all sorts of applications due to, among other things, its readable character and compact syntax (Python Software Foundation, 2015). The python codes are created, making use of the PyCharm 4.0.4 integrated development environment. PyCharm is a smart code editor which offers many advantages to coding in Python e.g. coding assistance, documentation, and debugging and testing (JetBrains s.r.o., 2015).

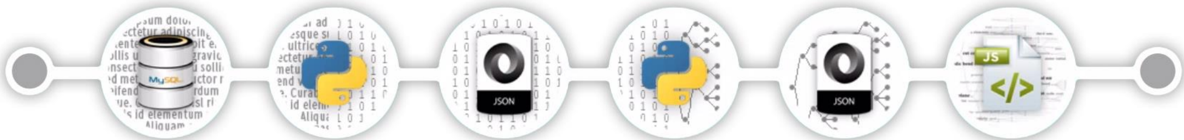


Figure 6.1 – Implementation Architecture

The implementation is based on three steps which are respectively Text-mining, The COBWEB conceptual clustering algorithm and visualization. Figure 6.1 shows the order of the implementation, starting with the input of patents and ending with the visualization. The data file that is the input of the Python COBWEB algorithm as well as its output data file, which is used as a bridge between the Python implementation and the visualization in JavaScript, are JSON (JavaScript Object Notation) files. JSON is a data-interchange format which can be easily parsed and generated by virtually all modern programming languages (ECMA International, 2013). This also gives the advantage of the output data being able for usage in combination with other software. The implementation structure will be explained by following the three main steps; text-mining in Section 6.1, conceptual clustering in Section 6.2, and visualization in Section 6.3. This Chapter is finalized by a structural validation of the conceptual clustering algorithm in Section 6.4.

6.1. STEP 1: TEXT-MINING

Before the patents can be parsed by the conceptual clustering algorithm, they need to be attributized. This means that a selection needs to be made off attributes which will be used for the conceptual clustering and adding to the individual patents whether or not they possess this attribute. For the initial implementation a Boolean expression is used (1 if the patent contains this attribute and 0 otherwise). This design choice does result in potential information represented in the number of occurrences. Future research should be done into the effects of including the number of occurrences to determine whether this influence is significant to the resulting output of the conceptual clustering technique. The attributizing is done by applying a text-mining program written in Python, Algorithm 6.1 shows the pseudocode of said program.

Algorithm 6.1 – Pseudocode for the Text-mining of patents

```
Define Patent fields to include
Define allowed word-types to use in attributizing
Define number of attributes
for all Patents do
  Select field to analyze
  Define word-types of words in field
  for all Words do
    if word-type is one of the allowed word-types then
      if word is and English word then
        if word is not a stop-word then
          Add word count to global list
        end if
      end if
    end if
  end for
end for
Sort the global word count list descending and select the defined number of attributes
for all Patents do
  for all At tributes do
    if Patent has Attribute then
      Set attribute to 1
    else
      Set attribute to 0
    end if
  end for
end for
Write the attributized patents to JSON file
```

Based on the preference/needs of the user the patent title and/or abstract can be selected for analysis. The choice to only include these bibliographical fields has been made to limit the needed parsing time as they are perceived to contain the key elements of the innovation as captured by the specific patent. From this fields, the sentences are semantically analyzed by using the nltk (Natural Language Toolkit) library. This library allows, among other things, for tokenizing the sentences into individual words and to determine what word-type (noun, verb, etc.) these individual words are (NLTK Project, 2015). Based on a selection by the user of what word-types to include, the words that are not of one of these types are excluded from being considered as an attribute. This user selection is added to use those words as attributes that can possess explanatory value, thus excluding non-explanatory word-types as e.g. numbers. Following the exclusion of irrelevant word-types is a check whether the word is existing in the English dictionary, which is done by using the Wordnet database (Princeton, 2014) of the nltk library. This step is added for those cases where for example in the abstract the same explanation in multiple languages is present. Finally as a last check, the word is compared with a list of common stop words that are considered to have no explanatory value, to exclude any stopwords that haven't been picked out by the previous steps. The resulting words are added to the global count list. When all patents are parsed, the global count list is ordered from the most frequent occurring word to the lowest occurring word. Based on the user's input for the number of attributes to include, the corresponding number of attributes are selected as the feature set. The final step is to again parse the individual patents, now comparing them with the feature set to see if the given patent contains the attribute, and consequentially attributize the patent (1 if the patent contains this attribute and 0 otherwise). A verification of operation is added in Appendix III.

When the patents are attributized, they are written into a JSON file. Together with the patents, meta data of the input-data, settings defined by the user and the date of operation are added to this JSON file to ensure traceability of the generated data. The input data is now made ready to be parsed through the conceptual clustering algorithm.

6.2. STEP 2: THE COBWEB ALGORITHM

As the interest lies with allowing for additional interactive visualizations, expectantly enhancing the amount of insight that the same set of patents can provide, the individual patents information needs to be maintained. However, the COBWEB algorithm nodes can individually contain the information from 1- ∞ patents. Hence there is the need for a separate object class which will contain the given patent related information. A visualization of the relation between the nodes and patents is shown in Figure 6.2. As the patents will be classified into the best fitting concept, the patents will always be linked to concepts that are at the end of the given tree branch (also called leafs). In this figure the concept nodes are represented by those containing 'C' and the patents by those containing 'P', although the patents are linked to a certain node they don't take part in the conceptual clustering and can therefore be seen as virtually linked.

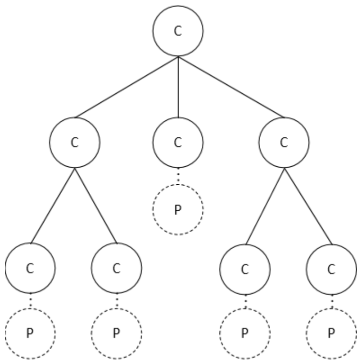


Figure 6.2 – Concept and Patent object relations

To be able to store the needed information of the patents as well as to create the hierarchical and relational structure, two types of object classes are created which are the Patent class and the Node class which can be seen in respectively Table 6.1 and in Table 6.2. The most important element for the Patent object is the Node ID, which links the given patent to a specific node. Where for the Node object, the Root and Children are the most important as these determine the complete structure of the concept tree.

Table 6.1 – Python implementation Node object class

Class Node	
Class attribute	**Explanation**
Unique ID	Assigned unique ID
Root	Parent Node object
Children	List [Node object,. . .]
Attributes	Dictionary: {[key, value],. . . }
Patents	List [Patent object,. . .]
Patents number	No. patents under node

Table 6.2 – Python implementation Patent object class

Class Patent	
Class attribute	**Explanation**
Unique ID	Assigned unique ID
Input ID	ID from input file
Input title	Patent title
Input patent orig. id	Patent code
Input company id	Company
Node ID	Concept Node object
Attributes	Dictionary: {[key, value],..}

The file that contains the COBWEB algorithm has four definable steps which are respectively Importing, Creating patent object, Creating node object and adding to the tree and Exporting. At the importing of the data from the JSON file step, for each entry being inputted a patent object is created, containing the patent related information. By creating a patent object, in turn a node object which is linked to the patent will be created. By creating a node object which has the patent information linked, the node object will run the COBWEB function, as shown in Algorithm 6.2. This function is ran until it finds its best suitable location within the hierarchy. When all the inputted entries are parsed and the hierarchy is created, the hierarchy needs to be put into the output format that is the JSON file. This is done in a nested manner, starting from the tree root node and continuously asking its children to include their information. An example of output structure is presented in Appendix IV.

Algorithm 6.2 – Pseudocode of the COBWEB core

```
function COBWEB(root )
  if root has children then
    Define Categorical Utilities of adding to children
    Define Categorical Utility of new class under root
    Define Categorical Utility of merging best 2 hosts
    Define Categorical Utility of splitting best host
    if CU of new class is highest then
      Create new concept under root and add patent
    else if CU of merging is highest then
      Merge 2 bests hosts
      COBWEB(Merged node)
    else if CU of splitting is highest then
      Split best host
      COBWEB(Current root)
    else
      COBWEB(Child with highest CU)
    end if
  else
    Add object as node under root concept
  end if
end function
```

To allow for data-layers being added to the visualization, in the JSON output file, information such as the patent title, filing company as well as the most relevant attributes are added to this file. This will allow the user to interactively explore different aspects of the created patent hierarchy to find that insight or those relations that are relevant. Now that the output file is finalized the visualization can be created.

6.3. STEP 3: VISUALIZATION

The visualization of the hierarchy created of the input patents is done in a web browser environment by making use of the D3 JavaScript library. The D3 library enables easy data visualization and live manipulation of the visualization (Bostock, 2015). By making use of a JavaScript based library every device with an internet browser installed could potentially access the visualization. This adds to the accessibility of the output data and allowing all users to easily explore this data. The created visualization exploits the use of the tree element in the D3 library which is based on the Reingold-Tilford algorithm for tidier drawings of trees (Reingold & Tilford, 1981). The page has been adapted to allow for the input of the JSON hierarchy structure and the additional functionalities for exploring the hierarchy space. The resulting visualization of the situation as presented in Figure 5.5 is shown in Figure 6.3. In this figure the patent nodes can be detected as they show parts of the patent relevant information, in this case the 'input patent original ID' field of the patent object (for the example filled with integers but normally with patent codes e.g. US8906979B2) and the 'input title' field are shown. An example of a slightly larger hierarchy tree is shown in Figure 6.4 which shows the hierarchy of 150 patents generated based on the 750 most frequent attributes.

To aid the user in interpreting the generated map, the choice has been made to display the most explanatory variables for a given concept cluster next to the concept node. An explanation will be given by using an example. Consider the situation presented in Figure 5.5, given that the attributes represented by [A_1, A_2, A_3, A_4, A_5] are [Light, LED, Halogen, Infrared, Car]. Given the information they contain the visualization should ideally present the output as shown in Figure 6.5.

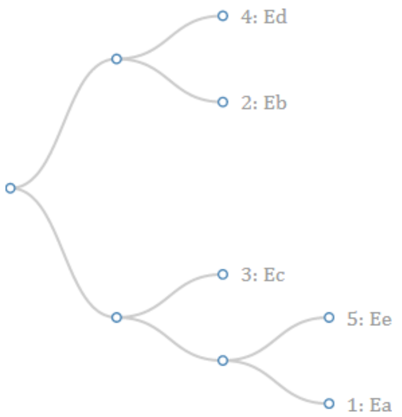


Figure 6.3 – Visualization in browser of Example

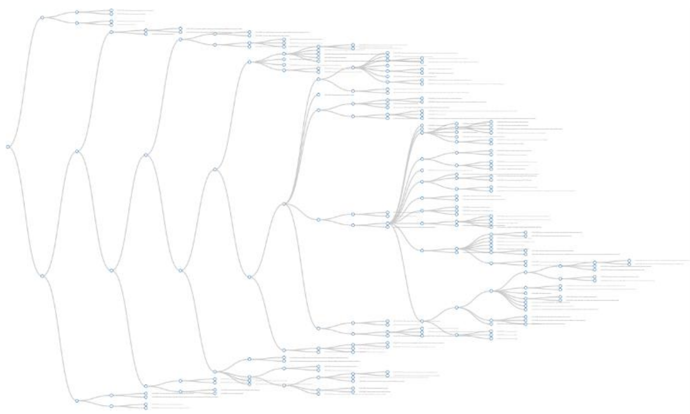


Figure 6.4 – Visualization in browser of 150 Patent entries

The entries under C_3 are grouped together as contrary to E_c they do contain the infrared attribute, in this case given the inheritance aspect, these elements ended up under C_3 as they contain both the LED as well as Infrared attribute. In displaying however, only C_1 should display LED as those under that concept have a relative high probability of containing that attribute (in this case 100%). As C_3 is a sub-concept of C_1 it inherits this attribute but this concept's relatively most defining attribute is that of Infrared. So by its parent displaying LED and itself displaying Infrared, it should be read as those under concept C_3 have a high probability of containing both LED and Infrared.

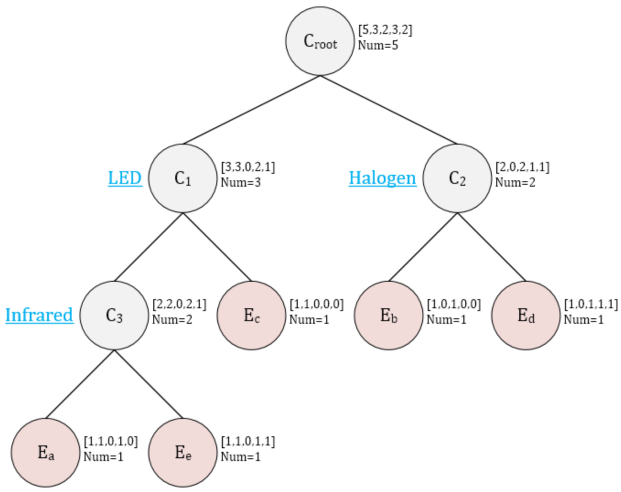


Figure 6.5 – Ideal attribute display in visualization

An initial implementation has been made by at the nodes displaying those attributes that have the largest relative percentage of the total occurrences of the given attribute. A calculation which is made by $n_{i\,local}/n_{i\,Tree\,Root}$ where n is the count of the attribute i for the local node under investigation as well as for the tree root node. Having defined the attributes relative representation, a selection function has been defined in addition to select which of these attributes to display at which node. This function first determines the relative percentages of occurrences for the individual attributes of the nodes. These are than ordered from high to low. Based on a user input on the number of attributes to display the corresponding top items are selected and presented.

However, there are situations in which certain attributes should not be displayed. First there are those situations in which a concept has only one patent with the given attribute present, in this situation this has no explanatory value for the concept and should not be displayed. Second, there is the notion of what will be described as the 'Carry Through Ratio' (CTR) in which the majority of the counts for the attribute are present in one of the concept children, in this case it doesn't have explanatory value of the concept under investigation but rather of said concept under it. For this reason the CTR of the attribute will be calculated for the concept's children by $CTR=n_{i\,child}/n_{i\,local}$. This CTR is compared with a user defined variable called the 'Carry Through Threshold' (set to e.g. 0.95), and when the CTR is equal or lower, the given attribute should not be displayed. Third, and strongly related to the second situation, is the notion of inheritance which will be described as the 'Trace Back Ratio' (TBR). This is added as when a concept attribute has a 100% probability of occurring but so does its parent it shouldn't be displayed. Therefore a step is added to check the relative share of the attribute counts of the concept under investigation related to its parent. This is done by first calculating the count share α by $\alpha = n_{i\ local} / n_{i\ parent}$, followed by the patent share β by $\beta = n_{patents\ local} / n_{patents\ parent}$. The TBR is then calculated by $TBR = \alpha / \beta$ which is compared with a user defined variable called the 'Trace Back Threshold' (set to e.g. 1.05). When the TBR is equal or lower the given attribute should not be displayed.

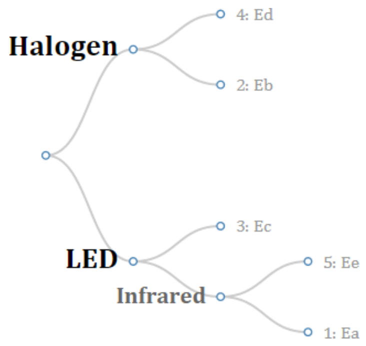


Figure 6.6 – Attribute display in actual visualization

Applying the described method of displaying attributes including the exceptions to the situation as shown in Figure 5.5 and based on the attributes [Light, LED, Halogen, Infrared, Car], results on the output as shown in Figure 6.6. In this visualization the size and the tone of grey of the attributes are linked to their relative number of occurrences (the more relatively present the darker and larger the text). As can be seen the attributes to display as determined by the prescribed function are equal to those described in Figure 6.5.

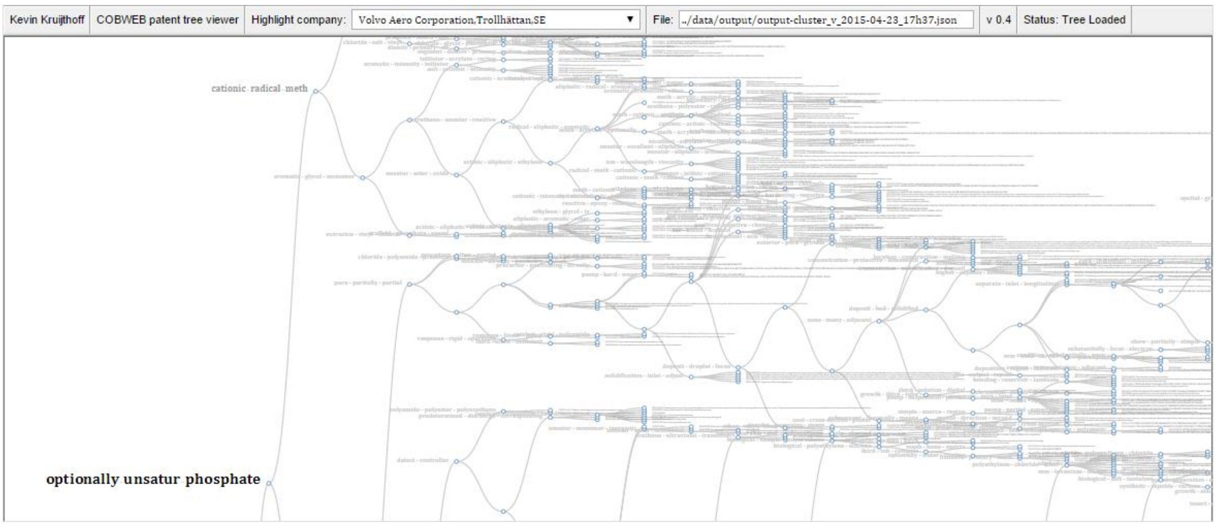


Figure 6.7 – Layout of the online tool for visualization

Figure 6.7 shows a screenshot of the developed online visualization tool for the interactive representation of the conceptual clustering output of inter-patent relations hierarchy.

6.4. CONCEPTUAL VERIFICATION OF THE STRUCTURE

To verify that the implementation of the COBWEB algorithm is performed correctly, a structural verification is performed. This is done by a comparative creation of a dataset in which one categorization is created by manually calculating the categorical utilities and deriving the structure and the other categorization is created by the algorithm. In this verification both the final structure as well as the individual calculations of the categorical utilities will be assessed.

For the structural verification the entries as shown in Table 5.1 will be used. However in comparison to the example shown in the Figure 5.5 for this verification all the COBWEB operators that are add, new, split and merge will be put into use (in the example resulting in Figure 5.5 only the add operator is used). The first two entries will generate two concepts under the tree root node. Therefore the first entry of interest is the third entry. As an example the calculations for the Categorical Utility (CU) of adding E_c into a concept together with E_a are shown.

$$\mathrm{CU} = \frac{\sum_{k=1}^{n} P(C_k)\left[\sum_i \sum_j \mathrm{P}\left(A_i = V_{ij} \middle| C_k\right)^2 - \sum_i \sum_j \mathrm{P}\left(A_i = V_{ij}\right)^2\right]}{n} \quad \text{Equation 6.1}$$

$$\mathrm{PCG}_{C_1} = \sum_i \sum_j \mathrm{P}\left(A_i = V_{ij} \middle| C_k\right)^2 = 1^2 + 0^2 + 1^2 + 0^2 + 0^2 + 1^2 + \frac{1^2}{2} + \frac{1^2}{2} + 0^2 + 1^2 = 4.5 \quad \text{Equation 6.2}$$

$$\mathrm{PCG}_{C_2} = \sum_i \sum_j \mathrm{P}\left(A_i = V_{ij} \middle| C_k\right)^2 = 1^2 + 0^2 + 0^2 + 1^2 + 1^2 + 0^2 + 0^2 + 1^2 + 0^2 + 1^2 = 5 \quad \text{Equation 6.3}$$

$$PCNK = \sum_i \sum_j \mathrm{P}\left(A_i = V_{ij}\right)^2 = 1^2 + 0^2 + \frac{2^2}{3} + \frac{1^2}{3} + \frac{1^2}{3} + \frac{2^2}{3} + \frac{1^2}{3} + \frac{2^2}{3} + 0^2 + 1^2 = 3.67 \quad \text{Equation 6.4}$$

$$\mathrm{CU} = \frac{\frac{2}{3}\left(\mathrm{PCG}_{C_1} - PCNK\right) + \frac{1}{3}\left(\mathrm{PCG}_{C_2} - PCNK\right)}{2} = \frac{\frac{2}{3}(4.5 - 3.67) + \frac{1}{3}(5 - 3.67)}{2} = 5 \quad \text{Equation 6.5}$$

Referring back to the Equation 5.3 is Equation 6.1 which will be entered for the virtually adding of E_c into a concept with E_a. The filled in equation can be seen in Equation 6.5. In this equation *PCNK* represents the part of the equation that encounts for the probability of correctly guessed with no knowledge which is calculated in Equation 6.4. PCG_{C1} represents the probability of correctly guessed attributes of concept 1 which virtually includes E_c and is calculated in Equation 6.2. PCG_{C2} represents those of concept 2 calculated in Equation 6.3, as it contains only one patent, this will automatically gain a maximum score equal to the amount of attributes (in this case 5). The *PNCK* and both the *PCG* variables are made up of 10 squared elements, which are 5 pairs. Each pair consists of the probability of the concept containing the attribute and the probability of the concept not containing the attribute. For example, E_b has a value of 1 for A_1 which results in $A_1^2 = 1^2$ and $(1 - A_1^2) = 0^2$. As all the equations performed in the verification are structurally alike and don't have further explanatory value, for the rest of the manually performed equations in the structural validation only the outcomes will be shown.

Table 6.3 – Manual calculation of the classification of the test dataset

Entry	CU add	CU new	CU split	CU merge	Action	Figure
E_a	n.a.	n.a.	n.a.	n.a.	New under Tree root	Figure 6.8
E_b	Add to E_a 0.000	0.750	n.a.	n.a.	New under Tree root	Figure 6.9
E_c	Add to E_a 0.498 Add to E_b 0.332	0.443	n.a.	0.000	Add to E_a in C_1	Figure 6.10
E_d	Add to C_1 0.270 Add to E_b 0.563	0.541	n.a.	0.000	Add to E_b in C_2	Figure 6.11
$E_{e\ \mathrm{step}\ 1}$	Add to C_1 0.493 Add to C_2 0.326	0.439	0.439	0.000	Add to C_1	
$E_{e\ \mathrm{step}\ 2}$	Add to E_a 0.283 Add to E_c 0.118	0.300	n.a.	0.000	New under C_1	Figure 6.12

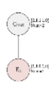


Figure 6.8 – E_a

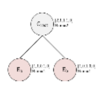


Figure 6.9 – E_b

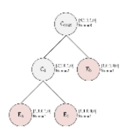


Figure 6.10 – E_c

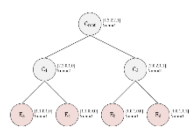


Figure 6.11 – E_d

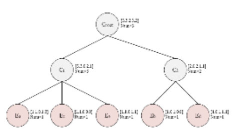


Figure 6.12 – E_e

The results of the manually calculated values for the test dataset is shown in Table 6.3 together with the corresponding visualization of the steps in Figure 6.8. To be able to verify the output of the implemented algorithm both the finalized output as well as the individual calculations need to be verified. To be able to perform the latter, a print is done when running the algorithm of each performed calculation. Listing 6.1 shows the printed outputs of the algorithm when parsing the test dataset. The generated hierarchy can be seen in Figure 6.13. When this is compared to Figure 6.12 it can be seen that the hierarchical structure of both the manually calculated categorization as that of the one generated by the algorithm are identical. When comparing the individual steps it can be seen that all made decisions are equal. There are some minimal differences in the actual values of the CUs but these can be explained by the rounding done in the manual calculations. Based on the verification it can be concluded that the prototype is structurally valid and can be used confidently for the categorizing of patents.

Listing 6.1 - Generated decisions for verification test 2

```
Adding patent: Ea
----
Add new node with ID: 1 under node with ID: 0
----------------------------------------------------------------
Adding patent: Eb
----
Cu: 0.000 as best addition under branch of node with ID: 1
Cu: 0.750 when adding new concept
----
Add new node with ID: 2 under node with ID: 0
----------------------------------------------------------------
Adding patent: Ec
----
Cu: 0.500 as best addition under branch of node with ID: 1
Cu: 0.444 when adding new concept
Cu: 0.000 when merging
----
Add to child with node ID: 1
----------------------------------------------------------------
Adding patent: Ed
----
Cu: 0.562 as best addition under branch of node with ID: 2
Cu: 0.542 when adding new concept
Cu: 0.000 when merging
----
Add to child with node ID: 2
----------------------------------------------------------------
Adding patent: Ee
----
Cu: 0.493 as best addition under branch of node with ID: 1
Cu: 0.440 when adding new concept
Cu: 0.000 when merging
CU: 0.440 when best split
----
Add to child with node ID: 1
----
Cu: 0.278 as best addition under branch of node with ID: 4
Cu: 0.296 when adding new concept
Cu: 0.000 when merging
----
Add new node with ID: 7 under node with ID: 1
```

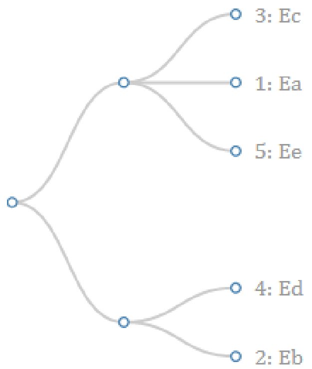


Figure 6.13 – Generated structure for verification test 2

Having developed a prototype visual patent set representation system built around the conceptual clustering of patents and having done a structural validation of its operation. It has been proven that the conceptual clustering technique can successfully be applied to a patent set. Now, to determine whether this technique's output results in a realistic representation of a technological field as well as to determine for which aspects of the strategic technology planning it could potentially be used. The developed prototype will be evaluated for use within the STP process, which is described in Chapter 7.

— To know what you know and what you do not know, that is true knowledge.
Confucius

7

Conceptual Clustering of Patents in practice: The case of Additive Manufacturing

This chapter will present an evaluation based on the design as presented in Section 4.6. It will evaluate the implemented elements and will present implementable solutions for those elements that have not been included in the developed prototype. The evaluation will be done per individual group block of Design Parameters in Section 7.1. The illustration of this evaluation is done based on a specific set of patents, to create a realistic sized technological field conceptual clustering hierarchy. The specific case is that of additive manufacturing and will be further elaborated in Section 7.2. Having evaluated the designed elements, Section 7.3 will finally refer the Design Parameters back to the customer needs and based on this elaborate for which of these needs the developed prototype could potentially be applied. In addition it will present the conclusions to be drawn from the case study as well as the scope under which the evaluation is valid.

7.1. Design parameter Group based evaluation

In this section an evaluation will be provided for each individual group block of Design Parameters as presented in Section 4.6. For those elements that not have been included in the developed prototype an implementation suggestion is provided when reasonably applicable. These will be combined and presented in Section 7.3 showing an overview of the perceived costs and benefits of implementing the individual elements.

7.1.1. A. Text mining

A-1 Setup & attributizing

This design parameter group block is aimed at transforming the text file that contains the patent entries into attributized patent entries. This is implemented as described in Section 6.1. The verification of the text-mining software's compliance with the designed process variables is added in Appendix III.

A-2 Subject Action Object structures

As the development of the prototype focused on the analysis to answer the question whether or not it would be possible to apply the conceptual clustering technique to patents. The most basic method of attributizing has been applied first, which is that of a binary method resulting in a yes or no per attribute. The extraction of Subject-Action-Object (SAO) based structures, as introduced in Section 4.3.1, is however suspected to be an improvement which suits multiple purposes. As the SAO structures are aimed at capturing the structural relationships among technical components within the patent, resulting in a combined set capturing a more detailed picture of the patents key-findings (Park et al., 2011). The first improvement can therefore be the use of SAO structures as attributes instead of single words. Secondly it forms the base for the implementation of other process variables that are necessary for the search for general solutions for specific problems as captured by customer need CA 5, which can be found in Table 4.5.

Potential implementation can be done based on the implementation presented by Park (Park et al., 2011) as described in Section 4.3.1 and will require an adjustment to the text-mining software. This implementation includes semantic analysis done by applying Knowledgist™ 2.5, a commercial NLP program, as well as the use of the WordNet database (Princeton, 2014) for the inclusion of cognitive synonyms. This will require a significant effort to implement, but the resulting extra functionalities are manifold.

A-3 Exporting

The exporting process variable is represented in the text-mining software which transforms the attributized patents captured in a dictionary into a JSON structure and writes it to a text file in the designated folder. The resulting JSON structure of the attributized patents can be seen in Appendix III which covers the evaluation of the text-mining software's compliance with the process variables.

7.1.2. B. Clustering

B-1 Importing

The attributized patent data that is stored in the JSON file as a result of the text-mining software, is the data input to the developed COBWEB Python software. The individually stored patents are iteratively read to be placed in the conceptual clustering hierarchy.

B-2 Problem input

When a generic solution is searched for a specific problem, an input option needs to be present. In the developed prototype this is not the case, as there is currently no SAO structural analysis present in the text-mining software. However when this is would be added, the software should allow the user to enter the needed Subject, Action and Object combination. This can be obtained by simply adding three variables in the COBWEB software which should be able to be filled in by the user.

B-3 SAO pairwise comparison

In the current conceptual clustering software the SAO pairwise comparison is not included as these structures are currently not generated in the text-mining phase. As a generic solution is searched for a specific problem, the problem input variables will all need to be checked for synonyms as well, this can be done via implementation of the WordNet database cognitive synonyms (Princeton, 2014). Section 4.3.1 has presented multiple solutions for the implementation of semantic technological similarity analysis, of which the implementation presented by Park (Park et al., 2011) seems to be most promising solution for this situation. In their method they use an inter-relational 2D representation. However as it concerns only the similarity to the inputted technological solution searched for, this step can be removed from the process. Instead, a list of the highest ranking similarity presenting patents can be kept and transferred along into the visualization.

Important to mention here is that the results will be strongly dependent on the selected dataset and the user should be made aware of this. This is of importance when this feature is used for TRIZ type applications where a generic solution is searched for a specific problem. As the selected data set is most likely related to that of the technology that is being developed, the chance of finding a similar functional solution in a different technological field is not present.

B-4 Conceptual clustering

As described in Section 6.2, the COBWEB conceptual clustering algorithm has been implemented in the software. Its use results in a hierarchical representation of the entered patents which contains the structure, the linked patents and the individual nodes' counts of attributes. It has been conceptually verified in Section 6.4 and Figure 7.9 shows an image of a complete structure (containing 9360 patents) generated with the software.

The result of the conceptual clustering script can be improved, resulting in representations that better capture the actual inter-patent relationships. This can be done by the implementation of a fifth COBWEB operator as described by Fisher to only promote the best host when merging two nodes (Fisher, 1987). This can be easily added as the current software allows for this additional functionality to be implemented. A second improvement can be done by making use of weighted attributes based of the number of occurrences of an attribute. This requires a rewriting of the text-mining and conceptual clustering software and is therefore of more impact than adding the operator. A final improvement could be made by applying SAO type structures instead of words as attributes, as described in Section 7.1.1/A-2 Subject Action Object structures, which are expected to better capture the actual innovations embedded in patents.

Next to the potential improvements to the clustering operation, additions can also be made in this stage to allow for more interaction with the data. This could include the rewarding of penalizing of tree shapes to reduce the depth or the width of the hierarchy structure. Which can be realized by applying weights in the determination of the categorical utility which are then based on the depth of the node within the hierarchical structure. Another option is the seeding of attributes, forcing the starting point of a patent that contains a certain attribute to follow a given path at the start.

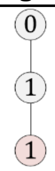
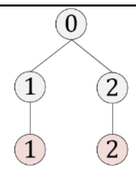
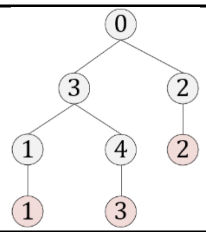
B-5 SIBLING SIMILARITY

The determination and storing of similarities between patents is not implemented in the developed prototype. There are however multiple applicable solutions that can be implemented. Based on the patent's position within the hierarchy, patents can first be selected for similarity evaluation. This would mean that within the software this module needs to be placed after the actual creating of the hierarchy. Section 4.3.1 contains an overview of semantic similarity (potential infringement) analysis solutions. However in this situation a similarity measure can also be calculated based on the attributes of the patents, which are already present. It is however questionable whether this will capture the truly similar technologies, and as of the importance of its outcome further research into this is advised.

B-6 CHANGE LOG

The current implementation of the COBWEB software does not keep track of changes made to the hierarchical structure and the placing of the patents it contains. When however the evolution of the hierarchy is wished to be studied, this functionality should be added. This can be done by keeping track of the additions and removals of nodes as well as of any changes in under which parent node a patent or node is situated. An example of a change log is presented in Table 7.1, in which the column showing the resulting situation is illustrating the resulting hierarchy after the step changes have been made.

Table 7.1 – Example change log

Step	Sub-Step	Element	Action	Target	Resulting situation
1	1.1.	Node 1	Add	Node 0	
	1.2.	Patent 1	Add	Node 1	
2	2.1.	Node 2	Add	Node 0	
	2.2.	Patent 2	Add	Node 2	
3	3.1.	Node 3	Add	Node 0	
	3.2.	Node 1	Change parent	Node 3	
	3.3.	Node 4	Add	Node 3	
	3.4.	Patent 3	Add	Node 4	

The proposed method of login the changes and additions made to the hierarchical structure can be added to the COBWEB software as an additional module. The resulting log table can be exported to e.g. a MySQL, JSON or

XML structure based on the visualization implementation needs. The addition of the change log and an export capability require mediocre additions to the currently developed prototype software.

B-7 Extracting & Exporting

Once all the patents have been parsed through the conceptual clustering algorithm, the resulting hierarchical structure is extracted within the same software and written to a JSON file. The structural extraction is done via the use of a function that loops through all the parent-child relationships of the nodes, starting from the root node. A complete description of these functionalities can be found in the software manual in Appendix V and an example of the created JSON file of a resulting conceptual clustering hierarchy is added in Appendix IV.

7.1.3. C. Visualization

C-1 Import & Structure

The created conceptual clustering hierarchy that is stored in the JSON file as a result of the COBWEB software, is the data input to the developed online visualization tool. The hierarchy is loaded into the JavaScript based viewer by a JSON loader. Which is then visualized exploiting the use of a d3.js library based tree viewer to create the hierarchy, as described in Section 6.3, from a JSON ordered file into a tree structure.

C-2 Node/Structural explanation

As can be seen in Figure 7.1, the developed prototype contains an interpretation enhancing visualization of attributes. Based on the node's relative most defining attributes the words are selected and based on their size and color, represented in importance. The number of words and the number of different scales in size and color are adjustable.

A point of improvement is detected in the situation of the screen being zoomed out to observe the overall structure, the current implementation doesn't scale up the text which results in no explanation interpretable in this mode. This could be improved by linking the text size to the zoom handler in the JavaScript implementation, and when doing so, selecting those with high scores to be visualized when space is limited.

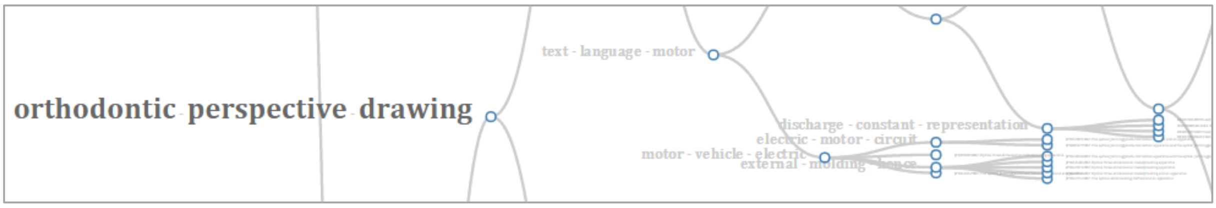


Figure 7.1 – Example of defining attribute visualization

C-3 Evolution

The visualization of the evolution of the hierarchical structure over time is currently not present in the developed prototype. Before implementation it prerequisites the logging in the conceptual clustering software of the changes to the hierarchical structure over time, as described in B-6 Change log. This will the form the base for a visualization which can show the evolving of the field over iterations of patents being added. This will however not require an extension of the current method of visualization but rather a complete new setup of implementation. This is due to the fact that it requires a completely different core functionality and it is therefore advised that when this functionality is wished for, to develop a new visualization method for this specific use.

7.1.4. D. Exploration

To increase the interpretability during exploration, as well as to assist during communication by not displaying irrelevant aspects of the structure. The developed prototype contains a function to hide everything under a certain node, which can be seen in comparison in Figure 7.2 and Figure 7.3. The node with the red circle added for clarification, of which its children (within the added red rectangle) are still visible in Figure 7.2 is hidden by clicking on the specific node. The result can be seen in Figure 7.3, in which the specific node now has no children visible. To indicate that this node does have children underneath, the specific node has changed in color from white to blue.

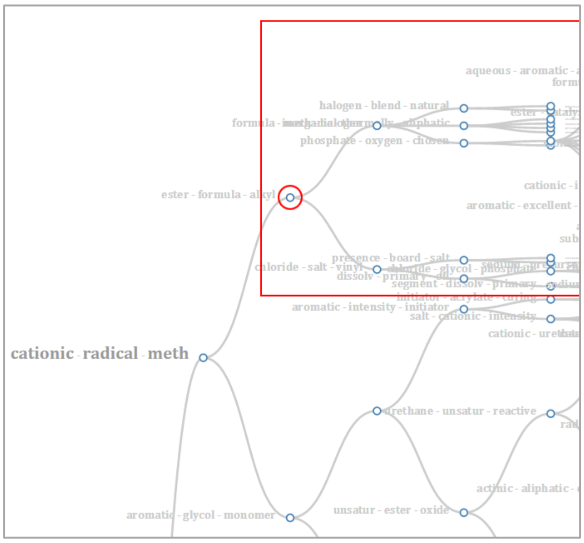


Figure 7.2 – Node children visible

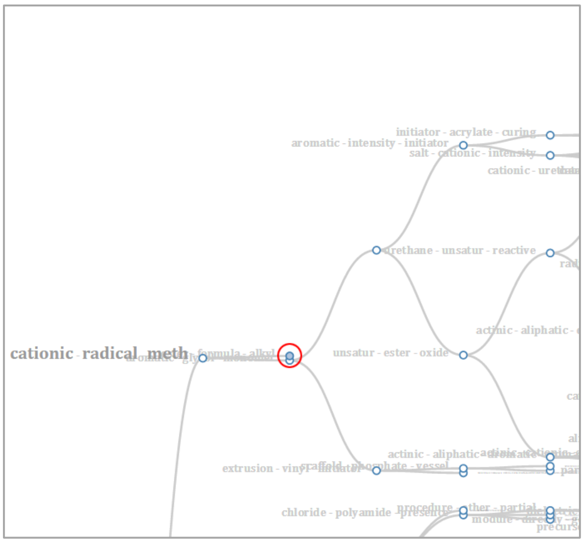


Figure 7.3 – Node children hidden

To assist the user in exploring the hierarchical structure, handling operators are added to enable the zooming in on and dragging off the structure. These are created to work on a computer by using the mouse scroll for zooming as well as for click and drag functionality. On mobile devices with touch screens these can be operated by the pinch-to-zoom and touch-and-drag functionalities.

7.1.5. E. ANALYSIS

Within the analysis group block, two of the three design parameters have been included in the developed prototype. These are implemented to highlight company specific patents and to visualize individual patent information. The visualization of the individual patent information can be seen in Figure 7.4 and Figure 7.5 which show patent specific bibliographic data fields (in this case the publication code, the title of the patent and the patent owning company). The latter is a screenshot of the visualization which has the setting switched on to visualize the patent owning company.

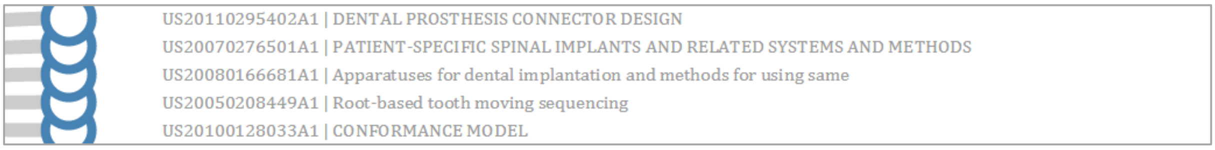


Figure 7.4 – Individual patent bibliographical data

Figure 7.5 – Individual patent bibliographical data including company

Regarding the highlighting of company specific patents, a selection box containing all the company names that are represented in the visualization is present[7]. Figure 7.6 shows what happens when in this case 'Align Technology Inc., Santa Clara, CA, US' is selected for the highlighting of individual patents. All the patents that have this specific company listed as their first owner have their appearance altered. In the developed prototype a bold

[7] For this prototype the first company of the patents is selected to be used in the visual representation

text and adjusted text color is applied as well as the highlighting of the node by altering its color. As this is determined by a style setting, the method of highlighting the selected patents can be implemented according to the wish of the user.

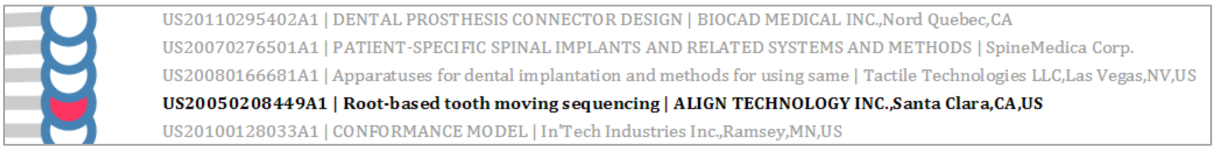


Figure 7.6 – Individual patent bibliographical data including company and highlighting patents of a certain company

In the created visualization, the bibliographical data used for the visualization is hard-coded within the JSON file that stores the hierarchy. To save up on data storage and to allow for the visualization of user requested data fields, the analysis representation could be improved by making use of the retrieval of data from the patents based on their unique publication code.

The highlighting of company strategy specific technology areas has not been included in the developed prototype. It is also questionable whether this would be implementable within reasonable effort, which will be further elaborated in Section 7.1.6. When however the decision is made to include this in the visualization, this could be added by making use of a node lists. Containing the technological fields according to the strategy that are represented by certain nodes that need to be highlighted. In a contrasting manner, the company owned patents should be highlighted, as such that they can be observed simultaneously and conclusions can be drawn about compliance with the set strategy.

Furthermore, extra analysis options could be added to the visual representation to provide even more insight, such as statistical descriptors e.g. showing which company owns the most patents, owns the most influential patents or show geographical location information.

7.1.6. F. Strategy Positioning

The determination of strategy positioning for visualization has, as mentioned in Section 7.1.5, not been included in the developed prototype. To be able to do so, first an automatic indexing of the created hierarchy would be needed. Based on which the company strategy technologies could be located on the hierarchy structure and should then be reverse analyzed whether or not the company is represented in this area accordingly. This results in such a specific and large addition that it is questionable whether it wouldn't be better off as an isolated form of analysis instead of trying to include it in the developed prototype.

7.1.7. G. Quality Analysis connection

The determined process variables that should be added in this group to comply with the functional requirements are not included in the developed prototype. It covers the ability of displaying the linking and citation overview of the patents which should be visualized on request of the user on top of the visualization. This results in two objectives; the retrieval of the data and the visualization. The retrieval of the information can be done 'live' in the visualization by exploiting an Asynchronous JavaScript and XML (AJAX) function running an XML interface with for example the European Patent Office (EPO) database. The EPO offers a service called the Open Patent Services (OPS) which allows for this method of data retrieval to be embedded into web-based applications (European Patent Office, 2014). Ideally the visualization would show connections between the target patent and the patents it links to and its forward and backward cites. However there is a chance that these patents are not part of the selected set of patents that are used to visualize the specific patent field. This can for example be due to them being in a different technology field. By applying this visualization it can thus potentially result in a skewed perspective on reality. Also, in the current setup it is impossible to actually show the links between patents that are linked by citations. To realize this an additional layer should be added on top of the visualization. An easier to implement option would be to highlight those patents that the given patent links to instead of visualizing an actual connection between them. This could quite easily be done by making use of the patent highlight function that is now being used for highlighting patents of a certain company and thus requires minimal extra functionality added to the existing visualization. This will however still leave the non-visualization of out of scope links and citations. A solution in the form of a list presenting all links and citations might therefore be applied here to overcome this issue.

7.1.8. H. COMPANY DATABASE CONNECTION

To display the inter-company relationship with the owner of a patent, a connection to an external source is necessary as this information is not embedded within the patents. By exploiting the use of an external source the threat level or the likelihood to exchange or obtain licenses can be displayed. In the developed prototype this is not implemented, this can however be easily implemented by making use of an AJAX/PHP/MySQL addition to the current visualization prototype. Which means that the visualization can in real-time extract data (using AJAX and PHP) from a table (which can be stored in a MySQL database). This data which will be entered by its users based on the relationship with certain companies can be linked by company name or an assigned company ID. The method of implementation will be dependent on the user needs, but it could for example be based on an ordinal scale of low, medium or high risk of law suits when potentially infringing said company's patent.

7.1.9. CONSTRAINTS

C 1. EASE OF USE

The developed Python scripts are intended to be used by Technology scouts. They are perceived to be highly educated and to have at least mediocre to good computer skills. It is therefore expected not to result in any issues regarding the execution of these scripts. When however deemed necessary (for operation or commercial purposes), the usage of these scripts can be made more user-friendly by creating a software shell around it that incorporates these scripts and which guides the user in its operation.

By making use of the web-based visualization, the resulting hierarchy can be easily explored independent of operating system or device, as long as a web-browser with JavaScript is present. It has however been detected that when accessed on mobile devices (phones and tablets) and viewing a large hierarchy structure (c.a. 10000 entries) the page will not load due to the devices available allocated memory for the browser. Based on the intended type of use this should not be a problem as it is aimed to be used on computers. However, for future communication purposes the intended usage might include tablet users, for which a detection and an optimization should then be realized e.g. in the form of a hierarchy reduced to its most important structure.

C 2. ROBUSTNESS OF OUTCOME

The robustness of the outcome will depend on the applicability of multiple sources. Therefore, on its own the developed prototype cannot fulfil this constraint. This holds however for all individual techniques and is therefore not a reason to decline this specific method for application. However, by applying it alongside the currently available analysis methods, the robustness of the decision making process enhances.

C 3. REPRODUCIBILITY OF OUTCOME

The reproducibility constraint is met by the developed prototype. This is inherent to the method of conceptual clustering. Which uses the same method of operation in each try, resulting in a constant output. This means that when ran on the same data file, containing the patent set, irrespective of the location or operating system the software is ran, the final outcome will be constant.

C 4. EXPLAINABILITY / INTERPRETABILITY OF OUTCOME

Because of the resulting hierarchy structure and the fact that the individual nodes contain the attribute counts of all the patents underneath, the reason why a patent is at a certain location can be completely traced back. To show the possibility of this functionality, in the developed prototype the 10 attributes that have the most relative influence for any given node (as explained under C-2 Node/Structural explanation) can be made visual, as can be seen in Figure 7.7.

The visualization of these attributes occurs when the cursor is placed on the node. This will result in a display of the number of patents under this node, followed by the 10 most influential attributes. The attribute name is shown followed by the percentage off all occurrences of this attribute in the complete structure that are under this node. In addition the probability that any given patent underneath this node contains this specific attribute is displayed. To illustrate this, an example will be given from Figure 7.7; there are 1964 patents under this specific node, of which 16% (314 patents) contain the attribute “optionally” which entail for 83,7% of the total occurrences of this attribute (375 patents) throughout the whole hierarchy.

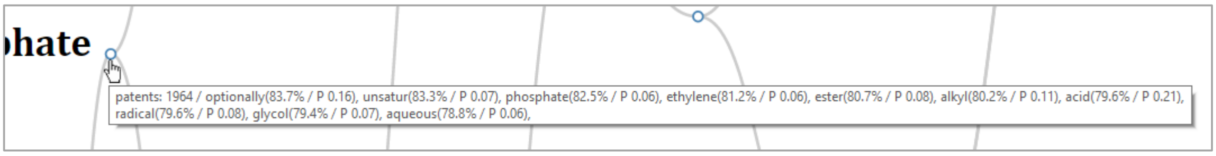


Figure 7.7 – Traceability of positioning via nodes

By making use of the developed and implemented method of labeling the nodes in this manner, combined with the hierarchical aspect of the conceptual clustering technique, the patents location can be explained and justified. This can be done to the point where one can say based on a specific combination of attributes a patent possesses what the probability is that it will be placed under a certain node. This is a clear advantage when compared to other representation methods as it is completely self-explanatory. Which results in the outcome to be more easily acceptable, as it is based on a clear method which can be completely traced back.

A potential point of critique though is that the method of representation might however mislead the reader into thinking that certain patents that are placed higher up in the hierarchy are also a higher level technologies which in practice not need to be true. This is something the user should be aware of when interpreting the resulting visualization.

C 5. Usability as a communication tool

As the visualization allows for adjustment based on specific needs at the moment of use, the developed prototype can be used to explain different aspects regarding the patent set. In comparison to other methods of visualization, this method allows for visualization of the overall structure as well as the inspection at the level of individual patents. Learn from the patents positioning as well as the retrieval of patent specific data. This can aid the user from different roles to get together and by interacting with the visualization, learn by exploring.

7.2. The case of Additive Manufacturing

As an addition to the evaluation of the design parameter groups, a realistically sized case has been parsed through the software and visualized using the developed prototype. The used patent set is a selection of 9360 patents related to additive manufacturing (3D-printing) retrieved from the Derwent patent database from Thomson Reuters which has been provided by SKF. The complete provided explanation of the search done to retrieve this patent set has been added in Appendix II. The actual received set contained a few more patents than the 9360 mentioned, but some have been excluded for not containing any information in the abstract field. Illustrating this field, Figure 7.8 presents a representation of the patent set generated with the self-organizing maps based software ThemeScape™, and has been provided by SKF.

Within the developed prototype, the text-mining has been done by assigning 750 attributes (empirically selected to be of high explanatory value) to each patent based on the analysis of their abstracts. In which nouns, verbs, adjectives and adverbs are included to be considered as attribute. This has been inputted in the conceptual clustering software, of which the resulting hierarchical structure has been uploaded to be visualized. The result of which can be seen in Figure 7.9, showing a zoomed out screenshot to capture the complete structure. To explore and interact with the structure go to http://www.kevinkruijthoff.nl.

7.2.1. Comparison of outputs and Enhancing capacities

The visualization of the conceptual clustering output as presented in Figure 7.9 has been analyzed to determine the high level branching structure, which can be seen in Figure 7.10. The construction of which is based on the visual analysis of the represented hierarchy as well as by exploiting the interpretation assisting labeling which is present in the prototype. The overview starts out in the center which represents the root node containing the full patent set underneath. The branches are arising from this point and the numbers at the arrow indicate the number of patents that are in the given sub-branch. The boxes content is determined by the relatively most defining attributes. For example from the root node, the first branch entails 1964 patents which are in the branch related to synthetic materials, chemicals, acids and techniques in which these are handled or used for additive manufacturing. The empty boxes represent a node at which a significant split occurs but for which the prior splits are too small (and of lower influence) to enlist them all. The blue labeling with numbers has been added for referencing purpose, which will be covered in the next paragraph.

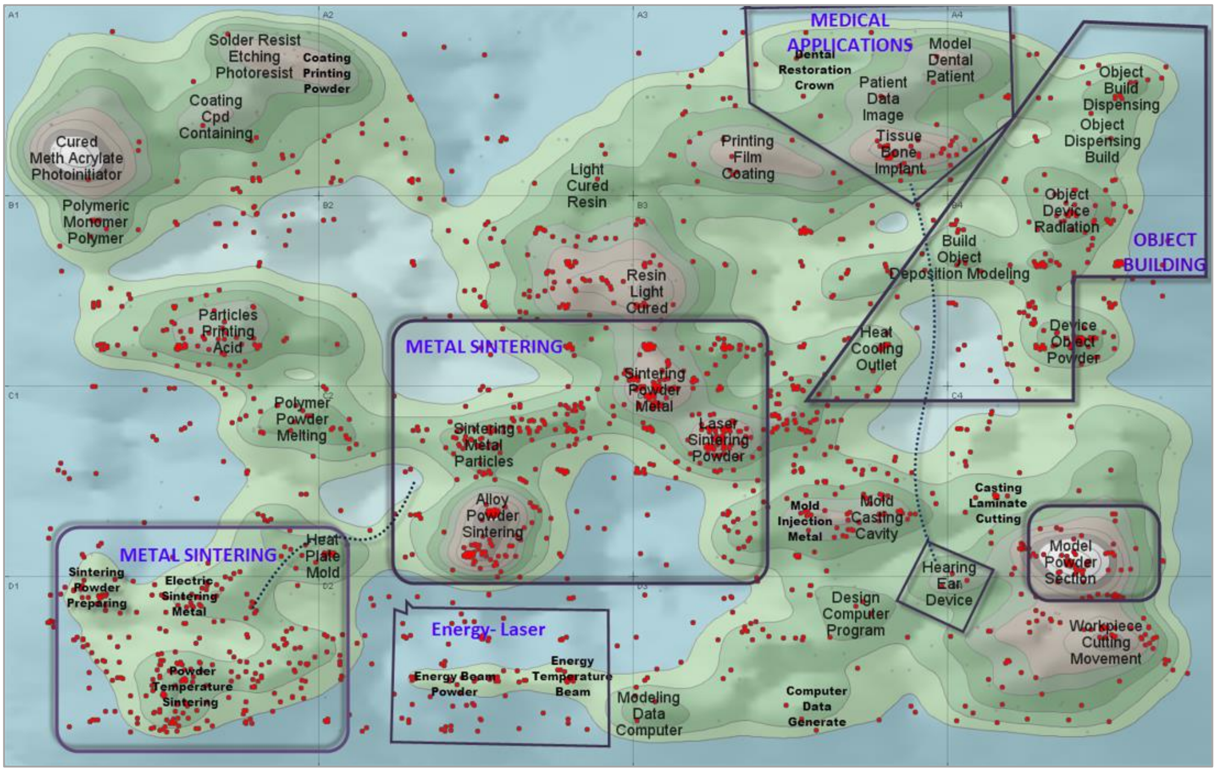


Figure 7.8 – Resulting Self-Organizing map of the additive manufacturing case patent set

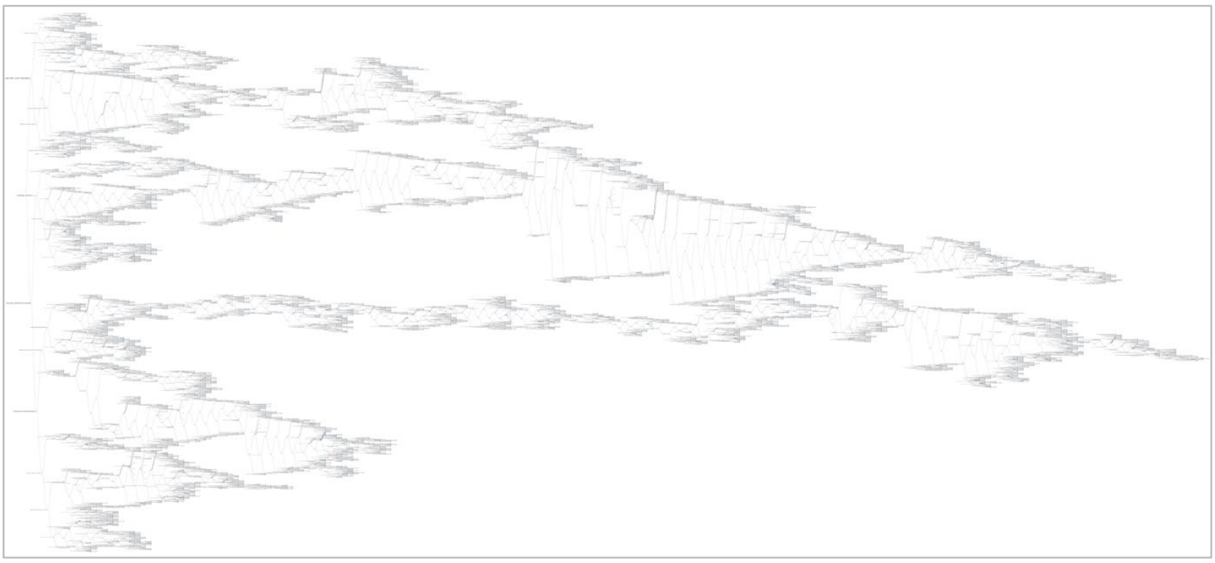


Figure 7.9 – Resulting Conceptual Clustering hierarchy of the additive manufacturing case patent set

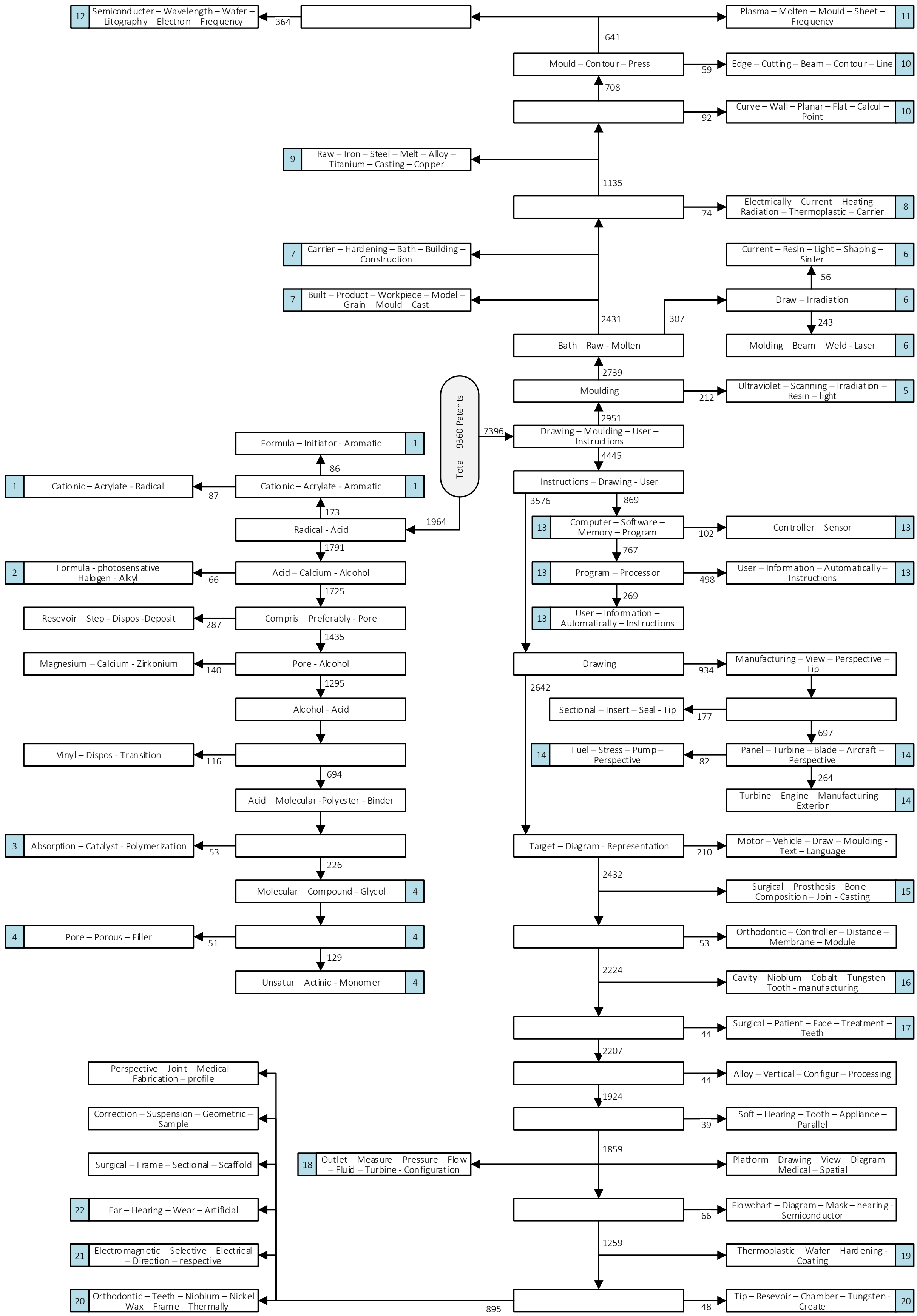


Figure 7.10 – Analysis of the conceptual clustering output

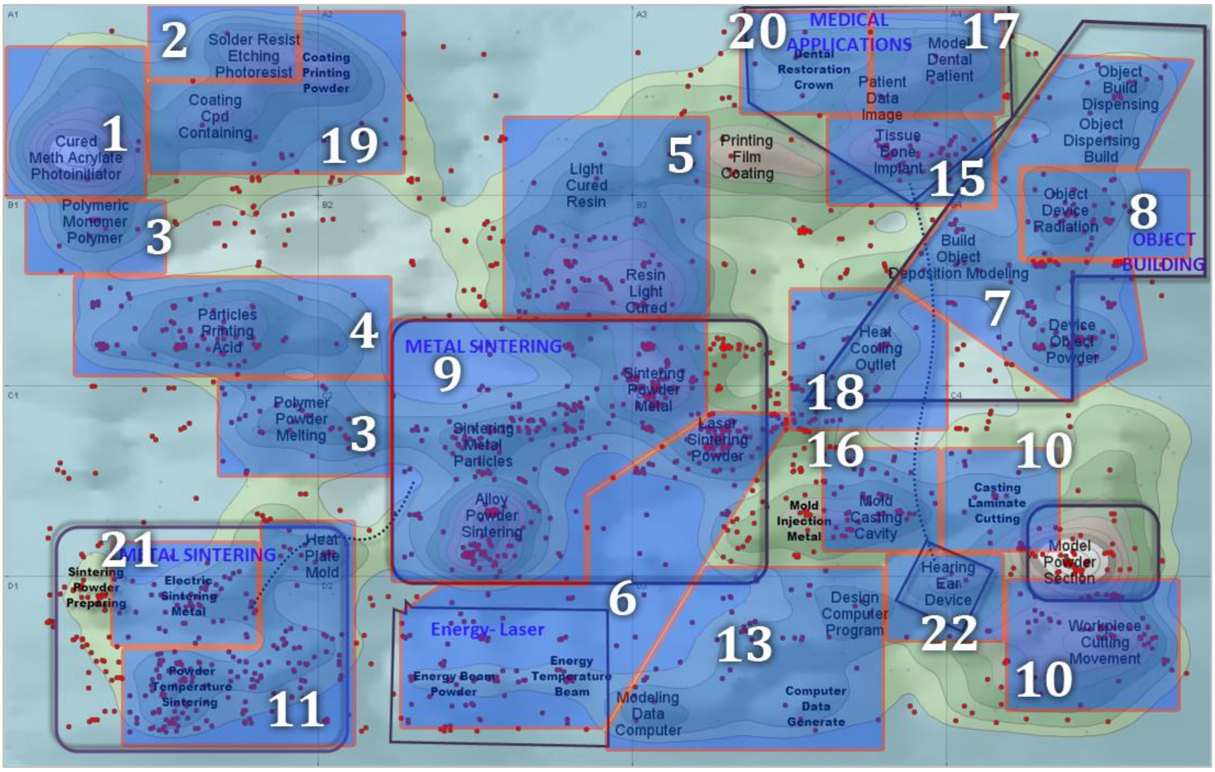


Figure 7.11 – Mapping of conceptual clustering clusters onto self-organizing map

Figure 7.11 shows the mapping of the conceptual clustering clusters onto the self-organizing map. This comparison shows that the main branches of the tree are represented by the hills in the self-organizing map, meaning that the local optima seem to match. However, two significantly sized branches of the tree are not represented in the self-organizing map. These are the application field of semiconductors (12) and the application field of aerospace (14). Furthermore, some other smaller application fields as to be found in smaller branches are also not represented.

Analyzing the structure presented in Figure 7.10, there are 4 main branches that can be determined. The first main branch (a) is covering the aforementioned synthetic materials, chemicals, acids and techniques in which these are handled or used for additive manufacturing, such as the use of light, esters or electrolysis for binding of materials. The second branch arising from the root node splits into two branches of which the first (b) contains patents related to metals and techniques to process or handle them. Which is why the moulding related attributes as well as those related to methods for its sintering[8], are the key determinants for the entrance of patents to this branch. The aforementioned cluster related to the semiconductor field is also to be found here, which is to be expected as they exploit the use of metals as their main elements for their conductive as well as control parts. Following the other branch from the second branch originating from the root node, two mayor sub-branches can be detected. These are firstly (c) a branch containing patents related to computer related concepts such as user operating systems as well as sensors and controllers. Secondly, (d) the patents related to drawing and specific branches such as medical and aerospace applications. Here sub-branches related to these application fields can also be detected, such as teeth cavity filling or the custom hearing appliances. The two mentioned branches of computer related patents and that of the specific applications seem to be related by their common exploitation of individual product customization at a high level and the specific interaction that this requires.

Figure 7.12 shows the four main branches described in the previous paragraph, mapped onto the self-organizing map as presented in Figure 7.8. As can be seen in (a), the group of local optima clusters of the conceptual clustering output of the first branch, can be also seen to group together in the self-organizing map, increasing the robustness of the conviction that this optima indeed are conceptually closely linked to each other. However positioned in the self-organizing map in close proximity to sintering or coating approaches, this might

[8] Sinter: "Make (a powdered material) coalesce into a solid or porous mass by heating it (and usually also compressing it) without liquefaction." (Oxford University Press, 2015b)

not be the case conceptually. As in the conceptual clustering output, this group is the first to branch of, thus showing the least similarity with the other groups. This is based on the types of technical approaches for manufacturing as well as the subject materials used for construction. However in the self-organizing map it can be the case that they are linked to sintering by the cause of both exploiting the fact that they heat up certain area to bind the subject material. This can be explained to be fitted in this manner from the perspective of the dimensional reduction that self-organizing maps applies.

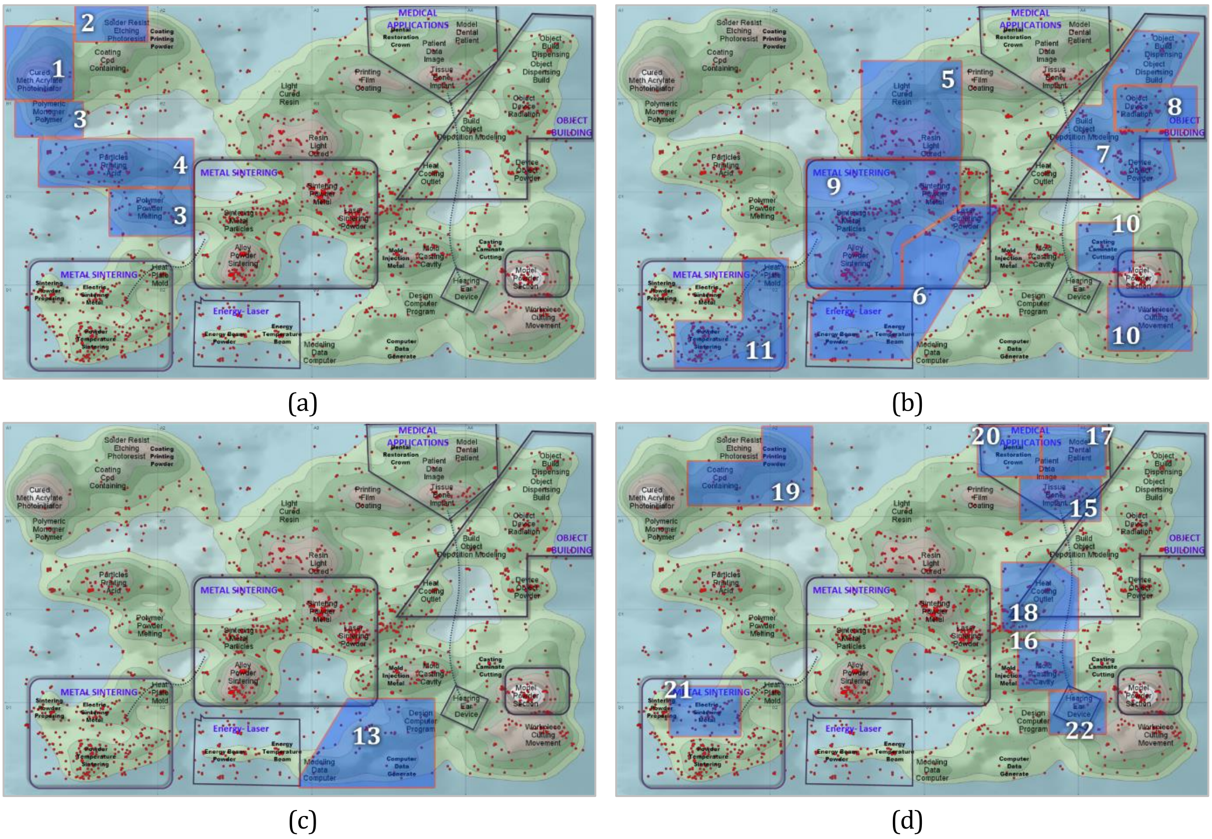


Figure 7.12 – Mapping of main branches onto the Self-Organizing map

In (b) the mapping of the second branch two separated groups can be detected when comparing it with the self-organizing map. Which are those of the metal material related concepts as well as the moulding and building related concepts. Which seem to be separated by the concepts of the fourth branch, as shown in (d). A mayor difference between the two methods is the location of the concepts related to the handling of the (mostly metal) materials. Which includes, among others, the cutting of edges and contours and the shapes of the workpieces such as curvatures or planar shapes. In the self-organizing map these are situated in the right bottom corner of the map where, when interpreting the conceptual clustering output it would be expected to lay more close to the metal sintering concepts as they seem to be related to the material and to the technical operations performed on it.

In (c) the grouping of the computer related concepts such as user operating systems as well as sensors and controllers. The grouping seems to show similarities between the two results. However the conceptual clustering output shows a significant cluster in this grouping which is related to user related concepts such as instructions and the presenting and entering of information, which is not directly detectable in the self-organizing map. The positioning in the self-organizing maps next to the laser related concepts can be explained by their exploitation of sensors and controllers which are grouped under the computer related concepts in the conceptual clustering output. The actual laser concepts themselves are in the conceptual clustering positioned within the second branch due to the metal subject matters they are used for. This can explain the resulting positioning of the laser related concepts of the self-organizing map.

In (d) the medical application branch, a local optima grouping can be detected as a string which is seemingly separating the metal sintering applications from the building applications within the self-organizing map, which may seem unexpected as noted earlier. It is interesting to see that the grouping related to casting and moulding for cavities (16) is situated relatively far from the other dental application concept groupings (17, 20). This can be because of the techniques applied in this group are focusing on casting, hence explaining the location on the self-organizing map. Where in the conceptual clustering the application field had more influence on its final position. Interesting to notice is that two sub-branches are placed far from the main group which are firstly those concepts related to the use of selective electrical and or electromagnetics methods of application. This could be because of the fact that this exploits the use of metallic subject matters resulting in a positioning in close proximity of the metal sintering concepts. Where in the conceptual clustering they are situated at the given position due to their application use in the medical field or the specific metals that they are aiming to bind which are for medical purposes. Secondly those concepts related to coatings and hardening seem to be positioned separately from the grouping. This can be due to the fact that it exploits the use of heating or uses specific chemicals. Where in the conceptual clustering they seem to be closely related to use in the dental applications such as cavities.

7.2.2. FINDINGS BASED ON THE COMPARATIVE CASE OF ADDITIVE MANUFACTURING

The comparative case has shown that on the high level the same optima seem to arise from the patent set. However the inter-relations of these optima are not always depicted in the same manner. One of the aspects of the self-organizing maps is the notion of dimensional reduction to show highly dimensional data on in this case a 2-D field. This can however result in the forcing of concepts to fit the designated 2-D field, which might result in concepts to be put in proximity to each other where in the high dimensional space this is actually not the case. The additional benefit of using conceptual clustering is here, to have another representation of the same patent space, which is based on a different technique of dealing with the high dimensional space. This can results in new insights into clusters or in verification of the idea about the patent space. Resulting in an increase of robustness and increasing the trustworthiness of the notion of concept optima's to have similarities. This can be of great assistance to patent attorneys that use the maps based on the patent claim fields to determine similarity in concepts for infringement analysis within the freedom to operate process, given their need for low risks.

It can be seen that applying the conceptual clustering method, results next to the notion of the main techniques more into branches of specific application areas. Which can be explained by the use of the categorical utility function, as introduced in Section 5.2, which acts upon differences between groups. It can therefore be seen that under the concept of metal sintering approaches we see a branch appear which has concepts related to semiconductors. The same holds for the branch containing the medical and aerospace related concepts, in which we can find specific application field branches. Such as the custom fabrication of joints which are more surgical related or the fabrication of bone prostheses. The added value of applying the conceptual clustering technique can here be a different view on the patent field under which the optima are further deviated according to their specific application field.

The combination of methods results in a more powerful tool for unveiling the inter-conceptual relations of the patents by comparing and overlaying their results. Which forces the user to question the resulting outputs that don't seem to have a one to one relation between the two visualizations. Furthermore, by exploiting the use of the conceptual clustering technique, explanations can be made for the positioning of optima on the self-organizing map. This is due to the fact that this technique allows for the investigation of the structure's resulting inter-relations by investigating the nodes' attributes throughout the tree.

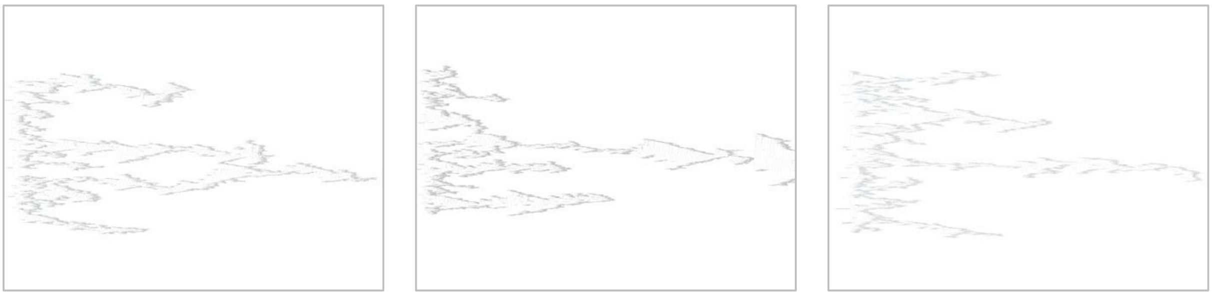
7.2.3. FURTHER FINDINGS OF USE

When analyzing the structure of the conceptual clustering output it can be seen that the depth of the structure relates mainly to the extent of similarity between concepts. It must therefore be concluded that the expectation that this method might be able to capture the hierarchical aspect of technology is not one to one relatable to the outcome of the conceptual clustering. This is expected to be due to the fact that the technique exploits the use of the categorical utility and therefore looks at best fitting locations based on similarities.

Regarding the analysis of the tree the expectation, as mentioned in Section C-2 Node/Structural explanation, regarding the ease of interpretation of the labels when exploring the complete structure is difficult due to their size. The presented solution of the adjustment of font-size based on the zoom-level could therefore be applied. However, a decision can also be made to develop the option of visualizing pruned down trees, showing the high level branching. And loading that part of the tree that is of interest when the user decides it wants to explore this.

This can for example be done by applying rough set theory for attribute reduction, which aims to reduce all redundant objects and attributes to find the minimum subset of attributes (Hassanien, 2004), or by exploiting the use of a tree pruning algorithm, which aims at combining similar branches (Patel & Upadhyay, 2012). Advantages would be the ease of interpretation and the computational needs. The downsides would be that due to the simplifications needed, important sub-branches might be removed as well as missing out on the possibility of seeing the complete structure and visually detect points of interest.

As mentioned before, multiple representations can be created of the same patent set based on the needs of the user and the decisions he needs support for. To illustrate this difference, multiple representations of the same dataset have been made, to show the effect of the specific choices made. The resulting high level structures, all based on 600 attributes, are shown below. In which Figure 7.13 is created based on the patent abstract and Figure 7.13 on the patent novelty claims. Figure 7.13 has been created for comparative reasons based on the patent abstract using the text-mining software as presented by Cunningham and Kwakkel (Cunningham & Kwakkel, 2015). To be able to better interpret the visualizations of these structures please visit the created online environment.



(a) (b) (c)

Figure 7.13 – Different representations of the same patent set

Empirical testing has shown that the inclusion of more attributes does not necessarily add to the insightfulness of the visualization. This has to do with the weighting of the actual information a patent contains and the method of operation of the conceptual clustering algorithm. For this we need to consider again the categorical utility function, as introduced in Section 5.2, on which the conceptual clustering algorithm relies. Being a summation of the individual attributes contributions, and all attributes having the same weight in this summation, the relative impact of those attributes containing the relevant information is reduced. For this reason it is important when conducting any research applying conceptual clustering to make multiple representations based on indexations done with different numbers of attributes. This does not mean that other attributes should not be taken into account. It is rather advisable to gain an even better insight, to first perform the high level attribute clustering. From this the branch of interest can be further examined to unveil more sub-branches. This can be done by after the run of the complete set selecting that node from which the branch of interest starts and extract the patents under this node into a new subset. This can for example be done via the hierarchical structure which is present in the JSON file, which structure is ideal for the selection of subsets. This subset can now be parsed through the text-mining and conceptual clustering software again. An example of this can be seen in Figure 7.14 in which a visualization is shown of a subset from the additive manufacturing case. In this case that of the 869 patents that are in the group of computer and software related concepts (group 13). The earlier mentioned branching into application fields can herein be detected e.g. such fields as hearing aids and electrical circuits. This will further define the inter-relations between the concepts in this subset as now a parameter set is made specifically for this subset, which will result in an even more informative branching. Another additional advantage of this approach is that when evaluating the more precisely conceptualized subset, the computational needs for the visualization will also be lower.

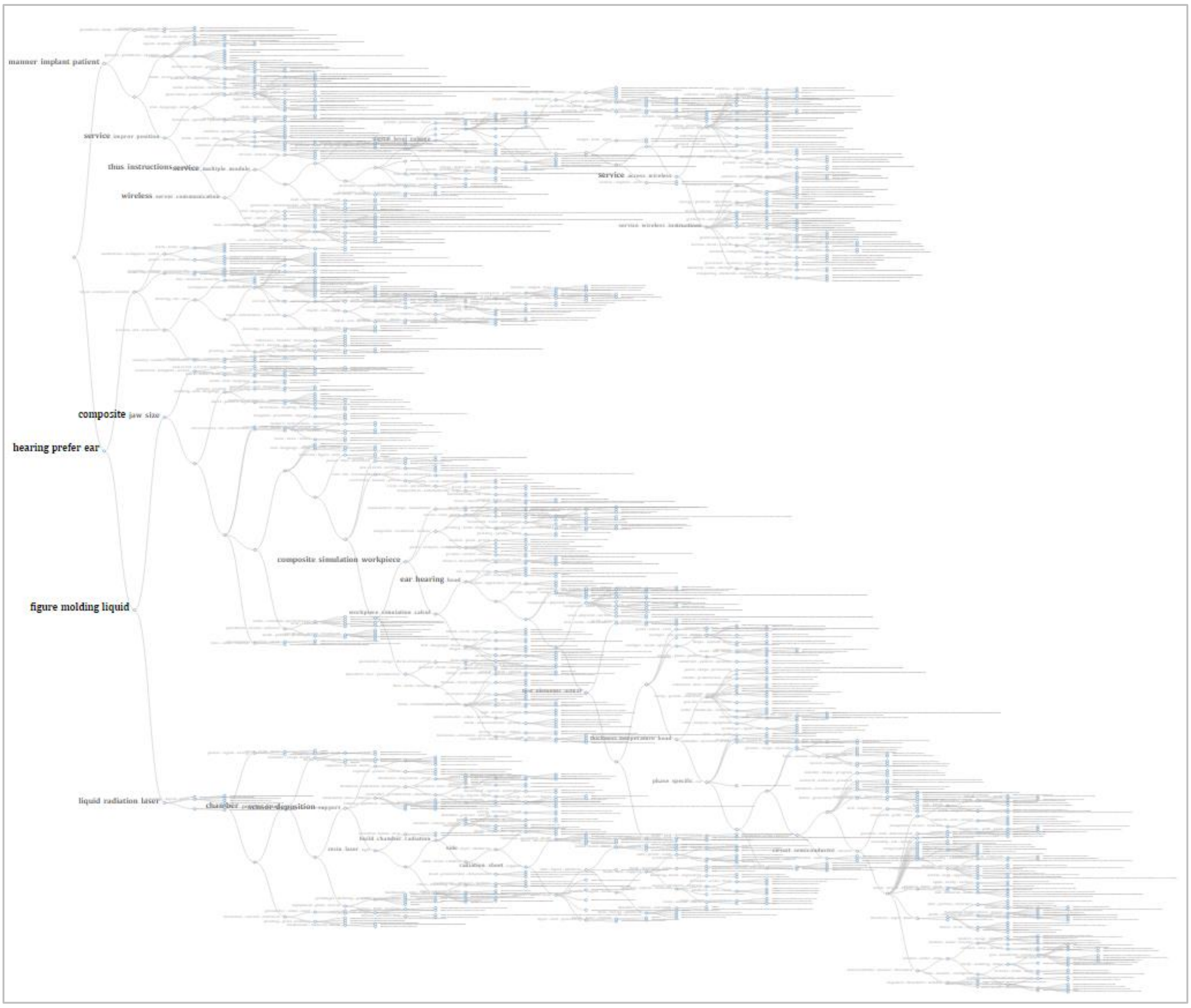


Figure 7.14 – Visualization of a selected subset of the additive manufacturing set

7.3. ANALYSIS OF APPLICABILITY FOR THE DECISION MAKING SETTING

Having performed the evaluation of the design parameter groups as well as a comparative case study into the field of additive manufacturing, this section will present the conclusions that can be drawn from this. This will be done by first mapping the evaluation results onto the axiomatic design to determine which questions can be answered in Section 7.3.1. Secondly, Section 7.3.2 presents the conclusions to be drawn from the comparative case study. Finally Section 7.3.3 will present the scope under which this evaluation are valid.

7.3.1. CONCLUSIONS BASED ON DESIGN PARAMETER GROUP EVALUATION

Having evaluated the design parameters per group block, the created design can now be used to trace back through the mappings and see to which customer needs the developed prototype complies. Table 7.2 presents the design parameters and the created group blocks as presented in Table 4.9. However, in this overview the traced back customer needs are added with marks indicating which process variables need to be met to be able to satisfy the specific customer needs. Those process variables that are met are represented in green and those that are not included in the developed prototype or don't meet the designed process variables, are represented in orange.

Table 7.2 – Trace back of implementation to customer needs

CA 1.	CA 2.	CA 3.	CA 4.	CA 5.	CA 6.	CA 7.	PV 1.	PV 2.	PV 3.	PV 4.	PV 5.	PV 6.	PV 7.	PV 8.	PV 9.	PV 10.	PV 11.	PV 12.	PV 13.	PV 14.	PV 15.	PV 16.	PV 17.	PV 18.	PV 19.	PV 20.	PV 21.	PV 22.	PV 23.	PV 24.	PV 25.	PV 26.	PV 27.	PV 28.	PV 29.	PV 30.	PV 31.
x	x	x	x	x	x	x	X																													
x	x	x	x	x	x	x		X																												
x	x	x	x	x	x	x			X																											
x	x	x	x	x	x	x				X																										
x	x	x	x	x	x	x					X																									
x	x	x	x	x	x	x						X																								
				x											X																							
				x												X																						
				x													X																					
				x														X																				
x	x	x	x	x	x	x											X																				
x	x	x	x	x	x	x												X																			
x	x	x	x	x	x	x													X																		
x	x	x	x	x	x	x														X																	
x	x	x	x	x	x	x															X																
			x																					X																	
			x																						X																
						x																				X															
						x																					X														
x	x	x	x	x	x	x																				X											
x	x	x	x	x	x	x																					X										
x	x	x	x	x	x	x																						X									
x	x	x	x	x	x	x																							X								
x	x	x	x	x	x	x																								X							
	x	x	x	x	x																												X						
		x	x	x																														X					
					x																																X				
					x																																	X			
			x																																				X		
			x																																					X	
			x																																						X

Table 7.3 – Customer needs codes and descriptions

CA	Description
CA 1.	How is the technological field structured?
CA 2.	What is in the portfolio of a specific company?
CA 3.	Who are competitors in certain technological areas?
CA 4.	What are the risks of developing a certain technology?
CA 5.	Is there an applicable solution for a specific problem?
CA 6.	Does the portfolio comply with strategy policy?
CA 7.	Where is the technological field expectedly heading?

For the process variables that are currently not represented in the prototype, an overview has been created in Table 7.4. This overview presents the specific process variable with expected costs (implementation time) and the perceived benefits that the addition of this specific PV has based on the design and evaluation. These estimations are made based on the functional implementation of the process variable at hand, not taking into account the optimization for user dependent needs e.g. a Graphical User Interface (GUI) shell on top of the text-mining and conceptual clustering software.

Table 7.4 – Expected costs and benefits of implementing designed elements not represented in prototype

To comply		Functionality	T.M.	C.C.	Vis.	New	Don't include	Expected investment	Expected return	Requires
CA 5.	PV 7.	SAO structure attributize	X					60 – 80 h	• Improved clustering • Enabling TRIZ functionalities	
	PV 8.	Enter functionality search		X				1 h	• Enabling TRIZ functionalities	PV 7.
	PV 9.	SAO similarity		X				60 – 80 h	• Enabling TRIZ functionalities	PV 7., PV 8.
	PV 10.	Store SAO similarity		X				1 h	• Enabling TRIZ functionalities	PV 7., PV 8., PV 9.
CA 7.	PV 18.	Change log		X				16 h	• Enabling Trend analysis	
	PV 19.	Visualize evolution				X		40 – 60 h	• Enabling Trend analysis	PV 19.
CA 6.	PV 27.	Highlight comp specific tech areas					X			
	PV 28.	Determine strategy positioning					X			
CA 4.	PV 16.	Patent similarity		X				24 h	• Allow for FTO	
	PV 17.	Store patent similarity		X				1 h	• Allow for FTO	PV 16.
	PV 29.	Patent links			X			20 h	• Allow for FTO	
	PV 30.	Patent citations			X			20 h	• Allow for FTO	
	PV 31.	Company database connection			X			40 h	• Allow for FTO • Search for (cross-) licensing	

T.M.: Text-mining software C.C.: Conceptual clustering software Vis.: Visualization

Based on the presented information in this chapter, conclusions about the developed prototype regarding the customer needs and the potential additions are:

- The developed prototype complies with the needs to fulfil, answering; how the technological field is structured, what is in the portfolio of a certain company, and who competitors are in certain technological areas.
- Additions can be made to answer what the risks are to develop a certain technology. The presented solutions would make useful additions to the already existing functionalities and can make full use of the developed structural relations as well as of the patent individual visualization options.
- To find an applicable solution to a specific problem, additions can be made. However as mentioned in Section 7.1.2 it needs to be clear to the user that the used dataset will have a large influence on the result of this search. This has influence on the extent to which this method can provide an addition to the answering of this customer need.
- As explained in Section 7.1.6 the addition to answer the question whether the company portfolio complies with the company's set strategy, is not recommended to be added. This is discouraged as it requires a completely different setup in both the analysis as well as the visualization segments.
- To answer where the technological field is expectedly heading the suggested additions are advised to be implemented as it requires a minimal addition to the conceptual clustering software, and the to be created visualization is expected to provide extra insight on the resulting hierarchy.

Taking into account the presented conclusions regarding the customer needs and the Strategic Technology Planning as presented in Figure 4.1, multiple conclusions can be drawn regarding the applicability to enhance the current STP process.

The developed prototype can in the technology research phase be applied for Exploration, Competitor analysis and Portfolio analysis. In the Technology Road-mapping and Freedom to Operate stages it can partially be used for competitor analysis but would benefit from the addition of the possibility to access an inter-company relation database. The same holds for the Exploration analysis in the Technology road-mapping stage.

When the proposed extensions are included, the additional application for the tool in the Technology research phase is that it allows for TRIZ analysis capabilities (however with the mentioned limitations). In the Freedom to Operate stage the additions to the tool will allow for the use of the Similarity detection, Competitor analysis as well as the Patent Quality Identification. Based on the fact that one can examine the field down to the individual patents, which can be individually examined, promises to provide additional insights to this process. In the Technology road-mapping stage the proposed additions will add the possibility to provide TRIZ analysis capabilities (here too the mentioned limitations hold true).

7.3.2. CONCLUSIONS BASED ON THE CASE OF ADDITIVE MANUFACTURING

Regarding the case study as presented in Section 7.2, multiple additional conclusions can be drawn about the conceptual clustering technique in combinational use with the self-organizing mapping technique:

- The high level arising optima are equal in the two presented methods. However, not all the inter-relations match. This results in more confidence in those conceptual relations that are similar and which requires the user to investigate why the other optima have different inter-relations. Enhancing the insight into the technical field.
- The conceptual clustering method shows more groupings of domains within a branch instead of focusing on the technical applications. This is due to the method of operation of the technique which rewards mutual similarities. This can offer a different perspective on the technological field than the technical application focus.
- Multiple representations can be created based on the same patent set using different patent fields as well as decision criteria, based on the specific user' need for insight. This can assist the versatility of different users as technology scouts or patent attorneys.
- The number of attributes has influence on the interpretability of the visualization. This is due to the fact that the categorical utility function which the conceptual clustering technique uses, rewards all attributes equally. Meaning that when more attributes are added, the individual influence becomes less. It is therefore advised to use a high level clustering first to detect the main clusters. Then select a subset (a specific branch) and perform a new round of attributizing and clustering on this specific subset of patents, in order to more specifically unveil its sub-branches.
- Using both tools simultaneously on the same data set, fields of interest can be detected in one and be further investigated in the other. The conceptual clustering technique can be used for its explanatory value, due to the storing of decision relevant information in its node hierarchy, resulting in a powerful combination in explanatory value.

7.3.3. SCOPE OF THE EVALUATION

The presented conclusions regarding the complying of the prototype with the proposed design as well as those for the potential use within the strategic technology planning process, are when considering the evaluation methods as presented in Section 2.2, valid for:

- In respect to the three faceted approach the prototype is technically as well as partly subjectively evaluated. The completion of the subjective evaluation, as well as the performing of the empirical evaluation can take place once the DSS is turned into a full-fledged DSS tuned to the organizational processes.

- In respect to the Sequential approach, the prototype is evaluated for the identification of criteria, for the formative evaluation and partly for the evaluation of the system outcome. The prototype has not had a summative evaluation, as this requires a full integration into a specific organizational process.
- In respect to the general approach to DSS evaluation, the prototype is not evaluated as this requires a full implementation into the organizational process, and the focus is completely on the user.

Concluding it can be said that these evaluation conclusions are technically valid. Based on a specific company's needs and wishes, the prototype can be further developed. Meaning that the present functionalities that are technically or formatively valid, should be adapted to the specific wishes of the users and their process needs, to be able to fully be subjectively or summatively evaluated as well.

— There is no real ending. It's just the place where you stop the story.

Frank Herbert

8

Conclusions & Recommendations

8.1. Conclusions

This research has focused on the novel application of representing the technological field based on patents by exploiting the use of the conceptual clustering technique. With the goal to evaluate its potential applicable use within the Strategic Technology Planning (STP) process in technological R&D companies. Chapter 2 has shown that Decision Support Systems (DSSs) take an important role in this process, due to the vast amount of complex information that needs to be processed by decision makers. They are a human-machine system in which the machine part deals with the structuring and processing of data in order to present the human with a representation on which he can make decisions. The ultimate goal of DSSs is the improvement of the efficiency of the decision making as well as the quality of the resulting decision itself. The ever present need for more insight and risk reduction, alongside a methodological interest in the application of conceptual clustering has formed the base for the research into the extension of the Decision Support System capabilities. For which the research question has been stated as: For which aspect and in what manner could conceptual clustering enhance the current patent analysis process for Strategic Technology Planning in technical R&D companies?

Insight into the hierarchical structure of technology as presented by the field of data-science, together with an explanation of the role of patents and their structure in Chapter 3, formed the introduction into the analysis of the STP process. This analysis, as presented in Section 4.2, has shown the multiple possible uses of patent analysis within the multiple phases of the STP process and what their functional needs are. The following analysis of currently applied solutions in Section 4.4, has shown that improvements can be made on the aspects of explainability, reproducibility and representation of the hierarchical aspect of technology. These aspects support the choice to apply conceptual clustering, as it was expected that this method could deliver a potential solution to them. The functional needs for patent analysis within the STP process have been combined and captured in 7 core questions which users seek answers for, as shown in Table 4.5. These core questions have been defined based on an extensive literature review as well as on expert inputs.

To gain insight into the different needs for patent analysis within the STP process, as well as to provide a source for evaluation, a design has been made for the Decision Support System by applying the Axiomatic Design method as introduced in Section 4.5. The inputs for this design are customer needs which are the aforementioned 7 core questions for patent analysis in the STP process. The application of Axiomatic Design in the development of a decision support system for technological R&D has been proven to be a useful one. As the method forces the designer to think through all the steps to be able to end with a complete design. The aspects actually relevant to the eventual users are guaranteed to be included, without the addition of unnecessary additions. Additionally the zig-zagging between domains guarantees a complete design, which is efficient in time when compared to trial-and-error type design implementations. The design process results in the created design, from which the Decision Parameter Grouping, as presented in Table 4.11, can be used for implementation and evaluation purposes.

The conceptual clustering technique has been explained in Chapter 5, and its application to patents has been identified as a new application branch under the visualization approaches, in the taxonomy of patent analysis approaches in Section 5.4.

To assess whether the application of the conceptual clustering to patents is possible, a prototype Decision Support System has been developed, which has been presented in Chapter 6. The implementation of the conceptual clustering functionality is done based on the COBWEB algorithm as proposed by Fisher. The developed prototype entails two software products written in Python, which are respectively a text-mining and a conceptual clustering program. These software tools are combined with a developed online visualization tool based on the D3 JavaScript library. The online visualization tool, as described in Section 6.3, has been extended with interactivity to enable the user to interact with the data. In addition a node attribute representation method has been developed to show the user the relative important attributes that define the nodes' concept. By the use of a technical verification, presented in Section 6.4, it has been proven that the application of conceptual clustering to patents is indeed possible and operates according to theory.

To evaluate for which aspect in the STP process the developed tool can be used to enhance the currently available patent analysis tools, the Design Parameter Groups as presented in Table 4.11, are used. Through evaluation based on the design, summarized in Table 7.2, the prototype has been proven to be technically usable in Exploration analysis, Competitor analysis and Portfolio analysis. The highest added value to the R&D division's toolkit is perceived to lay in the explainability of the outcome, the accessibility through the online visual representation and the possibility of using the decision support system in an interactive way. Allowing the technology scouts and managers to interact with the data, discussing and continuing exploration based on newly gathered insights. Furthermore implementable solutions are presented in a cost-benefit estimation, presented in Table 7.4, allowing to extend the developed prototype's applicability for use to the inclusion of Freedom to operate analysis capabilities, trend analysis and to a limited extend inventive problem solving analysis capabilities. The current prototype can now answer 3 out of the 7 core questions, however it must be noticed that the weights of these questions implementation needs are the highest, as can be derived from Table 4.5. Meaning that the addition of functionality to answer the other questions are far lower than those that have been implemented already.

To gain further insights for evaluation, in Section 7.2, a complete selection of patents related to additive manufacturing (3D printing) retrieved from Thomson Reuters' Derwent patent database is used. This patent set (containing 9360 patents) has been parsed through-, and visualized by-, the developed prototype, showing a representation of this technological fields inter-patent relational structure. The outcome has been comparatively used alongside a self-organizing map of the same patent set. Further conclusions based on this analysis are: (1) the fact that the same high level optima arise among the two methods, however not all inter-relations match. Forcing the user to further investigate, increasing robustness of knowledge obtained. (2) The conceptual clustering shows more application domains in addition to the technical concepts. (3) Multiple representations can be made of the same set based on user needs. (4) There is an influence of the attribute number on the outputs interpretability, for which it is suggested to perform a high level search first. Then select a subset of patents for reprocessing and visualization, in order to obtain more insight into sub-branching. (5) The combinational use of the two methods enhances the insight, due to the explanatory value of the nodes in the conceptual clustering output.

As shown in Section 7.3.3, the presented conclusions regarding the complying of the prototype with the proposed design as well as those for the potential use within the strategic technology planning process, are valid for: (1) The technical facet and partly the subjective facet when considering the three faceted approach for evaluation of DSSs. (2) The identification of criteria as well as formatively when considering the sequential approach for evaluation of DSSs.

8.2. CONTRIBUTIONS AND LIMITATIONS

Regarding the patent analysis' theoretical field, it has been shown that the application of conceptual clustering is a novel application. Which should be placed in a new category in the visualization methods when adapting the taxonomy of patent analysis as presented by Abbas, Zhang and Khan (Abbas et al., 2014). Discussing the theory of conceptual clustering in Chapter 5, it has been shown that this method is fully explanatory as well as mutually exclusive. Which are some of the main critiques on the currently used methods in the patent analysis field which are lacking this. It therefore offers a self-explanatory alternative to the currently available black box models.

Furthermore empirical study has shown that the conceptual clustering technique seems to show more application field information in addition to the grouping of technical concepts, when compared to the conventional analysis.

A limitation regarding the theoretical aspects of patent analysis, is that no exclusive answer has been given on the aspect of the hierarchical aspect of technology being present. Which is presented as a critique towards the existing methods and was one of the interests triggering the research into the application of conceptual clustering to patents. As besides a visual analysis, in which it seems that this is not present, no statistical test has been performed to verify this within the existing prototype outputs.

Considering the theoretical fields of Strategic Technology Planning and Decision Support Tools, contributions have been made by summarizing the high level needs for which patent analysis is currently applied. This is done by performing an extensive literature study as well as the use of expert inputs. These needs have been converted into seven core questions for which answers are sought within the Strategic Technology Planning process. Based on these core questions a design has been created for a Decision Support System, using the Axiomatic Design method, which can be used for evaluation of implementations.

Looking at the practical field of Strategic Technology Planning, a prototype of applying conceptual clustering to a patent set has been created. The created prototype has been technically and formatively evaluated for use as a Decision Support System. Dependent on the needs is can be adjusted to fit into a specific company's organizational processes. The added value to the process has been illustrated by the use of a comparative case study of additive manufacturing, using the self-organizing maps and conceptual clustering techniques. This has shown added value in the detection of similar grouping of optima and the forcing of the user to question the difference in inter-relations, enhancing the robustness of the final knowledge. The option for making multiple representations of the same dataset based on different patent fields, allows multiple types of users to interact with the same dataset, retrieving more relevant knowledge. Using multiple methods simultaneously on the same patent set, has proven to provide more insight into the relevant groupings of concepts and the inter-relations of these groups, by exploiting the explanatory value of the conceptual clustering technique.

A limitation to the backing of the evaluation data is the small sample of experts interviewed for input. This is due to the exploratory nature of the thesis. As it is was set out to discover firstly whether the method of conceptual clustering could be applied to patents at all, and for which aspects of patent analysis within the Strategic Technology Planning it might prove to be useful based on carefully defined criteria. The value of expert insight would however be higher when a full implementation has been done. As, up till this point the research has shown that it is technically feasible to comply with 6 out of the 7 questions. However further development in usability needs to be done before the experts opinions on the achieving of the goal of making better and more efficient decisions can really be answered. Due to time limits, the step towards this full implementation and its successive usability testing, was infeasible for this thesis.

8.3. RECOMMENDATIONS

Recommendations based on the research done can be divided into improvements of the developed decision support system, additions to extend its possible applications of use, improving its usability, and further research on the theory of the notion of similarity.

The improvements of the developed Decision Support System focus on improving the ability of the conceptual clustering software to extract the inter-patent relations. The current applied version of the COBWEB algorithm is that based on the four operators as described by Fisher. In this version when a concept is split, all its children are being promoted. To reduce the influence of the promoting of those siblings who are better off staying clustered together Fisher describes the addition of a fifth operator (Fisher, 1987). This fifth operator only promotes the best host, keeping the other siblings clustered together. The current implementation of the conceptual clustering software allows for the addition of this operator with minimal effort. And is therefore suggested to be applied first, when improvement of the conceptual clustering is searched for. Another possible improvement could be the use of weighted attributes instead of the Boolean attributes that are present in the current implementation. However, when an improvement is searched for in the adjustment of the attributes, the advice is given to focus on the implementation of semantic analysis in the form of Subject-Action-Object (SAO) structures. Which are proven to better capture the technical innovations embedded in patents. This can be combined with a synonym referencing, to eliminate the influence of wording or writing style. This method of implementation has the benefit of not having to adjust the conceptual clustering software as the attributizing can be done based on whether or not a patent contains a certain SAO structure. Additionally, the attributization method could also be considered for improvements, as in the current version only single words are used as attributes. Combinations of words or parts of sentences could also be considered to be used, for which Cunningham and Kwakkel present an implementable solution in their book (Cunningham & Kwakkel, 2015).

The proposed additions, as presented in Table 7.4, to extend the possible applications of use of the developed prototype, can be implemented according to preference, need or interest. For each of the three questions for which additional functionality could potentially be added, the needed set of elements to comply is presented. Of these three, the addition of the elements to answer the question what the risks and benefits are of developing a certain technology is suggested to be added first. As this is expected to provide the most added insights. Secondly the addition of the elements enabling the study of trends are suggested to be added. Lastly, the addition to find a generic problem for a specific problem. Which on its own is expected not to be of the level of added insight as the previous two. However, by adding its elements, the previously mentioned attributization of SAO structures becomes available as well. Thus, inherently improving the resulting conceptual clustering of the inter-patent relationship hierarchy.

Regarding usability there are potential improvements that can be made on the developed prototype. As this first implementation was aimed on exploring the potential usage there was less focus on the usability by all the potential users. To optimize this, a self-explanatory graphical user interface can be created as a software shell, assisting the users to set the right parameters and assisting them in the execution of the software. Furthermore, research can be done into the visual representation of the explainability of the hierarchical structure, which is expected to improve the interpretability of the representation. This can be done by making use of a simplified representation by exploiting tree reduction techniques. Suggested approaches are for example applying rough set theory for attribute reduction, which aims to reduce all redundant objects and attributes to find the minimum subset of attributes (Hassanien, 2004), or by exploiting the use of a tree pruning algorithm, which aims at combining similar branches (Patel & Upadhyay, 2012). After adding usability elements as well as the DSS integration into the specific organizational processes. The application of the general approach to DSS evaluation for user side evaluation is suggested. To evaluate the extent of the achieving of the DSS goals of more efficient decision making and the enhancing of the decisions made, it is suggested to perform a summative evaluation or apply the general approach to DSS evaluation, as introduced in Section 2.2.

In addition, to get a better insight into the interchangeability as well as the impact of similarity of technologies represented by patents in hierarchical representations. Further research is advised to be done into semantic similarity detection. For which the work of Bergmann (Bergmann et al., 2008) and Park (Park et al., 2011), are expected to provide a useful starting path.

8.4. REFLECTION

Criticizing on the application of Axiomatic Design, the creativity aspect must be put in perspective. As Axiomatic Design concentrates more on the definition of the problem and making sure that the final solution complies with the axioms and is implementable, than on the actual proposing of solutions to resolve issues in situations of coupled or complex designs (Ogot, 2011). A second critique is that the information axiom tells the designer to use the solution which possesses the highest probability of success and the lowest number of elements. However, due to notion of bounded rationality one cannot always foresee all possible solutions and their expected repercussions and thus one cannot always guarantee to comply with this axiom. Third, due to the high level requirements that are the result of the Axiomatic Design the evaluation of the result can come out to be pictured more negatively. Nonetheless, the axiomatic design method does forces the designer to think through the mapping between all of the domains, making sure that at least all relevant aspects of the chosen design elements are covered. Thus, when applied, an effective and efficient realization of customer needs can be obtained.

However advised to further evaluate the Decision Support System on its usability, there are noticeable issues with DSS evaluations. Since the decisions for which these DSSs are used, are of such high complexity. The measurements of the increase of efficiency of the decision making process as well as the resulting quality of the decision, could easily be influenced by alternative causes than the DSS (Alavi & Joachimsthaler, 1992; Rhee & Rao, 2008). It has been claimed that the effect of reliance, when a DSS is present, improves decision quality (Barr & Sharda, 1997). Furthermore, Khazanchi pointed out the paradox of the evaluation of the correctness of the presented solutions by a DSS (Khazanchi, 1991). As, the semi-structured issues can be judged as good, bad or reasonable, but not as right or wrong (Mason & Mitroff, 1973), which is up to the subjective opinion of the expert. This is why indirect quality measures of the decisions are done by applying technical evaluations (Rhee & Rao, 2008).

Considering the research performed and looking back at the applied approach. It must be said that it would have been wiser to put in even more time resources in the literature review and expert input on the specific needs, to have this clear even sooner. I have found myself working on a first implementation of the prototype in parallel to this process, which resulted in numerous revisions. This might be considered the downside of the applied iterative approach. But that said, it did gave a good insight into the potential applicability, which in turn resulted in multiple leads for literature searches.

Bibliography

Abbas, A., Zhang, L., & Khan, S. U. (2014). A literature review on the state-of-the-art in patent analysis. *World Patent Information*, *37*, 3–13. doi:10.1016/j.wpi.2013.12.006

Adam, F., Fahy, M., & Murphy, C. (1998). A framework for the classification of DSS usage across organizations. *Decision Support Systems*, *22*(1), 1–13. doi:10.1016/S0167-9236(97)00039-0

Adelman, L. (1992). *Evaluating decision support and expert systems*. New York: John Wiley & Sons, Inc.

Alavi, M., & Joachimsthaler, E. A. (1992). Revisiting DSS Implementation Research: A Meta-Analysis of the Literature and Suggestions for Researchers. *MIS Quarterly*, *16*(1), 95–116. Retrieved from http://www.jstor.org/stable/249703

Allison, J. R., Lemley, M. A., Moore, K. A., & Trunkey, R. D. (2004). Valuable patents. *Georgetown Law Journal*, *92*, 435–479.

Arthur, W. B. (2007). The structure of invention. *Research Policy*, *36*(2), 274–287. doi:10.1016/j.respol.2006.11.005

Barr, S. H., & Sharda, R. (1997). Effectiveness of decision support systems: development or reliance effect? *Decision Support Systems*, *21*(2), 133–146. doi:10.1016/S0167-9236(97)00021-3

Barry, B. K., Domb, E., & Slocum, M. S. (2008). TRIZ - What Is TRIZ ? Retrieved from http://www.triz-journal.com/triz-what-is-triz/

Bergmann, I., Butzke, D., Walter, L., Fuerste, J. P., Moehrle, M. G., & Erdmann, V. A. (2008). Evaluating the risk of patent infringement by means of semantic patent analysis: The case of DNA chips. *R&D Management*, *38*(5), 550–562.

Blockeel, H., Bruynooghe, M., Dzeroski, S., Ramon, J., & Struyf, J. (2002). Hierarchical multi-classification. In *Proceedings of the ACM SIGKDD Workshop on Multi- Relational Data Mining* (pp. 21–35).

Bodenhausen, G. H. C. (1986). *Paris Convention for the Protection of Industrial Property*. *Koln I971*.

Bonczek, R. H., Holsapple, C. W., & Whinston, A. B. (2014). *Foundations of decision support systems*. Academic Press.

Bonino, D., Ciaramella, A., & Corno, F. (2010). Review of the state-of-the-art in patent information and forthcoming evolutions in intelligent patent informatics. *World Patent Information*, *32*(1), 30–38. doi:10.1016/j.wpi.2009.05.008

Bostock, M. (2015). D3.js. Retrieved from http://d3js.org/

Cambridge Dictionaries Online. (2015). Axiomatic definition, meaning. Retrieved May 29, 2015, from http://dictionary.cambridge.org/dictionary/british/axiomatic

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81.

Campbell, R. S. (1983). Patent Trends as a Technological Forecasting Tool. *World Patent Information*, *5*(3), 137–143.

Cascini, G., & Zini, M. (2008). Measuring patent similarity by comparing inventions functional trees. In *Computer-aided innovation (CAI)* (pp. 31–42). Retrieved from http://link.springer.com/chapter/10.1007/978-0-387-09697-1_3

Cha, S.-H. (2007). Comprehensive survey on distance/similarity measures between probability density functions. *International Journal of Mathematical Models and Methods in Applied Sciences*, *1*(4).

Chandrasekaran, B. (2005). *From Optimal to Robust COAs : Challenges in Providing Integrated Decision Support for Simulation-Based COA Planning A White Paper*. Columbus.

Chang, P.-L., Wu, C.-C., & Leu, H.-J. (2009). Using patent analyses to monitor the technological trends in an emerging field of technology: a case of carbon nanotube field emission display. *Scientometrics*, *82*(1), 5–19. doi:10.1007/s11192-009-0033-y

Corter, J., & Gluck, M. (1985). Machine Generalization and Human Categorization: An Information-Theoretic View. In *Proceedings of the First Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-85)* (pp. 201–207). Corvallis, Oregon: AUAI Press.

Courtney, J. F., & Paradice, D. B. (1993). Studies in managerial problem formulation systems. *Decision Support Systems*, *9*(4), 413–423. doi:10.1016/0167-9236(93)90050-D

Cunningham, S. W., & Kwakkel, J. H. (2015). *Analytics and Tech Mining for Engineering Managers*. New York: Momentum Press.

Cyert, R. M., & March, J. G. (1963). A behavioral theory of the firm. *Englewood Cliffs, NJ*, *2*.

De Bruijn, J. A., & ten Heuvelhof, E. F. (2008). *Management in Networks: On multi-actor decision making*. Routledge.

De Rassenfosse, G., & van Pottelsberghe de la Potterie, B. (2009). A policy insight into the R&D–patent relationship. *Research Policy*, *38*(5), 779–792. doi:10.1016/j.respol.2008.12.013

Dunn, W. N. (1994). *Public Policy Analysis: an introduction*. Prentice Hall, Inc.

ECMA International. (2013). *Standard ECMA-404 - The JSON Data Interchange Format*. Geneva.

Enserink, B., Hermans, L., Kwakkel, J., Thissen, W., Koppenjan, J., & Bots, P. (2010). *Policy analysis of multi-actor systems*. Lemma The Hague.

Ernst, H. (2003). Patent information for strategic technology management. *World Patent Information*, *25*(3), 233–242. doi:10.1016/S0172-2190(03)00077-2

European Patent Office. (2014). *Open Patent Services RESTful Web Services Reference Guide*. Retrieved from http://documents.epo.org/projects/babylon/eponet.nsf/0/7AF8F1D2B36F3056C1257C04002E0AD6/$File/OPS_v3.1_documentation_version_1.2.14_en.pdf

European Patent Office. (2015). Patent Information Tour V 1.3. Retrieved from http://application.epo.org/wbt/pi-tour/

Fantoni, G., Apreda, R., Dell'Orletta, F., & Monge, M. (2013). Automatic extraction of function–behaviour–state information from patents. *Advanced Engineering Informatics*, *27*(3), 317–334. doi:10.1016/j.aei.2013.04.004

Firat, A. K., Woon, W. L., & Madnick, S. (2008). *Technological Forecasting – A Review*. Cambridge, United States. Retrieved from http://web.mit.edu/smadnick/www/wp/2008-15.pdf

Fisher, D. H. (1987). Knowledge acquisition via incremental conceptual clustering. *Machine Learning*, *2*(2), 139–172.

Gallini, N. T. (1984). Deterrence by Market Sharing : A Strategic Incentive for Licensing. *The American Economic Review*, *74*(5), 931–941.

Gediga, G., Hamborg, K.-C., & Düntsch, I. (1999). The IsoMetrics usability inventory: An operationalization of ISO 9241-10 supporting summative and formative evaluation of software systems. *Behaviour & Information Technology*, *18*(3), 151–164. doi:10.1080/014492999119057

Gorry, G. a, & Morton, M. S. S. (1971). A framework for management information systems. *Sloan Management Review*, *13*, 55–70. doi:10.1016/S0167-

Hassanien, a E. (2004). Rough set approach for attribute reduction and rule generation: A case of patients with suspected breast cancer. *Journal of the American Society for Information Science and Technology*, *55*(11), 954–962. doi:10.1002/asi.20042

Hegde, D., & Sampat, B. (2009). Examiner citations, applicant citations, and the private value of patents. *Economics Letters*, *105*(3), 287–289. doi:10.1016/j.econlet.2009.08.019

Jaffe, A. B., & Trajtenberg, M. (2002). *Patents, citations, and innovations: A window on the knowledge economy*. MIT press.

JetBrains s.r.o. (2015). Python IDE & Django IDE for Web developers : JetBrains PyCharm. Retrieved from https://www.jetbrains.com/pycharm/

Kasravi, K., & Risov, M. (2009). Multivariate Patent Similarity Detection. In *Proceedings of the 42nd Hawaii International Conference on System Sciences* (pp. 1–8). Washington: IEEE Computer Society. doi:10.1109/HICSS.2009.318

Keen, P. G. W. (1987). Decision support systems: the next decade. *Decision Support Systems*, *3*, 253–265. doi:10.1016/0167-9236(87)90180-1

Khazanchi, D. (1991). Evaluating decision support systems: a dialectical perspective. *Proceedings of the Twenty-Fourth Annual Hawaii International Conference on System Sciences*, *iii*(July). doi:10.1109/HICSS.1991.184131

Kim, Y. G., Suh, J. H., & Park, S. C. (2008). Visualization of patent analysis for emerging technology. *Expert Systems with Applications*, *34*(3), 1804–1812. doi:10.1016/j.eswa.2007.01.033

Kirakowski, J., & Corbett, M. (1990). *Effective Methodology for the Study of Hci (Human Factors in Information Technology)*. Amsterdam: North-Holland.

Knoke, D., & Kuklinski, J. (1982). *Network analysis* (Vol. 07–028). Beverly Hills, CA: Sage Publications, Inc.

Kohonen, T. (1982). Self-Organized Formation of Topologically Correct Feature Maps. *Biological Cybernetics*, *43*, 59–69.

Kolodner, J. L. (1983). Reconstructive Memory: A Computer Model. *Cognitive Science*, *7*(4), 281–328. doi:10.1207/s15516709cog0704_2

Kulak, O., Cebi, S., & Kahraman, C. (2010). Applications of axiomatic design principles: A literature review. *Expert Systems with Applications*, *37*(9), 6705–6717. doi:10.1016/j.eswa.2010.03.061

Lamy, J., Ellini, A., Nobécourt, J., Venot, A., & Zucker, J. (2010). Testing Methods for Decision Support Systems. In C. S. Jao (Ed.), *Decision Support Systems* (pp. 87–98). Rijeka: InTech.

Lanjouw, J. O., & Schankerman, M. (2001). Characteristics of patent litigation : a window on competition. *RAND Journal of Economics*, *32*(1), 129–151.

Lebowitz, M. (1987). Experiments with Incremental Concept Formation: UNIMEM. *Machine Learning*, *2*(2), 103–138. doi:10.1023/A:1022800624210

Lee, C., Song, B., & Park, Y. (2013). How to assess patent infringement risks: a semantic patent claim analysis using dependency relationships. *Technology Analysis & Strategic Management*, *25*(1), 23–38. doi:10.1080/09537325.2012.748893

Lee, S., Yoon, B., & Park, Y. (2009). An approach to discovering new technology opportunities: Keyword-based patent map approach. *Technovation*, *29*, 481–497. doi:10.1016/j.technovation.2008.10.006

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics* (pp. 281–297). Berkeley, CA: University of California Press. Retrieved from http://projecteuclid.org/euclid.bsmsp/1200512992

March, J. G., & Simon, H. A. (1958). *Organizations*. Wiley.

Marsden, P. V., & Laumann, E. O. (1984). Mathematical ideas in social structural analysis. *Journal of Mathematical Sociology*, *10*, 271–294.

Martinsons, M. G., & Davison, R. M. (2007). Strategic decision making and support systems: Comparing American, Japanese and Chinese management. *Decision Support Systems*, *43*(1), 284–300. doi:10.1016/j.dss.2006.10.005

Mason, R. O., & Mitroff, I. I. (1973). A Program for Research on Management Information Systems. *Management Science*, *19*(5), 475–487. Retrieved from http://www.jstor.org/stable/10.2307/2629445

Maynard, S., Burstein, F., & Arnott, D. (2001). A multi-faceted decision support system evaluation approach. *Journal of Decision Systems*, *10*(3-4), 395–428.

McNamee, R. C. (2013). Can't see the forest for the leaves: Similarity and distance measures for hierarchical taxonomies with a patent classification example. *Research Policy*, *42*(4), 855–873. doi:10.1016/j.respol.2013.01.006

McNurlin, B. C., Sprague, R. H., & Bui, T. X. (1989). *Information systems management in practice*. Prentice-Hall International.

Michalski, R. S. (1980). Knowledge acquisition through conceptual clustering: A theoretical framework and an algorithm for partitioning data into conjunctive concepts. *Journal of Policy Analysis and Information Systems*, *4*(3), 219–244.

Mitroff, I. I. (2008). Knowing: how we know is as important as what we know. *Journal of Business Strategy*, *29*(3), 13–22. doi:10.1108/02756660810873173

Mitroff, I. I., & Linstone, H. A. (1993). *The unbounded mind: Breaking the chains of traditional business thinking*. Oxford University Press.

Murmann, J. P., & Frenken, K. (2006). Toward a systematic framework for research on dominant designs, technological innovations, and industrial change. *Research Policy*, *35*(7), 925–952. doi:10.1016/j.respol.2006.04.011

NLTK Project. (2015). NLTK 3.0 documentation. Retrieved from http://www.nltk.org/

Noyons, E. (2001). Bibliometric mapping of science in a science policy context. *Scientometrics*, *50*(1), 83–98.

Ogot, M. (2011). Conceptual design using axiomatic design in a TRIZ framework. *Procedia Engineering*, *9*, 736–744. doi:10.1016/j.proeng.2011.03.163

Oxford University Press. (2015a). Hypernym - definition. Retrieved June 27, 2015, from http://www.oxforddictionaries.com/definition/english/hypernym

Oxford University Press. (2015b). Sinter - definition. Retrieved July 30, 2015, from http://www.oxforddictionaries.com/definition/english/sinter

Oxford University Press. (2015c). Superordinate - definition. Retrieved June 27, 2015, from http://www.oxforddictionaries.com/definition/english/superordinate

Paradice, D. B., & Courtney, J. F. (1989). Organizational knowledge management structure. *Information Resources Management Journal*, *2*(3), 1–14. Retrieved from http://proquest.umi.com/pqdweb?did=793598781&Fmt=7&clientId=20931&RQT=309&VName=PQD

Park, H., Ree, J. J., & Kim, K. (2013). Identification of promising patents for technology transfers using TRIZ evolution trends. *Expert Systems with Applications*, *40*(2), 736–743. doi:10.1016/j.eswa.2012.08.008

Park, H., Yoon, J., & Kim, K. (2011). Identifying patent infringement using SAO based semantic technological similarities. *Scientometrics*, *90*(2), 515–529. doi:10.1007/s11192-011-0522-7

Patel, N., & Upadhyay, S. (2012). Study of Various Decision Tree Pruning Methods with their Empirical Comparison in WEKA. *International Journal of Computer Applications*, *60*(12), 20–25.

Pearson, J. M., & Shim, J. P. (1995). An empirical investigation into DSS structures and environments. *Decision Support Systems*, *13*(2), 141–158. doi:10.1016/0167-9236(93)E0042-C

Porter, A. L., & Cunningham, S. W. (2004). Analytical families. In *Tech Mining: Exploiting New Technologies for Competitive Advantage* (pp. 173–180). Hoboken, New Jersey: John Wiley & Sons, Inc.

Princeton. (2014). What is WordNet? Retrieved from http://wordnet.princeton.edu/

Python Software Foundation. (2015). About Python™ | Python.org. Retrieved from https://www.python.org/about/

Rafols, I., Porter, A. L., & Leydesdorff, L. (2010). Science Overlay Maps : A New Tool for Research Policy. *Journal of the American Society for Information Science and Technology*, *61*(9), 1871–1887. doi:10.1002/asi.21368

Reingold, E. M., & Tilford, J. S. (1981). Tidier Drawings of Trees. *IEEE Transactions on Software Engineering*, *SE-7*(2), 223–228. doi:10.1109/TSE.1981.234519

Rhee, C., & Rao, H. R. (2008). Evaluation of Decision Support Systems. In Burstein, Frada, Holsapple, & Clyde (Eds.), *Handbook on Decision Support Systems 2* (1st ed., p. pp 313–327). Springer Berlin Heidelberg. doi:10.1007/978-3-540-48716-6_15

Rip, A. (1997). Qualitative conditions of scientometrics: The new challenges. *Scientometrics*, *38*(1), 7–26. doi:10.1007/BF02461120

Schickel-zuber, V., & Faltings, B. (n.d.). OSS : A Semantic Similarity Function based on Hierarchical Ontologies, 551–556.

Schröder, O. (2013). *European patents and the grant procedure*. Munich, Germany.

Segev, A., & Kantola, J. (2012). Identification of trends from patents using self-organizing maps. *Expert Systems with Applications*, *39*(18), 13235–13242. doi:10.1016/j.eswa.2012.05.078

Shadish, W. R. (1993). Critical multiplism: A research strategy and its attendant tactics. *New Directions for Program Evaluation*, *1993*(60), 13–57. doi:10.1002/ev.1660

Shapiro, C. (2001). Navigating the Patent Thicket : Cross Licenses , Patent Pools , and Standard Setting. In A. B. Jaffe, J. Lerner, & S. Stern (Eds.), *Innovation Policy and the Economy* (Volume 1., Vol. 1, pp. 119–150). MIT Press.

Shepard, A. (1987). Licensing to enhance demand for new technologies. *The RAND Journal of Economics*, *18*(3), 360–368.

Shim, J. P., Warkentin, M., Courtney, J. F., Power, D. J., Sharda, R., & Carlsson, C. (2002). Past, present, and future of decision support technology. *Decision Support Systems*, *33*(2), 111–126. doi:10.1016/S0167-9236(01)00139-7

Shin, J., & Park, Y. (2005). Generation and Application of Patent Claim Map : Text Mining and Network Analysis. *Journal of Intellectual Property Rights*, *10*(May), 198–205.

Shirwaiker, R. a., & Okudan, G. E. (2011). Contributions of TRIZ and axiomatic design to leanness in design: An investigation. *Procedia Engineering*, *9*, 730–735. doi:10.1016/j.proeng.2011.03.162

Simon, H. A. (1982). *Models of bounded rationality: Empirically grounded economic reason* (Vol. 3.). MIT Press.

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, *24*(4), 265–269. doi:10.1002/asi.4630240406

Snow, R., Jurafsky, D., & Ng, A. Y. (2004). Learning syntactic patterns for automatic hypernym discovery. In L. K. Saul, Y. Weiss, & L. Bottou (Eds.), *Advances in Neural Information Processing Systems 17* (Vol. 17, pp. 1297–1304). Cambridge, MA: MIT Press. Retrieved from http://books.nips.cc/papers/files/nips17/NIPS2004_0887.pdf

Sternitzke, C., Bartkowski, A., & Schramm, R. (2008). Visualizing patent statistics by means of social network analysis tools. *World Patent Information*, *30*(2), 115–131. doi:10.1016/j.wpi.2007.08.003

Suh, N. P. (2001). *Axiomatic Design - Advances and Applications* (1st ed.). New York: Oxford University Press.

Thomson Reuters. (2008). Thomson Innovation ThemeScape Quick Tour. Retrieved June 27, 2015, from http://ip-science.thomsonreuters.com/winningmove/secure/TI_Themescape_QT.html

Trajtenberg, M., Henderson, R., & Jaffe, A. (1997). *University Versus Corporate Patents: A Window On The Basicness Of Invention*. *Economics of Innovation and New Technology* (Vol. 5). doi:10.1080/10438599700000006

Trappey, A. J. C., Trappey, C. V., Wu, C.-Y., Fan, C. Y., & Lin, Y.-L. (2013). Intelligent patent recommendation system for innovative design collaboration. *Journal of Network and Computer Applications*, *36*(6), 1441–1450. doi:10.1016/j.jnca.2013.02.035

Trappey, A. J. C., Trappey, C. V., Wu, C.-Y., & Lin, C.-W. (2012). A patent quality analysis for innovative technology and product development. *Advanced Engineering Informatics*, *26*(1), 26–34. doi:10.1016/j.aei.2011.06.005

Trippe, A. J. (2003). Patinformatics: Tasks to tools. *World Patent Information*, *25*(3), 211–221. doi:10.1016/S0172-2190(03)00079-6

Tseng, Y.-H., Lin, C.-J., & Lin, Y.-I. (2007). Text mining techniques for patent analysis. *Information Processing and Management*, *43*(5), 1216–1247. doi:10.1016/j.ipm.2006.11.011

United States Patent and Trademark Office. (2013). "Kind codes" Included on the USPTO Patent Documents. Retrieved from http://www.uspto.gov/learning-and-resources/support-centers/electronic-business-center/kind-codes-included-uspto-patent

United States Patent and Trademark Office. (2014). General Information Concerning Patents. Retrieved from http://www.uspto.gov/patents-getting-started/general-information-concerning-patents

Van de Riet, O. A. W. T. (2003). *Policy analysis in a multi-actor policy settings: navigating between negotiated nonsense & superfluous knowledge*. TU Delft, Delft University of Technology.

Verhaegen, P. -a., D'hondt, J., Vertommen, J., Dewulf, S., & Duflou, J. R. (2009). Relating properties and functions from patents to TRIZ trends. *CIRP Journal of Manufacturing Science and Technology*, *1*(3), 126–130. doi:10.1016/j.cirpj.2008.09.010

Warkentin, M. E., Sayeed, L., & Hightower, R. (1997). Virtual Teams versus Face-to-Face Teams: An Exploratory Study of a Web-based Conference System. *Decision Sciences*, *28*(4), 975–996. doi:10.1111/j.1540-5915.1997.tb01338.x

Watanabe, C., Zhu, B., & Miyazawa, T. (2001). Hierarchical impacts of the length of technology waves: An analysis of technolabor homeostasis. *Technological Forecasting and Social Change*, *68*(1), 81–104. doi:10.1016/S0040-1625(00)00106-2

White, M. A., & Bruton, G. D. (2007). Management of Technology and Innovation: An Overview. In *The Management of Technology and Innovation: A Strategic Approach* (pp. 6–33). Thomson South-Western.

World Intellectual Property Organization. (2014a). *International Patent Classification: Guide to the IPC*. Geneva, Switzerland. Retrieved from http://www.wipo.int/classifications/ipc/en/

World Intellectual Property Organization. (2014b). *World Intellectual Property Indicators*. Geneva, Switzerland.

Yoon, B. (2010). Strategic visualisation tools for managing technological information. *Technology Analysis & Strategic Management*, *22*(3), 377–397. doi:10.1080/09537321003647438

Yoon, B., & Park, Y. (2004). A text-mining-based patent network: Analytical tool for high-technology trend. *The Journal of High Technology Management Research*, *15*(1), 37–50. doi:10.1016/j.hitech.2003.09.003

Yoon, B.-U., Yoon, C.-B., & Park, Y.-T. (2002). On the development and application of a self-organizing feature map-based patent map. *R&D Management*, *32*(4), 291–300. doi:10.1111/1467-9310.00261

Yoon, J., & Kim, K. (2011). An automated method for identifying TRIZ evolution trends from patents. *Expert Systems with Applications*, *38*(12), 15540–15548. doi:10.1016/j.eswa.2011.06.005

Abbreviations

AJAX	Asynchronous JavaScript and XML
CA	Customer needs
CTR	Carry Through Ratio
CTT	Carry Through Threshold
CU	Categorical Utility
DP	Design Parameter
DSS	Decision Support System
ECLA	European Patent Classification System
EPO	European Patent Office
FR	Functional Requirements
GSS	Group Support System
GUI	Graphical User Interface
IP	Intellectual Property
IPC	International Patent Classification
INID	Internationally agreed Numbers for the Identification of bibliographic Data
NLTK	Natural Language Toolkit
OPS	Open Patent Services
PV	Process Variable
SAO	Subject–Action–Object
SOFM	Self-Organizing Feature Map
SOM	Self-organizing maps
STP	Strategic Technology Planning
TBT	Trace Back Threshold
TRIZ	Theory of Inventive Problem Solving
USPC	US Patent Classification System
USPTO	United States Patent and Trademark Office
WIPO	World Intellectual Property Organization

Figures

Tables

Appendix

I. Expert input discussion summary

Meeting goal: Open discussion on the developed prototype so far

Date: Thursday 27 March 2015, 13:00 – 15:30

Attending: Alejandro Sanz – SKF
Scott Cunningham – Associate professor at BA
Kevin Kruijthoff – EPA student

Key words: Prototype, Conceptual expansion ideas

Key discussion information

- Explained the model implemented as described in Fisher's paper adjusted to patents
- Abstract based overview output different from title based overview
- The output is consistent
  - Testing for different order of entries
- Alejandro curiosity interest; where are my patents located as they are as general and transversal as possible
  - Does it nest more patents like it or will it end up alone and at what level
- Added is showing 50% of all occurrences of this attribute are under this node
  - Doesn't show anything for lowest nodes due to low counts
    - Ranking of percentage contained could be applied as a solution
- Alejandro expects main families to appear from his technical bias
  - Dental, Low resolution, air space, high precision, techniques as laser etc.
- Morphology of the tree
  - Very deep, sometime very broad
  - Are those under the same node technical alternatives
- Does this analysis tool add value?
  - Alejandro: A lot, if you can play with the information, current method shows counts (high counts hills) that's it, but how can I refine it? Placing seeds, manipulate and control, this method could highlight on words concepts etc., it could find complementary assets (GE and their acquisitions), once you have the structure you can manipulate it in the way you want
- Industrially and academically interesting
- Useful in which part of process?
  - Who?
    - In terms of IPC classes -> Expert from EPO
    - Technology -> Expert from industry
    - Companies -> Expert in market
  - Same data can provide multiple insights
    - Priorities set when creating the model -> can thus be made for the specific audience
- Suppose there was a tool like this, a dynamic object, would it facilitate the communication between amongst teams or layers in companies like yours?
  - Alejandro; if I was able to manipulate the data (you need to simplify the interface etc. but the rough material would still be the same), I would buy it. Because my main problem is that I need data representation that is readable by technical experts, translate it into views for the IP people that want to see what are the most dangerous claims (I will send you a table that we have developed) relate to licensing -> tool that can visualize where the risks are, you could create a more intelligent analysis system with this
- Is there consistency with IPC classes
  - if not explain why

- When advising, Solid recommendations could be given, “Because”
- Methodological people will want to know when does this break what are the assumptions how reliable is it how different is it, where the applied oriented will want to know how can I use it? they don’t care if it doesn’t do everything or gets the entire world right but what more can it do, that I can’t do already, and how do I expend to make it more usable
    - Choose
        - Map the field better, validate, limits for the software
        - Enhance the robustness of the process

Ideas

- Seed at the begin of the cluster to create custom trees based on needed insight
- Ask user to decide at moments when the output for the decision where to go is (almost) similar
    - Could be done by comparing categorical utilities
        - Downside are the subjectivity as well as overruling the selection of best option (add, new, split, merge)
- Guidance for where to look
    - Show at the branches what it contains
- Possibly test for a correlation between patent IPC code and depth in hierarchy
- Morphology
    - Reduce levels
        - Validation based on this
            - Ask experts to place certain patents in the tree
    - Random forests
        - Random ordering
            - Are reduced outputs significantly the same
                - Reduced map truncation should be done in a smart way based on a criteria
- Companies
    - Visualize risks
- Patent families
    - Where are they? Are they determining the structure?
        - Families, original is listed in the patent information
- Dynamic growing over time
    - Can you visualize -> Slider
        - Keep track of changes -> tree construction language could be developed
            - Logging of events can already be interesting -> history of splits and merging can already be interesting
- Add legal status to visualization
    - Example add a color related to code (fixed categories: abandon, enforce and applications)
        - Explore trust in different branches, where do people see potential
- Strength indicators could be added
    - Strong medium weak etc.

II. ADDITIVE MANUFACTURING SEARCH STRATEGY

Project ID:2013-136-TI-IN-AdditiveManufacturing				
Period of search: All Years				
Countries Covered: All				
Date of search: 22 Feb 2013. Revised on 18 Dec 2014				
S.No	**Search String used in Thomson Innovation**	**No. of hits**	**Family reduction (1/ family)**	**Remarks**
1	(AIC=(G05B00194099) OR EC=(G05B00194099))	2.925	702 INPADOC Families	Surface or curve machining, making 3D objects, e.g. desktop manufacturing
2	(AIC=(B22F0003105) OR EC=(B22F0003105S))	7.985	2511 INPADOC Families	Sintering by using electric current; Selective sintering, i.e. stereolithography
3	(ACP=((B22F00031055) OR (B29C00670051) OR (B29C00670055) OR (B29C00670059) OR (B29C00670062) OR (B29C00670066) OR (B29C0067007) OR (B29C00670074) OR (B29C00670077) OR (B29C00670081) OR (B29C00670085) OR (B29C00670088) OR (B29C00670092) OR (B29C00670096)))	16.026	3174 INPADOC Families	**Includes:** Selective sintering, i.e. stereolithography Rapid manufacturing and prototyping of 3D objects by additive depositing, agglomerating or laminating of plastics material, e.g. by stereolithography or selective laser sintering
4	(UC=(264/401 OR 700/118 OR 700/119 OR 700/120 OR 264/497))	3.710	1754 INPADOC Families	**Includes:** Plastic + STEREOLITHOGRAPHIC SHAPING FROM LIQUID PRECURSOR Plastic + Using laser sintering of particulate material to build three-dimensional product (e.g., SLS, selective laser sintering, etc.) Three-dimensional product forming Rapid prototyping (e.g., layer-by-layer, material deposition) Stereolithography
5	(EC=(B29C006700R* OR B22F0003105S OR G03F000770H6 OR L22F0003105S OR L29C006700R* OR S03F000770H6 OR L22F0003105S OR L22F0003105S1 OR L22F0003105S1B OR L22F0003105S1D OR L22F0003105S1F)) ;	12.478	2345 INPADOC Families	Includes: Rapid manufacturing and prototyping of 3D objects by additive depositing, agglomerating or laminating of plastics material, e.g. by stereolithography or selective laser sintering Selective sintering, i.e. stereolithography Stereolithography, 3D printing, rapid prototyping
6	(MC=((X25-A08))) ;	1406	672 INPADOC Families	RAPID PROTOTYPING AND SOLID FREEFORM FABRICATION
7	(FTC=(4F213WL*)) ;	4.527	2563INPADOC Families	TECHNIQUE FOR THREE-DIMENSIONAL MOULDING
8	(IC=(B22F0003105 OR B29C006704 OR B29C006700) AND ALL=(((3D ADJ printing) OR (Three ADJ Dimension* ADJ Printing) OR (Rapid ADJ Prototyping) OR (Rapid-Prototyping) OR (3D OR (Three ADJ Dimension*)) OR (DIRECT ADJ METAL ADJ FABRICATION) OR (three-dimension*) OR (3-dimension*) OR (Additive ADJ (Manufacturing OR Fabrication*)) OR (Stereolithograph*) OR (Stereo ADJ lithograph*) OR (Direct ADJ Metal ADJ Laser ADJ Sintering) OR (solid ADJ free ADJ form ADJ fabricat*) OR (Optical ADJ Fabrication)))) ;	19.178	6067 INPADOC Families	IPC codes for Metal Sintering, plastic sintering and keywords 3069

9	(CTB=((Fused ADJ deposition ADJ modeling) OR (Direct ADJ metal ADJ laser ADJ sintering) OR (Electron ADJ beam ADJ melting) OR (Rapid ADJ Prototyping) OR (Rapid-Prototyping) OR (Selective ADJ heat ADJ sintering) OR (Selective ADJ laser ADJ sintering) OR (solid ADJ free ADJ form ADJ fabrication) OR (Laminated ADJ object ADJ manufacturing) OR (Stereo ADJ lithograph*) OR (Stereolithography) OR ((3d OR (Three ADJ Dimension*)) ADJ printing) OR (Additive ADJ (Manufacturing OR Fabrication*)))) ;	19.841	7662 INPADOC Families	
10	(AIC=(B29 OR B22F) AND ALL=(((Fused ADJ deposition ADJ modeling) OR (Direct ADJ metal ADJ laser ADJ sintering) OR (Electron ADJ Beam ADJ Sintering) OR (Laser ADJ Net ADJ Shape ADJ Manufacturing) OR (Laser ADJ Engineered ADJ Net ADJ Shaping) OR (Direct ADJ Metal ADJ Deposition) OR (Electron ADJ beam ADJ melting) OR (Rapid ADJ Prototyping) OR (Rapid-Prototyping) OR (Selective ADJ heat ADJ sintering) OR (Selective ADJ laser ADJ sintering) OR (solid ADJ free ADJ form ADJ fabrication) OR (Laminated ADJ object ADJ manufacturing) OR (Stereo ADJ lithograph*) OR (Stereolithography) OR ((3d OR (Three ADJ Dimension*)) ADJ printing) OR (Additive ADJ (Manufacturing OR Fabrication*)))))	16.991	5298 INPADOC Families	
11	CTB=((Electron ADJ Beam ADJ Sintering) OR (Laser ADJ Engineered ADJ Net ADJ Shaping) OR (Laser ADJ Net ADJ Shape ADJ Manufacturing) OR (Direct ADJ Metal ADJ Deposition)) ;	301	122 INPADOC Families	
12	(AIC=(B23K00150086 AND B22F299800)) ;	34	6 INPADOC Families	
13	CTB=(Additive ADJ (Manufacturing OR Fabrication*))	1.869	943 INPADOC Families	
14	CTB=((Stereolithograph*) OR (Stereo ADJ lithograph*))	7.873	2208 INPADOC Families	
15	CTB=((Direct ADJ Metal ADJ Laser ADJ Sinter*))	179	77 INPADOC Families	
16	CTB=(solid ADJ free ADJ form ADJ fabricat*)	476	121 INPADOC Families	
17	CTB=(Optical ADJ Fabrication)	122	49 INPADOC Families	
18	CTB=(Fus* ADJ deposit* ADJ model*)	1045	341 INPADOC Families	
19	CTB=((Electron ADJ beam ADJ melting))	1594	790 INPADOC Families	
20	CTB=((Laminated ADJ object ADJ manufactur*))	285	115 INPADOC Families	
21	CTB=(Electron ADJ Beam ADJ Sinter*)	48	19 INPADOC Families	
22	CTB=((Laser ADJ Engin* ADJ Net ADJ Shaping))	98	48 INPADOC Families	
23	CTB=((Laser ADJ Net ADJ Shape ADJ Manufacturing))	33	7 INPADOC Families	
24	CTB=((Direct ADJ Metal ADJ Deposit*))	161	73 INPADOC Families	
25	1 OR 2 OR 3 OR 23 OR 24	46.488	9765 INPADOC Families	

Date of search: 22 Feb 2013. Revised on 18 Dec 2014

S.No	Search String used in Thomson Innovation (INSPEC)	No. of hits	Remarks
1	TI=(Additive ADJ (Manufacturing OR Fabrication));	388	
2	ALL=((3D ADJ printing) OR (Three ADJ Dimensional ADJ Printing))	2.184	
3	TI=((DIRECT ADJ METAL ADJ FABRICATION) OR (Stereolithograph*) OR (Stereo ADJ lithograph*) OR (Direct ADJ Metal ADJ Laser ADJ Sintering) OR (solid ADJ free ADJ form ADJ fabricat*));	616	
4	TI=((Rapid ADJ Prototyping) OR (Rapid-Prototyping));	2.856	
5	1 OR 2 OR 3 OR 4	5.852	

S.No	Search String used in Thomson Innovation (WOS, Conferences & Current Cone)	No. of hits	Remarks
1	TI=(Additive ADJ (Manufacturing OR Fabrication));	560	

2	ALL=((3D ADJ printing) OR (Three ADJ Dimensional ADJ Printing))	2261	
3	TI=((DIRECT ADJ METAL ADJ FABRICATION) OR (Stereolithograph*) OR (Stereo ADJ lithograph*) OR (Direct ADJ Metal ADJ Laser ADJ Sintering) OR (solid ADJ free ADJ form ADJ fabricat*));	1070	
4	TI=((Rapid ADJ Prototyping) OR (Rapid-Prototyping));	2811	
5	1 OR 2 OR 3 OR 4	6526	
S.No	**Search String used in Thomson Innovation (Compendex)**	**No. of hits**	**Remarks**
1	1003 articles found in Compendex for 1969-2015: ((((Additive ONEAR Manufacturing)) WN TI) OR (((Additive ONEAR Fabrication)) WN TI))	1003	
2	200 articles found in Compendex for 1969-2015: (((3D ONEAR printing) WN All fields) AND ((Three ONEAR Dimensional ONEAR Printing) WN All fields))	200	
3	947 articles found in Compendex for 1969-2015: ((((((Direct ONEAR Metal ONEAR Laser ONEAR Sintering) WN TI) OR ((solid ONEAR free ONEAR form ONEAR fabrication) WN TI)) OR ((DIRECT ONEAR METAL ONEAR FABRICATION) WN TI)) OR ((Stereo ONEAR lithography) WN TI)) OR ((Stereolithography) WN TI))	947	
4	3951 articles found in Compendex for 1969-2015: ((Rapid ONEAR Prototyping) WN TI)	3951	
5	1 OR 2 OR 3 OR 4	6039	

III. TEXT-MINING OPERATION AND EXPORTING VERIFICATION TEST

The verification test for text-mining software is done based on the following three entries as representing patent abstracts:

1. The tin can is picked up by a mechanical arm and transported onto a conveyor belt.
2. The operation of the conveyor belt is done via a computer interface which controls the mechanical arm.
3. Based on the weight of a tin can the computer selects which belt path to follow.

With the settings:

- number_of_attributes = 15
- include_title = False
- include_abstract = True
- include_nov = False
- include_nouns = True
- include_verbs = True
- include_adjectives = True
- include_adverbs = True

It can be seen in Listing III that only the selected word-types are represented as attributes and words such as 'the' and 'a' are excluded as they should. In this case the in the top 15 words, there are also words that only occur once of which there are more, in this situations they are picked at random to fill on the attribute list.

Listing III – Text-mining example output JSON structure


```
{
  "Patents": [
    {
      "patent_title": "Test title 1",
      "company_id": "Test company 1",
      "id": 0,
      "Attributes": [
        {
          "mechanical": 1,
          "operation": 0,
          "interface": 0,
          "follow": 0,
          "weight": 0,
          "computer": 0,
          "base": 0,
          "belt": 1,
          "conveyor": 1,
          "arm": 1,
          "tin": 1,
          "done": 0,
          "path": 0
        }
      ],
      "input_id": "TEST1"
    },
    {
      "patent_title": "Test title 2",
      "company_id": "Test company 2",
      "id": 1,
      "Attributes": [
        {
          "mechanical": 1,
          "operation": 1,
          "interface": 1,
          "follow": 0,
          "weight": 0,
          "computer": 1,
          "base": 0,
          "belt": 1,
          "conveyor": 1,
```

```
          "arm": 1,
          "tin": 0,
          "done": 1,
          "path": 0
        }
      ],
      "input_id": "TEST2"
    },
    {
      "patent_title": "Test title 3",
      "company_id": "Test company 3",
      "id": 2,
      "Attributes": [
        {
          "mechanical": 0,
          "operation": 0,
          "interface": 0,
          "follow": 1,
          "weight": 1,
          "computer": 1,
          "base": 1,
          "belt": 1,
          "conveyor": 0,
          "arm": 0,
          "tin": 1,
          "done": 0,
          "path": 1
        }
      ],
      "input_id": "TEST 3"
    }
  ],
  "Metadata": [
    {
      "includetitle": false,
      "include_verbs": true,
      "includeabstract": true,
      "include_nouns": true,
      "include_adverbs": true,
      "date": "1-7-2015",
      "author": "Kevin Kruijthoff",
      "time": "15:14",
      "inputfile": "data/raw/raw_3_entries_test.txt",
      "numberofattributes": 15,
      "include_adjectives": true
    }
  ]
}
```

IV. COBWEB EXPORT JSON STRUCTURAL EXAMPLE

Listing IV – Conceptual Clustering output JSON structure example


```
{
  "children": [
    {
      "children": [
        {
          "company": "Albany International Corp.,Albany,NY,US",
          "patent": 131,
          "patent_original_id": "US8822009B2",
          "patent_title": "Industrial fabric, and method of making thereof"
        },
        {
          "company": "Hexagon Technology Center GmbH,Heerbrugg,CH",
          "patent": 48,
          "patent_original_id": "US8875655B2",
          "patent_title": "Graphical application system"
        }
      ],
      "description": "patents: 2 / spatter(100.0% / P 0.5), spattering(100.0% / P 0.5), 
drum(50.0% / P 0.5), floor(50.0% / P 0.5), handheld(50.0% / P 0.5), dynamically(50.0% / P 
0.5), sand(33.3% / P 0.5), movable(33.3% / P 0.5), equipment(33.3% / P 0.5), nozzle(33.3% / P 
1.0), ",
      "name": "<tspan class='cat3'>nozzle</tspan><tspan class='cat4'> - </tspan><tspan 
class='cat3'>additional</tspan><tspan class='cat4'> - </tspan><tspan 
class='cat4'>sheet</tspan>"
    },
    {
      "company": "Wake Forest University Health Sciences,Winston-Salem,NC,US",
      "patent": 110,
      "patent_original_id": "US20140350499A1",
      "patent_title": "DEVICES AND METHODS FOR TREATING SPINAL CORD TISSUE"
    },
    {
      "company": "National Oilwell Varco L.P.,Houston,TX,US",
      "patent": 141,
      "patent_original_id": "US8814968B2",
      "patent_title": "Thermally conductive sand mould shell for manufacturing a matrix bit"
    }
  ],
  "description": "Root node"
}
```

V. CONCEPTUAL CLUSTERING SOFTWARE DOCUMENTATION

1. INTRODUCTION

This documentation is based on the IEEE – std 830-1993 and complying with its template of SRS section 3 organized by object.

1.1. PURPOSE

This document is created as an addition to the thesis document and acts as a guide to the software created for the application of the conceptual clustering implementation of the COBWEB algorithm. It is aimed to allow the conducting of further research by extending the realized software implementation. The target audience of this document are software developers as well as researchers interested in patent analysis applications.

1.2. SCOPE

The scope of this document is the Python implementation of the conceptual clustering software. In which the processing is done by the conceptual clustering software, which is based on the COBWEB algorithm (Fisher, 1987), and aims to transform the individual attributized patents into a hierarchical inter-relational structure. The interface between the input for the conceptual clustering software, as well as between the conceptual clustering software and the visualization is done via JSON format files.

1.3. REFERENCES

Fisher, D. H. (1987). Knowledge acquisition via incremental conceptual clustering. *Machine Learning*, *2*(2), 139–172.

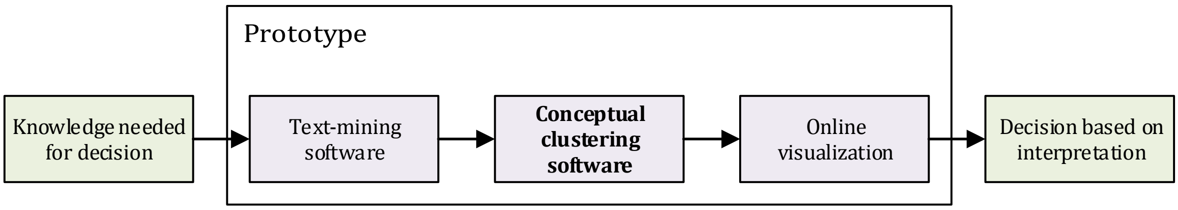
1.4. OVERVIEW

The software documentation is structured by first describing the overall use of the software and its goals in Section 2, followed by the specific setup of the conceptual clustering software in Section 3.

2. OVERALL DESCRIPTION

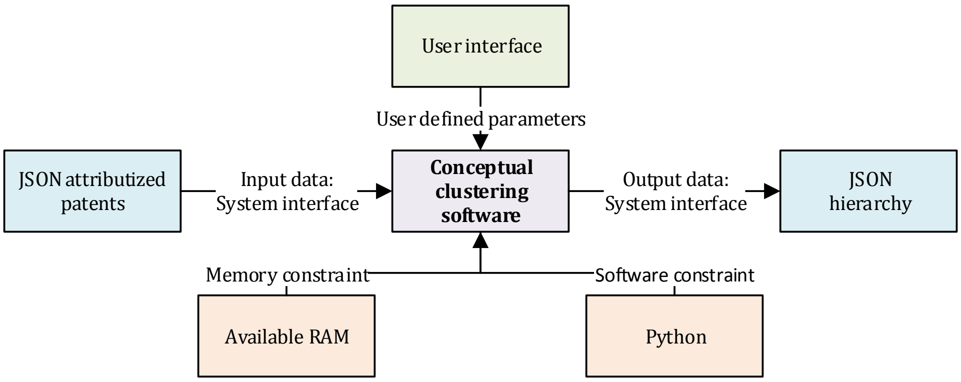
2.1. PRODUCT PERSPECTIVE

The developed software aims to aid users in the strategic technology planning process. It has been created for technical evaluative purposes needed for the thesis, but can be extended to be used in organizational use. Software-Figure 1, shows a block diagram of the positioning of the conceptual clustering within the context of the larger system.



Software-Figure 1

Software-Figure 2, shows the software's interfaces. The software input system interface is with a JSON file which contains the individual patent information as well as the assigned attributes. These are Boolean type attributes, meaning they are 1 when the patent contains the attribute, and 0 otherwise. The output system interface is also with a JSON file which contains the hierarchical structure created by the software.



Software-Figure 2

2.2. PRODUCT FUNCTIONS

The main functions of the software is to determine the relation between the entered patents based on the conceptual clustering algorithm. Secondly the software prepares the resulting hierarchy for interpretation in the visualization step. This includes the extraction of the hierarchy as well as the determination of which attributes to show at the individual nodes to assist the user in interpreting the structure. It is developed in light of the thesis to be used by its creator to evaluate the potential of conceptual clustering applied to patents and its potential enhancement of the patent analysis for Strategic Technology Planning.

2.3. USER CHARACTERISTICS

There are multiple potential users for the developed software. Which include scholars performing research into the visualization of the patent fields and technology scouts and R&D managers. All users are expected to be well educated as well as having (mediocre) experience and knowledge of programming, as well as of the patent analysis field.

2.4. CONSTRAINTS

A	Regulatory policies	Non applicable
B	Hardware limitations	Applicable amount of RAM is required when operating the software due to the attributes stored in the nodes as well as in the patent attributes.
C	Interfaces to other applications	The input as well as the output make use of JSON, therefore the JSON library must be installed in Python.
D	Parallel operations	Impossible due to the hill climbing search applied
E	Audit functions	Non applicable
F	Control functions	The operators for control have been defined in the user interface. These control which operators are allowed to be used.
G	Higher order language requirements	Non applicable
H	Signal handshake protocol	Non applicable
I	Reliability requirements	Non applicable
J	Criticality of the application	Non applicable
K	Safety and security considerations	Non applicable

2.5. ASSUMPTIONS AND DEPENDENCIES

The assumption is made that the user has a computer with a windows operating system, which has a python compiler installed and enough processing capacity. The processing capacity needed is dependent on the number of attributes used to attributize the individual patents and the number of patents entered. Furthermore, for the visualization a webserver is required for which upload and overwriting access are required.

2.6. APPORTIONING OF REQUIREMENTS

Requirements that may be delayed till later versions of the software might include the addition of the 5th operator as described by Fisher (Fisher, 1987), the storing/logging of changes over iterations, as well as the determination of a similarity indication among leafs. For explanations on the potential additions please refer to the thesis document.

3. SPECIFIC REQUIREMENTS

3.1. EXTERNAL INTERFACE REQUIREMENTS

3.1.1.	User interfaces	Non applicable
3.1.2.	Hardware interfaces	Non applicable
3.1.3.	Software interfaces	Non applicable
3.1.4.	Communications interfaces	Non applicable

3.2. CLASSES/OBJECTS

3.2.1. PATENT

3.2.1.1. PATENT ATTRIBUTES

Attribute	Direct / inherited	Description
patent_id_counter	Inherited	Incremental counter to assign unique IDs at creation of new patent objects
patents_total_counts	Inherited	The total combined attribute counts of the patent objects
patent_unique_id	Direct	The ID that is uniquely assigned to the object when the patent object is created
patent_input_id	Direct	The ID that has been read from the input
patent_input_title	Direct	The title that has been read from the input and has been retrieved from the patent
patent_input_patent_original_id	Direct	The ID that has been read from the input and has been retrieved from the patent
patent_input_company_id	Direct	The company ID that has been read from the input and has been retrieved from the patent
patent_node_id	Direct	The node ID to which this patent object has to be linked
patent_attributes	Direct	The patent specific counts of attributes

3.2.1.2. PATENT FUNCTIONS

Function	Description
__init__	Creates a new patent object
new_id	Returns a new unique id to assign to the object

3.2.2. NODE

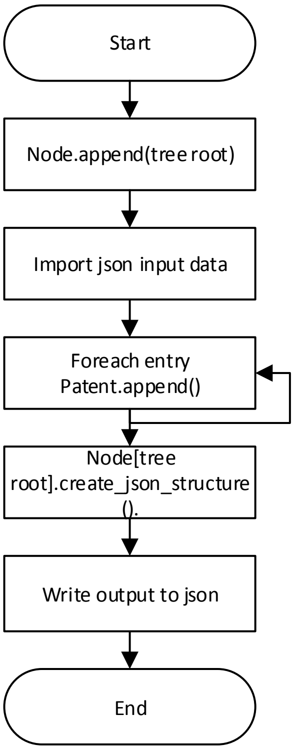
3.2.2.1. NODE ATTRIBUTES

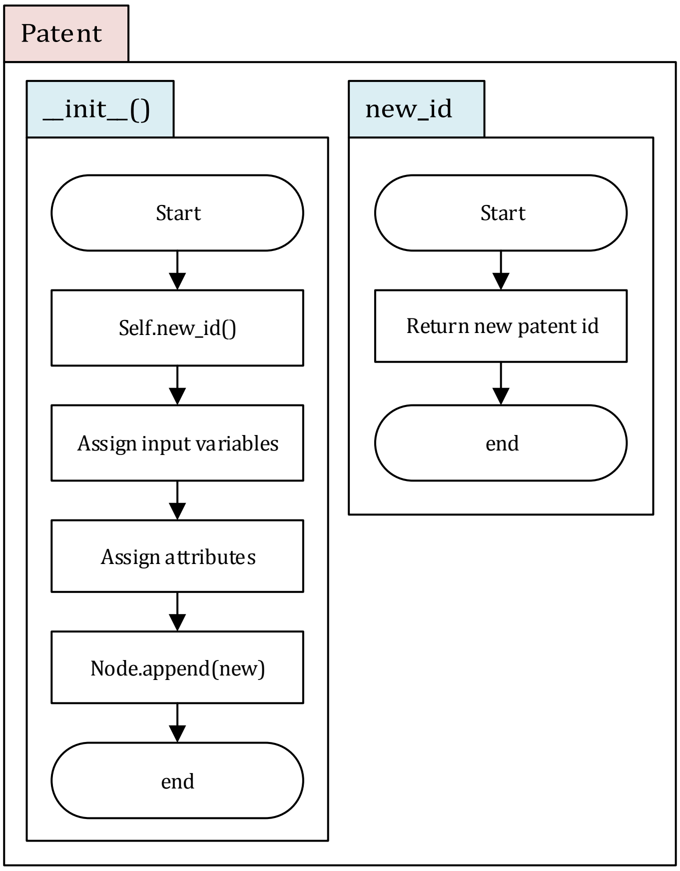
Attribute	Direct / inherited	Description
node_id_counter	Inherited	Incremental counter to assign unique IDs at creation of new node objects
tree_root	Inherited	The object that is the tree root node
patents_number	Inherited	The number of patents that are in the tree
patents_total_counts	Inherited	The total combined attribute counts of the objects that are in the tree
node_unique_id	Direct	The ID that is uniquely assigned to the object when the node object is created
node_root	Direct	The node object that is the parent of this specific node object within the tree hierarchy
node_children	Direct	List containing the objects that are the children of this specific node object within the tree hierarchy
node_attributes	Direct	Dictionary containing the total combined attribute counts of this node object or that of all the node objects that are under this specific node object within the tree hierarchy
node_patents	Direct	The patent object that is linked to this specific node
node_patents_number	Direct	The number of patent objects that are linked to this node object or to node objects that are under this specific node object within the tree hierarchy

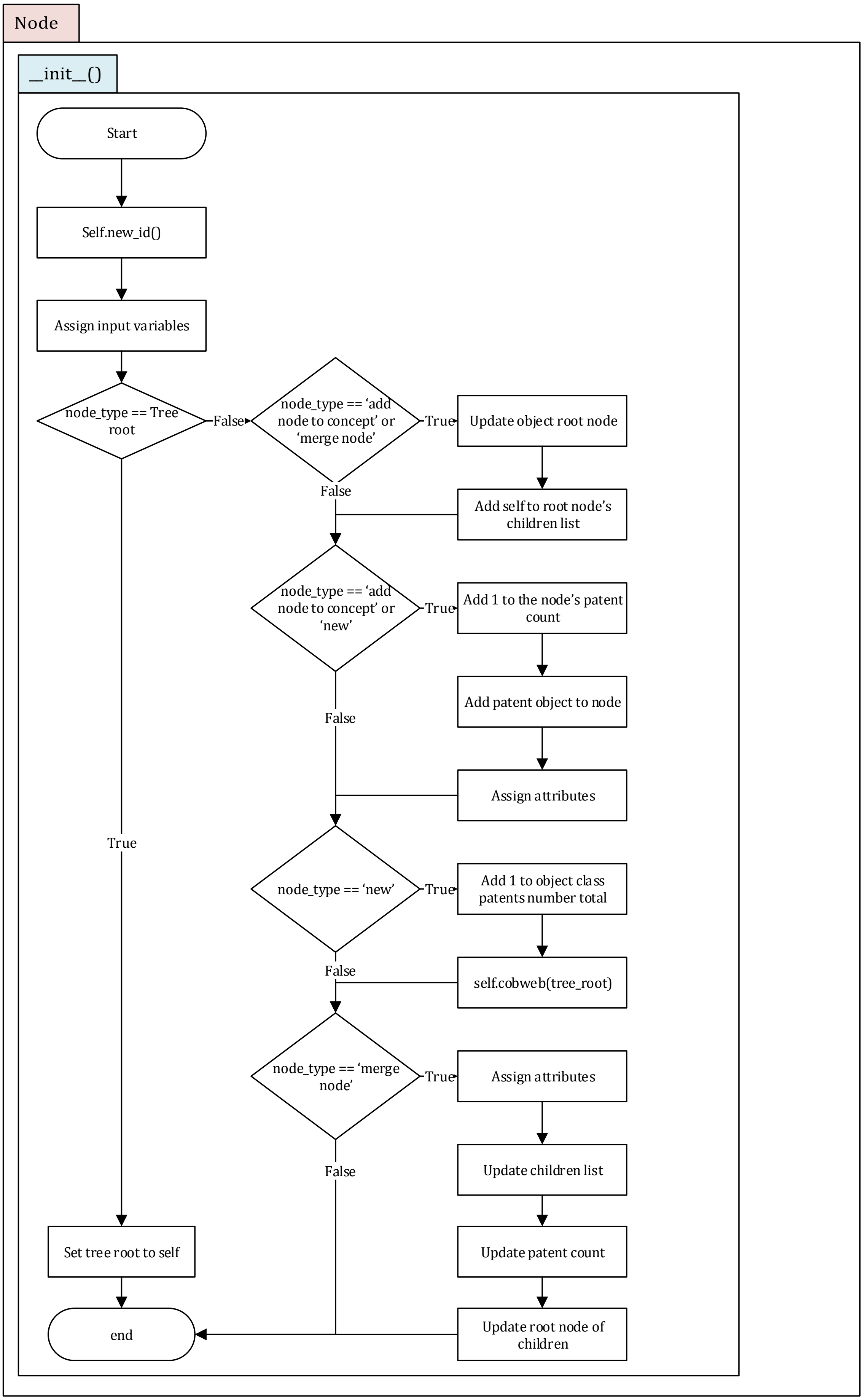
3.2.2.2. NODE FUNCTIONS

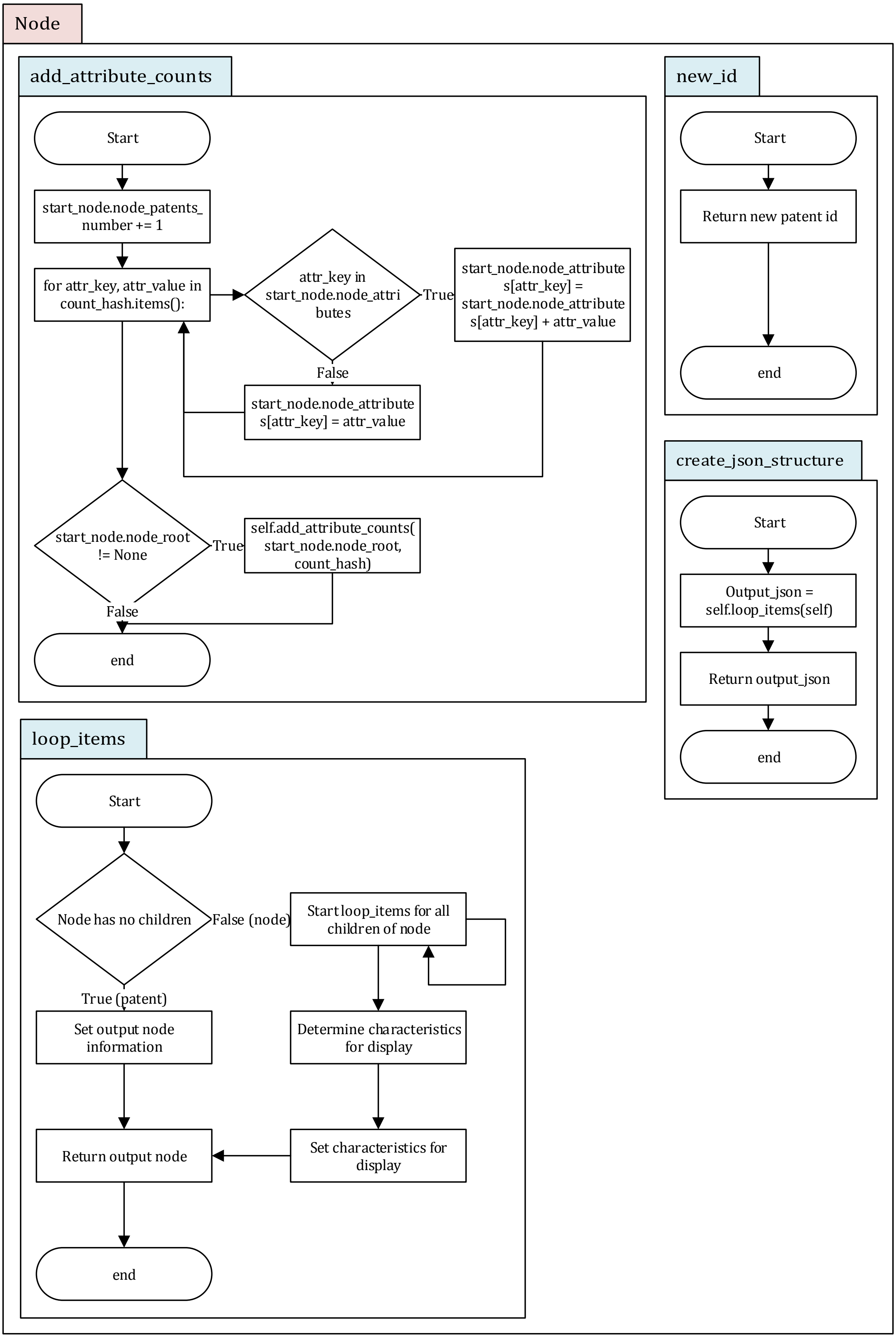
Function	Description
__init__	Creates a new node object and links the relevant patent objects to it
new_id	Returns a new unique id to assign to the object
cobweb	Determines the position of the newly entered patent object
add_attribute_counts	Adjust the attributes of the node based on the newly added patent
create_json_structure	Start the extraction of the resulting structure
loop_items	Walk through the structure from the root down the branches and create the hierarchical structure, the node descriptions and patent descriptions

3.3. SOFTWARE OPERATION & OBJECTS FUNCTION FLOWS

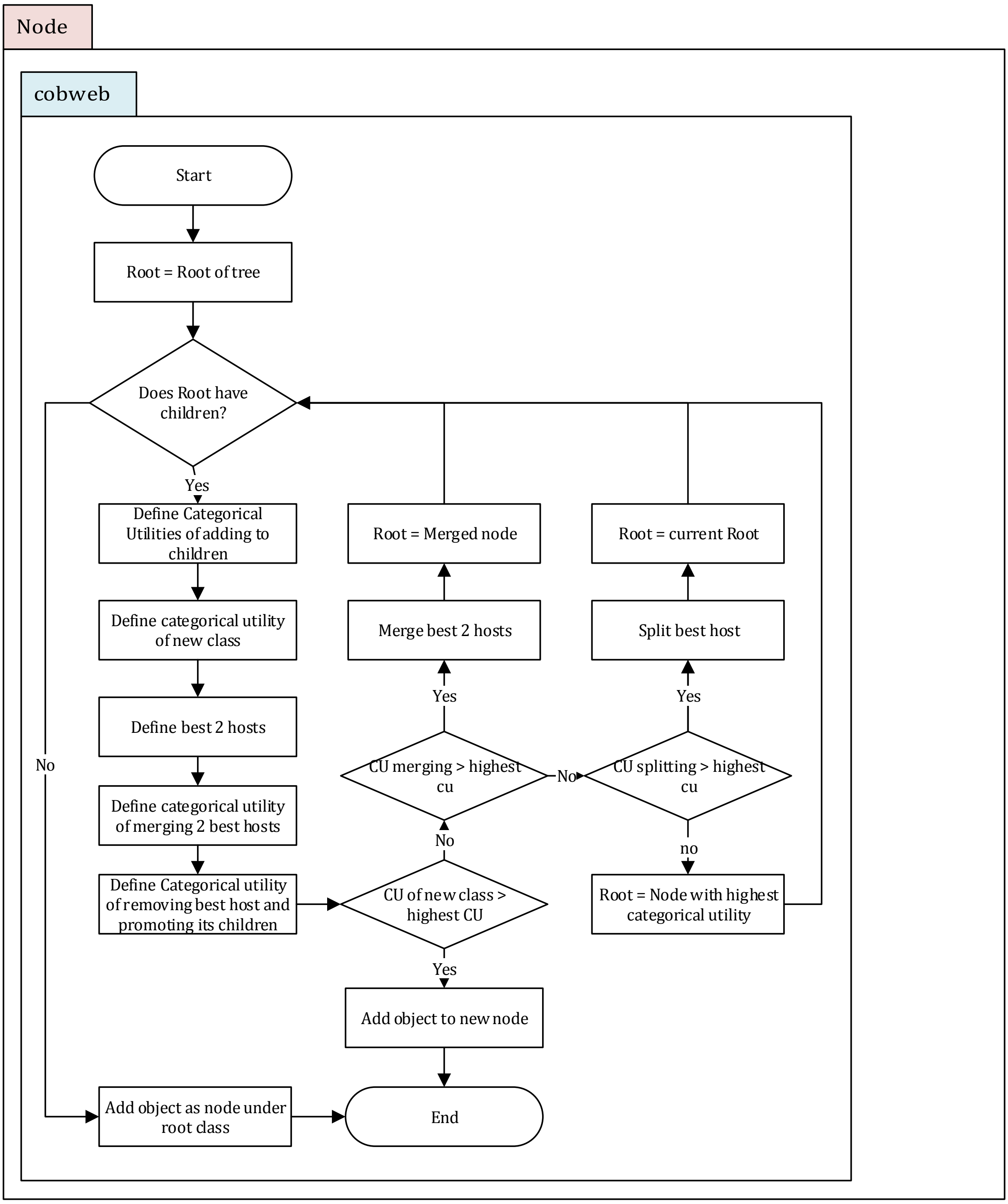





Node
__init__()
Start
Self.new_id()
Assign input variables
node_type == Tree root
False
node_type == 'add node to concept' or 'merge node'
True
Update object root node
Add self to root node's children list
False
node_type == 'add node to concept' or 'new'
True
Add 1 to the node's patent count
Add patent object to node
Assign attributes
False
True
node_type == 'new'
True
Add 1 to object class patents number total
self.cobweb(tree_root)
False
node_type == 'merge node'
True
Assign attributes
Update children list
Update patent count
Update root node of children
False
Set tree root to self
end



Node
add_attribute_counts
Start
start_node.node_patents_number += 1
for attr_key, attr_value in count_hash.items():
attr_key in start_node.node_attributes
True
start_node.node_attributes[attr_key] = start_node.node_attributes[attr_key] + attr_value
False
start_node.node_attributes[attr_key] = attr_value
start_node.node_root != None
True
self.add_attribute_counts(start_node.node_root, count_hash)
False
end
new_id
Start
Return new patent id
end
create_json_structure
Start
Output_json = self.loop_items(self)
Return output_json
end
loop_items
Start
Node has no children
False (node)
Start loop_items for all children of node
True (patent)
Set output node information
Determine characteristics for display
Set characteristics for display
Return output node
end


Node
cobweb
Start
Root = Root of tree
Does Root have children?
Yes
No
Define Categorical Utilities of adding to children
Define categorical utility of new class
Define best 2 hosts
Define categorical utility of merging 2 best hosts
Define Categorical utility of removing best host and promoting its children
CU of new class > highest CU
Yes
No
Add object to new node
Add object as node under root class
End
CU merging > highest cu
Yes
No
Merge best 2 hosts
Root = Merged node
CU splitting > highest cu
Yes
no
Split best host
Root = current Root
Root = Node with highest categorical utility